

# 目 录

第一章 引言 .....	1
§ 1 解常微分方程为什么要研究数值方法 .....	1
§ 2 建立数值方法的基本思想与途径 .....	2
§ 3 一些基本概念 .....	5
第二章 常用的单步法 .....	7
§ 1 Euler 方法 .....	7
§ 2 Runge-Kutta 方法 .....	15
§ 3 Richardson 外推法 .....	27
第三章 线性多步法 .....	35
§ 1 Adams 显式公式 .....	35
§ 2 Adams 隐式公式 .....	39
§ 3 初始出发值的计算 .....	41
§ 4 Adams 公式的截断误差 .....	44
§ 5 隐式公式的迭代解法 .....	45
第四章 预测-校正法 .....	51
§ 1 最简单的预测-校正法 .....	51
§ 2 Milne 方法 .....	55
§ 3 Hamming 方法 .....	59
第五章 常微分方程组及高阶微分方程的数值解法 .....	62
§ 1 常微分方程组简介 .....	62
§ 2 Runge-Kutta 方法 .....	64
§ 3 Hamming 方法 .....	76
§ 4 不显含一阶导数的二阶方程的特殊计算方法 .....	81

<b>第六章</b>	<b>数值方法的相容性、收敛性和稳定性</b>	87
§ 1	单步法的相容性和收敛性	87
§ 2	多步法的相容性和收敛性	90
§ 3	数值稳定性问题	92
§ 4	绝对稳定性	95
<b>第七章</b>	<b>坏条件方程组简介</b>	105
§ 1	什么是坏条件方程组	105
§ 2	适合于不同情况的解坏条件方程的线性方法	107
§ 3	非线性方法	118
§ 4	关于阶数、步长和方法的选择	122
<b>第八章</b>	<b>边值问题的数值解法</b>	128
§ 1	解线性边值问题的差分方法	129
§ 2	样条函数简介及其在两点边值问题上的应用	139
§ 3	试射法	143
§ 4	适合于非线性方程的差分方法	148
<b>附录 I</b>	<b>差分方程简介</b>	154
§ 1	一般差分方程	154
§ 2	线性差分方程	154
§ 3	线性常系数差分方程	156
<b>附录 II</b>	<b>第六章 定理 2 的证明</b>	158

# 第一章 引言

## § 1 解常微分方程为什么要研究数值方法

在《常微分方程》一书中，我们已经对一些典型的微分方程（例如线性方程、某些特殊的一阶非线性方程等）介绍了一些求解析解的基本方法。有了解析解，就可以（至少从理论上说）根据初值问题或边值问题的条件把其中的任意常数完全确定下来。然而，在生产实际和科学研究中所遇到的微分方程往往很复杂，在很多情况下都不可能给出解的解析表达式。有时候即使能求出封闭形式的解，也往往因计算量太大而不实用。例如，计算收敛缓慢的幂级数解，为了满足规定的精确度，就得计算很多的项。又如，容易求出初值问题

$$\begin{cases} y' = 1 - 2xy, \\ y(0) = 0 \end{cases}$$

的解

$$y = e^{-x^2} \int_0^x e^{t^2} dt,$$

但是要计算它的值，还需应用数值积分的方法，而如果要求对于许多  $x$  值计算解  $y(x)$  的值，那么工作量就十分庞大了。就连简单的初值问题

$$\begin{cases} y' = y, \\ y(0) = 1, \end{cases}$$

其解为  $y = e^x$ ，虽有表可查，但对表上所没有给出的  $e^x$  值也还得利用插值方法来计算；如果对大量的  $x$  值要进行这种插值的话，其工作量将是很可观的。此外，线性常系数微分方程看

来够简单的了,只要求出特征方程的根,就可以得到通解.可是,高次代数方程求根也并不容易,按照这种途径来求解析解那就更为困难了.

根据以上的讨论,可以看出,用求解析解的方法来计算微分方程的数值解往往是不适宜的,甚至是很难办到的.

实际上,对于解微分方程问题,一般只要求得到解在若干个点上的近似值或者解的便于计算的近似表达式(只要满足规定的精度就行了).所以,研究解微分方程的数值方法就显得十分必要了.

## §2 建立数值方法的基本思想与途径

### 2.1. 离散化 我们以一阶微分方程的初值问题

$$\begin{cases} y' = f(t, y), & a \leq t \leq b, \\ y(a) = \eta \end{cases} \quad (1)$$

为例来说明建立数值方法的基本思想.初值问题(1)的解 $y(t)$ 是区间 $[a, b]$ 上的连续变量 $t$ 的函数,因而问题(1)实际上是一个连续性的问题.求这个初值问题的数值解,就是要求在区间 $[a, b]$ 上的若干个离散的点处,例如

$$a \leq t_0 < t_1 < \cdots < t_n \leq b,$$

计算出解 $y(t)$ 的近似值

$$y_1, y_2, \cdots, y_n.$$

一般常取 $t_0, t_1, \cdots, t_n$ 为等距离的,即

$$t_1 - t_0 = t_2 - t_1 = \cdots = t_n - t_{n-1} = h \text{ (设),}$$

或

$$t_i = a + ih, \quad i = 0, 1, \cdots, n,$$

我们称 $h$ 为步长.建立数值方法的第一步工作就是把连续性问题(1)通过一定的方法化为在给定的 $n+1$ 个点上的近似的

差分方程的初值问题. 我们称这个过程为离散化. 关于离散化的方法, 将在以下介绍.

**2.2. 化导数为差商的方法** 在点  $t_i$  处的导数  $y'(t_i)$  可以近似地表示成差商

$$y'(t_i) \approx \frac{y_{i+1} - y_i}{h},$$

从而就把微分方程初值问题(1)化为差分方程的初值问题

$$\begin{cases} \frac{y_{i+1} - y_i}{h} = f(t_i, y_i), & i=0, 1, \dots, \\ y_0 = \eta, \end{cases} \quad (2)$$

其中,  $y_i$  表示解  $y(t)$  在点  $t_i$  处的近似值, 即

$$y_i \approx y(t_i).$$

应当指出, 用差商来近似地表示导数, 方法不是唯一的, 这里所用的是前差, 也可以用另外的方法得出具有更高精确度的差商近似.

用差商代替导数是微分方程问题离散化的一种基本方法.

**2.3. Taylor 展式法** 在一个点 (例如  $t_i$ ) 的附近,  $y(t)$  的同次数的近似多项式中以 Taylor 多项式

$$y(t_i + h) \approx y(t_i) + hy'(t_i) + \dots + \frac{h^p}{p!} y^{(p)}(t_i) \quad (3)$$

为最好. 其中,  $p$  为一正整数. 通过微分方程

$$y' = f(t, y),$$

可以逐次把各阶导数  $y', y'', \dots$  在  $t_i$  处的值表示出来, 例如,

$$y'(t_i) = f(t_i, y(t_i))$$

$$y''(t_i) = f'_t(t_i, y(t_i)) + f'_y(t_i, y(t_i))y'(t_i)$$

.....

从而可知, 各阶导数的近似值近似地满足下列关系:

$$\begin{cases} y'_i = f(t_i, y_i) \\ y''_i = f'_t(t_i, y_i) + f'_y(t_i, y_i)y'_i \\ \dots\dots\dots, \end{cases} \quad (4)$$

其中,  $y_i$  表示  $y(t_i)$  的近似值, 而  $y'_i, y''_i, \dots$  分别表示各阶导数  $y'(t_i), y''(t_i), \dots$  的近似值, 因此, 可以把微分方程初值问题 (1) 利用 Taylor 多项式 (3) 化为差分方程的初值问题:

$$\begin{cases} y_{i+1} = y_i + hy'_i + \frac{h^2}{2} y''_i + \dots + \frac{h^p}{p!} y^{(p)}_i, \quad i=0, 1, \dots \\ y_0 = \eta. \end{cases} \quad (5)$$

若取  $p=1$ , 则 (5) 化为

$$\begin{cases} y_{i+1} = y_i + hy'_i \approx y_i + hf_i, \\ y_0 = \eta, \end{cases}$$

其中  $f_i$  表示函数  $f(t, y(t))$  在  $t_i$  点的近似值  $f(t_i, y_i)$ . 这就是前一小节中的 (2). 应当指出, 用 Taylor 多项式 (3) 来建立差分方程, 当  $p$  稍大时, 差分方程就非常复杂, 因而不宜于直接应用.

#### 2.4. 数值积分方法 把微分方程

$$y' = f(t, y)$$

在区间  $[t_i, t_{i+1}]$  上求积分便得

$$y(t_{i+1}) - y(t_i) = \int_{t_i}^{t_{i+1}} f(t, y(t)) dt, \quad (6)$$

于是, 初值问题 (1) 便可以近似地化为:

$$\begin{cases} y_{i+1} = y_i + \int_{t_i}^{t_{i+1}} f(t, y(t)) dt, \\ y_0 = \eta. \end{cases} \quad (7)$$

关于 (6) 式右端的积分, 可以用数值积分方法计算其近似值.

例如, 在  $[t_i, t_{i+1}]$  上利用矩形公式, 则有

$$\int_{t_i}^{t_{i+1}} f(t, y(t)) dt \approx hf_i,$$

这时(7)就化为

$$\begin{cases} y_{i+1} = y_i + hf_i, \\ y_0 = \eta. \end{cases}$$

这又得到了(2). 若使用较精确的梯形公式,

$$\int_{t_i}^{t_{i+1}} f(t, y(t)) dt \approx \frac{h}{2} (f_i + f_{i+1}),$$

这时, (7)化为

$$\begin{cases} y_{i+1} = y_i + \frac{h}{2} (f_i + f_{i+1}), \\ y_0 = \eta. \end{cases} \quad (8)$$

在矩形求积公式中若取

$$\int_{t_i}^{t_{i+1}} f(t, y(t)) dt \approx hf_{i+1},$$

则(1)近似地化为

$$\begin{cases} y_{i+1} = y_i + hf(t_{i+1}, y_{i+1}), \\ y_0 = \eta. \end{cases} \quad (9)$$

我们称(2)为 Euler 公式, 而称(9)为后退 Euler 公式.

### § 3 一些基本概念

我们以 Euler 公式

$$\begin{cases} y_{i+1} = y_i + hf_i, \\ y_0 = \eta \end{cases} \quad (10)$$

为例, 来介绍数值方法中的一些基本概念. 在 Euler 公式中, 若已知  $y_0$ , 便可以由递推公式(2)来求出  $y_1$ , 然后由  $y_1$  求出  $y_2$ , 依次类推, 由  $y_n$  可求出  $y_{n+1}$ ; 为了求后一点  $t_{n+1}$  上的数值  $y_{n+1}$ , 只要知道前一点  $t_n$  上的数值  $y_n$  就够了, 这种方法称为单步法, 显然它是一个自开始的公式, 即由微分方程给出的初始值  $y(0) = \eta = y_0$  作为它的出发值, 就可以按所给公式计算出

后面的函数值  $y_1, y_2, \dots$ , 而不需其他的信息.

Euler 公式是一个显式公式, 因为所要求的量  $y_{n+1}$  可以用  $y_n$  和由  $y_n$  求得的  $f_n = f(t_n, y_n)$  来表示. 相反, 后退 Euler 公式则是一个隐式公式, 因为所要求的  $y_{n+1}$  也包含在  $f_{n+1}$  中, 当  $f$  为  $y$  的非线性函数时, 一般来讲, 是不可能将  $y_{n+1}$  表示成  $y_n$  的函数. 故使用隐式公式时需由其他方法提供一个初次逼近值  $y_{n+1}^{(0)}$ , 然后再通过适当的迭代过程来求得  $y_{n+1}$  的改进值, 所以它不是自开始的公式.

为了定义方法的截断误差及阶数, 我们先来比较数值解  $y_n$  所满足的 Euler 公式与精确解  $y(t)$  所满足的带余项的 Taylor 公式

$$y(t_{n+1}) = y(t_n) + hy'(t_n) + \frac{h^2}{2} y''(\xi_n),$$

其中  $\xi_n$  为区间  $(t_n, t_{n+1})$  中的某一点, 因为  $y'(t_n) = f(t_n, y(t_n))$ , 上式可以改写为

$$y(t_{n+1}) = y(t_n) + hf(t_n, y(t_n)) + \frac{h^2}{2} y''(\xi_n). \quad (11)$$

由此可见, Euler 公式就是由精确解满足的方程中截去了  $\frac{h^2}{2} y''(\xi_n)$  这一项而得到的近似公式. 我们称  $\frac{h^2}{2} y''(\xi_n)$  为 Euler 公式的截断误差或称局部离散误差, 它是与  $h^2$  同阶的量, 记之为  $O(h^2)$ . 又由于丢掉的项含有  $y(t)$  的二阶导数, 所以 Euler 方法是一个一阶的方法. 我们定义: 为所有不超过  $p$  次的多项式都精确满足的公式为  $p$  阶的方法. 在第二章中我们将介绍其他一些常用的单步法, 我们将会发现方法的阶数正好比方法的截断误差中  $h$  的幂次低 1, 即若截断误差为  $O(h^{p+1})$ , 则方法为  $p$  阶的. 所以方法的阶数可以作为衡量方法精确度的一个重要标志.



## 第二章 常用的单步法

### § 1 Euler 方法

Euler 方法是最简单的单步法, 它是一阶的, 精度较差, 但由于公式简单, 且有明显的几何解释, 有利于初学者从直观上了解数值  $y_n$  怎样逼近微分方程精确解  $y(x)$  的, 所以有必要对这个方法进行详细的讨论.

#### 1.1. Euler 公式及其几何解释 解初值问题

$$\begin{aligned} y' &= f(t, y), \\ y(0) &= \eta \end{aligned} \quad (1)$$

的 Euler 方法是由递推公式

$$\begin{aligned} y_{n+1} &= y_n + hf_n, \\ y_0 &= \eta \end{aligned} \quad (2)$$

来定义的. 微分方程  $y' = f(t, y)$  的右端函数  $f(t, y)$  在  $(t, y)$  平面上的带形区域  $R: a \leq t \leq b, -\infty < y < +\infty$  内确定了一个方向场. 求解初值问题(1), 其几何意义相当于确定一条通过初始点  $(t_0, y_0)$  的曲线  $y = y(t)$ , 使曲线上的每一点的切线方向与已给向量场在该点的方向一致. 我们从 Euler 法的递推公式中可以看到, 从  $(t_0, y_0)$  点出发求  $(t_1, y_1)$  点的过程, 实际上是从  $(t_0, y_0)$  出发, 沿方向场在这一点的方向  $f_0$  画一线段, 它与直线  $t = t_1$  的交点, 便是所求的  $(t_1, y_1)$  点, 同样从  $(t_1, y_1)$  点出发, 沿方向场在这一点的方向  $f_1$  画一线段, 它与直线  $t = t_2$  的交点, 便是所求的  $(t_2, y_2)$  点. 依次类推, 从  $(t_{n-1}, y_{n-1})$  出发, 便得到所要求的  $(t_n, y_n)$  点. 以  $A_0, A_1, A_2, \dots, A_n$

分别表示这些点. 把这些横坐标间隔为  $h$  的点  $A_0, A_1, A_2, \dots, A_n$ , 分别用直线段联起来, 便得到了一条折线  $\overline{A_0A_1}, \overline{A_1A_2}, \dots, \overline{A_{n-1}A_n}$ , 我们将它记作  $P_0$ . 同样, 以  $\frac{h}{2}$  为步长按上述方法, 从  $(t_0, y_0)$  这一点出发, 便可得到  $(t_1^{(1)}, y_1^{(1)}), (t_2^{(1)}, y_2^{(1)}), \dots$ . 把这些点联接起来, 就得另一条折线  $P_1, \overline{A_0^{(1)}A_1^{(1)}}, \overline{A_1^{(1)}A_2^{(1)}}, \dots, \overline{A_{n-1}^{(1)}A_n^{(1)}}, \dots$ . 依次类推, 以  $\frac{h}{2^k}$  作步长得到折线  $P_k: \overline{A_0^{(k)}A_1^{(k)}}, \overline{A_1^{(k)}A_2^{(k)}}, \dots, \overline{A_{n-1}^{(k)}A_n^{(k)}} \dots$ . 这样, 便得到了一个折线序列  $\{P_n\}$ , 其中  $A_0 = A_0^{(1)} = \dots = A_0^{(k)} = \dots$ .

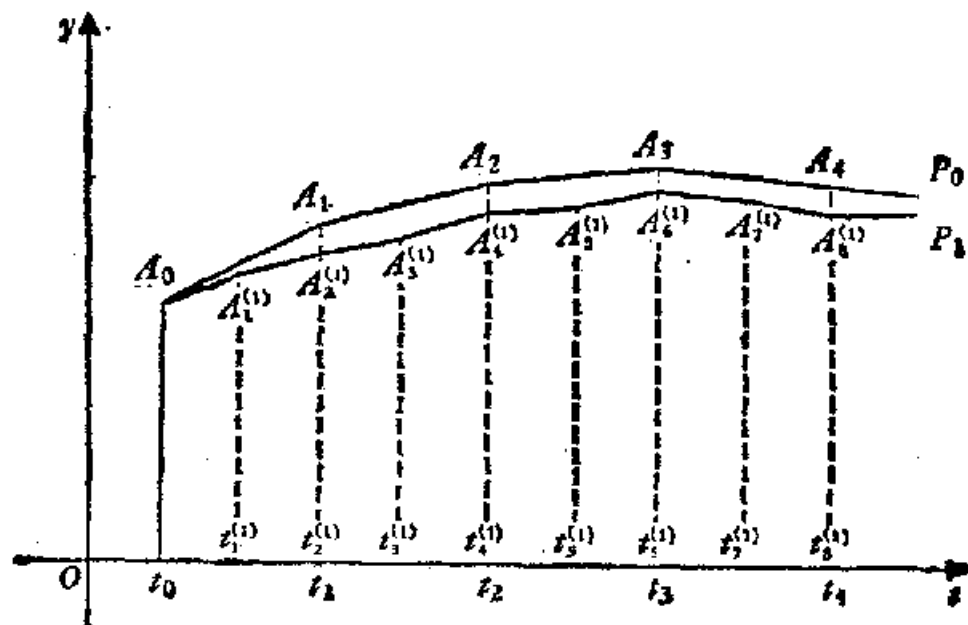


图 2.1

Euler 法便是以折线序列  $\{P_n\}$  来逼近微分方程精确解的方法, 所以 Euler 法又称为折线法.

**1.2. 用 Euler 法解  $y' = \lambda y$  的收敛性** 从前一节对 Euler 法的讨论, 我们可以看到这样一个事实: 折线序列  $\{P_n\}$  中每一条折线的每一个小线段都是以左端点处方向场所规定的方向为斜率的, 所以如果我们观察  $(a, b)$  中的某一点  $t$ , 以  $t - a = h$  作为步长, 用 Euler 公式计算一次, 便得到  $t$  这一点

$y(t)$  的近似值  $y_i^{(0)}$ , 下标  $i$  表示横坐标为  $t$ . 如以  $\frac{h}{2}$  为步长, 使用 Euler 公式计算两次, 所得到的近似值记为  $y_i^{(1)}$ .  $y_i^{(1)}$  应较  $y_i^{(0)}$  更接近  $y(t)$ . 因为在中点用微分方程所规定的方向修正了一次, 参看图 2.2.

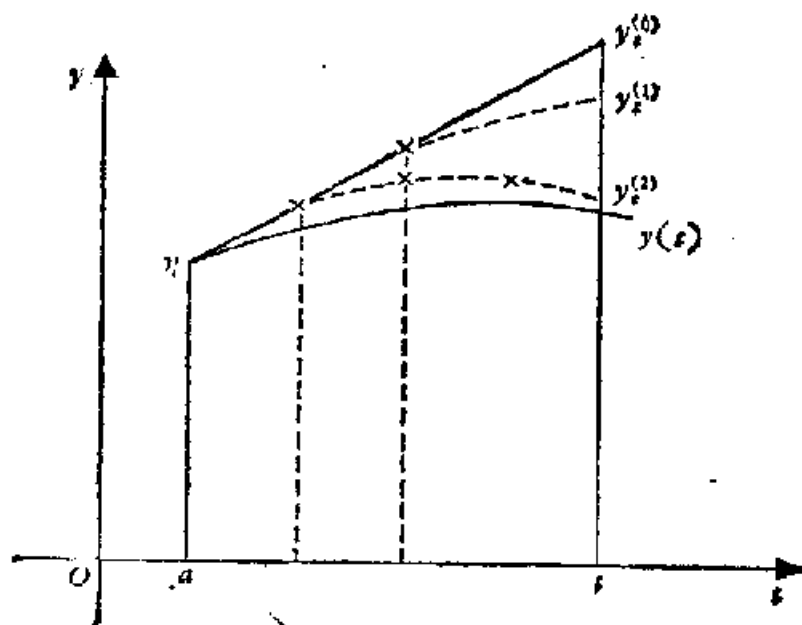


图 2.2

如果以  $\frac{h}{4}$  为步长用 Euler 公式计算四次, 所得到的  $t$  这一点的近似值记为  $y_i^{(2)}$ , 则  $y_i^{(2)}$  较之  $y_i^{(1)}$  更接近于  $y(t)$  了. 所以, 对固定的任意点  $t$  以越来越小的步长反复多次使用 Euler 公式所得到的  $t$  点的数值解, 就会越来越接近微分方程的解在这一点的确切值  $y(t)$ . 今以初值问题

$$\begin{cases} y' = \lambda y, \\ y(0) = 1 \end{cases} \quad (3)$$

为例来证明这一事实.

这时由于  $f(t, y) = \lambda y$ , Euler 公式便化为

$$y_{n+1} = y_n + \lambda h y_n = (1 + \lambda h) y_n, \quad n = 0, 1, 2, \dots$$

由初始条件  $y(0) = 1$ , 得  $y_0 = 1$ , 因此有  $y_1 = 1 + \lambda h$ ,  $y_2 = (1 + \lambda h)y_1 = (1 + \lambda h)^2$ , 一般有

$$y_n = (1 + \lambda h)^n, \quad n = 1, 2, \dots$$

由于  $n = t/h$ , 所以解在  $t_n = t$  的近似值可以表示成

$$y_n = (1 + \lambda h)^{\frac{t}{h}} = [(1 + \lambda h)^{\frac{1}{\lambda h}}]^{\lambda t}.$$

令  $h \rightarrow 0$ , 即  $n \rightarrow \infty$  便得到  $y_n \rightarrow e^{\lambda t}$ . 这就说明了微分方程初值问题(3)的解在  $t$  点的值可由 Euler 公式所得到的数值解任意地逼近.

在下节中我们将对一般方程  $y' = f(t, y)$  来证明 Euler 公式的收敛性.

### 1.3. 误差分析简介 Euler 法离散误差的估计

理论上由 Euler 法所得的解  $y_n$ , 当  $h \rightarrow 0$  ( $nh = t$ ) 时收敛于微分方程的精确解  $y(t)$ . 由于我们总是以一定的步长进行计算的 (其他的数值公式也是这样的), 所以用数值方法求得的解在  $t_n$  点的近似值  $y_n$  与微分方程的解  $y(t_n)$  之间就有差别, 记为

$$e_n = y_n - y(t_n).$$

我们称  $e_n$  为 Euler 法在  $t_n$  点的离散误差. 它是由于微分方程的离散化而产生的, 所以又称为方法误差.

此外, 由于计算机的字长有限, 在计算过程中不可避免地会产生舍入误差, 因此由近似公式计算得到的值  $\tilde{y}_n$  和 Euler 法的理论解  $y_n$  也有差别, 记为

$$r_n = \tilde{y}_n - y_n,$$

我们称  $r_n$  为舍入误差.

如果利用三角不等式, 则可得总误差

$$|\tilde{y}_n - y(x_n)| = |(\tilde{y}_n - y_n) + (y_n - y(x_n))| \leq |e_n| + |r_n|.$$

借助于解析工具, 我们可以估计数值公式所带来的误差, 即离散误差, 而舍入误差却复杂得多, 因为它的产生是多因素的, 概括起来大致可以分为下面几方面:

(i) 计算机的字长;

(ii) 机器所使用的数字系统;

(iii) 定点或浮点运算;

(iv) 数的运算次序;

(v) 计算 $f(x, y)$ 所用子程序的精确度;

以及其它的因素. 因此讨论舍入误差是很复杂的, 但还是可以设法估计出它的界. 例如, 可以把舍入误差当作随机变量, 采用统计方法, 可得较好的结果. Henrioi 对定点运算作出了系统的研究<sup>1)</sup>.

下面讨论 Euler 方法的离散误差. 对一阶常微分方程的初值问题

$$\begin{cases} y' = f(t, y), \\ y(0) = \eta, \end{cases} \quad (1)$$

估计用 Euler 法所产生的离散误差的界. 此处, 假设(1)中的函数  $f(t, y)$  和它的导数在区域  $R$ :  $a \leq t \leq b$ ,  $-\infty < y < +\infty$  内连续且有界. 这时, 存在唯一的解  $y(t)$ . 于是必定存在两个常数  $M$  和  $K$ , 使得不等式

$$\begin{aligned} |y''(t)| &= |f'(t, y(t))| \\ &= \left| \frac{\partial f(t, y)}{\partial t} + f(t, y) \frac{\partial f(t, y)}{\partial y} \right| \leq M \end{aligned} \quad (4)$$

及

$$|f(t, y) - f(t, y^*)| = \left| \frac{\partial f(t, \alpha)}{\partial y} \right| |y - y^*| \leq K |y - y^*| \quad (5)$$

成立. (5)式是由微分学的中值定理而推得, 其中  $\alpha$  为界于  $y, y^*$  之间的一个量.  $(x, y)$  和  $(x, y^*)$  为区域  $R$  中的两个点.

在点  $t_{n+1}$  和  $t_n$  上的离散误差分别为

---

1) P. Henrioi, Discrete Variable Method in Ordinary Differential Equations, 1962.

$$e_{n+1} = y_{n+1} - y(t_{n+1}),$$

$$e_n = y_n - y(t_n).$$

两式相减即得

$$e_{n+1} - e_n = y_{n+1} - y_n - [y(t_{n+1}) - y(t_n)]. \quad (6)$$

假设初始值是精确的, 即  $e_0 = 0$ , 由 Euler 公式

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

及 Taylor 展式

$$y(t_{n+1}) - y(t_n) = hf(t_n, y(t_n)) + \frac{h^2}{2!} f'(\xi_n, y(\xi_n)),$$

$$t_n < \xi_n < t_{n+1},$$

可得

$$e_{n+1} - e_n = h[f(t_n, y_n) - f(t_n, y(t_n))] - \frac{h^2}{2!} f'(\xi_n, y(\xi_n)). \quad (7)$$

由(4), (5)两式得

$$|e_{n+1} - e_n| \leq hK |e_n| + \frac{h^2}{2!} M. \quad (8)$$

又因

$$|e_{n+1}| \leq |e_{n+1} - e_n| + |e_n|,$$

故(8)式又可改写为

$$|e_{n+1}| \leq (1 + hK) |e_n| + \frac{M}{2!} h^2. \quad (9)$$

递推下去可以得到

$$|e_n| \leq \frac{Mh}{2K} [(1 + hK)^n - 1]. \quad (10)$$

为了证明这个不等式, 需要用到下面的引理.

**引理** 假设数列  $\xi_n$  满足不等式

$$|\xi_{n+1}| \leq A |\xi_n| + B, \quad n = 0, 1, 2, \dots, N-1, \quad (11)$$

则有

$$|\xi_n| \leq A^n |\xi_0| + \begin{cases} \frac{A^n - 1}{A - 1} B, & A \neq 1, \\ nB, & A = 1, \end{cases} \quad (12)$$

其中  $A$  和  $B$  是与  $n$  无关的非负常数.

证明 用数学归纳法加以证明, 当  $n=1$  时, 不等式 (12) 就是不等式 (11), 由假设它是成立的. 现假设 (12) 对  $n < N$  的一切  $n$  都成立, 应用不等式 (11), 对于  $A \neq 1$ , 我们有

$$\begin{aligned} |\xi_{n+1}| &\leq A \left\{ A^n |\xi_0| + \frac{A^n - 1}{A - 1} B \right\} + B \\ &= A^{n+1} |\xi_0| + \left( A \frac{A^n - 1}{A - 1} + 1 \right) B, \\ &= A^{n+1} |\xi_0| + \frac{A^{n+1} - 1}{A - 1} B. \end{aligned}$$

所以, 不等式 (12) 对  $n+1$  也成立.

当  $A=1$  时,

$$|\xi_{n+1}| \leq |\xi_0| + nB + B = |\xi_0| + (n+1)B.$$

于是, 引理便得到了证明.

对 (9) 使用上述引理, 立即就得到 (10).

为了得到离散误差  $e_n$  的界, 利用关系  $1+hK \leq e^{hK}$  与 (10) 便可进一步得到下面的估计式

$$|e_n| \leq \frac{Mh}{2K} (e^{nKh} - 1) \quad (18)$$

设所讨论的区间为  $[a, b]$ , 则由于  $0 \leq nh = t_n - a \leq b - a$ , 不等式 (13) 又可改写为

$$|e_n| \leq \frac{Mh}{2K} (e^{(b-a)K} - 1).$$

这就是所要求的离散误差的界, 综上所述, 我们有

**定理** 假设初值问题 (1) 中的右端函数  $f(t, y)$  及其导数在  $R$  内连续且有界. 并设初始值是精确的, 即  $e_0=0$ , 则以

Euler 公式所得到的近似解  $y_n$  的离散误差  $e_n$  满足不等式

$$|e_n| \leq \frac{Mh}{2K} (e^{lK} - 1), \quad (14)$$

其中  $M, K$  为常数, 而  $l$  为积分区间的全长, 它们都是与  $h$  无关的常量.

**推论** 在上述定理的假设下, 用 Euler 公式所得到的初值问题 (1) 在点  $t=t_n$  的理论数值解  $y_n$ , 收敛于它的精确解  $y(t)$ .

事实上, 我们有

$$\lim_{h \rightarrow 0} |e_n| \leq \lim_{h \rightarrow 0} \frac{Mh}{2K} (e^{lK} - 1) = 0,$$

亦即当  $h \rightarrow 0$  时  $y_n \rightarrow y(t)$ .

**数值例子** 估计用 Euler 法解初值问题

$$\begin{cases} y' = t + y \\ y(0) = 0 \end{cases} \quad (15)$$

的离散误差.

**解** 初值问题 (15) 的解析解为

$$y = e^t - t - 1.$$

$M$  为  $f'(t, y)$  在  $[0, t_n]$  上的界, 故有

$$M = |f'(t, y)|_{\max} = |e^{t_n}|_{\max} = e^{t_n}$$

$K$  为  $f_y$  的界, 所以

$$K = |f_y(t, y)|_{\max} = 1.$$

取  $h=0.1$ , 据不等式 (10) 得

$$|e_n| \leq \frac{0.1e^{t_n}}{2} [(1+0.1)^n - 1]. \quad (16)$$

从 (16) 的右端可计算出在各个点  $t_n$  上的  $|e_n|_{\max}$ . 由 Euler 公式

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

以及  $y(t_n) = e^{t_n} - t_n - 1$  计算出离散误差  $e_n = y_n - y(t_n)$ . 计算结果见下表:



$n$	$t_n$	$ e_n _{\max}$	$e_n = y_n - y(t_n)$
0	0.0	—	0.0000
1	0.1	0.0055	-0.0052
2	0.2	0.0128	-0.0114
3	0.3	0.0222	-0.0189
4	0.4	0.0349	-0.0277
5	0.5	0.0502	-0.0382
6	0.6	0.0704	-0.0505
7	0.7	0.0930	-0.0651
8	0.8	0.1226	-0.0819
9	0.9	0.1678	-0.1017
10	1.0	0.2175	-0.1246

从表 1 中可以看出,用估计式(10)得出的离散误差的界,要较实际所得的离散误差要大.但是,对于给定的允许误差,不等式(10)可以帮助我们选取能保证精确度的步长  $h$ .

Euler 法优点是简单易行,每前进一步只要计算一个函数值  $f(t_n, y_n)$ ,但它是一个一阶的方法,精确度较低.为了满足实际需要有必要寻求精确度较高的方法.

## § 2 Runge-Kutta 方法

**2.1. 显式 Runge-Kutta 公式** 由于不宜于直接使用 Taylor 展式的方法,今设法间接使用 Taylor 展式,以求得精确度较高的数值方法. Runge 首先提出了间接使用 Taylor 展式的方法,即用在  $n$  个点上函数值  $f$  的线性组合来代替  $f$  的导数,然后按 Taylor 公式展开,确定其中的系数,以提高方法的阶数,这样既能避免计算  $f$  的导数,同时又保证了精确度,其法如下: 令

$$y_{n+1} = y_n + \sum_{i=1}^v w_i b_i, \quad (17)$$

其中  $w_i$  为待定的权因子,  $v$  为所使用的  $f$  值的个数,  $k_i$  满足下列方程:

$$\begin{aligned} k_i &= hf(t_n + c_i h, y_n + \sum_{j=1}^{i-1} a_{ij} k_j), \\ c_i &= 0 \quad i=1, 2, \dots, v \end{aligned} \quad (18)$$

即

$$\begin{cases} k_1 = hf(t_n, y_n), \\ k_2 = hf(t_n + c_2 h, y_n + a_{21} k_1), \\ k_3 = hf(t_n + c_3 h, y_n + a_{31} k_1 + a_{32} k_2), \end{cases} \quad (19)$$

余类推. 当  $v=1$  时, 只需要一个  $k_1$ , 这时公式(17)就化为前面所讲的 Euler 公式. (18)和(19)中出现的参数  $w_1, \dots, w_v; c_2, \dots, c_v, a_{21}, \dots$ , 为提高方法的阶数创造了条件.

将  $y_{n+1}$  展成  $h$  的幂级数并与微分方程精确解的 Taylor 展式相比较, 例如, 我们取  $v=2$ , 将  $k_1, k_2$  在同一点  $(t_n, y_n)$  展开, 则有

$$\begin{cases} k_1 = hf_n \\ k_2 = hf_n + h^2(c_2 f'_{tn} + a_{21} f_n f'_{yn}) + O(h^3) \end{cases} \quad (20)$$

将(20)代入(17)并与  $y(t_n + h)$  在  $t_n$  点的 Taylor 展式:

$$\begin{aligned} y(t_n + h) &= y(t_n) + hf(t_n, y(t_n)) + \frac{h^2}{2!}(f'_{tn} + f'_{yn} f_n) \\ &+ \frac{h^3}{3!}[(f''_{tntn} + 2f''_{tnyn} + f''_{nn} f'_{yn} f_n) + f'_{yn}(f'_{tn} + f_n f'_{yn})] + \dots \end{aligned} \quad (21)$$

逐项进行比较, 令  $h, h^2$  项的系数相等, 便得到三个等式

$$w_1 + w_2 = 1, \quad w_2 c_2 = 1/2, \quad w_2 a_{21} = 1/2 \quad (22)$$

取  $c_2$  作为自由参数来定  $w_1$  及  $w_2$ . 取  $c_2 = 1/2, 1/3, 1$  就得到相应的  $(w_1, w_2)$  分别为  $(0, 1), (\frac{1}{4}, \frac{3}{4})$  及  $(\frac{1}{2}, \frac{1}{2})$ , 相应的(17)式分别为

$$y_{n+1} = y_n + hf\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hf_n\right), \quad (23)$$

$$y_{n+1} = y_n + \frac{h}{4}\left[f(t_n, y_n) + 3f\left(t_n + \frac{2}{3}h, y_n + \frac{2}{3}hf_n\right)\right], \quad (24)$$

$$y_{n+1} = y_n + \frac{h}{2}[f(t_n, y_n) + f(t_n + h, y_n + hf_n)]. \quad (25)$$

这是三个典型的二阶 Runge-Kutta 公式, 它们分别称为中点公式, Heun 公式和改进的 Euler 公式.

对于  $v=3$  及  $v=4$  的情形我们也可以完全仿上述方法推导出三阶和四阶的 Runge-Kutta 公式, 这里我们仅将几个最著名的公式列举如下:

三阶 Runge-Kutta 公式:

$$\begin{cases} y_{n+1} = y_n + \frac{1}{6}[k_1 + 4k_2 + k_3], \\ k_1 = hf(t_n, y_n), \\ k_2 = hf\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right), \\ k_3 = hf(t_n + h, y_n - k_1 + 2k_2), \end{cases} \quad (26)$$

及

$$\begin{cases} y_{n+1} = y_n + \frac{1}{4}[k_1 + 3k_3] \\ k_1 = hf(t_n, y_n), \\ k_2 = hf\left(t_n + \frac{1}{3}h, y_n + \frac{1}{3}k_1\right), \\ k_3 = hf\left(t_n + \frac{2}{3}h, y_n + \frac{2}{3}k_2\right), \end{cases} \quad (27)$$

及

$$\begin{cases} y_{n+1} = y_n + \frac{1}{9}[2k_1 + 3k_2 + 4k_3], \\ k_1 = hf(t_n, y_n), \\ k_2 = hf\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right), \\ k_3 = hf\left(t_n + \frac{3}{4}h, y_n + \frac{3}{4}k_2\right). \end{cases} \quad (28)$$

四阶的 Runge-Kutta 公式:

1) 古典形式:

$$\begin{cases} y_{n+1} = y_n + \frac{1}{6} [k_1 + 2k_2 + 2k_3 + k_4], \\ k_1 = hf(t_n, y_n), \\ k_2 = hf\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right), \\ k_3 = hf\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}k_2\right), \\ k_4 = hf(t_n + h, y_n + k_3). \end{cases} \quad (29)$$

2) Kutta 公式:

$$\begin{cases} y_{n+1} = y_n + \frac{1}{8} [k_1 + 3k_2 + 3k_3 + k_4], \\ k_1 = hf(t_n, y_n), \\ k_2 = hf\left(t_n + \frac{1}{3}h, y_n + \frac{1}{3}k_1\right), \\ k_3 = hf\left(t_n + \frac{2}{3}h, y_n + \frac{1}{3}k_1 + k_2\right), \\ k_4 = hf(t_n + h, y_n + k_1 - k_2 + k_3). \end{cases} \quad (30)$$

3) Gill 公式(它有减少舍入误差的优点):

$$\begin{cases} y_{n+1} = y_n + \frac{1}{6} [k_1 + (2 - \sqrt{2})k_2 + (2 + \sqrt{2})k_3 + k_4], \\ k_1 = hf(t_n, y_n), \\ k_2 = hf\left(t_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right), \\ k_3 = hf\left[t_n + \frac{1}{2}h, y_n + \frac{\sqrt{2}-1}{2}k_1 + \left(1 - \frac{\sqrt{2}}{2}\right)k_2\right], \\ k_4 = hf\left[t_n + h, y_n - \frac{\sqrt{2}}{2}k_2 + \left(1 + \frac{\sqrt{2}}{2}\right)k_3\right]. \end{cases} \quad (31)$$

上述公式称为显式的或开端点的 Runge-Kutta 公式, 因为在计算  $k_{i+1}$  时只用到  $k_i, k_{i-1}, \dots, k_1$ .

四阶以下的 Runge-Kutta 公式,其阶数与每前进一步所需要计算的函数值  $f$  的次数是一致的,四阶以上的 Runge-Kutta 公式,所需计算的  $f$  值的次数要比阶数多,若用  $N(v)$  表示只需计算  $v$  次  $f$  值所获得的方法的最高阶数,则有

$$N(v) = v, \quad 1 \leq v \leq 4, \quad (32-1)$$

而

$$\begin{cases} N(5) = 4, & N(6) = 5, & N(7) = 6, \\ N(8) = 6, & N(9) = 7, \end{cases} \quad (32-2)$$

所以四阶以上的 Runge-Kutta 方法计算  $f$  值的工作量将大大增加. 对于大量的实际问题,四阶的 Runge-Kutta 公式已可满足对精度的要求,如使用者有特殊需要可以查看 Butcher 的著作<sup>1)</sup> 对于  $p \leq 8$  所列的 Taylor 级数系数及误差项的表.

**2.2. 显式 Runge-Kutta 公式的局部截断误差** 为了控制 Runge-Kutta 公式的误差,可以通过适当地选取步长  $h$  来实现,所以有必要研究 2.1 中所列出的三个公式的局部截断误差.

1) 二阶的 Runge-Kutta 公式的局部截断误差为

$$\begin{aligned} T(t, h) = & h^3 \left( \frac{1}{6} - \frac{c_2}{4} \right) (f''_{tt} + 2f''_{tv} + f''_{vv} f^2) \\ & + \frac{h^3}{6} (f'_t f'_v + f'^2_v f). \end{aligned} \quad (33)$$

根据  $f' = f'_t + f f'_v$ , (33) 可以改写为

$$T(t, h) = h^3 \left( \frac{1}{6} - \frac{c_2}{4} \right) f'' + \frac{h^3}{4} c_2 f'_v f'.$$

对于  $c_2 = 2/3$  的 Heun 公式,截断误差的第一项正好为零. 显然截断误差与微分方程有关与所选具体方法和步长  $h$  有关.

2) 三阶的 Runge-Kutta 公式的局部离散误差为

1) Butcher, J. G., *J. Austral. Math. Soc.*, 3, 185 (1963).

$$T(t, h) = (h^4/4!) \{ [1 - 4(c_2^3 w_2 + c_3^3 w_3)] D^3 f \\ + (1 - 12c_2^2 a_{32} w_3) f'_v D^2 f \\ + (3 - 24c_2 c_3 w_3) Df Df'_v + f_v'^2 Df \}. \quad (34)$$

其中,  $Df$  表示  $\frac{df}{dt}$ ,  $D^2 f$  表示  $\frac{d^2 f}{dt^2}$ .

3) 四阶的 Runge-Kutta 公式的局部离散误差为

$$T(t, h) = h^5 \left\{ \left[ \frac{1}{120} - \frac{w_3 c_2^4 + w_3 c_3^4 + w_4 c_4^4}{24} \right] D^4 f \right. \\ + \left[ \frac{1}{20} - \frac{w_3 c_2 c_3^2 a_{32} + w_4 c_4^2 (c_2 a_{42} + c_3 a_{43})}{2} \right] D^2 f_v Df \\ + \left[ \frac{1}{30} - \frac{w_3 a_{32} c_2^2 c_3 + w_4 c_4 (a_{42} c_2^2 + a_{43} c_3^2)}{2} \right] Df_v D^2 f \\ + \left[ \frac{1}{120} - \frac{w_4 a_{43} a_{32} c_2^2}{2} \right] f_v'^2 D^2 f \\ + \left[ \frac{1}{40} - \frac{w_3 a_{32}^2 c_2^2 + w_4 (a_{43} c_3 + a_{42} c_2)^2}{2} \right] f_{vv}'' D^2 f \\ + \left[ \frac{1}{120} - \frac{w_3 a_{32} c_2^3 + w_4 (a_{43} c_3^3 + a_{42} c_2^3)}{6} \right] f_v' D^3 f \\ + \left[ \frac{7}{120} - w_4 a_{43} a_{32} c_2 (c_3 + c_4) \right] f_v' Df_v Df \\ \left. + \frac{1}{120} f_v'^3 Df \right\} \quad (35)$$

其中  $Df = f'_t + f'_v f$ ,  $D^n f = \underbrace{D \cdots D}_n f$ .

四阶的 Runge-Kutta 公式的局部截断误差由于表达式太繁复, 使用不便. 在 §3 节中我们将介绍在计算过程中用数值结果的线性组合来判断误差的大小, 并用它来自动选取步长, 从而达到控制误差在允许范围以内的目的.

**2.3. 隐式 Runge-Kutta 公式** 在介绍隐式 Runge-Kutta 公式以前, 我们先回顾一下显式 Runge-Kutta 公式

(17) 与 (18).

$$y_{n+1} = y_n + \sum_{i=1}^v w_i k_i, \quad (17)$$

$$k_i = hf \left( t_n + c_i h, y_n + \sum_{j=1}^{i-1} a_{ij} k_j \right),$$

$$c_1 = 0, \quad i = 1, 2, \dots, v. \quad (18)$$

这里所涉及到的矩阵  $A = [a_{ij}]$  为下三角阵型. 现将所待定的系数列表如下:

0				
$c_2$	$a_{21}$			
$c_3$	$a_{31}$	$a_{32}$		
$\vdots$	$\vdots$	$\vdots$		
$c_v$	$a_{v1}$	$a_{v2}$	$\dots\dots\dots a_{vv-1}$	
	$w_1$	$w_2$	$\dots\dots\dots w_v$	

使用矩阵和向量的记号, 上表可简写为

$$\begin{array}{c|c} \mathbf{c} & A_L \\ \hline & \mathbf{w}^T \end{array}$$

其中的  $\mathbf{c}$  和  $\mathbf{w}$  表示列向量, 上标  $T$  表示转置,  $A_L$  表示元素为  $a_{ij}$  的矩阵, 它是一个主对角元均为零的下三角矩阵. 显式 Runge-Kutta 公式的任一个  $k_i$  的表达式中只含有  $k_1, \dots, k_{i-1}$ , 因为求和的上限仅为  $i-1$ . 如果  $A$  的主对角元素非零, 则相应的 Runge-Kutta 公式称为半隐式公式. 如果把 (18) 改为

$$y_{n+1} = y_n + \sum_{i=1}^v w_i k_i, \quad (36)$$

则每一个  $k_i$  中都含有全部的  $v$  个  $k_j, j=1, 2, \dots, v$ , 即

$$k_i = hf \left( t_n + c_i h, y_n + \sum_{j=1}^v a_{ij} k_j \right), \quad i = 1, 2, \dots, v. \quad (37)$$

这种形式的 Runge-Kutta 公式, 称为是隐式的. 它的待定系数如下表

$c_1$	$a_{11}$	$a_{12}$	.....	$a_{1v}$
$c_2$	$a_{21}$	$a_{22}$	.....	$a_{2v}$
$\vdots$	$\vdots$			
$c_v$	$a_{v1}$	$a_{v2}$	.....	$a_{vv}$
<hr/>				
	$w_1$	$w_2$	.....	$w_v$

或简记为

$$\frac{\begin{array}{c|c} \mathbf{c} & A \end{array}}{\mathbf{w}^T}$$

其中的  $A$  是一个一般的  $v \times v$  阶矩阵,  $\mathbf{c} = [c_1, c_2, \dots, c_v]^T$ ,  $\mathbf{w} = [w_1, w_2, \dots, w_v]^T$ .

隐式 Runge-Kutta 公式的建立, 需要确定参量  $\mathbf{c}$ ,  $\mathbf{w}$  和  $A$ , 常用的有下列两种方法, 其一是将 (37) 在  $(t_n, y_n)$  点附近展开, 并代入 (36) 中, 和直接把  $y(t_n+h)$  利用微分方程在  $t_n$  点附近展成的 Taylor 级数相比较, 以确定  $\mathbf{c}$ ,  $\mathbf{w}$  和  $A$  中的元素; 另一种途径是将微分方程化成等价的积分方程, 然后取阶数较高的数值积分公式, 以求得  $y_{n+1}$  的表达式, 把它和 (37) 的 Taylor 展式代入 (36) 中的结果相比较, 确定诸参数.

对于每一个 Runge-Kutta 方法, 相应的用  $w_1, \dots, w_v$  及  $c_1, \dots, c_v$  所确定的积分公式, 是一个 Gauss 型数值积分公式. 在 Gauss-Legendre 积分的情况下, Butcher 曾证明,  $c_1, c_2, \dots, c_v$  为  $P_v(2c-1)=0$  的根, 其中  $P_v(t)$  为  $v$  阶 Legendre 多项式. 对  $c$  而言, 它的定义区间是  $(0, 1)$ . 每步计算  $v$  个函数值  $f$  的隐式 Runge-Kutta 方法, 其阶数  $p=2v$ , 较显式 Runge-



Kutta 公式大大提高. 确定隐式 Runge-Kutta 公式中系数的步骤如下:

1. 求出  $v$  阶 Legendre 多项式  $P_v(2c-1)$  的  $v$  个零点  $c_1, \dots, c_v$ , (可以查表);

2. 对于每一个  $i=1, 2, \dots, v$ , 解线性方程组

$$\sum_{j=1}^v a_{ij} c_j^{k-1} = \left(\frac{1}{k}\right) c_i^k, \quad k=1, 2, \dots, v,$$

3. 从线性方程组

$$\sum_{j=1}^v w_j c_j^{k-1} = 1/k, \quad k=1, 2, \dots, v,$$

来确定系数  $w_j$  ( $j=1, 2, \dots, v$ ), 在这些方程中  $c_i \neq 0$ .

我们对公式的推导不准备作进一步的介绍, 仅就常用的三个公式把它们的系数列表如下:

I.  $v=1, p=2$  的系数表:

$\frac{1}{2}$	$\frac{1}{2}$
1	

故具体公式为:

$$\begin{cases} y_{n+1} = y_n + k_1, \\ k_1 = hf\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1\right). \end{cases} \quad (38)$$

这是一个二阶方法.

II.  $v=2, p=4$  的系数表:

$(3-\sqrt{3})/6$	$1/4$	$(3-2\sqrt{3})/12$
$(3+\sqrt{3})/6$	$(3+2\sqrt{3})/12$	$1/4$
$1/2$		$1/2$

具体公式为

$$\left\{ \begin{array}{l} y_{n+1} = y_n + \frac{1}{2} (k_1 + k_2), \\ k_1 = hf \left( t_n + \frac{(3 - \sqrt{3})}{6} h, \right. \\ y_n + \frac{k_1}{4} + \frac{(3 - 2\sqrt{3})}{12} k_2 \Big), \\ k_2 = hf \left( t_n + \frac{(3 + \sqrt{3})}{6} h, \right. \\ y_n + \frac{(3 + 2\sqrt{3})}{12} k_1 + \frac{k_2}{4} \Big). \end{array} \right. \quad (39)$$

III.  $v=3, p=6$  的系数表:

$(5 - \sqrt{15})/10$	$5/36$	$(10 - 3\sqrt{15})/45$	$(25 - 6\sqrt{15})/180$
$1/2$	$(10 + 3\sqrt{15})/72$	$2/9$	$(10 - 3\sqrt{15})/72$
$(5 + \sqrt{15})/10$	$(25 + 6\sqrt{15})/180$	$(10 + 3\sqrt{15})/45$	$5/36$
	$5/18$	$4/9$	$5/18$

具体公式为

$$\left\{ \begin{array}{l} y_{n+1} = y_n + \frac{5}{18} k_1 + \frac{4}{9} k_2 + \frac{5}{18} k_3, \\ k_1 = hf \left( t_n + \frac{5 - \sqrt{15}}{10} h, \right. \\ y_n + \frac{5}{36} k_1 + \frac{10 - 3\sqrt{15}}{45} k_2 + \frac{25 - 6\sqrt{15}}{180} k_3 \Big), \\ k_2 = hf \left( t_n + \frac{h}{2}, y_n + \frac{10 + 3\sqrt{15}}{72} k_1 \right. \\ \left. + \frac{2}{9} k_2 + \frac{10 - 3\sqrt{15}}{12} k_3 \Big), \end{array} \right. \quad (40)$$

$$\left\{ \begin{aligned} k_3 &= hf\left(t_n + \frac{5 + \sqrt{15}}{10} h, \right. \\ &\quad \left. y_n + \frac{25 + 6\sqrt{15}}{180} k_1 + \frac{10 + 3\sqrt{15}}{45} k_2 + \frac{5}{36} k_3 \right). \end{aligned} \right.$$

计算同样多次  $f$  值的显式 Runge-Kutta 公式与隐式 Runge-Kutta 公式相比较, 隐式的精确度要比显式高得多. 另一方面, 隐式的 Runge-Kutta 方法具有较好的数值稳定性, 所以在解坏条件方程组时可以使用它. 但隐式 Runge-Kutta 公式在使用时, 由于每个  $k_i$  中都含有  $k_1, \dots, k_i, k_{i+1}, \dots, k_v$ , 需要用迭代法, 计算量较大.

**2.4. 半隐式 Runge-Kutta 方法** 前面所讲的隐式 Runge-Kutta 方法虽然具有较好的数值稳定性, 但每一步都需要进行迭代, 计算量较大. 现在, 我们介绍一种所谓半隐式的 Runge-Kutta 方法, 它既有隐式 Runge-Kutta 方法的数值稳定性, 同时又不需要迭代, 这种方法是由 Rosenbrock 所提出<sup>1)</sup>, 而为 Haines 加以完善的<sup>2)</sup>.

设所讨论的微分方程初值问题仍为 (1), 令

$$y_{n+1} = y_n + w_1 k_1 + \dots + w_v k_v$$

其中

$$\begin{aligned} k_i &= h[f(x_n + b_i h, y_n + b_{i1} k_1 + b_{i2} k_2 + \dots + b_{i, v-1} k_{v-1}) \\ &\quad + a_i A(x_n + c_i h, y_n + c_{i1} k_1 + \dots + c_{i, v-1} k_{v-1}) k_i] \\ &\quad i = 1, 2, \dots, v \end{aligned} \quad (41)$$

$b_1 = c_1 = b_{10} = c_{10} = 0$ . 而  $A(y) = f'_y$ , 若  $a_i = 0, i = 1, 2, \dots, v$  则 (41) 就还原为显式的 Runge-Kutta 公式, 公式中的  $w_1, \dots, w_v, a_i, b_i, c_i, b_{ij}, c_{ij}, j < i, i = 1, 2, \dots, v, j = 1, 2, \dots, i-1$  的确定, 可仿照显式 Runge-Kutta 公式求得的过程

1) Rosenbrock, H. H., *Comput. J.* 5, 329 (1963).

2) Haines, C. F., *Comput. J.*, 12, 183 (1969).

那样,利用 Taylor 级数比较系数来确定.

从表面看来,确定  $k_1, \dots, k_v$  的方程组(41)是隐式的,因为  $k_i$  的表达式中包含了他本身,但是,由于它的表达式和  $k_{i+1}, \dots, k_v$  无关,并且确定  $k_i$  的方程组都是线性的,所以很容易把它们的明显表达式求出来:

$$k_i = h[1 - ha_i A(x_n + c_i h, y_n + c_{i1}k_1 + \dots + c_{i, i-1}k_{i-1})]^{-1} \\ \cdot f(x_n + b_i h, y_n + b_{i1}k_1 + \dots + b_{i, i-1}k_{i-1}), \\ i = 1, 2, \dots, v.$$

从上面的公式中我们清楚地看到,在求  $k_i$  时右端所含  $k_j, j < i$ , 它们都是已知量. 所以我们称这样的公式是半隐式的(或假隐式的)

特别是,当  $v=2$  时上述半隐式公式化为

$$y_{n+1} = y_n + w_1 k_1 + w_2 k_2 \\ k_1 = h[1 - ha_1 A(x_n, y_n)]^{-1} f(x_n, y_n) \\ k_2 = h[1 - ha_2 A(x_n + c_2 h, \\ y_n + c_{21}k_1)]^{-1} f(x_n + b_2 h, y_n + b_{21}k_1).$$

当  $v=2, p=3$  时, Rosenbrock 半隐式公式中的参数值如下:

$$a_1 = 1 + \frac{\sqrt{6}}{6} = 1.40824829,$$

$$a_2 = 1 - \frac{\sqrt{6}}{6} = 0.59175171,$$

$$b_{21} = c_{21} = (-6 - \sqrt{6} + \sqrt{58 + 20\sqrt{6}})/(6 + 2\sqrt{6}) \\ = 0.17378667,$$

$$b_2 = c_2 = 0.17378667,$$

$$w_1 = -0.41315432,$$

$$w_2 = 1.41315432.$$

### § 3 Richardson 外推法

在 2.2 中提供的二阶、三阶、四阶显式 Runge-Kutta 公式的局部截断误差(离散误差)公式,虽然对具体的公式,有确定的常数,又对具体的微分方程来说  $f$  是已知的,故可对  $f$  的各阶导数进行计算,但计算太繁. 而对于一般的单步法, Henrici 在比较弱的条件下,对局部截断误差(或离散误差)的积累给出了估计式. 为了叙述这个结论,我们把一般的单步法概括为

$$y_{n+1} = y_n + h\Phi(t_n, y_n; h). \quad (42)$$

(42) 中的  $\Phi(t_n, y_n; h)$  表示增量函数,而称关于微分方程的精确解  $y(t)$  的增量函数

$$\frac{y(t_n+h) - y(t_n)}{h} = \Delta(t_n, y(t_n); h) \quad (43)$$

为精确增量,我们假定  $\Phi(t, y; h)$  在区域  $D: \{t \in [a, b], y \in (-\infty, \infty), 0 \leq h \leq h_0, (h_0 > 0)\}$  中为变量  $t$  和  $y$  的连续函数,并且  $\Phi(t, y; h)$  对  $y$  来说还满足 Lipschitz 条件,即存在常数  $L > 0$ . 对于  $(t, y), (t, y^*) \in D$  都有下面的关系式

$$|\Phi(t, y; h) - \Phi(t, y^*; h)| \leq L|y - y^*|.$$

**定理** 设存在有  $N \geq 0$  及  $p \geq 0$  使得对于一切  $t \in [a, b]$ ; 及正数  $h \leq h_0$ , 都有

$$|\Phi(t, y(t); h) - \Delta(t, y(t); h)| \leq Nh^p, \quad (44)$$

则对于  $t_n \in [a, b]$  及任意的  $h \leq h_0$ , 不等式

$$|y_n - y(t_n)| \leq h^p N E_L(t_n - a) \quad (45)$$

成立,其中  $E_L$  表示 Lipschitz 函数

$$E_L(t) = \begin{cases} \frac{e^{Lt} - 1}{L}, & L > 0 \\ t, & L = 0 \end{cases}$$

这个对于离散误差的估计,相应地说明,若相对局部离散误差为  $O(h^p)$ , 则(绝对)累积离散误差与之同阶. 并且还说明了当  $h \rightarrow 0$  时( $nh=t$ )方法为收敛的.

但是由(45)式来确定  $h$  也是不方便的, 由于  $N$  不易估计, 更因为这种界都是偏于保守的, 就是能确定  $h$  也是不适当的. Henrici 进一步在较强条件下证明的关于累积离散误差的渐近公式. 下面我们将以定理形式给出, 它对我们在计算过程中估计误差和自动选择步长都是很有用的.

**定理** 若  $f$  充分可微, 则累积离散误差满足

$$e_n = h^p \varepsilon(t) + O(h^{p+1}), \quad (46)$$

其中的  $\varepsilon(t)$  称为主误差函数, 它为初值问题

$$\begin{cases} \varepsilon'(t) = f_y'(t, y(t)) \varepsilon(t) - \frac{1}{(p+1)!} y^{(p+1)}(t) \\ \varepsilon(t_0) = 0 \end{cases} \quad (47)$$

的解.

下面我们要介绍, 利用这个渐近公式和 Richardson 外推法便可以实现自动选择步长的要求. 并且通过不大的工作量便可以使解的精确度显著提高.

**8.1. Richardson 外推法思想及其在提高精度方面的应用** Richardson 的外推法思想是, 如果某种离散化的近似公式所得到的数值解与精确公式解之间的误差有简单的渐近公式的话, 则可用不同步长  $h$  与  $\frac{h}{2}$ , 所求得的数值解的线性组合来构成新的解. 它与精确解的误差, 由于在线性组合过程中消除了误差的主要部分, 因而阶数有所提高. 这种思想可以反复使用, 分别取步长为  $\frac{h}{4}$ ,  $\frac{h}{8}$ , 在每次的线性组合中消去剩余误差中的主误差项. 计算结果的精确度就会逐步有所提高. 这个方法称 Richardson 外推法. 计算数值积分的

Romberg 方法, 实际上就是这种方法的一种应用.

对于前述显式单步公式, 它的积累离散误差如上所述有简单的渐近公式(46)

$$e_n \approx h^p \varepsilon(t_n) + O(h^{p+1}).$$

虽然在一般情况下, 由于微分方程的精确解  $y(t)$  为未知的, 所以  $\varepsilon(t)$  也无法求得. 但按 Richardson 外推思想依据(46)式可由  $h$  为步长所得的数值解  $y_n(h)$  与以  $\frac{h}{2}$  为步长所得的数值解  $y_n(\frac{h}{2})$  的线性组合来提高解的精确度. 具体作法如下:

以  $h$  为步长计算  $t_n$  点的数值解  $y_n(h)$ , 它满足渐近公式(46)

$$y_n(h) - y(t_n) = h^p \varepsilon(t_n) + O(h^{p+1})$$

设以  $\frac{h}{2}$  为步长计算得到  $t_n$  点的数值解为  $y_n(\frac{h}{2})$ , 它亦满足渐近公式(46)

$$y_n(\frac{h}{2}) - y(t_n) = (\frac{h}{2})^p \varepsilon(t_n) + O(h^{p+1}). \quad (48)$$

用  $2^p$  乘(48)的两端并和(46)相减便得到

$$\frac{2^p y_n(\frac{h}{2}) - y_n(h)}{2^p - 1} = y(t_n) + O(h^{p+1}). \quad (49)$$

从而可以看出  $y_n(\frac{h}{2})$  与  $y_n(h)$  的线性组合

$$y_n = \frac{2^p y_n(\frac{h}{2}) - y_n(h)}{2^p - 1} \quad (50)$$

作为新的近似值. 它与  $y(t_n)$  的误差为  $O(h^{p+1})$ , 阶数有所提高.

**3.2. 用 Richardson 外推法进行误差估计** 前面已经提到, 利用事先的误差估计来决定适当的步长是不方便的, 因此研究在计算过程中自动选择步长的方法, 从而得到提高计

算精确度的目的是十分必要的。

利用(49)便得到  $y_n(h)$  与  $y(t_n)$  之间的误差, 它实际上与  $y(t_n)$  无关仅与数值解  $y_n(h)$  和  $y_n\left(\frac{h}{2}\right)$  有关, 因为我们有

$$y_n(h) - y(t_n) = \frac{2^p \left[ y_n(h) - y_n\left(\frac{h}{2}\right) \right]}{2^p - 1} + O(h^{p+1}),$$

略去高阶量  $O(h^{p+1})$  不计, 就得到累积离散误差的近似估计式

$$y_n(h) - y(t_n) \approx \frac{2^p \left[ y_n(h) - y_n\left(\frac{h}{2}\right) \right]}{2^p - 1}. \quad (51)$$

仿此可得

$$y_n\left(\frac{h}{2}\right) - y(t_n) \approx \frac{y_n(h) - y_n\left(\frac{h}{2}\right)}{2^p - 1}. \quad (52)$$

在计算过程中使用这个估计式随时可以进行判断. 如果它超过允许误差, 就应继续缩小步长. 如比允许误差小得多就可以放大步长. 这样作法的基本依据是, 在计算结果的总误差中舍入误差比离散误差要小得多从而可以略去不计. 在计算过程的后半期, 当舍入误差占主导地位时, 再用缩小步长的方法就不可能达到减少误差的目的了. 这时就要选择精度更高的方法.

**3.3. 有理外推法** 由于所有介绍过的方法均为收敛, 即

$$\lim_{\substack{h \rightarrow 0 \\ nh = t}} y_n(h) = y(t).$$

因此也可以根据已经计算出的一些函数值进行多项式插值或有理插值进一步改善计算结果.

为了记号简单起见, 我们用  $T(h)$  表示某个问题用某种离散化方法对应于步长  $h$  得到的在某一点处的近似数值解. 根



据方法的收敛性,

$$\lim_{h \rightarrow 0} T(h) = T(0). \quad (53)$$

$T(0)$  就是该问题的精确解. 此时, 根据已经算出的一些结果  $(h_i, T(h_i))$  作插值函数  $\hat{T}(h)$ , 进而令  $h \rightarrow 0$  取极限就得到  $T(0)$  的更好的近似值  $\hat{T}(0)$ .

现在我们介绍使用有理插值的外推法 (称有理式外推). 设初值问题为

$$\begin{cases} y' = f(t, y), \\ y(t_0) = y_0. \end{cases}$$

给定步长  $h \neq 0$ ,  $t = t_0 + h$ . 现用某种离散化方法以步长  $h_0, h_1, h_2, \dots, h_m$  ( $h_0 > h_1 > h_2 > \dots > h_m$ ) 分别进行计算, 得到  $y$  在点  $t$  处的近似值  $T(h_0, t), T(h_1, t), T(h_2, t), \dots, T(h_m, t)$ .

设  $\hat{T}_m^{(i)}(h)$  为有理式

$$\hat{T}_m^{(i)}(h) = \frac{p_0^{(i)} + p_1^{(i)}h^2 + \dots + p_\mu^{(i)}h^{2\mu}}{q_0^{(i)} + q_1^{(i)}h^2 + \dots + q_\nu^{(i)}h^{2\nu}}, \quad (54)$$

其中的  $\mu = \left[ \frac{m}{2} \right]$  为  $m/2$  的整数部分.  $\mu + \nu = m$ . 我们假定

$\hat{T}_m^{(i)}(h)$  满足插值条件

$$\hat{T}_m^{(i)}(h_k) = T(h_k, t), \quad k = i, i+1, \dots, i+m.$$

则可以证明<sup>1)</sup>, 外推值

$$T_m^{(i)} = \hat{T}_m^{(i)}(0) \approx T(0, t).$$

关于  $T_k^{(i)}$  的值可以按下列公式来计算:

$$\begin{cases} T_1^{(i)} = 0, \\ T_0^{(i)} = T(h_i, t), \\ T_k^{(i)} = T_{k-1}^{(i+1)} + \frac{T_{k-1}^{(i+1)} - T_{k-1}^{(i)}}{\left(\frac{h_i}{h_{i+k}}\right)^2 \left[1 - \frac{T_{k-1}^{(i+1)} - T_{k-1}^{(i)}}{T_{k-1}^{(i+1)} - T_{k-2}^{(i+1)}}\right] - 1}, \quad k \geq 1 \end{cases} \quad (55)$$

这些  $T_{k-2}^{(i+1)}, T_{k-1}^{(i)}, T_{k-1}^{(i+1)}, T_k^{(i)}$  可按菱形规则排成下表:

1) Bulirsch, R., Stoer, U. J., *Num. Math.*, **6**, 413—427 (1964).

$$\begin{array}{ccccccc}
& & T_0^{(0)} & & & & \\
T_{-1}^{(1)} & & & T_1^{(0)} & \cdots & & \\
& T_0^{(1)} & & T_1^{(1)} & \cdots & T_{m-1}^{(0)} & \\
T_{-1}^{(2)} & & T_0^{(2)} & & T_1^{(2)} & & T_m^{(0)} \\
& T_0^{(3)} & & T_1^{(3)} & \cdots & T_{m-1}^{(1)} & \\
\vdots & & \vdots & & \vdots & & \\
T_{-1}^{(m)} & & T_0^{(m)} & & T_1^{(m-1)} & \cdots & 
\end{array}$$

为了避免多次重复计算(55)中的差, 可以把公式(55)加以变形. 令

$$\Delta T_k^{(i)} = T_k^{(i)} - T_{k-1}^{(i+1)},$$

$$C_k^{(i)} = T_k^{(i)} - T_{k-1}^{(i)},$$

$$W_k^{(i)} = C_k^{(i)} - \Delta T_k^{(i-1)} (\equiv T_k^{(i)} - T_k^{(i-1)}),$$

则容易得到等价于(55)的公式

$$\Delta T_0^{(m)} = T(h_m, t),$$

$$C_0^{(m)} = T(h_m, t),$$

$$\Delta T_k^{(m-k)} = \frac{C_{k-1}^{(m-k+1)} W_{k-1}^{(m-k+1)}}{\left(\frac{h_{m-k}}{h_m}\right)^2 \Delta T_{k-1}^{(m-k)} - C_{k-1}^{(m-k+1)}}, \quad k=1, 2, \dots, m.$$

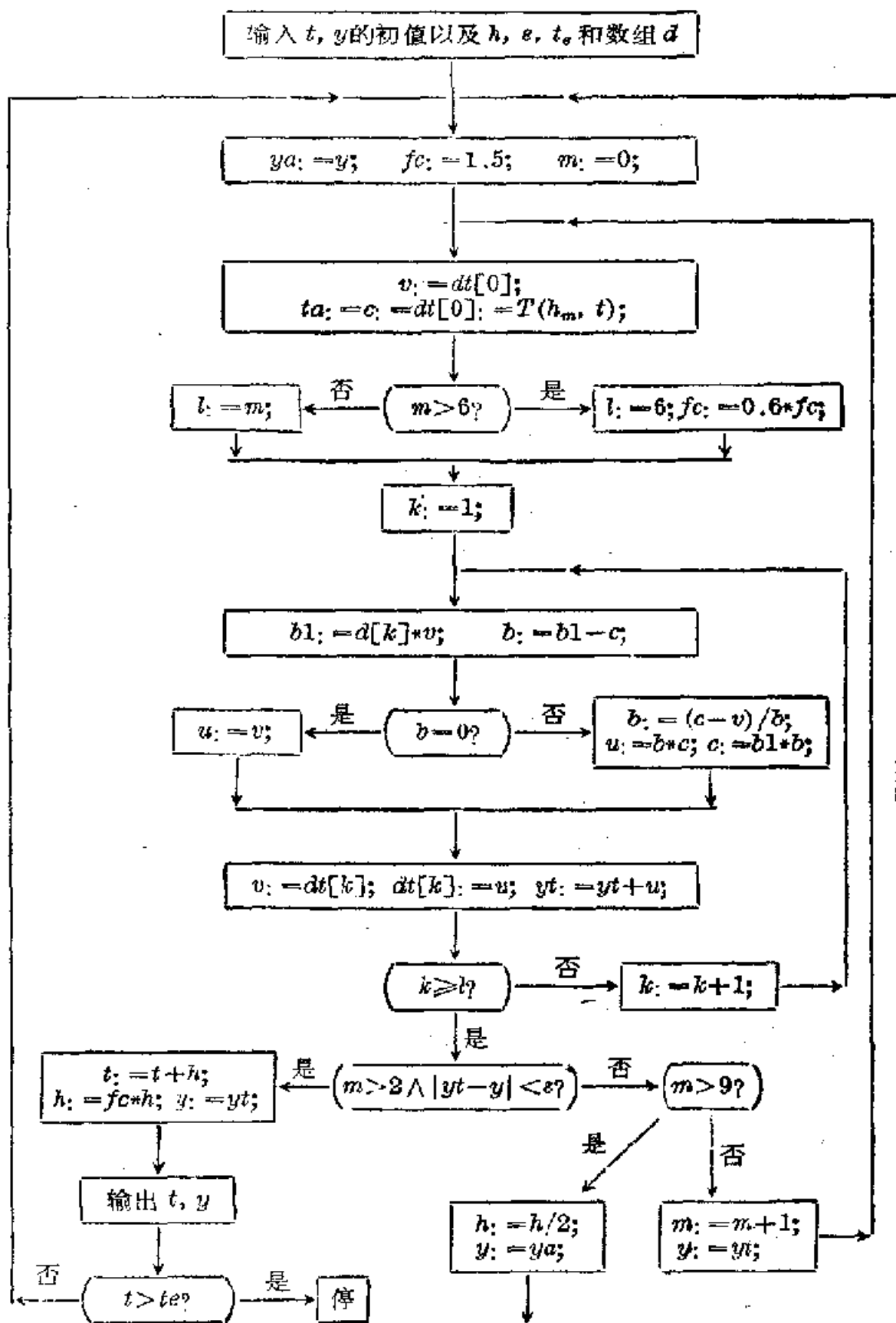
$$C_k^{m-k} = \frac{\left(\frac{h_{m-k}}{h_m}\right)^2 \Delta T_{k-1}^{(m-k)} W_{k-1}^{(m-k+1)}}{\left(\frac{h_{m-k}}{h_m}\right)^2 \Delta T_{k-1}^{(m-k)} - C_{k-1}^{(m-k+1)}}$$

$$= W_{k-1}^{(m-k+1)} + \Delta T_k^{(m-k)}, \quad k=1, 2, \dots, m.$$

$$T_m^{(0)} = \sum_{k=0}^m \Delta T_k^{m-k}.$$

这些公式依  $m=0, 1, \dots$  的顺序进行计算. (注意当  $m=0$  时将不计算  $\Delta T_k^{(m-k)}$  及  $C_k^{(m-k)}$ ).

许多计算实践表明, 有理外推法是目前初值问题求解的一个较好的方法. 下面是它的计算框图:



框图说明:

1. 框图中  $T(h_m, t)$  是指某种离散化方法对方程  $y' = f(t, y)$  以步长  $h_m$  进行积分在点  $t+h$  处得到的值, 且其展式为  $h^3$  的函数<sup>1)</sup> (例如 Gragg 修改的中点法). 其中  $h_m (m=0, 1, 2, \dots)$  是严格单调递减趋于零的步长序列. 例如, 如果取序列为  $\{h, h/2, h/3, h/4, \dots\}$  的话, 则  $h_0, h_1, h_2, \dots$ , 分别为  $h, h/2, h/3, \dots$ .

2. 由于有理外推法对舍入误差的敏感性在外推的过程中逐渐增大需要适当限制菱形图中求值的列数. 本框图只计算  $k \leq 6$  的  $T_k^{(0)}$ , 且取最大的  $m$  为 9, 并把菱形图中  $m > 6$  的  $T_0^{(m-6)}$  作为  $T_m^{(0)}$  的近似值.

3. 为了自动调节步长, 在外推 6 次内满足精度时, 便把步长因子  $f_0$  取为 1.5; 否则, 每外推一次时要把步长因子  $f_0$  缩小为  $0.6 \times f_0$ .

4. 在外推 9 次仍不能满足所需精度时, 将步长减半后从新开始外推.

5. 框图中另外一些标识符的含义:

$t$ : 初始点;  $y$ : 初始点值;  $h$ : 步长;  $\epsilon$ : 精度控制常数;  $te$ : 积分区间右端点值;  $d$ : 数组, 用以存放计算  $\Delta T_k^{(m-k)}$  时用的  $(h_{m-k}/h_m)^2$ . 因为一旦选择了一个步长序列以后, 它是可以完全确定的, 是一些不依赖于  $h$  的常数. 但要注意, 它不但随  $h$  而变, 也随  $m$  而变.  $v$ : 存放  $\Delta T_{k-1}^{(m-k)}$ ;  $c$ : 存放  $C_{k-1}^{(m-k+1)}$ ;  $u$ : 存放  $\Delta T_k^{(m-k)}$ ;  $dt$ : 数组, 用以寄存  $\Delta T_k^{(m-k)}$ , 为下步计算服务. 进行计算时, 它的初值为任意.  $yt$ : 存放外推值  $T_m^{(0)}$ .

<sup>1)</sup> O. W. 吉尔, 常微分方程初值问题的数值解法, 科学出版社, 1978.

### 第三章 线性多步法

在这一章中我们将着重介绍线性多步法中 Adams 公式, 以及各种可能的推导过程. 这对其他多步法的推导有普遍适用性.

线性多步法的一般计算公式为:

$$\begin{aligned} & \alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n \\ & = h \{ \beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \cdots + \beta_0 f_n \}, \end{aligned} \quad (1)$$
$$n=0, 1, 2, \dots,$$

其中  $\alpha_\mu, \beta_\mu, \mu=0, 1, \dots, k$  为与  $n$  无关的常数,  $f_\mu = f(t_\mu, y_\mu)$ . 一般设  $\alpha_k \neq 0, |\alpha_0| + |\beta_0| > 0$ , 由于在计算  $y_{n+k}$  时需要知道  $y_{n+k-1}, \dots, y_n$  等共  $k$  个函数值, 所以称 (1) 所确定的方法为  $k$  步法. 当  $k=1$  时就是单步法.  $k>1$  时就是多步法. 由于公式中出现的  $y_\mu$  及  $f_\mu$  都是线性的, 所以又称为线性多步法. 当  $\beta_k=0$  时, 公式为显式, 否则为隐式的. 本章只限于研究只含有  $y$  和  $y'$ , 而不含有  $y$  的更高阶导数的公式.

#### § 1 Adams 显式公式

Adams 显式公式又称 Adams-Bashforth 公式. 为讨论方便起见, 在本章中仍假设讨论的是一阶微分方程的初值问题

$$\begin{cases} y' = f(t, y), \\ y(a) = \eta. \end{cases} \quad (2)$$

然而所介绍的方法不难推广到多个方程即方程组的情形.

设  $y(t)$  为 (2) 的解, 即

$$\begin{cases} y'(t) = f(t, y(t)), \\ y(a) = \eta. \end{cases}$$

对所给微分方程两端求积分, 从  $t_n$  积分到  $t_{n+1}$ , 就得到等价的积分方程

$$y(t_{n+1}) - y(t_n) = \int_{t_n}^{t_{n+1}} f(t, y(t)) dt. \quad (3)$$

为了把(3)离散化以得到数值公式, 其方法之一是用插值多项式来代替右端的被积函数. 设插值基点为  $t_{n-q}, t_{n-q+1}, \dots, t_n$ . 由于我们的计算是从区间的左端向右端推进的, 在求  $y_{n+1}$  时,  $y_{n-q}, \dots, y_n$  均为已知的, 故使用 Newton 后差公式是方便的. 以  $p(t)$  表示函数  $f$  的插值多项式, 它在  $t_{n-q}, \dots, t_n$  诸点上取值  $f_{n-q}, \dots, f_n$ . 于是,

$$\begin{aligned} p(t) = & f_n + \frac{t-t_n}{1!h} \nabla f_n + \frac{(t-t_n)(t-t_{n-1})}{2!h^2} \nabla^2 f_n + \dots \\ & + \frac{(t-t_n) \cdots (t-t_{n-q+1})}{q!h^q} \nabla^q f_n. \end{aligned}$$

为记号简单起见, 引进辅助变量  $s = (t - t_n)/h$ , 这时, 上述插值公式就可简写为

$$p(t) = \sum_{m=0}^q (-1)^m \binom{-s}{m} \nabla^m f_n, \quad (4)$$

其中  $\binom{-s}{m}$  表示广义的二项式系数.

我们在(3)中用插值多项式  $p(t)$  来代替  $f(t, y(t))$ , 便得到公式

$$y_{n+1} - y_n = \int_{t_n}^{t_{n+1}} p(t) dt = h \sum_{m=0}^q \gamma_m \nabla^m f_n \quad (5)$$

其中诸系数

$$\begin{aligned} \gamma_m = & (-1)^m \frac{1}{h} \int_{t_n}^{t_{n+1}} \binom{-s}{m} dt = (-1)^m \int_0^1 \binom{-s}{m} ds. \\ & m=0, 1, \dots, q. \end{aligned} \quad (6)$$

它们与函数  $f$  无关, 计算每个  $\gamma_m$  都需要对一个  $m$  次多项式进行积分, 当  $m$  较大时积分是不方便的, 为便于计算  $\gamma_m$  起见, 我们引进产生  $\gamma_m$  的母函数  $G(\tau)$ , 由它便可方便地导出系数  $\gamma_m$  所满足的递推公式. 令

$$\begin{aligned} G(\tau) &= \sum_{m=0}^{\infty} \gamma_m \tau^m = \sum_{m=0}^{\infty} (-\tau)^m \int_0^1 \binom{-s}{m} ds \\ &= \int_0^1 \sum_{m=0}^{\infty} (-\tau)^m \binom{-s}{m} ds = \int_0^1 (1-\tau)^{-s} ds \\ &= \int_0^1 e^{-s \ln(1-\tau)} ds = \frac{-1}{\ln(1-\tau)} \left[ \frac{1}{(1-\tau)^s} \right]_0^1 \\ &= \frac{-\tau}{(1-\tau) \ln(1-\tau)}, \end{aligned}$$

或

$$-\frac{\ln(1-\tau)}{\tau} G(\tau) = \frac{1}{1-\tau}. \quad (7)$$

利用展式

$$\frac{1}{1-\tau} = 1 + \tau + \tau^2 + \dots \quad |\tau| < 1 \quad (8)$$

$$\frac{-\ln(1-\tau)}{\tau} = 1 + \frac{\tau}{2} + \frac{\tau^2}{3} + \dots \quad -1 \leq \tau < 1, \tau \neq 0, \quad (9)$$

将它们代入(7)便得到关系式

$$\begin{aligned} &\left(1 + \frac{\tau}{2} + \frac{\tau^2}{3} + \dots\right)(\gamma_0 + \gamma_1 \tau + \gamma_2 \tau^2 + \dots) \\ &= 1 + \tau + \tau^2 + \dots \end{aligned}$$

比较  $\tau$  的同次项系数, 便得到  $\gamma_m$  所满足的递推关系

$$\begin{aligned} \gamma_m + \frac{1}{2} \gamma_{m-1} + \frac{1}{3} \gamma_{m-2} + \dots + \frac{1}{m+1} \gamma_0 &= 1, \\ m &= 0, 1, 2, \dots \end{aligned} \quad (10)$$

由此递推式, 便可以逐个地计算出  $\gamma_0, \gamma_1, \dots, \gamma_m, \dots$ . 在下表中给出  $\gamma_m$  的部分数值:

$m$	0	1	2	3	4	5	6
$\gamma_m$	1	1/2	5/12	3/8	251/720	95/288	19087/60480

由等式 (5), 如果已知  $y_n, y_{n-1}, \dots, y_{n-q}$ , 便可计算  $f_n, f_{n-1}, f_{n-2}, \dots, f_{n-q}$ , 以及  $f_n$  的各阶后差  $\nabla^0 f_n, \nabla f_n, \dots, \nabla^q f_n$ , 从而计算出  $y_{n+1}$ . 由于  $y_{n+1}$  只出现在 (5) 式的左端, 右端不再包含它. 所以 (5) 是一个显式的公式. 称为 Adams 显式公式. 这种以后差表示的 Adams 公式 (5), 便于在台式计算机上实现, 可以在计算过程中随精度的要求而决定  $q$  的选取. 但它在电子计算机上实现是不方便的, 因为它需要输入一个差分表. 为了得到便于在电子计算机上使用的公式, 我们利用以函数值来表示差分的公式

$$\nabla^m f_n = \sum_{\rho=0}^m (-1)^\rho \binom{m}{\rho} f_{n-\rho}.$$

于是 (5) 便可以化成

$$\begin{aligned} y_{n+1} - y_n &= h \sum_{m=0}^q \gamma_m \nabla^m f_n \\ &= h \sum_{m=0}^q \gamma_m \sum_{\rho=0}^m (-1)^\rho \binom{m}{\rho} f_{n-\rho} \\ &= h \sum_{\rho=0}^q (-1)^\rho \sum_{m=0}^q \gamma_m \binom{m}{\rho} f_{n-\rho} \\ &= h \sum_{\rho=0}^q (-1)^\rho \sum_{m=\rho}^q \gamma_m \binom{m}{\rho} f_{n-\rho}. \end{aligned}$$

这个公式可进一步简化为

$$y_{n+1} - y_n = h \sum_{\rho=0}^q \beta_{q\rho} f_{n-\rho}, \quad (11)$$

其中

$$\begin{aligned} \beta_{q\rho} &= (-1)^\rho \left\{ \binom{\rho}{\rho} \gamma_\rho + \binom{\rho+1}{\rho} \gamma_{\rho+1} + \dots + \binom{q}{\rho} \gamma_q \right\}, \\ \rho &= 0, 1, \dots, q; \quad q = 0, 1, \dots. \end{aligned}$$



注意到  $\beta_{qp}$  是依赖于两个参量  $q$  及  $p$  的,  $q$  一旦选定后, 对于  $p=0, 1, \dots, q$  便得到一组系数, 在下表中给出  $\beta_{qp}$  的部分数值:

$p$	0	1	2	3	4	5
$\beta_{0p}$	1					
$2\beta_{1p}$	3	-1				
$12\beta_{2p}$	23	-16	5			
$24\beta_{3p}$	55	-59	37	-9		
$720\beta_{4p}$	1901	-2774	2616	-1274	251	
$1440\beta_{5p}$	4277	-7923	9482	-6798	2627	-425

Adams 显式公式可以方便地用于方程组的情况, 这里不作详述.

## § 2 Adams 隐式公式

Adams 隐式公式又称为 Adams-Moulton 公式.

把微分方程

$$y'(t) = f(t, y(t))$$

从  $t_{n-1}$  到  $t_n$  求积分, 便得

$$y(t_n) - y(t_{n-1}) = \int_{t_{n-1}}^{t_n} f(t, y(t)) dt.$$

利用与 § 1 中相同的方法, 便得到公式

$$y_n - y_{n-1} = h \sum_{m=0}^q \gamma_m^* \nabla^m f_n, \quad (12)$$

其中系数

$$\gamma_m^* = (-1)^m \frac{1}{h} \int_{t_{n-1}}^{t_n} \binom{-s}{m} dt = (-1)^m \int_{-1}^0 \binom{-s}{m} ds. \quad (13)$$

为了计算  $\gamma_m^*$  方便起见, 仿前可令产生  $\gamma_m^*$  的母函数为:

$$\begin{aligned}
 G^*(\tau) &= \sum_{m=0}^{\infty} \gamma_m^* \tau^m = \sum_{m=0}^{\infty} (-\tau)^m \int_{-1}^0 \binom{-s}{m} ds \\
 &= \int_{-1}^0 \sum_{m=0}^{\infty} (-\tau)^m \binom{-s}{m} ds = \int_{-1}^0 (1-\tau)^{-s} ds \\
 &= \frac{-\tau}{\ln(1-\tau)},
 \end{aligned}$$

或

$$-\frac{\ln(1-\tau)}{\tau} G^*(\tau) = 1. \quad (14)$$

利用(9)得

$$\left(1 + \frac{1}{2}\tau + \frac{1}{3}\tau^2 + \cdots\right)(\gamma_0^* + \gamma_1^*\tau + \cdots + \gamma_m^*\tau^m + \cdots) = 1.$$

比较两端  $\tau$  的同次项, 便得到计算  $\gamma_m^*$  的递推公式

$$\gamma_m^* + \frac{1}{2}\gamma_{m-1}^* + \frac{1}{3}\gamma_{m-2}^* + \cdots + \frac{1}{m+1}\gamma_0^* = \begin{cases} 1, & m=0, \\ 0, & m=1, 2, 3, \dots \end{cases} \quad (15)$$

由递推公式(15), 可以逐个地计算出  $\gamma_0^*, \gamma_1^*, \gamma_2^*, \dots, \gamma_m^*, \dots$ .  
在下表中给出  $\gamma_m^*$  的部分数值:

$m$	0	1	2	3	4	5	6
$\gamma_m^*$	1	$-\frac{1}{2}$	$-\frac{1}{12}$	$-\frac{1}{24}$	$-\frac{19}{720}$	$-\frac{3}{160}$	$-\frac{85}{6048}$

比较(8)及(14)可得  $G(\tau)$  与  $G^*(\tau)$  的关系式

$$\frac{1}{1-\tau} G^*(\tau) = G(\tau),$$

从而有

$$(1 + \tau + \tau^2 + \cdots)(\gamma_0^* + \gamma_1^*\tau + \gamma_2^*\tau^2 + \cdots) = \gamma_0 + \gamma_1\tau + \gamma_2\tau^2 + \cdots$$

比较等式两端  $\tau^m$  的系数, 便得到  $\gamma_m$  与  $\gamma_0^*, \gamma_1^*, \dots, \gamma_m^*$  之间的关系式

$$\sum_{i=0}^m \gamma_i^* = \gamma_m, \quad m=0, 1, 2, \dots \quad (16)$$

为了使 Adams 隐式公式便于在电子计算机上实现, 它又可改写为

$$y_n - y_{n-1} = h \sum_{\rho=0}^q \beta_{q\rho}^* f_{n-\rho}, \quad (17)$$

其中

$$\begin{aligned} \beta_{q\rho}^* &= (-1)^\rho \left\{ \binom{\rho}{\rho} \gamma_\rho^* + \binom{\rho+1}{\rho} \gamma_{\rho+1}^* + \dots + \binom{q}{\rho} \gamma_q^* \right\} \\ &= (-1)^\rho \sum_{m=\rho}^q \binom{m}{\rho} \gamma_m^*, \quad \begin{matrix} \rho=0, 1, 2, \dots, q, \\ q=0, 1, 2, \dots. \end{matrix} \end{aligned}$$

在表 4 中给出它的部分数值

$\rho$	0	1	2	3	4	5
$\beta_{00}^*$	1					
$2\beta_{10}^*$	1	1				
$12\beta_{20}^*$	5	8	-1			
$24\beta_{30}^*$	9	19	-5	1		
$720\beta_{40}^*$	251	646	-264	106	-19	
$1440\beta_{50}^*$	475	1427	-798	482	-173	27

在应用 Adams 隐式公式时, 常使用 Adams 显式公式对它提供一个零次近似值  $y_n^{(0)}$ . 后者实际上起预测作用, 前者通过迭代不断地校正它. 这样显式公式与隐式公式联合使用的方法称为预测-校正法.

### § 3 初始出发值的计算

多步公式既然需要知道多个出发值才能进行计算, 例如在求  $y_{n+1}$  时就需知  $y_n, y_{n-1}, \dots, y_{n-q}$ , 微分方程的初始值只提供了一个出发值, 尚缺少  $q$  个出发值, 这需要通过其他方法来计算. 今介绍两种计算出发值的方法于下.

(i) 用单步法来求所需要的出发值, 例如, 可用 Runge-Kutta 方法, 为了保证精确度, 最好将步长减半来进行计算.

(ii) 使用微分方程精确解  $y(t)$  的 Taylor 展开式. 如果  $f(t, y)$  的解析表达式比较简单, 可以根据初值问题

$$\begin{cases} y' = f(t, y), \\ y(a) = \eta. \end{cases}$$

逐步计算出  $y'(a), y''(a), \dots, y^{(s)}(a)$ , 直到 Taylor 展式

$$\begin{aligned} y(t) = & y(a) + \frac{(t-a)}{1!} y'(a) + \frac{(t-a)^2}{2!} y''(a) + \dots \\ & + \frac{1}{s!} (t-a)^s y^{(s)}(a) + R_{s+1} \end{aligned} \quad (18)$$

的余项  $R_{s+1}$  不会影响到所需的精度 (当然这里假设级数是收敛的).

例: 用 Taylor 级数求初值问题

$$\begin{cases} y' = t - y^2, \\ y(0) = 0 \end{cases}$$

的前三个点的值 ( $h=0.1$ ).

解: 由于  $f(t, y) = t - y^2$ , 所以

$$\begin{aligned} y'(t) &= t - y^2, \\ y''(t) &= 1 - 2yy', \\ y'''(t) &= -2(y'^2 + yy''), \\ y^{(4)}(t) &= -2(3y'y'' + yy'''), \\ y^{(5)}(t) &= -2(3y''^2 + 4y'y''' + yy^{(4)}), \\ y^{(6)}(t) &= -2(10y''y''' + 5y'y^{(4)} + yy^{(5)}), \\ y^{(7)}(t) &= -2(10y'''^2 + 15y''y^{(4)} + 6y'y^{(5)} + yy^{(6)}), \\ y^{(8)}(t) &= -2(35y'''y^{(4)} + 21y''y^{(5)} + 7y'y^{(6)} + yy^{(7)}). \end{aligned}$$

当  $t=a=0$  时得

$$\begin{aligned} y(0) = y'(0) = y'''(0) = y^{(4)}(0) = y^{(6)}(0) = y^{(7)}(0) &= 0, \\ y''(0) = 1, y^{(5)}(0) = -6, y^{(8)} = 252, \dots \end{aligned}$$

由(18)式得

$$y(t) = \frac{1}{2} t^2 - \frac{1}{20} t^5 + \frac{1}{160} t^8 - \dots$$

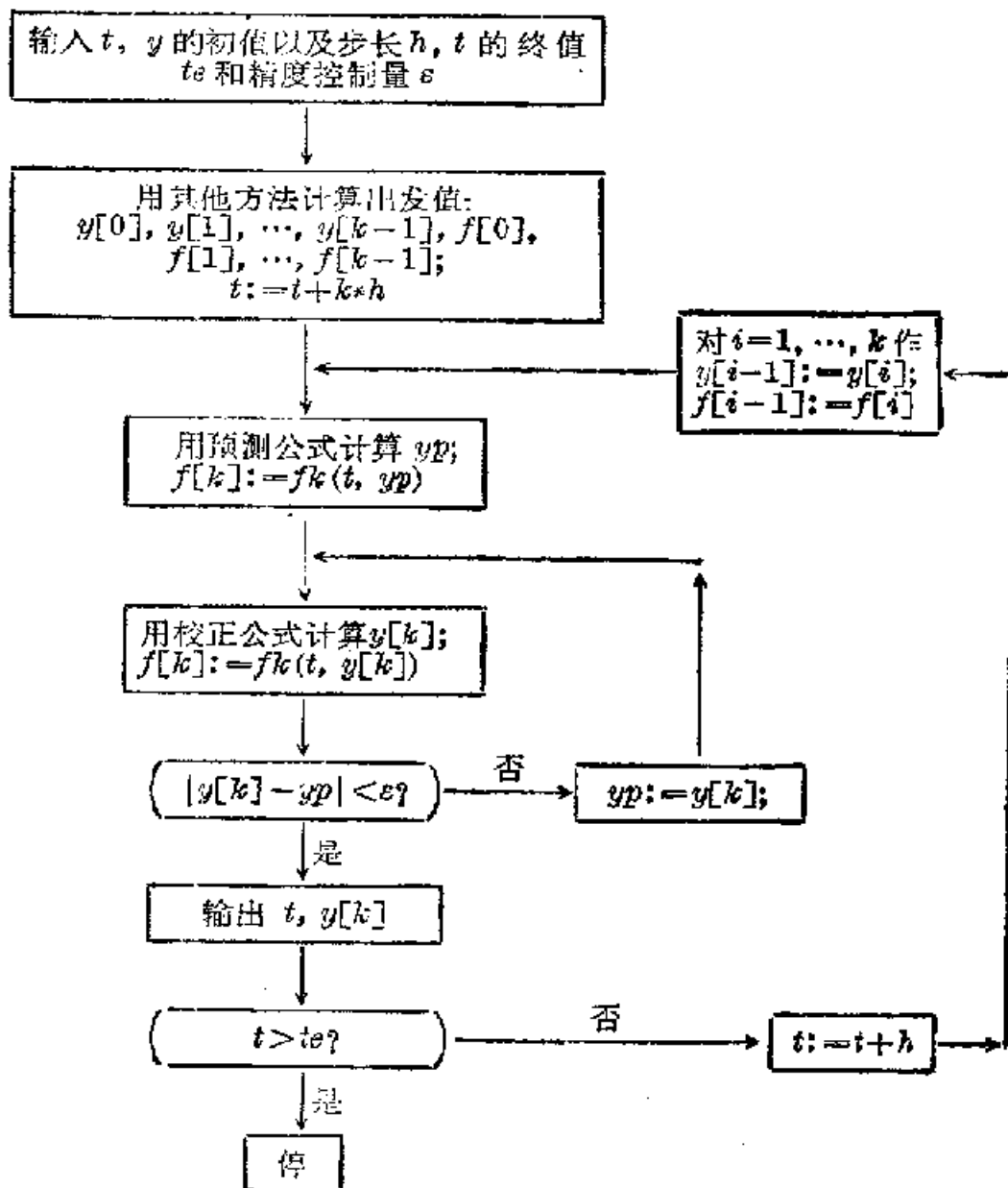
为精确起见, 选取与精确初值  $t=t_0=0$  对称的两个数值  $t_{-1}=-h$  及  $t_1=h$  ( $h=0.1$ ) 得

$$y_{-1} = 0.0050005000,$$

$$y_0 = 0,$$

$$y_1 = 0.0049995000.$$

Adams 方法框图



框图说明:

框图中的  $k$  是所选择的步数,  $f_k(t, y)$  是计算方程右端函数的函数过程.

#### § 4 Adams 公式的截断误差

为了讨论 Adams 公式的截断误差, 我们来考虑带余项的牛顿插值公式

$$y'(t) = \sum_{m=0}^q (-1)^m \binom{-s}{m} \nabla^m y'(t_n) + (-1)^{q+1} \binom{-s}{q+1} h^{q+1} y^{(q+2)}(\xi), \quad (19)$$

其中的  $\xi$  为界于  $\min(t, t_n, t_{n-q})$  及  $\max(t, t_n, t_{n-q})$  之间的一个量. 将上式两端从  $t_n$  到  $t_{n+1}$  求积分, 并由  $\gamma_m$  的定义得

$$y(t_{n+1}) - y(t_n) = h \sum_{m=0}^q \gamma_m \nabla^m y'(t_n) + R_q^{(1)}. \quad (20)$$

这是微分方程精确解  $y(t)$  所满足的公式, 把它和 Adams 显式公式(5)相比较, 注意到  $y'(t) = f(t, y(t))$  便知 Adams 显式公式的截断误差为  $R_q^{(1)}$ . 显然

$$R_q^{(1)} = (-1)^{q+1} h^{q+1} \int_{t_n}^{t_{n+1}} \binom{-s}{q+1} y^{(q+2)}(\xi) dt.$$

注意到  $\binom{-s}{q+1}$  在区间  $t_n \leq t \leq t_{n+1}$  中保持定号,  $y^{(q+2)}(\xi)$  为  $t$  的连续函数, 利用积分第二中值定理, 即得

$$R_q^{(1)} = (-1)^{q+1} h^{q+1} y^{(q+2)}(\xi') \int_{t_n}^{t_{n+1}} \binom{-s}{q+1} dt.$$

其中的  $\xi'$  是  $t$  在  $(t_n, t_{n+1})$  中所对应的一个  $\xi$  值,  $t_{n-q} < \xi' < t_{n+1}$ , 根据  $\gamma_{q+1}$  的定义,  $R_q^{(1)}$  又可写成

$$R_q^{(1)} = h^{q+2} y^{(q+2)}(\xi') \gamma_{q+1}. \quad (21)$$

这就是 Adams 显式公式的截断误差.

完全类似地可求得 Adams 隐式公式的截断误差  $R_q^{(2)}$ , 它满足等式:

$$y(t_n) - y(t_{n-1}) = h \sum_{m=0}^q \gamma_m^* \nabla^m y'(t_n) + R_q^{(2)}, \quad (22)$$

所以

$$R_q^{(2)} = h^{q+2} y^{(q+2)}(\xi) \gamma_{q+1}^*, \quad (23)$$

其中  $\xi \in (t_{n-q}, t_n)$ .

当  $q=3$  时 Adams 显式公式为

$$y_{n+1} - y_n = h \left\{ f_n + \frac{1}{2} \nabla f_n + \frac{5}{12} \nabla^2 f_n + \frac{3}{8} \nabla^3 f_n \right\},$$

或

$$y_{n+1} - y_n = \frac{h}{24} \{ 55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3} \},$$

其余项为

$$R_3^{(1)} = \frac{251}{720} h^5 y^{(5)}(\xi).$$

Adams 隐式公式为

$$y_n - y_{n-1} = \frac{h}{24} \left\{ f_n - \frac{1}{2} \nabla f_n - \frac{1}{12} \nabla^2 f_n - \frac{1}{24} \nabla^3 f_n \right\},$$

或

$$y_n - y_{n-1} = \frac{h}{24} \{ 9f_n + 19f_{n-1} - 5f_{n-2} + f_{n-3} \},$$

其余项为

$$R_3^{(2)} = -\frac{19}{720} h^5 y^{(5)}(\xi).$$

所以  $q=3$  时的 Adams 公式是四阶的方法.

## § 5 隐式公式的迭代解法

### 5.1. Jacobi 迭代法

单个方程的情形 在介绍一般的隐式公式的 Jacobi 迭代

法之前,我们先以 Adams 公式为例来说明迭代的步骤. 对于 Adams 公式(5)

$$y_n = y_{n-1} + h \sum_{m=0}^q \gamma_m^* \nabla^m f_n,$$

假设我们已由某种方法求得  $y_n$  的首次近似值  $y_n^{(0)}$ , 于是就能计算出各阶有限差

$$f_n^{(0)} = f(t_n, y_n^{(0)}), \quad \nabla f_n^{(0)} = f_n^{(0)} - f_{n-1}, \quad \nabla^2 f_n^{(0)} = \nabla f_n^{(0)} - \nabla f_{n-1},$$

再由公式

$$y_n^{(1)} = y_{n-1} + h \sum_{m=0}^q \gamma_m^* \nabla^m f_n^{(0)} \quad (24)$$

就能得到较  $y_n^{(0)}$  更为精确的近似值  $y_n^{(1)}$  (上式中出现的  $y_{n-1}$  没有上标, 表示第  $n-1$  个值已是迭代完毕后的终值) 同样, 我们将上式右端的  $y_n^{(0)}$  都换成  $y_n^{(1)}$ , 便可以得到  $y_n^{(2)}$ , 余类推. 重复这个过程, 便得到迭代公式

$$y_n^{(s)} = y_{n-1} + h \sum_{m=0}^q \gamma_m^* \nabla^m f_n^{(s-1)} \quad s=0, 1, 2, \dots \quad (25)$$

其中  $f_n^{(s)} = f(t_n, y_n^{(s)})$ . 下面我们要证明在一定限制下, 迭代公式所产生的序列是收敛的, 它收敛于差分方程的解  $y_n$ . 所以从迭代公式中所得的  $s$  次迭代值从理论上讲要较  $s-1$  次更为精确.

实际上, 由

$$y_n^{(s+1)} = y_{n-1} + h \sum_{m=0}^q \gamma_m^* \nabla^m f_n^{(s)},$$

和(25)式两端相减得

$$y_n^{(s+1)} - y_n^{(s)} = h \sum_{m=0}^q \gamma_m^* (\nabla^m f_n^{(s)} - \nabla^m f_n^{(s-1)}).$$

由于

$$\nabla^m f_n^{(s)} - \nabla^m f_n^{(s-1)} = f_n^{(s)} - f_n^{(s-1)},$$

所以



$$y_n^{(s+1)} - y_n^{(s)} = h(f_n^{(s)} - f_n^{(s-1)}) \sum_{m=0}^q \gamma_m^*.$$

由关系式(16)上式可以改写为

$$y_n^{(s+1)} - y_n^{(s)} = h\gamma_q(f_n^{(s)} - f_n^{(s-1)}). \quad (26)$$

由假设  $f(t, y)$  是满足 Lipschitz 条件的, 故有

$$|f_n^{(s)} - f_n^{(s-1)}| \leq K |y_n^{(s)} - y_n^{(s-1)}|.$$

令

$$\delta^{(s)} = y_n^{(s+1)} - y_n^{(s)}.$$

对(26)式两端取绝对值得到不等式

$$|\delta^{(s)}| \leq h\gamma_q K |\delta^{(s-1)}|.$$

于是得

$$|\delta^{(s)}| \leq (h\gamma_q K)^s |\delta^{(0)}|.$$

如果

$$h\gamma_q K < 1,$$

则级数

$$|\delta^{(0)}| \sum_{s=0}^{\infty} (h\gamma_q K)^s$$

收敛, 从而由比较判别法知级数

$$y_n^{(0)} + \sum_{s=0}^{\infty} \delta^{(s)}$$

也收敛, 故它的部分序列

$$s_p = y_n^{(0)} + \sum_{s=0}^p \delta^{(s)} = y_n^{(p+1)}, \quad p=1, 2, \dots$$

收敛, 设其极限为  $y_n$ ,  $s_p \rightarrow y_n$ .  $y_n$  便是隐式公式(12)的解. 这只需对(25)式两端求极限就可以证明. 此外不难证明  $y_n$  就是差分方程的唯一解.

在计算过程中考虑到关系式(26), 当  $f_n^{(s)} - f_n^{(s-1)}$  可以略去不计时, 即相邻二次计算结果变化不大时, 便可以停止计算. 此时所得到的  $f_n^{(s)}$  值可当  $f_n$  的终值, 由它再计算出

$\nabla^m f_n$ , 代入公式(12)便得  $y_n$  的终值.

## 5.2. 一般的隐式方法 现在来考虑

$$y_{n+1} - h\beta_0 f(x_{n+1}, y_{n+1}) - \sum_{i=1}^k (\alpha_i y_{n+1-i} + h\beta_i f_{n+1-i}) = 0 \quad (27)$$

的情形. 由于(27)式的最后一项为已知的, 若将它记为  $W_n$ , 则(27)式可以写成

$$y_{n+1} - h\beta_0 f(x_{n+1}, y_{n+1}) - W_n = 0. \quad (28)$$

由(28)式我们来确定  $y_{n+1}$ .

Jacobi 的计算格式为

$$y_{n+1}^{(s+1)} - h\beta_0 f(x_{n+1}, y_{n+1}^{(s)}) - W_n = 0. \quad (29)$$

反复迭代

设  $y_{n+1}^*$  为(28)式的理论精确解, 则

$$y_{n+1}^* - h\beta_0 f(x_{n+1}, y_{n+1}^*) - W_n = 0. \quad (30)$$

由(29)式减去(30)得到

$$y_{n+1}^{(s+1)} - y_{n+1}^* = h\beta_0 (f_{n+1}^{(s)} - f_{n+1}^*). \quad (31)$$

利用中值定理, 便得到

$$y_{n+1}^{(s+1)} - y_{n+1}^* = h\beta_0 (f_{\bar{y}})_y (y_{n+1}^{(s)} - y_{n+1}^*), \quad (32)$$

其中  $\bar{y}$  为介于  $y_{n+1}^{(s)}$  与  $y_{n+1}^*$  中的一个量, 若我们假设  $f_y$  为有界, 则存在有 Lipschitz 常数  $L$  使  $|f_y| < L$ , 由(32)式可得

$$|y_{n+1}^{(s+1)} - y_{n+1}^*| \leq h\beta_0 L |y_{n+1}^{(s)} - y_{n+1}^*|.$$

用归纳法就得到

$$|y_{n+1}^{(s+1)} - y_{n+1}^*| \leq (h\beta_0 L)^{s+1} |y_{n+1}^{(0)} - y_{n+1}^*|. \quad (33)$$

从而可知迭代收敛的主要条件为:

$$|h\beta_0 L| < 1. \quad (34)$$

在(34)式成立的情况下, Jacobi 迭代收敛于唯一的解, 它落在包含精确解在内的一个足够大的区域中. 条件(34)用于已有的 Adams 隐式公式可以得到下面的表:

$$q=1, \quad hL < 2$$

$$q=2, \quad hL < \frac{12}{5}$$

$$q=3, \quad hL < \frac{8}{3}$$

$$q=4, \quad hL < \frac{720}{251}$$

我们讨论的一般隐式方法的 Jacobi 迭代方法可以方便地用于方程组的情况, 只要将所出现的标量  $y, f$ , 改为向量,  $f_v$  改为 Jacobi 矩阵  $f_v$ , 如取谱范数则  $L$  相应地换成  $f_v$  的最大特征值, 此时条件(34)便化为

$$h\beta_0 |\lambda_{\max}| < 1. \quad (35)$$

从(35)式便可以看出当  $|\lambda_{\max}|$  比较大时,  $h$  就得选择得很小. 这样就要大量增加计算工作量, 就象坏条件方程组遇到的问题那样.

**5.3. Newton-Raphson 迭代法** 解非线性方程的一个著名的方法叫做 Newton-Raphson 迭代法. 今介绍于下. 为了记号简化, 我们用  $f_{n+1}$  来表示(28)式中的  $f(x_{n+1}, y_{n+1})$ , 这时, (28)式便可以简写为:

$$y_{n+1} - h\beta_0 f_{n+1} - W_n = 0.$$

于是便得到基本的迭代公式

$$y_{n+1}^{(s+1)} - h\beta_0 f_{n+1}^{(s+1)} - W_n = 0, \quad (36)$$

其中  $f_{n+1}^{(s+1)} = f(x_{n+1}, y_{n+1}^{(s+1)})$ .

Newton-Raphson 方法的基本思想在于把  $y_{n+1}$  的非线性函数  $f$  在  $y_{n+1}^{(s)}$  的附近线性化, 即取其近似表达式为:

$$f_{n+1} = A_{n+1}^{(s)}(y_{n+1} - y_{n+1}^{(s)}) + f_{n+1}^{(s)}, \quad (37)$$

其中的  $A_{n+1}^{(s)}$  为  $f_v$  在  $y_{n+1}^{(s)}$  上取的值, 据(37)便有

$$f_{n+1}^{(s+1)} = A_{n+1}^{(s)}(y_{n+1}^{(s+1)} - y_{n+1}^{(s)}) + f_{n+1}^{(s)}. \quad (38)$$

将它代入(36)中便得到线性化了的近似迭代关系:

$$y_{n+1}^{(s+1)} - h\beta_0 [A_{n+1}^{(s)} (y_{n+1}^{(s+1)} - y_{n+1}^{(s)}) + f_{n+1}^{(s)}] - W_n = 0,$$

从而便得到

$$\begin{aligned} y_{n+1}^{(s+1)} - y_{n+1}^{(s)} - h\beta_0 A_{n+1}^{(s)} (y_{n+1}^{(s+1)} - y_{n+1}^{(s)}) \\ + y_{n+1}^{(s)} - h\beta_0 f_{n+1}^{(s)} - W_n = 0. \end{aligned}$$

这个线性关系可以进一步表示成

$$(1 - h\beta_0 A_{n+1}^{(s)}) (y_{n+1}^{(s+1)} - y_{n+1}^{(s)}) + y_{n+1}^{(s)} - h\beta_0 f_{n+1}^{(s)} - W_n = 0.$$

所以有

$$y_{n+1}^{(s+1)} = y_{n+1}^{(s)} + (1 - h\beta_0 A_{n+1}^{(s)})^{-1} (h\beta_0 f_{n+1}^{(s)} - y_{n+1}^{(s)} + W_n). \quad (39)$$

这个公式便叫做 Newton-Raphson 迭代公式, 由给定的初始值  $y_{n+1}^{(0)}$ , 据 (39) 便可依次计算出  $y_{n+1}^{(1)}$ ,  $y_{n+1}^{(2)}$ ,  $\dots$ . 关于 Newton-Raphson 迭代的收敛性问题讨论比较复杂, 从略.

同样 Newton-Raphson 迭代公式可以方便地用于方程组的情况, 这时  $A_{n+1}^{(s)}$  为向量函数  $f$  关于变量  $y$  的分量的 Jacobi 矩阵  $f_y$  取值于  $y_{n+1}^{(s)}$  上. (39) 中的  $(1 - h\beta_0 A_{n+1}^{(s)})^{-1}$  改为  $(I - h\beta_0 A_{n+1}^{(s)})^{-1}$ , 其中  $I$  表单位矩阵.

在根据迭代公式 (39) 进行迭代时, 为了减少 Jacobi 矩阵及其逆矩阵的繁重的计算量, 在每一步迭代时, 可以固定 Jacobi 矩阵  $A_{n+1}^{(0)}$  在迭代过程中保持不变, 这样, 每一步便只需要计算一次 Jacobi 矩阵, 计算一次逆矩阵, 计算量大为减少, 其中  $A_{n+1}^{(0)}$  是在  $y_n$  上算得的值, 由于在相邻两步之间,  $y_n$  和  $y_{n+1}$  变化不大, 这样的作法是合理的, 而实践证明也是有效的, 按这种方式进行迭代的方法, 叫做修改的 (或改进的) Newton-Raphson 迭代法.

如果在某一步经过迭代多次尚不收敛, 则可以作如下的处理: 其一是缩短步长  $h$ , 因为收敛缓慢的原因有可能是由于步长过大; 其二是在迭代过程中修改一次 Jacobi 矩阵及所需要的逆矩阵, 再照前迭代下去, 以观后效.

## 第四章 预测-校正法

在上一章中介绍的 Adams 方法, 如果将显式公式与隐式公式联合使用, 前者提供预测值, 后者将预测值加以校正, 使其更精确, 这类方法就称为预测-校正法. 所用的两个公式, 分别称为预测公式和校正公式. 下面我们先以单个方程的初值问题

$$\begin{cases} y' = f(t, y), \\ y(a) = \eta \end{cases}$$

为例, 来介绍常用的预测-校正法.

### § 1 最简单的预测-校正法

可以把中点公式

$$\begin{cases} y_{n+1} = y_{n-1} + 2hy'_n, \\ T(t, h) = h^3/3y'''(\xi'), \\ t_{n-1} < \xi' < t_{n+1} \end{cases} \quad (1)$$

及梯形公式

$$\begin{cases} y_{n+1} = y_n + \frac{h}{2} [y'_{n+1} + y'_n], \\ T(t, h) = -\frac{h^3}{12} y'''[\xi''], \\ t_n < \xi'' < t_{n+1} \end{cases} \quad (2)$$

联合使用. 中点公式为一个显式公式, 用它作为预测公式, 梯形公式为隐式公式, 可用它作为校正公式, 以校正预测值. 这

两个公式是同阶的,且二者余项的系数,就绝对值来说,前者为  $1/3$ , 后者为  $1/12$ , 可望校正后的结果,精度有所改善. 这种方法,是一种最简单的预测校正法. 使用这种预测-校正公式可以有下面三种方案:

$$1) \quad \begin{cases} \bar{y}_{n+1} = y_{n-1} + 2hf_n, \\ \bar{f}_{n+1} = f(t_{n+1}, \bar{y}_{n+1}), \\ y_{n+1} = y_n + \frac{h}{2}(\bar{f}_{n+1} + f_n), \\ f_{n+1} = f(t_{n+1}, y_{n+1}). \end{cases} \quad (3)$$

然后进入下一点的计算,这样每一步上二次使用微分方程. 亦即对右端函数进行了二次估计. 称为 PECE 公式(预测-方程-校正-方程). 这种方案每一步的最终结果  $y_{n+1}$  一般并不准确满足校正公式. 我们希望得到的计算结果  $y_{n+1}^*$ , 应满足校正公式

$$y_{n+1}^* = y_n + \frac{h}{2}(y_{n+1}' + y_n').$$

所以,为了提高计算结果的精度,还需要用迭代的方法对  $y_{n+1}$  加以改进使之更为接近  $y_{n+1}^*$ .

于是我们有第二种方案:

$$2) \quad \begin{cases} y_{n+1}^{(0)} = y_{n-1} + 2hf_n, \\ f_{n+1}^{(0)} = f(t_{n+1}, y_{n+1}^{(0)}), \\ y_{n+1}^{(s+1)} = y_n + \frac{h}{2}[f_{n+1}^{(s)} + f_n], \\ f_{n+1}^{(s+1)} = f(t_{n+1}, y_{n+1}^{(s+1)}), \\ s = 0, 1, 2, \dots \end{cases} \quad (4)$$

由于在每一步的计算过程中,我们都使用了迭代法,所以,我们称这种方法为迭代的预测-校正法,而称第一种方案为简单的预测-校正法.

这种带迭代的预测-校正方法,虽然能够改善计算结果,

但由于每一步都要进行若干次迭代, 这就使计算工作量大为增加. 另外, 由于预测公式和校正公式的截断误差为同阶, 它们仅系数不同, 因此, 便可以用预测值和校正值的组合来表示截断误差. 于是便可以用计算结果来修正公式以达到提高精确度的目的. 这也就是 Richardson 外推法的基本思想, 其具体做法如下:

我们用  $p_n$  和  $c_n$  分别表示第  $n$  步  $y_n$  的预测值和校正值. 根据所用的预测公式和校正公式的截断误差公式, 我们有下列关系式:

$$y(t_{n+1}) - p_{n+1} = \frac{1}{3} h^3 y'''(\xi'_n), \quad t_{n-1} < \xi'_n < t_{n+1}, \quad (5)$$

$$y(t_{n+1}) - c_{n+1} = -\frac{1}{12} h^3 y'''(\xi''_n), \quad t_n < \xi''_n < t_{n+2}, \quad (6)$$

(5), (6) 两端分别相减得

$$c_{n+1} - p_{n+1} = \frac{1}{3} h^3 y'''(\xi'_n) + \frac{1}{12} h^3 y'''(\xi''_n). \quad (7)$$

假定微分方程的解  $y(t)$  的三阶导数  $y'''(t)$  在所述区间内是连续的, 则在  $t_{n-1}$  和  $t_{n+1}$  之间必存在一数  $\xi_n$ , 使

$$\begin{aligned} & \frac{1}{3} h^3 y'''(\xi'_n) + \frac{1}{12} h^3 y'''(\xi''_n) \\ &= h^3 \left[ \frac{1}{3} y'''(\xi'_n) + \frac{1}{12} y'''(\xi''_n) \right] \\ &= h^3 \left[ \frac{1}{3} y'''(\xi_n) + \frac{1}{12} y'''(\xi_n) \right] = \frac{5}{12} h^3 y'''(\xi_n), \end{aligned}$$

于是, (7) 就可以简写为:

$$c_{n+1} - p_{n+1} = \frac{5}{12} h^3 y'''(\xi_n), \quad t_{n-1} < \xi_n < t_{n+1}. \quad (8)$$

从而得

$$h^3 y'''(\xi_n) = \frac{12}{5} (c_{n+1} - p_{n+1}).$$

假定微分方程的解  $y(t)$  的四阶导数  $y^{(iv)}(t)$  在  $(t_{n-1}, t_{n+1})$  内

存在并有界, 则

$$y'''(\xi'_n) = y'''(\xi_n) + O(h), \text{ 当 } h \rightarrow 0,$$

$$y''(\xi''_n) = y''(\xi_n) + O(h), \text{ 当 } h \rightarrow 0.$$

据此, 便可以把预测公式和校正公式中的截断误差  $T_1$  和  $T_2$  分别表示为

$$\begin{cases} T_1 = \frac{1}{3} h^3 y'''(\xi_n) + h^3 O(h) \\ \quad \approx \frac{1}{3} h^3 y'''(\xi_n) = \frac{4}{5} (c_{n+1} - p_{n+1}); \\ T_2 = -\frac{1}{12} h^3 y'''(\xi_n) + h^3 O(h) \\ \quad \approx -\frac{1}{12} h^3 y'''(\xi_n) = -\frac{1}{5} (c_{n+1} - p_{n+1}). \end{cases} \quad (9)$$

于是, 我们便可以给出实现预测-校正法的另一种方案.

$$\begin{cases} 3) \quad \begin{cases} \text{预测: } p_{n+1} = y_{n-1} + 2hf_n; \\ \text{修改: } m_{n+1} = \dot{p}_{n+1} - \frac{4}{5} (p_n - c_n); \\ \quad \quad \quad m'_{n+1} = f(t_{n+1}, m_{n+1}); \\ \text{校正: } c_{n+1} = y_n + \frac{h}{2} (m'_{n+1} + f_n); \\ \text{修改: } y_{n+1} = c_{n+1} + \frac{1}{5} (p_{n+1} - c_{n+1}), \\ \text{计算导数: } f_{n+1} = c_{n+1}, \end{cases} \end{cases} \quad (10)$$

其中  $m_n$  表示  $y_n$  的修正值.

由于开始时无预测值和校正值可以利用, 故令  $p_0 = c_0 = 0$ . 以后就可以按上列步骤进行计算. 量  $p_n - c_n$  还可以用来调整步长. 如果它的绝对值非常小, 说明步长  $h$  还可以放大; 如果  $p_n - c_n$  的绝对值比较大, 说明步长  $h$  取得过大, 需要进一步缩小. 如果  $p_n - c_n$  出现突然的变化, 说明计算中发生问题, 应检查其原因.



## § 2 Milne 方法

**2.1. 预测公式和校正公式的建立** 前面已经讲过, 一阶微分方程的初值问题可以化为等价的积分方程. 设所考虑的区间为  $[t_{n-q}, t_{n+1}]$ , 则微分方程  $y' = f(t, y)$  的初值问题, 在这个区间上可以化为积分方程

$$y_{n+1} = y_n + \int_{t_{n-q}}^{t_{n+1}} f(t, y(t)) dt.$$

在介绍 Adams 方法时, 我们曾使用多项式插值的方法, 使上述积分方程化为一个差分方程. 就实际应用来说, 使所得公式达到四阶就可以了. 所以可取  $q=3$ , 并以

$$t_{n-3}, t_{n-2}, t_{n-1}, t_n$$

作为插值基点. 这时得到的差分方程是

$$y_{n+1} = y_{n-3} + \frac{4}{3} h (2y'_n - y'_{n-1} + 2y'_{n-2}). \quad (11)$$

它的离散误差是

$$T_1 = \frac{14}{45} h^5 y^{(5)}(\xi'), \quad t_{n-3} < \xi' < t_{n+1}. \quad (12)$$

公式(11)是显式的, 可以作为预测公式来使用, 由于它是一个多步法, 计算  $y_{n+1}$  时需要四个出发值  $y_n, y_{n-1}, y_{n-2}$  和  $y_{n-3}$ . 初值问题本身, 只提供了一个, 其余三个得用其他方法来计算, 例如用 Runge-Kutta 方法.

如果用

$$t_{n-q}, t_{n-q+1}, \dots, t_{n+1}$$

作为插值多项式的基点, 因为它包含了  $t_{n+1}$ , 得到的公式将是隐式的. 为了得到四阶的公式, 可取  $q=1$ , 这时得到的差分方程是

$$y_{n+1} = y_{n-1} + \frac{h}{3} (y'_{n+1} + 4y'_n + y'_{n-1}), \quad (13)$$

它的离散误差是

$$T_2 = -\frac{1}{90} h^5 y^{(5)}(\xi''), \quad t_{n-1} < \xi'' < t_{n+1}. \quad (14)$$

这个隐式公式可以用来作为校正公式。

用 (11) 作预测公式, 而用 (14) 作校正公式的预测-校正法, 就是 Milne 方法。

**2.2. Milne 方法的实现和修改** 按照 §1 中所介绍的使用迭代的预测-校正法(第二种方案), Milne 方法可以按下列步骤进行计算:

$$\text{预 测: } y_{n+1}^{(0)} = y_{n-3} + \frac{4}{3} h (2y'_n - y'_{n-1} + 2y'_{n-2});$$

$$\text{计算导数: } f_{n+1}^{(0)} = f(t_{n+1}, y_{n+1}^{(0)});$$

$$\text{迭代校正: } y_{n+1}^{(s+1)} = y_{n-1} + \frac{h}{3} (f_{n+1}^{(s)} + 4y'_n + y'_{n-1});$$

$$f_{n+1}^{(s+1)} = f(t_{n+1}, y_{n+1}^{(s+1)}),$$

$$s = 0, 1, 2, \dots$$

关于 Milne 方法的修改:

前面已经指出, 可以根据预测公式和校正公式的截断误差对预测-校正法加以修改, 以提高计算的精确度, 同时又避免使用迭代而减少计算工作量。具体方法如下。

根据截断误差公式(12)和(14), 容易求得校正值  $c_{n+1}$  和预测值  $p_{n+1}$  跟截断误差的关系式:

$$\begin{aligned} c_{n+1} - p_{n+1} &\approx \frac{28}{90} h^5 y^{(5)}(\xi) + \frac{h^5}{90} y^{(5)}(\xi) \\ &= \frac{29}{90} h^5 y^{(5)}(\xi), \quad t_{n-3} < \xi < t_{n+1}. \end{aligned} \quad (15)$$

于是有

$$T_1 \approx \frac{28}{29}(c_{n+1} - p_{n+1}), \quad (16)$$

$$T_2 \approx -\frac{1}{29}(c_{n+1} - p_{n+1}). \quad (17)$$

如果在预测值  $p_{n+1}$  中加上第  $n$  步对截断误差的修正项

$$\frac{28}{29}(c_n - p_n), \quad (18)$$

预测值便得到了进一步改善,把修正后的结果记为  $m_{n+1}$ , 则

$$m_{n+1} = p_{n+1} - \frac{28}{29}(p_n - c_n), \quad (19)$$

$$m'_{n+1} = f(t_{n+1}, m_{n+1}). \quad (20)$$

把  $m_{n+1}$  代入校正公式的右端,就得到校正值  $c_{n+1}$

$$c_{n+1} = y_{n-1} + \frac{h}{3}(m'_{n+1} + 4y'_n + y'_{n-1}). \quad (21)$$

根据(17)再进行一次修正,便得到了最终结果

$$y_{n+1} = c_{n+1} + \frac{1}{29}(p_{n+1} - c_{n+1}). \quad (22)$$

综上所述,修改的 Milne 预测-校正法的计算步骤可归结如下:

1) 预测:  $p_{n+1} = y_{n-3} + \frac{4}{3}h(2y'_n - y'_{n-1} + y'_{n-2});$

2) 修正:  $m_{n+1} = p_{n+1} - \frac{28}{29}(p_n - c_n),$

$$m'_{n+1} = f(t_{n+1}, m_{n+1});$$

3) 校正:  $c_{n+1} = y_{n-1} + \frac{h}{3}(m'_{n+1} + 4y'_n + y'_{n-1});$

4) 终值:  $y_{n+1} = c_{n+1} + \frac{1}{29}(p_{n+1} - c_{n+1}),$

$$y'_{n+1} = f(t_{n+1}, y_{n+1}).$$

应当指出,在开始上述计算时,必须应用适当的方法先计

算出所需要的出发值.

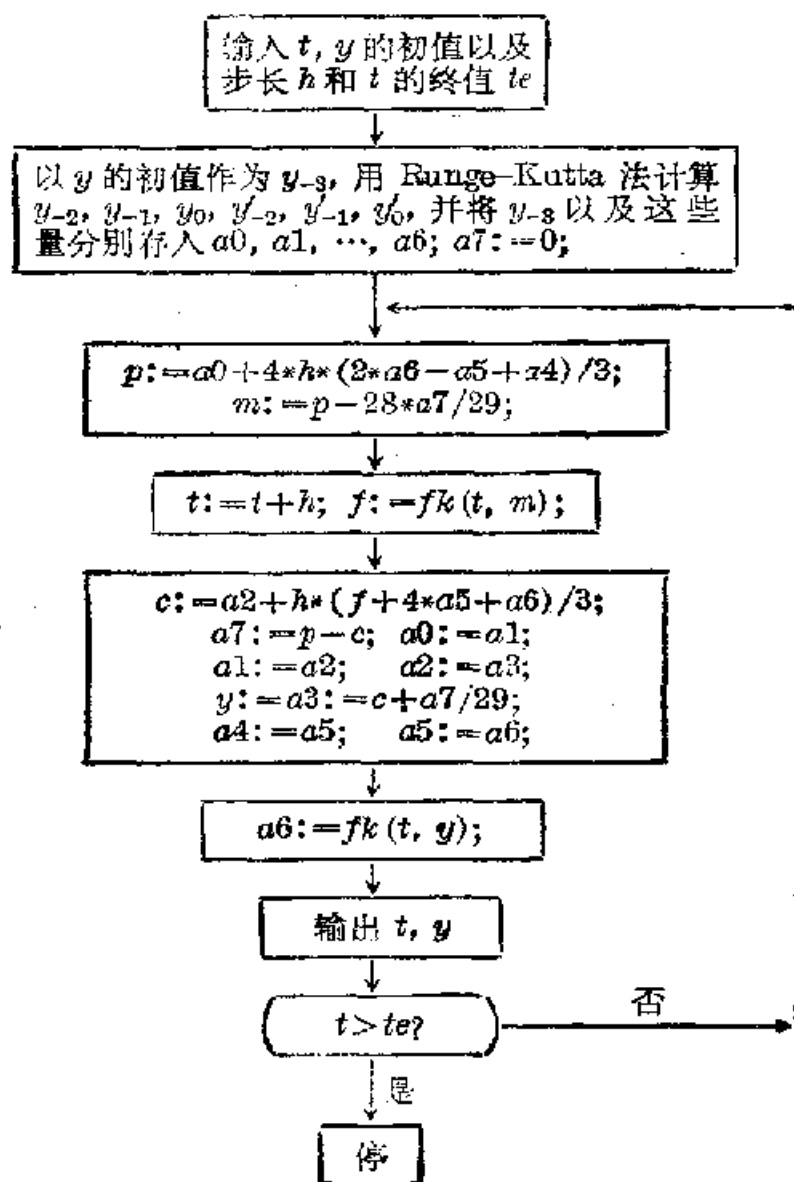
修改的 Milne 方法有下列优点:

1°. 它是一个四阶的方法, 每一步只要算两个导数值, 而同是四阶的 Runge-Kutta 方法, 则每一步却需要计算四个导数值.

2°. 由于对预测值和校正值都作了修正, 这样便减少了不使用迭代而产生的误差;

3°. 差  $p_n - c_n$  还可以用来调整步长, 使用方便.

修改的 Milne 方法的框图如下:



### § 3 Hamming 方法

**3.1. 对 Milne 方法校正公式的修改** 使用迭代的 Milne 方法, 后来发现不是(强)稳定的<sup>1)</sup>, 即在计算步数增多的时候, 舍入误差的增长很快. Hamming 改变了 Milne 方法中的校正公式, 得到了一个具有较好的数值稳定性的预测-校正法. Hamming 建立校正公式的思想如下:

把校正公式先写成比较一般的形式

$$y_{n+1} = ay_n + by_{n-1} + cy_{n-2} + h(dy'_{n+1} + ey'_n + fy'_{n-1}). \quad (23)$$

要想这个公式是四阶的, 它应对直到四次的多项式都完全准确. 分别令  $y=1, t, t^2, t^3$  和  $t^4$  代入(23), 便得到诸系数  $a, b, c, d, e$  和  $f$  所满足的方程, 解之, 便得到下列关系:

$$\begin{cases} a = \frac{27(1-b)}{24}, & d = \frac{9-b}{24}, \\ c = \frac{-3(1-b)}{24}, & e = \frac{18+14b}{24}, \\ f = \frac{-9+17b}{24}, \end{cases} \quad (24)$$

其中的  $b$  可以是任意的. 当  $b=1$  时就得到 Milne 方法中的校正公式. Hamming 分别对  $b=1, 9/17, 1/9, 0, -1/7, -9/31$  和  $-6/10$  进行了试验, 发现以  $b=0$  为最好. Ralston 就  $b$  值和稳定性的关系问题, 也作了大量实验, 结果表明,  $b=0$  接近于最好的选择.

令  $b=0$ , 据(24)便得到

1) 关于稳定性问题在第六章专门讨论.

$$\begin{cases} a = \frac{9}{8}, & b = 0, & c = -\frac{1}{8}, \\ d = \frac{3}{8}, & e = \frac{3}{4}, & f = -\frac{3}{8}. \end{cases} \quad (25)$$

于是便得到 Hamming 的校正公式:

$$y_{n+1} = \frac{1}{8} [9y_n - y_{n-2} + 3h(y'_{n+1} + 2y'_n - y'_{n-1})], \quad (26)$$

它的截断误差公式为:

$$T = -\frac{1}{40} h^5 y^{(5)}(\xi), \quad t_{n-1} < \xi < t_{n+1}. \quad (27)$$

**3.2. 修改的 Hamming 预测-校正法** 根据前一小节的推导, 得到了 Hamming 所提出的一种预测-校正法, 它的预测公式和校正公式分别为:

预测公式:

$$y_{n+1} = y_{n-3} + \frac{4}{3} h (2y'_n - y'_{n-1} + 2y'_{n-2});$$

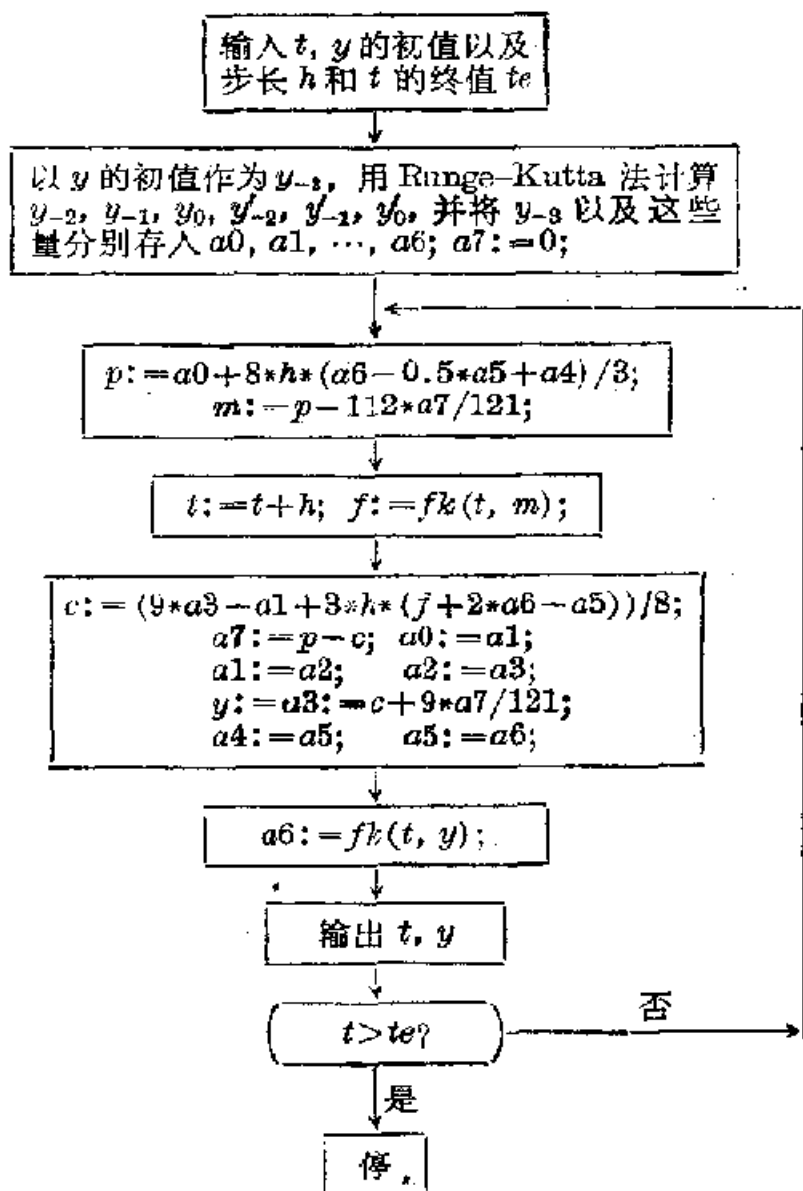
校正公式:

$$y_{n+1} = \frac{1}{8} [9y_n - y_{n-3} + 3h(y'_{n+1} + 2y'_n - y'_{n-1})].$$

按照前面介绍的据截断误差公式对预测值和校正值修正的方法, 修改的 Hamming 方法的计算步骤如下:

1. 预测:  $p_{n+1} = y_{n-3} + \frac{4}{3} h (2y'_n - y'_{n-1} + 2y'_{n-2});$
2. 修正:  $m_{n+1} = p_{n+1} - \frac{112}{121} (p_n - c_n),$   
 $m'_{n+1} = f(t_{n+1}, m_{n+1});$
3. 校正:  $c_{n+1} = \frac{1}{8} \{9y_n - y_{n-3} + 3h[m'_{n+1} - 2y'_n - y'_{n-1}]\};$
4. 终值:  $y_{n+1} = c_{n+1} + \frac{9}{121} (p_{n+1} - c_{n+1}),$   
 $y'_{n+1} = f(t_{n+1}, y_{n+1}).$

修改的 Hamming 方法的框图如下:



框图说明:

1°  $a_0, a_1, \dots, a_7$  是用来存放有关的积分信息.  $a_0$  放  $y_{i-3}$ ,  $a_1$  放  $y_{i-2}$ ,  $a_2$  放  $y_{i-1}$ ,  $a_3$  放  $y_i$ ,  $a_4$  放  $y'_{i-2}$ ,  $a_5$  放  $y'_{i-1}$ ,  $a_6$  放  $y'_i$ ,  $a_7$  放  $p_i - c_i$  (初值为零).

2°  $fk(t, y)$  为计算方程右端函数值的函数过程.





为书写方便起见, 我们可将方程组 (1) 表示成向量的形式:

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y}),$$

其中  $\mathbf{y} = [y_1, y_2, \dots, y_s]^T$ ,  $\mathbf{f} = [f_1, f_2, \dots, f_s]^T$ ; 相应的初始条件 (3) 写成:

$$\mathbf{y}(a) = \boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_s]^T.$$

**1.2. 常微分方程组的来源** 今就下列三个方面来说明常微分方程组的来源.

(a) 在理论上, 高于一阶的常微分方程都能化为一阶的常微分方程组. 设给定  $m$  阶的微分方程如下:

$$y^{(m)} = f(x, y, y', y'', \dots, y^{(m-1)}), \quad (4)$$

其中  $f$  为  $m+1$  个变元的已知函数. 令

$$y = y_1, y' = y_2, y'' = y_3, \dots, y^{(m-1)} = y_m, \quad (5)$$

于是, (4) 便化成为关于函数  $y_1, \dots, y_m$  的一阶微分方程组:

$$\begin{cases} y_1' = y_2, \\ y_2' = y_3, \\ \dots\dots\dots \\ y_m' = f(x, y_1, y_2, \dots, y_m). \end{cases} \quad (6)$$

这是 (1) 的特殊情况. 若  $y_1, y_2, \dots, y_m$  满足 (6) 则  $y(x) = y_1$  同样满足微分方程 (4).

例如, 考虑二阶微分方程

$$y'' = -y. \quad (7)$$

令

$$y = y_1, y' = y_2,$$

则 (7) 就化为方程组

$$\begin{cases} y_1' = y_2, \\ y_2' = -y_1. \end{cases}$$

(b) 在物理和力学中, 常常出现微分方程组. 例如有几个自由度的力学问题, 陀螺仪的运动, 外弹道基本方程以及火

箭的飞行等等，特别是在天体力学中，要讨论二阶微分方程组，即所谓摄动方程。

(c) 此外还能把偏微分方程化为常微分方程组，而求其近似解。例如，我们要求一个函数  $u(x, t)$  满足抛物型偏微分方程

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad 0 \leq x \leq L, \quad t \geq 0, \quad (8)$$

以及附加条件

$$u(x, 0) = f(x), \quad 0 \leq x \leq L,$$

$$u(0, t) = g(t), \quad u(L, t) = h(t), \quad (t \geq 0),$$

其中的  $f(x)$ ,  $g(t)$  和  $h(t)$  是已知函数。

对于  $t \geq 0$  和  $x = x_n = nk$ ,  $k = \frac{L}{N}$ ,  $n = 1, 2, \dots, N-1$ , 我们有

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{k^2} [u(x_{n+1}, t) - 2u(x_n, t) + u(x_{n-1}, t)].$$

令

$$u_n(t) = u(x_n, t),$$

而在  $t$  的方向保持不变。于是偏微分方程(8)就化为

$$\frac{du_n}{dt} = k^{-2} [u_{n+1} - 2u_n + u_{n-1}], \quad n = 1, 2, \dots, N-1,$$

其中

$$u_0 = g(t), \quad u_N = h(t);$$

初始条件为

$$u_n(0) = f(x_n), \quad n = 1, 2, \dots, N-1.$$

这个方法对于更为复杂的偏微分方程也可以采用。

## § 2 Runge-Kutta 方法

### 2.1. 含两个方程的一阶微分方程组 考虑微分方程组

$$\begin{cases} \frac{dy}{dx} = f(x, y(x), u(x)), \\ \frac{du}{dx} = F(x, y(x), u(x)), \end{cases} \quad (9)$$

初始条件为：当  $x = x_0$  时，

$$y = y_0, \quad u = u_0. \quad (10)$$

Runge-Kutta 公式为

$$y_{i+1} = y_i + \frac{1}{6} (k_1 + 2k_2 + 2k_3 + k_4), \quad (11)$$

其中

$$\begin{cases} k_1 = (\Delta x) f(x_i, y_i, u_i), \\ k_2 = (\Delta x) f\left(x_i + \frac{\Delta x}{2}, y_i + \frac{k_1}{2}, u_i + \frac{q_1}{2}\right), \\ k_3 = (\Delta x) f\left(x_i + \frac{\Delta x}{2}, y_i + \frac{k_2}{2}, u_i + \frac{q_2}{2}\right), \\ k_4 = (\Delta x) f(x_i + \Delta x, y_i + k_3, u_i + q_3); \end{cases} \quad (12)$$

以及

$$u_{i+1} = u_i + \frac{1}{6} (q_1 + 2q_2 + 2q_3 + q_4), \quad (13)$$

其中

$$\begin{cases} q_1 = (\Delta x) F(x_i, y_i, u_i), \\ q_2 = (\Delta x) F\left(x_i + \frac{\Delta x}{2}, y_i + \frac{k_1}{2}, u_i + \frac{q_1}{2}\right), \\ q_3 = (\Delta x) F\left(x_i + \frac{\Delta x}{2}, y_i + \frac{k_2}{2}, u_i + \frac{q_2}{2}\right), \\ q_4 = (\Delta x) F(x_i + \Delta x, y_i + k_3, u_i + q_3). \end{cases} \quad (14)$$

先将  $y$  及  $u$  的初值  $y_0, u_0$  代入微分方程 (9) 中求出  $f$  和

$F$  的初值, 然后由公式(12)及(14)的第一式便知, 只要将  $f$  和  $F$  分别乘以步长  $\Delta x$ , 便得到  $k_1$  及  $q_1$ , 然后依次从(12), (14)的第二、第三、第四个公式中分别算出  $k_2, q_2, k_3, q_3$  以及  $k_4, q_4$ ; 然后再由(11)及(13)算出在  $x_1 = x_0 + \Delta x$  处的  $y_1$  及  $u_1$  的值. 最后以  $x_1$  处的  $y_1, u_1$  为初始值, 重复上面所述的过程, 便可依次地得到  $x_i$  处的  $y_i$  及  $u_i$  值.

**2.2. 一般的一阶微分方程组** 对具有二个以上方程的情形来说, 计算的方法也是一样的. 设一阶方程组的向量表达式为

$$\begin{aligned} \mathbf{y}' &= \mathbf{f}(x, \mathbf{y}), \\ \mathbf{y} &= [y_1, y_2, \dots, y_s]^T, \mathbf{f} = [f_1, \dots, f_s]^T. \end{aligned}$$

给定初值为

$$\mathbf{y}(x_0) = \boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_s]^T.$$

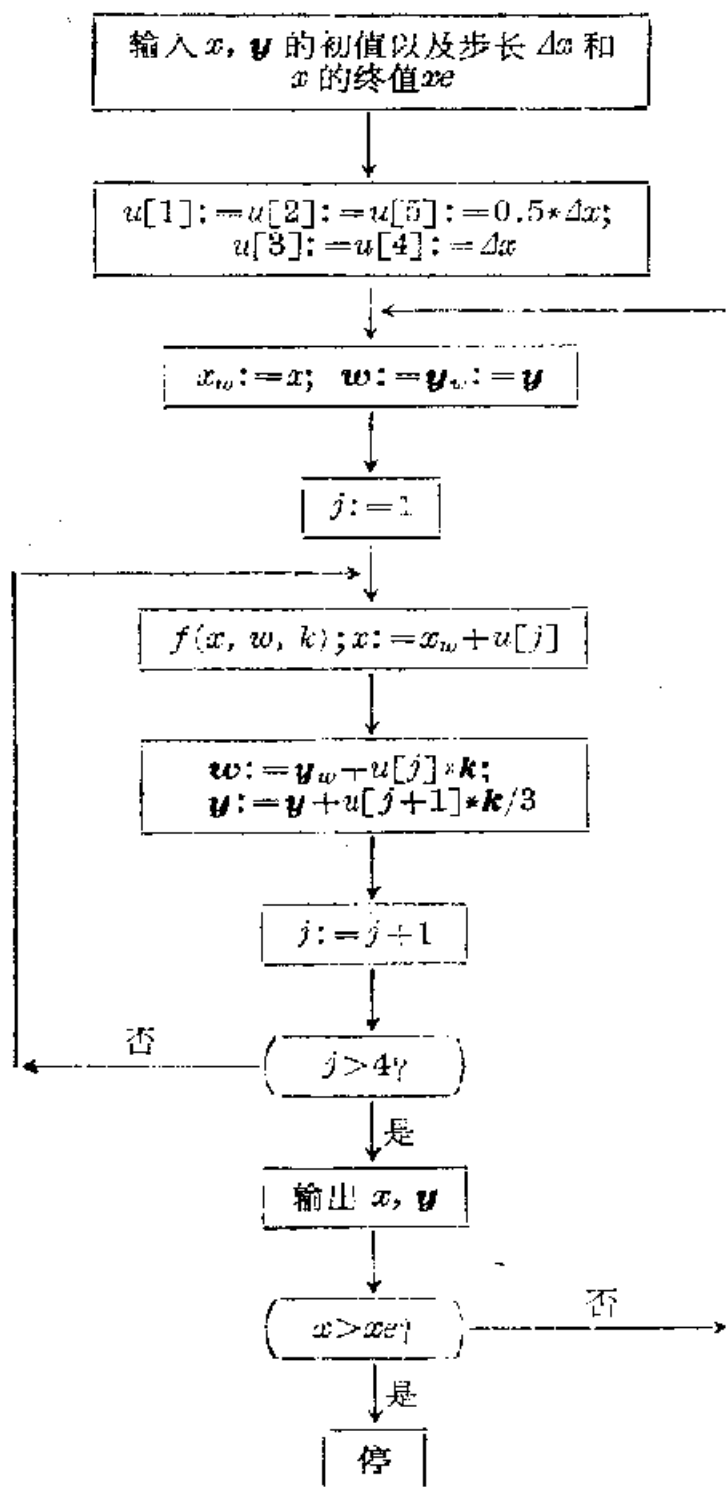
对于  $s$  个联立方程的四阶 Runge-Kutta 公式为

$$\mathbf{y}_{i+1} = \mathbf{y}_i + \frac{\Delta x}{6} (\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4), \quad (15)$$

其中

$$\begin{cases} \mathbf{k}_1 = (\Delta x) \mathbf{f}(x_i, \mathbf{y}_i), \\ \mathbf{k}_2 = (\Delta x) \mathbf{f}\left(x_i + \frac{\Delta x}{2}, \mathbf{y}_i + \frac{1}{2} \mathbf{k}_1\right), \\ \mathbf{k}_3 = (\Delta x) \mathbf{f}\left(x_i + \frac{\Delta x}{2}, \mathbf{y}_i + \frac{1}{2} \mathbf{k}_2\right), \\ \mathbf{k}_4 = (\Delta x) \mathbf{f}(x_i + \Delta x, \mathbf{y}_i + \mathbf{k}_3). \end{cases} \quad (16)$$

这里  $\mathbf{y}_{i+1}$  为微分方程在  $x_{i+1} = x_i + \Delta x$  处的数值解. 上述 Runge-Kutta 公式求解  $\mathbf{y}_1, \dots, \mathbf{y}_i$  的顺序与两个方程的情形相同, 叙述从略. 下面介绍框图:



框图说明:

$$\begin{aligned}
 (1) \text{ 注意到: } y_{i+1} &= y_i + \frac{\Delta x}{6} (k_1 + 2k_2 + 2k_3 + k_4) \\
 &= y_i + (u[2]k_1 + u[3]k_2 \\
 &\quad + u[4]k_3 + u[5]k_4) / 3
 \end{aligned}$$

其中  $u[1]=u[2]=u[5]=0.5\Delta x$ ,  $u[3]=u[4]=\Delta x$ .

(2) 框图中  $f$  为过程, 计算方程组右端函数, 其导引形式为: Procedure  $f(x, w, k)$ ; Real  $x$ ; Array  $w, k$ ; 其中  $x$  为自变量, 数组  $w[1:n]$  存放计算右端函数用的  $y$  值,  $k[1:n]$  存放计算结果, 这里  $n$  为方程个数.

**2.3. 高阶微分方程的情形** 对于  $n$  阶微分方程则可以先将它化为  $n$  个一阶微分方程, 应用含有  $n$  个公式的 Runge-Kutta 方法便可得到初值问题的数值解. 下面我们仍以二阶方程为例来加以说明. 在这种情况下, Runge-Kutta 公式可得到进一步的简化. 我们以一般的二阶微分方程为例.

设给定微分方程

$$\frac{d^2x}{dt^2} = f\left(t, x, \frac{dx}{dt}\right). \quad (17)$$

令  $v = \frac{dx}{dt}$ , 方程(17)可以化为两个一阶微分方程:

$$\begin{cases} \frac{dv}{dt} = f(t, x, v), \\ \frac{dx}{dt} = v. \end{cases} \quad (18)$$

令  $y = [v, x]^T$ ,  $f = [f, v]^T$ , 则(18)就可以表示成向量的形式

$$y' = f.$$

此时相应的 Runge-Kutta 公式(11—14)为

$$v_{i+1} = v_i + \frac{1}{6} (k_1 + 2k_2 + 2k_3 + k_4), \quad (19)$$

其中

$$\begin{cases} k_1 = (\Delta t)f(t_i, x_i, v_i), \\ k_2 = (\Delta t)f\left(t_i + \frac{\Delta t}{2}, x_i + \frac{q_1}{2}, v_i + \frac{k_1}{2}\right), \\ k_3 = (\Delta t)f\left(t_i + \frac{\Delta t}{2}, x_i + \frac{q_2}{2}, v_i + \frac{k_2}{2}\right), \\ k_4 = (\Delta t)f(t_i + \Delta t, x_i + q_3, v_i + k_3), \end{cases} \quad (20)$$

以及

$$x_{i+1} = x_i + \frac{1}{6}(q_1 + 2q_2 + 2q_3 + q_4), \quad (21)$$

其中

$$\begin{cases} q_1 = (\Delta t) F(v_i) = \Delta t(v_i), \\ q_2 = (\Delta t) F\left(v_i + \frac{k_1}{2}\right) = \Delta t\left(v_i + \frac{k_1}{2}\right), \\ q_3 = (\Delta t) F\left(v_i + \frac{k_2}{2}\right) = \Delta t\left(v_i + \frac{k_2}{2}\right), \\ q_4 = (\Delta t) F(v_i + k_3) = \Delta t(v_i + k_3). \end{cases} \quad (22)$$

我们将  $q_j (j=1, 2, 3, 4)$  代入 (20) 及 (21), 这样就可以得到进一步简化了的只含有  $k_j (j=1, 2, 3, 4)$  的 Runge-Kutta 公式:

$$\begin{cases} x_{i+1} = x_i + (\Delta t) v_i + \frac{\Delta t}{6}(k_1 + k_2 + k_3), \\ v_{i+1} = v_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \end{cases} \quad (23)$$

其中

$$\begin{cases} k_1 = \Delta t f(t_i, x_i, v_i), \\ k_2 = \Delta t f\left(t_i + \frac{\Delta t}{2}, x_i + \frac{\Delta t}{2}(v_i), v_i + \frac{k_1}{2}\right), \\ k_3 = \Delta t f\left(t_i + \frac{\Delta t}{2}, x_i + \frac{\Delta t}{2}(v_i) + \frac{\Delta t}{4}k_1, v_i + \frac{k_2}{2}\right), \\ k_4 = \Delta t f\left(t_i + \Delta t, x_i + \Delta t(v_i) + \frac{\Delta t}{2}k_2, v_i + k_3\right). \end{cases} \quad (24)$$

例 1: 跳伞员由一个在水平方向作直线飞行的飞机中跳出, 在降落伞张开前为一自由落体. 这段时间中跳伞员的速度可用矩形法则来描述 (图 5.1). 当降落伞张开时作用在跳伞员身上的反作用力与伞张开瞬间跳伞员下降的速度平方成比例, 故希望伞在自由下落速度最小时打开.

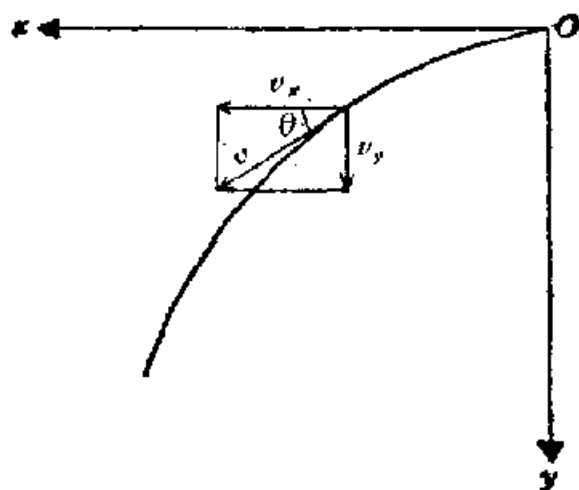


图 5.1

众所周知一个在有阻力的介质中自由下降的物体,如果允许它下降到足够远的话,最终会到达一个极限速度,即当流体介质阻力与落体的重量平衡时的速度.另一方面实践证明,对于任何一种水平降落速度,跳伞员自由地下落将经过一个既比降落速度也比极限速度更小的最小速度,故在此时将伞打开对跳伞员最为有利,即受冲击力为最小.

一般来讲,阻力  $D$  由于空气阻力而产生,它与物体在其运动垂直方向上的投影面,与空气的密度,与速度的平方成比例.通常,阻力  $D$  可表示为

$$D = \frac{c\gamma A v^2}{2}, \quad (25)$$

其中  $\gamma$ ——流体介质的物质密度(千克·秒<sup>2</sup>/米<sup>4</sup>),

$A$ ——在物体运动垂直方向上的投影面积(米<sup>2</sup>),

$c$ ——阻力系数(无量纲量),

$v$ ——速度(米/秒).

若设跳伞员的投影面积  $A$  为常数,阻力系数在所出现的速度范围内为常数,且空气密度为常数,则量  $c\gamma A$  可作为常数看待,此时

$$\frac{c\gamma A}{2} = \text{常数} = K.$$



从(25)得

$$D = K v^2.$$

跳伞员的运动服从牛顿第二定律,在直角坐标系中为

$$\begin{cases} \frac{W}{g} \frac{dv_x}{dt} = -K v^2 \cos \theta & (x \text{ 方向}), \\ \frac{W}{g} \frac{dv_y}{dt} = W - K v^2 \sin \theta & (y \text{ 方向}), \end{cases} \quad (26)$$

其中  $v_x$ ——水平分速度,

$v_y$ ——垂直分速度,

$v = \sqrt{v_x^2 + v_y^2}$ ——合速度,

$W$ ——跳伞员及全部配带物的重量,

$g$ ——重力加速度,

$K$ ——空气动力阻力系数.

为方便起见引入牛顿记号:

$\dot{x}$ ——水平分速度,

$\dot{y}$ ——垂直分速度,

$v = \sqrt{\dot{x}^2 + \dot{y}^2}$ ——合速度.

又由图 5.1 知  $v_x = v \cos \theta$ ,  $v_y = v \sin \theta$ , 并应用物体沿铅直方向降落时在极限速度时的平衡方程, 得

$$W = K (v_t)^2$$

或

$$K = W / v_t^2,$$

其中  $v_t$  为极限速度. 此时(26)式可以改写为

$$\begin{cases} \ddot{x} = -\frac{g}{v_t^2} \sqrt{\dot{x}^2 + \dot{y}^2} \dot{x} \equiv F_1(\dot{x}, \dot{y}), \\ \ddot{y} = g - \frac{g}{v_t^2} \sqrt{\dot{x}^2 + \dot{y}^2} \dot{y} \equiv F_2(\dot{x}, \dot{y}). \end{cases} \quad (27)$$

取物体下抛点为坐标系的原点,如图 5.1 所示,初始条件为

$$t=0 \begin{cases} x=0, \\ \dot{x}=134.11 \text{ 米/秒}, \\ y=0, \\ \dot{y}=0. \end{cases} \quad (28)$$

现在我们要计算出: 1) 在各个时刻由方程 (27) 及初值 (28) 所确定的  $x, y, \dot{x}, \dot{y}$ , 以及合速度  $v$ . 2) 求出在自由降落一段时间后最佳开伞时刻(即跳伞员的速度为最小时).

方程 (27) 为  $\dot{x}, \dot{y}$  的一阶微分方程组, 使用适合于方程组 (9) 的 Runge-Kutta 公式 (11) — (14) 于方程组 (27) 便得:

$$\dot{x}_{i+1} = \dot{x}_i + \frac{1}{6} (k_1 + 2k_2 + 2k_3 + k_4), \quad (29)$$

其中

$$\begin{cases} k_1 = (\Delta t) F_1(\dot{x}_i, \dot{y}_i), \\ k_2 = (\Delta t) F_1\left(\dot{x}_i + \frac{k_1}{2}, \dot{y}_i + \frac{q_1}{2}\right), \\ k_3 = (\Delta t) F_1\left(\dot{x}_i + \frac{k_2}{2}, \dot{y}_i + \frac{q_2}{2}\right), \\ k_4 = (\Delta t) F_1(\dot{x}_i + k_3, \dot{y}_i + q_3), \end{cases} \quad (30)$$

以及

$$\dot{y}_{i+1} = \dot{y}_i + \frac{1}{6} (q_1 + 2q_2 + 2q_3 + q_4), \quad (31)$$

其中

$$\begin{cases} q_1 = (\Delta t) F_2(\dot{x}_i, \dot{y}_i), \\ q_2 = (\Delta t) F_2\left(\dot{x}_i + \frac{k_1}{2}, \dot{y}_i + \frac{q_1}{2}\right), \\ q_3 = (\Delta t) F_2\left(\dot{x}_i + \frac{k_2}{2}, \dot{y}_i + \frac{q_2}{2}\right), \\ q_4 = (\Delta t) F_2(\dot{x}_i + k_3, \dot{y}_i + q_3). \end{cases} \quad (32)$$

由上面的公式可以逐个求出  $\dot{x}_i, \dot{y}_i, i=1, 2, \dots$ . 由于

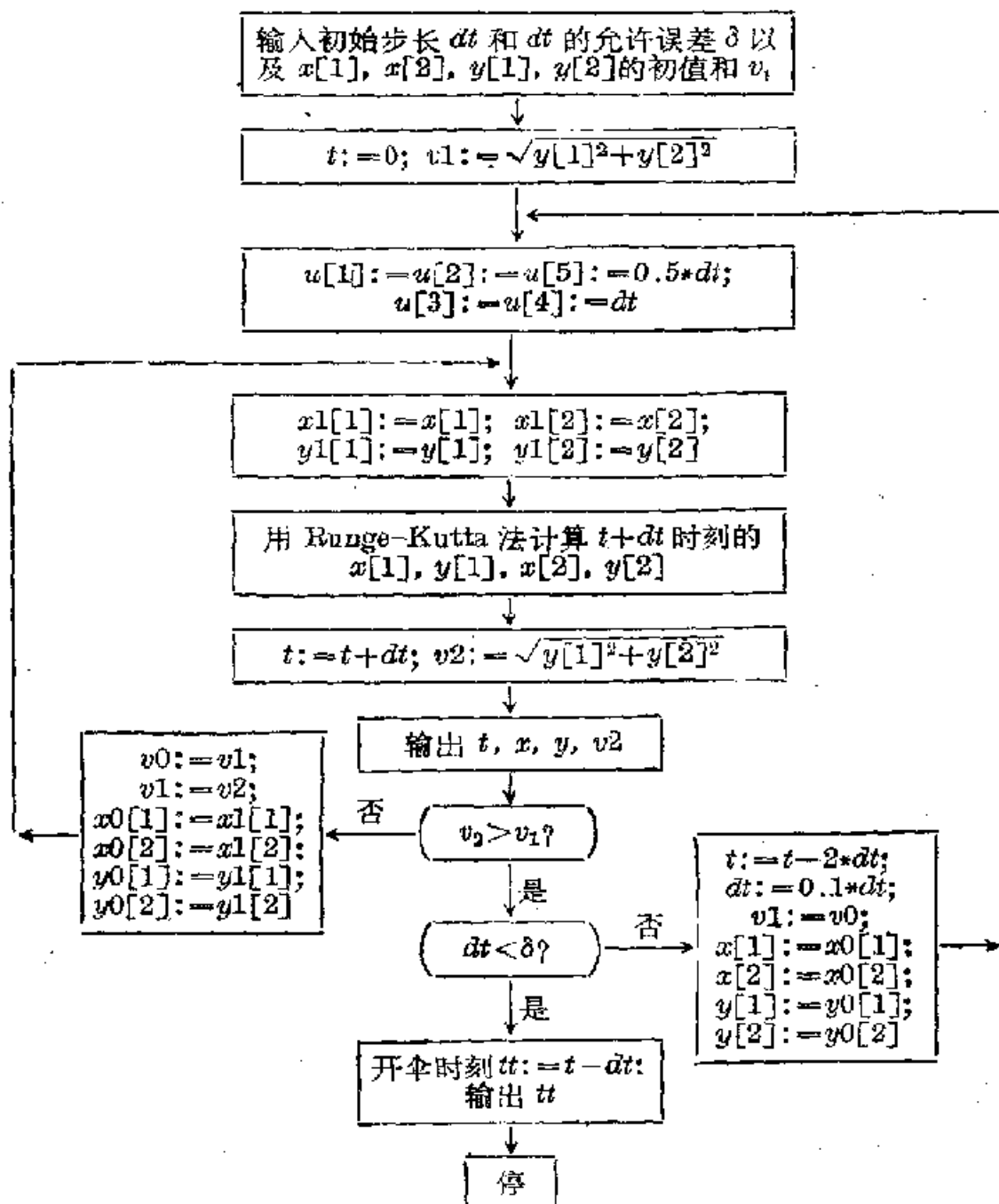
$$\frac{dx}{dt} = \dot{x}, \quad \frac{dy}{dt} = \dot{y},$$

应用(23)便得

$$x_{i+1} = x_i + (\Delta t) \dot{x}_i + \frac{\Delta t}{6} (k_1 + k_2 + k_3),$$

$$y_{i+1} = y_i + (\Delta t) \dot{y}_i + \frac{\Delta t}{6} (q_1 + q_2 + q_3).$$
(33)

框图如下:



# 框图说明:

1° 令  $y[1] = \dot{x}$ ,  $y[2] = \dot{y}$ ,  $x[1] = x$ ,  $x[2] = y$ ,  $dt$  为积分步长  $\Delta t$ .

2° 为了更精确地得到开伞时刻  $t$ , 在计算过程中出现  $v_2 > v_1$  时并不立即停止计算, 而是以  $t - 2\Delta t$  作为初始时刻并将步长缩小后再进行计算来得到  $t$ .

计算结果如下:

水平下跳速度 = 134.11 米/秒

时间	$x$ 方向的 位 移	$y$ 方向的 位 移	$x$ 方向的 分 速 度	$y$ 方向的 分 速 度	合 速 度
0.5	59.205	1.131	105.0632	4.3693	105.1540
1.0	106.749	4.257	86.3272	8.0529	86.7020
1.5	146.455	9.122	73.2025	11.3557	74.0781
2.0	180.509	15.573	63.4554	14.4167	65.0725
2.5	210.271	23.510	55.8876	17.3041	58.5052
3.0	236.641	32.854	49.7988	20.0507	53.6838
3.5	260.243	43.539	44.7535	22.6694	50.1675
4.0	281.521	55.502	40.4682	25.1603	47.6531
4.5	300.806	68.680	36.7525	27.5249	45.9171
5.0	318.347	83.004	33.4763	29.7502	44.7854
5.5	334.340	98.406	30.5478	31.8306	44.1175
6.0	348.942	114.810	27.9034	33.7601	43.7989
6.4	359.710	128.603	25.9614	35.1924	43.7321
6.5	362.283	132.140	25.4976	35.5349	43.7362
7.0	374.473	150.319	23.2975	37.1542	43.8544
7.5	385.611	169.268	21.2793	38.6199	44.0943
8.0	395.780	188.914	19.4246	39.9369	44.4103
8.5	405.060	209.132	17.7192	41.1119	44.7678
9.0	413.522	230.003	16.1513	42.1536	45.1419

(续表)

时间	$x$ 方向的 位 移	$y$ 方向的 位 移	$x$ 方向的 分 速 度	$y$ 方向的 分 速 度	合 速 度
9.5	421.233	251.315	14.7108	43.0717	45.5146
10.0	428.253	273.056	13.3886	43.8765	45.8738
10.5	434.639	295.174	12.1765	44.5787	46.2118
11.0	440.446	317.619	11.0667	45.1886	46.5240
11.5	445.722	340.349	10.0519	45.7162	46.8083
12.0	450.513	363.323	9.1250	46.1712	47.0642
12.5	454.860	386.509	8.2795	46.5621	47.2925
13.0	458.805	409.876	7.5090	46.8911	47.4944
13.5	462.381	433.395	6.8076	47.1804	47.6720
14.0	465.623	457.053	6.1605	47.4276	47.8270
14.5	468.560	480.820	5.6896	47.6354	47.9622
15.0	471.221	504.683	5.0639	47.8119	48.0792
15.5	473.631	508.627	4.5848	47.9615	48.1801
16.0	475.813	552.640	4.1510	48.0882	48.2670
16.5	477.739	576.712	3.7577	48.1953	48.3415
17.0	479.577	600.833	3.4011	48.2857	48.4053
17.5	481.195	624.995	3.0780	48.3620	48.4598
18.0	482.660	649.193	2.7853	48.4262	48.5062
18.5	483.985	673.420	2.5202	48.4803	48.5457
19.0	485.184	697.671	2.2802	48.5257	48.5793
19.5	486.269	721.944	2.0629	48.5639	48.6077
20.0	487.251	746.234	1.8662	48.5960	48.6318

上面的结果是在  $v_2 > v_1$  后将步长  $dt$  改变为原来的  $\frac{1}{5}$  再进行计算所得。如欲获得更准确的开伞时间，可按框图中所述，将时间步长改变成  $dt$  的  $\frac{1}{10}$ 。

### §3 Hamming 方法

由于 Hamming 方法具有较好的稳定性，在多步方法中它的优点较多，适应性较强。在这一节中我们将通过具体例子来说明 Hamming 方法在高阶方程上的应用。

例 2. 一个具有 1350 千克初始重量的小火箭，包括 1080

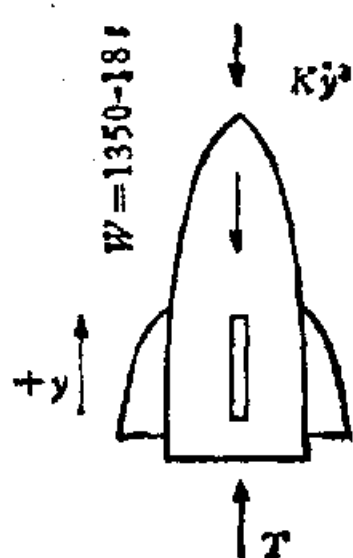


图 5.2

千克的燃料重，点燃后垂直向上运动，火箭中的燃料以每秒 18 千克的常速率燃烧，提供 3150 千克的推力，火箭上带有可记录加速度-时间，速度-时间以及位移-时间曲线的仪器。

我们希望提供一个理论上的解析模型，由它出发也可以得到上述的三种曲线，将仪器所测得的数据与理论数据进行比较，便可检验物理问题所归结的数学模型的正确程度，同时还可以用实验

数据来校正数学模型，从而修改它的假设及参量。数学模型可作如下的假设：

1. 阻力与速度的平方成比例： $D = K v^2$ ，

2. 空气动力阻力系数  $K$  取平均值  $0.039$  千克·秒<sup>2</sup>/米<sup>2</sup>。

作用在火箭上的力有：

$T$ ——推力：3150 千克，

$W$ ——在时刻  $t$  火箭燃料及总装配的总重量： $1350 - 18t$ ，

$D$ ——阻力： $0.039\dot{y}^2$  千克。

假设  $y$  向上为正，用牛顿第二定律得二阶变系数非线性微分方程：

$$\ddot{y} = \frac{gT}{W} - g - \frac{K\dot{y}^2g}{W}, \quad (34)$$

初始条件为

$$t=0 \text{ 时 } \begin{cases} y=0, \\ \dot{y}=0. \end{cases} \quad (35)$$

所需出发值可用 Runge-Kutta 方法求得. 对于二阶方程

$$\ddot{y} = f(t, y),$$

如本例中所述, Runge-Kutta 公式如下:

$$\begin{cases} y_{i+1} = y_i + (\Delta t) \dot{y}_i + \frac{(\Delta t)^2}{6} [k_1 + k_2 + k_3], \\ \dot{y}_{i+1} = \dot{y}_i + \frac{1}{6} [k_1 + 2k_2 + 2k_3 + k_4], \end{cases} \quad (36)$$

其中

$$\begin{cases} k_1 = (\Delta t) f(t_i, y_i), \\ k_2 = (\Delta t) f\left(t_i + \frac{\Delta t}{2}, y_i + \frac{k_1}{2}\right), \\ k_3 = (\Delta t) f\left(t_i + \frac{\Delta t}{2}, y_i + \frac{k_2}{2}\right), \\ k_4 = (\Delta t) f(t_i + \Delta t, y_i + k_3), \end{cases} \quad (37)$$

用上述公式的误差与  $(\Delta t)^5$  同阶.

用上一章中修改的 Hamming 方法来解这个右端不显含  $y$  的二阶微分方程所采取的公式为:

$$p(\dot{y}_{i+1}) = \dot{y}_{i-3} + \frac{4(\Delta t)}{3} [2\ddot{y}_i - \ddot{y}_{i-1} + 2\ddot{y}_{i-2}], \quad (38-1)$$

$$m(\dot{y}_{i+1}) = p(\dot{y}_{i+1}) - \frac{112}{121} [p(\dot{y}_i) - c(\dot{y}_i)], \quad (38-2)$$

$$m(\ddot{y}_{i+1}) = f[t_{i+1}, m(\dot{y}_{i+1})], \quad (38-3)$$

$$c(\dot{y}_{i+1}) = \frac{1}{8} \{9\dot{y}_i - \dot{y}_{i-2} + 3(\Delta t) [m(\ddot{y}_{i+1}) + 2\ddot{y}_i - \ddot{y}_{i-1}]\}, \quad (38-4)$$

$$\dot{y}_{i+1} = c(\dot{y}_{i+1}) + \frac{9}{121} [p(\dot{y}_{i+1}) - c(\dot{y}_{i+1})], \quad (38-5)$$

$$\ddot{y}_{i+1} = f(t_{i+1}, \dot{y}_{i+1}), \quad (38-6)$$

其中  $p$ ,  $c$  和  $m$  分别表示预测值, 校正值和修正值. (38-3) 说明以  $\bar{y}_{i+1}$  的修正值代入微分方程的右边得到同一时刻的  $\bar{y}_{i+1}$  修正值. 由于在方程中不显含  $y$ , 故在确定  $m(\bar{y}_{i+1})$  时不需要知道  $y_{i+1}$ , 因此在确定  $\bar{y}_{i+1}$  的终值时也不需知道相应的  $y_{i+1}$  值. 如果要计算位移的话, 可在每一步终了时用 Hamming 的校正公式来计算, 其公式为:

$$y_{i+1} = \frac{1}{8} \{9y_i - y_{i-2} + 3\Delta t [\bar{y}_{i+1} + 2\bar{y}_i - \bar{y}_{i-1}]\}. \quad (39)$$

为了获得  $y_{i+1}$  的可能最精确的值, 故以  $\bar{y}_{i+1}$  取代了原先校正公式中的  $m(\bar{y}_{i+1})$ . 需要注意的是: 在程序的第一个循环中所用到的  $p(\bar{y}_i)$  及  $c(\bar{y}_i)$  都取出发值  $\bar{y}_i$ .

按公式 (38), (39) 所得计算结果如第 79 页.

这里要指出, 如果二阶微分方程的右端显含有  $y$ , 则 Hamming 公式要作修改. 尚要求  $p(y_{i+1})$  及  $y_{i+1}$ , 故此时的 Hamming 公式为

$$p(y'_{i+1}) = y'_{i-3} + \frac{4\Delta t}{3} [2y''_i - y''_{i-1} + 2y''_{i-2}], \quad (40-1)$$

$$m(y'_{i+1}) = p(y'_{i+1}) - \frac{112}{121} [p(y'_i) - c(y'_i)], \quad (40-2)$$

$$p(y_{i+1}) = \frac{1}{8} \{9y_i - y_{i-2} + 3(\Delta t) [m(y'_{i+1}) + 2y'_i - y'_{i-1}]\}, \quad (40-3)$$

$$m(y''_{i+1}) = f[t_{i+1}, p(y_{i+1}), m(y'_{i+1})], \quad (40-4)$$

$$c(y'_{i+1}) = \frac{1}{8} \{9y'_i - y'_{i-2} + 3(\Delta t) [m(y''_{i+1}) + 2y''_i - y''_{i-1}]\}, \quad (40-5)$$

$$y'_{i+1} = c(y'_{i+1}) + \frac{9}{121} [p(y'_{i+1}) - c(y'_{i+1})], \quad (40-6)$$

$$y_{i+1} = \frac{1}{8} \{9y_i - y_{i-2} + 3(\Delta t) [y'_{i+1} + 2y'_i - y'_{i-1}]\}, \quad (40-7)$$

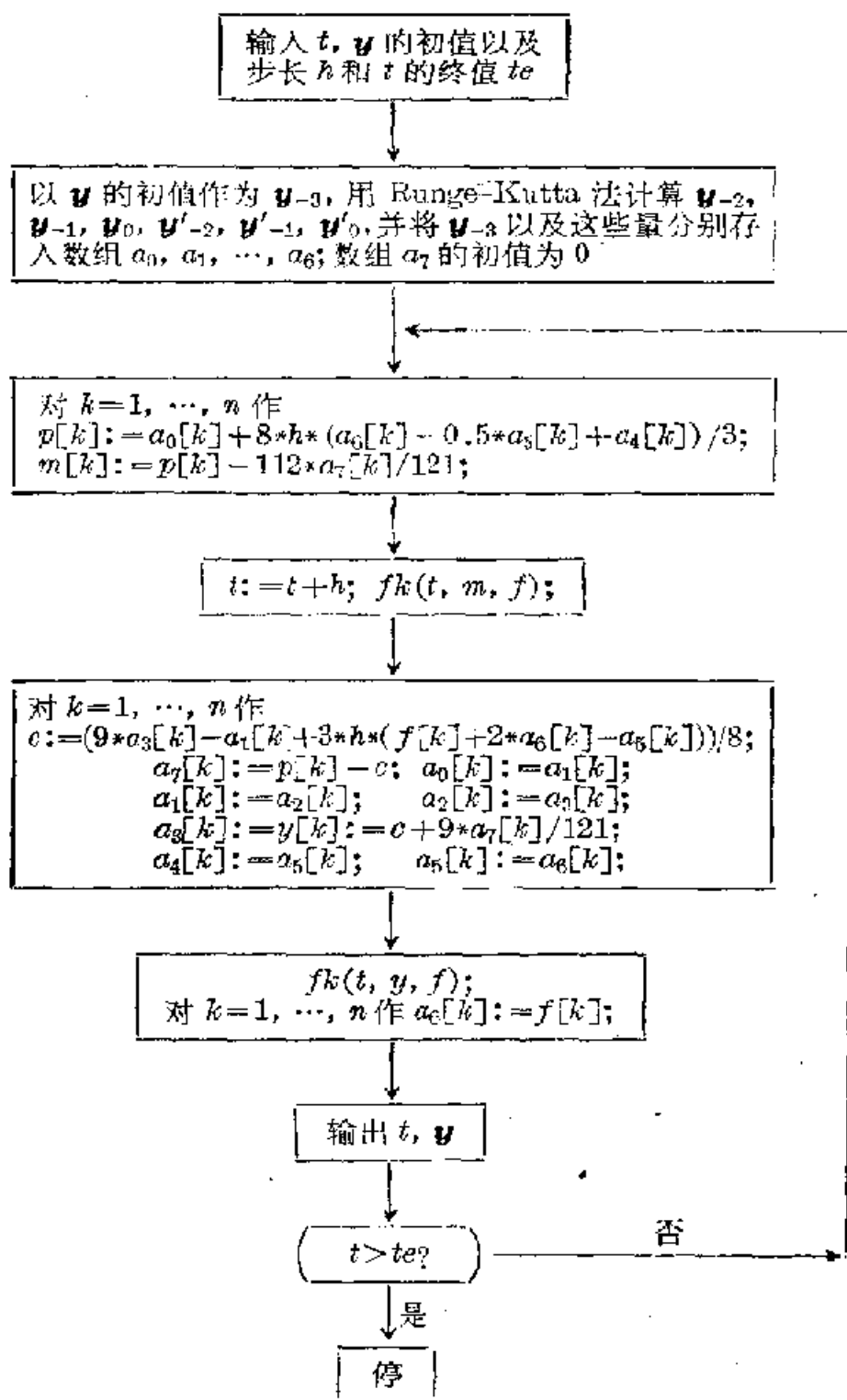
$$y''_{i+1} = f(t_{i+1}, y_{i+1}, y'_{i+1}). \quad (40-8)$$



时间	位 移	速 度	加速度	时间	位 移	速 度	加速度
0.10	0.07	1.31	13.10	30.00	4647.02	237.14	1.78
0.20	0.26	2.62	13.13	32.00	5124.68	240.44	1.54
0.30	0.59	3.93	15.15	34.00	5608.50	243.33	1.37
2.00	26.48	26.62	13.49	36.00	6097.81	245.93	1.24
4.00	106.77	53.67	13.49	38.00	6592.10	248.33	1.16
6.00	240.88	80.31	13.07	40.00	7091.05	250.59	1.10
8.00	427.15	105.69	12.26	42.00	7594.42	252.76	1.06
10.00	662.35	129.13	11.14	44.00	8102.04	254.85	1.03
12.00	942.03	150.11	9.83	46.00	8613.80	256.90	1.01
14.00	1260.93	168.39	8.44	48.00	9129.63	258.92	1.00
16.00	1613.75	183.92	7.09	50.00	9649.45	260.90	0.99
18.00	1994.92	196.84	5.85	52.00	10173.22	262.86	0.98
20.00	2399.53	207.44	4.77	54.00	10700.89	264.80	0.97
22.00	2823.34	216.04	3.86	56.00	11232.42	266.73	0.96
24.00	3262.61	222.99	3.12	58.00	11767.79	268.63	0.95
26.00	3714.44	228.64	2.55	60.00	12306.94	270.52	0.94
28.00	4176.50	233.27	2.11				

注意到公式(40-3), 以它来求出  $y_{i+1}$  的预测值, 然而由于这一点的  $m(y'_{i+1})$  是已知的, 故由 (40-3) 获得的是  $y_{i+1}$  的一个更好的预测值. 将它代入微分方程本身就得到  $y''_{i+1}$  的一个修正值. 如(40-4)所示, 而  $y_{i+1}$  的最终值仍以 Hamming 的校正公式(40-7)所获得.

至于一般的方程组, Hamming 方法的计算步骤, 则类似于第四章中所介绍的, 不同的只是其中的  $y$ ,  $p$ ,  $c$  和  $m$  都应是数组, 其分量的个数等于方程的个数, 方法的框图如下:



框图说明:

框图中,  $n$  为方程个数,  $t$  自变量,  $h$  积分步长, 数组  $y[1:n]$  积分后放终值. 数组  $a_0, a_1, \dots, a_7[1:n]$  都用来存放积分信息,  $a_0[k]$  放  $y_{k, t-3}$ ,  $a_1[k]$  放  $y_{k, t-2}$ ,  $a_2[k]$  放  $y_{k, t-1}$ ,  $a_3[k]$

放  $y_{k,i}$ ,  $a4[k]$  放  $y'_{k,i-2}$ ,  $a5[k]$  放  $y'_{k,i-1}$ ,  $a6[k]$  放  $y'_{k,i}$ ,  $a7[k]$  放  $p_{k,i} - C_{k,i}$  (初值为零), 计算函数的过程  $f^k(t, y, f)$ .

对于所述火箭问题, 只要将原方程(34)改写成一阶方程组的形成

$$y'_1 = y_2,$$

$$y'_2 = gT/W - g - Kgy^2/W.$$

令  $n=2$ , 且用另外的同阶方法求得出发值后, 是不难应用上述框图来解此问题的.

## § 4 不显含一阶导数的二阶方程的特殊计算方法

天体力学中出现的运动方程, 它的右端不出现一阶导数  $y'$ , 这种方程的一般形式为

$$y'' = f(x, y). \quad (41)$$

对于具有这种特殊形式的微分方程, 我们介绍二类公式. 一类是 Störmer 公式及 Cowell 公式. 另一类是使用非节点上的值的公式, 后者可以达到  $p=2k$  的精确度.

**4.1. Störmer 公式及 Cowell 公式的推导** 对(41)积分二次<sup>1)</sup> 我们得到

$$y(x+k) - y(x) = ky'(x) + \int_x^{x+k} (x+k-t)f(t, y(t))dt \quad (42)$$

这个公式可以看成带有积分余项的 Taylor 公式, 然而我们并不希望  $y'(x)$  在我们的公式中出现, 为此我们只要将公式中的  $k$  换成  $-k$  便得到

$$1) \text{ 利用 cauchy 积分公式 } \underbrace{\int_{x_0}^x dx \int_{x_0}^x dx \cdots \int_{x_0}^x f(x) dx}_{n \text{ 次}} = \frac{1}{(n-1)!} \int_{x_0}^x (x-s)^{n-1} f(s) ds.$$

$$y(x-k) - y(x) = -ky'(x) + \int_x^{x-k} (x-k-t)f(t, y(t))dt. \quad (43)$$

将(42), (43)相加, 并令  $f(t) = f(t, y(t))$ , 由于

$$\begin{aligned} & \int_x^{x+k} (x+k-t)f(t)dt + \int_x^{x-k} (x-k-t)f(t)dt \\ &= \int_x^{x+k} (x+k-t)[f(t) + f(2x-t)]dt, \end{aligned}$$

于是便得到恒等式

$$\begin{aligned} & y(x+k) - 2y(x) + y(x-k) \\ &= \int_x^{x+k} (x+k-t)[f(t) + f(2x-t)]dt. \end{aligned} \quad (44)$$

由(44)出发便可得到许多积分公式. 在(44)中将  $f(t)$  换成在节点  $x_p, \dots, x_{p-q}$  上的  $q$  次插值多项式, 对于不同的  $x, k$  和  $q$  的选法便能得到许多特殊的方法, 下面我们介绍常用的 Störmer 方法及 Cowell 方法.

#### (i) Störmer 方法

在公式(44)中令  $x = x_p$ ,  $x+k = x_{p+1}$ ,  $q \geq 0$  便得到公式

$$y_{p+1} - 2y_p + y_{p-1} = k^2 \sum_{m=0}^q \sigma_m \nabla^m f_p, \quad (45)$$

其中

$$\begin{aligned} \sigma_m &= \frac{(-1)^m}{k^2} \int_{x_p}^{x_{p+1}} (x_{p+1} - x) \left[ \binom{-s}{m} + \binom{s}{m} \right] dx \\ &= (-1)^m \int_0^1 (1-s) \left[ \binom{-s}{m} + \binom{s}{m} \right] ds, \quad s = \frac{x - x_p}{k}. \end{aligned} \quad (46)$$

令  $\sigma_m$  为某个母函数  $S(t)$  的 Melaurin 级数的系数, 即

$$S(t) = \sum_{m=0}^{\infty} \sigma_m t^m = \sum_{m=0}^{\infty} (-t)^m \int_0^1 (1-s) \left[ \binom{-s}{m} + \binom{s}{m} \right] ds.$$

为了得到  $\sigma_m$ , 我们交换求和与积分记号便得到

$$S(t) = \left[ \frac{t}{\log(1-t)} \right]^2 \frac{1}{1-t}. \quad (47)$$

又由

$$\begin{aligned}\frac{d}{dt} [\log(1-t)]^2 &= -\frac{2}{1-t} \log(1-t) \\ &= 2(1+t+t^2+\cdots) \left( t + \frac{1}{2}t^2 + \frac{1}{3}t^3 + \cdots \right) \\ &= 2(h_1t + h_2t^2 + h_3t^3 + \cdots),\end{aligned}\quad (48)$$

其中  $h_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$  表示调和级数的前  $m$  项部分和, 于是得到

$$\left[ \frac{\log(1-t)}{t} \right]^2 = 1 + \frac{2}{3}h_2t + \frac{2}{4}h_3t^2 + \cdots \quad (49)$$

以(49)乘(47)的两端得到

$$\left( 1 + \frac{2}{3}h_2t + \frac{2}{4}h_3t^2 + \cdots \right) (\sigma_0 + \sigma_1t + \cdots) = 1 + t + t^2 + \cdots.$$

由此便得到确定  $\sigma_m$  的递推关系式:

$$\begin{aligned}\sigma_0 &= 1 \\ \sigma_m &= 1 - \frac{2}{3}h_2\sigma_{m-1} - \frac{2}{4}h_3\sigma_{m-2} - \cdots - \frac{2}{m+2}h_{m+1}\sigma_0, \\ &\quad m=1, 2, \cdots.\end{aligned}\quad (50)$$

$\sigma_m$  的前几个值见下表:

$m$	0	1	2	3	4	5	6
$\sigma_m$	1	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{19}{240}$	$\frac{3}{40}$	$\frac{863}{12096}$

这是一个显式的公式, 只要  $y_r, \cdots, y_{r-q}$  的值知道, 便可求得  $y_{r+1}$ , 不需要迭代, 而所需的  $q+1$  个出发值, 可由单步法或 Taylor 公式来求.

为了减少工作量及增加精确度, 可选择对  $x_0$  来说为对称的出发值. 例如, 需要求出 5 个初值的话, 应该求出在  $x_{-2}, x_{-1}, x_0, x_1, x_2$  处的值. 当  $q=0$ , 及  $q=1$  时, 则 Störmer 公式便简化为下面的形式:

$$y_{p+1} - 2y_p + y_{p-1} = k^2 f_p. \quad (51)$$

(ii) Cowell 方法

令  $x = x_{p-1}$ ,  $x+k = x_p$ ,  $q \geq 2$ , 便可得

$$y_p - 2y_{p-1} + y_{p-2} = k^3 \sum_{m=0}^q \sigma_m^* \nabla^m f_p, \quad (52)$$

其中

$$\begin{aligned} \sigma_m^* &= \frac{(-1)^m}{k^2} \int_{x_{p-1}}^{x_p} (x_p - x) \left[ \binom{-s}{m} + \binom{s+2}{m} \right] dx \\ &= (-1)^m \int_{-1}^0 (-s) \left[ \binom{-s}{m} + \binom{s+2}{m} \right] ds, \end{aligned} \quad (53)$$

$\sigma_m^*$  的母函数  $S^*(t)$  的求法与求  $S(t)$  时一样, 经运算得

$$S^*(t) = \sum_{m=0}^{\infty} \sigma_m^* t^m = \left[ \frac{t}{\log(1-t)} \right]^2, \quad (54)$$

由于  $S^*(t) = (1-t)S(t)$ , 故对所有的  $m=1, 2, \dots$ , 有

$$\sigma_m^* = \sigma_m - \sigma_{m-1}. \quad (55)$$

用(49)乘(54)的两端得

$$\left(1 + \frac{2}{3} h_2 t + \frac{2}{4} h_3 t^2 + \dots\right) (\sigma_0^* + \sigma_1^* t + \sigma_2^* t^2 + \dots) = 1.$$

比较系数, 便得到递推关系式

$$\begin{aligned} \sigma_0^* &= 1 \\ \sigma_m^* &= -\frac{2}{3} h_2 \sigma_{m-1}^* - \frac{2}{4} h_3 \sigma_{m-2}^* - \dots - \frac{2}{m+2} h_{m+1} \sigma_0^*, \\ &\quad m=1, 2, \dots \end{aligned}$$

$\sigma_m^*$  前几个值见下表:

$m$	0	1	2	3	4	5	6
$\sigma_m^*$	1	-1	$\frac{1}{12}$	0	$-\frac{1}{240}$	$-\frac{1}{240}$	$-\frac{221}{60480}$

Cowell 公式不用于  $q=0$  的情况, 当  $q=1$  时, 它就是公式(51), 当  $q=2$  或  $q=3$  时便得到常用的公式

$$\begin{aligned}
 y_p - 2y_{p-1} + y_{p-2} &= h^2 \left\{ f_{p-1} + \frac{1}{12} \nabla^2 f_p \right\} \\
 &= \frac{1}{12} h^2 \{ f_p + 10f_{p-1} + f_{p-2} \}. \quad (56)
 \end{aligned}$$

当  $q \geq 2$  时 Cowell 公式为隐式的, 由于 Störmer 公式的左端与 Cowell 公式的左端一致, 故可用来作为它的预测公式.

**4.2. 含有非节点值的公式** Dyer<sup>1)</sup> 将含有非节点值可以提高方法阶数的思想用于特殊的二阶方程上, 得到精确度  $p=2k$  的公式.

他的预测公式为  $p=6, \theta=0.3$ ,

$$\bar{y}_{n+1-\theta} = \sum_{i=1}^4 \bar{\alpha}_i y_{n+1-i} + h^2 \sum_{i=1}^4 \bar{\gamma}_i y''_{n+1-i}, \quad (57)$$

其中

$$\begin{aligned}
 \bar{\alpha}_1 &= -1.9320993, & \bar{\gamma}_1 &= 0.862500525, \\
 \bar{\alpha}_2 &= 6.4719549, & \bar{\gamma}_2 &= 3.030380475, \\
 \bar{\alpha}_3 &= -3.4476119, & \bar{\gamma}_3 &= 0.433679775, \\
 \bar{\alpha}_4 &= -0.0922437, & \bar{\gamma}_4 &= -0.007217775.
 \end{aligned}$$

另一预测公式为

$$\begin{aligned}
 \bar{y}_{n+1} &= -16y_n + 34y_{n-1} - 16y_{n-2} - 1y_{n-3} \\
 &\quad + h[2.66\cdots 67y''_n + 14.66\cdots 67y''_{n-1} + 2.66\cdots 67y''_{n-2}]. \quad (58)
 \end{aligned}$$

校正公式:  $p=7$ ,

$$y_{n+1} = \sum_{i=1}^4 \alpha_i y_{n+1-i} + h^2 \sum_{i=0}^4 \gamma_i y''_{n+1-i} + h^2 \gamma y''_{n+1-\theta}, \quad (59)$$

其中

$$\begin{aligned}
 \alpha_1 &= 2.05804967222800, \\
 \alpha_2 &= -0.963457924410107, \\
 \alpha_3 &= -0.247233242848194, \\
 \alpha_4 &= 0.152641470035498,
 \end{aligned}$$

1) J. Dyer, *J. Assoc. Comput. Mach.*, **15**, 712(1968).

$$\gamma_0 = 0,$$

$$\gamma_1 = 0.704431783153855,$$

$$\gamma_2 = 0.0872587915934802,$$

$$\gamma_3 = 0.154458906599228,$$

$$\gamma = 0.1615975532325,$$

$$\gamma_4 = -0.951881093873159.$$

注意到(59)中右端下标为  $n+1$  的量的系数均为零(相当于  $\gamma_0=0$ )，所以这个校正公式实际上为显式的，所以每前进一步，只需计算二个  $f$  值。



## 第六章 数值方法的相容性、收敛性和稳定性

### § 1 单步法的相容性和收敛性

#### 1.1. 相容性条件 解一阶常微分方程初值问题

$$\begin{cases} y' = f(x, y), \\ y(a) = \eta \end{cases} \quad (1)$$

的单步法,其一般形式可以表示成

$$\begin{cases} y_{n+1} - y_n = h\Phi(x_n, y_n; h) \\ y_0 = \eta, \end{cases} \quad (2)$$

其中,  $h$  为步长,  $x_n = a + nh$ ,  $y_n$  为在点  $x_n$  处的解的近似值, 而  $\Phi(x_n, y_n; h)$  称为方法的增量函数, 它对所包含的变元是连续的. (2) 中的差分方程是由 (1) 中的微分方程经离散化而得到的, 用不同的离散化方法, 便得到不同的单步公式. 首先一个问题就是初值问题 (1) 的理论解  $y(x)$ , 在极限情形是否满足差分方程 (2) 的问题, 具体地讲, 当  $h \rightarrow 0$  时, 对求解区间中的任一固定的  $x$ , 是否存在有极限关系

$$\lim_{h \rightarrow 0} \frac{y(x+h) - y(x)}{h} = \lim_{h \rightarrow 0} \Phi(x, y(x); h).$$

显然, 由于上式左端的极限为  $y'(x)$ , 它应等于  $f(x, y(x))$ , 从而可知, 要想上述极限关系成立, 必须

$$\Phi(x, y; 0) = f(x, y). \quad (3)$$

如果这个关系成立, 我们称由 (2) 所确定的单步法与微分方程初值问题 (1) 为相容, 并称 (3) 为相容性条件.

**1.2. 相容性和方法的阶数** 设单步公式(2)为  $p$  阶的,  $y(x)$  为(1)的解, 则应有

$$y(x+h) - y(x) = h\Phi(x, y; h) + O(h^{p+1}). \quad (4)$$

如果单步公式(2)满足相容性条件, 据(4)可知

$$y'(x) = f(x, y) + \lim_{h \rightarrow 0} \frac{1}{h} O(h^{p+1}),$$

所以必有

$$\lim_{h \rightarrow 0} O(h^p) = 0.$$

这就是说, 单步公式(2)的阶数  $p$  必须为正数, 即  $p > 0$ . 由于我们所考虑的方法的阶数都是正整数, 所以  $p$  至少应为 1, 即满足相容性条件的方法, 其阶数至少为 1.

**1.3. 单步法的收敛性及其和相容性之间的关系** 单步公式(2)是用来计算初值问题(1)的近似解而提出来的, 如果(2)不满足相容性条件, 由它所算出的结果就不可能是(1)的近似解. 在(2)满足相容性条件的情况下, 还必须考虑单步法(2)的解在极限情形下是否收敛于微分方程初值问题(1)的解的问题. 关于这个问题, 我们有:

**定理 1** 设增量函数  $\Phi(x, y; h)$  在区域

$$a \leq x \leq b, \quad -\infty < y < \infty, \quad 0 \leq h \leq h_0 \quad (h_0 > 0)$$

中为连续, 并对变量  $y$  满足 Lipschitz 条件, 即存在有与  $x, h$  无关的常数  $L$ , 使不等式

$$|\Phi(x, y_2; h) - \Phi(x, y_1; h)| \leq L|y_2 - y_1|$$

在上述区域中均成立. 在这个前提下, 方法(2)收敛的充分必要条件是, 相容性条件成立.

先证相容性条件成立时, 方法(2)必收敛.

令  $\Phi(x, y; 0) = g(x, y)$ , 则  $g(x, y)$  在所述区域中为  $(x, y)$  的连续函数并对  $y$  满足 Lipschitz 条件. 因此, 根据微分方程的解的存在唯一性定理, 初值问题

$$\begin{cases} z' = g(x, z), \\ z(a) = \eta \end{cases} \quad (5)$$

对任意给定的  $\eta$ , 都存在有唯一的解  $z(x)$ , 现在我们来证明, 差分方程初值问题

$$\begin{cases} z_{n+1} = z_n + h\Phi(x_n, z_n; h) \\ z_0 = \eta \end{cases} \quad (6)$$

的解  $z_n$ , 当  $a + nh = x$ ,  $n \rightarrow \infty$  时收敛于 (5) 的解  $z(x)$ .  $z(x_n)$  满足

$$z(x_{n+1}) - z(x_n) = h\Delta(x_n, z(x_n); h), \quad (7)$$

令

$$z_n - z(x_n) = e_n.$$

(6) - (7), 便得

$$e_{n+1} = e_n + h[\Phi(x_n, z_n; h) - \Delta(x_n, z(x_n); h)]. \quad (8)$$

据微分学中值定理,

$$\begin{aligned} \Delta(x_n, z(x_n); h) &= \frac{z(x_{n+1}) - z(x_n)}{h} = z'(x_n + \theta h) \\ &= g(x_n + \theta h, z(x_n + \theta h)), \quad 0 < \theta < 1. \end{aligned} \quad (9)$$

方程 (8) 右端方括号中的差可以写成

$$\begin{aligned} &\Phi(x_n, z_n; h) - \Delta(x_n, z(x_n); h) \\ &= \Phi(x_n, z_n; h) - \Phi(x_n, z(x_n); h) \\ &\quad + \Phi(x_n, z(x_n); h) - \Phi(x_n, z(x_n); 0) \\ &\quad + g(x_n, z(x_n)) - g(x_n + \theta h, z(x_n + \theta h)), \end{aligned}$$

由于函数  $\Phi(x, y; h)$  在有界闭集  $x \in [a, b]$ ,  $y = z(x)$  上是连续的, 所以在这个有界闭集上也必为一致连续, 所以, 当  $h \rightarrow 0$  时

$$\Omega(h) = \max_{a \leq x \leq b} |\Phi(x, z(x); h) - \Phi(x, z(x); 0)| \rightarrow 0,$$

同样, 当  $h \rightarrow 0$  时

$$\omega(h) = \max_{\substack{a \leq x \leq b \\ 0 \leq h' \leq h}} |g(x, z(x)) - g(x + h', z(x + h'))| \rightarrow 0.$$

另外, 由于  $\Phi(x, y; h)$  对于变量  $y$  满足 Lipschitz 条件, 故得

$$|e_{n+1}| \leq |e_n| + h[L|e_n| + \Omega(h) + \omega(h)].$$

令  $1+hL=A$ ,  $h(\Omega(h)+\omega(h))=B$ , 据第一章中的引理可知

$$|e_n| \leq (\Omega(h) + \omega(h)) \frac{(1+hL)^n - 1}{L}.$$

由于  $1+hL \leq e^{hL}$ ,  $x_n - a = nh$ , 故上式又可改写为

$$|e_n| \leq (\Omega(h) + \omega(h)) \frac{e^{L(x_n-a)} - 1}{L}.$$

从而可知, 当  $h \rightarrow 0$  时,  $e_n \rightarrow 0$ , 这就是说  $z_n \rightarrow z(x)$ , 因为相容性条件成立, 所以  $g(x, y) = f(x, y)$ , 从而  $z(x) = y(x)$ . 这就证明了方法(2)的收敛性.

再证必要性. 设由  $\Phi(x, y; h)$  所定义的方法收敛, 而在某个点  $(x, y)$  处  $g(x, y) \neq f(x, y)$ , 则由存在定理知有这样的  $\eta$  使初值问题的解  $y(t)$  通过  $(x, y)$ , 而由(6)所定义的  $z_n$  一方面收敛于  $y(t)$ , 一方面由上述证明知  $z_n$  收敛于(5)的解  $z(t)$ , 若  $z(x) \neq y(x)$ , 立得矛盾, 若  $z(x) = y(x)$ , 则因  $z'(x) = g(x, y) \neq f(x, y) = y'(x)$ , 故  $z(t) \neq y(t)$ . 这是由假设在某点  $(x, y)$  处  $g(x, y) \neq f(x, y)$  带来的矛盾. 所以相容条件成立.

## §2 多步法的相容性和收敛性

**2.1. 多步法的相容性条件**  $k$  步公式的一般形式为

$$\alpha_k y_{n+k} + \cdots + \alpha_0 y_n = h(\beta_k f_{n+k} + \cdots + \beta_0 f_n), \quad (10)$$

其中,

$$\alpha_k \neq 0 \quad |\alpha_0| + |\beta_0| > 0.$$

从前节对单步法的讨论可知, 要想(10)和微分方程(11)相容, 则其阶数至少为 1, 即, 对  $y=1$ ,  $y=x$  时, (10)应准确成立. 令  $y=1$  得  $y'=0$ , 所以

$$\alpha_k + \cdots + \alpha_0 = 0. \quad (11)$$

令  $y=x$  得  $y'=1$ , 所以

$$\alpha_k(x+kh) + \cdots + \alpha_1(x+h) + \alpha_0 x = h(\beta_k + \cdots + \beta_0),$$

即

$$k\alpha_k + \cdots + \alpha_1 = \beta_k + \cdots + \beta_0. \quad (12)$$

所以, 我们称(11)—(12)为多步公式(10)的相容性条件.

**2.2. 多步法的收敛性问题**  $k$  步法需要  $k$  个出发值, 而初值问题(1)只提供一个

$$y_0 = y(a) = \eta = \eta_0.$$

在使用  $k$  步法时尚需用其它单步方法补充  $k-1$  个出发值, 今假定它们是

$$y_\mu = \eta_\mu(h), \quad \mu = 1, 2, \dots, k-1.$$

当  $h \rightarrow 0$  时, 这  $k-1$  个出发值都应收敛于共同的极限  $\eta$ , 即

$$\lim_{h \rightarrow 0} y_\mu = \lim_{h \rightarrow 0} \eta_\mu(h) = \eta,$$

$$\mu = 1, 2, \dots, k-1. \quad (13)$$

在考虑  $k$  步法的收敛性时, 我们假定(13)成立. 在这个前提下, 如果对任意的  $x \in [a, b]$ ,  $a + nh = x$ , 由  $k$  步公式计算的结果  $y_n$  随  $h \rightarrow 0$  而趋于初值问题(1)的解  $y(x)$ , 则称  $k$  步法是收敛的, 为了讨论  $k$  步法的收敛性问题, 需要用到多项式

$$\rho(\xi) = \alpha_k \xi^k + \cdots + \alpha_0, \quad (14)$$

$$\sigma(\xi) = \beta_k \xi^k + \cdots + \beta_0. \quad (15)$$

我们称(14)为  $k$  步公式(10)的特征多项式. 显然, 相容性条件(11)—(12)可以表示成

$$\rho(1) = 0, \quad \rho'(1) = \sigma(1). \quad (16)$$

**定理 2**  $k$  步法(10)收敛的充分必要条件为: 1) 相容性条件(16)成立; 2) 特征多项式(14)的零点的模不能大于 1,

并且在单位圆上的零点只能是单零点.

我们称 2) 为特征根条件.

### § 3 数值稳定性问题

**3.1. 稳定性概念** 在实际进行计算时,一方面出发值不一定是完全精确的,带有一定的误差,同时,由于计算机的字长有限,在运算中一般总会产生舍入误差,不论是单步法或多步法,在逐次计算下去的时候,初始数据的误差(或称之为摄动)以及在计算过程中产生的舍入误差,都会传播下去,对以后的计算结果产生影响.所谓稳定性问题,就是指误差的积累是否受到控制的问题.粗略地说,如果计算结果对初始数据的误差以及计算过程中的舍入误差不敏感,就说相应的计算方法是稳定的,否则就称之为不稳定的.现在,我们就  $k$  步法的情形对稳定性问题具体地加以阐述.

设  $y_j$  为  $k$  步法的理论结果,它们应严格满足方程

$$\sum_{j=0}^k \alpha_j y_{n+j} = h \sum_{j=0}^k \beta_j f(x_{n+j}, y_{n+j}), \quad (17)$$

而实际计算得到的结果  $\bar{y}_j$ , 所满足的方程则为

$$\sum_{j=0}^k \alpha_j \bar{y}_{n+j} = h \sum_{j=0}^k \beta_j f(x_{n+j}, \bar{y}_{n+j}) + \bar{\eta}_{n+k}, \quad (18)$$

其中的  $\bar{\eta}_{n+k}$  为第  $n+1$  步(注意由  $k$  步法所计算的第一个值为  $y_k$ )所产生的摄动.对于给定的任意正数  $\varepsilon$ , 如果总存在相应的正数  $\delta = \delta(\varepsilon)$ , 使得对给定的(1)的右端函数  $f(x, y)$ , 当

$$\sum_{j=0}^{k-1} |\bar{y}_j - y_j| + \sum_{j=k}^N |\bar{\eta}_j| \leq \delta, \quad 0 \leq h \leq h_0 \quad (19)$$

时,恒有

$$|\bar{y}_j - y_j| \leq \varepsilon, \quad j = k, \dots, N, \quad (20)$$

其中,  $a + Nh \leq b$ , 我们就说  $k$  步公式(2)对给定的  $f(x, y)$  是稳定的.

**3.2. 稳定性条件** 在这种稳定性的意义下, 我们有

**定理 3** 设  $f(x, y)$  在区域  $a \leq x \leq b, -\infty < y < \infty$  中为  $x, y$  的连续函数, 并对  $y$  满足 Lipschitz 条件, 则  $k$  步公式(10)对所有这样的  $f(x, y)$  为稳定的充分必要条件是  $k$  步公式满足特征根条件.

**3.3. 收敛性、稳定性和相容性之间的关系** 根据定理 2—3 所阐明的结论, 实际上在相容性、收敛性和稳定性之间, 存在着以下的等价性关系:

**定理 4** 对于前述函数类  $f(x, y)$ , 在相容性条件成立的前提下,  $k$  步法(2)的收敛性和稳定性是等价的<sup>1)</sup>.

推论: 在相容性条件成立时, 收敛性和稳定性的充分必要条件都是特征根条件成立.

根据这个定理, 在讨论具体的方法时就比较方便了, 例如,  $q=3$  的 Adams 外推公式为

$$y_{n+1} = y_n + \frac{h}{24} (55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}),$$

所以

$$\rho(\xi) = \xi^4 - \xi^3, \quad \sigma(\xi) = \frac{1}{24} (55\xi^3 - 59\xi^2 + 37\xi - 9).$$

易知  $\rho(1) = 0, \rho'(1) = \sigma(1)$ , 从而可知相容性条件成立. 另外, 由于特征多项式  $\rho(\xi)$  有三个零点是 0, 有一个是 1, 它在单位圆上, 但为单重的, 所以也满足特征根条件. 所以, 对于满足定理 3 条件的函数类  $f(x, y)$ , Adams 方法是收敛的和稳定的.

1) 推论: 相容性和稳定性与收敛性互为充要.

今再举一例以说明收敛性和稳定性之间的关系, 设用多步公式

$$-\frac{1}{2}y_{n+2} + 2y_{n+1} - \frac{3}{2}y_n = hf_n$$

来解初值问题

$$y' = y, \quad y(0) = 1.$$

相应的特征多项式为

$$\rho(\xi) = -\frac{1}{2}\xi^2 + 2\xi - \frac{3}{2},$$

它的两个零点是 1 和 3. 根据定理 2—3 可知, 所述方法是不稳定的也是不收敛的. 而实际上, 初值问题的解很容易求出, 它就是  $y(x) = e^x$ . 取步长  $h = 0.1$ , 出发值  $y_0 = 1.00000$ ,  $y_1 = 1.10517$ , 用所述方法计算的结果及其和理论解的误差见下表:

$x_n$	$y_n$	$y_n - e^{x_n}$
0	1.00000	0.00000
0.1	1.10517	0.00000
0.2	1.22068	-0.00072
0.3	1.34618	-0.00368
0.4	1.47854	-0.01329
0.5	1.60638	-0.04234
0.6	1.69419	-0.12993
0.7	1.63634	-0.37741
0.8	1.12395	-1.10159
0.9	-0.74049	-3.20009
1.0	-6.55860	-9.27688

从上表可以看出, 误差增长很快. 这个具体例子说明, 不稳定的方法也不会收敛.



## §4 绝对稳定性

**4.1. 绝对稳定性概念** 前一节中考虑的稳定性概念,是在  $h \rightarrow 0$  的情况下讨论的. 这样的稳定性称之为渐近稳定性 (或古典的稳定性). 然而, 实际上我们是取有限的固定步长  $h$  进行计算的, 它并不能随意地缩小. 因此, 重要的是, 在计算过程中所产生的摄动对以后的计算结果的影响不会步步增长. 这种稳定性概念, 就是通常所说的绝对稳定性.

**4.2. 单步法的绝对稳定性和绝对稳定区** 在考虑方法的绝对稳定性时, 一般只限于典型方程

$$y' = \lambda y. \quad (21)$$

为了保证方法为绝对稳定, 步长  $h$  和  $\lambda$  值都要受到一定的限制. 它们的允许范围, 就称为相应方法的绝对稳定区. 今以单步法为例, 说明如下.

例 1. Euler 法

$$y_{n+1} = y_n + hf_n, \quad (22)$$

这时,

$$y_{n+1} = (1 + \lambda h)y_n, \quad (23)$$

它的特征多项式为  $\xi - (1 + \lambda h)$ , 特征根为

$$\mu_1 = 1 + \lambda h.$$

由此可见, 要想 Euler 法为绝对稳定, 则应有

$$|\mu_1| = |1 + \lambda h| \leq 1.$$

$\lambda h$  的值, 从  $|\mu_1| < 1$  到  $|\mu_1| > 1$ , 必经过  $|\mu_1| = 1$ , 因而使  $|\mu_1| = 1$  的  $\lambda h$  值就构成绝对稳定区的边界.

象 Euler 法那样, 具有有限的绝对稳定区的边界, 我们称之为条件稳定的.

一个差分方法在应用于某一个微分方程时, 没有有限的

绝对稳定区的边界, 便称之为无条件稳定的.

例如, 一个多步法在用于微分方程  $y' = \lambda y$  时 ( $\lambda$  为一复数, 它的实部  $\operatorname{Re}(\lambda) < 0$ ), 若对固定的  $h$ , 当  $n \rightarrow \infty$  时多步法的解都趋于零, 这个多步法就是无条件稳定的.

一个多步法具有的上述无条件稳定性, 也称为  $A$ -稳定性.

例如, 后退 Euler 公式

$$y_{n+1} = y_n + h y'_{n+1},$$

在用于  $y' = \lambda y$ , ( $\operatorname{Re}(\lambda) < 0$ ) 时, 得到

$$y_{n+1} = \frac{1}{1 - \lambda h} y_n,$$

它的特征根

$$\mu_1 = \frac{1}{1 - \lambda h}.$$

因为  $\operatorname{Re}(\lambda) < 0$ , 所以恒有

$$|\mu_1| < 1,$$

从而可知, 后退 Euler 法对所述微分方程为  $A$ -稳定的.

我们称微分方程  $y' = \lambda y$ ,  $\operatorname{Re}(\lambda) < 0$ , 具有固有稳定性.

又如, 梯形公式

$$y_{n+1} = y_n + \frac{h}{2} (y'_{n+1} + y'_n),$$

在用于具有固有稳定性的微分方程时, 由于

$$y_{n+1} = \frac{1 + \frac{1}{2}\lambda h}{1 - \frac{1}{2}\lambda h} y_n,$$

所以

$$|\mu_1| = \left| \frac{1 + \frac{1}{2}\lambda h}{1 - \frac{1}{2}\lambda h} \right| < 1,$$

即梯形公式对固有稳定的微分方程也是  $A$ -稳定的.

**4.3. 微分方程组的情形** 今以含两个因变量的一阶线性常微分方程组为例, 来讨论差分方法的稳定性问题, 设

$$\begin{cases} \mathbf{y}' = A\mathbf{y}, \\ \mathbf{y}(0) = \mathbf{y}_0. \end{cases} \quad (24)$$

为给定的二维线性常系数微分方程组, 其中  $A = [a_{ij}]$  是一个  $2 \times 2$  阶矩阵, 我们假定它的两个特征值为  $\lambda_1$  和  $\lambda_2$ , 并且存在与其相应的线性独立的特征向量  $\mathbf{u} = [u_1, u_2]^T, \mathbf{v} = [v_1, v_2]^T$ , 即

$$A\mathbf{u} = \lambda_1\mathbf{u}, \quad A\mathbf{v} = \lambda_2\mathbf{v}. \quad (25)$$

设  $\mathbf{y} = [y_1(x), y_2(x)]^T$  为 (24) 的解, 则它可以唯一地表示成  $\mathbf{u}$  和  $\mathbf{v}$  的线性组合

$$\mathbf{y} = c_1\mathbf{u} + c_2\mathbf{v}, \quad (26)$$

其中,  $c_1, c_2$  为  $x$  的函数, 代入原微分方程得

$$\mathbf{y}' = c_1'\mathbf{u} + c_2'\mathbf{v} = A\mathbf{y} = c_1\lambda_1\mathbf{u} + c_2\lambda_2\mathbf{v}. \quad (27)$$

据此得

$$c_1' = \lambda_1 c_1 \quad c_2' = \lambda_2 c_2, \quad (28)$$

从而有

$$\begin{aligned} c_1(x) &= e^{\lambda_1 x} c_1(0), \\ c_2(x) &= e^{\lambda_2 x} c_2(0). \end{aligned} \quad (29)$$

代入 (26) 得<sup>1)</sup>

$$\mathbf{y} = c_1(0)e^{\lambda_1 x}\mathbf{u} + c_2(0)e^{\lambda_2 x}\mathbf{v}, \quad (30)$$

1) 这个表达式的推导可以直接从微分方程初值问题 (24) 的解  $\mathbf{y} = e^{Ax}\mathbf{y}_0$  着手, 因为  $\mathbf{y}_0$  可唯一地表示成

$$\mathbf{y}_0 = c_1\mathbf{u} + c_2\mathbf{v},$$

所以

$$\mathbf{y}' = e^{Ax}(c_1\mathbf{u} + c_2\mathbf{v}).$$

由于矩阵  $e^{Ax}$  的特征值为  $e^{\lambda_1 x}$  和  $e^{\lambda_2 x}$ , 相应的特征向量仍为  $\mathbf{u}$  和  $\mathbf{v}$ , 故得

$$\mathbf{y} = c_1 e^{\lambda_1 x} \mathbf{u} + c_2 e^{\lambda_2 x} \mathbf{v}.$$

其中  $c_1(0)$  和  $c_2(0)$  满足初始条件

$$c_1(0)u + c_2(0)v = y_0. \quad (31)$$

今考虑用 Euler 法解(24):

$$\begin{aligned} y_{1n+1} &= y_{1n} + h(a_{11}y_{1n} + a_{12}y_{2n}), \\ y_{2n+1} &= y_{2n} + h(a_{21}y_{1n} + a_{22}y_{2n}), \end{aligned} \quad (32)$$

其中,  $y_{1n}$ ,  $y_{2n}$  以及  $y_{1n+1}$ ,  $y_{2n+1}$  分别表示  $y_n$  和  $y_{n+1}$  的分量. 令(32)的解为

$$y_{1n} = w_1\mu^n, \quad y_{2n} = w_2\mu^n, \quad (33)$$

代入(32)得到

$$\begin{aligned} a_{11}w_1 + a_{12}w_2 &= -(1-\mu)w_1/h, \\ a_{21}w_1 + a_{22}w_2 &= -(1-\mu)w_2/h. \end{aligned} \quad (34)$$

若令

$$\frac{1-\mu}{h} = -\lambda, \quad \mu = 1 + \lambda h, \quad (35)$$

于是, (34)就化为

$$Aw = \lambda w, \quad (36)$$

其中,  $w = [w_1, w_2]^T$ , 这说明由(35)所规定的  $\lambda$  为矩阵  $A$  的特征值, 而  $w$  为与之相应的特征向量, 据前面所述,  $A$  的两个特征值为  $\lambda_1$  和  $\lambda_2$ ,  $u$  和  $v$  为与之相应的特征向量, 故可取  $w_1 = u$ ,  $w_2 = v$ , 从而得 Euler 法(32)的解为

$$\begin{aligned} y_{1n} &= c_1u_1(1 + \lambda_1h)^n + c_2v_1(1 + \lambda_2h)^n, \\ y_{2n} &= c_1u_2(1 + \lambda_1h)^n + c_2v_2(1 + \lambda_2h)^n. \end{aligned} \quad (37)$$

常系数  $c_1$  和  $c_2$  由初始条件所确定, 它们应与(30)中的  $c_1(0)$  和  $c_2(0)$  相同. 易知

$$1 + \lambda_1h \approx e^{\lambda_1h}, \quad 1 + \lambda_2h \approx e^{\lambda_2h}. \quad (38)$$

对固有稳定的微分方程,  $\operatorname{Re}(\lambda_1) < 0$ ,  $\operatorname{Re}(\lambda_2) < 0$ , 欲 Euler 法稳定, 则必须

$$|1 + \lambda_1h| < 1, \quad |1 + \lambda_2h| < 1, \quad (39)$$

否则, 当  $n \rightarrow \infty$  时, 对固定的步长  $h$ ,  $y_n$  将不趋于零. 据(39)可知, 绝对值较大的特征值决定步长  $h$  的选取, 而方法的精确性则决定于(38)中两个近似式的逼近程度, 就中与特征值较大者相应的近似程度较差.

4.4. 关于多步法的绝对稳定性 多步公式的一般形式为

$$\alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n + h \{ \beta_k y'_{n+k} + \cdots + \beta_0 y'_n \} = 0. \quad (40)$$

设所考虑的方程为  $y' = \lambda y$ , 则(40)式便可以表示成

$$\rho(E) y_n + \lambda h \sigma(E) y_n = 0, \quad (41)$$

其中

$$\begin{aligned} \rho(\xi) &= \alpha_k \xi^k + \alpha_{k-1} \xi^{k-1} + \cdots + \alpha_0, \\ \sigma(\xi) &= \beta_k \xi^k + \beta_{k-1} \xi^{k-1} + \cdots + \beta_0. \end{aligned}$$

$E$  表示移位算子, 其作用为  $E y_n = y_{n+1}$ ,  $E^k y_n = y_{n+k}$ , 由于有舍入误差, 所以实际数值解  $\tilde{y}_n$  满足的方程并非(41)而是

$$\rho(E) \tilde{y}_n + \lambda h \sigma(E) \tilde{y}_n + \tilde{\eta}_n = 0, \quad (42)$$

从(42)减去(41)得

$$\rho(E) [\tilde{y}_n - y_n] + \lambda h \sigma(E) [\tilde{y}_n - y_n] + \tilde{\eta}_n = 0.$$

为便于讨论, 我们假定  $\tilde{\eta}_n = \eta$  为常数, 令  $\tilde{y}_n - y_n = \varepsilon_n$ , 于是上式又可写为

$$\rho(E) \varepsilon_n + \lambda h \sigma(E) \varepsilon_n + \eta = 0. \quad (43)$$

这就是舍入误差  $\varepsilon_n$  所满足的线性常系数非齐次方程, 它的解由齐次方程的通解  $\varepsilon_{n_h}$  和非齐次方程的一个特解,

$$\varepsilon_{n_p} = \frac{-\eta}{\left( \sum_{i=0}^k \alpha_i + \lambda h \sum_{i=0}^k \beta_i \right)} \quad (44)$$

所组成. 我们要研究的是, 当  $n$  增大时, 在什么条件下  $\varepsilon_n = \varepsilon_{n_h} + \varepsilon_{n_p}$  有界. 由于  $\varepsilon_{n_p}$  为与  $n$  无关的常数, 所以我们只要研究齐次方程的通解就行了, 为此, 我们考虑特征方程

$$\rho(\mu) + \lambda h \sigma(\mu) = 0 \quad (45)$$

的根, 并研究要使方法为稳定的, 这些特征根所满足的条件. 设(45)的  $k$  个根  $\mu_1, \dots, \mu_k$  互异, 显然它们是  $\lambda h$  的函数

$$\mu_i = \mu_i(\lambda h) \quad i=1, \dots, k$$

于是, 齐次差分方程的通解为

$$\varepsilon_{n_H} = c_1 \mu_1^n + c_2 \mu_2^n + \dots + c_k \mu_k^n, \quad (46)$$

其中  $c_1, \dots, c_k$  可由出发值的误差  $\varepsilon_0, \dots, \varepsilon_{k-1}$  所确定. 因为误差方程(43)的特征方程(45)与  $y_n$  所满足的(41)的特征方程相同, 所以

$$y_n = d_1 \mu_1^n + d_2 \mu_2^n + \dots + d_k \mu_k^n,$$

其中  $d_1, \dots, d_k$  由出发值  $y_0, \dots, y_{k-1}$  所确定.

令  $\mu_1$  为逼近精确解  $y(x) = e^{\lambda x}$  ( $y' = \lambda y, y(0) = 1$  的解)的那个根, 则

$$\mu_1 = e^{\lambda h} + O(h^{p+1}).$$

当  $h \rightarrow 0$  时  $\mu_1 \rightarrow 1$ , 这个根称为主根, 其他  $k-1$  个根  $\mu_2, \dots, \mu_k$  都是因为用  $k$  阶差分方程逼近一阶微分方程所产生的, 所以称为外来根. 要求误差保持有界, 就必须  $|\mu_i| \leq 1$ , 满足这个条件的方法就称为绝对稳定的. 使  $|\mu_i| \leq 1$  成立的  $\lambda h$  值构成的区域称为稳定性域, 如  $|\mu_1| \geq |\mu_i| \quad i=2, 3, \dots, k$  成立, 则称方法为相对稳定的, 绝对稳定并不包含相对稳定, 换言之, 当  $|\mu_i| \leq 1, i=1, 2, \dots, k$  的情况下, 也可以有  $|\mu_1| < |\mu_i| \leq 1, i=2, 3, \dots, k$ . 对于固有稳定的方程  $y' = \lambda y, \operatorname{Re}(\lambda) < 0$ . 绝对稳定性重要, 因为精确解随  $n$  增大而趋于零; 对  $\operatorname{Re}(\lambda) > 0$  的方程, 相对稳定性重要, 因为这时只要数值解的分量  $\mu_i^n$  的增长都不比主根  $\mu_1^n$  增长得快, 解就不会失真.

在  $h \rightarrow 0$  时  $\mu_i = \mu_i(\lambda h) \rightarrow \tilde{\mu}_i$ , 而  $\tilde{\mu}_i$  为  $\rho(\xi)$  的根. Dahlquist 曾证明, 稳定的条件为  $|\tilde{\mu}_i| \leq 1$ , 且在单位圆上的根不能为重根, 这就是 § 2 中所述的渐近稳定性的定义.

上面各节中所述的误差主要来自方法的截断误差, 及舍入误差. 由于多步公式的初始出发值系由其他单步法或 Taylor 展式所提供, 这些出发值必然有误差. Dahlquist 曾证明在单位圆上, 如果除了  $\tilde{\mu}_1=1$  这个根外还有其他的根  $\tilde{\mu}=e^{i\theta}$ , 而  $\theta \neq 0$  的话, 则这种根的存在就会使初始出发值带来的误差产生振动, 对固有稳定的方程来说, 精确解随变量的增长而衰减. 有了这种振动就严重地影响到数值解的真实性, 这种方法称为弱稳定的, 或称数值不稳定性.

下面我们将以 Milne 方法和 Hamming 方法为例来说明上述的概念.

迭代的 Milne 方法其稳定域仅由其校正公式

$$\begin{aligned} y_{n+1} &= y_{n-1} + (h/3) [y'_{n+1} + 4y'_n + y'_{n-1}], \\ T(x, h) &= -(h^5/90)y^{(5)}(\xi) \end{aligned} \quad (47)$$

所决定, 为讨论稳定性方便起见, 我们假设所考虑的方程为

$$y' = \lambda y. \quad (48)$$

将(48)代入(47)便得到

$$\left(1 - \frac{h\lambda}{3}\right)y_{n+1} - \frac{4h\lambda}{3}y_n - \left(1 + \frac{h\lambda}{3}\right)y_{n-1} = 0. \quad (49)$$

(49)式的特征方程为

$$\left(1 - \frac{h\lambda}{3}\right)\mu^2 - \frac{4h\lambda}{3}\mu - \left(1 + \frac{h\lambda}{3}\right) = 0. \quad (50)$$

它是一个二次方程, 故有两个根

$$\begin{aligned} \mu_1 &= \left(1 - \frac{h\lambda}{3}\right)^{-1} \left[ \frac{2h\lambda}{3} + \left(1 + \frac{h^2\lambda^2}{3}\right)^{1/2} \right], \\ \mu_2 &= \left(1 - \frac{h\lambda}{3}\right)^{-1} \left[ \frac{2h\lambda}{3} - \left(1 + \frac{h^2\lambda^2}{3}\right)^{1/2} \right]. \end{aligned} \quad (51)$$

在(51)中将含有平方根的项展开, 便可得到

$$\mu_1 \approx e^{h\lambda}, \quad \mu_2 \approx e^{-h\lambda/3}.$$

显见  $\mu_2$  为外来根. 当  $\lambda > 0$  时  $\mu_1$  为逼近精确解的那个根.

$\mu_2$  这个根则由于  $|\mu_2| < 1$  而失去作用。但当  $\lambda < 0$  时, 一方面  $\mu_1$  就如精确解那样具有不断下降的性质, 而  $\mu_2$  则以不断增长的振幅进行振动。由于这一振动的存在, 就有可能掩盖精确解。所以 Milne 方法不具有实负稳定界, 故它不适合固有稳定的方程。在  $h \rightarrow 0$  的渐近情况下 (50) 便简化为

$$\mu^2 - 1 = 0.$$

它的两个根为  $\mu_1 = +1$ ,  $\mu_2 = -1$ 。Milne 方法在单位圆上由于除了  $\mu_1 = 1$  这个根外还有  $\mu_2 = e^{i\pi}$  这个根。按 Dahlquist 理论, 它是弱稳定的。

Hamming 针对了迭代的 Milne 公式的这一弱点提出了他的校正公式

$$y_{n+1} = 1/8 [9y_n - y_{n-2}] + (3h/8) [y'_{n+1} + 2y'_n - y'_{n-1}], \quad (52)$$

$$T(x, h) = -(9h^5/360)y^{(5)}(\xi).$$

现在我们来考虑迭代的 Hamming 公式的稳定域, 为此将 (48) 代入 (52) 便得到

$$\left(\frac{3h\lambda}{8} - 1\right)y_{n+1} + \left(\frac{9}{8} + \frac{3h\lambda}{4}\right)y_n - \left(\frac{3h\lambda}{8}\right)y_{n-1} - \frac{1}{8} = 0. \quad (53)$$

所以它的特征方程为

$$\mu^3 \left(\frac{3h\lambda}{8} - 1\right) + \mu^2 \left(\frac{9}{8} + \frac{3h\lambda}{4}\right) - \mu \left(\frac{3h\lambda}{8}\right) - \frac{1}{8} = 0. \quad (54)$$

Chase<sup>1)</sup> 对 (54) 进行了分析, 得出实负稳定域为  $-2.6 < h\lambda < 0$ 。在  $h \rightarrow 0$  的渐近情况下 (54) 简化为

$$8\mu^3 - 9\mu + 1 = 0. \quad (55)$$

(55) 的三个根, 分别为  $\mu_1 = 1$ ,  $\mu_{2,3} = \frac{1 \pm \sqrt{33}}{16}$ , 显然  $|\mu_2|$

1) Chase, P. E., Stability properties of predictor-corrector methods for ordinary differential equations. *J. Assoc. Comput. Mach.* 9, 457 (1962).



$<1$ ,  $|\mu_3| < 1$ , 所以按 Dahlquist 理论 Hamming 方法是强稳定的.

同样对于非迭代的预测-校正 Hamming 公式

$$\bar{y}_{n+1} = y_{n-3} + (4h/3) [2y'_n - y'_{n-1} + 2y'_{n-2}],$$

$$y_{n+1} = (1/8) [9y_n - y_{n-2}] + (3h/8) [\bar{y}'_{n+1} + 2y'_n - y'_{n-1}],$$

它的稳定域为  $-0.5 < h\lambda < 0$ , 若使用校正公式二次则它的稳定域就扩大成为  $-0.9 < h\lambda < 0$ .

对于修正的 Hamming 预测-校正公式

$$p_{n+1} = y_{n-3} + (4h/3) [2y'_n - y'_{n-1} + 2y'_{n-2}],$$

$$m_{n+1} = p_{n+1} - (112/121) (p_n - c_n),$$

$$c_{n+1} = (1/8) [9y_n - y_{n-2}] + (3h/8) [m'_{n+1} + 2y'_n - y'_{n-1}],$$

$$y_{n+1} = c_{n+1} + (9/121) (p_{n+1} - c_{n+1})$$

来说其稳定域为  $-0.85 < h\lambda < 0$ .

Milne 方法虽然有截断误差较小的优点, 但如在 Hamming 公式中使用较小的  $h$  (约减小 15%) 就能弥补这一损失. 反之由于预测-校正 Milne 公式的稳定域为  $-0.8 < h\lambda < -0.3$ , 修正的预测-校正 Milne 公式的稳定域反而缩小到  $-0.42h\lambda < -0.2$ , 而迭代的 Milne 公式甚至无实负稳定界, 所以相比之下 Hamming 公式要较 Milne 公式优越得多.

4.5. 多步法的绝对稳定性, 方程组的情形 现仍以下列二方程为例来说明:

$$\begin{cases} \dot{y} = Ay, \\ y(0) = y_0. \end{cases}$$

多步法公式为

$$\begin{aligned} & \alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n \\ & + h\Delta (\beta_k y_{n+k} + \cdots + \beta_0 y_n) = 0. \end{aligned} \quad (56)$$

(56) 可写成标量的形式

$$\sum_{i=0}^k \alpha_i \begin{bmatrix} y_{1, n+i} \\ y_{2, n+i} \end{bmatrix} + h \sum_{i=0}^k \beta_i \begin{bmatrix} a_{11} y_{1, n+i} + a_{12} y_{2, n+i} \\ a_{21} y_{1, n+i} + a_{22} y_{2, n+i} \end{bmatrix} = 0. \quad (57)$$

(57)的特征方程有  $2k$  个根:  $\mu_{ij}$ ,  $i=1, 2$ ,  $j=1, 2, \dots, k$ . 因此对应于  $y$  的二个分量均有一个由  $k$  个根组成的集合. 其中分别逼近  $e^{\lambda_1 h}$  及  $e^{\lambda_2 h}$  的主根记为  $\mu_{11}$  及  $\mu_{21}$ . 其截断误差由方法的阶数所决定, 其他的根  $\mu_{1j}$ ,  $\mu_{2j}$ ,  $j=2, \dots, k$ , 都是由用  $k$  阶差分方程去逼近微分方程而产生的外来根.

差分方程(57)的解为

$$\begin{aligned} y_{1n} &= c_{11}\mu_{11}^n + c_{12}\mu_{12}^n + \dots + c_{1k}\mu_{1k}^n \\ &\quad + d_{11}\mu_{21}^n + d_{12}\mu_{22}^n + \dots + d_{1k}\mu_{2k}^n, \\ y_{2n} &= c_{21}\mu_{11}^n + c_{22}\mu_{12}^n + \dots + c_{2k}\mu_{1k}^n \\ &\quad + d_{21}\mu_{21}^n + d_{22}\mu_{22}^n + \dots + d_{2k}\mu_{2k}^n, \end{aligned} \quad (58)$$

其中  $c_{11}=c_1u_1$ ,  $d_{11}=c_2v_1$ ,  $c_{21}=c_1u_2$ , 及  $d_{21}=c_2v_2$ .  $c_1, c_2$  由初始条件所确定.  $u=[u_1, u_2]^T$ ,  $v=[v_1, v_2]^T$  为对应  $A$  的特征根  $\lambda_1, \lambda_2$  的二个特征向量.

$|\mu_{ij}| \leq 1$ ,  $i=1, 2$ ;  $j=1, 2, \dots, k$  为绝对稳定的条件.

$|\mu_{11}| \geq |\mu_{1j}|$ ,  $|\mu_{21}| \geq |\mu_{2j}|$ ,  $j=2, 3, \dots, k$  为相对稳定的条件. 其中主根逼近精确解的分量.

$$\mu_{11} \approx e^{\lambda_1 h}, \quad \mu_{21} \approx e^{\lambda_2 h}. \quad (59)$$

方法的精确度, 取决于(59)式误差项的阶数.

(57)的根只与  $h$  和  $A$  的特征值有关而不依赖于  $A$  的个别元素. 又上述讨论亦完全适用于多个方程. 故不再作进一步的介绍.

## 第七章 坏条件方程组简介

### § 1 什么是坏条件(stiff)方程组

很多控制系统、化学非平衡反应、管式反应动力学、电子网络及模拟理论等方面的问题, 数学模型为常微分方程组的初值问题

$$\begin{cases} \mathbf{y}' = \mathbf{f}(t, \mathbf{y}), \\ \mathbf{y}(a) = \boldsymbol{\eta}, \end{cases}$$

其中  $\mathbf{y} = [y_1, \dots, y_n]^T$ ,  $\mathbf{f} = [f_1, \dots, f_n]^T$ ,  $\boldsymbol{\eta} = [\eta_1, \dots, \eta_n]^T$  为  $n$  维向量, 这种系统的物理特征, 是与它的 Jacobi 矩阵  $\frac{\partial \mathbf{f}}{\partial \mathbf{y}}$  的特征值  $\lambda_k = \alpha_k + i\beta_k$  ( $k=1, 2, \dots, n$ ) 所关联的. 特征值的实部  $\text{Re}\lambda_k = \alpha_k$ , 当  $\alpha_k > 0$  时表示运动振幅的增长; 反之,  $\alpha_k < 0$  则表示衰减, 特征值的虚部  $\text{Im}\lambda_k = \beta_k$  表示周期性振动的频率, 而  $|\text{Re}\lambda_k|$  是一与物理系统时间常数所关联的量, 它为物理学家用来说明衰减速率的. 例如, 微分方程  $y' = \lambda y$  的解为  $ce^{\lambda t}$ . 若  $\lambda$  为负的, 则  $y$  在时间  $-\frac{1}{\lambda}$  中减少了一个  $e^{-1}$  的因子, 所以  $-\frac{1}{\lambda}$  就称为时间常数, 其模  $|\lambda|$  越大, 时间常数就越小. 物理系统的解通常至少在局部都可用指数函数来描述的, 如果对于固有稳定问题即  $\text{Re}\lambda_k < 0$  ( $k=1, 2, \dots, n$ ), 且

$$s = \max_{1 \leq k \leq n} |\text{Re}\lambda_k| / \min_{1 \leq r \leq n} |\text{Re}\lambda_r|$$

相当大时, 求这类问题的数值解就成了一个困难的问题. 在使用通常的数值方法时, 要受到对稳定性要求的限制, 即

$|h\lambda_k|$  ( $k=1, 2, \dots, n$ ) 不得超过某个量, 所以, 在发生上述情况时, 步长  $h$  就需要取得很小, 其结果是造成机器时间的浪费和舍入误差的大量积累, 致使所得到的数值解失真. 然而从方法满足一定的精度的要求出发,  $h$  并不需要取得这么小, 从物理上来讲这更是非本质的, 这类  $s \gg 1$  的问题, 称为坏条件问题, 相应的方程组称为坏条件方程组.

早在五十年代前期在化学反应中已接触到了这类问题, 然而对它在实践上的普遍性和重要性以及困难性的认识还是近十年的事, 事实上它不仅在于常微分方程方面, 而且在计算数学的许多方面都有类似的坏条件问题.

今举一个简单的例子来说明这一问题. 线性常系数微分方程组

$$y' = Ay,$$

其中  $y = [y_1, y_2, y_3]^T$ ,  $A$  为  $3 \times 3$  阶矩阵

$$A = \begin{bmatrix} -0.1 & -49.9 & 0 \\ 0 & -50 & 0 \\ 0 & 70 & -120 \end{bmatrix}.$$

它的满足初始条件  $y(0) = [2, 1, 2]^T$  的解为

$$y_1(x) = e^{-0.1x} + e^{-50x},$$

$$y_2(x) = e^{-50x},$$

$$y_3(x) = e^{-50x} + e^{-120x},$$

其中  $\lambda_1 = -120$ ,  $\lambda_2 = -50$ ,  $\lambda_3 = -0.1$  为  $A$  的三个特征值, 于是  $y_1(x)$ ,  $y_2(x)$  和  $y_3(x)$  都有对应于  $\lambda_1$  或  $\lambda_2$  的一个衰减得很快的分量, 它们很快地便无足轻重, 然而从稳定性的要求出发,  $|h\lambda_i|$  ( $i=1, 2, 3$ ) 都需要小于一个小量, 例如对于 Euler 方法来说, 这个量就是 2, 所以  $|120h| < 2$ , 因此  $h$  的最大值只能为  $1/60$ , 虽然对于解的对应于  $\lambda_1$  的这个分量来说, 很快就失去了实际价值了, 但是在整个积分区间, 由于受到绝对稳定的

限制,  $h$  又必须取得这么小, 所以对这种坏条件问题, 尤其对  $s$  很大的方程组, 当积分区间很长时用通常的方法就不合适了.

## § 2 适合于不同情况的解坏条件 方程的线性方法

由于上述原因, 在解坏条件方程组时, 为了保证解的一定的精确度, 就需要寻求有效的解法. 本节我们主要介绍三种解法: 显式公式、隐式公式和预测校正型.

**2.1. 显式方法** 这类方法的稳定域为有限的, 所以将它用于解坏条件方程组时, 主要的着眼点是在保证精确度的前提下, 尽可能地扩大稳定区域, 由于许多坏条件方程都具有形式

$$y' = -p(y - \tilde{y}), \quad (1)$$

其中  $p$  为相当大的一个数,  $\tilde{y}$  为  $x$  的变化缓慢的函数, 它可以用  $\omega$  的含有未知参数的幂级数来逼近它, 而这些参数可在积分过程中来确定.

下面介绍 Treanor 方法, 它是建立在上述假设下适合于坏条件方程的 Runge-Kutta 公式的一种变形, 在正常条件下, 它还原为 Runge-Kutta 公式.

假设

$$y'_i = f_i(x, y).$$

在每个子区间  $[x_n, x_n + h]$  中可以用

$$y'_i = -(p_i)_n y_i + (a_i)_n + (b_i)_n x + (c_i)_n x^2 \quad (2)$$

来逼近. 其中的四个常数  $a_i, b_i, c_i, p_i$ , 可用  $f_i(x, y)$  在区间  $[x_n, x_n + h]$  中的四个点来确定它. 这里与四阶 Runge-Kutta

方法一样, 可选  $x_n, x_n + \frac{1}{2}h, x_n + \frac{1}{2}h$  及  $x_{n+1}$  这几个点. 于是向量  $y$  的每个分量可以用下面的公式来计算

$$\begin{aligned}
 y_{n+1/2}^{(1)} &= y_n + (h/2)f_n, \\
 f_{n+1/2}^{(1)} &= f(x_n + h/2, y_{n+1/2}^{(1)}), \\
 y_{n+1/2}^{(2)} &= y_n + (h/2)f_{n+1/2}^{(1)}, \\
 f_{n+1/2}^{(2)} &= f(x_n + h/2, y_{n+1/2}^{(2)}), \\
 y_{n+1}^{(3)} &= y_n + h[2f_{n+1/2}^{(2)}F_2 + f_{n+1/2}^{(1)}PhF_2 + f_n(F_1 - 2F_2)], \\
 f_{n+1}^{(3)} &= f(x_n + h, y_{n+1}^{(3)}), \\
 y_{n+1} &= y_n + hf_nF_1 + hv_3(Py_n + f_n) + hv_2(Py_{n+1/2}^{(1)} + f_{n+1/2}^{(1)}) \\
 &\quad + hv_2(Py_{n+1/2}^{(2)} + f_{n+1/2}^{(2)}) + hv_1(Py_{n+1}^{(3)} + f_{n+1}^{(3)}),
 \end{aligned} \tag{3}$$

其中

$$\begin{aligned}
 F_1 &= \frac{e^{-Ph} - 1}{-Ph}, & F_2 &= \frac{e^{-Ph} - 1 + Ph}{(Ph)^2}, \\
 F_3 &= \frac{e^{-Ph} - 1 + Ph - \frac{1}{2}(Ph)^2}{-(Ph)^3}
 \end{aligned} \tag{4}$$

以及

$$\begin{aligned}
 v_1 &= -F_2 + 4F_3, \\
 v_2 &= 2(F_2 - 2F_3), \\
 v_3 &= 4F_3 - 3F_2.
 \end{aligned} \tag{5}$$

在公式中出现的  $P$  为

$$p_i = - \frac{f_{i,n+1/2}^{(2)} - f_{i,n+1/2}^{(1)}}{y_{i,n+1/2}^{(2)} - y_{i,n+1/2}^{(1)}} \tag{6}$$

中的最大值, 而  $p_i$  为头二步值差之比, 对  $m$  个方程均使用同一个标量  $P$  是为了保证  $m$  个方程都具有同样的差分步长, 若所有的  $p_i$  均为负的, 亦即  $\frac{\partial f}{\partial y}$  的差分近似

$$(f_{i,n+1/2}^{(2)} - f_{i,n+1/2}^{(1)}) / (y_{i,n+1/2}^{(2)} - y_{i,n+1/2}^{(1)})$$

均为正的, 所以问题为非坏条件的, 可用寻常的方法积分, 此时, 我们令  $P=0$ . 且通过极限过程得到相应的

$$F_1=1, \quad F_2=1/2, \quad F_3=1/6, \\ v_1=1/6, \quad v_2=1/3, \quad v_3=-5/6,$$

此时 Treanor 公式就还原成四阶的 Runge-Kutta 公式.

Treanor 公式的具体计算步骤为:

1. 选取一个初始步长  $h$ ;
2. 由(5)的头二个方程计算  $y_{n+1/2}^{(1)}$  及  $y_{n+1/2}^{(2)}$ ;
3. 计算  $p_i$  的最大值  $P$ . 若所有的  $p_i < 0$ , 则令  $P=0$ ;
4. 由(3)的最后一个方程计算  $y_{n+1}$ ;
5. 若每一个  $|y_{n+1}-y_n|/|y_{n+1}| > \varepsilon_{\max}$ , 令  $h=h/2$  并转第二步; 若每一个  $|y_{n+1}-y_n|/|y_{n+1}| < \varepsilon_{\min}$ , 令  $h=2h$  并转第二步.

对于实的负值  $h\lambda$  的稳定域, Lomax 及 Bailey<sup>1)</sup> 曾计算过, 对于  $-2 < h\lambda_i < 0$ , 方法对任何  $P$  值都是稳定的, 又若有特征值使  $h\lambda_i < -2$ , 其稳定区域如图 7.1 所示.

Treanor 方法的优点是由于它是显式的便于程序设计, 特别, 对一般条件好的方程, 它就还原为四阶 Runge-Kutta 方法, 而对坏条件方程它又有增加稳定性的好处, 假如  $Ph=8$ , 则实稳定界为  $-10$ , 比四阶的 Runge-Kutta 公式的实稳定界  $-2.785$  要宽.

**2.2. 隐式方法** 众所周知隐式公式, 诸如后退 Euler 公式, 梯形公式等, 都是  $A$ -稳定的, 故都适合于固有稳定的坏条件方程组, 假如对一组  $h$  序列  $\{h_i\}$ ,  $h_i > h_{i+1} > \dots$  使用梯形公式, 算得区间终点处的数值解的序列为  $\{T_i\}$ , 然后对序列

1) Lomax, H. and Bailey, H. E., A critical analysis of various numerical integration methods for computing the flow of a gas in chemical non-equilibrium. NASA Technical Note, NASA TND-4109 (1967).

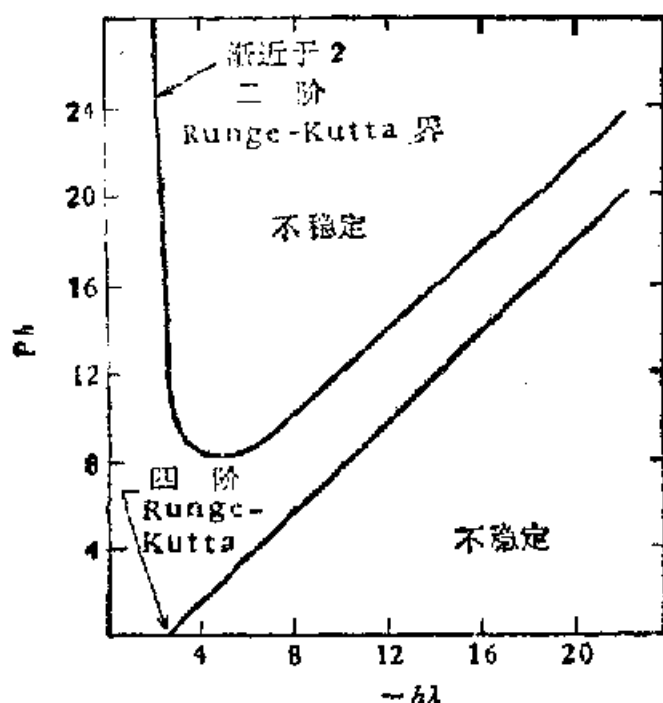


图 7.1 Treanor 方法关于实负  $h\lambda$  稳定界

$\{T\}$  进行外推就可以达到保持稳定性又提高精度的目的。同样半隐式的 Runge-Kutta 公式, 它既保持隐式 Runge-Kutta 公式的良好的数值稳定性, 而工作量却小得多, 这些方法的适应范围, 要较 Treanor 方法广。它对微分方程组右端函数的 Jacobi 矩阵没有特殊要求。具体公式已在第二章 Runge-Kutta 型公式中作了详细介绍。

**2.3. 预测-校正型多步公式(多值公式)** 我们知道,  $k$  步公式可以写成

$$L_h(y_n) \equiv \sum_{i=0}^k (\alpha_i y_{n-i} + h\beta_i f_{n-i}) = 0.$$

引进两个多项式

$$\rho(\xi) = \sum_{i=0}^k \alpha_i \xi^{k-i},$$

和

$$\sigma(\xi) = \sum_{i=0}^k \beta_i \xi^{k-i},$$



并考虑函数  $y(t) = e^{\lambda t}$ ,

$$\begin{aligned} L_h(y(t)) &= \sum_{i=0}^k (\alpha_i + \lambda h \beta_i) e^{\lambda(t-hi)} \\ &= \sum_{i=0}^k e^{\lambda(t-hk)} (\alpha_i + \lambda h \beta_i) (e^{\lambda h})^{k-i} \\ &= e^{\lambda(t-hk)} [\rho(e^{\lambda h}) + \lambda h \sigma(e^{\lambda h})]. \end{aligned} \quad (7)$$

设考虑的方法为  $r$  阶的, 又由  $e^{h\lambda} = 1 + O(h)$ , 则有

$$\begin{aligned} L_h(y(t)) &= C_{r+1} h^{r+1} y^{(r+1)}(t) + O((\lambda h)^{r+2}) \\ &= C_{r+1} (\lambda h)^{r+1} e^{\lambda t} + O((\lambda h)^{r+2}) \\ &= C_{r+1} (\lambda h)^{r+1} e^{\lambda(t-hk)} + O((\lambda h)^{r+2}). \end{aligned} \quad (8)$$

因此, 由(7)与(8)可得

$$\begin{aligned} &\rho(e^{\lambda h}) + \lambda h \sigma(e^{\lambda h}) \\ &= C_{r+1} (\lambda h)^{r+1} + O((\lambda h)^{r+2}). \end{aligned} \quad (9)$$

记  $\lambda h = \log(1+z)$  并注意到  $\log(1+z) = z + O(z^2)$ , 则有

$$\begin{aligned} &\rho(1+z) + \log(1+z) \sigma(1+z) \\ &= C_{r+1} z^{r+1} + O(z^{r+2}). \end{aligned} \quad (10)$$

可见(10)式为  $r$  阶方法的必要条件. 又若(10)成立,

$$L_h(e^{\lambda t}) = \sum_{q=0}^{\infty} C_q (\lambda h)^q e^{\lambda t} = O((\lambda h)^{r+1})$$

隐含着  $C_0 = C_1 = \dots = C_r = 0$ , 所以(10)这个条件也是方法为  $r$  阶的充分条件.

现在我们要用(10)式导出对坏条件方程组来说为稳定的方法. Gear 曾证明, 取  $\sigma(\xi) = \beta_0 \xi^k$  时得到的  $k$  步公式, 当  $2 \leq k \leq 6$  时, 对于相当一类的坏条件方程组, 其计算过程是稳定的. 为了得到一个  $k$  阶方法, 其具体作法如下: 对于一个固定的  $k$ , 利用  $\sigma(\xi) = \beta_0 \xi^k$  以及关系式(10), 我们可以按下法定出多项式  $\rho(\xi)$  的诸系数  $\alpha_i$ .

以  $k=2$ ,  $\sigma(\xi) = \beta_0^* \xi^2$  为例, 我们有

$$\begin{aligned}
\rho(1+z) &= -\log(1+z) \beta_0^* (1+z)^2 + O(z^3) \\
&= -\beta_0^* \left( z - \frac{1}{2} z^2 \right) (1+z)^2 + O(z^3) \\
&= -\beta_0^* \left( z + \frac{3}{2} z^2 \right) + O(z^3).
\end{aligned}$$

考虑到通常公式中  $\alpha_0 = -1$ , 取  $\beta_0^* = 2/3$ , 得到

$$\begin{aligned}
\rho(1+z) &= -\left( \frac{2}{3} z + z^2 \right) \\
&= -(1+z)^2 + \frac{4}{3} (1+z) - \frac{1}{3}.
\end{aligned}$$

从而可知

$$\rho(\xi) = -\xi^2 + \frac{4}{3} \xi - \frac{1}{3},$$

即  $k=2$  时校正公式的系数

$$\beta_0^* = 2/3, \alpha_0 = -1, \alpha_1 = 4/3, \alpha_2 = -1/3.$$

当  $k=1$  时我们便得到后退的尤拉公式.

用类似的方法, 对固定的  $k$  可得到诸  $\alpha_i (i=0, 1, \dots, k)$  和  $\beta_0$ ,

现给出  $2 \leq k \leq 6$  时校正公式的  $\beta_0$  及  $\alpha_i$  数值表 ( $\alpha_0 \equiv -1$ ):

$k$	2	3	4	5	6
$\beta_0$	2/3	6/11	12/25	60/137	60/147
$\alpha_1$	4/3	18/11	48/25	300/137	360/147
$\alpha_2$	-1/3	-9/11	-36/25	-300/137	-450/147
$\alpha_3$		2/11	16/25	200/137	400/147
$\alpha_4$			-3/25	-75/137	-225/147
$\alpha_5$				12/137	72/147
$\alpha_6$					-10/147

由于使用  $k$  步公式时在计算中想改变步长的大小通常很麻烦,而且,计算开始时还需要为它提供  $k$  个出发值. 为了克服这两个缺点,可作如下处理.

把  $k$  阶的预测公式和校正公式分别记为

$$y_{n,(0)} = \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \cdots + \alpha_k y_{n-k} + \beta_1 h y'_{n-1}, \quad (11)$$

$$y_{n,(m)} = \alpha_1^* y_{n-1} + \alpha_2^* y_{n-2} + \cdots + \alpha_k^* y_{n-k} + \beta_0^* h f(y_{n,(m-1)}, t_n), \quad (12)$$

其中下标  $(0)$ ,  $(m)$  分别表零次近似与  $m$  次近似值, 如有  $s$  个方程, 则  $y, f$  都表示  $s$  维向量的(任)一个分量.

定义向量:

$$Y_n = [y_n, h y'_n, y_{n-1}, \cdots, y_{n-k+1}]^T, \\ Y_{n,(m)} = [y_{n,(m)}, h y'_{n,(m)}, y_{n-1}, \cdots, y_{n-k+1}]^T,$$

其中

$$h y'_{n,(m)} = \begin{cases} (\beta_1/\beta_0^*) h y'_{n-1} + \sum_{i=1}^k \{[(\alpha_i - \alpha_i^*)/\beta_0^*] y_{n-i}\}, & m=0, \\ h f(y_{n,(m-1)}, t_n), & m \geq 1. \end{cases} \quad (13)$$

令

$$\gamma_i = (\alpha_i - \alpha_i^*)/\beta_0^*, \quad i=1, 2, \cdots, k, \\ \delta_1 = \beta_1/\beta_0^*.$$

引进矩阵  $B$ :

$$B = \begin{bmatrix} \alpha_1 & \beta_1 & \alpha_2 & \cdots & \alpha_{k-1} & \alpha_k \\ \gamma_1 & \delta_1 & \gamma_2 & \cdots & \gamma_{k-1} & \gamma_k \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

则有

$$Y_{n,(0)} = B Y_{n-1}. \quad (14)$$

由于

$$\begin{aligned} hy'_{n,(m+1)} &= hf(y_{n,(m)}, t_n) \\ &= hy'_{n,(m)} + h[f(y_{n,(m)}, t_n) - y'_{n,(m)}], \end{aligned}$$

又注意到由(11), (12), (13)有

$$y_{n,(m+1)} = y_{n,(m)} + \beta_0^* h[f(y_{n,(m)}, t_n) - y'_{n,(m)}],$$

因此若令

$$\begin{aligned} C &= [\beta_0^*, 1, 0, \dots, 0]^T, \\ F(Y_{n,(m)}) &= h[f(y_{n,(m)}, t_n) - y'_{n,(m)}], \end{aligned}$$

则有

$$Y_{n,(m+1)} = Y_{n,(m)} + CF(Y_{n,(m)}), \quad (15)$$

这里

$$CF(Y_{n,(m)}) = \begin{bmatrix} \beta_0^* h[f(y_{n,(m)}, t_n) - y'_{n,(m)}] \\ h[f(y_{n,(m)}, t_n) - y'_{n,(m)}] \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

在公式(14)和(15)的基础上,我们再作变换

$$\begin{cases} Z_n = QY_n, \\ Z_{n,(m)} = QY_{n,(m)}, \end{cases} \quad (16)$$

使得

$$Z_n = [y_n, hy'_n, \dots, h^k y_n^{(k)}/k!]^T.$$

关于确定  $Q$  矩阵的方法如下: 把

$$Y_n = [y_n, hy'_n, y_{n-1}, \dots, y_{n-k+1}]^T$$

的每一个分量在  $t=t_n$  处展开, 舍去  $O(h^{k+1})$  项, 然后利用矩阵乘法定出  $Q$  的每一个元素. 例如  $k=3$  时,

$$\begin{aligned} Y_n &= [y_n, hy'_n, y_{n-1}, y_{n-2}]^T, \\ Z_n &= [y_n, hy'_n, h^2 y_n''/2!, h^3 y_n^{(3)}/3!]^T. \end{aligned}$$

由于

$$y_{n-1} = y_n - hy'_n + \frac{h^2}{2!} y''_n - \frac{h^3}{3!} y^{(3)}_n + O(h^4),$$

$$y_{n-2} = y_n - 2hy'_n + \frac{4h^2}{2!} y''_n - \frac{8h^3}{3!} y^{(3)}_n + O(h^4),$$

得到

$$\frac{h^2 y''_n}{2!} = -\frac{7}{4} y_n + \frac{6}{4} hy'_n + 2y_{n-1} - \frac{1}{4} y_{n-2},$$

$$\frac{h^3 y^{(3)}_n}{3!} = -\frac{3}{4} y_n + \frac{1}{2} hy'_n + y_{n-1} - \frac{1}{4} y_{n-2}.$$

从而易得矩阵  $Q$  为

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -\frac{7}{4} & \frac{6}{4} & 2 & -\frac{1}{4} \\ -\frac{3}{4} & \frac{1}{2} & 1 & -\frac{1}{4} \end{bmatrix}.$$

通过变换 (16), 公式 (14), (15) 变为

$$Z_{n,(0)} = QBQ^{-1}Z_{n-1}, \quad (17)$$

$$\begin{aligned} Z_{n,(m+1)} &= Z_{n,(m)} + LF(Q^{-1}Z_{n,(m)}) \\ &= Z_{n,(m)} + LF(Z_{n,(m)}). \end{aligned} \quad (18)$$

这里,  $L = QC$ , 且关系式  $F(Q^{-1}Z_{n,(m)}) = F(Y_{n,(m)}) = F(Z_{n,(m)})$  成立的原因是  $Q^{-1}Z_{n,(m)} = Y_{n,(m)}$  而  $Y_{n,(m)}$  与  $Z_{n,(m)}$  有相同的头两个分量  $y_{n,(m)}$  和  $hy'_{n,(m)}$  而  $F(Y_{n,(m)}) = hf(y_{n,(m)}, t_n) - hy'_{n,(m)}$  仅仅依赖于这两个分量.

公式 (17) 中的矩阵  $QBQ^{-1}$  可以证明是 Pascal 矩阵, 即

$$QBQ^{-1} = A = \begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ & 1 & 2 & 3 & k \\ & & 1 & 3 & \vdots \\ & & & 1 & \vdots \\ & & & & \ddots \\ & & & & & k \\ & & & & & & 1 \end{bmatrix}.$$

(它的每一列均由二项式系数所组成), 从而, 预测公式(17)变为

$$Z_{n,(0)} = AZ_{n-1}. \quad (19)$$

可以看到, 由于矩阵的特殊性, 公式(19)的计算只要用加法而不需用乘法便能实现.

对公式(18)我们知道, 如果方法收敛, 它将收敛到

$$Z_n = Z_{n,(0)} + L\omega,$$

这里,  $\omega$  是这样的一个数, 它使

$$F(Z_n) = F(Z_{n,(0)} + L\omega) = 0 \quad (20)$$

成立. 我们用牛顿法来解(20),

$$\omega_{(m+1)} = \omega_{(m)} - \left[ \frac{\partial F}{\partial Z} \cdot L \right]^{-1} F(Z_{n,(0)} + L\omega_{(m)}), \quad (21)$$

如果我们记

$$Z_{n,(m)} = Z_{n,(0)} + L\omega_{(m)},$$

则(21)变成

$$Z_{n,(m+1)} = Z_{n,(m)} - L \left[ \frac{\partial F}{\partial Z} \cdot L \right]^{-1} F(Z_{n,(m)}), \quad (22)$$

因为

$$F(Z) = hf(z_0) - z_1$$

( $z_0$  为  $Z$  的第一个分量,  $z_1$  为  $Z$  的第二个分量), 因此我们有

$$W = \left[ \frac{\partial F}{\partial Z} \cdot L \right]^{-1} = \left[ -l_1 + hl_0 \frac{\partial f}{\partial y} \right]^{-1},$$

且容易知道  $l_1 = 1$ ,  $l_0 = \beta_0^*$ , 这样我们就有了新的校正公式

$$Z_{n, (m+1)} = Z_{n, (m)} - L \left[ -l_1 + hl_0 \frac{\partial f}{\partial y} \right]^{-1} F(Z_{n, (m)}). \quad (23)$$

(23) 中的矩阵  $W = \left[ -l_1 + hl_0 \frac{\partial f}{\partial y} \right]^{-1}$  依赖于方法的阶 (通过  $l_0$ )、步长  $h$  和  $\frac{\partial f}{\partial y}$ , 在步长和阶不改变且  $\frac{\partial f}{\partial y}$  变化缓慢 (实际问题中常常如此) 的情况下,  $W$  将不会改变很大, 在计算中如果注意到这一点, 可以减少很多计算量.

从 (19) 式和 (23) 式可以看到, 公式仅用到了前一步的  $y$  及其导数  $y', y'', \dots$  的信息, 而没有用到多个点上  $y$  的信息. 由 (19) 和 (23) 组成的公式称为多值公式. 由此可见, 我们经过一系列变换便将一个多步预测-校正公式化成了一个自开始、便于变阶、变步长的多值预测-校正公式.

附  $2 \leq k \leq 6$  时  $L=QO$  的数值表:

$k$	2	3	4	5	6
$l_0$	2/3	6/11	24/50	120/274	720/1764
$l_1$	1	1	1	1	1
$l_2$	1/3	6/11	35/50	225/274	1624/1764
$l_3$		1/11	10/50	85/274	735/1764
$l_4$			1/50	15/274	175/1764
$l_5$				1/274	21/1764
$l_6$					1/1764

Gear 已将上述算法编制了常微分方程的初值问题的自动积分程序<sup>1)</sup>.

1) Gear, C. W., Numerical Initial Value Problems in Ordinary Differential Equations, Prentice-Hall, Inc., Engelwood Cliffs, New Jersey (1971).

### §3 非线性方法

在前面几章中所介绍的求微分方程初值问题的数值解的方法,例如 Runge-Kutta 方法、多步法及预测-校正法等,它们的计算公式对于函数值和导数值都是线性的,因而在用于试验方程  $y' = Ay$  时所得到的方程组都是一般的线性代数方程组。所以我们称这一类方法是线性方法,在一般情况下,使用这些方法都能得到较好的效果。然而在遇到坏条件方程组时,情况就不同了。由于对数值稳定性有较强的要求,必须对原有的方法加以适当改进,或者被迫采用需要迭代的隐式方法,从而增大了计算工作量。在这一节中我们将介绍一种既能避免使用隐式方程,而又具有较好数值稳定性的非线性方法。这种方法的基本思想在于对初值问题的解能局部地用有理函数逼近。下面我们仍以单个方程为例来介绍这一方法,并讨论它的稳定性,以及与线性方法不同之处,即方法的阶数和局部截断误差依赖于所给初值问题。

#### 3.1. 方法的建立 设初值问题为

$$\begin{cases} y' = f(x, y), \\ y(a) = \eta \end{cases}$$

的理论解为  $y(x)$ , 在区间  $[x_n, x_{n+1}]$  上我们根据插值原理,用最简单的有理函数  $I(x) = A/(x+B)$  来逼近它。用  $y_n$  表示  $y(x_n)$  的已计算出的近似值,并令  $f_n = f(x_n, y_n)$ , 其中

$$x_n = a + nh.$$

根据插值条件

$$y_n = I(x_n), \quad f_n = I'(x_n) \quad (24)$$

来确定系数  $A$  和  $B$ , 然后令  $y_{n+1} = I(x_{n+1})$  便得到单步公式



$$y_{n+1} - y_n = \frac{hy_n f_n}{y_n - hf_n}. \quad (25)$$

这里应假定

$$|y(x)| + |y'(x)| \neq 0, \quad x \geq a, \quad (26)$$

否则, (25) 便退化为  $y_{n+1} = y_n$ . 如果对于某一步出现

$$y_n - hf_n = 0$$

的情况, 那就应该选择另一个  $h$ .

**3.2. 方程组的情形** 将前述方法用于方程组

$$\mathbf{y}' = \mathbf{f}(x, \mathbf{y})$$

上, 其中

$$\mathbf{y} = [y^1, y^2, \dots, y^s]^T, \quad \mathbf{f} = [f^1, f^2, \dots, f^s]^T,$$

则公式(25)化为

$$y_{n+1}^i - y_n^i = \frac{hy_n^i f_n^i}{y_n^i - hf_n^i}, \quad i=1, 2, \dots, s. \quad (27)$$

容易看出方程(25)和(26)的右端为关于  $y_n$  和  $f_n$  的非线性函数, 从而把它用于试验方程  $\mathbf{y}' = A\mathbf{y}$  时所得到的代数方程组是非线性的.

**3.3. 稳定性问题** 将公式(27)应用于试验方程

$$\mathbf{y}' = A\mathbf{y},$$

便得到

$$\frac{y_{n+1}^i}{y_n^i} = \frac{1}{1 - h\lambda_i}, \quad i=1, 2, \dots, s. \quad (28)$$

于是, 递推下去便得到关系

$$y_n^i = \left( \frac{1}{1 - h\lambda_i} \right)^n y_0^i.$$

在  $\text{Re}(\lambda_i) < 0$  时,  $|1 - h\lambda_i| > 1$ , 所以当  $n \rightarrow \infty$  时,  $y_n^i \rightarrow 0$  ( $i=1, 2, \dots, s$ ). 因此, 这种方法为  $A$ -稳定的.

这里对于非线性方法我们在进行理论研究时, 将限于特殊的试验方程  $\mathbf{y}' = A\mathbf{y}$ , 其中  $A = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , (它和

单个试验方程  $y' = \lambda y$ , 没有实质上的区别), 而不是用于一般的试验方程  $y' = Ay$ , 其中  $A$  为一般的矩阵.

对于线性方法来讲, 用  $y' = Ay$  作为试验方程和用  $y' = Ay$  作为试验方程, 实质上是一回事, 其中  $A$  为一般的矩阵, 它的特征值都在左半平面且为互异的, 例如, 若我们使用变换  $y = Hz$ , 其中  $H^{-1}AH = A = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , 则方程组  $y' = Ay$  就可以变换为  $z' = Az$ . Euler 法, 用于原方程组可得  $y_{n+1} = (I + hA)y_n$ , 它在进行了变换  $y_n = Hz_n$  后就变成了  $z_{n+1} = (I + hA)z_n$ , 就是 Euler 公式应用于变换后的方程组的情况. 但是对于非线性方法, 上述的情况就不成立了. 因为 (27) 用于这种试验方程时, 所得到的差分方程才是线性的, 而对于线性差分方程的通解结构已有完善的理论. 在  $A$  为一般的矩阵时, 所得的差分方程为非线性的, 而对于非线性方程的研究还很不够, 目前仅有下述的不完善的部分结论:

将方法 (27) 应用于  $y' = Ay$ , 我们假定  $A$  为实矩阵并具有互异的特征值  $\lambda_1, \lambda_2, \dots, \lambda_s$  与其相应的特征向量分别为  $c_1, c_2, \dots, c_s$ , 则由此而产生的 (非线性) 差分方程有  $s$  个独立的解

$$y_{n,j} = \frac{1}{(1 - h\lambda_j)^n} c_j, \quad j = 1, 2, \dots, s. \quad (29)$$

这个结果的不完善性表现在, 由  $s$  个独立的解不能组成非线性差分方程的通解.

**3.4. 方法的阶数与截断误差** 为了研究方法的阶数, 我们考虑由下式所定义的非线性算子

$$P[y(x); h] = y(x+h) - y(x) - \frac{hy(x)y'(x)}{y(x) - hy'(x)}, \quad (30)$$

其中  $y(x)$  为  $[a, b]$  上的具有一阶连续导数的函数且

$$|y(x)| + |y'(x)| \neq 0, \quad x \in [a, b].$$

若

$$P[y(x); h] = O(h^{p+1}),$$

我们称该方法为  $P$  阶的. 于是, 当  $y(x)$  为初值问题的理论解时,  $P[y(x_n); h]$  便定义了  $x_{n+1}$  处的局部截断误差  $T_{n+1}$ :

$$y(x_{n+1}) - y_{n+1} = T_{n+1}.$$

关于非线性方法的阶数和线性情形不同, 它和初值问题有关, 且在计算过程中会发生变化. 今以方法 (25)

$$y_{n+1} - y_n = \frac{hy_n f_n}{y_n - hf_n}$$

为例来说明这一点. 从 (25) 式可以看出, 不论  $f$  如何选取, 若  $y_n = 0$  就必有  $y_{n+1} = 0$ . 因此数值解总是以零来逼近理论解的. 为什么会产生这个现象? 从 (25) 式的截断误差能说明这一问题.

仿线性方法时那样我们将  $P[y(x_n); h]$  在  $x_n$  这一点展开, 写成主截断误差与余项的形式

$$T_{n+1} = h^2 \left[ \frac{1}{2} y'' - (y')^2 / y \right]_{x=x_n} + O(h^3). \quad (31)$$

所以一般来讲方法为二阶的. 注意到当  $y_n \approx 0$  时, 由于假设 (26) 不含有  $y' \approx 0$ , 所以此时  $T_{n+1}$  中的主截断误差, 当  $h$  很小时就变为不定形式. 这就不能保证截断误差  $T_{n+1}$  一定很小, 从而就无法估计计算结果的精确性, 这就有可能使方法失效.

但是, 实际上却出现有这样的例子. 当发生上述现象时计算的效果仍是很好的.

对  $y_n \approx 0$  时可能使方法失效的原因可作如下的分析: 分子与分母分别展开得

$$T_{n+1} = \frac{[yy'' - 2(y')^2] \frac{h^2}{2!} + [yy''' - 3y'y''] \frac{h^3}{3!} + O(h^4)}{y - hy'} \Big|_{s=2n} \quad (32)$$

从(32)中可以看出: 当  $y_n \neq 0$  及  $y_n - hy'_n \neq 0$  时,  $T_{n+1} = O(h^2)$ , 即方法至少为一阶的. 而当  $y_n = 0$  时, 则  $T_{n+1} = O(h)$ , 此时方法为零阶的, 从而便破坏了相容性条件. 用一个不相容的方法来解初值问题, 当然就不能保证数值结果的可靠性了.

从上面的讨论可以看出, 非线性方法的阶数是可能随初值问题及计算过程的推移而变的.

对于高阶方法, 在这里不准备作进一步介绍, 只列出有关文章<sup>1)</sup>.

#### § 4 关于阶数、步长和方法的选择

关于如何选择方法、阶数和步长应视问题及其要求而定. 在对具体问题进行计算时, 并不是说方法的阶数越高, 步长越小效果就越好. 很显然, 如果一个问题对计算结果本来要求不高, 而却硬要用四阶的 Runge-Kutta 方法来计算, 这只有造成机器时间的浪费, 这个问题应该引起注意.

如果我们将方法、阶数和步长都作为参量来看待的话, 则在整个积分过程中可按预定的要求来决定是否改变这些参量, 例如对物理图象变化剧烈的区间, 就需要用较小的步长  $h$  才能比较准确地来刻划它的变化. 相反, 在变化不大的地方, 就可以将步长  $h$  放大一些. 又例如, 对于坏条件问题, 选择方法时应特别考虑到问题对稳定性的要求. 这样才不会产生由

<sup>1)</sup> J. D. Lambert, Nonlinear Methods for Stiff Systems of Ordinary Differential Equations, Lecture Notes in Mathematics, 363. Conference on the Numerical Solution of Differential Equations.

于方法的选择不当,使得受绝对稳定域的限制只好选择较小的步长  $h$ ,从而使得固有稳定问题的解失真.总之,应全面地考虑这些参量,尽量做到以最小的工作量(自然是指人工与机器时间的总体,但这里我们主要考虑的是机器时间)来实现预期的目的.

**4.1. 阶数的选择** 对于给定的问题,究竟要选择多高阶数的方法应按所要求的精确度来确定才是合理的.现以单步方法为例来说明这一问题,假设我们要在一个固定的区间  $[a, b]$  上以一个确定阶数的方法进行求积,此外又假设我们有阶数为  $r=1, 2, \dots$  的多个方法可供使用.若预先知道对于每个方法应怎样借助  $h_r$  及  $\theta_r(t)$  来选得最好的步长,其中  $h_r$  表示最大步长,  $\theta_r(t_n)$  为满足  $0 < \Delta < \theta_r(t_n) \leq 1$  的变步长因子,下标  $r$  表示方法的阶数.对于  $r$  阶方法选择由  $t_n$  到  $t_{n+1}$  的步长  $h_n$  的公式可以表示成

$$h_n = h_r \theta_r(t_n).$$

上述的最好是指对于指定的允许误差该怎样选择  $h_r$  使工作量为最小.下面我们要介绍  $h_r$  与允许误差  $E$  的关系.

考虑到在区间中使用变步长时,可设  $h$  为最大步长,若我们假设存在着函数  $\theta(t)$  使得  $0 < \Delta \leq \theta(t) \leq 1$  对于  $t \in [a, b]$  成立.并假设从  $t_n$  到  $t_{n+1}$  所取的步长为

$$h_n = h \theta(t_n),$$

$$t_{n+1} = t_n + h_n,$$

便可以导出误差  $e_n$  的渐近表达式<sup>1)</sup>

$$e_n = h^r \delta(t_n) + O(h^{r+1}), \quad (33)$$

其中的  $\delta(t)$  满足微分方程

1) G. W. Gear, Numerical Initial Value Problems in Ordinary Differential Equations, § 4.7.

$$\delta'(t) = \frac{\partial f(y(t), t)}{\partial y} \delta(t) + \theta^r(t) \phi(y(t), t) \quad (34)$$

的解, 其中  $\phi(y, t)$  表示主误差函数. 它满足方程

$$\begin{aligned} h\psi(y(t), t; h) - (y(t+h) - y(t)) \\ = h^{r+1}\phi(y, t) + O(h^{r+2}), \end{aligned}$$

其中  $\psi(y, t; h)$  表增量函数, 公式(33), (34)中的  $h, \delta, \theta$  都省略了标志  $r$  阶方法的下标  $r$ .

对于方程组来说,

$$e_N = \delta_r(t_N) h_r^r + O(h_r^{r+1}), \quad (35)$$

$\delta_r(t_N)$  满足微分方程

$$\delta'(t) = \frac{\partial f}{\partial y} \delta(t) + \theta^r(t) \phi(y(t), t), \quad (36)$$

其中  $t_N = b$ .  $h_r$  的大小依赖于允许误差  $E$  和方法的阶数  $r$ . 又设  $\theta_r(t)$  不依赖于  $E$ , 将(35)中  $O(h_r^{r+1})$  这一项略去, 便得到估计式

$$E \geq \|e_N\| \approx \|\delta_r(t_N)\| h_r^r.$$

用  $W_r$  ( $r$  表方法的阶数) 表示计算到  $t_N = b$  处的总工作量,  $N = O(1/h_r)$ , 因此  $W_r = L_r/h_r$  近似地成立.  $L_r$  为与具体方法有关的量. 如果对不同阶的方法以  $\|e_N\|$  对  $W_r^{-1} = h_r/L_r$  来作图(图 7.2), 便可以看出下面二点: 其一, 对于同样的步长(或  $W_r^{-1}$ ), 阶数高的方法误差小, 或者说要使误差小可以提高方法的阶数. 其二, 对于充分小的  $\|e_N\|$ ,  $W_r^{-1}$  最大值的出现在最高阶的方法上, 而对于充分大的误差,  $W_r^{-1}$  的最大值出现在  $r=1$  的方法上, 这表明对精确度要求低的问题, 要使工作量  $W_r$  为最小, 则应选低阶的方法.

**4.2. 关于步长的选择** 选择步长以达到控制局部误差的作用, 应在计算过程中自动实现. 对单步法来说, 凡误差有渐近表达式者, 都可用已介绍过的 Richardson 外推法来控制

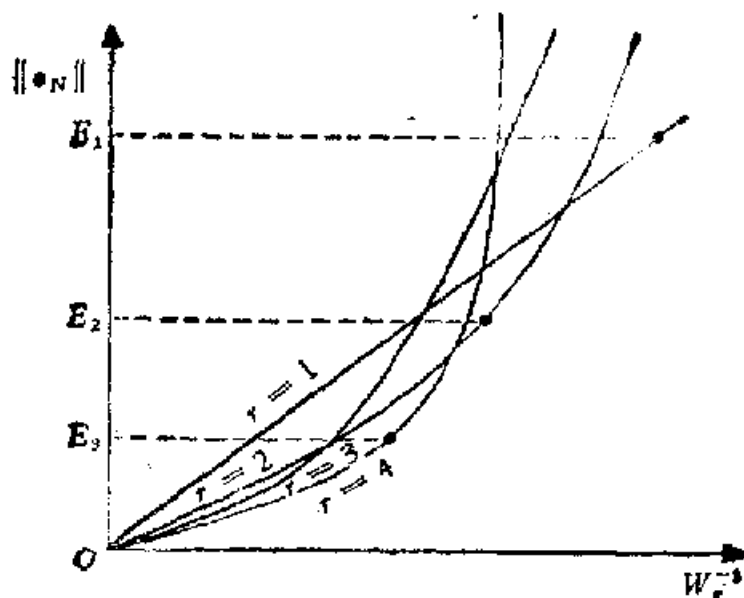


图 7.2 不同阶方法的误差与工作量倒数的关系

误差。特别当右端函数  $f(x, y)$  为  $y$  的线性函数时还可以用下面所介绍的 Runge-Kutta-Merson 公式方便地在计算过程中估计误差，从而决定缩小或放大步长，以达到控制误差的目的。

Runge-Kutta-Merson 方法是一个四阶的 Runge-Kutta 型公式，在计算过程中只要多算一个函数值，就能给出每一步上的误差，它的具体公式为

$$\begin{aligned}
 \eta_0 &= y_n, \quad k_0 = hf(\eta_0), \\
 \eta_1 &= \eta_0 + \frac{k_0}{3}, \quad k_1 = hf(\eta_1), \\
 \eta_2 &= \eta_0 + \frac{k_0 + k_1}{6}, \quad k_2 = hf(\eta_2), \\
 \eta_3 &= \eta_0 + \frac{k_0 + 3k_2}{8}, \quad k_3 = hf(\eta_3), \\
 \eta_4 &= \eta_0 + \frac{k_0 - 3k_2 + 4k_3}{2}, \quad k_4 = hf(\eta_4), \\
 y_{n+1} &= \eta_5 = \eta_0 + \frac{k_0 + 4k_3 + k_4}{6}.
 \end{aligned}$$

易证

$$\eta_1 = y\left(t_n + \frac{h}{3}\right) + O(h^2),$$

$$\eta_2 = y\left(t_n + \frac{h}{3}\right) + O(h^3),$$

$$\eta_3 = y\left(t_n + \frac{h}{2}\right) + O(h^4),$$

$$\eta_4 = y(t_n + h) + O(h^4),$$

最后

$$\eta_5 = y_{n+1} = y(t_{n+1}) + O(h^5).$$

若

$$f(t, y) = ay + bt,$$

则

$$\eta_4 = y(t_{n+1}) + O(h^5).$$

Merson 曾证明

$$\eta_4 = y(t_{n+1}) - \frac{1}{120} h^5 y^{(5)} + O(h^6),$$

$$\eta_5 = y(t_{n+1}) - \frac{1}{720} h^5 y^{(5)} + O(h^6).$$

因此,  $\eta_5 - \eta_4$  就可以用来估计局部误差, 如果这个量过大, 则可将  $h$  缩小, 反之, 则将  $h$  放大.

对于预测-校正公式来说, 可利用预测公式与校正公式的截断误差为同阶这一事实, 在计算过程中控制局部误差, 这个技巧在无迭代的 Hamming 公式中已作了介绍, 但多步公式在改变步长时应注意在新的结点上所需的  $k$  个出发值  $y$  及  $y'$  是否已经准备好.

**4.3. 方法的选取** 对于那种比较简单, 计算量不大的问题在选择方法上主要考虑到准备工作简单就行了, 此时选古典的 Runge-Kutta 方法最方便, 如果精度嫌高, 则可将步长加倍再往下算. 当然也可使用其他不需特别计算出发值的方法. 若问题虽简单(例如, 线性情况)但条件较坏, 此时若再用



Runge-Kutta 方法, 就会因为稳定性的要求, 将  $h$  限制得过小, 因而不适合的, 在这种情况下可用梯形法则, 它虽为隐式格式但若方程为线性的, 可以用直接方法解出.

对于光滑的非坏条件问题, 可采用外推法 (Bulirsch-stoer)<sup>1)</sup> 对于许多问题来说, 它是最快的一个方法, 不足之处是到目前为止还没有一个好的判别法来确定最好的阶数及步长. 便于自开始, 自动变步长与变阶的多值法, (Multivalue method) 对于区间较长而导数值相对大小变化不大时可以使用, 在一般情况下, Adams 公式是适用的.

对于那些导数有一系列间断点的问题, 例如火箭轨道, 当机器开动和关闭时, 二阶导数将出现间断. 使用 Taylor 展式所得到公式在经过这种点时, 阶数就必须加以限制, 例如火箭问题在计算到不连续点时就不能用高于一阶的方法, 如果使用  $k$  步公式, 当在  $[t_{n-k}, t_n]$  中出现间断点时, 也只能限于使用一阶方法. 所以对于间断问题使用单步公式最为合适. 对于这种问题 Runge-Kutta 法及外推法都是适合的.

最后, 对于坏条件方程组再补充几点: 如果方程组为线性的. 且次对角线上元素的模十分小的话, 可用介绍过的 Treanor 方法. 对于一般的非线性方程, 梯形法则加上适当的有理外推是行之有效的方法. 如果出现间断时可使用隐式 Runge-Kutta 公式, 也可使用 Axelsson 的适合于坏条件  $A$ -稳定的方法<sup>2)</sup>. 为了便于改变步长和阶数, 可用已介绍过的适合于坏条件的预测-校正多值方法.

---

1) Bulirsch, R. and Stoer, J. (1966), *Num. Math.*, 8, pp. 1~13.

2) Axelsson, O., *BIT.*, 9(1969), pp. 185~199.

## 第八章 边值问题的数值解法

本章将以二阶微分方程为例来讨论两点边值问题,即考虑形如

$$y'' = f(x, y, y'), \quad a < x < b \quad (1)$$

的微分方程,其边值条件可分为下面三类:

第一边值条件

$$y(a) = \alpha, \quad y(b) = \beta; \quad (1-1)$$

第二边值条件

$$y'(a) = \alpha, \quad y'(b) = \beta; \quad (1-2)$$

第三边值条件

$$\begin{aligned} y'(a) - \alpha_0 y(a) &= \alpha_1, \quad y'(b) + \beta_0 y(b) = \beta_1, \\ \alpha_0 \geq 0, \quad \beta_0 \geq 0, \quad \alpha_0 + \beta_0 &> 0, \end{aligned} \quad (1-3)$$

当(1)式为线性方程时,可写为

$$y'' + p(x)y' + q(x)y = f(x). \quad (2)$$

将上式两端乘以  $e^{\int p(x)dx}$ , 即得

$$(y'e^{\int p(x)dx})' + qe^{\int p(x)dx}y = fe^{\int p(x)dx}, \quad (3)$$

令  $t = \int e^{-\int p(x)dx} dx \equiv \varphi(x)$ , 则方程(2)就化为

$$\frac{d^2 y}{dt^2} + Q(t)y = R(t) \quad (4)$$

的形式. 容易验证,在这种变换下,边值条件的形状不变. 因此,下面考虑线性的情况时,可以假设方程中不显含  $y'$ .

本章讨论的内容共分三部分: 第一部分介绍边值问题的差分方法及解差分方程的稳定追赶法; 第二部分介绍样条

(Spline)函数在二点边值问题上的应用; 第三部分介绍适合于非线性方程边值问题的试射法及迭代解法.

## § 1 解线性边值问题的差分方法

**1.1. 差分方程的建立** 今就线性的情况来讨论, 这时微分方程的形式为

$$y'' - q(x)y = r(x), \quad q(x) \geq 0, \quad a < x < b. \quad (5)$$

用分点  $x_0 = a, x_1 = a + h, \dots, x_n = a + nh = b; h = \frac{b-a}{n}$  将区间  $[a, b]$  划分为  $n$  等分,  $x_i$  称为结点. 现在我们把在  $a < x < b$  中求解  $y(x)$  的问题化为在这些结点  $x_i$  上求  $y(x_i)$  的近似值的问题. 为此, 我们将(5)中的变量进行离散. 具体做法如下:

对于内部节点, 将二阶导数用二阶中心差商来表示, 便得到关系式

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} = y''(x_i) + \frac{h^2}{12} y^{(4)}(\xi_i),$$

$$i = 1, 2, \dots, n-1, \quad x_{i-1} < \xi_i < x_{i+1},$$

将它代入(5)便得到在结点  $x_i$  上  $y_i = y(x_i)$  所满足的关系式

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - q_i y_i = r_i + \frac{h^2}{12} y^{(4)}(\xi_i),$$

其中  $q_i = q(x_i), r_i = r(x_i)$ . 上式中略去最后一项, 我们就可得到方程(5)的近似差分方程

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - q_i y_i = r_i, \quad i = 1, 2, \dots, n-1. \quad (6)$$

这是含有  $n+1$  个未知数  $y_i (i=0, 1, \dots, n)$  的线性代数方程组, 方程的个数为  $n-1$ , 要使方程组(6)有唯一解, 还需要由边值条件补充两个方程.

对于第一边值条件,直接就得到了另二个方程

$$y_0 = \alpha,$$

$$y_n = \beta.$$

于是,就得到第一边值问题的差分方程组:

$$\begin{cases} \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - q_i y_i = r_i, & i=1, 2, 3, \dots, n-1, \\ y_0 = \alpha, y_n = \beta. \end{cases} \quad (7)$$

但对于第二及第三边值条件,由于包含了导数,故边值条件也必须用差商来近似表示. 因为我们不能利用在区间  $[a, b]$  以外的点,所以在引进  $y'_0$  (即  $y'(a)$ ) 或  $y'_n$  (即  $y'(b)$ ) 的近似式时,就不能利用中心差商. 如果只要求误差是  $O(h)$ , 自然可以用最简单的近似公式

$$y'_0 \approx \frac{y_1 - y_0}{h}, \quad y'_n \approx \frac{y_n - y_{n-1}}{h}. \quad (8)$$

但要使误差达到  $O(h^2)$ , 则需利用 Newton 等距插值公式,可以得到近似公式

$$y'_0 \approx \frac{1}{h} \left[ 4y_0 - \frac{1}{2} \Delta^2 y_0 \right] = \frac{-y_2 + 4y_1 - 3y_0}{2h}, \quad (9)$$

$$y'_n \approx \frac{1}{h} \left[ 4y_{n-1} + \frac{1}{2} \Delta^2 y_{n-2} \right] = \frac{3y_n - 4y_{n-1} + y_{n-2}}{2h}. \quad (10)$$

把这些近似式代入边值条件中,再与方程组(6)联立,就可以得到对应的差分方程组. 对于第三边值问题,采用近似式(9), (10), 就得到差分方程组

$$\begin{cases} \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - q_i y_i = r_i, & i=1, \dots, n-1, \\ \frac{-y_2 + 4y_1 - 3y_0}{2h} - \alpha_0 y_0 = \alpha_1, \\ \frac{3y_n - 4y_{n-1} + y_{n-2}}{2h} + \beta_0 y_n = \beta_1. \end{cases} \quad (11)$$

这样我们便通过离散化的过程将微分方程的边值问题化

为一个差分方程的边值问题了. 假设上述的差分方程已经建立, 此时人们自然会问, 这样建立起来的差分方程是否能提供微分方程的近似解, 以及用什么方法去得到差分方程的解. 也就是说, 还需要进一步研究下面的问题:

1° 差分方程的可解性以及解法.

2° 差分方程解对微分方程解的收敛性, 以及对精确度的估计.

**1.2. 差分方程的可解性** 从上面建立的差分方程可以看出, 不管是那一种边值条件, 它们都表示为  $y_0, \dots, y_n$  的一个线性代数方程组. 因此, 要证明它的解的存在唯一性, 只要证明对应的齐次方程组只有零解就行了. 为此, 我们引进下述的极值原理.

**引理 1** 设给定一组数  $y_0, y_1, \dots, y_n$ , 它们满足条件

$$l(y_i) = \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - q_i y_i \geq 0, \quad (12)$$

$$i = 1, 2, \dots, n-1; \quad q_i \geq 0.$$

如果  $y_i$  不全相等, 则这组数中正的最大值只能在  $y_0$  或  $y_n$  上达到.

**证明** 设  $M = \max_{0 \leq i \leq n} y_i > 0$ . 可用反证法来证明. 设  $M$  在  $y_1, y_2, \dots, y_{n-1}$  中达到. 因为  $y_i$  不全相等, 所以总有某一个  $i_0$ , ( $1 \leq i_0 \leq n-1$ ), 使得  $y_{i_0} = M$ , 而  $y_{i_0-1}$  及  $y_{i_0+1}$  中确实至少有一个小于  $M$ . 此时就有

$$l(y_{i_0}) = \frac{y_{i_0+1} - 2y_{i_0} + y_{i_0-1}}{h^2} - q_{i_0} y_{i_0}$$

$$< \frac{M - 2M + M}{h^2} - q_{i_0} M = -q_{i_0} M.$$

由于  $q_{i_0} \geq 0, M > 0$ , 因此  $l(y_{i_0}) < 0$ . 这与假设矛盾. 故引理得证.

类似地可以证明, 当  $l(y_i) \leq 0$  时, 如果  $y_i$  不全于零, 则它们中负的最小值也只能在  $y_0$  或  $y_n$  上达到.

差分方程组(7)对应的齐次方程是

$$\begin{aligned} l(y_i) &= 0, \quad i=1, 2, \dots, n-1, \\ y_0 &= 0, \\ y_n &= 0. \end{aligned}$$

据引理 1, 立得所有  $y_i (i=1, 2, \dots, n-1)$  都必须为零的结论, 于是非齐次方程组必存在唯一的解.

现在我们来证明第三边值问题的解也是存在的和唯一的. 相应齐次方程组是

$$\begin{aligned} l(y_i) &= \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - q_i y_i = 0, \\ i &= 1, 2, \dots, n-1. \end{aligned} \quad (13)$$

$$\frac{-y_2 + 4y_1 - 3y_0}{2h} - \alpha_0 y_0 = 0, \quad (14)$$

$$\frac{3y_n - 4y_{n-1} + y_{n-2}}{2h} + \beta_0 y_n = 0, \quad (15)$$

$$\alpha_0 \geq 0, \quad \beta_0 \geq 0, \quad -\alpha_0 + \beta_0 > 0.$$

只要证明它们不可能有非零解就行了. 事实上, 从(13)中  $i=1$  的方程与(14)联立消去  $y_2$  即得

$$y_1 = \frac{1 + \alpha_0 h}{1 - \frac{1}{2} q_1 h^2} y_0 \quad (16)$$

同样从(13)中  $i=n-1$  的方程与(15)联立消去  $y_{n-2}$ , 可得

$$y_{n-1} = \frac{1 + \beta_0 h}{1 - \frac{1}{2} q_{n-1} h^2} y_n. \quad (17)$$

因为  $q_i \geq 0$ , 我们可设  $h$  充分小, 使得  $0 < 1 - \frac{1}{2} q_1 h^2 \leq 1$  及  $0 < 1 - \frac{1}{2} q_{n-1} h^2 \leq 1$ , 故  $y_1$  与  $y_0$  同号,  $y_{n-1}$  与  $y_n$  同号. 又由

于  $\alpha_0 \geq 0, \beta_0 \geq 0$ , 故  $|y_1| \geq |y_0|, |y_{n-1}| \geq |y_n|$ . 由上面的引理, 只有所有的  $y_i$  全相等, 即  $y_0 = y_1 = \cdots = y_n$ , 代入(16)及(17)式中, 因为  $\alpha_0, \beta_0$  中至少有一个不是零. 故

$$\frac{1 + \alpha_0 h}{1 - \frac{1}{2} q_1 h^2} \text{ 与 } \frac{1 + \beta_0 h}{1 - \frac{1}{2} q_{n-1} h^2}$$

至少有一个大于 1. 因此就只有  $y_0 = y_1 = \cdots = y_n = 0$ , 这正是我们要证明的, 这就证明了第三边值问题差分方程解的存在唯一性.

**1.3. 差分方程解的收敛性** 今以第一边值问题为例来讨论. 我们设  $y_i$  为差分方程(7)的解.  $y(x_i)$  表示微分方程的精确解在  $x_i$  点的数值. 一般来讲  $y(x_i)$  不满足(7)除非它是不大于 3 次的多项式. 事实上  $y(x_i)$  满足的方程应为:

$$\begin{cases} \frac{y(x_{i+1}) - 2y(x_i) + y(x_{i-1}))}{h^2} - q_i y(x_i) = r_i + \frac{h^2}{12} y^{(4)}(\xi_i), \\ x_{i-1} < \xi_i < x_{i+1}, \\ y(x_0) = \alpha, y(x_n) = \beta. \end{cases} \quad (18)$$

这里我们假设四阶导数  $y^{(4)}(x)$  是存在的, 我们引进误差记号

$$e_i = y(x_i) - y_i, \quad i = 0, 1, \dots, n,$$

并对它进行估计. 为此我们从(18)减去(7)即得误差  $e_i$  满足的差分方程

$$\begin{cases} l(e_i) = \frac{e_{i+1} - 2e_i + e_{i-1}}{h^2} - q_i e_i = \frac{h^2}{12} y^{(4)}(\xi_i), \\ e_0 = e_n = 0. \end{cases} \quad (19)$$

由于  $\xi_i$  一般来说并不知道, 故不能得到  $e_i$  的精确表达式, 为了得到它的估计式还得借助下面的引理:

**引理 2** 如果给定两个数组  $y_0, y_1, \dots, y_n$  及  $Y_0, Y_1, \dots, Y_n$ , 它们满足关系

$$\begin{cases} l(Y_i) \leq -|l(y_i)|, & i=1, 2, \dots, n-1, \\ Y_0 \geq |y_0|, Y_n \geq |y_n|, \end{cases} \quad (20)$$

则必有

$$Y_i \geq |y_i|, \quad i=1, 2, \dots, n-1.$$

证明 条件(20)可写成

$$l(Y_i - y_i) \leq 0, \quad l(Y_i + y_i) \leq 0, \quad i=1, 2, \dots, n-1,$$

$$Y_0 - y_0 \geq 0, \quad Y_0 + y_0 \geq 0,$$

$$Y_n - y_n \geq 0, \quad Y_n + y_n \geq 0.$$

因此由引理1即知

$$Y_i - y_i \geq 0, \quad Y_i + y_i \geq 0, \quad i=1, 2, \dots, n-1,$$

即  $Y_i \geq |y_i|$ .

利用引理2, 我们令  $M_4 = \max_{a \leq x \leq b} |y^{(4)}(x)|$ , 考虑下列方程

$$\begin{cases} l(\eta_i) = \frac{\eta_{i+1} - 2\eta_i + \eta_{i-1}}{h^2} - q_i \eta_i = -\frac{h^2}{12} M_4, \\ \eta_0 = \eta_n = 0, \quad i=1, 2, \dots, n-1, \end{cases} \quad (21)$$

则有  $\eta_i \geq |\varepsilon_i|$ , (21)虽比(19)简单些, 但由于左边含  $q_i \eta_i$  这一项, 仍不易求解. 注意到  $q_i \geq 0$  的特点, 我们还可以引进更简单的方程

$$\begin{cases} \lambda(\rho_i) = \frac{\rho_{i+1} - 2\rho_i + \rho_{i-1}}{h^2} = -\frac{h^2}{12} M_4, \\ \rho_0 = \rho_n = 0. \end{cases} \quad (22)$$

由(22)—(21)得到

$$\lambda(\rho_i - \eta_i) = -q_i \eta_i \leq 0, \quad i=1, 2, \dots, n-1,$$

而  $\rho_0 - \eta_0 = \rho_n - \eta_n = 0$ , 故由极值原理(注意  $q_i \equiv 0$  时  $l(y_i)$  即成为  $\lambda(y_i)$ )知  $\rho_i - \eta_i \geq 0$ , 因此  $|\varepsilon_i| \leq \eta_i \leq \rho_i$ . 可以验证方程(22)的解是二次多项式

$$\rho(x) = \frac{h^2}{24} M_4 (x-a)(b-x).$$



在  $x_i$  点上的值, 因为  $\rho(x)$  在点  $x = \frac{a+b}{2}$  达到最大值,

$$\rho\left(\frac{a+b}{2}\right) = \frac{h^2 M_4 (b-a)^2}{96}.$$

因此就得到估计式

$$|\varepsilon_i| \leq \rho_i = \rho(x_i) \leq \frac{h^2 M_4 (b-a)^2}{96}.$$

从上式可以看出, 当  $h \rightarrow 0$  时  $\varepsilon_i \rightarrow 0$ , 这也就证明: 当  $h \rightarrow 0$  时差分方程组的解收敛到原微分方程的解.

1.4. 解差分方程组的追赶法 我们先考虑第一边值问题的差分解法. 将(7)中的  $y_0, y_n$  消去可得

$$\begin{cases} -(2+q_1 h^2)y_1 + y_2 = h^2 r_1 - \alpha, & (23-1) \\ y_1 - (2+q_2 h^2)y_2 + y_3 = h^2 r_2, & (23-2) \end{cases}$$

.....

$$y_{n-3} - (2+q_{n-2} h^2)y_{n-2} + y_{n-1} = h^2 r_{n-2}, \quad (23-n-2)$$

$$y_{n-2} - (2+q_{n-1} h^2)y_{n-1} = h^2 r_{n-1} - \beta. \quad (23-n-1)$$

方程组(23)的系数矩阵为三对角的. 又因  $q_i \geq 0$ , 故可按自然顺序消元, 在用一般的主元素消去法来解上述方程组时, 可以利用方程组的特点将步骤加以简化.

先从(23-1)中解出  $y_1$

$$y_1 = \frac{1}{2+q_1 h^2} y_2 - \frac{h^2 r_1 - \alpha}{2+q_1 h^2} = u_1 y_2 + v_1. \quad (24-1)$$

将(24-1)代入(23-2)并解出  $y_2$

$$\begin{aligned} y_2 &= -\frac{1}{(2+q_2 h^2) - u_1} y_3 \\ &\quad + \frac{v_1 - h^2 r_2}{(2+q_2 h^2) - u_1} = u_2 y_3 + v_2. \end{aligned} \quad (24-2)$$

一般可设

$$y_i = u_i y_{i+1} + v_i, \quad i=1, 2, \dots, n-1. \quad (24-i)$$

为了求得  $u_i$  及  $v_i$  的递推公式, 可将 (24-i) 代入 (23-i+1), 解出  $y_{i+1}$

$$y_{i+1} = \frac{1}{(2+q_{i+1}h^2) - u_i} y_{i+2} + \frac{v_i - h^2 r_{i+1}}{(2+q_{i+1}h^2) - u_i}.$$

因此有

$$u_{i+1} = \frac{1}{(2+q_{i+1}h^2) - u_i}, \quad v_{i+1} = u_{i+1}(v_i - h^2 r_{i+1}). \quad (25)$$

(25) 为  $u_i, v_i$  的递推关系式, 解出  $u_i, v_i$  之后由 (24) 就能求得  $y_i$ , 从 (24) 式可以看出,  $u_0, v_0$  应满足关系式

$$y_0 = u_0 y_1 + v_0.$$

已知  $y_0 = \alpha$ , 故应假设  $u_0 = 0, v_0 = \alpha$ , 这样便可逐个地按下标从小到大的次序求得  $u_i$  和  $v_i$  ( $i=1, \dots, n-1$ ), 又因  $y_n = \beta$  故由 (24) 式可以按下标从大到小的次序逐个求得  $y_i$  ( $i=n-1, \dots, 1$ ), 求  $u_i, v_i$  的过程下标由小到大, 称为追的过程, 求  $y_i$  的过程正好相反, 下标由大到小称为赶的过程, 这就是解差分方程的追赶法. 上述计算过程实际上与主元素消去法是一致的. 利用公式 (25) 计算  $u_i$  和  $v_i$  的过程相当于消元过程, 而用 (24) 计算  $y_i$  的过程相当于回代过程.

下面我们将简单地证明, 追赶法有误差在传播中不增长的优点, 也就是说在第  $i$  个节点上对  $u_i, v_i, y_i$  进行计算时产生的误差, 那么在第  $i+1$  个节点或  $i-1$  个节点 (注意到由  $y_i$  来计算  $y_{i-1}$  的) 这个误差不会增长.

我们首先证明, 若在计算系数  $u_i$  中产生一误差  $\delta u_i$ , 由此影响到计算  $u_{i+1}$  中也产生了误差  $\delta u_{i+1}$ , 则有  $|\delta u_{i+1}| \leq |\delta u_i|$ .

事实上, 由于  $u_i + \delta u_i$  与  $u_{i+1} + \delta u_{i+1}$  满足方程 (25)

$$u_{i+1} + \delta u_{i+1} = \frac{1}{[(2+q_{i+1}h^2) - (u_i + \delta u_i)]}. \quad (26)$$

将 (26) 与 (25) 相减, 便得到

$$\delta u_{i+1} = \frac{\delta u_i}{[(2+q_{i+1}h^2) - (u_i + \delta u_i)][(2+q_{i+1}h^2) - u_i]} \\ \approx u_{i+1}^2 \delta u_i.$$

我们用归纳法可证  $0 < u_{i+1} < 1$ . 如果  $0 < u_i < 1$ , 则因  $q_{i+1} \geq 0$ , 所以  $2 + q_{i+1}h^2 - u_i > 1$ , 因此由(25)知  $0 < u_{i+1} < 1$ . 于是

$$|\delta u_{i+1}| < |\delta u_i|.$$

再看求  $v_i$  及  $y_i$  的过程, 与上面推导一样. 由(25)及(24)可知

$$\delta v_{i+1} \approx u_{i+1} \delta v_i, \quad \delta y_i \approx u_i \delta y_{i+1}.$$

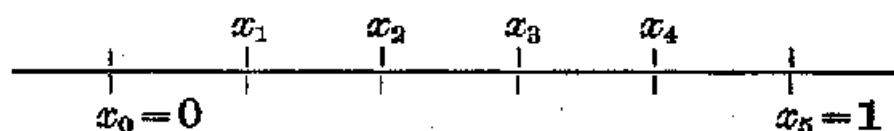
于是由  $0 < u_i < 1$  即知  $|\delta v_{i+1}| < |\delta v_i|$ ,  $|\delta y_i| < |\delta y_{i+1}|$ .

这就说明追赶法在计算过程中是稳定的, 故它是解差分方程组的有效方法.

例: 考虑边值问题

$$y'' = 0, \quad y(0) = 0, \quad y(1) = 1.$$

取  $h = 0.2$ , 则区间  $[0, 1]$  被分成 5 等分,  $x_1, x_2, x_3, x_4$  为其内点.



在内点  $x_i (i = 1, 2, 3, 4)$  上我们以二阶中心差商

$$\frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}$$

来代替二阶导数  $y''$ , 便得差分方程

$$y_{n+1} - 2y_n + y_{n-1} = 0, \quad n = 1, 2, 3, 4 \\ y_0 = 0, \quad y_5 = 1,$$

它是一个线性方程组, 可将它化成

$$y_n = u_n y_{n+1} + v_n \tag{27}$$

的形式, 其中

$$u_{n+1} = \frac{1}{2-u_n}, \quad v_n = 0. \quad (28)$$

由于  $y_0=0$ , 由 (27) 式得  $u_0=v_0=0$ . 由 (28) 令  $n=1$  立得  $u_1=\frac{1}{2}$ , 然后可逐个求得  $u_2=\frac{2}{3}$ ,  $u_3=\frac{3}{4}$ ,  $u_4=\frac{4}{5}$ . 再由 (27) 与  $y_5=1$  逐个求得  $y_4=\frac{4}{5}$ ,  $y_3=\frac{3}{5}$ ,  $y_2=\frac{2}{5}$ ,  $y_1=\frac{1}{5}$ .

现在我们来考虑第三边值问题的差分解法, 这时

$$\begin{aligned} y'' &= q(x)y + r(x), \quad a < x < b, \\ y'(a) &= \alpha_0 y(a) + \alpha_1, \\ y'(b) &= -\beta_0 y(b) + \beta_1. \end{aligned}$$

可以类似地建立差分方程, 所不同的是需要采取新的分点

$$x_i = a + \left(i - \frac{1}{2}\right)h, \quad h = \frac{b-a}{n}, \quad i = 0, 1, \dots, n+1,$$

在内点上的差分方程为

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - q_i y_i = r_i, \quad i = 1, 2, \dots, n.$$

在边点上均以差商来代替导数, 于是从边值条件便又得两个方程

$$\begin{aligned} \frac{y_1 - y_0}{h} &= \alpha_0 \frac{y_1 + y_0}{2} + \alpha_1, \\ \frac{y_{n+1} - y_n}{h} &= -\beta_0 \frac{y_{n+1} + y_n}{2} + \beta_1. \end{aligned}$$

这样便得到  $n+2$  个方程的线性方程组, 经过整理后可以写成

$$y_{i+1} - (2 + q_i h^2) y_i + y_{i-1} = h^2 r_i, \quad i = 1, 2, \dots, n, \quad (29)$$

$$y_0 = \frac{1 - \frac{h}{2} \alpha_0}{1 + \frac{h}{2} \alpha_0} y_1 - \frac{\alpha_1 h}{1 + \frac{h}{2} \alpha_0}, \quad (30)$$

$$y_n = \frac{1 + \frac{h}{2}\beta_0}{1 - \frac{h}{2}\beta_0} y_{n+1} - \frac{\beta_1 h}{1 - \frac{h}{2}\beta_0}, \quad (31)$$

与第一边值问题一样，我们把上述方程组用 Gauss 主元素消去法化成一组递推公式

$$y_i = u_i y_{i+1} + v_i, \quad i = n+1, \dots, 0, \quad (32)$$

其中

$$u_{i+1} = \frac{1}{(2 + q_{i+1}h^2) - u_i}, \quad v_{i+1} = u_{i+1}(v_i - h^2 r_{i+1}), \quad (33)$$

$$u_0 = \frac{1 - \frac{h}{2}\alpha_0}{1 + \frac{h}{2}\alpha_0}, \quad v_0 = \frac{-\alpha_1 h}{1 + \frac{h}{2}\alpha_0}.$$

先由(33)解出  $u_i$  及  $v_i$  ( $i=1, 2, \dots, n+1$ )，然后再由(32)求出  $y_i$  ( $i=n+1, \dots, 0$ )。所不同的是现在  $y_{n+1}$  是未知的，需将(32)中的

$$y_n = u_n y_{n+1} + v_n$$

与(31)式联立起来解出  $y_n, y_{n+1}$  然后就可由(32)逐个求出  $y_i$  ( $i=n-1, \dots, 0$ )。与前一样，也可以证明计算过程是稳定的。同样可先证  $0 < u_i < 1$ ，即得  $|\delta u_{i+1}| < |\delta u_i|$ ，以及  $|\delta v_{i+1}| < |\delta v_i|$ ， $|\delta y_i| < |\delta y_{i+1}|$ 。

## §2 样条函数简介及其在两点边值问题上的应用

本节仅列出样条函数的定义，并通过一个实例来说明如何用样条函数来解常微分方程的两点边值问题，至于样条函数的进一步性质和如何来进行函数插值可参阅《数值逼近方法》一书。

样条(spline)是绘图员用来划光滑曲线的一种简单工具,为了使一些指定点(称为样点)联结成一条光滑曲线,常常用细长木条(称为样条)把它用压铁压在样点上,必要时调整其位置,使样条自动弯成一条通过样点的光滑曲线.在工程上的所谓船体、车体的放样就是如此作的.

下面给出样条函数的数学描述

**2.1. 样条函数的定义** 假设给定  $n+1$  个点  $a=x_0 < x_1 < \dots < x_n=b$ , 称为结点, 如果函数  $S(x)$  满足下列条件:

1) 在每个子区间  $[x_{k-1}, x_k]$  ( $k=1, 2, \dots, n$ ) 上,  $S(x)$  是一个不超过  $m$  次的多项式;

2) 函数  $S(x)$  在整个区间  $[a, b]$  上具有直到  $m-1$  阶的连续导数; 则称  $S(x)$  为以  $\{x_k\}$  为结点的  $m$  次样条函数.

在样条函数中, 三次样条函数是最常用的, 下面的讨论仅限于此.

由样条函数的定义可知, 样条函数的基本思想是很简单的, 它无非是把区间分段用低次多项式来逼近使得整个函数成为“装配式”的, 同时又保证联接处有一定的光滑性, 这样它既保留了多项式在表达形式上的简便性, 又克服了高次多项式插值中的问题.

**2.2. 样条函数在解两点边值问题上的应用** 现介绍用三次样条函数来求线性常微分方程二点边值问题的近似解, 考察二阶线性微分方程

$$u'' + p(x)u' + q(x)u = r(x), \quad (34)$$

其边界条件是

$$\alpha_0 u + \beta_0 u' = r_0, \quad x = x_0 = a, \quad (35-1)$$

$$\alpha_n u - \beta_n u' = r_n, \quad x = x_n = b. \quad (35-2)$$

设用分点

$$a = x_0 < x_1 < \dots < x_n = b$$

将区间  $[a, b]$  分成  $n$  段,  $S(x)$  是以  $\{x_k\}$  为结点的三次样条函数, 可以证明,  $S(x)$  可表示为

$$S(x) = e + f(x - x_0) + \frac{1}{2} g(x - x_0)^2 + \frac{1}{6} \sum_{k=0}^{n-1} d_k (x - x_k)_+^3, \\ a \leq x \leq b$$

的形式, 其中引进的记号  $x_+$ , 其定义为:

$$x_+ = \begin{cases} x, & x \geq 0, \\ 0, & x < 0; \end{cases}$$

而  $e, f, g, d_0, d_1, \dots, d_{n-1}$  是  $n+3$  个待定常数.

根据定义,  $S'(x), S''(x)$  都是  $[a, b]$  上的连续函数, 且

$$S'(x) = f + g(x - x_0) + \frac{1}{2} \sum_{k=0}^{n-1} d_k (x - x_k)_+^2,$$

$$S''(x) = g + \sum_{k=0}^{n-1} d_k (x - x_k)_+.$$

把  $S(x), S'(x), S''(x)$  代入 (34), 便得到关系式

$$g + \sum_{k=0}^{n-1} d_k (x_i - x_k)_+ + p_i \left[ f + g(x_i - x_0) + \frac{1}{2} \sum_{k=0}^{n-1} d_k (x_i - x_k)_+^2 \right] \\ + q_i \left[ e + f(x_i - x_0) + \frac{1}{2} g(x_i - x_0)^2 \right. \\ \left. + \frac{1}{6} \sum_{k=0}^{n-1} d_k (x_i - x_k)_+^3 \right] = r_i, \quad i = 0, 1, \dots, n, \quad (36)$$

其中  $p_i = p(x_i), q_i = q(x_i), r_i = r(x_i)$  等等, 而边界条件 (35-1), (35-2) 此时为

$$\alpha_0 e + \beta_0 f = r_0, \quad (37)$$

$$\alpha_n \left[ e + f(x_n - x_0) + \frac{1}{2} g(x_n - x_0)^2 \right. \\ \left. + \frac{1}{6} \sum_{k=1}^{n-1} d_k (x_n - x_k)_+^3 \right] \\ - \beta_n \left[ f + g(x_n - x_0) + \frac{1}{2} \sum_{k=1}^{n-1} d_k (x_n - x_k)_+^2 \right] = r_n, \quad (38)$$

于是 (36), (37), (38) 联在一起就构成了一组关于  $e, f, g,$

$d_0, d_1, \dots, d_{n-1}$  的  $n+3$  个未知量的方程组, 从而可求得(34), (35)的一个近似解  $S(x)$ .

例: 给定微分方程

$$u'' + u + 1 = 0 \quad (39)$$

和边界条件

$$u = 0, \quad x = 0, \quad (40-1)$$

$$u = 0, \quad x = 1. \quad (40-2)$$

这个方程有简单的解析解, 它在区间  $[0, 1]$  的中点  $x=0.5$  处的值是 0.13949. 现在把区间  $[0, 1]$  分成三等分, 分点是 0,  $1/3$ ,  $2/3$ , 1, 步长  $h=1/3$ . 根据前面的讨论, 有 6 个未知数  $e, f, g, d_0, d_1, d_2$ . 由边界条件 (40-1) 很容易推出  $e=0$ . 现设

$$S(x) = fx + gx^2 + d_0x_+^3 + d_1\left(x - \frac{1}{3}\right)_+^3 + d_2\left(x - \frac{2}{3}\right)_+^3,$$

则有

$$S'(x) = f + 2gx + 3d_0x_+^2 + 3d_1\left(x - \frac{1}{3}\right)_+^2 + 3d_2\left(x - \frac{2}{3}\right)_+^2,$$

$$S''(x) = 2g + 6d_0x + 6d_1\left(x - \frac{1}{3}\right)_+ + 6d_2\left(x - \frac{2}{3}\right)_+.$$

以此代入 (40-2) 和 (39), 并取  $x$  为 1,  $2/3$ ,  $1/3$ , 0, 则得下列关系式:

$$d_2 + 8d_1 + 27d_0 + 27g + 27f = 0,$$

$$55d_2 + 116d_1 + 189d_0 + 81g + 275f = -27,$$

$$55d_1 + 116d_0 + 66g + 18f = -27,$$

$$55d_0 + 57g + 9f = -27,$$

$$2g = -1.$$

这组方程的解为:

$$f = 0.540814, \quad g = -0.5, \quad d_0 = -0.061224,$$

$$d_1 = 0.061224, \quad d_2 = 0.061224.$$



因此  $u$  的近似解为

$$\begin{aligned}\bar{u}(x) = & 0.540814x - 0.5x^2 - 0.061224x^3 \\ & + 0.061224\left(x - \frac{1}{3}\right)_+^3 + 0.061224\left(x - \frac{2}{3}\right)_+^2, \\ & 0 \leq x \leq 1.\end{aligned}$$

而  $\bar{u}\left(\frac{1}{2}\right) = 0.1436797$  与精确解相差 3%.

这个极简单的例子, 阐明了用三次样条函数解常微分方程二点边值问题的方法.

### §3 试 射 法

试射法(shooting)可用来解二阶或高阶的线性或非线性的常微分方程. 这个方法的实质在于把边值问题化为初值问题来解. 此时可以采用已讨论过的各种初值问题的单步法或多步法进行求解. 下面将以第一边值问题与第三边值问题为例来介绍试射法是怎样用的.

今考虑二阶方程的第一边值问题

$$\begin{cases} y'' = f(x, y, y'), \end{cases} \quad (41)$$

$$\begin{cases} y(a) = \alpha, y(b) = \beta. \end{cases} \quad (42)$$

试射法的思想就是设法确定  $y'(a)$  的值  $m$  使满足初值问题  $y(a) = \alpha, y'(a) = m$  的解也满足另一边值条件  $y(b) = \beta$ . 也就是要从微分方程(41)的经过点  $(a, \alpha)$  而具有不同斜率的积分曲线中, 去寻找一条经过点  $(b, \beta)$  的曲线. 首先我们可以根据经验, 或对方程作定性分析, 或按照实际存在的运动规律, 选取一个斜率  $m_1$ . 我们就用这个斜率进行试算, 即解初值问题

$$y'' = f(x, y, y'), y(a) = \alpha, y'(a) = m_1.$$

这样便得到一个解  $y_1(x)$ , 如果  $y_1(b) = \beta$  或  $|y_1(b) - \beta| < \varepsilon$  ( $\varepsilon$  为允许误差), 则  $y_1(x)$  即为所求的解. 否则, 可根据  $\beta_1 = y_1(b)$  与  $\beta$  的差距来适当地将  $m_1$  修改为  $m_2$  (例如取  $m_2 = \frac{\beta}{\beta_1} m_1$ ). 这时, 再解初值问题

$$y'' = f(x, y, y'), \quad y(a) = \alpha, \quad y'(a) = m_2.$$

于是又可得到另一个解  $y_2(x)$ . 仿前, 如果  $y_2(b) = \beta$  或满足不等式  $|y_2(b) - \beta| < \varepsilon$ , 则  $y_2(x)$  即所求的解, 否则再对  $m$  值作适当修改.

事实上, 假定  $\tilde{y}(x)$  为初值问题

$$y'' = f(x, y, y'), \quad y(a) = \alpha, \quad y'(a) = m$$

的解, 显然, 在  $a, \alpha$  为确定的情况下, 终值  $\tilde{y}(b)$  是初始斜率  $m$  的函数:

$$\tilde{y}(b) = F(m).$$

于是, 问题就变成求  $m$  使

$$F(m) = \beta.$$

这是一个代数方程(线性或非线性的), 但由于  $F(m)$  的具体表达式往往是不明确的, 所以解这方程一般是困难的. 最简单的办法是由  $m_1$  及  $m_2$  用线性插值法来求出新的  $m$  值

$$m = m_1 + \frac{m_2 - m_1}{\beta_2 - \beta_1} (\beta - \beta_1).$$

当然, 用线性插值的依据是不足的, 可是却往往用它来调节初值, 因为它是方便易行. 值得指出的是, 如果有更适当的插值公式可利用的话, 那末就有可能使尝试的次数有效地减少.

例. 解边值问题

$$\begin{cases} y'' = -10y^3, & 0 < x < 1, \\ y(0) = 0, & y(1) = 1. \end{cases}$$

我们取第一次尝试解  $y_1(x)$  在  $(0, 0)$  点的斜率为  $m_1 = 1$ ,

通过计算得到  $y_1(b) = \beta_1 = 0.646$ , 它与  $\beta$  相差尚远, 显然不符合要求. 接下去我们取  $m_2 = \frac{\beta}{\beta_1} m_1 = 1.55$ , 得到  $\beta_2 = 0.610$ . 此时由于仍不满足要求, 所以就用线性插值公式

$$m_3 = m_1 + \frac{m_2 - m_1}{\beta_2 - \beta_1} (\beta - \beta_1),$$

得到  $m_3 = -4.41$ , 相应的  $\beta_3 = 0.685$ . 它离  $\beta = 1$  又近了一些.

我们再从  $m_3$  重新开始按比例取  $m_4 = \frac{\beta}{\beta_3} m_3 = -6.44$ .

求得  $\beta_4 = 1.616$ , 它大于 1. 再由  $m_3, m_4$  作线性插值, 得  $m_5 = -5.1$ . 通过计算得  $\beta_5 = 1.052$ . 由于此值已比较接近 1,

故可再按比例取  $m_6 = \frac{\beta}{\beta_5} m_5 = -4.847$ , 算得  $\beta_6 = 1.023$ . 再

按比例取  $m_7 = -4.74$ , 算得  $\beta_7 = 0.868$ . 再由  $m_6, m_7$  作线性插值求出  $m_8 = -4.83$ . 此时算得  $\beta_8 = 0.999$ . 它与所要求的  $\beta = 1$  只相差 0.001, 计算可到此结束. 计算结果见第 146 页表. 初值问题的解开始二点是用 Runge-Kutta 公式计算的, 后面各点是按 Störmer 公式计算的.

上述具体步骤主要考虑怎样更快地达到要求, 在做法上似乎缺少规律性. 从程序设计方便的角度来看, 可以在求得  $m_1, m_2$  后, 若所得  $\beta_1, \beta_2$  都不符合要求, 则随后的  $m_3, m_4, \dots$  都按线性插值公式进行修改, 直到  $\beta$  值满足要求时停止计算. 如果一直不满足要求, 则还需灵活运用例中所采取的步骤, 或用非线性的插值公式.

对于更一般的第三边值条件也可用试射法进行求解. 第三边值条件为:

$$p_1 y(a) + q_1 y'(a) = \alpha_1, \quad (43)$$

$$p_2 y(b) + q_2 y'(b) = \alpha_2, \quad (44)$$

其中  $p_i, q_i, \alpha_i, i=1, 2$  均为常数. 对于这类条件的试射法是

	$m_1=1$	$m_2=1.55$	$m_3=-4.41$	$m_4=-6.44$	$m_5=-5.1$	$m_6=-4.847$	$m_7=-4.74$	$m_8=-4.83$
$x$	$y_1(x)$	$y_2(x)$	$y_3(x)$	$y_4(x)$	$y_5(x)$	$y_6(x)$	$y_7(x)$	$y_8(x)$
0	0	0	0	0	0	0	0	0
0.1	0.100	0.155	-0.441	-0.643	-0.510	-0.484	-0.474	-0.482
0.2	0.200	0.309	-0.869	-1.246	-1.000	-0.952	-0.931	-0.948
0.3	0.299	0.460	-1.227	-1.644	-1.384	-1.327	-1.302	-1.323
0.4	0.395	0.601	-1.595	-1.592	-1.496	-1.463	-1.448	-1.471
0.5	0.485	0.720	-1.294	-1.161	-1.281	-1.292	-1.294	-1.316
0.6	0.563	0.801	-0.988	-0.591	-0.872	-0.919	-0.940	-0.987
0.7	0.623	0.831	-0.591	0.015	-0.398	-0.473	-0.506	-0.470
0.8	0.659	0.804	-0.170	0.629	0.089	-0.009	-0.054	0.012
0.9	0.667	0.726	0.256	1.218	0.581	-0.459	0.403	0.497
1.0	0.646	0.160	0.685	1.616	1.052	1.023	0.868	0.939

选择精确满足条件(43)的  $y(a)$ ,  $y'(a)$ , 然后用解初值问题的方法来计算  $y(b)$  和  $y'(b)$ , 看它们是否满足(44). 如不满足, 则重选二个满足条件(43)的  $y(a)$  和  $y'(a)$ , 然后再重复上述的过程, 直至满足(44)为止. 事实上, 在选择  $y(a)$  和  $y'(a)$  时一个是任意的, 另一个则应由(43)来定. 例如在  $p_1 \neq 0$  时可选  $y'(a) = m$ , 则  $y(a) = (\alpha_1 - q_1 m) / p_1$ . 于是, 终值  $y(b)$  和  $y'(b)$  显然都是  $m$  的函数, 因而  $p_2 y(b) + q_2 y'(b)$  也是  $m$  的函数, 即

$$p_2 y(b) + q_2 y'(b) = G(m).$$

所以问题象第一边值问题那样变为找  $m$  使

$$G(m) = \alpha_2.$$

由此可见, 前面所介绍的在解第一边值问题时所用的插值法和步骤均可应用. 试射法可以图示如下:

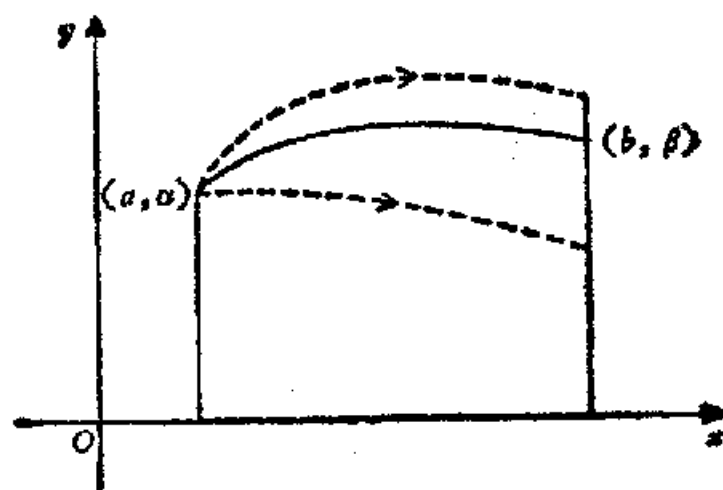


图 8.1

从理论上来说, 这个方法对于高阶的微分方程或方程组也是适用的, 然而在实际使用中可能遇到不只一个为任意的未知初值要作选择, 这时尝试的工作量将会大大地增加. 因此, 试射法通常仅用在只有一个初值是未知的情况.

## §4 适合于非线性方程的差分方法

### 4.1 非线性差分方程的迭代解法

今考虑方程

$$y'' = f(x, y, y') \quad (45)$$

的第一边值问题, 边界条件为

$$y(a) = \alpha, \quad y(b) = \beta. \quad (46)$$

取分点

$$x_i = a + ih, \quad h = \frac{b-a}{n}, \quad i = 0, 1, \dots, n.$$

在内点  $x_i$   $i = 1, \dots, n-1$  处可将方程近似地换成

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} = f\left(x_i, y_i, \frac{y_{i+1} - y_{i-1}}{2h}\right),$$
$$i = 1, 2, \dots, n-1, \quad (47)$$

边值条件为

$$y_0 = \alpha, \quad y_n = \beta, \quad (48)$$

(47) 与 (48) 组成了一个关于  $y_0, y_1, \dots, y_n$  的方程组, 但由于它是非线性的, 所以不能直接求解. 但可通过一个迭代过程求其近似解. 即先将近似值  $y_i^{(0)}$  ( $i = 0, 1, \dots, n$ ) 代入 (47) 式的右端使它为已知量, 这时 (47) 式便是关于  $y_i^{(0)}$  ( $i = 0, 1, \dots, n$ ) 的线性代数方程组, 它可以按通常的方法直接求解. 设求得的解为  $y_i^{(1)}$  ( $i = 0, 1, \dots, n$ ) 然后再通过迭代过程来求 (47) 式的更好的近似值. 其具体做法如下: 如果已选定了它的一个适当的近似解  $y_i^{(0)}$  ( $i = 0, 1, \dots, n$ ), 用它作为初始值, 然后进行如下的迭代:

$$\frac{y_{i+1}^{(r+1)} - 2y_i^{(r+1)} + y_{i-1}^{(r+1)}}{h^2} = f\left(x_i, y_i, \frac{y_{i+1}^{(r)} - y_{i-1}^{(r)}}{2h}\right),$$
$$i = 1, 2, \dots, n-1,$$

$$\begin{aligned} y_0^{(r+1)} &= \alpha, \\ y_n^{(r+1)} &= \beta. \end{aligned} \quad (49)$$

如果  $y_i^{(r)}$  为已知, 则方程组 (49) 是一个三对角型的线性方程组, 可以采用上节中介绍的适合于线性代数方程的追赶法去求解, 这种迭代方法实质上是 Jacobi 迭代. 可以证明, 在一定的条件下, 这样的迭代过程是收敛的, 而且还可以证明方程组 (47) (48) 的解当  $h \rightarrow 0$  时也是收敛到边值问题 (45) (46) 的解的 (参看 И. С. Березин, Н. П. Жидков, 计算方法第九章 (俄文), 莫斯科, 1959)

## 4.2. Newton-Raphson 迭代

若所考虑的边值问题为

$$y' = f(x, y), \quad (50)$$

$$y(a) = \alpha, \quad y(b) = \beta, \quad (51)$$

其中  $-\infty < a < b < \infty$ ,  $\alpha$  及  $\beta$  为任意常数, 其右端函数  $f(x, y)$  除了满足

- 1) 在  $a \leq x \leq b$ ,  $-\infty < y < \infty$  中为  $x, y$  的连续函数,
- 2) 对于  $x \in [a, b]$ , 任意二个  $y, y^*$ , 存在着一个常数  $L$ , 使得

$$|f(x, y) - f(x, y^*)| \leq L|y - y^*|$$

成立, 此外, 还满足

- 3)  $f_y(x, y)$  连续且  $f_y(x, y) \geq 0$ , 对于  $a \leq x \leq b$ ,  $-\infty < y < \infty$  成立, 则可用 Newton-Raphson 迭代方法求得其数值解.

为此我们必须先建立数值解所满足的差分方程, 同样我们将区间  $[a, b]$  分成  $n$  等份, 其分点为  $x_i = a + ih (i=0, 1, \dots, n)$ , 其中  $h = \frac{b-a}{n}$ ,  $n$  为一适当的整数, 为求得在  $x_i$  处 (50) 的精确解  $y(x_i)$  的近似值  $y_i$ , 我们在每一个内点  $x_i (i=1, \dots,$

$n-1$ )上建立一个差分方程

$$-y_{i-1} + 2y_i - y_{i+1} + h^2\{\beta_0 f_{i-1} + \beta_1 f_i + \beta_2 f_{i+1}\} = 0, \quad (52)$$

$$i=1, \dots, n-1.$$

欲使 (52) 所定义的方法其阶数  $p > 0$ , 我们假设  $\beta_0 + \beta_1 + \beta_2 = 1$ .

常用的公式为

$$-y_{i-1} + 2y_i - y_{i+1} + h^2 f_i = 0, \quad p=2 \quad (53)$$

和

$$-y_{i-1} + 2y_i - y_{i+1} + \frac{1}{12} h^2 (f_{i-1} + 10f_i + f_{i+1}) = 0, \quad p=4 \quad (54)$$

为简便起见, 引用矩阵和向量的记号

$$\mathbf{y} = (y_1, y_2, \dots, y_{n-1})^T,$$

$$\mathbf{f}(\mathbf{y}) = (f(x_1, y_1), f(x_2, y_2), \dots, f(x_{n-1}, y_{n-1}))^T,$$

$$\mathbf{W} = (\alpha - \beta_0 h^2 f(x_0, \alpha), 0, \dots, 0, \beta - \beta_2 h^2 f(x_n, \beta))^T,$$

及

$$J = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix},$$

$$B = \begin{bmatrix} \beta_1 & \beta_2 & & & \\ \beta_0 & \beta_1 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_0 & \beta_1 & \beta_2 \\ & & & \beta_0 & \beta_1 \end{bmatrix}.$$



于是(52)式便可写成

$$Jy + h^2 Bf(y) - W = 0. \quad (55)$$

(55)为  $y$  的非线性方程.

现在我们介绍比 4.1 节中所采用的迭代方法更好的 Newton-Raphson 迭代方法. 它的主要思想可通过对单个方程  $f(x) = 0$  的求解来说明, 在一个初始的近似解  $x^{(0)}$  的附近将  $f(x)$  线性化. 以求解  $f(x^{(0)}) + f'(x^{(0)})\Delta x = 0$  来代替求解  $f(x) = 0$ , 即  $\Delta x = -\frac{f(x^{(0)})}{f'(x^{(0)})}$ , 由此而产生的  $x^{(1)} = x^{(0)} + \Delta x$

可望为精确解的一个更好的近似值. 将这一思想用在(55)上, 便可得到解边值问题的 Newton-Raphson 迭代方法.

为此记

$$r(y) = Jy + h^2 Bf(y) - W, \quad (56)$$

它表示余向量.  $r(y)$  在  $y^{(0)}$  处的导数为  $J + h^2 BF(y^{(0)})$ , 其中

$$F(y) = \begin{bmatrix} f_1(x_1, y_1) & & & \\ & f_2(x_2, y_2) & & \\ & & \ddots & \\ & & & f_n(x_{n-1}, y_{n-1}) \end{bmatrix},$$

(56) 经过线性化后化为

$$r(y^{(0)}) + (J + h^2 BF(y^{(0)}))\Delta y = 0.$$

令  $A(y) = J + h^2 BF(y)$ . 如果  $A(y^{(0)})^{-1}$  存在, 便可解出

$$\Delta y = \Delta y^{(0)} = -A(y^{(0)})^{-1}r(y^{(0)}). \quad (57)$$

此时  $y^{(1)} = y^{(0)} + \Delta y^{(0)}$  为精确解的一个较好的近似值, 此时所获的余向量  $r(y^{(1)})$  可望更小一些, 以  $y^{(1)}$  代替  $y^{(0)}$  重复上述过程直到收敛为止.

对于边值问题(50)(51), 当  $h$  充分小时, (52)式存在唯一的解  $y$ , 而 Newton-Raphson 迭代方法所提供的向量序列

$y^{(v)}$ ,  $v=1, 2, \dots$ , 在适当选择初始逼近值  $y^{(0)}$  的情况下能很快地收敛于  $y$ . 这里不准备作详细介绍<sup>1)</sup>. 值得注意的是在具体使用上述 Newton-Raphson 迭代方法时, 由于

$$A(y^{(0)}) \Delta y = -r(y^{(0)}) \quad (58)$$

为  $\Delta y$  的分量的线性代数方程, 而  $A(y)$  又是三对角的, 所以在求  $\Delta y$  时并不需要解(57), 也就是说需求  $A(y^{(0)})^{-1}$ , 直接解(58)就行了. 事实上, 若设  $A(y^{(0)}) = (a_{ij})$ , 则有

$$a_{i, i-1} = -1 + h^2 \beta_0 f_y(x_{i-1}, y_{i-1}^{(0)}), \quad i=2, \dots, n-1,$$

$$a_{i, i} = 2 + h^2 \beta_1 f_y(x_i, y_i^{(0)}), \quad i=1, \dots, n-1,$$

$$a_{i, i+1} = -1 + h^2 \beta_2 f_y(x_{i+1}, y_{i+1}^{(0)}), \quad i=1, \dots, n-2.$$

其他各元素均为零, 所以每前进一步只要计算余向量  $r(y)$  和偏导数  $f_y(x_i, y_i)$  ( $i=1, 2, \dots, n-1$ ), 当余向量  $r(y^{(v)})$  很小时, 或是当  $\Delta y^{(v)}$  在误差允许范围内便可停止计算.

例. 对于边值问题

$$y'' = -2 + \sin hy, \quad y(0) = 0, \quad y(1) = 0, \quad (59)$$

用差分方程

$$-y_{i-1} + 2y_i - y_{i+1} + h^2 f_i = 0, \quad (60)$$

以  $h=0.1$  求其数值解.

此时共有 9 个未知量  $y_1, y_2, \dots, y_9$ , 但由于微分方程的精确解满足  $y(x) = y(1-x)$ , 故可假设  $y_n = y_{10-n}$ , 所以只需计算  $y_1, \dots, y_5$ . 利用  $y_6 = y_4$  (60) 便化成

$$2y_1 - y_2 + 0.01(-2 + \sin hy_1) = 0,$$

$$-y_1 + 2y_2 - y_3 + 0.01(-2 + \sin hy_2) = 0,$$

$$-y_2 + 2y_3 - y_4 + 0.01(-2 + \sin hy_3) = 0,$$

$$-y_3 + 2y_4 - y_5 + 0.01(-2 + \sin hy_4) = 0,$$

$$-2y_4 + 2y_5 + 0.01(-2 + \sin hy_5) = 0.$$

1) 见 P. Henrici, Discrete Variable Methods in Ordinary Differential Equations, § 7.2—4, 1962.

我们将(59)的非线性项  $\sin hy$  略去, 获得的近似解

$$y(x) = x(1-x),$$

以它作为零次近似值, 此时得余向量为

$$\mathbf{r}(\mathbf{y}^{(0)}) = (0.00090122, 0.00160684, 0.00211547, \\ 0.00242311, -0.00252612)$$

矩阵

$$A(\mathbf{y}^{(0)}) = \begin{bmatrix} 2.01004053 & -1 & & & \\ & -1 & 2.01012827 & -1 & \\ & & -1 & 2.01022131 & -1 \\ & & & -1 & 2.01028939 & -1 \\ & & & & -1 & 2.01031413 \end{bmatrix}.$$

$A$  中未标出的元素为零, 所得到的校正量  $\Delta \mathbf{y}^{(0)}$  为

$$\Delta \mathbf{y}^{(0)} = (-0.0075325, -0.0142394, -0.0194837, \\ -0.0228116, -0.0239511)$$

由新的近似值  $\mathbf{y}^{(1)} = \mathbf{y}^{(0)} + \Delta \mathbf{y}^{(0)}$  所产生的余向量

$$\mathbf{r}(\mathbf{y}^{(1)}) = (0.0000000, 0.0000002, 0.0000003, \\ 0.0000006, 0.0000008).$$

再一次应用上述方法所得的向量为

$$\mathbf{y}^{(2)} = (0.0824662, 0.1457580, 0.1905125, \\ 0.2171837, 0.2260438).$$

此时  $\mathbf{r}(\mathbf{y}^{(2)}) = 0$ .

类似此例, 在  $A(\mathbf{y})$  对  $\mathbf{y}$  的依赖性并不太灵敏的问题中, 只需在  $\mathbf{y} = \mathbf{y}^{(0)}$  处对  $A(\mathbf{y})$  进行计算, 以后的近似值均可以由

$$A(\mathbf{y}^{(0)}) \Delta \mathbf{y}^{(v)} = -\mathbf{r}(\mathbf{y}^{(v)}), \quad v = 1, 2, \dots$$

求得. 对于适当选择的初始值  $\mathbf{y}^{(0)}$ , 所得  $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots$ , 还是收敛于差分方程的精确解  $\mathbf{y}$ , 但收敛速度要较在每一次迭代均使用重新计算的  $A(\mathbf{y}^{(v)})$  值要慢.

## 附录 I 差分方程简介

### § 1 一般差分方程

表达式

$$F(t; y(t), \Delta y(t), \dots, \Delta^k y(t)) = 0 \quad (1)$$

称为未知函数  $y(t)$  的差分方程, 其中  $y(t)$  定义在  $\{t=0, 1, 2, \dots\}$  上,  $F$  是已知函数,  $\Delta y(t) = y(t+1) - y(t)$ ,

$$\Delta^2 y(t) = \Delta y(t+1) - \Delta y(t) = y(t+2) - 2y(t+1) + y(t),$$

.....

$$\Delta^k y(t) = \Delta^{k-1} y(t+1) - \Delta^{k-1} y(t)$$

$$= \sum_{m=0}^k (-1)^m \binom{k}{m} y(t+k-m).$$

方程(1)中各阶差分若都以函数值表示, 则方程(1)可写为

$$Q(t; y(t), y(t+1), \dots, y(t+k)) = 0. \quad (2)$$

若  $y(t)$  及  $y(t+k)$  均含于方程(2)中, 则称方程(1)或方程(2)是  $k$  阶差分方程, 若  $y(t)$  或  $y(t+k)$  至少有一个在(2)中不出现, 则方程(1)或(2)是低于  $k$  阶的.

### § 2 线性差分方程

方程

$$a_k(n)y_{n+k} + a_{k-1}(n)y_{n+k-1} + \dots + a_0(n)y_n = b(n), \quad (3)$$

其中  $a_i(n)$ ,  $b(n)$  是正整数变量  $n$  的已知函数, 若  $a_k(n) \neq 0$ ,

$a_0(n) \neq 0$ , 则称(3)为未知函数  $y(n) = y_n$  的  $k$  阶线性差分方程. 若已给定  $k$  个初始值  $y_0, y_1, \dots, y_{k-1}$ , 便可由方程(3)逐个地求出  $y_k, y_{k+1}, \dots$ , 若  $b(n) = 0$ , 则称相应的方程

$$a_k(n)y_{n+k} + a_{k-1}(n)y_{n+k-1} + \dots + a_0(n)y_n = 0 \quad (4)$$

为齐次的; 否则, 称为非齐次的.

齐次方程与非齐次方程的解之间有下列重要性质 (证明从略):

1. 若  $y_n^{(1)}, y_n^{(2)}, \dots, y_n^{(k)}$  是方程(4)的特解, 则它们的任意线性组合

$$z_n = C_1 y_n^{(1)} + C_2 y_n^{(2)} + \dots + C_k y_n^{(k)}$$

也是方程(4)的解.

2. 若 Wronski 行列式

$$\begin{vmatrix} y_0^{(1)} & y_1^{(1)} & \dots & y_{k-1}^{(1)} \\ y_0^{(2)} & y_1^{(2)} & \dots & y_{k-1}^{(2)} \\ \dots & \dots & \dots & \dots \\ y_0^{(k)} & y_1^{(k)} & \dots & y_{k-1}^{(k)} \end{vmatrix} \neq 0, \quad (5)$$

则  $y_n^{(1)}, y_n^{(2)}, \dots, y_n^{(k)}$  为线性无关.

若  $y_n^{(1)}, y_n^{(2)}, \dots, y_n^{(k)}$  是方程(4)的解且满足(5), 则称  $y_n^{(1)}, y_n^{(2)}, \dots, y_n^{(k)}$  是方程(4)的基本解组. (4)的任何一个特解都可以表示为

$$z_n = C_1 y_n^{(1)} + C_2 y_n^{(2)} + \dots + C_k y_n^{(k)}, \quad (6)$$

亦即(6)是(4)的通解.

3. 非齐次方程(3)的通解可以用它的一个特解和齐次方程(4)的通解之和来表示.

4. 设  $y_n^{(1)}, y_n^{(2)}, \dots, y_n^{(k)}$  是方程(4)的特解且满足条件(5), 则

$$z_n = \sum_{i=0}^{n-1} \begin{vmatrix} y_{i+1}^{(1)} & y_{i+1}^{(2)} & \cdots & y_{i+1}^{(k)} \\ y_{i+2}^{(1)} & y_{i+2}^{(2)} & \cdots & y_{i+2}^{(k)} \\ \cdots & \cdots & \cdots & \cdots \\ y_{i+k-1}^{(1)} & y_{i+k-1}^{(2)} & \cdots & y_{i+k-1}^{(k)} \\ y_n^{(1)} & y_n^{(2)} & \cdots & y_n^{(k)} \\ y_{i+1}^{(1)} & y_{i+1}^{(2)} & \cdots & y_{i+1}^{(k)} \\ y_{i+2}^{(1)} & y_{i+2}^{(2)} & \cdots & y_{i+2}^{(k)} \\ \cdots & \cdots & \cdots & \cdots \\ y_{i+k}^{(1)} & y_{i+k}^{(2)} & \cdots & y_{i+k}^{(k)} \end{vmatrix} b(i)$$

是非齐次方程(3)的特解(与常数变易法相类似)。

### §3 线性常系数差分方程

若方程(3)或(4)的系数  $a_i(n)$ ,  $b(n)$  与  $n$  无关, 则称此方程是线性常系数差分方程。这时可以把(3)和(4)写成

$$a_k y_{n+k} + a_{k-1} y_{n+k-1} + \cdots + a_0 y_n = b, \quad a_k \neq 0, a_0 \neq 0, \quad (7)$$

$$a_k y_{n+k} + a_{k-1} y_{n+k-1} + \cdots + a_0 y_n = 0. \quad (8)$$

设常系数齐次差分方程(8)的解具有以下形式:

$$y_n = z^n, \quad z \neq 0.$$

将它代入(8), 便得到关于  $z$  的代数方程

$$a_k z^k + a_{k-1} z^{k-1} + \cdots + a_1 z + a_0 = 0. \quad (9)$$

方程(9)称为方程(8)的特征方程, 方程(9)的根称为特征根。

若方程(9)的  $k$  个根  $z_1, z_2, \cdots, z_k$  互异, 则  $z_1^n, z_2^n, \cdots, z_k^n$  是方程(8)的线性无关解组。(8)的通解  $u$  为  $z_1^n, z_2^n, \cdots, z_k^n$  的线性组合

$$u = C_1 z_1^n + C_2 z_2^n + \cdots + C_k z_k^n, \quad (10)$$

其中  $C_1, \cdots, C_k$  为可由初始条件加以确定的任意常数。若其中有一个特征根为  $m$  重, 设其为  $z_1$ , 则齐次方程的解可以表

示成

$$u = (C_1 + C_2 n + \cdots + C_{m-1} n^{m-1}) z_1^n + C_m z_m^n + \cdots + C_k z_k^n. \quad (11)$$

若  $a_0, a_1, \dots, a_{k-1}$  都为实数, 如有复根则必成对出现, 对于这样的复根  $z_1 = v_R + i v_I$  及  $z_2 = v_R - i v_I$ , 我们令

$$\begin{aligned} r^2 &= v_R^2 + v_I^2, \\ \theta &= \tan(v_I/v_R). \end{aligned}$$

此时(8)的通解为

$$u = C_1 r^n \cos n\theta + C_2 r^n \sin n\theta + C_3 z_3^n + \cdots + C_k z_k^n \quad (12)$$

若特征根互异且不等于1, 则非齐次方程(7)的解为

$$v = b / (a_k + a_{k-1} + \cdots + a_0) \quad (13)$$

上述的概念可以推广到线性常系数差分方程组

$$A_k \mathbf{y}_{n+k} + A_{k-1} \mathbf{y}_{n+k-1} + \cdots + A_0 \mathbf{y}_n = \mathbf{b} \quad (14)$$

其中  $\mathbf{y}_j$  表向量,  $A_j$  表  $m \times m$  矩阵, 方程(14)利用记号  $E^k \mathbf{y}_n = \mathbf{y}_{n+k}$  又可以写成  $G(E) \mathbf{y}_n = \mathbf{b}$ , 其中  $G$  为一个  $m \times m$  矩阵. (14)式的特征方程为

$$\det G(\mu) = 0. \quad (15)$$

由于  $G$  是一个  $m \times m$  矩阵, 且在  $G(\mu)$  中  $\mu$  的幂次最高为  $k$ , 所以(15)的特征根有  $km$  个, 若这些根都是互异的, 则(14)的通解可表为

$$\begin{aligned} y_{in} &= \sum_{j=1}^m \sum_{l=1}^k C_{ij}(\mu_{jl})^n + \{[A_k + A_{k-1} + \cdots + A_0]^{-1} \mathbf{b}\}_i, \quad (16) \\ i &= 1, 2, \dots, m. \end{aligned}$$

## 附录 II 第六章 定理 2 的证明

设  $k$  步公式的一般形式为

$$\begin{aligned} & \alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n \\ & = h(\beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \cdots + \beta_0 f_n), \end{aligned} \quad (1)$$

与其相关连的二个多项式为

$$\begin{aligned} \rho(\zeta) &= \alpha_k \zeta^k + \alpha_{k-1} \zeta^{k-1} + \cdots + \alpha_0, \\ \sigma(\zeta) &= \beta_k \zeta^k + \beta_{k-1} \zeta^{k-1} + \cdots + \beta_0. \end{aligned}$$

$\rho(\zeta)$  称为特征多项式.

**定理 2**  $k$  步法(1)收敛的充分必要条件为 1) 相容性条件  $\rho(1)=0$ ,  $\rho'(1)=\sigma(1)$ , 2) 特征多项式  $\rho(\zeta)$  的零点的模不能大于 1, 并且在单位圆上的零点只能是单零点.

条件 2) 又称为根条件.

**证明** 先证在收敛的假设下, 根条件成立. 设方法(1)是收敛的, 由收敛的定义. 知其对初值问题  $y'=0$ ,  $y(0)=0$  也成立, 显然该初值问题的精确解为  $y(x)=0$ , 在这种情况下多步法(1)简化成

$$\alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n = 0. \quad (2)$$

按收敛的定义, 对于任何  $x>0$  和(1)的所有满足条件

$$\lim_{h \rightarrow 0} \eta_\mu(h) = 0, \quad \mu = 0, 1, \dots, k-1 \quad (3)$$

的解  $\{y_n\}$  均有

$$\lim_{n \rightarrow \infty} y_n = 0, \quad h = x/n, \quad (4)$$

其中  $y_\mu = \eta_\mu(h)$ . 令  $\zeta = re^{i\varphi}$  ( $r \geq 0$ ,  $0 \leq \varphi < 2\pi$ ) 为  $\rho(\zeta)$  的一个



根. 易知<sup>1)</sup>

$$y_n = hr^n \cos n\varphi \quad (5)$$

定义了(2)的一个解, 它亦满足(3)式. 若(1)所定义的方法为收敛的, 则(4)必须成立. 当  $\varphi=0$  或  $\varphi=\pi$  时, 立得  $r \leq 1$  必须成立的结论. 若  $\varphi \neq 0$ ,  $\varphi \neq \pi$ , 则有下列的关系

$$\frac{y_n^2 - y_{n+1}y_{n-1}}{\sin^2 \varphi} = h^2 r^{2n} \quad (6)$$

成立. 由于当  $n \rightarrow \infty$  ( $h=x/n$ ) 时(6)的左端趋于零, 所以右端也必如此, 再一次得到  $r \leq 1$ . 现进一步证明  $\rho(\zeta)$  在  $|\zeta|=1$  上的零点的次数不得超过1. 设  $\zeta = re^{i\varphi}$  为  $\rho(\zeta)$  的重零点, 其次数超过1, 同样可知<sup>1)</sup>此时

$$y_n = h^{1/2} n r^n \cos n\varphi \quad (7)$$

也是(2)的解, 对于  $|\eta_\mu(h)| = |y_\mu| \leq h^{1/2} \mu r^n$  ( $\mu=0, 1, \dots, k-1$ ), 它们是满足(3)的. 由于方法是收敛的, 所以(4)必需成立, 当  $\varphi=0$  或  $\varphi=\pi$  时, 由于  $h=x/n$ , 故  $|y_n| = x^{1/2} n^{1/2} r^n$ , 立得  $r < 1$ , 又若  $\varphi \neq 0$ ,  $\varphi \neq \pi$ , 我们从关系式

$$\frac{z_n^2 - z_{n+1}z_{n-1}}{\sin^2 \varphi} = r^{2n}, \quad (8)$$

其中  $z_n = n^{-1} h^{-1/2} y_n$ , 由于当  $n \rightarrow \infty$  时  $z_n \rightarrow 0$ , 故(8)的左端当  $n \rightarrow \infty$  时趋于零. 再一次得到  $r < 1$  的结论, 综上所述当(1)定义的方法为收敛时,  $\rho(\zeta)$  必满足根条件, 所以方法是稳定的.

现进一步证明若由(1)所定义的多步法为收敛的, 则它也一定是相容的. 由相容性的定义知相容性与由(1)所定义的差分算子的阶数  $p \geq 1$  必需成立为等价的, 亦即  $C_0 = C_1 = 0$ . 它可以用多项式  $\rho(\zeta)$  与  $\sigma(\zeta)$  表示如下:

$$\rho(1) = 0, \quad \rho'(1) = \sigma(1), \quad (9)$$

1) P. Henrici, Discrete Variable Methods in Ordinary Differential Equations, 1962.

亦即定理中的条件 1). 由于方法(1)为收敛的, 所以它对具有精确解  $y(x)=1$  的初值问题  $y'=0, y(0)=1$  也是收敛的, 此时相应的多步法(1)便简化为

$$\alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n = 0. \quad (10)$$

因为(10)的取精确初始出发值  $y_\mu = 1$  ( $\mu=0, 1, \dots, k-1$ ) 的解  $\{y_n\}$  必须满足当  $h \rightarrow 0, nh=x$  时  $y_n \rightarrow 1$  的条件, 又因为在这种情况下  $y_n$  并不依赖于  $h$ , 所以这个条件又可叙述为当  $n \rightarrow \infty$  时  $y_n = 1$ . 在(10)中令  $n \rightarrow \infty$  时  $y_n \rightarrow 1$ , 便得到了

$$\alpha_k + \alpha_{k-1} + \cdots + \alpha_0 = 0.$$

于是便有  $C_0 = \rho(1) = 0$  亦即方法的阶数  $p \geq 0$ .

为了证明  $C_1 = 0$ , 亦即  $\rho'(1) = \sigma(1)$  我们考虑具有精确解  $y(x)=x$  的初值问题  $y'=1, y(0)=0$ . 此时相应的差分方程(1)变为

$$\begin{aligned} & \alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n \\ & = h(\beta_k + \beta_{k-1} + \cdots + \beta_0). \end{aligned} \quad (11)$$

对于(11)的每一个取初始出发值  $y_\mu = \eta_\mu(h)$ , 满足

$$\lim_{h \rightarrow 0} \eta_\mu(h) = 0, \quad \mu = 0, 1, \dots, k-1 \quad (12)$$

的解  $\{y_n\}$ , 则由于方法为收敛的所以必需有

$$\lim_{h \rightarrow 0} y_n = x, \quad x_n = x$$

成立. 又由收敛的方法必是稳定的故知

$$k\alpha_k + (k-1)\alpha_{k-1} + \cdots + \alpha_1 = \rho'(1) \neq 0.$$

现定义  $\{y_n\}$  为  $nhK$ , 其中

$$K = \frac{\beta_k + \beta_{k-1} + \cdots + \beta_0}{k\alpha_k + (k-1)\alpha_{k-1} + \cdots + \alpha_1}. \quad (13)$$

这个数列显然满足(12), 且易证它也是(11)的解, 所以

$$\lim_{h \rightarrow 0} nhK = x$$

必需成立, 又因  $nh=x$ , 所以  $K$  必须为 1, 即

$$\frac{\beta_k + \beta_{k-1} + \cdots + \beta_0}{k\alpha_k + (k-1)\alpha_{k-1} + \cdots + \alpha_1} = 1,$$

亦即  $\rho'(1) = \sigma(1)$ , 或可写成  $C_1 = 0$ , 此时方法的阶数  $p \geq 1$ . 所以方法(1)为相容的, 必要性证明完毕.

接下来我们要证明在条件 1) 和 2) 成立的前题下由(1)定义的方法为收敛的. 亦即相容性与根条件成立为收敛的充分条件, 为此我们需要下面的二个引理.

引理 1 令多项式  $\rho(\zeta) = \alpha_k \zeta^k + \alpha_{k-1} \zeta^{k-1} + \cdots + \alpha_0$  满足根条件且令系数  $\gamma_l (l=0, 1, 2, \cdots)$  为由

$$\frac{1}{\alpha_k + \alpha_{k-1}\zeta + \cdots + \alpha_0 \zeta^k} = \gamma_0 + \gamma_1 \zeta + \gamma_2 \zeta^2 + \cdots \quad (14)$$

所定义的, 则

$$\Gamma = \sup_{l=0, 1, \dots} |\gamma_l| < \infty. \quad (15)$$

证明 假若  $\hat{\rho}(\zeta) = \alpha_k + \alpha_{k-1}\zeta + \cdots + \alpha_0 \zeta^k = \zeta^k \rho(\zeta^{-1})$ , 则  $\hat{\rho}(\zeta)$  的零点为  $\rho(\zeta)$  的零点的倒数. 由于  $\rho(\zeta)$  的零点满足根条件, 所以  $\rho(\zeta)$  在  $|\zeta|=1$  外无零点. 故  $\frac{1}{\hat{\rho}(\zeta)}$  在  $|\zeta| < 1$  中为全纯的. 又由于  $\rho(\zeta)$  在  $|\zeta|=1$  上的零点  $\zeta_1, \zeta_2, \dots, \zeta_m$  为单的, 所以  $\hat{\rho}(\zeta)^{-1}$  在  $|\zeta|=1$  上的极点也是单的. 因此存在着常数  $A_1, A_2, \dots, A_m$  使得

$$f(\zeta) = \frac{1}{\hat{\rho}(\zeta)} = \frac{A_1}{\zeta - \zeta_1^{-1}} - \cdots - \frac{A_m}{\zeta - \zeta_m^{-1}} \quad (16)$$

在  $|\zeta| \leq 1$  上为全纯的, 于是用 Cauchy 的估计<sup>1)</sup> 知  $f(\zeta)$  在  $\zeta=0$  处 Taylor 展式的系数为有界的. 又由于(16)右端的每一项  $A_\mu / (\zeta - \zeta_\mu^{-1})$  在  $\zeta=0$  处的 Taylor 展式的系数为有界的, 于是得

$$\begin{aligned} \frac{1}{\hat{\rho}(\zeta)} &= f(\zeta) + \frac{A_1}{\zeta - \zeta_1^{-1}} + \cdots + \frac{A_m}{\zeta - \zeta_m^{-1}} \\ &= \gamma_0 + \gamma_1 \zeta + \gamma_2 \zeta^2 + \cdots \end{aligned}$$

1) Ahlfors, L. V., Complex Analysis, 1953.

的系数为有界的. 引理证毕.

引理 2 考虑非齐次线性差分方程

$$\begin{aligned} & \alpha_k z_{m+k} + \alpha_{k-1} z_{m+k-1} + \cdots + \alpha_0 z_m \\ & = h \{ \beta_{k,m} z_{m+k} + \beta_{k-1,m} z_{m+k-1} + \cdots + \beta_{0,m} z_m \} + \lambda_m. \end{aligned} \quad (17)$$

$\rho(\zeta) = \alpha_k \zeta^k + \alpha_{k-1} \zeta^{k-1} + \cdots + \alpha_0$ , 它的零点满足根条件. 令  $B^*$ ,  $\beta$  及  $\Lambda$  为非负常数, 满足下面的不等式

$$\begin{aligned} & |\beta_{k,n}| + |\beta_{k-1,n}| + \cdots + |\beta_{0,n}| \leq B^* \\ & |\beta_{k,n}| \leq \beta, \quad |\lambda_n| \leq \Lambda, \\ & n = 0, 1, 2, \dots, N, \end{aligned} \quad (18)$$

且令  $0 \leq h \leq |\alpha_k| \beta^{-1}$ . 这时, (17) 的每一个满足

$$|z_\mu| \leq Z, \quad \mu = 0, 1, \dots, k-1 \quad (19)$$

的解满足关系式

$$|z_n| \leq K^* e^{nhL^*}, \quad n = 0, 1, \dots, N, \quad (20)$$

其中

$$L^* = \Gamma^* B^*, \quad K^* = \Gamma^* (N\Lambda + AZk),$$

$$A = |\alpha_k| + |\alpha_{k-1}| + \cdots + |\alpha_0|, \quad \Gamma^* = \frac{\Gamma}{1 - h|\alpha_k|^{-1}\beta}. \quad (21)$$

证明: 在(17)中令  $m = n - k - l$ , 并以  $\gamma_l$  乘之, 再令  $l = 0, 1, \dots, n - k$ , 并把所得的  $n - k + 1$  个方程相加, 以  $S_n$  记这个和, 于是左端相加得

$$\begin{aligned} S_n &= (\alpha_k z_n + \alpha_{k-1} z_{n-1} + \cdots + \alpha_0 z_{n-k}) \gamma_0 \\ &+ (\alpha_k z_{n-1} + \alpha_{k-1} z_{n-2} + \cdots + \alpha_0 z_{n-k-1}) \gamma_1 \\ &+ \cdots + (\alpha_k z_k + \alpha_{k-1} z_{k-1} + \cdots + \alpha_0 z_0) \gamma_{n-k}. \end{aligned}$$

将上式中各项重新按  $z_l$  整理得

$$\begin{aligned} S_n &= \alpha_k \gamma_0 z_n + (\alpha_k \gamma_1 + \alpha_{k-1} \gamma_0) z_{n-1} \\ &+ \cdots + (\alpha_k \gamma_{n-k} + \alpha_{k-1} \gamma_{n-k-1} + \cdots + \alpha_0 \gamma_{n-2k}) z_k \\ &+ (\alpha_{k-1} \gamma_{n-k} + \cdots + \alpha_0 \gamma_{n-2k+1}) z_{k-1} \\ &+ \cdots + \alpha_0 \gamma_{n-k} z_0. \end{aligned}$$

由于

$$\alpha_k \gamma_l + \alpha_{k-1} \gamma_{l-1} + \cdots + \alpha_0 \gamma_{l-k} = \begin{cases} 1, & l=0, \\ 0, & l>0 \end{cases} \quad (22)$$

成立, 这里假定  $l < 0$  时  $\gamma_l = 0$ . 这个恒等式由 (14) 式二端乘以  $\alpha_k + \alpha_{k-1} \zeta + \cdots + \alpha_0 \zeta^k$  并比较  $\zeta$  的同幂次系数便可以得到. 所以  $S_n$  又简化成

$$S_n = z_n + (\alpha_{k-1} \gamma_{n-k} + \cdots + \alpha_0 \gamma_{n-2k+1}) z_{k-1} + \cdots + \alpha_0 \gamma_{n-k} z_0. \quad (23)$$

右端所得到的和

$$\begin{aligned} S_n = & h \{ \beta_{k, n-k} \gamma_0 z_n + (\beta_{k-1, n-k} \gamma_0 + \beta_{k, n-k-1} \gamma_1) z_{n-1} \\ & + \cdots + (\beta_{0, n-k} \gamma_0 + \cdots + \beta_{k, n-2k} \gamma_k) z_{n-k} \\ & + \cdots + \beta_{0, 0} \gamma_{n-k} z_0 \} + \lambda_{n-k} \gamma_0 + \lambda_{n-k-1} \gamma_1 \\ & + \cdots + \lambda_0 \gamma_{n-k}. \end{aligned} \quad (24)$$

联立 (23) 及 (24), 并利用 (18) (19) 式便得到估计式

$$|z_n| \leq h\beta |\alpha_k^{-1}| |z_n| + h\Gamma B^* \sum_{m=0}^{n-1} |z_m| + N\Gamma\Delta + A\Gamma Zk.$$

从上式中解出  $|z_n|$  便得到

$$|z_n| \leq hL^* \sum_{m=0}^{n-1} |z_m| + K^*, \quad (25)$$

其中  $L^*$  及  $K^*$  由 (21) 式给出. 由于  $A\Gamma \geq 1$ , 所以  $K^* \geq Z$ , 于是估计式

$$|z_m| \leq K^* (1 + hL^*)^m \quad (26)$$

对于  $m=0, 1, \dots, k-1$  成立. 现按数学归纳法设 (26) 式对  $m=0, 1, \dots, n-1$  成立. 将它用于 (25) 的右端, 我们得到

$$|z_n| \leq hL^* K^* \frac{(1+hL^*)^n - 1}{hL^*} + K^* = K^* (1+hL^*)^n.$$

故 (26) 式对于  $m=n$  成立, 于是 (26) 式对  $m=0, 1, \dots, N$  都成立. 进一步利用关系式  $1+hL^* \leq e^{hL^*}$ , 便得引理的结论.

利用上面的引理, 我们便能证明由 (1) 所定义的线性多步

法如果满足相容性及根条件那末它就是收敛的.

设  $f(x, y)$  在带状区域  $a \leq x \leq b, -\infty < y < \infty$  中连续, 对  $y$  来说满足 Lipschitz 条件, 设  $\eta$  为任意常数, 又设  $y(x)$  为初值问题  $y' = f(x, y), y(a) = \eta$  的解. 令  $y_n (n=0, 1, 2, \dots)$  为差分方程 (1) 取初始出发值  $y_\mu = \eta_\mu(h), \mu=0, 1, \dots, k-1$  的解, 我们定义

$$\delta = \delta(h) = \max_{\mu=0, 1, \dots, k-1} |\eta_\mu(h) - y(a + \mu h)|, \quad (27)$$

并假设

$$\lim_{h \rightarrow 0} \delta(h) = 0, \quad (28)$$

现在我们来证明对任何的  $x \in [a, b]$

$$\lim_{\substack{h \rightarrow 0 \\ x_n = x}} y_n = y(x).$$

为此我们定义  $e_m = y_m - y(x_m), m=0, 1, 2, \dots$  为了估计  $e_m$ , 我们必需先建立  $e_m$  满足的方程. 设函数  $y'(x) = f(x, y(x))$  在闭区间上连续, 对于  $\varepsilon \geq 0$ , 我们定义

$$\omega(\varepsilon) = \max_{\substack{|x^* - x| < \varepsilon \\ x, x^* \in [a, b]}} |y'(x^*) - y'(x)|.$$

对于  $\mu=0, 1, 2, \dots, k$  我们可以得到

$$y'(x_{m+\mu}) = y'(x_m) + \theta_\mu \chi(\mu h),$$

其中  $|\theta_\mu| \leq 1$ . 此外由于

$$y(x_{m+\mu}) = y(x_m) + \mu h y'(\xi_\mu),$$

其中  $x_m < \xi_\mu < x_{m+\mu}$ , 我们得到

$$y(x_{m+\mu}) = y(x_m) + \mu h [y'(x_m) + \theta'_\mu \chi(\mu h)],$$

其中  $|\theta'_\mu| \leq 1$ , 因此有

$$\begin{aligned} & L[y(x_m); h] \\ &= \alpha_k y(x_m + kh) + \alpha_{k-1} y(x_m + (k-1)h) + \dots \\ &+ \alpha_0 y(x_m) - h \{ \beta_k y'(x_m + kh) + \beta_{k-1} y'(x_m + (k-1)h) + \dots \\ &+ \beta_0 y'(x_m) \} = (\alpha_0 + \alpha_1 + \dots + \alpha_k) y(x_m) + (\alpha_1 + 2\alpha_2 + \dots \end{aligned}$$

$$+k\alpha_k)y'(x_m)h+\theta'(|\alpha_1|+2|\alpha_2|+\cdots$$

$$+k|\alpha_k|)\chi(kh)h-(\beta_0+\beta_1+\cdots$$

$$+\beta_k)y'(x_m)h-\theta(|\beta_0|+\cdots+|\beta_k|)\chi(kh)h,$$

其中  $|\theta|\leq 1$ ,  $|\theta'|\leq 1$ . 由假设算子  $L$  为相容的, 所以有

$$\alpha_0+\alpha_1+\cdots+\alpha_k=0, \alpha_1+2\alpha_2$$

$$+\cdots+k\alpha_k-\beta_0-\beta_1-\cdots-\beta_k=0.$$

因此

$$|L[y(x_m); h]| \leq K\chi(kh)h, \quad (29)$$

其中

$$K=|\alpha_1|+2|\alpha_2|+\cdots+k|\alpha_k|+|\beta_0|+\cdots+|\beta_k|.$$

现在我们从  $y_m$  所满足的方程

$$\alpha_k y_{m+k} + \cdots + \alpha_0 y_m - h\{\beta_k f_{m+k} + \cdots + \beta_0 f_m\} = 0$$

中减去  $L[y(x_m); h]$ , 并令

$$g_m = \begin{cases} [f(x_m, y_m) - f(x_m, y(x_m))]e_m^{-1}, & e_m \neq 0, \\ 0 & e_m = 0, \end{cases}$$

于是便得到  $e_m$  所满足的方程

$$\alpha_k e_{m+k} + \cdots + \alpha_0 e_m - h\{\beta_k g_{m+k} e_{m+k} + \cdots + \beta_0 g_m e_m\} \\ = \theta_m K\chi(kh)h.$$

其中  $|\theta_m|\leq 1$ . 由于  $f(x, y)$  满足 Lipschitz 条件, 所以  $|g_m|\leq L$ ,  $m=0, 1, 2, \cdots$  成立. 现在我们可以应用引理 2, 为此令  $z_m=e_m$ ,  $Z=\delta(h)$ ,  $A=K\chi(kh)h$ ,  $N=(x_n-a)/h$ , 及  $B^*=BL$ , 其中  $B=|\beta_0|+|\beta_1|+\cdots+|\beta_k|$ . 于是便得到

$$|e_n| \leq \Gamma^*[A\delta(h) + (x_n-a)K\chi(kh)]\exp[(x_n-a)L\Gamma^*B], \quad (30)$$

其中

$$A=|\alpha_0|+|\alpha_1|+\cdots+|\alpha_k|,$$

$$\Gamma^* = \frac{\Gamma}{1-h|\alpha_k^{-1}\beta_k|L}. \quad (31)$$

由于  $y'(x)$  在  $[a, b]$  上为一致连续的, 所以当  $h \rightarrow 0$  时  $\chi(kh) \rightarrow 0$ . 从 (30) 便得到当  $h \rightarrow 0$  时, 对于任何  $x_n \in [a, b]$ ,  $|e_n| \rightarrow 0$  的结论. 收敛性证明完毕.

定理 3 的证明从略, 读者可参阅有关文献<sup>1)</sup>. 由定理 2, 3 的结果, 定理 4 及其推论就是显而易见的.

---

1) G. Dahlquist, *Math. Scand.*, vol. 4, pp. 33—53, 1956; John Todd, *A Survey of Numerical Analysis*, 1962.