教学班：周四下午 6.7.8

浙江大学

2022 数值计算方法大作业报告

题　　目：美国人口增长模型

姓　　名：吕智

学　　院：电气工程学院

专　　业：自动化（电气）

学　　号：3200105004

2022 年 4 月 22 日

题目：美国的人口增长模型

一、问题描述

1800-2000 年美国每年的人口记录如下表所示：

时间(年)	1800	1810	1820	1830	1840	1850	1860
人口（百万人）	5.3	7.2	9.6	12.9	17.1	23.2	31.4
时间(年)	1870	1880	1890	1900	1910	1920	1930
人口（百万人）	38.6	50.2	62.9	76.0	92.0	106.5	123.2
时间(年)	1940	1950	1960	1970	1980	1990	2000
人口（百万人）	131.7	150.7	179.3	204.0	226.5	251.4	281.4

利用以上数据和数值方法建立适合的人口增长模型，观察美国人口数量的变化规律，并对接下来的每隔 10 年进行 5 次人口预测，并对已有的 2010 和 2020 数据进行对比分析。（数据来源美国人口调查局）2010 年人口数据为 308.7 百万人，2020 年人口为 329.4 百万人。

二、问题分析

对于人口增长的问题，其影响因素很多，比如：人口基数，出生率，死亡率，迁入迁出，年龄组成，男女比例，人口的生育率和生育模式，国家医疗发展情况，国家的政治政策等众多的因素。但是如果把这些因素全部考虑进去，根本无从下手。因此，我们在下面提出一些假设来帮助我们更好的解决这些问题。

上述人口数据是以十年为一组离散的数据，我们可以对它先进行散点图绘制和插值图像绘制，来观察模型的大致情况，然后通过指数增长模型（Malthus 模型）分析，进一步优化为一个阻滞增长模型（Logistic 模型），进而进行之后五十年的数据预测。

三、问题基本假设

1. 假设所给的数据真实可靠
2. 人口数随时间连续变化并可导
3. 人口增长仅和出生率和死亡率相关
4. 各个年龄段的性别比例大致保持不变
5. 记时刻 t 时的人口为 p(t)，单位为百万人，p(t)为一个连续可微的函数，将 1800 年记为 t=0，且此时人口记为 P0，将之后的 1810，1820 直到 2000 记为 1，2……20，单位为 10 年，则 2010，2020 即为 21，22。

四、模型建立过程

1.利用表中的数据，使用 matlab 绘制散点图

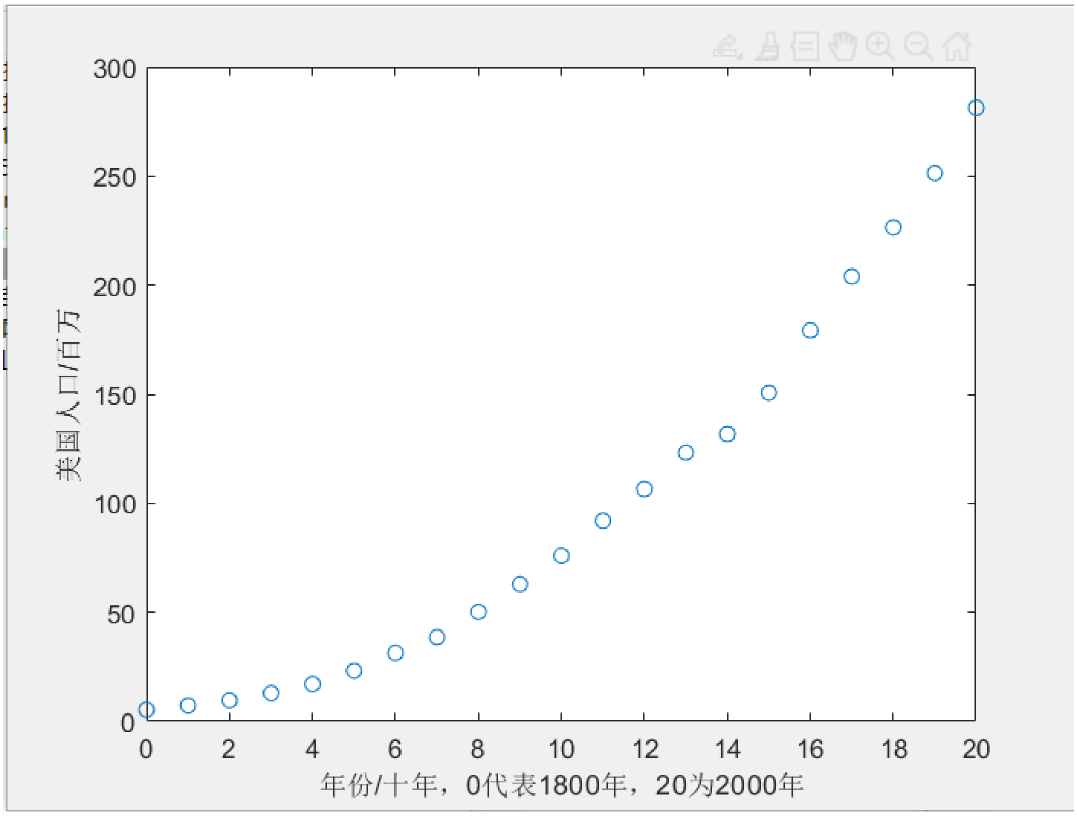

图 1：1800-2000 年的美国人口的散点数据图

2.利用三次样条插值方法绘制图像，观察趋势(函数在附录中)

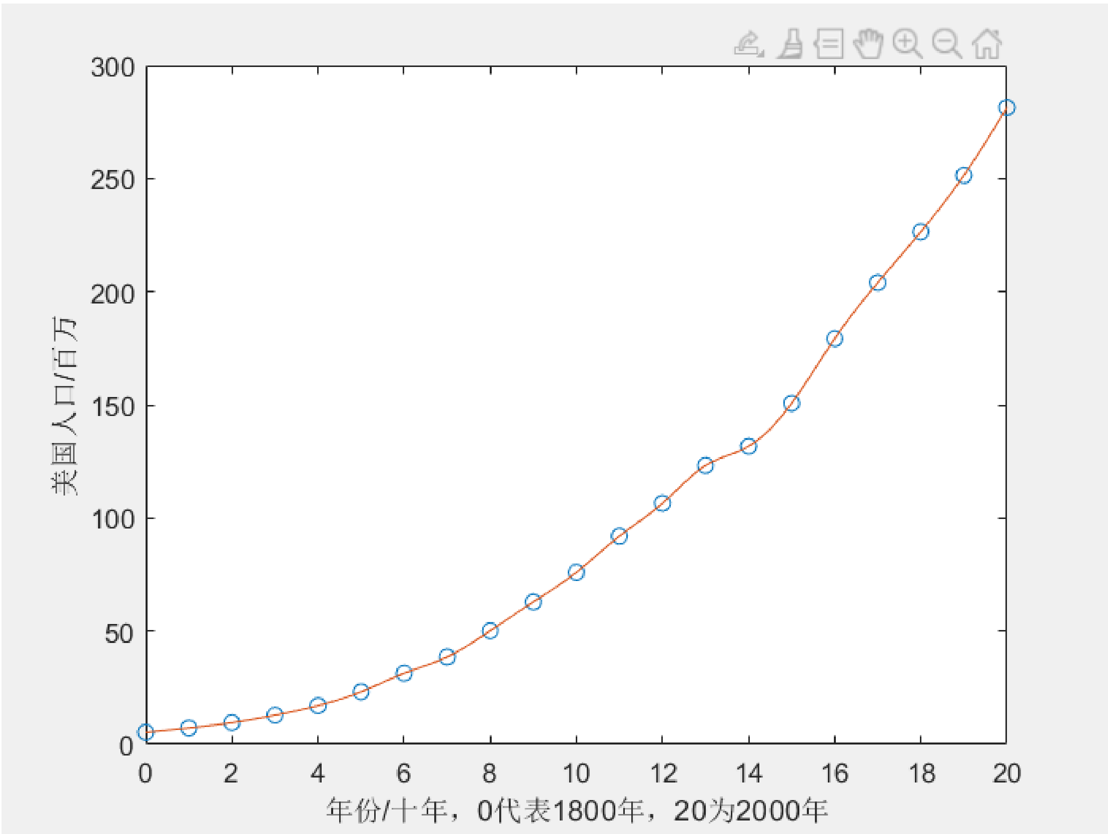

图 2：1800-2000 年的美国人口的三次样条插值数据图

观察上述散点图和三次样条插值图可知，美国人口数据的增长趋势基本符合指数增长，即指数增长模型（Malthus 模型）。

3. Malthus 模型

（1）记时刻 t 时的人口为 p(t)，单位为百万人，p(t)为一个连续可微的函数，将 1800 年记为 t=0，且此时人口记为 P0，将之后的 1810，1820 直到 2000 记为 1，2……20，单位为 10 年。

（2）假设人口增长率为常数 r，即单位时间内 x(t)的增长量满足下面的微分方程：

$$\frac{dP}{dt}=rP，P(0)=P_0 \quad \text{（a）}$$

解方程，可得$P(t)=P_0e^{rt}$，当 r>0 时，此式将按照指数规律随时间不断增长。

4. 阻滞增长模型（Logistic 模型）

（1）由于自然资源，环境条件等因素对人口增长起着阻滞作用，且随着人口的增加，阻滞作用越来越大。

（2）阻滞增长模型在指数增长模型的基本假设之上，引入了自然资源和环境条件能够容纳的最大人口数 M，称为人口容量，λ(M − P)为增长率。

可得常微分方程如下：

$$\frac{dP}{dt}=\lambda(M-P)P，P(0)=P_0 \quad \text{（b）}$$

解该常微分方程可得到 $P(t)=\frac{P_0Me^{M\lambda t}}{M-P_0+P_0e^{M\lambda t}}$ （c）

五、模型求解

1.马尔萨斯模型求解

a.算法设计：$P(t)=P_0e^{rt}$ 由已知得，P_0为 1800 年的人口数据即 5.3 百万人，但是仍有一个未知数 r 是未知的，r 可以通过实际数据得线性最小二乘法求解，对于原方程直接求解是比较麻烦的，因此再两边取对数，可得$\ln P(t)=\ln P_0+rt$,lnP_0已知，利用最小二乘法拟合的方法把 r 求出来，具体操作如下：

原式可以看成 y =ax + b，且 b 已知的形式，满足最小二乘法的条件为残差和最小，此时满足$\sum_{i=1}^{n}(ax_i+b-y_i)x_i=0$，代入美国人口数值可求得 a = 0.2265，即 r 为 0.2265.

马尔萨斯模型表达式为$P(t)=5.3e^{0.2265t}$，图像如下页图 3 所示，红线表示马尔萨斯人口模型数据，蓝色圆圈代表实际人口数据。(此部分具体程序请看附录部分)

b. 结果与分析：从图中可以看到，马尔萨斯模型在前面几个点的吻合度较高，在后面的点距离曲线偏差较大，拟合效果并不好。通过表达式也可以对 2010——2050 年的数据进行计算，结果如下表

年份	2010	2020	2030	2040	2050
数据	616.057	772.628	968.992	1215.3	1524.1
真实数据	308.7	329.4			
误差	308	443			

可以看到，误差是相当大的，这是由于指数函数有着指数爆炸的性质，随着 t 的增加，函数值增加的非常的快，所以用此模型计算结果并不准确。同时，我们知道人的增长不可能是无限的，所以需要对其加以限制，这就引出了我们的阻滞增长模型。

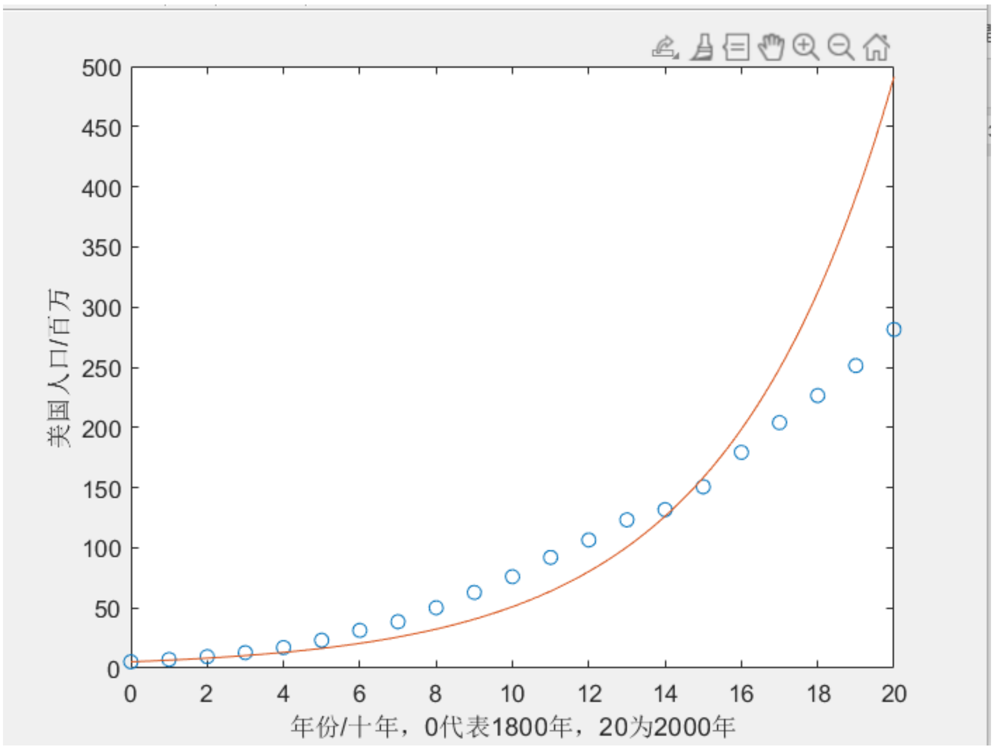

图 3：马尔萨斯模型数据图

2. Logistic模型求解

(1)通过离散模型进行参数拟合

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}\boldsymbol{t}}=\boldsymbol{\lambda}(\boldsymbol{M}-\boldsymbol{P})\boldsymbol{P}$$

a.算法设计：上式为人口的阻滞模型的表达式，其中有两个参数未知，λ和M。我们可以将原方程写乘离散形式，即$\frac{\boldsymbol{P(n+1)-P(n)}}{\boldsymbol{P(n)}}=-\boldsymbol{\lambda P(n)}+\boldsymbol{\lambda M}$，其中可以将$\frac{\boldsymbol{P(n+1)-P(n)}}{\boldsymbol{P(n)}}$看成y，P(n)看为x，即写成了线性拟合的表达形式：y=kx+b，k=$-\boldsymbol{\lambda}$，b=$\boldsymbol{\lambda M}$。以1860为起始时间，对1860——2000年的数据进行拟合可以得到k = -6.887×10^{-4}，b=0.2559。即得到λ = 6.887×10^{-4}，M = 371.9399。将$\boldsymbol{\lambda}$和M代入进常微分方程式中，利用四阶龙格库塔法对每年数据求解，绘制曲线图如下页所示。

b. 结果与分析：可以看到logistic曲线图和马尔萨斯模型图相比更加缓慢，数据也更加接近，曲线图更像一个“S”型，而马尔萨斯模型图像更像一个“J”型。

对 2010——2050 年的数据通过 Logistic 模型计算结果如下表所示：

年份	2010	2020	2030	2040	2050
数据	301.5607	315.0201	326.2951	335.5944	343.1662
真实数据	308.7	329.4			
误差	2.3%	4.2%			

可以看到，2010 年的误差仅为 2.3%，2020 年的误差为 4.2%，也在 5%以内，已经较为精确。

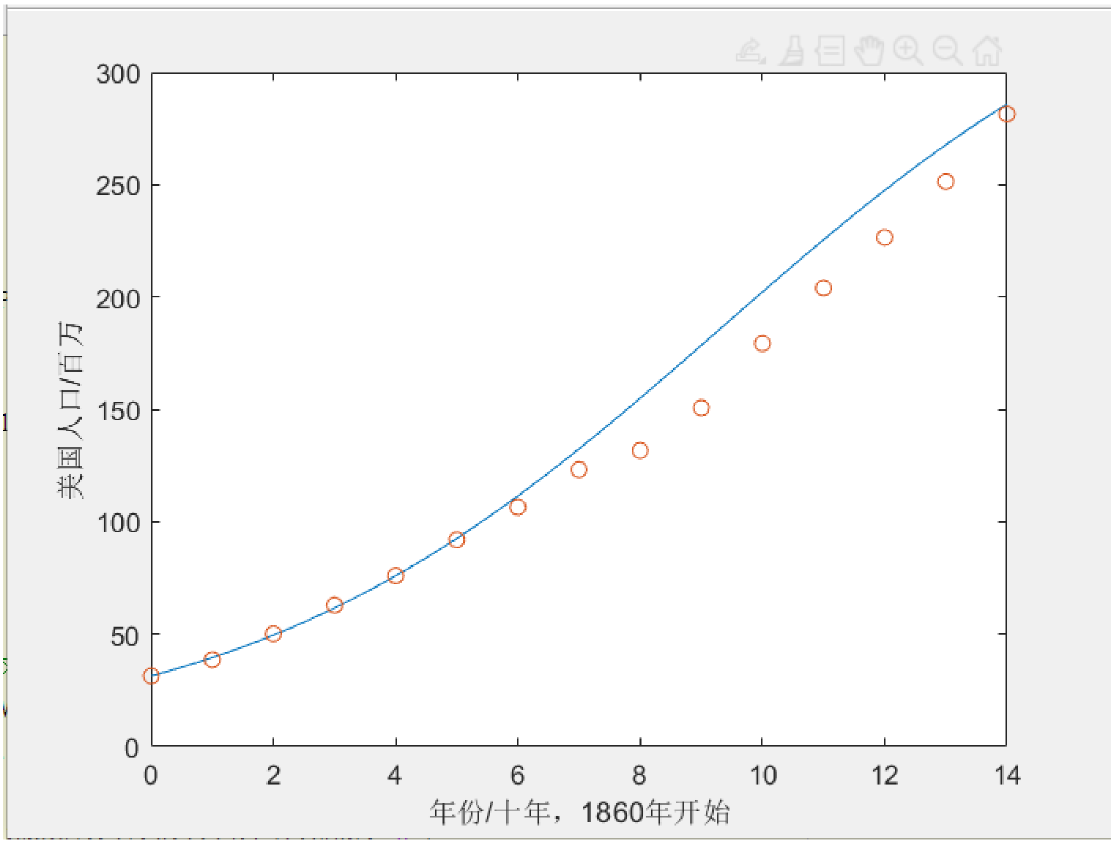

图 4：离散阻滞增长模型曲线图

(2)利用数值微分近似导数来参数拟合

a.算法设计： 原常微分方程可以变换为$\frac{\frac{dP}{dt}}{P}=\lambda M-\lambda P$，可通过数值微分即差商近似微商求得$\frac{dP}{dt}$。然后同样可以将$\frac{\frac{dP}{dt}}{P}$看作y，P看为x，进行y=kx+b的最小二乘法线性参数拟合，其中k=-λ，b = λM。以1860——2000年的数据进行拟合，首先先对每一个节点使用数值微分的三点公式求得$\frac{dP}{dt}$，然后获得$\frac{\frac{dP}{dt}}{P}$序列，进而线性拟合。可以求得k = -4.517×10^{-4}，b=0.2145。即得到 λ =4.517×10^{-4}，M = 474.9105。将$\boldsymbol{\lambda}$和M代入进常微分方程式中，利用四阶龙格库塔法对每年数据求解，绘制曲线图如下页所示。

b. 结果与分析： 可以看到此图像和真实情况更加吻合，图像效果比离散模型要好。

对 2010——2050 年的数据通过数值微分 Logistic 模型计算结果如下表所示：

年份	2010	2020	2030	2040	2050
数据	303.3764	326.1207	347.1195	366.1432	383.0841
真实数据	308.7	329.4			
误差	1.5%	0.99%			

可以看到，2010 年的误差仅为 1.5%，2020 年的误差仅为 0.99%，可以看到预测效果非常好，可以认为该模型是相当让人满意的。最终我们预测 2030 年美国人口将 347.1195 百万人，2040 年美国人口 366.1432 百万人，2050 年美国人口为 383.0841 百万人。

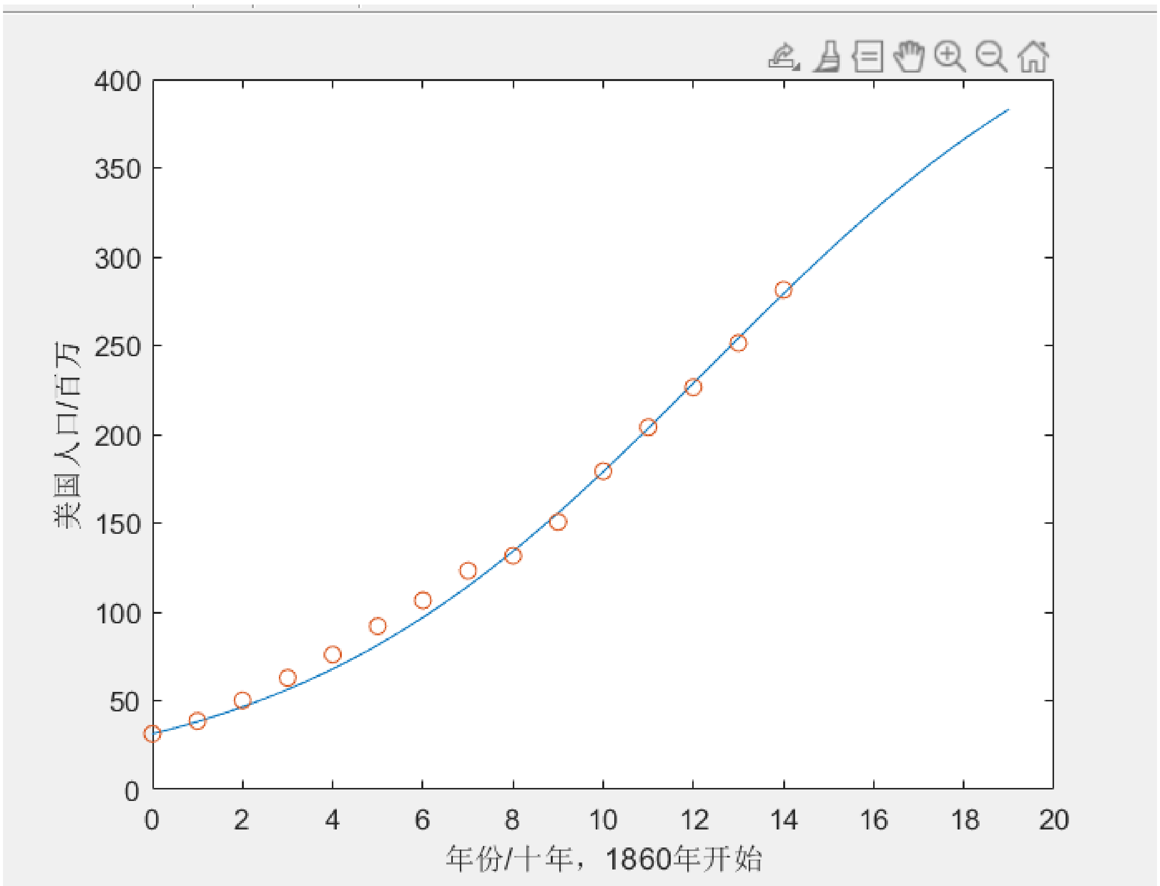

图 5：数值微分方法的阻滞增长模型曲线图

六、模型评价与推广

Malthus 数学模型在短期内具有较好的准确度，简单易行，但是不能准确的预测出人口长期的发展趋势，不具有预测人口长期增长数量的能力。

在人口增长过程中 logistic 模型预测的数据与题中数据更为接近，能够较好地进行预测，且 2010 和 2020 年的数据结果，与真实结果相差不大。

但是 logistic 模型也并不完善，实际情况下，还会涉及到国家宏观调控，科技进步带来的人口容量提升，生育率改变等多方面的综合影响，所以想要更加准确的预测未来的人口趋势变化，目前的 logistic 模型还远远不足。

七、数值计算方法的使用

1. 三次样条插值的使用

$S_i(x) = a_i x^3 + b_i x^2 + ci^x + d_i$， $x \in [x_{i-1}, x_i]$,i=1,2,3,…,n

式中存在 4n 个未知数，满足如下条件：

1.端点值相等，i=1,2,3,…,n-1

2.端点的一阶导值相等，i=1,2,3,…,n-1

3.端点的二阶导值相等，i=1,2,3,…,n-1

4. $S_i(x) = f_i(x)$，i=0,2,3,…,n

以上共 4n-2 个式子，无法满足条件，所以需要边界条件的限制，本例中假设了在起始点和最终点的二阶导数已知且都为 0 的自然边界条件。

算法设计：
利用三弯矩法实现三次样条插值
利用 S(x)在节点 x_i 处的二阶导数值 $M_i=S''(x_i)$表示 $S_i(x)$。用 S'(x)在内节点 x_i 上的连续性和边界条件来确定 M_i，利用二阶导数连续性和插值条件端点值相等和二阶导数相等
得到此式：$u_i M_{i-1} + 2M_i + \lambda_i M_{i+1} = g_i$　i=1,2,3,…,n-1

其中：$u_i = \frac{h_i}{h_i+h_{i+1}}$　$\lambda_i = 1 - u_i$　$g_i = 6f[x_{i-1}, x_i, x_{i+1}]$　$h_i = x_i - x_{i-1}$

再根据自然边界条件可得到如下矩阵：

$$A = \begin{pmatrix} 2 & \lambda 1 & & \cdots & & & \\ \mu 2 & 2 & \lambda 2 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ & & & & u_{n-2} & 2 & \lambda_{n-2} \\ 0 & 0 & 0 & \cdots & & & \\ & & & & u_{n-1} & & 2 \end{pmatrix}$$

X = [M1 ,…,M_{n-1}]
B = [g1，g2，…,g_{n-1}]

求解出 X，将参数代入到三次样条插值函数公式

$$S_i(x) = \frac{1}{6h_i}[M_{i-1}(x_i - x)^3 + M_i(x - x_{i-1})^3] + (f_{i-1} - \frac{M_{i-1}h_i^2}{6})\frac{x_i - x}{h_i} + (f_i - \frac{M_i h_i^2}{6})\frac{x - x_{i-1}}{h_i}$$

绘制图像即可。

2. 最小二乘法曲线拟合

构造一个能逼近列表数据的近似的数学表达式，使各数据点从总体上最贴近，而不一定要求构造的函数曲线通过所给数据点

算法设计：线性拟合方法即$\varphi(x) = a + bx$

建立此方程组：$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$ 利用高斯消元法或高斯约当法等线性方程组求解方法求解系数 a 和 b

3.四阶龙格库塔法求解常微分方程

$$\begin{cases} y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ k_1 = hf(x_i, y_i) \\ k_2 = hf(x_i + \frac{h}{2}, y_i + \frac{k_1}{2}) \\ k_3 = hf(x_i + \frac{h}{2}, y_i + \frac{k_2}{2}) \\ k_4 = hf(x_i + h, y_i + k_3) \end{cases}$$

4.数值微分

两点差商近似：

向前差商： $f'(x_0) \approx \frac{f(x_0+h)-f(x_0)}{h}$

向后差商： $f'(x_0) \approx \frac{f(x_0)-f(x_0-h)}{h}$

中间差商： $f'(x_0) \approx \frac{f(x_0+h)-f(x_{0-\mathrm{h}})}{2h}$

数值微分三点公式

中间差商公式和上面方法相同。

三点向前差商公式：$f'_{(x_0)} = \frac{1}{2h}[-3f(x_0) + 4f(x_1) - f(x_2)]$

三点向后差商公式：$f'_{(x_2)} = \frac{1}{2h}[f(x_0) - 4f(x_1) + 3f(x_2)]$

八、参考资料

1. https://zhuanlan.zhihu.com/p/181717510 人口增长模型的文本介绍
2. 数据建模之人口模型（3）——模型的参数拟合_哔哩哔哩_bilibili 人口增长模型的视频介绍

附录：

1. Mian.m 主函数，所有操作在这里进行，分节运行

```
%%
clear;clc;
%散点图绘制
T = 0:1:20;%表示年份，题目所给的数据是以每10年为期的
P = [5.3,7.2,9.6,12.9,17.1,23.2,31.4,38.6,50.2,62.9,76.0,92.0,106.5,123.2,131.7,150.7,179.3,204.0,226.5,251.4,281.4];%人口数据
plot(T,P,'o');%绘制散点图观察趋势
xlabel('年份/十年，0代表1800年，20为2000年');
ylabel('美国人口/百万');
```

```
%%
%三次样条插值图像绘制
M1 = [T' P'];
figure(2);
Interpolation_Spine3(M1);
xlabel('年份/十年，0代表1800年，20为2000年');
ylabel('美国人口/百万');
```

```
%%
%马尔萨斯模型得线性拟合
lnP = log(P);%转换为对数形式
sum1 = 0;%sum1表示T(i)*lnP(i)得和
sum2 = 0;%sum2表示T(i)的平方和
sum3 = 0;
n = length(lnP);
for i = 1:n
    sum1 = sum1 + T(i)*lnP(i);
    sum2 = sum2  + T(i)^2;
    sum3 = sum3 + lnP(1)*T(i);
end
r =(sum1-sum3)/sum2;%最小二乘法算出增长率r

PMf = @(t)P(1)*exp(r*t);%马尔克斯模型函数

t = [];
PM = [];
for i = 0:0.1:n-1
    t(end+1) = i;
    PM(end+1) = PMf(i);
end

plot(T,P,'o',t,PM);%绘制图像
xlabel('年份/十年，0代表1800年，20为2000年');
ylabel('美国人口/百万');
```

```
%%
%Logistic模型
%离散模型参数拟合
n = length(P);
d_P = [];%保存(P(i+1) - P(i)) / P(i)
for i = 7:n-1
    d_P(end+1) = (P(i+1) - P(i)) / P(i);
end

[k,b] = fit(P(7:n-1),d_P);%7代表从1860年开始，线性拟合返回参数

%
lamuda = -k;
M = b/lamuda;

%利用四阶龙格库塔法求解常微分方程
f = @(x,y)lamuda*(M-y)*y;
output = RK4(0,n-7,P(7),f,0.1);%四阶龙格库塔法函数，返回一个矩阵

plot(output(:,2),output(:,1),0:1:14,P(7:end),'o');
xlabel('年份/十年，1860年开始');
ylabel('美国人口/百万');
```

```
%%
n = length(P);
dP_P = [];%保存dP/dt /P
for i = 7:n
    if i == 7
        temp = df3(i,1,P);%前向差商公式
    elseif i == n
        temp = db3(i,1,P);%后向差商公式
    else
        temp = dm(i,1,P);%中心差商公式
    end
    dP_P(end+1) = temp/P(i);
end

[k,b] = fit(P(7:n),dP_P);%线性拟合返回参数

lamuda = -k;
M = b/lamuda;

%利用四阶龙格库塔法求解常微分方程
f = @(x,y)lamuda*(M-y)*y;
output = RK4(0,n-2,P(7),f,0.1);

plot(output(:,2),output(:,1),0:1:14,P(7:end),'o');
xlabel('年份/十年，1860年开始');
ylabel('美国人口/百万');
```

2.Interpolation_Spine3.m 三次样条插值函数

```
function Interpolation_Spine3(M)
    %假设满足x0处的二阶导和xn处的二阶导均为0，自然边界条件
    h = [];
    [n,~] = size(M);
    for i = 2:n
        h(end+1) = M(i,1) - M(i-1,1);
    end
    lamuda = [];
    miu = [];%lamuda和miu为记录值
    g = [];%g为方程右边值
    for i = 1:n-2
        lamuda(end+1) = h(i+1)/(h(i) + h(i+1));
        miu(end+1) = h(i)/(h(i) + h(i+1));
        g(end+1) = 6*(((M(i+2,2)-M(i+1,2))/(M(i+2,1) - M(i+1,1)) - (M(i+1,2)-M(i,2))/(M(i+1,1) - M(i,1)))/(h(i)+h(i+1)));
    end

    A = diag(diag(2*ones(n-2)));

    A(1,2) = lamuda(1);
    %构建系数矩阵
```

```
    for i=2:n-3
        A(i,i-1) = miu(i);
        A(i,i+1) = lamuda(i);
    end
    A(n-2,n-3) = miu(n-2);

    %求解一阶导数向量M1
    M1 = Guass_Jordan(A,n-2,g');
    M1 = [0;M1;0];
    %三次样条函数
    X = [];
    Y = [];
    k = 1;
    for i = M(1:1):0.1:M(n,1)
        if i == M(k+1,1)
            k = k + 1;
        end
        X(end+1) = i;
        if k~= n
            Y(end+1) = 1/(6*h(k))*(M1(k)*((M(k+1,1)-i)^3) + M1(k+1)*((i-M(k,1))^3)) + (M(k,2)-M1(k)*(h(k)^2)/6)*(M(k+1,1)-i)/h(k) + (M(k+1,2) - M1(k+1)*(h(k)^2)/6)*(i-M(k,1))/h(k);
        else
            Y(end+1) = M(n,2);
        end
    end

    %绘制图像
    plot(M(:,1),M(:,2),'o',X,Y)

end
```

3.Guass_Jordan 高斯约当法解线性多项式，在拟合和三次样条插值样条函数里都会用到

```
function [X] = Guass_Jordan(A,n,b)
    A = [A b];%增广矩阵
    for i=1:n
        x = A(i,i);
        for j = 1:n+1
            A(i,j) = A(i,j)/x;%除以主元
        end

        for j = 1:n%标准化
            y = A(j,i);
            if j~= i
                for k = 1:n+1
                    A(j,k) = A(j,k) - y*A(i,k);
                end
            end
        end
    end
    X = A(:,n+1);%获得增广矩阵的最后一列即为所求
end
```

4.fit.m 进行线性拟合，返回斜率和截距两个参数

```
function [a,b] = fit(x,y)  %最小二乘法拟合曲线,传入的x和y为两个行向量
    %求解最小二乘法的系数a和b
    %构建系数矩阵和方程右边列向量
    A = zeros(2,2);
    [~,n] = size(x);
    A(1,1) = n;
    A(1,2) = sum(x);
    A(2,1) = A(1,2);
    B = zeros(2,1);
    B(1) = sum(y);
    for i = 1:n
        A(2,2) = A(2,2) + x(i)^2;
        B(2) = B(2) + x(i)*y(i);
    end

    X = Guass_Jordan(A,2,B);%高斯约旦法求解出一个[a b]的2×1的列向量
    a = X(2);%a为斜率
    b = X(1);%b为截距


end
```

5.RK4.m 四阶龙格库塔法

```
function output = RK4(a,b,y,f,h)%传入参数为起始点，终点，初值，f(x,y)，步长
    %经典的四阶龙格库塔法
    output = [];%最终返回一个包含每一步步长计算的矩阵
    output(end+1,1) = y;
    output(end,2) = a;
    for x=a:h:b-h            %迭代计算，直至终点
        k1 = f(x,y);
        k2 = f(x+h/2,y+h/2*k1);
        k3 = f(x+h/2,y+h/2*k2);
        k4 = f(x+h,y+h*k3);          %四阶龙格库塔法的经典做法
        y = y + (k1 + 2*k2 + 2*k3 + k4)/6*h;
        output(end+1,1) = y;
        output(end,2) = x+h;
    end
end
```

6.df3.m 和 db3.m 和 dm.m 数值微分三点差商公式 传入参数均为起始点，步长，函数

```
function [output] = df3(x,h,f)
    %三点向前差商公式
    output = (-3*f(x) + 4*f(x+h) - f(x+2*h))/2/h;
end
```

```
function [output] = db3(x,h,f)
    %三点向后差商公式
    output = (f(x-2*h) - 4*f(x-h) + 3*f(x))/2/h;
end
```

```
function [output] = dm(x,h,f)%参数为被导的x，步长h，函数f
    output = (f(x+h)-f(x-h))/h/2;
end
```