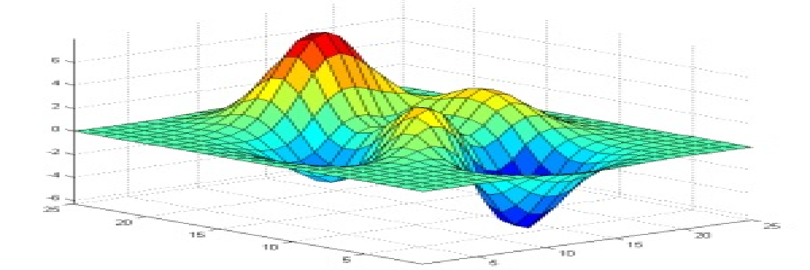
国家级精品课程

★★★★★2010★★★★★

计算方法

COMPUTATIONAL METHODS



李东方

数学与统计学院

SCHOOL OF MATHEMATICS AND STATISTICS

dfli@mail.hust.edu.cn

目　录

第一章 绪　论

科学技术发展到今天，电子计算机的应用已渗透到社会生活的各个领域。其中，数值计算是电子计算机处理实际问题的一种关键手段，从宏观天体运动学到微观分子细胞学说，从工程系统到非工程系统，无一能离开数值计算。数值计算这门学科的诞生，使科学发展产生了巨大飞跃，它使各科学领域从定性分析阶段走向定量分析阶段，从粗糙走向精密。由此可见，数值计算方法是当今每一位从事科学研究与应用的人不可缺少的知识。

本章主要介绍数值算法的预备知识及其基本思想。

§1.1 数值算法概论

一个实际问题当采用计算机来求解时，主要分下面几个步骤：

$\boxed{\text{实际问题}} \Rightarrow \boxed{\text{建立数学模型}} \Rightarrow \boxed{\text{构造数值算法}} \Rightarrow \boxed{\text{编程上机}} \Rightarrow \boxed{\text{获取近似结果}}$

由此可知，**数值算法**是利用计算机求解数学问题近似解的方法。其中，所获近似解也称为原问题的**数值解**或**逼近解**。当构造一个数值算法时，它既要面向数学模型，使算法能尽可能地仿真原问题；同时，它也要面向计算机及其程序设计，要求算法具递推性、简洁性及必要的准确性，使其能借助于计算机最终在尽可能少的时间内获得合符原问题精度要求的数值解。

例1.1 计算积分

$$I_n = \int_0^1 \frac{x^n}{x+5} dx, \quad n = 0, 1, 2, ..., 30.$$

解 通过直接计算可产生递推关系

$$I_n = -5I_{n-1} + \frac{1}{n}, \quad I_0 = \ln\frac{6}{5} \approx 1.8232e-001. \tag{1}$$

且由经典微积分知识可推得I_n具如下性质：

1) $I_n > 0$,

2) I_n单调递减,

3) $\lim\limits_{n\to\infty} I_n = 0$,

4) $\frac{1}{6n} < I_{n-1} < \frac{1}{5n} \quad (n > 1)$.

下面我们用两种算法计算I_n:

算法A：按公式(??)，自$n=1$计算到$n=30$，产生如下计算结果：

n	1	2	3	4	5
I_n	8.8392e-002	5.8039e-002	4.3139e-002	3.4306e-002	2.8468e-002
n	6	7	8	9	10
I_n	2.4325e-002	2.1233e-002	1.8837e-002	1.6926e-002	1.5368e-002
n	11	12	13	14	15
I_n	1.4071e-002	1.2977e-002	1.2040e-002	1.1229e-002	1.0522e-002
n	16	17	18	19	20
I_n	9.8903e-003	9.3719e-003	8.6960e-003	9.1515e-003	4.2426e-003
n	21	22	23	24	25
I_n	2.6406e-002	-8.6575e-002	4.7635e-001	-2.3401e+000	1.1740e+001
n	26	27	28	29	30
I_n	-5.8664e+001	2.9336e+002	-1.4667e+003	7.3338e+003	-3.6669e+004

由上表可见，该算法产生的数值解自$n=18$开始并未呈现单调递减性质且出现负值和大于1的数，这显然与I_n的固有性质相矛盾，因此本算法所得数值解不符合原问题要求。

究其原因，按公式(??)计算，其从I_{n-1}到I_n 每向前推进一步，计算值的舍入误差便增长5倍，误差由此积累传播导致最终数值解与原问题真解相悖的结果。为克服算法A的缺陷，我们改进该算法成算法B:

第1步，由性质(4)，取

$$I_{30} \approx \frac{\frac{1}{6\times 31} + \frac{1}{5\times 31}}{2} = 5.9140e - 003,$$

第2步，用递推公式

$$I_{n-1} = -\frac{I_n}{5} + \frac{1}{5n}, \tag{2}$$

自n=30计算到n=1。该算法每向后推进一步，其舍入误差便减少5倍。因此获得合符原积分模型性态的如下数值结果：

n	29	28	27	26	25
I_n	5.4839e-003	5.7998e-003	5.9829e-003	6.2108e-003	6.4501e-003
n	24	23	22	21	20
I_n	6.7100e-003	6.9913e-003	7.2974e-003	7.6314e-003	7.9975e-003
n	19	18	17	16	15
I_n	8.4005e-003	8.8462e-003	9.3419e-003	9.8963e-003	1.0521e-002
n	14	13	12	11	10
I_n	1.1229e-002	1.2040e-002	1.2977e-002	1.4071e-002	1.5368e-002
n	9	8	7	6	5
I_n	1.6926e-002	1.8837e-002	2.1233e-002	2.4325e-002	2.8468e-002
n	4	3	2	1	0
I_n	3.4306e-002	4.3139e-002	5.8039e-002	8.8392e-002	1.8232e-001

对上述例子，我们采用的是由原模型精确解的递推关系来实现计算机求解的，这种数值求解方法称为**直接法**。

在大多数情况下，我们只能获得原模型解的近似递推关系，借助于这种近似递推关系求解的方法称为离散变量法。

例1.2 考察两点边值问题

$$\begin{cases} y'' + p(x)y' + q(x)y = f(x), \quad x \in (a,b) \\ y(a) = \alpha, \quad y(b) = \beta, \end{cases} \tag{3}$$

其中$p(x)$、$q(x)$及$f(x)$是(a,b)上的给定函数，α、β为已知常数，且设问题(1.3)在[a,b]上恒有唯一解y(x)。试给出其一种离散变量法。

解 其方法构造如下：

1) 将区间[a,b]离散化，即将[a,b]N等分，所得节点为

$$x_i = a + ih \ \ (i = 0, 1, 2, ..., N; \ x_0 = a, \ x_N = b),$$

其中$h = \frac{b-a}{N}$称为方法的步长;

2) 将问题(3)离散化。由于

$$\frac{y(x_{i+1})-y(x_{i-1})}{2h}=y'(x_i)+\mathcal{O}(h^2),$$

$$\frac{y(x_{i+1})-2y(x_i)+y(x_{i-1})}{h^2}=y''(x_i)+\mathcal{O}(h^2),$$

故可略去上两式中的余项$\mathcal{O}(h^2)$，并取$y_i\approx y(x_i)$，即得

$$y'(x_i)\approx\frac{y_{i+1}-y_{i-1}}{2h},\quad \text{（一阶中心差商）} \tag{4}$$

$$y''(x_i)\approx\frac{y_{i+1}-2y_i+y_{i-1}}{h^2}\quad \text{（二阶中心差商）} \tag{5}$$

将(4)、(5)代入(3)中得差分格式

$$\begin{cases}\frac{y_{i+1}-2y_i+y_{i-1}}{h^2}+p_i\frac{y_{i+1}-y_{i-1}}{2h}+q_iy_i=f_i, \quad i=1,2,\ldots,N-1,\\ y_0=\alpha,\ \ y_N=\beta,\end{cases} \tag{6}$$

其中$p_i=p(x_i)$, $q_i=q(x_i)$及$f_i=f(x_i)$. (6)实质是含$N-1$个未知数$y_1,y_2,\cdots,y_{N-1}$的线性方程组，由此可解得数值解$y_1,y_2,\cdots,y_{N-1}$.

§1.2 预备知识

§1.2.1 范数

定义1.1 称n维实空间R^n上的一个非负函数$\|\cdot\|$为范数，若其满足

1) $\|x\|=0$当且仅当$x=0$, $x\in R^n$,

2) $\|\alpha x\|=|\alpha|\|x\|$, $\forall\alpha\in R$, $\forall x\in R^n$,

3) $\|x+y\|\le\|x\|+\|y\|$, $\forall x,y\in R^n$.

对一维实空间R而言，$\|x\|$即为绝对值$|x|$。下面我们将主要涉及l_p范数

$$\|x\|_p=(\sum_{i=1}^n|x_i|^p)^{\frac{1}{p}},\quad x=(x_1,x_2,\cdots,x_n)^T\in R^n.$$

特别，l_∞范数即为:$\|x\|_\infty=\lim_{p\to\infty}\|x\|_p=\max_{1\le i\le n}|x_i|$.

定理1.1 (范数等价定理) 若$\|\cdot\|$与$\|\cdot\|'$为R^n上的任意两种范数，则存在正常数$C_2 \ge C_1$使得

$$C_1\|x\| \le \|x\|' \le C_2\|x\|, \quad \forall x \in R^n.$$

定义1.2 设有向量序列$\{x^{(k)} \in R^n |\, x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)})^T\}$, 若

$$\lim_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \cdots, n,$$

则称序列$\{x^{(k)}\}$收敛于向量$x = (x_1, x_2, \cdots, x_n)^T$.

定理1.2 在空间R^n中，序列$\{x^{(k)}\}$收敛于向量x 的充要条件是存在范数$\|\cdot\|$，使得$\lim_{k\to\infty}\|x^{(k)} - x\| = 0$.

证 一方面，若序列$\{x^{(k)}\}$收敛于向量x，则$\lim_{k\to\infty}\|x^{(k)} - x\|_\infty = 0$; 另一方面，若存在范数$\|\cdot\|$ 使得$\lim_{k\to\infty}\|x^{(k)} - x\| = 0$, 则由定理1.1，存在常数$C_2 \ge C_1 > 0$，使得

$$C_1\|x^{(k)} - x\| \le \|x^{(k)} - x\|_\infty \le C_2\|x^{(k)} - x\|.$$

因此, 由夹逼定理知$\lim\limits_{k\to\infty}\|x^{(k)} - x\|_\infty = 0$, 故$\{x^{(k)}\}$收敛于$x$.

定义1.3 设A为n级方阵，$\|\cdot\|$为R^n中的某范数，则称

$$\max_{\|x\|=1}\|Ax\| \quad (x\in R^n)$$

为矩阵A的从属于该向量范数的范数，记为$\|A\|$.

利用定义1.3可直接推得其矩阵范数具有如下性质:

1) 对任意n级方阵A有$\|A\|\geq 0$; 且$\|A\|=0$当且仅当$A=0$;

2) 对任意实数k及任意n级方阵A，有$\|kA\|=|k|\|A\|$;

3) 对任意两个n级方阵A,B，有

$$\|A+B\|\leq\|A\|+\|B\|,\ \ \|AB\|\leq\|A\|\|B\|;$$

4) 对$\forall x\in R^n$及任意n级方阵A，有$\|Ax\|\leq\|A\|\|x\|$.

由矩阵范数的定义及其性质可知，矩阵范数与向量范数之间存在着一定的对应关系.

定理1. 3 设有n级实方阵$A=(a_{ij})$，则与l_1、l_2、l_∞范数相容的矩阵范数分别为

$$\|A\|_1=\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|, \tag{7}$$

$$\|A\|_2=\sqrt{\rho(A^TA)}, \tag{8}$$

$$\|A\|_\infty=\max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|, \tag{9}$$

其中$\rho(\cdot)$为矩阵谱半径，其满足$\rho(B)=\max_{1\le i\le n}|\lambda_i^B|$, λ_i^B为方阵B的特征值.

证式(8): 由于A^TA为对称非负定阵，则其特征值$\lambda_i\ (i=1,2,\cdots,n)$非负，且存在$n$维正交方阵$H$，使得$A^TA=H^T\mathrm{diag}(\lambda_i)H$, 其中

$$\mathrm{diag}(\lambda_i)=\begin{bmatrix}\lambda_1 & & & 0\\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n\end{bmatrix}.$$

$\forall x\in\{x|\ \ \|x\|_2=1\}$, 若记$Hx=y=(y_1,y_2,\cdots,y_n)^T$, 则有$\|y\|_2^2=x^TH^THx=\|x\|_2^2=1$ 及

$$\|Ax\|_2^2=x^TA^TAx=(Hx)^T\mathrm{diag}(\lambda_i)(Hx)=\sum_{i=1}^n\lambda_iy_i^2\le(\max_{1\le i\le n}\lambda_i)\|y\|_2^2.$$

从而$\|Ax\|_2\le\sqrt{\rho(A^TA)}$. 若$\rho(A^TA)$对应矩阵$A^TA$的单位特征向量为$\tilde{x}$，则

$$\|A\|_2^2\ge\|A\tilde{x}\|_2^2=\tilde{x}^TA^TA\tilde{x}=\tilde{x}^T\rho(A^TA)\tilde{x}=\rho(A^TA)\|\tilde{x}\|_2^2=\rho(A^TA),$$

即$\|A\|_2\ge\sqrt{\rho(A^TA)}$. 故式(8)得证。

定理1.4 设A为n级方阵，则对任意矩阵范数$\|\cdot\|$ 有$\rho(A) \le \|A\|$.

证 设λ为阵A的任一特征值，x为其相应的特征向量，则

$$|\lambda|\|x\| = \|\lambda x\| = \|Ax\| \le \|A\|\|x\|,$$

即$|\lambda| \le \|A\|$. 故$\rho(A) \le \|A\|$.

此外矩阵范数与谱半径之间还存在如下关系:

定理1.5 $\forall \varepsilon > 0$,必存在$R^{n\times n}$ 中的某范数$\|\cdot\|$，使得

$$\|A\| \le \rho(A) + \varepsilon, \quad A\text{为任意}n\text{级方阵}.$$

记$R^{n\times m}$ 为全体实$n \times m$级矩阵的集合，在矩阵范数的概念下，我们可讨论矩阵序列的收敛性。

定义1.4 设有矩阵序列$\{A^{(k)}|\ A^{(k)} = (a_{ij}^{(k)})\} \subset R^{n\times m}$，若

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}, \ \ i = 1, 2, \cdots, n; \ j = 1, 2, \cdots, m,$$

则称矩阵序列$\{A^{(k)}\}$收敛于矩阵$A = (a_{ij})$, 记为$\lim\limits_{k\to\infty} A^{(k)} = A$.

矩阵序列有类似于向量序列的收敛性结果。

定理1.6 设有矩阵序列$\{A^{(k)}\} \subset R^{n\times n}$, 则$\lim\limits_{k\to\infty} A^{(k)} = A$的充要条件是存在矩阵范数$\|\cdot\|$, 使得$\lim_{k\to\infty} \|A^{(k)} - A\| = 0$.

定理1.7 $\forall A \in R^{n\times n}$, $\lim\limits_{m\to\infty} A^m = 0$的充要条件是$\rho(A) < 1$.

证 根据定理1.6，本定理仅需证明：对某矩阵范数$\|\cdot\|$, $\lim\limits_{m\to\infty} \|A^m\| = 0$的充要条件是$\rho(A) < 1$. 事实上，一方面由定理1.4，有

$$[\rho(A)]^m = \rho(A^m) \le \|A^m\| \to 0 \ \ (m \to \infty),$$

从而$\rho(A) < 1$. 另一方面，由于$\rho(A) < 1$，则存在$\varepsilon > 0$，使得

$$\rho(A) + \varepsilon < 1.$$

进一步，由定理1.5存在$R^{n\times n}$中某范数$\|\cdot\|$使得

$$\|A^m\| \le \|A\|^m \le (\rho(A) + \varepsilon)^m \to 0 \ \ (m \to \infty).$$

由此得$\lim\limits_{m\to\infty} \|A^m\| = 0$.

定理1.8 设有n级方阵A，若存在矩阵范数$\|\cdot\|$，使得$\|A\| < 1$，则$I - A$非奇异，且有

$$\|(I-A)^{-1}\| \le \frac{1}{1-\|A\|},$$

其中I为n级单位阵.

证 由于

$$\rho(A) = \max_{1\le i\le n} |\lambda_i^A| \le \|A\| < 1,$$

则$I - A$非奇异，且$\lim\limits_{k\to\infty} A^k = 0$. 又由

$$(I-A)(I+A+A^2+\cdots+A^k) = I - A^{k+1}$$

有

$$I+A+A^2+\cdots+A^k = (I-A)^{-1}(I-A^{k+1}),$$

令$k\to\infty$得

$$(I-A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

因此

$$\|(I-A)^{-1}\| \le \sum_{k=0}^{\infty} \|A\|^k = \frac{1}{1-\|A\|}.$$

例1.3 证明摄动定理：设矩阵$A, B \in \mathbb{R}^{n\times n}$，其中$A$ 可逆，且存在矩阵范数$\|\cdot\|$ 使得

$$\|A^{-1}\| \le \alpha,\ \ \|A-B\| \le \beta,\ \ \alpha\beta < 1,$$

那么矩阵B 也可逆，且

$$\|B^{-1}\| \le \frac{\alpha}{1-\alpha\beta}.$$

证 由已知条件有

$$\|I - A^{-1}B\| = \|A^{-1}(A-B)\| \le \alpha\beta < 1.$$

又$A^{-1}B = I - (I - A^{-1}B)$。因此，据定理1.8$A^{-1}B$可逆，从而$B$也可逆，且有

$$\|B^{-1}\| = \|[I-(I-A^{-1}B)]^{-1}A^{-1}\| \le \|[I-(I-A^{-1}B)]^{-1}\|\|A^{-1}\| \le \frac{\alpha}{1-\alpha\beta}.$$

§1.2.2 差分方程

形如

$$F(t;x(t),x(t+1),\cdots,x(t+k))=0, \tag{10}$$

的方程称为k**阶差分方程**。以下，我们将重点讨论t取有理数或整数情形的线性差分方程

$$\sum_{j=0}^{k} a_j(n)x_{n+j}=b_n,\quad n=0,1,2,\cdots, \tag{11}$$

其中系数$a_j(n)$、$b_n\in C$，且$a_k(n)a_0(n)\neq 0$. 若给定其k个初始值$x_0,x_1,\cdots,x_{k-1}$，则由(??)即可求出其解序列$\{x_n\}$. 特别，若$b_n=0\ (n=0,1,2,\cdots)$, 则称之为**齐次的**，否则称为**非齐次的**。线性差分方程具有与线性常微分方程相类似的性质。

定理1.9 若(11)为齐次差分方程，$x_n^{(1)}, x_n^{(2)}, \cdots, x_n^{(m)}$为该方程的一组特解，则其任意线性组合$\sum\limits_{i=1}^{m} c_i x_n^{(i)}$仍为该方程的解，其中$c_i$为任意常数。

定理1.10 若(11)为齐次差分方程，$x_n^{(1)}, x_n^{(2)}, \cdots, x_n^{(k)}$为其线性无关的特解(称为**基本解组**)，则$\sum\limits_{i=1}^{k} c_i x_n^{(i)}$为该方程的通解。

定理1.11 若(11)为齐次差分方程，则其解$x_n^{(1)}, x_n^{(2)}, \cdots, x_n^{(k)}$线性无关的充要条件是相应的Wronski行列式$\det(W) \neq 0$，其中

$$W = \begin{bmatrix} x_0^{(1)} & x_0^{(2)} & \cdots & x_0^{(k)} \\ x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(k)} \\ \vdots & \vdots & & \vdots \\ x_{k-1}^{(1)} & x_{k-1}^{(2)} & \cdots & x_{k-1}^{(k)} \end{bmatrix}.$$

定理1.12 非齐次差分方程(11)的通解可表示为它的任一特解与相应的齐次方程的通解之和。

以上性质可由差分方程解的定义直接获证。

若(11)中系数$a_i(n)$均与n无关，则得k阶常系数差分方程

$$\sum_{j=0}^{k} a_j x_{n+j} = b_n, \quad n = 0, 1, 2, \cdots. \tag{12}$$

特别取$b_n = 0\ (n = 0, 1, 2, \cdots)$, 即得相应齐次差分方程

$$\sum_{j=0}^{k} a_j x_{n+j} = 0, \quad n = 0, 1, 2, \cdots. \tag{13}$$

设(13)有形如$x_n = y^n\ (y \neq 0)$的解，则将其代入(13)即得代数方程

$$\sum_{j=0}^{k} a_j y^j = 0, \tag{14}$$

我们称(14)为方程(13)的**特征方程**，而其根称为**特征根.** 因此，我们有

定理1.13 当y为特征根时，$x_n = y^n (n = 0, 1, 2, \cdots)$为齐次方程(13)的解。

定理1.15 若特征方程(14)的全体互异根为$\{y_i\}_{i=1}^p$, 其中y_i的重数为m_i，则$\{y_i^n, ny_i^n, \cdots, n^{m_i-1}y_i^n\}_{i=1}^p$ 为(13)的基本解组，且此时(13)的通解可表为

$$x_n = \sum_{i=1}^{p}\sum_{j=0}^{m_i-1} c_{ij}n^j y_i^n,\ \text{其中诸}c_{ij}\text{为任意常数}.$$

定理1.16 在定理1.15的假设条件下，非齐次差分方程(12)的通解可表为$x_n = \sum\limits_{i=1}^{p}\sum\limits_{j=0}^{m_i-1} c_{ij}n^j y_i^n + x_n^*$, 其中$x_n^*$为(12)的某个特解。

(12)的特解可取为以$x_0 = x_1 = \cdots = x_{k-1} = 0$为初值的解$x_n^* = \sum\limits_{q=0}^{n-k} b_q\tilde{x}_{n-q-1}$ 其中$\tilde{x}_n$为相应常系数齐次差分方程初值问题

$$\begin{cases} \sum\limits_{j=0}^{k} a_j x_{n+j} = 0, & n = 0, 1, 2, \cdots, \\ x_0 = x_1 = \cdots = x_{k-2} = 0, & x_{k-1} = \frac{1}{a_k} \end{cases}$$

的解，特别，若$b_n \equiv b$，且$\sum\limits_{j=0}^{k} a_j \neq 0$, 则(12)的特解可取为$x_n^* = b/\sum\limits_{j=0}^{k} a_j$.

例1.2 试求解差分方程初值问题

$$\begin{cases} x_{n+4}+2x_{n+3}+3x_{n+2}+2x_{n+1}+x_n=9, \\ x_0=x_1=x_3=0, \quad x_2=-1. \end{cases}$$

解 其差分方程的特征方程为

$$y^4+2y^3+3y^2+2y+1=0,$$

即

$$(y^2+y+1)^2=0,$$

解之得

$$y_{1,\,2}=\cos\frac{2\pi}{3}-i\sin\frac{2\pi}{3}, \quad y_{3,\,4}=\cos\frac{2\pi}{3}-i\sin\frac{2\pi}{3}.$$

因此, 其差分方程对应的齐次方程的通解为

$$\begin{aligned} x_n &= (\hat{c}_1+\hat{c}_2 n)(\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3})^n+(\hat{c}_3+\hat{c}_4 n)(\cos\frac{2\pi}{3}-i\sin\frac{2\pi}{3})^n \\ &= (c_1+c_2 n)\cos\frac{2\pi n}{3}+(c_3+c_4 n)\sin\frac{2\pi n}{3}. \end{aligned}$$

而原差分方程有特解$x_n=1$, 则该方程的通解为

$$x_n=(c_1+c_2n)\cos\frac{2\pi n}{3}+(c_3+c_4n)\sin\frac{2\pi n}{3}+1.$$

将初始条件代入其中得

$$\begin{cases} c_1=-1, \\ (c_1+c_2)\cos\frac{2\pi}{3}+(c_3+c_4)\sin\frac{2\pi}{3}=-1 \\ c_1+3c_2=-1, \\ (c_1+2c_2)\cos\frac{4\pi}{3}+(c_3+2c_4)\sin\frac{4\pi}{3}=-2. \end{cases}$$

解之得

$$c_1=-1,\ \ c_2=0,\ \ c_3=-\frac{11\sqrt{3}}{3},\ \ c_4=\frac{8\sqrt{3}}{3}.$$

故其初值问题的解为

$$x_n=-\cos\frac{2\pi n}{3}-\frac{\sqrt{3}}{3}(11-8n)\sin\frac{2\pi n}{3}+1.$$

§1.3 误 差

利用数值方法求得的数值解是一个近似结果，因此总给人们一个不严格的感觉。其实，现实世界各种问题的求解过程中，误差产生是绝对的，而精确存在是相对的。当然，数值解的近似程度必须是原问题所容许的、合理的，否则，其计算结果将毫无意义。误差产生的原因是多样的，就数学建模而言，有可能忽略问题的一些次要因素，因而产生模型误差。此外，模型的原始数据是通过观测获得的，因而又有观测误差。在本课程的学习中，我们仅考虑数值计算带来的误差。

§1.3.1 绝对误差与相对误差

定义1.5 设x^*是某量的精确值，x是其近似值，则称差$e := x^* - x$为x的绝对误差。

由于问题的精确解往往是未知的，因此要直接确定e通常是困难的。实际应用中常用满足$|e| \leq \varepsilon$的较小正数ε来表征绝对误差，而称ε为**绝对误差限**. 若某实际问题中精确解x^*的近似值x有绝对误差限ε，则记$x = x^* \pm \varepsilon$。值得注意的是绝对误差概念一般只能用于比较相同条件下近似值的精度，如：甲、乙两个学生分别做满分为100分和150分的试题，甲得90分，乙得139分，显然若从绝对误差角度来衡量两者成绩优劣，则会导致乙比甲成绩差的不合理结果。

为什么这种比较不合理呢？原因是乙做150分题才错11分，而甲做100分题就错了10分。这里，人们在考虑绝对误差的同时也将其与原量的精确值大小进行了比较。为此，我们引入衡量近似值精度另一尺度。

定义1.6 在定义1.5的条件下，称比值$e_r = e/x^*$为近似值的相对误差。上例中甲的相对误差为10%，乙的相对误差约为7.3%，因此乙的成绩优于甲。同样，我们通常用满足不等式$|e_r| \leq \varepsilon_r$的**相对误差限**ε_r来表征相对误差，且在实际应用中常近似地取$\varepsilon_r = \varepsilon/|x|$.

实际工作中，人们也总结出用近似值的有效数字位数来表征其精度。

定义1.7 若x^*的近似值$x = \pm 0.x_1x_2\cdots x_n \times 10^m$，其中$x_1 \neq 0$，诸$x_i \in \{0,1,2,\cdots,9\}$，$m$为整数，且

$$|x - x^*| \leq \frac{1}{2} \times 10^{m-p},\ \ 1 \leq p \leq m,$$

则称近似值x有p位有效数字或称x准确到10^{m-p}位。

如设有圆周率π的近似数为$x = 3.142$，由于其绝对误差

$$|\pi - x| = 0.000407\cdots < 0.0005 = \frac{1}{2} \times 10^{-3},$$

则近似数x有4位有效数字。此外，相对误差与有效数字具下述关系。

定理1.17 若近似值x具有定义1.7中的形式，且有p位有效数字，则其相对误差限$\varepsilon_r = \frac{1}{2x_1} \times 10^{-p+1}$.

一个算法的优劣除需考虑其误差外，其误差传播也是不容忽视的。刻画误差传播状况的概念是数值稳定性，若误差传播是可控的，则称之为**数值稳定**的，否则称之为**数值不稳定**.

§1.3.2 误差的Richardson外推估计法

Richardson外推法是提高数值解精度及估计误差的常用方法。记$y(t,h)$ 为某数值方法以步长h求解某定解问题真解$y(t)$的近似值。设$y(t,h)$关于h可展开为

$$y(t,h) = y(t) + \sum_{i=1}^{\infty} A_i h^i, \tag{15}$$

若将方法的步长减半为$\frac{1}{2}$, 则有

$$y(t,\frac{h}{2}) = y(t) + \sum_{i=1}^{\infty} \frac{A_i}{2^i} h^i. \tag{16}$$

由(15)、(16)得

$$2y(t,\frac{h}{2}) - y(t,h) = y(t) - \frac{1}{2}A_2h^2 - (1-\frac{1}{2^2})A_3h^3 + \cdots .$$

故当取$\tilde{y}(t) := 2y(t,\frac{h}{2}) - y(t,h) \approx y(t)$时，其截断误差为$\mathcal{O}(h^2)$。

进一步，若$y(t,h)$的截断误差为$\mathcal{O}(h^p)$，则有

$$y(t,h)=y(t)+\sum_{i=p}^{\infty}A_ih^i,\quad y(t,\frac{h}{2})=y(t)+\sum_{i=p}^{\infty}\frac{A_i}{2^i}h^i. \tag{17}$$

由上两式可得

$$2^py(t,\frac{h}{2})-y(t,h)=(2^p-1)y(t)-\frac{1}{2}A_{p+1}h^{p+1}+\cdots,$$

从而$y(t)$的近似值

$$\tilde{y}(t)=\frac{1}{2^p-1}[2^py(t,\frac{h}{2})-y(t,h)], \tag{18}$$

其截断误差为$\mathcal{O}(h^{p+1})$。上述方法即为**Richardson外推法**。

利用Richardson外推法也可获得数值解的误差估计。设所用方法的截断误差为$\mathcal{O}(h^p)$，并记$\varepsilon_h=y(t)-y(t,h)$, $\varepsilon_{\frac{h}{2}}=y(t)-y(t,\frac{h}{2})$, 由(17)得

$$\varepsilon_{\frac{h}{2}}=\frac{1}{2^p}\varepsilon_h+\mathcal{O}(h^{p+1}).$$

将该式代入$\varepsilon_h = y(t) - y(t,h)$得

$$y(t) = y(t,h) + 2^p \varepsilon_{\frac{h}{2}} + \mathcal{O}(h^{p+1}).$$

再将上式代入$\varepsilon_{\frac{h}{2}} = y(t) - y(t, \frac{h}{2})$得

$$\varepsilon_{\frac{h}{2}} = \frac{1}{2^p - 1}[y(t, \frac{h}{2}) - y(t,h)] + \mathcal{O}(h^{p+1}).$$

由此即得方法的截断误差主项

$$\tilde{\varepsilon}_{\frac{h}{2}} = \frac{1}{2^p - 1}[y(t, \frac{h}{2}) - y(t,h)]. \tag{19}$$

由于(19)仅是方法的截断误差主项，因此实际估计误差时, 人们保守地以$\triangle := |y(t, \frac{h}{2}) - y(t,h)|$ 作为误差。利用该量可获得实际计算的终止准则：设$\varepsilon > 0$为预定精度且$\triangle > \varepsilon$，则反复将步长折半进行计算，直至$\triangle < \varepsilon$，这时取步长最终折半后的计算结果作为欲求数值解。