

Chapter 1

2022 年上海疫情数据数值拟合

1.1 introduction

在已有部分数据的条件下,使用数值方法对现实问题进行模拟预测是数值科学在现实生活中的常见应用。今年三月以来,上海市疫情爆发引起社会各界的广泛关注。虽然目前疫情已经基本得到有效控制,但是本文以学习为目的使用数值方法对疫情数据进行模拟。

本报告中,基于表 1.1 的数据,使用最小二乘拟合和基于传染病模型 SIR 方程组的数值求解对上海疫情以来的数据进行数值模拟。由于确诊感染数与无症状感染数差距过于悬殊,该文全部数据取自百度实时确诊/治愈数据。

日期	3.17	3.24	3.31	4.7	4.14	4.21	4.28	5.5	5.12	5.19	5.26
累计确诊	4944	5135	6452	9334	20244	39643	56300	60113	61741	62521	62911
累计治愈	4281	4779	4998	5270	7936	15836	31782	50871	56424	58539	60499

表 1.1: 上海疫情无症状感染人数统计

1.2 数值模拟

1.2.1 最小二乘拟合

出于各种理论和实践原因,寻求一些高维物体的低维近似是常见的。这样做我们可以消除错误或忽略不相关的细节,如从嘈杂的数据中提取信号或趋势,或将大量数据减少到更易于管理的数量,或用简单的近似值替换一些复杂的函数。我们并不期望这样的近似值是准确的,实际上,在大多数情况下,我们不希望它是准确的,但我们希望保留一些与原始数据的相似性。在线性代数术语中,我们希望将向量从高维空间投影到低维子空间。实现这一点的最常用且计算方便的方法之一是最小二乘法,我将在本节中使用的方法。

Definition 1.1. 给定点集 $(t_i, y_i), i = 1, \dots, m$ 。我们希望找到一个 n 维向量 x 函数 $f(t, x) : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ 满足”最好”地拟合给定数据, ”最好”的定义维最小二乘意义下:

$$\min_x \sum_{i=1}^m (y_i - f(t_i, x))^2.$$

其中线性的表示 f, x 是线性关系,如下所示

$$f(t, x) = x_1 + x_2 t + x_3 t^2 + \dots + x_n t^{n-1}.$$

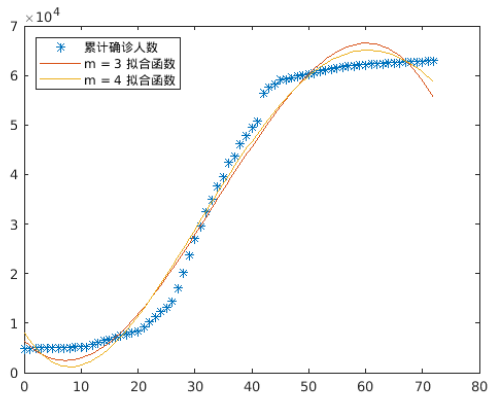
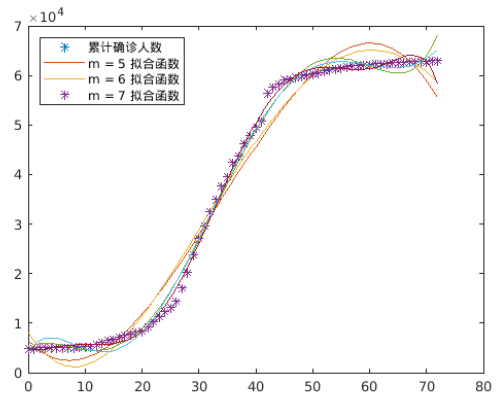
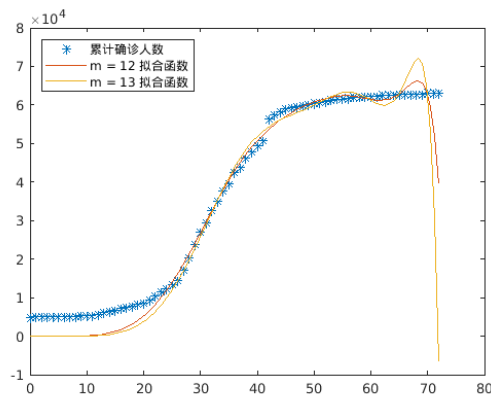
(a) $m = 3, 4$ 时的最小二乘拟合结果(b) $m = 5, 6, 7$ 时的最小二乘拟合结果

图 1.1: 上海无症状感染者拟合结果

图 1.2: 当 m 过大时的拟合结果

对于真实数据拟合时, 转化为求解微分方程组, 在 matlab 中矩阵除法即得结果 x 。

当 $m = 3, 4, 5, 6, 7$ 的计算结果如图所示 1.1(a)1.1(b) 从图中可以看出当 m 不够大 ($m = 3, 4$, 左图) 时最小二乘拟合结果不理想。但是当 m 足够大时继续提升阶数对拟合结果影响不明显 (右图), 并且继续提高 m 会在边界处引起类似拉格朗日龙格现象的震荡如下图 1.2.1。对于该数据, $m = 5$ 时最合适的拟合阶数。

1.3 SIR 模型模拟

分隔模型是一种非常通用的建模技术。它们通常用于传染病的数学建模。所有人被分配到带有标签的人群, 例如 S、I 或 R (易感、感染或恢复)。人可能会在人群之间变化。标签的顺序通常显示人群之间的流动模式; 例如, SIR 意味着易感, 感染, 恢复。

SIR 模型最简单的传染病模型之一, 很多更深的模型从中衍生发展, 这个模型分为三个部分, $S(t)$ 表示易感人群的人数, $I(t)$ 是感染人群人数, $R(t)$ 是恢复人群。

通常模型做如下几个假设

- 人口数量始终不变 (不考虑人口的出生、死亡和流动等种群动力因素)

$$N(t) = S(t) + I(t) + R(t) = N.$$

- 一个感染人一旦与易感者接触就会具有一定的传染力, 假设该传染概率为 β , 易感人数减少的速度等于转

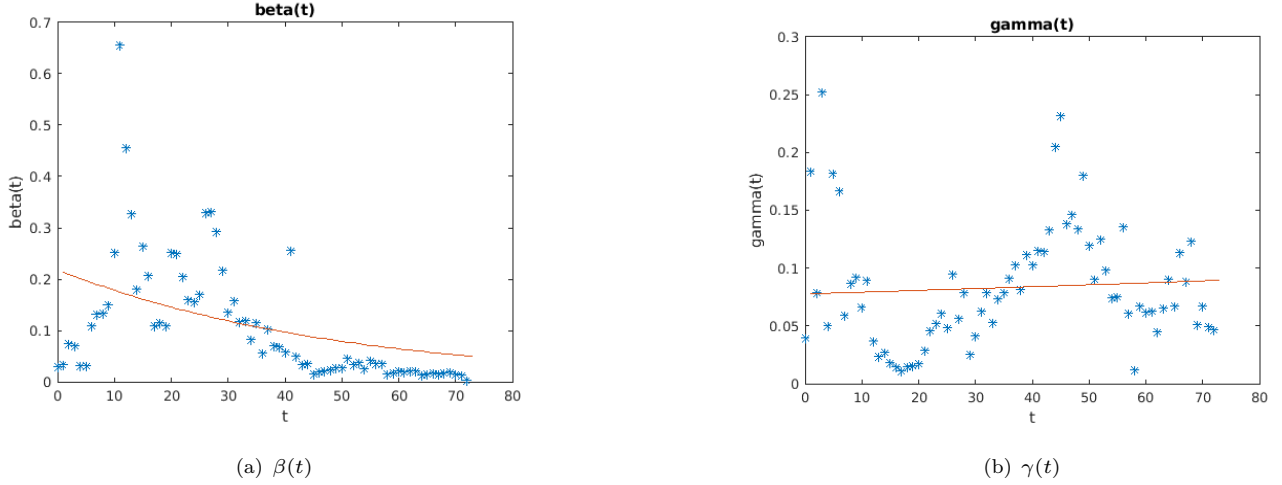


图 1.3: 参数拟合结果

变为感染人数的速度，即

$$\frac{dS(t)}{dt} = -\beta \frac{I(t)}{N(t)} S(t).$$

- 单位时间内从感染者中恢复的人数与感染者数量成正比，以 γ 来表示从感染到恢复的比例，从而

$$\frac{dR(t)}{dt} = \gamma I(t).$$

基于以上这三个假设，我们还可以得到感染人数增加的速度等于易感者增加的速度减去恢复人数增加的速度，即

$$\frac{dI(t)}{dt} = \beta \frac{I(t)}{N(t)} S(t) - \gamma I(t).$$

综上，我们有偏微分方程组

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta \frac{I(t)}{N(t)} S(t) \\ \frac{dI(t)}{dt} &= \beta \frac{I(t)}{N(t)} S(t) - \gamma I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t), \end{aligned}$$

初值为 $S(0) = 25000000, I(0) = 100, R(0) = 0$. 为数值求解该方程组还需要两个参数值 β, γ ，采用百度实时疫情动态数据确诊与治愈人数 $I(t), R(t)$ ，拟合估算动态 $\beta(t), \gamma(t)$ ，

因感染人数相较于上海市常住人口 2500 万较小，不妨设 $N(t) = S(t)$. 可得如图 1.3.

使用 matlab 自带的龙格库塔步进可得如图 1.3. 可以看出结果非常不理想. 1.3 中拟合 β, γ 的源数据分布不佳，可以推断上海疫情数据不适合使用 SIR 模型进行模拟。

1.4 模拟结果分析

使用最小二乘拟合较为成功的模拟了疫情发展情况，而 SIR 模型失败了。我认为原因在于最小二乘没有对疫情发展做过多的相关性假设，导致在模拟过程中错误的假设对模拟产生负面的影响。

而在 SIR 模型中，假设了当感染人群 $I(t)$ 数量上升时对易感人群 $S(t)$ 感染造成线性增长的正相关关系。虽然在疫情前期缺乏管控期间，理论上确实如此发展，但是在疫情大部分和主要增长的中后期，政府的防控措施将疫情和绝大部分感染人群 $I(t)$ 控制在疫区密接人群 $M(t)$ 当中，该人群数量远小于整个上海的易感人群

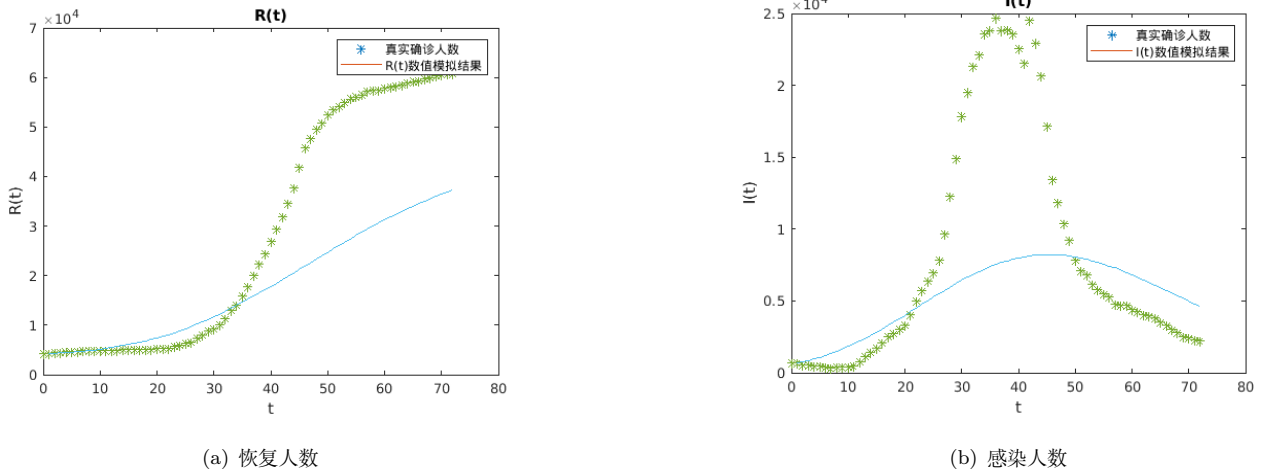


图 1.4: 参数拟合结果

$M(t) \ll S(t) \approx N(t)$, 因此方程组第一个方程造成了不良影响。另一方面, 由于新冠病毒的高突变概率, 恢复人群 $R(t)$ 不能完全避免再度感染, 因此第二方程也不能正确反映感染人群 $I(t)$ 的变化情况。最后, 由于在政府管控手段下恢复人群 $R(t)$ 远没有达到群体免疫的状况下, 疫情就接近停息, 所以疫情终止也不是由于 SIR 模型中的假设的群体免疫。因此综上, SIR 较为失败得拟合上海疫情数据结果是情理之中。

1.5 附录

该节包含两种数值模拟的 matlab 源代码

1.5.1 最小二乘拟合

```
clear
xx = linspace(0,72,73)';
ss = [4944;4964;4986;5027;5066;5080;5094;5135;
      5182;5233;5293;5400;5729;6087;6452;6714;
      7158;7591;7862;8177;8506;9334;10351;11366;
      12283;13281;14471;17044;20244;23838;27078;29498;
      32582;35077;37712;39642;42379;43780;46252;47913;
      49519;50811;56300;57550;58339;59068;59342;59605;
      59868;60113;60366;60583;60907;61141;61370;61514;
      61741;61936;62104;62174;62251;62348;62431;62521;
      62605;62658;62714;62772;62816;62865;62911;62950;
      62982;62988];

cc = [4281;4307;4428;4472;4612;4635;4716;4779;
      4800;4833;4870;4898;4943;4972;4998;5037;
      5068;5098;5125;5166;5213;5270;5389;5616;
      5914;6305;6646;7387;7936;8911;9288;10021;
      11238;12923;14100;15836;17715;19958;22409;24352;
      26992;29302;31782;34593;37649;41891;45873;47742;
```

```

49481;50871;52535;53476;54121;54968;55575;56008;
56424;57146;57439;57497;57812;58087;58354;58539;
58798;59142;59379;59759;60027;60371;60499;60661;
60775;60879];

% yy = ss(2:end) - ss(1:end-1);
yy = ss(1:end-1);

plot(xx,yy,'*')
% axis([0 72 -30 300])
hold on
dist = 5;
for m=5:1:7
x = xx(1:dist:end);
y = yy(1:dist:end);
A=[];
a=[];
for i=0:1:m
    A = [A x.^i];
end
a = A\y;
u = linspace(0,72)';
f = a(1);
for i=1:1:m
    f = f + a(i+1) * u.^i;
end
plot(u, f,'-');
% plot(u,a(1)+a(2)*u+a(3)*u.^2+a(4)*u.^3 + a(5)*u.^4 + a(6)*u.^5,'-')
hold on
str(m) = "m = " + num2str(m) + " 拟合函数";
end

legend('累计确诊人数', str(5), str(6), str(7) , 'Location', 'NorthWest')

```

1.5.2 SIR 数值模拟

```

clear
xx = linspace(0,72,73)';
ss = [4944;4964;4986;5027;5066;5080;5094;5135;
5182;5233;5293;5400;5729;6087;6452;6714;
7158;7591;7862;8177;8506;9334;10351;11366;
12283;13281;14471;17044;20244;23838;27078;29498;
32582;35077;37712;39642;42379;43780;46252;47913;
49519;50811;56300;57550;58339;59068;59342;59605;

```

```

59868;60113;60366;60583;60907;61141;61370;61514;
61741;61936;62104;62174;62251;62348;62431;62521;
62605;62658;62714;62772;62816;62865;62911;62950;
62982;62988];

cc = [4281;4307;4428;4472;4612;4635;4716;4779;
4800;4833;4870;4898;4943;4972;4998;5037;
5068;5098;5125;5166;5213;5270;5389;5616;
5914;6305;6646;7387;7936;8911;9288;10021;
11238;12923;14100;15836;17715;19958;22409;24352;
26992;29302;31782;34593;37649;41891;45873;47742;
49481;50871;52535;53476;54121;54968;55575;56008;
56424;57146;57439;57497;57812;58087;58354;58539;
58798;59142;59379;59759;60027;60371;60499;60661;
60775;60879];

I = ss(1:end) - cc(1:end);

dI = I(2:end) - I(1:end-1);
dR = cc(2:end) - cc(1:end-1);

i = I(1:end-1);

gamma = dR ./ i;
beta = (dR + dI) ./ i;

figure(1);
plot(xx, beta, '*'); hold on;
title("beta(t)"); xlabel t; ylabel beta(t);
figure(2);
plot(xx, gamma, '*'); hold on;
title("gamma(t)"); xlabel t; ylabel gamma(t);

syms b(t) g(t)

b = fit(xx, beta, 'exp1');
g = fit(xx, gamma, 'exp1');
figure(1); plot(b(xx));
figure(2); plot(g(xx));

f=@(t,y)[y(1)*(b(t)-g(t));g(t)*y(1)];
[x,y]=ode45(f,1:xx(end),[i(1) cc(1)]);

```

```
figure(3);
plot(xx, I(1:end-1), '*');
hold on; plot(x, y(:, 1));
title("I(t)"); xlabel t; ylabel I(t);
legend('真实确诊人数', "I(t)数值模拟结果");
figure(4);
plot(xx, cc(1:end-1), '*');
hold on; plot(x, y(:, 2));
title("R(t)"); xlabel t; ylabel R(t);
legend('真实确诊人数', "R(t)数值模拟结果");
```