



中国海洋大学
OCEAN UNIVERSITY OF CHINA

中国海洋大学

水产学院

数值分析

Author:

韩方成

Student ID:

18060013010

数值分析笔记

留一份不足，可得无限美好

2021 年 7 月 30 日

目录

1	基础	1
1.1	多项式计算	1
1.2	二进制	2
1.3	实数的浮点数表示法	3

1 基础

1.1 多项式计算

计算多项式

$$P(x) = 2x^4 + 3x^3 - 3x^2 + 5x - 1$$

在 $x = 1/2$ 的最好的方法是什么？我们假设多项式的系数和数字 $1/2$ 都存储在内存中，我们尽量去减少计算的加法和乘法。我们不考虑从内存中取值所耗费的时间。

方法 1 首先最直接的方法是

$$P\left(\frac{1}{2}\right) = 2 * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} + 3 * \frac{1}{2} * \frac{1}{2} * \frac{1}{2} - 3 * \frac{1}{2} * \frac{1}{2} + 5 * \frac{1}{2} - 1 = \frac{5}{4} \quad (1.1)$$

此方法共执行乘法 10 次，加法 4 次。很显然，有比此方法更简单的方法。

方法二 首先先计算 $x = 1/2$ 的幂，存储以后使用：

$$\begin{aligned} \frac{1}{2} * \frac{1}{2} &= \left(\frac{1}{2}\right)^2 \\ \left(\frac{1}{2}\right)^2 * \frac{1}{2} &= \left(\frac{1}{2}\right)^3 \\ \left(\frac{1}{2}\right)^3 * \frac{1}{2} &= \left(\frac{1}{2}\right)^4 \end{aligned}$$

只有，我们可以将这些项加起来：

$$P\left(\frac{1}{2}\right) = 2 * \left(\frac{1}{2}\right)^4 + 3 * \left(\frac{1}{2}\right)^3 - 3 * \left(\frac{1}{2}\right)^2 + 5 * \frac{1}{2} - 1 = \frac{5}{4}$$

我们执行了 7 次乘法和 4 次加法。虽然相比上个方法执行运算的次数减少了，但是还有更好的方法。

方法三 (嵌套乘法)

$$\begin{aligned} P(x) &= -1 + x(2x^3 + 3x^2 - 3x + 5) \\ &= -1 + x(5 + x(2x^2 + 3x - 3)) \\ &= -1 + x(5 + x(-3 + x(2x + 3))) \end{aligned} \quad (1.2)$$

这套方法被称为嵌套法或霍纳法。一个 d 次多项式需要计算 d 次加法和 d 次乘法。

当多项式的标准格式为 $c_1 + c_2x + c_3x^2 + c_4x^3 + c_5x^4$ ，可以被写为嵌套形式：

$$c_1 + x(c_2 + x(c_3 + x(c_4 + x(c_5)))) \quad (1.3)$$

一些应用需要更一般的形式。特别的，插值计算需要以下形式：

$$c_1 + (x - r_1)(c_2 + (x - r_2)(c_3 + (x - r_3)(c_4 + (x - r_4)(c_5)))) \quad (1.4)$$

其中 r_1, r_2, r_3, r_4 为基点。注意令式 (1.4) 的 $r_1 = r_2 = r_3 = r_4 = 0$ 我们将得到式 (1.3)。

Listing 1: **Nest.m**

```

1 function y = Nest(d, c, x, b)
2 if nargin < 4
3     b = zeros(d, 1);
4 end
5 y = c(d+1);
6 for i = d:-1:1
7     y = y.*(x - b(i)) + c(i);
8 end

```

1.2 二进制

二进制数表示为

$$\cdots b_2b_1b_0.b_{-1}b_{-2}\cdots$$

每个二进制位或比特是 0 或 1。10 进制数等同于

$$\cdots b_22^2 + b_12^1 + b_02^0 + b_{-1}2^{-1} + b_{-2}2^{-2}\cdots$$

precision	sign	exponent	mantissa
single	1	8	23
double	1	11	52
long double	1	15	64

表 1.1: 浮点数的三种表示方法

1.3 实数的浮点数表示法

当使用有限精度的计算机存储单元来表示真实的、无限精度的数字时，舍入误差是不可避免的。

浮点数格式 IEEE 标准由一组实数的二进制表示法组成。浮点数由三部分组成：符号 (+ 或 -)、尾数 (包含有效位) 和指数。这三个部分一起存储在一个计算机字节中。

浮点数有三种常用的精度级别：单精度、双精度和扩展精度，也称为长双精度。分配给三种格式的每个浮点数的位数分别为 32、64 和 80。按如下方式在各部分之间划分比特：