前言

说来有趣，这份资料一开始叫《工程数值方法学习笔记》，本来是给自己复习用的，所以只是零零星星地写了一些关键点.

后来发现光记知识点没用，学会做题才是第一要务，于是就增添了一些例题及解法.

再后来，我发现期末考试的范围和题型比较固定，所以就摘取了“考试重点”，并把资料改名为《工程数值方法速通教程》，希望读者可以借此“速通”这门课.

不过我发现，“速通”的内容还是过于简略了，不利于理解，于是就一点一点地扩充了原先的内容.

到现在，这份资料已经基本完工，就差一些校对工作. 与历史版本相比，主要是增加了更多的例题，以及把23-24夏学期的回忆卷拆散，分摊到各个章节并附上我自己的解答. 所以，这里的“答案”并不一定准确，还需要各位一起校对.

不知不觉间，已经变成50多页的“皇皇巨著”了呢. 我想，这个时候再称之为“速通”教程可能并不准确了. 既然是面向期末考试的资料，不妨就叫它《工程数值方法考试指南》吧.

这里的内容并不需要全部理解，根据自己的能力和目标，以及所剩的复习时间，综合考虑该如何作取舍吧. 毕竟这门课的期末考试只占30%，所以不要有太大的压力.

打*的内容是我认为不会考的，不过并不保证. 这份资料用LaTeX编写，目录上的章节标题是可以跳转的.

本来我想狂妄地说“祝卷面满分”的，但是回忆卷的某些题属实让我颜面扫地. 讲求实际一些，那就祝考试顺利吧！

by 折一只纸鹭(Komeiji Ren)

【附】2023-2024夏学期《工程数值方法》考试大纲

1. 数值计算：误差、有效数字（3小题，15分）
2. 线性方程组：向量范数、矩阵条件数、雅可比迭代法和高斯迭代法（3小题，15分）
3. 插值：拉格朗日插值、牛顿插值（2小题，10分）
4. 非线性方程求根：二分法、牛顿法（2小题，15分）
5. 曲线拟合：最小二乘法（15分）
6. 数值积分：中点法、梯形法、辛普森法则等（15分）
7. 微分方程数值解：欧拉法、经典龙格库塔法（15分）

目录

1 误差与有效数字 **4**
1.1 *误差的来源 4
1.2 误差的基本概念 4
1.3 有效数字 5
1.4 误差传播 7
1.5 *误差传播的注意事项 7

2 线性方程组 **8**
2.1 *高斯消元法 8
2.2 *误差和剩余向量 9
2.3 向量范数 9
2.4 矩阵范数 10
2.5 逆矩阵 11
2.6 矩阵的条件数 14
2.7 雅可比迭代 15
2.8 高斯-赛德尔迭代 16

3 插值 **19**
3.1 插值概述 19
3.2 拉格朗日插值（全阶多项式插值） 19
3.3 用插值来近似函数 20
3.4 插值余项 20
3.5 差商 21
3.6 牛顿插值 22
3.7 牛顿插值法的插值余项 22
3.8 差商表 23
3.9 差分 24
3.10 等距节点的牛顿插值法 25
3.11 差分表 25

4 曲线拟合 **28**
4.1 最小二乘直线拟合 28
4.2 最小二乘算法的一般形式 28

4.3 *最小二乘幂函数拟合 30
4.4 最小二乘指数拟合 30

5 非线性方程求根 32
5.1 二分法 32
5.2 牛顿法 33
5.3 *牛顿迭代法的收敛性判断 36
5.4 *近似割线法 37

6 数值积分 38
6.1 数值积分概述 38
6.2 中点法 38
6.3 梯形法 38
6.4 辛普森法则 38
6.5 复合辛普森法则（区间二等分） 39
6.6 复合法 39
6.7 数值积分公式的精度 40
6.8 *牛顿-科斯特(Newton-Cotes)法则 41

7 微分方程数值解 42
7.1 数值微分 42
7.2 微分方程数值解的概述 42
7.3 常微分方程的一些基本概念 42
7.4 欧拉法 43
7.5 截断误差与精度分析 44
7.6 后退(隐式)欧拉法 45
7.7 改进欧拉法 46
7.8 泰勒展开 47
7.9 龙格-库塔(Runge-Kutta)法的构建思路 47
7.10 二阶龙格-库塔法 48
7.11 四阶经典龙格-库塔法(RK4) 50

1 误差与有效数字

【考纲要求】误差、有效数字

1.1 *误差的来源

1. 模型误差（Modeling Error）：从实际问题中抽象出数学模型会产生误差
2. 观测误差（Measurement Error）：观测得到某些参数或物理量的值会产生误差
3. 方法误差（截断误差，Truncation Error）：通过数值方法而非解析方法求得的是近似解
4. 舍入误差（Roundoff Error）：由于机器字长有限，舍入（如四舍五入）不可避免

1.2 误差的基本概念

设x为**真值**（精确值），x^*为近似值，则有如下概念：

（1）**绝对误差**（简称**误差**，absolute error），可正可负

$$e^* = x^* - x$$

（2）**绝对误差限**（accuracy），误差绝对值的上界，正的

$$|e^*| = |x^* - x| \le \varepsilon^*$$

式中的ε^*就是绝对误差限

（3）**相对误差**（relative error），可正可负

$$e_r^* = \frac{e^*}{x} = \frac{x^* - x}{x}$$

但是，在实际计算时，相对误差常取

$$e_r^* = \frac{e^*}{x^*} = \frac{x^* - x}{x^*}$$

（4）**相对误差限**（relative accuracy），相对误差绝对值的上界，正的

$$|e_r^*| = \left|\frac{e^*}{x^*}\right| = \left|\frac{x^* - x}{x^*}\right| \le \varepsilon_r^*$$

式中的ε_r^*就是相对误差限

1.3 有效数字

数值计算里的有效数字和我们平常所说的有效数字不太一样.

严谨的定义：若某数的近似值x^*的误差不大于该数某一位数字的半个单位，该位到x^*最左边的第一位非零数字都是该数的有效数字，其个数称为该数的有效位数。

也就是说，我们要寻找某一位数，使得

$$|e^*| = |x^* - x| \leq \frac{1}{2} \times 10^{-n}$$

我知道你听不懂，举个例子就懂了. 我们熟知的$\pi = 3.1415926...$，这是它的真值，有如下分析：

（1）若$\pi^* = 3.14$，则π^*的有效位数是3位，因为它的误差

$$|\pi^* - \pi| = 0.0015926... < \frac{1}{2} \times 10^{-2}$$

所以定位到10^{-2}这一位，用颜色标出，就是3.14，然后看它到最左边一共有3位非零数字（3, 1, 4），所以其有效位数是3位.

看起来和平常所说的有效数字也没什么不同？别急，且看下面这个.

（2）若$\pi^* = 3.141$，则π^*的有效位数是3位（而不是4位），因为它的误差

$$|\pi^* - \pi| = 0.0005926... > \frac{1}{2} \times 10^{-3}$$

超了，只能退而求其次

$$|\pi^* - \pi| = 0.0005926... < \frac{1}{2} \times 10^{-2}$$

所以它也是定位到10^{-2}这一位，用颜色标出，就是3.141，然后看它到最左边一共有3位非零数字（3, 1, 4），所以其有效位数是3位.

（3）若$\pi^* = 3.142$，则π^*的有效位数是4位，因为

$$|\pi^* - \pi| = 0.0004074... < \frac{1}{2} \times 10^{-3}$$

所以定位到10^{-3}这一位，用颜色标出，就是3.142，然后看它到最左边一共有4位非零数字（3, 1, 4, 2），所以其有效位数是4位.

例1.1【23-24夏试题改编】求下面三个数值的有效数字位数：

(1) $a = 0.335$，其真值为$\frac{1}{3}$.

(2) $b = 0.721$，其绝对误差限为0.005.

(3) $c = 1.618$，其真值为1.985.

解 (1)误差

$$|e^*| = |a^* - a| = 0.00166... < \frac{1}{2} \times 10^{-2}$$

因此有效数字2位（0.33）.

(2)误差

$$|e^*| \leq 0.005 = \frac{1}{2} \times 10^{-2}$$

因此有效数字2位（0.72）.

(3)误差

$$|e^*| = |c^* - c| = 0.367 < \frac{1}{2} \times 10^{0}$$

因此有效数字1位（1）.

例1.2【PPT】为使π^*的相对误差小于0.001%，应取几位有效数字？

解 设取n位有效数字，因为整数部分有1位有效数字，因此小数部分有$(n-1)$位有效数字. 于是绝对误差满足

$$|\pi^* - \pi| \leq \frac{1}{2} \times 10^{-(n-1)}$$

则相对误差

$$|e_r^*| = \frac{|\pi^* - \pi|}{\pi} \leq \frac{0.5 \times 10^{-n+1}}{3} = \frac{1}{6} \times 10^{-n+1} < 10^{-5}$$

解得$n > 6 - \lg 6$，因此$n \geq 6$，取6位有效数字，即$\pi^* = 3.14159$.

例1.3【23-24夏试题】要使$\sqrt{70}$的近似值的相对误差小于0.01%，则应取几位有效数字？

解 设取n位有效数字，因为整数部分有1位有效数字，因此小数部分有$(n-1)$位有效数字. 于是绝对误差满足

$$|x^* - x| \leq \frac{1}{2} \times 10^{-(n-1)}$$

则相对误差

$$|e_r^*| = \frac{|x^* - x|}{x} \leq \frac{0.5 \times 10^{-n+1}}{8} = \frac{1}{16} \times 10^{-n+1} < 10^{-4}$$

解得$n > 5 - \lg 16$，因此$n \geq 4$，取4位有效数字，即$x^* = 8.367$.

1.4 误差传播

函数$y=f(x)$的绝对误差

$$|e^*(y)| \approx |f'(x^*)| \cdot |e^*(x)| = |f'(x^*)| \cdot |x^* - x|$$

其中，$|f'(x^*)|$叫做**绝对条件数**.

相对误差

$$|e_r^*(y)| \approx \left|\frac{f'(x^*)}{f(x^*)}\right| \cdot |e^*(x)| = \left|\frac{x^* \cdot f'(x^*)}{f(x^*)}\right| \cdot |e_r^*(x)|$$

其中，$\left|\frac{x^* \cdot f'(x^*)}{f(x^*)}\right|$叫做**相对误差条件数**.

例1.4【PPT】若$x=10\pm5\%$，求函数$f(x)=\sqrt[n]{x}$的相对误差限.

解 近似值$x^*=10$，绝对误差限$|e^*(x)|=5\%=0.05$，则

$$|e_r^*(y)| \approx \left|\frac{f'(x^*)}{f(x^*)}\right| \cdot |e^*(x)| = \left|\frac{\frac{1}{n}(x^*)^{\frac{1}{n}-1}}{(x^*)^{\frac{1}{n}}}\right| \cdot |e^*(x)| = \frac{|e^*(x)|}{nx^*} = \frac{0.05}{10n} = \frac{0.005}{n}$$

例1.5【23-24夏试题】已知x的相对误差限为δ，且$x>0$，求$f(x)=\ln x$的绝对误差限（用含δ的式子表示）.

解 x的相对误差限

$$\left|\frac{x^*-x}{x^*}\right| = \delta$$

因为$x>0$，所以

$$|x^*-x| = \delta x^*$$

故$f(x)=\ln x$的绝对误差限

$$|e^*(y)| \approx |f'(x^*)| \cdot |x^*-x| = \frac{1}{x^*} \cdot \delta x^* = \delta$$

1.5 *误差传播的注意事项

1. 避免相近两数相减
2. 避免小分母，因为会造成浮点溢出
3. 避免大数吃小数
4. 先化简再计算，减少步骤，避免误差累计
5. 选用稳定的算法

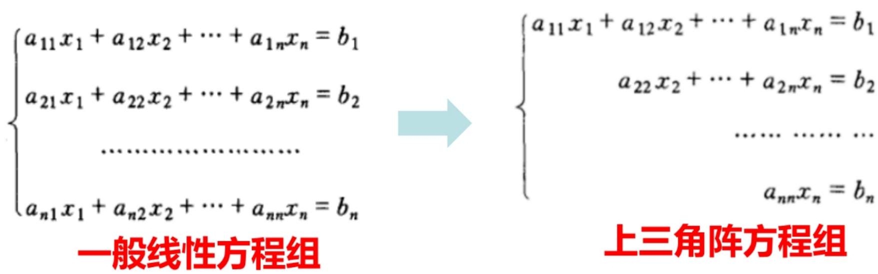
2 线性方程组

【考纲要求】向量范数、矩阵条件数、雅可比迭代法和高斯迭代法

2.1 *高斯消元法

包含前向消元、选主元、反向回代3个主要步骤.

前向消元和反向回代在线性代数里已经学过了，就是把一般的线性方程组变成阶梯形线性方程组，然后自下而上求解. 这里放个图：



至于选主元，这是为了减少计算过程中的舍入误差的影响. 在每次消元前，应选择绝对值大的元素作为约元的主元素，否则会出现小数消大数的情况，需要乘一个很大的数，放大了误差. 这是数值计算特有的，解析解不会有这样的顾虑.

虽然我感觉不会考，但保险起见还是拿一道题练练手吧，复习一下线性代数里的高斯消元法.

例2.1 求方程

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 4 \\ 0 & 6 & 7 \end{bmatrix} \boldsymbol{X} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

的解析解.

解 写出增广矩阵

$$\overline{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 2 & 3 & 4 & 1 \\ 0 & 6 & 7 & 0 \end{bmatrix}$$

高斯消元，化成阶梯形矩阵

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ & 3 & 2 & 1 \\ & & 3 & -2 \end{bmatrix}$$

从下往上逐一回代，解得

$$\boldsymbol{X} = \begin{bmatrix} \frac{2}{3} & \frac{7}{9} & -\frac{2}{3} \end{bmatrix}^{\mathrm{T}}$$

2.2 *误差和剩余向量

设方程

$$\boldsymbol{Ax} = \boldsymbol{b}$$

的理论解/解析解（theoretical solution）为$\boldsymbol{x}$，计算解/数值解（computed solution）为$\boldsymbol{x}^*$. 评价计算解的好坏有两种指标：

（1）**误差**（error）

$$\boldsymbol{e} = \boldsymbol{x} - \boldsymbol{x}^*$$

（2）**剩余向量**（residual）

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{Ax}^*$$

2.3 向量范数

向量范数用来度量向量的大小，其一般形式

$$||\boldsymbol{x}||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

不过这个一般形式没啥用，主要是用下面这3种：

（1）**1-范数**，又叫曼哈顿范数，是各元素的绝对值之和

$$||\boldsymbol{x}||_1 = \sum_{i=1}^{n} |x_i|$$

（2）**2-范数**，又叫欧几里得距离，是向量的模长

$$||\boldsymbol{x}||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

（3）**∞-范数**，又叫切比雪夫范数，是各元素绝对值的最大值

$$||\boldsymbol{x}||_\infty = \max |x_i|$$

例2.2【23-24夏试题改编】已知向量

$$\boldsymbol{x} = \begin{bmatrix} 7 & 3 & 5 & 5 & 6 & 0 & 8 \end{bmatrix}$$

求它的曼哈顿范数、欧几里得距离、切比雪夫范数.

解 曼哈顿范数

$$||\boldsymbol{x}||_1 = \sum_{i=1}^{n} |x_i| = 7+3+5+5+6+0+8 = 34$$

欧几里得距离

$$||\boldsymbol{x}||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{7^2+3^2+5^2+5^2+6^2+0^2+8^2} = \sqrt{208} \approx 14.42$$

切比雪夫范数

$$||\boldsymbol{x}||_\infty = \max |x_i| = 8$$

2.4 矩阵范数

矩阵范数的一般形式就不写了，没什么意义. 主要是用下面这3种：

（1）**1-范数**，又叫列和范数，其做法：拆成列向量，求它们的曼哈顿范数的最大值

$$||\boldsymbol{A}||_1$$

（2）**2-范数**，设λ为$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的最大特征值，则

$$||\boldsymbol{A}||_2 = \sqrt{\lambda}$$

（3）**∞-范数**，又叫行和范数，其做法：拆成行向量，求它们的曼哈顿范数的最大值

$$||\boldsymbol{A}||_\infty$$

记忆技巧：∞是横过来的，所以它是行和范数.（不对啊，这是半开卷，记它干嘛）

例2.3【PPT】设矩阵 $\boldsymbol{A} = \begin{bmatrix} 4 & -3 \\ 2 & 1 \end{bmatrix}$ ，求$||\boldsymbol{A}||_1$、$||\boldsymbol{A}||_2$、$||\boldsymbol{A}||_\infty$.

解 $||\boldsymbol{A}||_1$是列和范数，因此把矩阵拆成两个列向量

$$\boldsymbol{a}_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \boldsymbol{a}_2 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

分别求得它们的曼哈顿范数为6和4，于是取二者的最大值

$$||\boldsymbol{A}||_1 = \max\{6, 4\} = 6$$

同理得行和范数为

$$||\boldsymbol{A}||_\infty = \max\{7, 3\} = 7$$

而$||\boldsymbol{A}||_2$需要计算$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$的特征值$\lambda$. 由特征值的定义

$$|\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}| = 0$$

其中，$\boldsymbol{I}$是单位矩阵. 于是

$$\left|\begin{bmatrix} \lambda & \\ & \lambda \end{bmatrix} - \begin{bmatrix} 4 & 2 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 4 & -3 \\ 2 & 1 \end{bmatrix}\right| = 0$$

解得

$$\lambda = 15 \pm 5\sqrt{5}$$

取最大特征值开根号，得

$$||\boldsymbol{A}||_2 = \sqrt{\lambda_{\max}} = \sqrt{15 + 5\sqrt{5}} \approx 5.12$$

2.5 逆矩阵

下一节的矩阵条件数需要用到逆矩阵，苏芮老师的PPT上要求会手算矩阵的条件数，也就是说，求逆矩阵是必须会的前置知识.

如果你的计算器比较厉害，可以直接用计算器算逆矩阵. 如果没有这个功能，也不用紧张，考试给的矩阵应该是比较简单的，像23-24夏学期就给了个二阶矩阵，手算也方便的.

不是我看不起你，我相信你一定忘记怎么求逆矩阵了（因为我也忘了）. 没关系，我来教你. 求逆矩阵有两种方式，第一种是通过初等行变换，它的式子是

$$\left[\begin{array}{c:c} \boldsymbol{A} & \boldsymbol{E} \end{array}\right] \xrightarrow{\text{初等行变换}} \left[\begin{array}{c:c} \boldsymbol{E} & \boldsymbol{A}^{-1} \end{array}\right]$$

看不懂没事，拿一道例题你就懂了.

例2.4 用初等行变换求矩阵

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 4 \\ 0 & 6 & 7 \end{bmatrix}$$

的逆矩阵$\boldsymbol{A}^{-1}$.

解 将矩阵$\boldsymbol{A}$与单位矩阵$\boldsymbol{E}$合并，得新矩阵

$$\left[\begin{array}{c:c}\boldsymbol{A} & \boldsymbol{E}\end{array}\right]=\left[\begin{array}{ccc:ccc}1 & 0 & 1 & 1 & & \\ 2 & 3 & 4 & & 1 & \\ 0 & 6 & 7 & & & 1\end{array}\right]$$

通过初等行变换将左边变成单位矩阵，得

$$\left[\begin{array}{ccc:ccc}1 & & & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ & 1 & & -\frac{14}{9} & \frac{7}{9} & -\frac{2}{9} \\ & & 1 & \frac{4}{3} & -\frac{2}{3} & \frac{1}{3}\end{array}\right]$$

右边这块就是我们要求的逆矩阵

$$\boldsymbol{A}^{-1}=\begin{bmatrix}-\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{14}{9} & \frac{7}{9} & -\frac{2}{9} \\ \frac{4}{3} & -\frac{2}{3} & \frac{1}{3}\end{bmatrix}$$

第二种方法是利用伴随矩阵，它的式子是

$$\boldsymbol{A}^{-1}=\frac{\boldsymbol{A}^{*}}{|\boldsymbol{A}|}$$

其中，伴随矩阵

$$\boldsymbol{A}^{*}=\begin{pmatrix}A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn}\end{pmatrix}$$

在这里，A_{ij}是行列式$|\boldsymbol{A}|$中元素a_{ij}的代数余子式. 代数余子式的求法：把矩阵中元素a_{ij}所在的行和列删除，剩下的元素可构成一个新的行列式，求出这个行列式，再乘以系数$(-1)^{i+j}$.

例2.5 用伴随矩阵求矩阵

$$\boldsymbol{A}=\begin{bmatrix}1&0&1\\2&3&4\\0&6&7\end{bmatrix}$$

的逆矩阵$\boldsymbol{A}^{-1}$.

解 第一步，求伴随矩阵$\boldsymbol{A}^*$.

先求A_{11}，就是把矩阵$\boldsymbol{A}$中元素a_{11}所在的行和列删除，剩下的元素可构成一个新的行列式

$$\begin{vmatrix}3&4\\6&7\end{vmatrix}=3\times7-4\times6=-3$$

然后再乘以系数$(-1)^{1+1}=1$，得$A_{11}=-3$.

同样的，可以得到

$$A_{12}=(-1)^{1+2}\begin{vmatrix}2&4\\0&7\end{vmatrix}=-14$$

$$A_{13}=(-1)^{1+3}\begin{vmatrix}2&3\\0&6\end{vmatrix}=12$$

其他计算略，给出最终结果（注意代数余子式在伴随矩阵中的排列顺序）

$$\boldsymbol{A}^*=\begin{bmatrix}-3&6&-3\\-14&7&-2\\12&-6&3\end{bmatrix}$$

第二步，求行列式$|\boldsymbol{A}|$.

我相信你一定忘记怎么求三阶行列式了（因为我也忘了）. 没关系，可以将行列式按第一行展开

$$|\boldsymbol{A}|=a_{11}A_{11}+a_{12}A_{12}+a_{13}A_{13}=1\times(-3)+0\times(-14)+1\times12=9$$

第三步，求逆矩阵.

$$\boldsymbol{A}^{-1}=\frac{\boldsymbol{A}^*}{|\boldsymbol{A}|}=\begin{bmatrix}-\dfrac{1}{3}&\dfrac{2}{3}&-\dfrac{1}{3}\\[2ex]-\dfrac{14}{9}&\dfrac{7}{9}&-\dfrac{2}{9}\\[2ex]\dfrac{4}{3}&-\dfrac{2}{3}&\dfrac{1}{3}\end{bmatrix}$$

2.6 矩阵的条件数

定义矩阵的条件数为自身与其逆矩阵的范数乘积

$$\kappa(\boldsymbol{A}) = ||\boldsymbol{A}|| \cdot ||\boldsymbol{A}^{-1}||$$

也写作$\text{cond}(\boldsymbol{A})$，一般会指明是用1-范数、2-范数、$\infty$-范数中的哪一种，没有的话自己选.

对于一个线性方程组$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$，条件数$\kappa(\boldsymbol{A})$反映了它的稳定性，若其值远大于1，则说明这个线性方程组是病态的，微小的数值干扰就会引起解的极大波动.

例2.6 判断线性方程组

$$\begin{cases} x_1 + x_2 = 2 \\ x_1 + 1.0001x_2 = 2 \end{cases}$$

的稳定性.

解 系数矩阵及其逆矩阵为

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}, \boldsymbol{A}^{-1} = \begin{bmatrix} 10001 & -10000 \\ -10000 & 10000 \end{bmatrix}$$

若采用∞-范数计算，则其条件数

$$\kappa(\boldsymbol{A})_\infty = ||\boldsymbol{A}||_\infty \cdot ||\boldsymbol{A}^{-1}||_\infty = 2.0001 \times 20001 \approx 40004 \gg 1$$

说明这个方程组是病态的，非常不稳定. 若方程组的系数出现微小波动（比如把1.0001改成0.9999），则方程组的解就会出现巨大波动.

其实这个是比较好理解的，因为此时系数矩阵$\boldsymbol{A}$接近奇异（行列式趋近于0）.

例2.7【23-24夏试题】已知矩阵

$$\boldsymbol{A} = \begin{bmatrix} 10 & 30 \\ 20 & 40 \end{bmatrix}$$

求条件数$\text{cond}(\boldsymbol{A})_1, \text{cond}(\boldsymbol{A})_\infty$.

解 先求逆矩阵

$$\boldsymbol{A}^{-1} = \begin{bmatrix} -0.2 & 0.15 \\ 0.1 & -0.05 \end{bmatrix}$$

于是

$$\text{cond}(\boldsymbol{A})_1 = 70 \times 0.3 = 21$$

$$\text{cond}(\boldsymbol{A})_\infty = 60 \times 0.35 = 21$$

2.7 雅可比迭代

雅可比迭代（Jacobi Iteration）是用迭代法解方程组. 不用讲解，直接上题目就行.

例2.8【PPT】用雅可比迭代法求解方程组

$$\begin{cases} 3x_1 + x_2 = 5 \\ x_1 + 2x_2 = 5 \end{cases}$$

解 第一步，把每个方程改写成 $\begin{cases} x_1 = \ldots \\ x_2 = \ldots \end{cases}$ 的形式，即

$$\begin{cases} x_1 = -\dfrac{1}{3}x_2 + \dfrac{5}{3} \\ x_2 = -\dfrac{1}{2}x_1 + \dfrac{5}{2} \end{cases}$$

第二步，等号左边的都是$(k+1)$项，等号右边的都是(k)项，即

$$\begin{cases} x_1^{(k+1)} = -\dfrac{1}{3}x_2^{(k)} + \dfrac{5}{3} \\ x_2^{(k+1)} = -\dfrac{1}{2}x_1^{(k)} + \dfrac{5}{2} \end{cases}$$

第三步，取一个初始值，开始迭代，这里取 $\begin{cases} x_1^{(0)} = 0 \\ x_2^{(0)} = 0 \end{cases}$ ，于是

$$\begin{cases} x_1^{(1)} = \dfrac{5}{3} = 1.6667 \\ x_2^{(1)} = \dfrac{5}{2} = 2.5 \end{cases}$$

继续计算，得

$$\begin{cases} x_1^{(2)} = -\dfrac{1}{3} \times 2.5 + \dfrac{5}{3} = 0.8333 \\ x_2^{(2)} = -\dfrac{1}{2} \times 1.6667 + \dfrac{5}{2} = 1.6667 \end{cases}$$

如此往复，当$k=10$时，有

$$\begin{cases} x_1^{(10)}=0.9998 \\ x_2^{(10)}=1.9997 \end{cases}$$

分析可得，这个方程组的理论解为

$$\begin{cases} x_1=1 \\ x_2=2 \end{cases}$$

随着迭代过程的不断进行，$||\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}||$会不断减小，直到满足要求后可以收手.

2.8 高斯-赛德尔迭代

雅可比迭代的效率其实是不够高的，因为最新的信息没有被使用. 作为改进，有了高斯-赛德尔迭代（Guass-Seidel Iteration）.

例2.9 用高斯-赛德尔迭代法重新求解例2.8.

解 第一步与雅可比迭代法相同. 第二步将方程组改写为

$$\begin{cases} x_1^{(k+1)}=-\dfrac{1}{3}x_2^{(k)}+\dfrac{5}{3} \\ x_2^{(k+1)}=-\dfrac{1}{2}x_1^{(k+1)}+\dfrac{5}{2} \end{cases}$$

为什么要这样做？因为上面这个方程组通过第一个式子已经求出了$x_1^{(k+1)}$的值，因此第二个式子不必再用旧的$x_1^{(k)}$，而是可以用新的$x_1^{(k+1)}$. 这样可以提高迭代效率.

后续做法相同. 第三步，取 $\begin{cases} x_1^{(0)}=0 \\ x_2^{(0)}=0 \end{cases}$ 作为初始值，于是

$$\begin{cases} x_1^{(1)}=\dfrac{5}{3}=1.6667 \\ x_2^{(1)}=-\dfrac{1}{2}\times 1.6667+\dfrac{5}{2}=1.6667 \end{cases}$$

继续计算直到满足要求即可，后续过程略.

例2.10【23-24夏试题】分别用雅可比迭代法和高斯-赛德尔迭代法求解方程组

$$\begin{cases} 2x_1 - 2x_2 = 0 \\ x_1 - 3x_2 = -1 \end{cases}$$

要求$||\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}||_\infty < 0.01$.

解 (1)雅可比迭代法，先写出方程

$$\begin{cases} x_1^{(k+1)} = x_2^{(k)} \\ x_2^{(k+1)} = \dfrac{1}{3}x_1^{(k)} + \dfrac{1}{3} \end{cases}$$

再列出迭代表

k	0	1	2	3	...	8	9
x_1	0	0	0.3333	0.3333	...	0.4938	0.4938
x_2	0	0.3333	0.3333	0.4444	...	0.4938	0.4979

此时

$$||\boldsymbol{x}^{(9)} - \boldsymbol{x}^{(8)}||_\infty = 0.4979 - 0.4938 = 0.0041 < 0.01$$

满足要求，因此取方程组的解为

$$\begin{cases} x_1 = 0.4938 \\ x_2 = 0.4979 \end{cases}$$

(2)高斯-赛德尔迭代法，先写出方程

$$\begin{cases} x_1^{(k+1)} = x_2^{(k)} \\ x_2^{(k+1)} = \dfrac{1}{3}x_1^{(k+1)} + \dfrac{1}{3} \end{cases}$$

再列出迭代表

k	0	1	2	3	4	5	6
x_1	0	0	0.3333	0.4444	0.4815	0.4938	0.4979
x_2	0	0.3333	0.4444	0.4815	0.4938	0.4979	0.4993

此时

$$||\boldsymbol{x}^{(6)} - \boldsymbol{x}^{(5)}||_\infty = 0.4979 - 0.4938 = 0.0041 < 0.01$$

满足要求，因此取方程组的解为

$$\begin{cases} x_1 = 0.4979 \\ x_2 = 0.4993 \end{cases}$$

(3)显然高斯迭代法收敛更快. 进一步分析可得，方程组的理论解为

$$\begin{cases} x_1 = 0.5 \\ x_2 = 0.5 \end{cases}$$

3 插值

【考纲要求】拉格朗日插值、牛顿插值

3.1 插值概述

所谓插值（Interpolation），就是我给你几个点(x_i, y_i)，你要给出一条曲线$y = f(x)$，要求曲线必须经过这些点. 比如说，给你两个点，让你用直线插值，那就是求过这两点的直线方程. 再比如说，给你三个点，让你用抛物线插值，那就是求过这三点的抛物线方程.

插值与拟合的区别：插值要求曲线必须经过这些点，而拟合不要求经过这些点，只要点与曲线之间的误差最小即可.

3.2 拉格朗日插值（全阶多项式插值）

拉格朗日插值多项式为

$$p(x) = \sum_k \left(\prod_{j \neq k} \frac{x - x_j}{x_k - x_j} \right) y_k$$

等号右边只有x是变量，其他都是常量. 举个例子，假设给了四个点$(x_i, y_i)(i = 0, 1, 2, 3)$，那么拉格朗日插值多项式为

$$\begin{aligned} p(x) = & \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} y_0 + \frac{(x - x_0)(x - x_2)(x - x_3)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)} y_1 \\ & + \frac{(x - x_0)(x - x_1)(x - x_3)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)} y_2 + \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} y_3 \end{aligned}$$

化简后就是一个多项式（这里$n = 3$）

$$p(x) = a_0 + a_1 x + ... + a_n x^n$$

可见，$(n + 1)$个点进行拉格朗日插值得到的是n阶多项式（多项式的最高次数为n次）. 若$n = 1$，则为**线性插值**（两点确定一条直线）；若$n = 2$，则为**抛物线插值**（三点确定一条抛物线）.

例3.1【PPT】已知

x	0	1	2	3
y	-5	-6	-1	16

求拉格朗日插值多项式.

解 将值代入上面的式子即可，具体过程略，化简后的结果为$p(x) = x^3 - 2x - 5$.

3.3 用插值来近似函数

有时需要用近似函数来求值，此时就需要用到插值. 比如说，告诉你$e^{0.1}$和$e^{0.2}$的值，要你求$e^{0.18}$的近似值，你怎么求？你自然会想到把$(0.1, e^{0.1})$和$(0.2, e^{0.2})$这两点用直线$y = f(x)$连起来，然后求$f(0.18)$. 于是，这条直线$y = f(x)$就成为了函数$y = e^x$在**插值区间**$[0.1, 0.2]$上的**插值函数**.

3.4 插值余项

插值函数作为近似函数，与原函数之间毕竟有些差距，而这个差距就是**插值余项**. 若原函数为$f(x)$，插值函数为$p_n(x) = a_0 + a_1x + ... + a_nx^n$，插值区间为$[a, b]$，则插值余项

$$R_n(x) = f(x) - p_n(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}\prod_{i=0}^{n}(x - x_i)$$

其中，$\zeta \in (a, b)$且依赖于x（类似拉格朗日中值定理），$f^{(n+1)}(\zeta)$表示该点的$(n+1)$阶导数.

例3.2【PPT】给定函数表如下：

x	...	0.1	0.2	0.3	0.4	0.5	...
e^x	...	1.1052	1.2214	1.3499	1.4918	1.6487	...

试用线性插值与抛物线插值求$e^{0.285}$的近似值，并估计截断误差.

解 插值点要尽可能接近$x = 0.285$，因此线性插值取$x = 0.2, x = 0.3$两点，抛物线插值取$x = 0.2, x = 0.3, x = 0.4$三点.

线性插值函数

$$L_1(x) = 1.2214 \times \frac{x - 0.3}{0.2 - 0.3} + 1.3499 \times \frac{x - 0.2}{0.3 - 0.2}$$

因此得到近似值

$$e^{0.285} \approx L_1(0.285) \approx 1.3306$$

插值余项

$$R_1(x) = \frac{f''(\zeta)}{2!}(x - x_0)(x - x_1) = \frac{e^{\zeta}}{2}(x - 0.2)(x - 0.3)$$

其中$\zeta \in (0.2, 0.3)$，故截断误差

$$|R_1(0.285)| \leq \frac{e^{0.3}}{2}|(0.285 - 0.2)(0.285 - 0.3)| < 0.0009$$

抛物线插值同理，过程略，结果为

$$e^{0.285} \approx L_2(0.285) \approx 1.3298$$

截断误差

$$|R_2(0.285)| \le \frac{\mathrm{e}^{0.4}}{6}|(0.285-0.2)(0.285-0.3)(0.285-0.4)| < 0.00004$$

参考：$\mathrm{e}^{0.285}$的准确值为1.329762...

例3.3【23-24夏试题】对$(0,1),(1,2),(2,3)$三点使用拉格朗日二次插值，写出计算结果和最终表达式.

解 直接套公式

$$\begin{aligned} p(x) &= \frac{(x-1)(x-2)}{(0-1)(0-2)}\times 1 + \frac{(x-0)(x-2)}{(1-0)(1-2)}\times 2 + \frac{(x-0)(x-1)}{(2-0)(2-1)}\times 3 \\ &= \frac{1}{2}\left(x^2-3x+2\right) - 2\left(x^2-2x\right) + \frac{3}{2}\left(x^2-x\right) \\ &= \left(\frac{1}{2}-2+\frac{3}{2}\right)x^2 + \left(-\frac{3}{2}+4-\frac{3}{2}\right)x + 1 \\ &= x+1 \end{aligned}$$

解后反思 二次项没了，因为这三个点在一条直线上. 其实这个直线方程一眼就能看出来.

3.5 差商

接下来的牛顿插值会用到差商，因此有必要在这里先讲解差商的概念. 差商是导数的近似，用于描述函数的局部变化率. 对函数$f(x)$，定义**一阶差商**为

$$f[x_i,x_j] = \frac{f(x_j)-f(x_i)}{x_j-x_i}$$

二阶差商为（注意分母）

$$f[x_i,x_j,x_k] = \frac{f[x_j,x_k]-f[x_i,x_j]}{x_k-x_i}$$

n**阶差商**为（注意分母）

$$f[x_{i_0},x_{i_1},...,x_{i_n}] = \frac{f[x_{i_1},...,x_{i_n}]-f[x_{i_0},...,x_{i_{n-1}}]}{x_{i_n}-x_{i_0}}$$

特别地，规定**零阶差商**为

$$f[x_i] = f(x_i)$$

不难发现，在一阶差商中，若取$(x_j-x_i)\to 0$，则这个极限就是一阶导数.

3.6 牛顿插值

拉格朗日插值多项式的形式具有对称性，既便于记忆，又便于应用与编写程序. 但是，由于公式依赖于全部插值节点，导致在增加或删除节点时，必须全部重新计算. 为了克服这个缺点，牛顿插值法应运而生.

n次牛顿插值多项式为

$$\begin{aligned} N_n(x) = a_0 &+ a_1(x-x_0) \\ &+ a_2(x-x_0)(x-x_1) + ... \\ &+ a_n(x-x_0)(x-x_1)...(x-x_{n-1}) \end{aligned}$$

其中，各项系数为

$$\begin{aligned} a_0 &= f[x_0] \\ a_1 &= f[x_0,x_1] \\ a_2 &= f[x_0,x_1,x_2] \\ &... \\ a_n &= f[x_0,x_1,...,x_n] \end{aligned}$$

与拉格朗日插值相同，$(n+1)$个点（$x_0,x_1,...,x_n$）进行牛顿插值得到的是n阶多项式. 虽然牛顿插值多项式里没有显含x_n，但是最后一项的系数a_n与x_n有关.

3.7 牛顿插值法的插值余项

本章3.4节给出的公式依旧有效

$$R_n(x) = f(x) - N_n(x) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}\prod_{i=0}^{n}(x-x_i)$$

在实际计算中，特别是在$f(x)$的高阶导数比较复杂或$f(x)$的表达式没有给出时，常用差商表示余项

$$R_n(x) = f(x) - N_n(x) = f[x_0,x_1,...,x_n,x]\prod_{i=0}^{n}(x-x_i)$$

3.8 差商表

构造差商表是好习惯. 以三次牛顿插值为例，假设有x_0, x_1, x_2, x_3四个点，则差商表可以构建如下：

	$f(x)$	一阶差商	二阶差商	三阶差商
x_0	$f(x_0)$			
x_1	$f(x_1)$	$f[x_0, x_1]$		
x_2	$f(x_2)$	$f[x_1, x_2]$	$f[x_0, x_1, x_2]$	
x_3	$f(x_3)$	$f[x_2, x_3]$	$f[x_1, x_2, x_3]$	$f[x_0, x_1, x_2, x_3]$

其中，对角线上的元素就是牛顿插值法的系数. 比方说想要计算a_3，那就是

$$a_3 = f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$$

注意分母是$(x_3 - x_0)$而不是$(x_3 - x_2)$，这个挺容易弄错的.

例3.4【PPT】给定函数表如下：

x	1	2	4	6	7
$f(x)$	4	1	0	1	1

求四次牛顿插值多项式，并写出插值余项.

解 构造差商表（表中标注颜色的数据是我自己计算的结果，与ppt上的不同，我觉得自己应该算的没问题，大家也可以验算一下）

		$f(x)$	一阶差商	二阶差商	三阶差商	四阶差商
x_0	1	4				
x_1	2	1	-3			
x_2	4	0	$-\frac{1}{2}$	$\frac{5}{6}$		
x_3	6	1	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{7}{60}$	
x_4	7	1	0	$-\frac{1}{6}$	$-\frac{1}{12}$	$\frac{1}{180}$

于是，四次牛顿插值多项式为（懒得化简了）

$$\begin{aligned}N_4(x) = 4 - 3(x-1) \\ &+ \frac{5}{6}(x-1)(x-2) \\ &- \frac{7}{60}(x-1)(x-2)(x-4) \\ &+ \frac{1}{180}(x-1)(x-2)(x-4)(x-6)\end{aligned}$$

牛顿插值余项为

$$R_4(x) = f(x) - N_4(x) = \frac{f^{(5)}(\zeta)}{5!}(x-1)(x-2)(x-4)(x-6)(x-7),$$
$$\zeta \in (\min\{x, 1\}, \max\{x, 7\})$$

当然，也可以用差分形式写

$$R_4(x) = f(x) - N_4(x) = f[1, 2, 4, 6, 7, x](x-1)(x-2)(x-4)(x-6)(x-7)$$

3.9 差分

在实际计算中，由于便利性，常常遇到等距节点的情况，即

$$x_k = x_0 + kh \quad (k = 0, 1, 2, ...)$$

其中，h叫做**步长**（$h > 0$）.

借助差分，可以简化牛顿插值多项式，因此有必要讲解一下差分的概念.

差分有**向前差分**和**向后差分**两种，其中向前差分可以简称为**差分**（向后差分不能简称）.
定义**一阶差分**为

$$\Delta y_k = y_{k+1} - y_k$$

二阶差分为

$$\Delta^2 y_k = \Delta(\Delta y_k) = \Delta y_{k+1} - \Delta y_k$$

n**阶差分**为

$$\Delta^n y_k = \Delta(\Delta^{n-1} y_k) = \Delta^{n-1} y_{k+1} - \Delta^{n-1} y_k$$

类似地，我们定义**一阶向后差分**为

$$\nabla y_k = y_k - y_{k-1}$$

n**阶向后差分**为

$$\nabla^n y_k = \nabla(\nabla^{n-1} y_k) = \nabla^{n-1} y_k - \nabla^{n-1} y_{k-1}$$

3.10 等距节点的牛顿插值法

在等距节点的场合，差分与差商满足

$$f[x_0, x_1, ..., x_n] = \frac{\Delta^n y_0}{n!\ h^n}$$

令$x = x_0 + th$（此时$t > 0$），得**前插公式**

$$N_n(x_0 + th) = y_0 + t\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 + ... + \frac{t(t-1)...(t-n+1)}{n!}\Delta^n y_0$$

其插值余项为

$$R_n(x_0 + th) = \frac{t(t-1)...(t-n)}{(n+1)!}h^{n+1}f^{(n+1)}(\zeta), \quad \zeta \in (x_0, x_n)$$

同样的，向后差分与差商满足

$$f[x_n, x_{n-1}, ..., x_{n-m}] = \frac{\nabla^m y_n}{m!\ h^m}$$

令$x = x_n + th$（此时$t < 0$），得**后插公式**

$$N_n(x_n + th) = y_n + t\nabla y_n + \frac{t(t+1)}{2!}\nabla^2 y_n + ... + \frac{t(t+1)...(t+n-1)}{n!}\nabla^n y_n$$

其插值余项为

$$R_n(x_n + th) = \frac{t(t+1)...(t+n)}{(n+1)!}h^{n+1}f^{(n+1)}(\zeta), \quad \zeta \in (x_0, x_n)$$

3.11 差分表

与差商表一样，为了更加清晰明确，我们可以构建差分表：

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	
x_0	y_0					**1**
		Δy_0				**t**
x_1	y_1		$\Delta^2 y_0$			$\frac{1}{2!}t(t-1)$
		Δy_1		$\Delta^3 y_0$		
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$	$\frac{1}{3!}t(t-1)(t-2)$
		Δy_2		$\Delta^3 y_1$		
x_3	y_3		$\Delta^2 y_2$			
		Δy_3				$\frac{1}{4!}t(t-1)(t-2)(t-3)$
x_4	y_4					

如果是后插的场合，我们可以列出向后差分表：（感谢嘎嘎的勘误）

x_i	y_i	∇y_i	$\nabla^2 y_i$	$\nabla^3 y_i$	$\nabla^4 y_i$
x_0	y_0				
		∇y_1			
x_1	y_1		$\nabla^2 y_2$		
x_2	y_2	∇y_2	$\nabla^2 y_3$	$\nabla^3 y_3$	
x_3	y_3	∇y_3			$\nabla^4 y_4$
			$\nabla^2 y_4$	$\nabla^3 y_4$	
x_4	y_4	∇y_4			
	1	**t**	$\frac{1}{2!}t(t+1)$	$\frac{1}{3!}t(t+1)(t+2)$	$\frac{1}{4!}t(t+1)(t+2)(t+3)$

在等距节点的场合使用差分可以避开分母的易错点.

例3.5【PPT】给定等距节点的函数表如下：

i	0	1	2	3
x_i	0.4	0.6	0.8	1.0
y_i	1.5	1.8	2.2	2.8

试求$N_3(0.5)$和$N_3(0.9)$的值.

解 为了演示，本题计算$N_3(0.5)$用前插，$N_3(0.9)$用后插. 构造向前差分表

i	x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$
0	**0.4**	**1.5**			
			0.3		
1	**0.6**	**1.8**		**0.1**	
			0.4		**0.1**
2	**0.8**	**2.2**		**0.2**	
			0.6		
3	**1.0**	**2.8**			

于是，前插结果为

$$N_3(0.4+0.2t) = 1.5 + t \times 0.3 + \frac{t(t-1)}{2!} \times 0.1 + \frac{t(t-1)(t-2)}{3!} \times 0.1$$

代入$t = 0.5$，得

$$N_3(0.5) = 1.64375$$

后插同理，结果为

$$N_3(1.0 + 0.2t) = 2.8 + t \times 0.6 + \frac{t(t+1)}{2!} \times 0.2 + \frac{t(t+1)(t+2)}{3!} \times 0.1$$

代入$t = -0.5$，得

$$N_3(0.9) = 2.46875$$

解后反思

(1) 解题过程只构造了向前差分表，没有构造向后差分表，为什么？

事实上，向后差分表与向前差分表是一样的，只不过前插用了表格上面的数据（1.5, 0.3, 0.1, 0.1），后插用了表格下面的数据（2.8, 0.6, 0.2, 0.1）.

(2) 如果用前插算$N_3(0.9)$，用后插算$N_3(0.5)$，结果一样吗？

经过计算，是一样的. 所以，对于前插和后插，我觉得是随便选的，爱用哪个就用哪个.

例3.6【23-24夏试题】对$(0,0),(1,5),(2,27)$三点使用牛顿插值，写出结果.

解 列出差商表

		$f(x)$	一阶差商	二阶差商
x_0	0	0		
x_1	1	5	5	
x_2	2	27	22	8.5

于是，牛顿二次插值多项式为

$$\begin{aligned} N_2(x) = 0 + 5(x-0) \\ + 8.5(x-0)(x-1) \end{aligned}$$

化简得

$$N_2(x) = 8.5x^2 - 3.5x$$

4 曲线拟合

【考纲要求】最小二乘法

4.1 最小二乘直线拟合

新高考考区的同学在高中阶段就学过. 假设有n个点，那么最小二乘直线拟合的方程

$$y = Ax + B$$

其中，斜率为

$$A = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sum x_i^2 - n\overline{x}^2}$$

截距为

$$B = \overline{y} - A\overline{x}$$

PPT上的做法是构建**法方程组**

$$\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

把这个方程组解出来就是上面写的A, B，所以我建议你直接用上面写出来的解.

4.2 最小二乘算法的一般形式

假设你要拟合的函数是$y = f(x)$，最小二乘算法的原则是残差平方和最小，因此要让

$$S = \sum \left(f(x_i) - y_i\right)^2$$

取得最小值.

举个例子，假设你要用双曲线来拟合某组数据，设想函数

$$f(x) = \frac{1}{ax + b}$$

这里有两个待定的参数a, b，为了求解这两个参数，需要让目标函数

$$S(a, b) = \sum \left(\frac{1}{ax_i + b} - y_i \right)^2$$

取得最小值. 这是一个二元函数，想取最小值需要让每个参数的偏导为0，即

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases}$$

这就是一般情形下的**法方程组**，由此可以解出a, b.

例4.1【22-23夏试题改编】给定四个点：

x	0	1	2	3
y	0.9	2.1	5.2	9.8

用曲线$y = a + bx^2$拟合，写出结果.

解 目标函数

$$S(a,b) = \sum_{i=1}^{n} (f(x_i) - y_i)^2 = \sum_{i=1}^{n} (a + bx_i^2 - y_i)^2$$

求偏导，列方程

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases} \Longrightarrow \begin{cases} \sum (a + bx_i^2 - y_i) = 0 \\ \sum (a + bx_i^2 - y_i)x_i^2 = 0 \end{cases} \Longrightarrow \begin{cases} an + b\sum x_i^2 = \sum y_i \\ a\sum x_i^2 + b\sum x_i^4 = \sum x_i^2 y_i \end{cases}$$

其中$n = 4$，其他数据通过计算得到

$$\sum x_i^2 = 14, \quad \sum x_i^4 = 98, \quad \sum y_i = 18, \quad \sum x_i^2 y_i = 111.1$$

代入数据得

$$\begin{cases} 4a + 14b = 18 \\ 14a + 98b = 111.1 \end{cases}$$

解得

$$a \approx 1.064, \quad b \approx 0.982$$

所以拟合的曲线为

$$y = 1.064 + 0.982x^2$$

4.3 *最小二乘幂函数拟合

假设有n个点，最小二乘幂函数的拟合曲线为

$$y = Ax^M$$

其中M是预先给定的，则系数

$$A = \frac{\sum x_i^M y}{\sum x_i^{2M}}$$

这个系数同样可以通过4.2节所述的一般情形推得，没什么意思.

4.4 最小二乘指数拟合

假设有n个点，最小二乘指数函数的拟合曲线为

$$y = C\mathrm{e}^{Ax}$$

这里有两个参数C, A. 为了求解这两个参数，对方程两边取对数

$$\ln y = Ax + \ln C$$

利用换元法，令

$$Y = \ln y, \quad X = x, \quad B = \ln C$$

则方程转化为

$$Y = AX + B$$

这一过程称为**数据线性化**. 接下来只要按最小二乘直线拟合做就行了.

例4.2【PPT】【23-24夏试题改编】给定五个点：

x	0	1	2	3	4
y	1.5	2.5	3.5	5.0	7.5

用曲线$y = C\mathrm{e}^{Ax}$拟合，写出结果.

解 数据线性化，得

X	x	0	1	2	3	4
Y	$\ln y$	0.40547	0.91629	1.25276	1.60944	2.01490

计算各项数据的结果为

$$\overline{X} = 2, \quad \overline{Y} = 1.23977, \quad \sum X_i Y_i = 16.30973, \quad \sum X_i^2 = 30$$

由4.1节给出的斜率和截距计算公式，得

$$A = 0.39120, \quad B = 0.45737, \quad C = \mathrm{e}^B = 1.57991$$

所以数据线性化的结果为

$$Y = 0.39120X + 0.45737$$

指数拟合曲线为

$$y = 1.57991\mathrm{e}^{0.39120x}$$

5 非线性方程求根

【考纲要求】二分法、牛顿法

5.1 二分法

求方程的数值解的基本方法：第一步，找到有根区间；第二步，不断接近准确值.

二分法的基本思路：用对分区间的方法，保留有根区间，舍去无根区间，并且如此不断地对分下去，以逐步逼近方程的根.

二分法的理论依据是**零点存在定理**：若函数$f(x)$在闭区间$[a,b]$上连续，且满足$f(a)f(b) < 0$，则开区间(a,b)内至少存在一个点x_0，使得$f(x_0) = 0$.

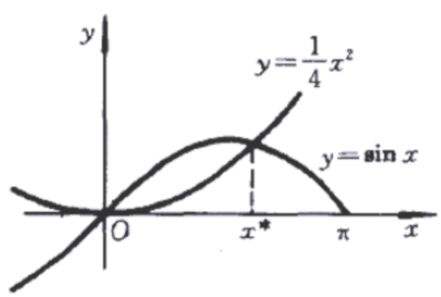
直接来道题就清楚了.

例5.1【PPT】用二分法求方程

$$f(x) = \sin x - \frac{x^2}{4} = 0$$

的非零实根的近似值，使误差不超过10^{-2}.

解 画出大致图像



试算，得

$$f(1.5) \approx 0.435 > 0, \quad f(2) \approx -0.091 < 0$$

于是取$a_0 = 1.5, b_0 = 2$，即初始区间为$(1.5, 2)$. 取区间中点$x_0 = (a_0 + b_0)/2 = 1.75$，计算得

$$f(x_0) = f(1.75) \approx 0.218 > 0$$

因此$a_1 = 1.75, b_1 = 2$，即下一个区间为$(1.75, 2)$. 再次取区间中点$x_1 = (a_1 + b_1)/2 = 1.875$，计算得

$$f(x_1) = f(1.875) \approx 0.075 > 0$$

所以$a_2 = 1.875, b_2 = 2$. 还是取区间中点$x_2 = (a_2 + b_2)/2 = 1.9375$，计算得

$$f(x_2) = f(1.9375) \approx -0.00496 < 0$$

所以$a_3 = 1.875, b_3 = 1.9375$. 接下来只要继续取区间中点$x_3 = (a_3 + b_3)/2$，计算其函数值$f(x_3)$的正负，然后替换$a_4$或$b_4$就行.

经过多次迭代后，区间缩小为$a_5 = 1.921875, b_5 = 1.9375$. 此时的区间中点$x_5 = (a_5 + b_5)/2 = 1.9296875$. 因为解的准确值一定在区间$(a_5, b_5)$内，所以此时的误差

$$e = |x_5 - x_{准确}| < |x_5 - a_5| = 0.0078125 < 10^{-2}$$

其精度满足题目要求，因此可将$x_5 = 1.9296875$作为方程$f(x) = 0$的非零实根的近似值.

注：迭代过程可以列个表，这样更加清晰明确.

k	a_k	b_k	x_k	$f(x_k)$
0	1.5	2	1.75	0.218361
1	1.75	2	1.875	0.0751795
2	1.875	2	1.9375	−0.00496228
3	1.875	1.9375	1.90265	0.0404208
4	1.90625	1.9375	1.921875	0.156014
5	1.921875	1.9375	1.9296875	0.00536340

解后反思 不妨判断一下有效位数，进一步分析误差

$$e = |x_5 - x_{准确}| < |x_5 - a_5| = 0.0078125 < 0.5 \times 10^{-1}$$

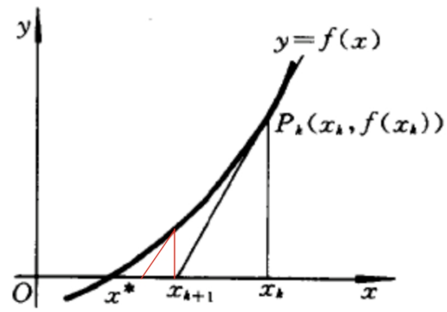
因此取10^{-1}，即上面求得的解$x_5 = 1.9296875$只有2位有效数字，保留有效数字的结果为1.9.

5.2 牛顿法

二分法收敛比较慢，所以我们需要更快的牛顿法. 牛顿法的迭代式

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

其几何意义也比较明确，就是借助了切线，如下图所示：



例5.2 用牛顿法求解例5.1，要求$|x_{k+1}-x_k|<10^{-5}$.

解 计算导数

$$f'(x)=\cos x-\frac{x}{2}$$

因此

$$x_{k+1}=x_k-\frac{\sin x_k-x_k^2/4}{\cos x_k-x_k/2}$$

取初始值$x_0=2$，则

$$x_1=x_0-\frac{\sin x_0-x_0^2/4}{\cos x_0-x_0/2}=2-\frac{\sin 2-1}{\cos 2-1}\approx 1.93595$$

继续迭代

$$x_2\approx 1.9337564$$

$$x_3\approx 1.9337538$$

此时

$$|x_3-x_2|=0.26\times 10^{-5}<10^{-5}$$

符合题目所给条件，因此取$x^*=x_3=1.9337538$.

解后反思

(1) 我们也来分析一下它的有效数字，其误差

$$e=|x_3-x_{准确}|<|x_3-x_2|=0.26\times 10^{-5}<0.5\times 10^{-5}$$

因此取10^{-5}，即$x_3=$1.9337538有6位有效数字，保留有效数字的结果为1.93375.

也许你会质疑，$|x_3-x_{准确}|<|x_3-x_2|$真的成立吗？我不知道，那个计算误差的式子真的很复杂，就当它成立吧，反正都收敛了，应该跑不掉的.

(2) 同一道题，二分法迭代到x_5才满足10^{-2}的精度，而牛顿法迭代到x_3就满足了10^{-5}的精度. 一目了然，不言而喻. 顺带一提，牛顿法是平方收敛，即每迭代一次，新的误差在量级上约等于前一次误差的平方，所以收敛速度非常快.

例5.3【23-24夏试题】给定方程

$$f(x) = 5x^3 + 10x - 1$$

(1) 给定初始区间$[0,1]$，用二分法计算方程的数值解，给出结果x和$f(x)$，保留4位有效数字.

(2) 给定初始值$x_0 = 1$，用牛顿法计算方程的数值解，给出结果x和$f(x)$，保留4位有效数字.

解 (1) 这题出的诗人握持，我现在只寄希望于回忆有误. 试算一下，显然有$f(0.1) > 0$，所以它的有根区间是$[0, 0.1]$. 也就是说，它的第一位非零数字在10^{-2}上，保留4位有效数字就意味着绝对误差限要小于0.5×10^{-5}，即

$$\frac{0.5}{2^n} < 0.5 \times 10^{-5}$$

解得$n > 5/\lg 2 \approx 16.6$，即$n \geq 17$. 也就是说，这道题要整整迭代17次！不做了，随缘吧.

(2) 写出$f(x)$并求导

$$f(x) = 5x^3 + 10x - 1$$
$$f'(x) = 15x^2 + 10$$

一步一步迭代就行，计算过程略

$$x_0 = 1$$
$$x_1 = 0.44$$
$$x_2 \approx 0.1435090$$
$$x_3 \approx 0.0998703$$
$$x_4 \approx 0.0995074$$
$$x_5 \approx 0.0995074$$

此时的误差

$$e < |x_5 - x_4| = 0 < 0.5 \times 10^{-5}$$

满足要求，保留4位有效数字的结果为$x^* = 0.09951$，$f(x^*) = 2.686 \times 10^{-5}$.

5.3 *牛顿迭代法的收敛性判断

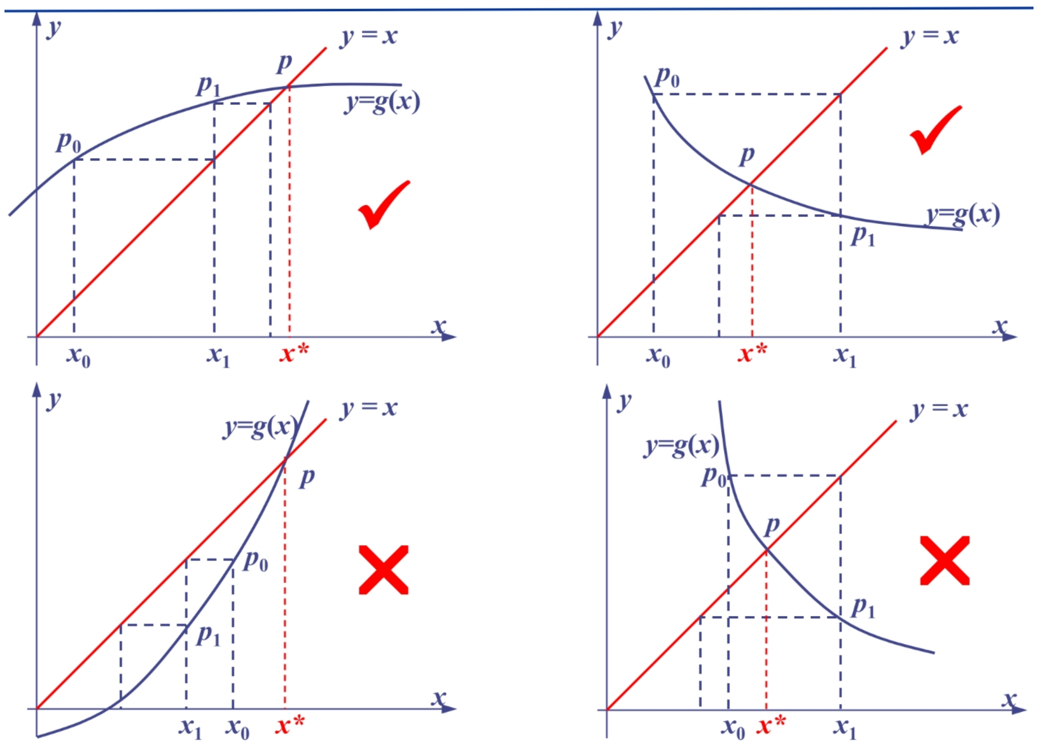
牛顿迭代法虽然收敛很快，但是它却不一定收敛. 为了分析其敛散性，令迭代函数

$$g(x) = x - \frac{f(x)}{f'(x)}$$

则牛顿法迭代可以表示为

$$x_{k+1} = g(x_k)$$

如果高中对数列的研究比较深入，那么一定会接触到**不动点**和**蛛网图**，用于分析数列的迭代过程，并可由此判断数列的敛散性.



上图展示了四种敛散性的分析，分别是：一个方向收敛、回字形收敛、一个方向发散、回字形发散. 从图中可以直观看出，牛顿迭代法的收敛条件是

$$|g'(x)| < 1$$

具体的分析不多讲了，我觉得不至于考这种东西，毕竟这门课是数值计算而非数学分析.

5.4 *近似割线法

牛顿法也有其缺陷：如果函数的导数$f'(x)$不易求出，那么用牛顿法求解就会步履维艰. 此时可用割线斜率近似代替切线斜率，即用

$$S_k = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$$

替代原式中的$f'(x_k)$.

6 数值积分

【考纲要求】中点法、梯形法、辛普森法则等

6.1 数值积分概述

我们知道，求定积分可以用牛顿—莱布尼茨公式

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a)$$

其中，$F(x)$是$f(x)$的原函数. 但是，你有没有想过，原函数不是那么好求的，比如这个：

$$\int \frac{1}{1+x^5}\mathrm{d}x = ?$$

这是一道经典的钓鱼题，它的原函数相当复杂，自己去网上找一下答案就知道了. 另外有一些时候，函数$f(x)$干脆不演了，它的原函数压根就不能用初等函数表示. 所以，数值积分应运而生，它的基本准则就是避免求原函数. 下面是一些数值积分常用的方法.

6.2 中点法

$$\int_a^b f(x)\mathrm{d}x \approx f(\frac{a+b}{2}) \cdot (b-a) = f(\frac{a+b}{2}) \cdot h$$

其中，$h = b - a$.

6.3 梯形法

$$\int_a^b f(x)\mathrm{d}x \approx \frac{f(a)+f(b)}{2} \cdot (b-a) = \frac{f(a)+f(b)}{2} \cdot h$$

其中，$h = b - a$.

6.4 辛普森法则

$$\int_a^b f(x)\mathrm{d}x \approx \frac{h}{6} \cdot [f(a) + 4f(c) + f(b)] = S$$

其中，$c = \dfrac{a+b}{2}$，$h = b - a$.

6.5 复合辛普森法则（区间二等分）

$$\int_a^b f(x)\mathrm{d}x \approx \frac{h}{12} \cdot [f(a) + 4f(d) + 2f(c) + 4f(e) + f(b)] = S_2$$

其中，$c = \dfrac{a+b}{2}, d = \dfrac{a+c}{2}, e = \dfrac{c+b}{2}$，$h = b - a$.

6.6 复合法

复合法就是等分区间. 比如上面的复合辛普森法则，它的由来就是把区间$[a, b]$等分成两个区间$[a, c], [c, b]$，分别对两个区间用辛普森法则积分，然后相加，即

$$\begin{aligned}\int_a^b f(x)\mathrm{d}x &\approx \frac{h/2}{6} \cdot [f(a) + 4f(d) + f(c)] + \frac{h/2}{6} \cdot [f(c) + 4f(e) + f(b)] \\ &= \frac{h}{12} \cdot [f(a) + 4f(d) + 2f(c) + 4f(e) + f(b)]\end{aligned}$$

复合法不限于辛普森法则，还可以是复合梯形法等其他方法；它也不限于区间二等分，还可以是区间三等分等其他方法. 所以上面的复合辛普森法则只是区间二等分的特例. 对于其他复合情况，这里以复合梯形法则为例，将区间等分为

$$[a, x_1], [x_1, x_2], ..., [x_{n-1}, b]$$

共n个区间，令h为区间长度，即

$$h = x_k - x_{k-1} = \frac{b-a}{n}$$

则可以得到**复合梯形公式**

$$\begin{aligned}\int_a^b f(x)\mathrm{d}x &\approx \frac{h}{2} \cdot [f(a) + f(x_1)] + \frac{h}{2} \cdot [f(x_1) + f(x_2)] + ... + \frac{h}{2} \cdot [f(x_{n-1}) + f(b)] \\ &= \frac{h}{2} \cdot \left[f(a) + 2\sum_{k=1}^{n-1} f(x_k) + f(b)\right]\end{aligned}$$

这个式子没必要记，因为很好理解，考试让你求个区间三等分已经差不多了，不会刁难人的. 要是A4纸空间还很多的话，那你把它抄上去也无妨.

例6.1【23-24夏试题】待计算积分为

$$\int_0^3 \frac{x}{2+x^2}\mathrm{d}x$$

(1) 用梯形法和中点法计算近似值.

(2) 用复合梯形法和复合辛普森法计算近似值，区间三等分.

解 (1) 梯形法

$$原式 \approx \frac{f(0)+f(3)}{2} \times 3 = \frac{0+\frac{3}{11}}{2} \times 3 \approx 0.409$$

中点法

$$原式 \approx f(1.5) \times 3 = \frac{1.5}{2+1.5^2} \times 3 \approx 1.059$$

(2) 复合梯形法

$$原式 \approx \frac{1}{2}[f(0)+2f(1)+2f(2)+f(3)] = \frac{1}{2}[0+2\times\frac{1}{3}+2\times\frac{2}{6}+\frac{3}{11}] \approx 0.803$$

复合辛普森法

$$原式 \approx \frac{1}{6}[f(0)+4f(0.5)+2f(1)+4f(1.5)+2f(2)+4f(2.5)+f(3)] \approx 0.853$$

6.7 数值积分公式的精度

上面这些近似计算的法则本质上都可以概括成这样一个式子：

$$\int_a^b f(x)\mathrm{d}x \approx \sum_{k=1}^{n} A_k f(x_k)$$

若这个式子对$f(x)=x^m$准确成立，但对$f(x)=x^{m+1}$却不能准确成立，则其代数精度为m.

例6.2 求梯形法则的代数精度.

解 先写出梯形法则的式子

$$\int_a^b f(x)\mathrm{d}x \approx \frac{f(a)+f(b)}{2}\cdot(b-a)$$

然后计算$f(x)=1$的情况

$$左边 = \int_a^b \mathrm{d}x = b-a$$

$$右边 = \frac{1+1}{2}(b-a) = b-a$$

两者相等. 再计算$f(x)=x$的情况

$$左边 = \int_a^b x\mathrm{d}x = \frac{b^2-a^2}{2}$$

$$右边 = \frac{a+b}{2}(b-a) = \frac{b^2-a^2}{2}$$

两者相等. 再计算$f(x)=x^2$的情况

$$\text{左边}=\int_a^b x^2\mathrm{d}x=\frac{b^3-a^3}{3}$$
$$\text{右边}=\frac{a^2+b^2}{2}(b-a)$$

显然两者不相等，所以梯形法则的代数精度为1.

例6.3【PPT】确定求积公式

$$\int_{-1}^{1}f(x)\mathrm{d}x\approx A_{-1}f(-1)+A_0f(0)+A_1f(1)$$

中的系数，使其具有尽可能高的代数精度.

解 分别代入$f(x)=1,f(x)=x,f(x)=x^2$，可以得到三个方程，联立一下再求解就行. 具体过程略，结果为

$$A_{-1}=A_1=\frac{1}{3},A_0=\frac{4}{3}$$

6.8 *牛顿-科斯特(Newton-Cotes)法则

简单来说就是等分区间，然后利用第三章的插值去近似. 我觉得不会考，没什么意思.

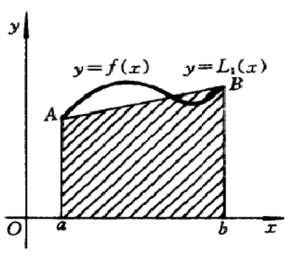
$$\int_a^b f(x)\mathrm{d}x\approx(b-a)\sum_{k=0}^{n}C_k^{(n)}f(x_k)$$

Cotes系数

$$\boxed{C_k^{(n)}=\frac{(-1)^{n-k}}{n\cdot k!\cdot(n-k)!}\int_0^n t(t-1)\cdots(t-k+1)(t-k-1)\cdots(t-n)\mathrm{d}t}$$

特例：$n=1$为梯形法则，$n=2$为辛普森法则.

当 $n=1$ 时有



$$C_0^{(1)}=-\int_0^1(t-1)\mathrm{d}t=\frac{1}{2}$$

$$C_1^{(1)}=\int_0^1 t\,\mathrm{d}t=\frac{1}{2}$$

$$\int_a^b f(x)\mathrm{d}x\approx\frac{b-a}{2}[f(a)+f(b)]$$

7 微分方程数值解

【考纲要求】欧拉法、经典龙格库塔法

7.1 数值微分

与数值积分一样，数值微分的出现是为了解决求导困难的问题. 由于导数的定义

$$\begin{aligned}
f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \qquad \text{（向前差商）}\\
&= \lim_{h\to 0}\frac{f(x)-f(x-h)}{h} \qquad \text{（向后差商）}\\
&= \lim_{h\to 0}\frac{f(x+h)-f(x-h)}{2h} \qquad \text{（中心差商）}
\end{aligned}$$

是取差商的极限，所以我们用差商近似代替导数.

7.2 微分方程数值解的概述

与求解定积分一样，微分方程的解析解不是那么容易就能求出来的（想必各位学机械的很少有学过常微分方程的吧，杂鱼~）. 如果你学过常微分方程，那你就会知道，现有的解法只能针对特定的一类常微分方程进行求解（可分离变量、不显含x或y、积分因子、齐次微分方程等），大多数微分方程是没法求解析解的. 所以，我们也需要用数值方法来求解微分方程.

7.3 常微分方程的一些基本概念

鉴于各位杂鱼应该没学过常微分方程，这里简单说一些概念，以便后续理解. 举个例子，放个非常简单的方程

$$\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{1}{x}$$

它的解法是分离变量（y都放左边，x都放右边）

$$\mathrm{d}y=\frac{1}{x}\mathrm{d}x$$

两边积分

$$\int \mathrm{d}y=\int \frac{1}{x}\mathrm{d}x$$

等号两边是不定积分，求出来

$$y=\ln|x|+C$$

其中，C是任意常数.

上面求出来的这个式子叫做原方程的**通解**. 有时方程会限制**初始条件**，比如我给定一个条件$y(1)=0$，代入通解就能得到$C=0$，于是求得了原方程在该**初值问题**下的**特解**

$$y=\ln|x|$$

7.4 欧拉法

欧拉法就是根据数值微分原理，用差商代替导数. 对于初值问题

$$\begin{cases} \dfrac{\mathrm{d}y}{\mathrm{d}x}=f(x,y), x\in[a,b] \\ y(a)=y_0 \end{cases}$$

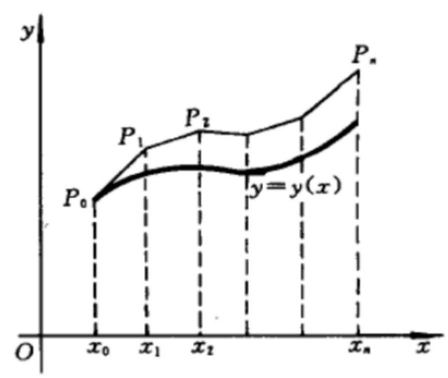
把它等分成n个区间，每个区间的长度$h=\dfrac{b-a}{n}$. 在每个区间$[x_i,x_{i+1}]$内，用差商代替导数

$$\frac{y(x_{i+1})-y(x_i)}{h}\approx f(x_i,y(x_i))$$

于是有

$$y(x_{i+1})\approx y(x_i)+hf(x_i,y(x_i))$$
$$y_{i+1}=y_i+hf(x_i,y_i)$$

为了更加清晰明确，这里放一张图. 上式的$y(x_i)$指的是曲线$y=y(x)$上的值，而y_i指的是折线段上的值，所以知道为什么一个是等于而另一个是约等于了吧. 换句话说，$y(x_i)$是符合解析解的准确值，而y_i是我们求得的数值解/近似值. 欧拉法就是用折线近似代替曲线.



例7.1【PPT】用欧拉法求初值问题

$$\begin{cases} y' = x - y + 1, x \in [0, 0.5] \\ y(0) = 1 \end{cases}$$

的数值解（取$h = 0.1$），并将计算结果与准确解$y(x) = x + \mathrm{e}^{-x}$进行比较.

解 本题$f(x, y) = x - y + 1$，由欧拉公式得数值解的计算公式

$$y_{i+1} = y_i + 0.1 \times x_i - y_i + 1$$

化简得

$$y_{i+1} = 0.1x_i + 0.9y_i + 0.1$$

将计算结果列表如下

x_i	y_i	$y(x_i)$	$\lvert y(x_i) - y_i \rvert$
0.0	1.000000	1.000000	0.000000
0.1	1.000000	1.004837	0.004837
0.2	1.010000	1.018731	0.008731
0.3	1.029000	1.040818	0.011818
0.4	1.056100	1.070320	0.014220
0.5	1.090490	1.106531	0.016041

7.5 截断误差与精度分析

与欧拉法的分析过程一样，设y_i是近似值，$y(x_i)$是准确值. 假定$y_i = y(x_i)$成立，则称

$$R_{i+1} = y(x_{i+1}) - y_{i+1}$$

为该数值方法在x_{i+1}处的**局部截断误差**. 简单来说就是准确值减去近似值.

如果一个数值方法的局部截断误差为$O(h^{p+1})$，则称这个方法的**阶数**为p.

例7.2 求证：欧拉法是一阶精度方法.

解 根据泰勒展开，准确值满足

$$y(x_{i+1}) = y(x_i) + hy'(x_i) + \frac{h^2}{2}y''(\zeta_i) \tag{1}$$

根据微分方程的条件和前提假设

$$y'(x_i) = f(x_i, y(x_i))$$
$$y_i = y(x_i)$$

(1)式可以写成

$$y(x_{i+1}) = y_i + hf(x_i, y_i) + \frac{h^2}{2}y''(\zeta_i) \tag{2}$$

根据欧拉公式，近似值满足

$$y_{i+1} = y_i + hf(x_i, y_i) \tag{3}$$

由(2)(3)式得局部截断误差

$$R_{i+1} = y(x_{i+1}) - y_{i+1} = \frac{h^2}{2}y''(\zeta_i) = O(h^2)$$

故欧拉法阶数为1，即欧拉法是一阶精度方法. 这个精度显然是比较低的，因此我们需要寻找精度更高的方法.

7.6 后退(隐式)欧拉法

前面所讲的欧拉法又叫**前进(显式)欧拉法**，是用向前差商代替导数. 如果用向后差商代替导数，则可以得到

$$\frac{y(x_i) - y(x_{i-1})}{h} \approx f(x_i, y(x_i))$$

于是有

$$y(x_i) \approx y(x_{i-1}) + hf(x_i, y(x_i))$$

为了与前进(显式)欧拉法比较，用$(i+1)$代替i，用i代替$(i-1)$，并改写成数值解的表达式，得

$$y_{i+1} = y_i + hf(x_{i+1}, y_{i+1})$$

上式的等号两边都含有待求的y_{i+1}，是一个隐式而非显式，所以叫做**后退(隐式)欧拉法**.

隐式用起来不方便，所以在实际计算中，一般先用前进欧拉法求出临时的$\overline{y}_{i+1}$，即

$$\overline{y}_{i+1} = y_i + hf(x_i, y_i)$$

然后将其代入后退欧拉法的等号右边，即

$$y_{i+1} = y_i + hf(x_{i+1}, \overline{y}_{i+1})$$

7.7 改进欧拉法

我知道，上面的讲解非常乏味，我自己都不想看. 所以，有必要用一种直观的方式解释前进欧拉法和后退欧拉法.

将微分方程

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x, y)$$

变换为

$$\mathrm{d}y = f(x, y)\mathrm{d}x$$

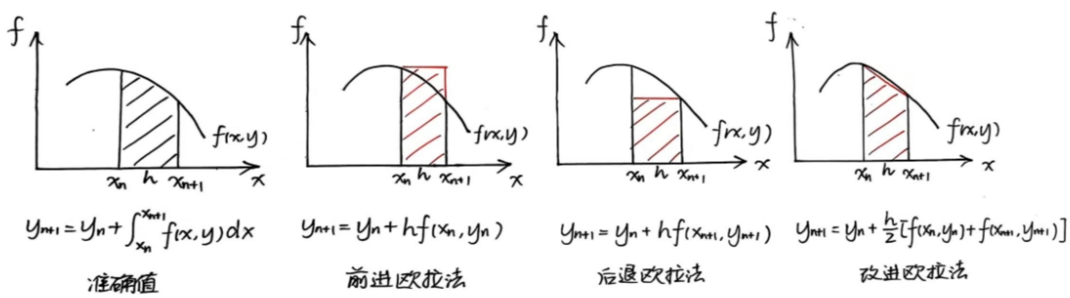
两边积分，得

$$\int_{y(x_1)}^{y(x_2)} \mathrm{d}y = \int_{x_1}^{x_2} f(x, y)\mathrm{d}x$$

于是

$$y(x_2) - y(x_1) = \int_{x_1}^{x_2} f(x, y)\mathrm{d}x$$

下面放一张图，联想数值积分章节的知识，你马上就能理解了.



一目了然，不言而喻. **改进欧拉法**的式子是

$$y_{i+1} = y_i + \frac{h}{2}\left[f(x_i, y_i) + f(x_{i+1}, y_{i+1})\right]$$

不过，它与后退欧拉法一样，会遇到隐式求解困难的问题，因此先用前进欧拉法求出临时的$\overline{y}_{i+1}$（称为**预测**），再用改进欧拉法求得最后的y_{i+1}（称为**矫正**）. 用式子表示为

$$\begin{cases} \overline{y}_{i+1} = y_i + hf(x_i, y_i) \\ y_{i+1} = y_i + \dfrac{h}{2}\left[f(x_i, y_i) + f(x_{i+1}, \overline{y}_{i+1})\right] \end{cases}$$

因此改进欧拉法又叫**预测-矫正法**.

改进欧拉法是二阶精度方法，证明：`https://www.zhihu.com/question/436480848`.

7.8 泰勒展开

精度证明需要用到泰勒展开，因此用大记忆恢复术回想一下. **一元函数的泰勒展开：**

$$f(x+h)=f(x)+f'(x)h+\frac{f''(x)}{2!}h^2+O(h^3)$$

二元函数的泰勒展开：

$$\begin{aligned}f(x+h,y+k)=f(x,y)&+f'_x(x,y)h+f'_y(x,y)k\\&+\frac{1}{2!}\left[f''_{xx}(x,y)h^2+2f''_{xy}(x,y)hk+f''_{yy}(x,y)k^2\right]\\&+O(h^3+k^3)\end{aligned}$$

7.9 龙格-库塔(Runge-Kutta)法的构建思路

根据**微分中值定理**（拉格朗日中值定理）

$$y(x_{i+1})-y(x_i)=y'(\zeta)(x_{i+1}-x_i),\quad \zeta\in[x_i,x_{i+1}]$$

根据微分方程的条件又有

$$y'(x)=\frac{\mathrm{d}y}{\mathrm{d}x}=f(x,y)$$

因此有

$$y(x_{i+1})=y(x_i)+hf(\zeta,y(\zeta))$$

记**平均斜率**（**最优斜率**）为

$$K^*=f(\zeta,y(\zeta))$$

因此有

$$y(x_{i+1})=y(x_i)+hK^*$$

求微分方程的近似解，本质上是用便于计算的方法得到 K^* 的近似值. 如果用左端点 x_i 处的斜率近似代替平均斜率，即令

$$K_1=f(x_i,y_i)$$

则可以得到欧拉公式

$$\begin{cases}y_{i+1}=y_i+hK_1\\K_1=f(x_i,y_i)\end{cases}$$

更进一步，如果使用右端点 x_{i+1} 处的斜率，令

$$K_2=f(x_{i+1},y_{i+1})=f(x_i+h,y_i+hK_1)$$

然后将它们的算术平均值 $(K_1+K_2)/2$ 近似代替平均斜率，则可以得到改进欧拉公式

$$\begin{cases} y_{i+1} = y_i + \dfrac{h}{2}(K_1+K_2) \\ K_1 = f(x_i, y_i) \\ K_2 = f(x_i+h, y_i+hK_1) \end{cases}$$

龙格-库塔法的基本思路：在区间 $[x_i, x_{i+1}]$ 内多预估几个点（通常将左端点 x_i 考虑在内）上的斜率值 $K_1, K_2, ..., K_m$，然后用它们的加权平均值作为平均斜率 K^* 的近似值，即

$$\begin{cases} y_{i+1} = y_i + h(\alpha_1 K_1 + \alpha_2 K_2 + \cdots + \alpha_m K_m) \\ K_1 = f(x_i, y_i) \\ K_2 = f(x_i + \lambda_2 h, y_i + \mu_2 h K_1) \\ \vdots \\ K_m = f(x_i + \lambda_m h, y_i + \mu_m h K_{m-1}) \end{cases}$$

其中， α, λ, μ 为待定参数，应使公式的截断误差的阶数尽量高.

7.10 二阶龙格-库塔法

根据上面的分析，我们可以得到**二阶龙格-库塔公式**为

$$\begin{cases} y_{i+1} = y_i + h(a_1 K_1 + a_2 K_2) \\ K_1 = f(x_i, y_i) \\ K_2 = f(x_i + \lambda_2 h, y_i + \lambda_2 h K_1) \end{cases}$$

若 a_1, a_2, λ_2 的取值得当，则二阶龙格-库塔法是二阶精度方法.

为了分析局部截断误差，将 $y(x_{i+1})$ 在 x_i 处泰勒展开，得

$$y(x_{i+1}) = y(x_i) + y'(x_i)h + \frac{1}{2}y''(x_i)h^2 + O(h^3)$$

注意到 $y'(x_i) = f(x_i, y_i) = K_1$ ， $y_i = y(x_i)$，所以有

$$y(x_{i+1}) = y_i + hK_1 + \frac{1}{2}y''(x_i)h^2 + O(h^3) \tag{4}$$

再将 K_2 在 (x_i, y_i) 处泰勒展开，得

$$\begin{aligned}K_2 = f(x_i, y_i) + (\lambda_2 h) f'_x(x_i, y_i) + (\lambda_2 h K_1) f'_y(x_i, y_i) \\ + \frac{1}{2!}\left[(\lambda_2 h)^2 f''_{xx}(x_i, y_i) + 2(\lambda_2 h)(\lambda_2 h K_1) f''_{xy}(x_i, y_i) + (\lambda_2 h K_1)^2 f''_{yy}(x_i, y_i)\right] \\ + O(h^3)\end{aligned}$$

注意到 $y'(x_i) = f(x_i, y_i) = K_1$ ， $y_i = y(x_i)$，所以有

$$K_2 = K_1 + \lambda_2 h y''(x_i) + \frac{\lambda_2^2 h^2}{2} y'''(x_i) + O(h^3)$$

代入方程 $y_{i+1} = y_i + h(a_1 K_1 + a_2 K_2)$ ，得

$$y_{i+1} = y_i + (a_1 + a_2) h K_1 + (a_2 \lambda_2) y''(x_i) h^2 + O(h^3) \tag{5}$$

由(4)(5)式得局部截断误差

$$R_{i+1} = y(x_{i+1}) - y_{i+1} = [1 - (a_1 + a_2)] \cdot h K_1 + \left[\frac{1}{2} - (a_2 \lambda_2)\right] \cdot y''(x_i) h^2 + O(h^3)$$

因此，当且仅当

$$\begin{cases} a_1 + a_2 = 1 \\ a_2 \lambda_2 = \dfrac{1}{2} \end{cases}$$

成立时，二阶龙格-库塔法是二阶精度方法.

如果取 $a_1 = a_2 = \frac{1}{2}, \quad \lambda_2 = 1$ ，则可以得到改进欧拉公式

$$y_{i+1} = y_i + \frac{h}{2}[f(x_i, y_i) + f(x_{i+1}, y_i + h f(x_i, y_i))]$$

如果取 $a_1 = 0, \quad a_2 = 1, \quad \lambda_2 = \frac{1}{2}$ ，则可以得到中点公式

$$y_{i+1} = y_i + h f\left[x_i + \frac{1}{2}h, y_i + \frac{h}{2} f(x_i, y_i)\right]$$

所以改进欧拉公式是二阶龙格-库塔法的一个特例，证明二阶龙格-库塔法的精度也就证明了改进欧拉公式的精度.

例7.3【23-24夏试题】对于微分方程

$$\begin{cases} y' = f(x,y), x \in [a,b] \\ y(0) = a \end{cases}$$

给出两种具有二阶精度的求解公式，并证明其方法精度.

解 改进欧拉法、中点法、二阶龙格-库塔法都是的. 证明过程已在上文给出，因此省略.

另外，我实在不理解为什么要在一门数值方法的课里考这种证明. 要是考试还出一样的证明那就直接开抄，但如果考了其他证明的话......我建议还是放弃吧.

7.11 四阶经典龙格-库塔法(RK4)

这个就没什么好说的了，应该只会考计算. 这里直接给出公式

$$\begin{cases} y_{k+1} = y_k + \dfrac{h(f_1 + 2f_2 + 2f_3 + f_4)}{6} \\ f_1 = f(x_k, y_k) \\ f_2 = f\left(x_k + \dfrac{h}{2}, y_k + \dfrac{h}{2} f_1\right) \\ f_3 = f\left(x_k + \dfrac{h}{2}, y_k + \dfrac{h}{2} f_2\right) \\ f_4 = f\left(x_k + h, y_k + h f_3\right) \end{cases}$$

它的形式与数值积分章节里的辛普森公式比较相似. 其中， f_2 使用的是前进欧拉公式， f_3 使用的是后退欧拉公式.

例7.4【23-24夏试题】已知微分方程

$$\begin{cases} y' = x + y, x \in [0,2] \\ y(0) = 1 \end{cases}$$

分别用欧拉法（取 $h = 0.5$ ）和RK4法（取 $h = 1$ ）求解，并与解析解 $y = -x - 1 + 2e^x$ 对比.

解 (1) 欧拉法比较简单，过程不写了，这里直接列个表

x_i	y_i	$y(x_i)$	$\lvert y(x_i) - y_i \rvert$
0	1	1	0
0.5	1.5	1.7974	0.2974
1	2.5	3.4367	0.9367
1.5	4.25	6.4634	2.2134
2	7.125	11.7781	4.6531

(2) 用RK4法计算 y_1 的过程如下

$$
\begin{aligned}
f_1 &= f(0,1) = 1 \\
f_2 &= f(0.5, 1+0.5f_1) = 2 \\
f_3 &= f(0.5, 1+0.5f_2) = 2.5 \\
f_4 &= f(1, 1+f_3) = 4.5 \\
y_1 &= 1 + \frac{f_1+2f_2+2f_3+f_4}{6} \approx 3.4167
\end{aligned}
$$

同理可得 y_2 的计算过程

$$
\begin{aligned}
f_1 &= 4.4167 \\
f_2 &= 7.12505 \\
f_3 &= 8.479225 \\
f_4 &= 13.895925 \\
y_2 &= 3.4167 + \frac{f_1+2f_2+2f_3+f_4}{6} \approx 11.6702
\end{aligned}
$$

将结果列成表格如下

x_i	y_i	$y(x_i)$	$\lvert y(x_i)-y_i\rvert$
0	1	1	0
1	3.4167	3.4367	0.02
2	11.6702	11.7781	0.1079

解后反思 以这道题为契机，结合本章的分析，不难发现以下结论：

1. RK4法的精度明显高于欧拉法；
2. 误差 $|y(x_i)-y_i|$ 是不断累计的；
3. 对于同一种方法，步长 h 越小，误差越小.

- E - N - D -