

## 一、问题的重述

考虑航天器在仅受到地球万有引力、航天器自身发动机作用力的作用下作平面运动，将地球和航天器视为质点，建立航天器运动的数学模型。

显然这样的数学模型在精度上是远远不能满足实际需要的，在其他要求精确制导等有关高科技的实际问题中，我们都面临着类似的问题：我们必须建立高精度的数学模型，必须高精度地估计模型中的大批参数，因为只有这样的数学模型才能解决实际问题，而不会出现差之毫厘，结果却失之千里的情况。由于航天器的问题太复杂，本题仅考虑较简单的确定高精度参数问题。

假设有一个生态系统，其中含有两种生物，即：A 生物和 B 生物，其中 A 生物是捕食者，B 生物是被捕食者。假设  $t$  时刻捕食者 A 的数目为  $x(t)$ ，被捕食者 B 数目为  $y(t)$ ，它们之间满足以下变化规律：

$$\begin{cases} x'(t) = x(t)[\alpha_1 + \alpha_2 y(t)] \\ y'(t) = y(t)[\alpha_3 + \alpha_4 x(t)] \end{cases}$$

初始条件为：

$$\begin{cases} x(t_0) = \alpha_5 \\ y(t_0) = \alpha_6 \end{cases}$$

其中  $\alpha_k (1 \leq k \leq 6)$  为模型的待定参数。

通过对此生态系统的观测，可以得到相关的观测数据。要利用有关数据，解决以下问题：

- 1) 在观测数据无误差的情况下，若已知  $\alpha_2$ ，求其它 5 个参数  $\alpha_k (k=1,3,4,5,6)$ ？
- 2) 若  $\alpha_2$  也未知，至少需要多少组观测数据，才能确定参数  $\alpha_k (1 \leq k \leq 6)$ ？
- 3) 在观测资料有误差（时间变量不含有误差）的情况下，确定参数  $\alpha_k (1 \leq k \leq 6)$  在某种意义下的最优解，并与仿真结果比较，进而改进数学模型。
- 4) 假设连观测资料的时间变量也含有误差，确定参数  $\alpha_k$  在某种意义下的最优解。

## 二、航天器运动模型的建立

考虑航天器在仅受到地球万有引力、航天器自身发动机作用力的作用下作平面运动，将地球和航天器视为质点，由理论力学可知，一个刚体在空间的运动可以看作质心的移动，因此可以应用质心运动定理来研究刚体质心的移动规律。以地球中心为原点，建立直角坐标系，航天器绕地球飞行，可以出现在该直角坐标系中四个象限的任意一个之内。平面直角坐标系如图 1。

符号说明如下：

$v_x$  —— 航天器在  $x$  方向的速度

$v_y$ ——航天器在  $y$  方向的速度

$F_1$ ——万有引力， $F_1 = G \frac{Mm}{r^2} = G \frac{Mm}{x^2 + y^2}$

$F_2$ ——航天器发动机作用力，为控制变量

$\alpha$ ——万有引力与  $x$  轴正方向的夹角

$\beta$ ——航天器发动机作用力与  $x$  轴正方向的夹角

$t_0$ ——初始时刻

$x_0, y_0$ ——航天器初始位置

$v_{x0}$ ——航天器  $x$  方向初速度

$v_{y0}$ ——航天器  $y$  方向初速度

航天器受的万有引力  $F_1$  方向指向地球中心（原点），航天器受推力  $F_2$  的方向与  $x$  轴正方向成  $\beta$  角。将  $F_1$  和  $F_2$  投影到该直角坐标系上，见图 1

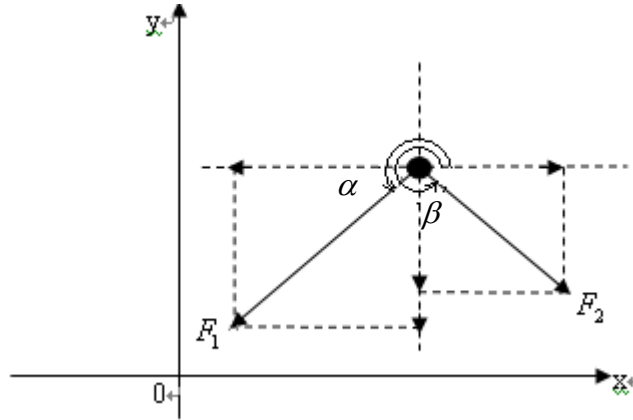


图 1 航天器受力分解图

其中， $F_x = F_2 \cos \beta - F_1 \cos \alpha = F_2 \cos \beta - G \frac{Mm}{x^2(t) + y^2(t)} \frac{x}{\sqrt{x^2(t) + y^2(t)}}$

$F_y = F_2 \sin \beta + F_1 \sin \alpha = F_2 \sin \beta + G \frac{Mm}{x^2(t) + y^2(t)} \frac{y}{\sqrt{x^2(t) + y^2(t)}}$

初始条件为

$$\begin{cases} x(t_0) = x_0 \\ y(t_0) = y_0 \\ v_x(t_0) = v_{x0} \\ v_y(t_0) = v_{y0} \end{cases}$$

$x(t)$ ,  $y(t)$  都是关于时间的位置函数, 航天器在  $x$  方向的分速度即  $x(t)$  对时间  $t$  求导, 航天器在  $y$  方向的分速度即  $y(t)$  对时间  $t$  求导, 航天器在  $x$  方向的加速度即  $x(t)$  对时间  $t$  求二阶导, 航天器在  $y$  方向的加速度即  $y(t)$  对时间  $t$  求二阶导, 根据牛顿第二定律有方程 (3) 和 (4)。由此建立的航天器模型如下:

$$\begin{cases} x'(t) = v_x \\ y'(t) = v_y \\ v'_x(t) = \frac{F_x}{m} = \frac{F_2 \cos \beta}{m} - G \frac{M}{x^2(t) + y^2(t)} \frac{x(t)}{\sqrt{x^2(t) + y^2(t)}} \\ v'_y(t) = \frac{F_y}{m} = \frac{F_2 \sin \beta}{m} - G \frac{M}{x^2(t) + y^2(t)} \frac{y(t)}{\sqrt{x^2(t) + y^2(t)}} \end{cases} \quad (1)$$

显然这样的数学模型在精度上是远远不能满足实际需要的, 在其他要求精确制导等有关高科技的实际问题中, 我们都面临着类似的问题: 我们必须建立高精度的数学模型, 必须高精度地估计模型中的大批参数, 因为只有这样的数学模型才能解决实际问题, 而不会出现差之毫厘, 结果却失之千里的情况。这时所建立数学模型的精度就成了数学模型的生命线。例如上述问题中的航天器还要受到地球质量分布不均匀所引起的摄动力, 大气阻力, 日、月及其它星球的摄动引力的影响, 以及航天器发动机为调整航天器自身姿态运作时作用力的影响。这样不但数学模型十分复杂, 而且在这些数学模型中还要涉及到许多重要的参数, 如地球的引力场模型就有许多待定参数。不仅如此, 在对航天器进行测量时, 还涉及到观测站的地理位置以及设备的系统误差等参数。为此人们要设法利用长期积累的丰富的观测资料, 高精度确定这些重要的参数。

由于航天器的问题太复杂, 下面本题仅考虑较简单的确定高精度参数问题。

### 三、捕食者与被捕食者生态系统问题的分析

题中假设有一个生态系统, 含有两种生物, A 生物和 B 生物, A 生物是捕食者, B 生物是被捕食者。假设  $t$  时刻捕食者 A 的数目为  $x(t)$ , 被捕食者 B 数目为  $y(t)$ , 它们之间满足以下变化规律:

$$\begin{cases} x'(t) = x(t)[\alpha_1 + \alpha_2 y(t)] \\ y'(t) = y(t)[\alpha_3 + \alpha_4 x(t)] \end{cases}$$

初始条件为：

$$\begin{cases} x(t_0) = \alpha_5 \\ y(t_0) = \alpha_6 \end{cases}$$

该模型中， $\alpha_1$  捕食者独自存在时死亡率， $\alpha_1 < 0$ ； $\alpha_2$  被捕食者对捕食者的供养能力  $\alpha_2 > 0$ ； $\alpha_3$  是被捕食者的独立生存增长率， $\alpha_3 > 0$ ； $\alpha_4$  是捕食者掠取被捕食者的能力， $\alpha_4 < 0$ 。<sup>[2]</sup>

这个方程就是生态系统中被捕食者与捕食者的 volterra 模型， $\alpha_k (1 \leq k \leq 6)$  为模型的待定参数。对于该模型理论上不存在解析解，因此我们不能通过参数拟合确定模型的参数。Volterra 模型在给定参数和初始值的情形下可以采用数值积分获得任意时间点的数值解。

根据 volterra 模型进行一些公式推导如下：

$$\begin{cases} \frac{dx(t)}{dt} = x(t)[\alpha_1 + \alpha_2 y(t)] \\ \frac{dy(t)}{dt} = y(t)[\alpha_3 + \alpha_4 x(t)] \end{cases} \quad (2)$$

两个方程相除得：

$$\frac{dy(t)}{dx(t)} = \frac{y(t)[\alpha_3 + \alpha_4 x(t)]}{x(t)[\alpha_1 + \alpha_2 y(t)]}$$

移项得：

$$\frac{[\alpha_1 + \alpha_2 y(t)]}{y(t)} dy(t) = \frac{[\alpha_3 + \alpha_4 x(t)]}{x(t)} dx(t)$$

两边积分：

$$\int_{y_0}^y \frac{[\alpha_1 + \alpha_2 y(t)]}{y(t)} dy(t) = \int_{x_0}^x \frac{[\alpha_3 + \alpha_4 x(t)]}{x(t)} dx(t) \quad (3)$$

得到相轨方程：

$$\alpha_1(\ln y - \ln y_0) + \alpha_2(y - y_0) - \alpha_3(\ln x - \ln x_0) - \alpha_4(x - x_0) = 0 \quad (4)$$

移项得：

$$\alpha_1 \ln y + \alpha_2 y - \alpha_3 \ln x - \alpha_4 x = \alpha_1 \ln y_0 + \alpha_2 y_0 - \alpha_3 \ln x_0 - \alpha_4 x_0$$

该式右边为

$$\alpha_1 \ln y_0 + \alpha_2 y_0 - \alpha_3 \ln x_0 - \alpha_4 x_0$$

只与系统初始状态有关，令

$$\alpha_1 \ln y_0 + \alpha_2 y_0 - \alpha_3 \ln x_0 - \alpha_4 x_0 = C \quad (5)$$

易知  $C \neq 0$ ，将 (5) 式代入 (4) 式得到

$$\alpha_1 \ln y + \alpha_2 y - \alpha_3 \ln x - \alpha_4 x = C \quad (6)$$

式中方程两边同除以  $C$ ，得：

$$\frac{\alpha_1}{C} \ln y + \frac{\alpha_2}{C} y - \frac{\alpha_3}{C} \ln x - \frac{\alpha_4}{C} x = 1 \quad (7)$$

在 (7) 式中，令

$$\alpha'_1 = \frac{\alpha_1}{C}, \quad \alpha'_2 = \frac{\alpha_2}{C}, \quad \alpha'_3 = \frac{\alpha_3}{C}, \quad \alpha'_4 = \frac{\alpha_4}{C}$$

得到

$$\alpha'_1 \ln y + \alpha'_2 y - \alpha'_3 \ln x - \alpha'_4 x = 1$$

这一方程体现了 Volterra 模型中两个变量之间的变化关系，我们称此方程为相轨方程。进一步研究相轨方程，可以发现 Volterra 模型中两个变量呈现周期性变化。

**第一问**，对  $\alpha'_k$  来说，相轨方程是一个 4 未知数的方程，DATA1 中有 6 组数据，用 6 组数据确定 4 个  $\alpha'_k$  可以采用极小范数最小二乘解。又因为  $\alpha_2$  已知， $C$  可求，从而  $\alpha_k$  ( $1 \leq k \leq 4$ ) 可求，由于各观测值真实准确， $\alpha_5, \alpha_6$  可取 DATA1 中任意一组数据，不失一般性，我们取第一组数据为初始值。

**第二问**，我们可以证明参数  $C$  与系统周期成反比，由参考资料可以知道 volterra 模型中  $\alpha_k$  的含义，从而确定  $C$  的正负性，在  $\alpha_2$  未知的情况下，求  $C$  可以从  $C$  与时间的关系入手，我们先在 DATA1 的 6 组数据中取 4 组算出  $\alpha'_k$ ，然后设计一个  $C' = -1$  的新生态系统，以无误差的观测数据 DATA1 为准，设计搜索算法找到与 DATA1 中  $x, y$  值极为接近的数值点，找到对应的观测时间，得到观测间隔，这个观测间隔与 DATA1 中已知的观测间隔一起可以求出  $C$ ，从而得

到  $\alpha_k (1 \leq k \leq 4)$ ,  $\alpha_5$ ,  $\alpha_6$  同第一问。

**第三问**, 用所有数据求得极小范数最小二乘解, 可以确定  $\alpha'_k$ 。经过与第二问类似的方法获得 C。进一步求出  $\alpha_k (1 \leq k \leq 4)$ ,  $\alpha_5$ ,  $\alpha_6$  可取 DATA2 中观测初始时间的值, 这一套  $\alpha_k (1 \leq k \leq 6)$  就是我们所求的最小二乘意义下的最优解。将这一组参数带入 volterra 模型, 获得各观测点上的仿真结果。通过与观测结果比较, 我们发现误差普遍较大。于是我们改进了参数估计模型, 改为求取均方误差意义下的最优解。获得了较好的效果。

**第四问**, 我们采取了第三问中的改良算法, 以求取使 x, y, t 三者均方误差最小的参数组为目标, 进行计算, 然而误差较大。经过判断, 我们认为这是由于时间和数据均存在误差导致搜索结果不够精确, 我们改进搜索算法, 结果大为改善。

## 四、模型的建立及求解

**问题一**,  $\alpha_2$  已知, 求其它 5 个参数  $\alpha_k (k=1,3,4,5,6)$

结合 DATA1.TXT 中 6 组无误差的观测数据 (包括了观测时刻  $t_j$ 、A 生物数目  $x(t_j)$ 、B 生物数目  $y(t_j)$  ), (7) 式含有 4 个未知数, 而题中提供了 6 组数据, 写为矩阵形式即:

$$\begin{bmatrix} \ln y_1 & y_1 & -\ln x_1 & -x_1 \\ \ln y_2 & y_2 & -\ln x_2 & -x_2 \\ \ln y_3 & y_3 & -\ln x_3 & -x_3 \\ \ln y_4 & y_4 & -\ln x_4 & -x_4 \\ \ln y_5 & y_5 & -\ln x_5 & -x_5 \\ \ln y_6 & y_6 & -\ln x_6 & -x_6 \end{bmatrix} \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \alpha'_3 \\ \alpha'_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

记为矩阵形式  $\mathbf{AX} = \mathbf{b}$ , 其中  $\mathbf{A} = \begin{bmatrix} \ln y_1 & y_1 & -\ln x_1 & -x_1 \\ \ln y_2 & y_2 & -\ln x_2 & -x_2 \\ \ln y_3 & y_3 & -\ln x_3 & -x_3 \\ \ln y_4 & y_4 & -\ln x_4 & -x_4 \\ \ln y_5 & y_5 & -\ln x_5 & -x_5 \\ \ln y_6 & y_6 & -\ln x_6 & -x_6 \end{bmatrix}$ ,  $\mathbf{X} = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \alpha'_3 \\ \alpha'_4 \end{bmatrix}$ ,  $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ ,

矛盾方程组  $\mathbf{AX} = \mathbf{b}$  的惟一极小范数最小二乘解为  $\mathbf{X} = \mathbf{A}^+ \mathbf{b}$ , 采用极小范数最小二乘解, 得到  $\alpha'_k (1 \leq k \leq 4)$  的值, 如表 1 所示

表 1 最小二乘的  $\alpha'_k (1 \leq k \leq 4)$  的值

$\alpha'_1$	$\alpha'_2$	$\alpha'_3$	$\alpha'_4$
0.14471560254457	-0.01447146313960	-0.86831792402425	0.07236037259773

因为  $\alpha'_2 = \frac{\alpha_2}{C}$ ，由题意  $\alpha_2 = \frac{1}{5}$ ，而从 **volterra** 模型本身出发， $\alpha'_2$  是捕食者掠取被捕食者的能力，所以利用 **DATA1** 中数据算出的  $\alpha'_2 < 0$ ，所以  $C < 0$ ，这一问题中  $C = -13.82030262390574$ ，把  $C$  代入

$$\alpha'_1 = \frac{\alpha_1}{C}, \quad \alpha'_3 = \frac{\alpha_3}{C}, \quad \alpha'_4 = \frac{\alpha_4}{C}$$

得到

表 2  $\alpha_2$  已知， $\alpha_k (k=1,3,4)$  的值

$\alpha_1 = -2.00001342156688$	$\alpha_3 = 12.00041648377672$	$\alpha_4 = -1.00004224727917$
--------------------------------	--------------------------------	--------------------------------

对于  $\alpha_5, \alpha_6$  来说， $x(t_0) = \alpha_5, y(t_0) = \alpha_6$ ，由观测数据  $x(t_j), y(t_j)$  已知，而  $t_0$  未知，所以， $\alpha_5, \alpha_6$ ，可以是观测数据中任意一组  $x(t_j), y(t_j)$  的值。不失一般性我们取  $t_0 = 0$ ，由 **DATA1** 知道， $x(t_0) = 10, y(t_0) = 60$ ，所以  $\alpha_5 = 10, \alpha_6 = 60$ 。

**问题二**， $\alpha_2$  未知，至少需要几组数据才能确定  $\alpha_k (1 \leq k \leq 6)$  的值

### 1. $\alpha'_k$ 的求解

由问题一的分析可知，为了求取  $\alpha'_k$ ，至少需要四个方程，即四组观测数据。列为线性方程

$$\begin{bmatrix} \ln y_1 & y_1 & -\ln x_1 & -x_1 \\ \ln y_2 & y_2 & -\ln x_2 & -x_2 \\ \ln y_3 & y_3 & -\ln x_3 & -x_3 \\ \ln y_4 & y_4 & -\ln x_4 & -x_4 \end{bmatrix} \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \alpha'_3 \\ \alpha'_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

可以解得

$$\alpha'_1 = \frac{\alpha_1}{C}, \quad \alpha'_2 = \frac{\alpha_2}{C}, \quad \alpha'_3 = \frac{\alpha_3}{C}, \quad \alpha'_4 = \frac{\alpha_4}{C}$$

### 2. 观测系统时间变换分析

对于同样系统的观测，当选取的观测时间起点和观测时间单位不同时，得到同一物理系统的两种不同时间观测结果。不失一般性，我们将两种观测之间的时间变换关系表示为  $t_2 = at_1 + b$ ，其中  $t_1, t_2$  表示两种观测系统的时间单位数量， $b$  反映了两个观测系统时间起点上的差异，而  $a$  反映了两个观测系统观测时间单位的比例，线性关系是由时间的均匀性确定的。

**定理一：**对于参数  $\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4$  完全相同的生态系统，其系统常数  $C$  与观测时间单位长度  $k$  存在反比例关系，即  $Ck = \text{Const}$ 。

证明：在 volterra 模型中：

$$\frac{dx(t)}{dt} = x(t)[\alpha_1 + \alpha_2 y(t)]$$

把  $\alpha_1 = C\alpha'_1, \alpha_2 = C\alpha'_2$  代入得：

$$\frac{dx(t)}{dt} = x(t)[C\alpha'_1 + C\alpha'_2 y(t)] \quad (8)$$

构建对于上述生态过程的两个观测系统，其时间轴变换关系为  $t' = kt$ ， $k$  为常数。(8) 即为

$$\frac{dx(t)}{dt'} = x(t)[C'\alpha'_1 + C'\alpha'_2 y(t)] \quad (9)$$

对 (9) 进行推导

$$\frac{dx(t)}{kdt} = x(t)[C'\alpha'_1 + C'\alpha'_2 y(t)]$$

$$\frac{dx(t)}{dt} = x(t)[kC'\alpha'_1 + kC'\alpha'_2 y(t)] \quad (10)$$

所以， $C = kC'$ ， $k$  为常数，结合  $t' = kt$ ，得到  $Ct = C't'$ ，这就证明了在参数  $\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4$  完全相同的生态系统，其系统常数  $C$  与观测时间间隔参数  $k$  存在

反比例关系，即  $Ca = \text{Const}$ 。在 volterra 模型中，对  $\frac{dy(t)}{dt} = y(t)[\alpha_3 + \alpha_4 x(t)]$

的推导与上面相同，结论也相同。

### 3. $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 求解

在  $\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4$  确定的情况下，我们只要得到系统常数  $C$  就可以确定生态系统参数  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 。

对于一个生态系统，当  $\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4$  确定时，由方程



$\alpha'_1 \ln y + \alpha'_2 y - \alpha'_3 \ln x - \alpha'_4 x = 1$  可知相轨线是完全确定的。对于观测数据中的相邻两组数据  $(t_i, x(t_i), y(t_i))$  和  $(t_{i+1}, x(t_{i+1}), y(t_{i+1}))$ ，其演变过程遵循系统方程

$$\begin{cases} \frac{dx(t)}{dt} = x(t)[\alpha'_1 + \alpha'_2 y(t)] \\ \frac{dy(t)}{dt} = y(t)[\alpha'_3 + \alpha'_4 x(t)] \end{cases}$$

即选择  $(x(t_i), y(t_i))$  作为起始点根据系统方程演化到  $(x(t_{i+1}), y(t_{i+1}))$ 。当  $C$  值不同时，从  $(x(t_i), y(t_i))$  演化到  $(x(t_{i+1}), y(t_{i+1}))$  的过程不同。下面我们在定理一中证明  $C$  值和演化时间存在反比关系。

我们首先用 DATA1 中的 4 组数据确定  $\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4$ ,

表 3  $\alpha'_k (1 \leq k \leq 4)$  的值

$\alpha'_1$	$\alpha'_2$	$\alpha'_3$	$\alpha'_4$
0.14471560254457	-0.01447146313960	-0.86831792402425	0.07236037259773

## 1, 模型的建立与求解

### 龙格-库塔方法求解微分方程

对于 Volterra 模型，没有显式的符号解，因此我们采用四阶龙格-库塔方法求解常微分方程组的数值解。求解方法介绍如下：

volterra 模型可写为

$$\begin{cases} \frac{dx}{dt} = x(t)[\alpha'_1 + \alpha'_2 y(t)] = f_1(t, x, y), & x(t_0) = \alpha_5 \\ \frac{dy}{dt} = y(t)[\alpha'_3 + \alpha'_4 x(t)] = f_2(t, x, y), & y(t_0) = \alpha_6 \end{cases} \quad (11)$$

令下标表示步数，则解此方程组的欧拉方法为

$$\begin{cases} x_{n+1} = x_n + hf_1(t_n, x_n, y_n) \\ y_{n+1} = y_n + hf_2(t_n, x_n, y_n) \end{cases} \quad (12)$$

引进向量记号

$$\mathbf{y} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \quad \mathbf{y}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}, \quad \mathbf{F}_n = \begin{pmatrix} f_{1, n} \\ f_{2, n} \end{pmatrix}$$

则式 (11) 与式 (12) 可分别写成

$$\frac{dy}{dt} = F(x, y), y(t_0) = y_0$$

$$y_{n+1} = y_n + hF_n$$

此时，常用的四阶龙格-库塔方法取形式

$$\begin{cases} y_{n+1} = y_n + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4) \\ K_1 = F(t_n, y_n) \\ K_2 = F(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hK_1) \\ K_3 = F(t_n + \frac{1}{2}h, y_n + \frac{1}{2}hK_2) \\ K_4 = F(t_n + h, y_n + hK_3) \end{cases}$$

采用龙格-库塔方法，我们可以求出微分方程的数值解，数值解十分密集，在图 2 上表现为连续曲线。

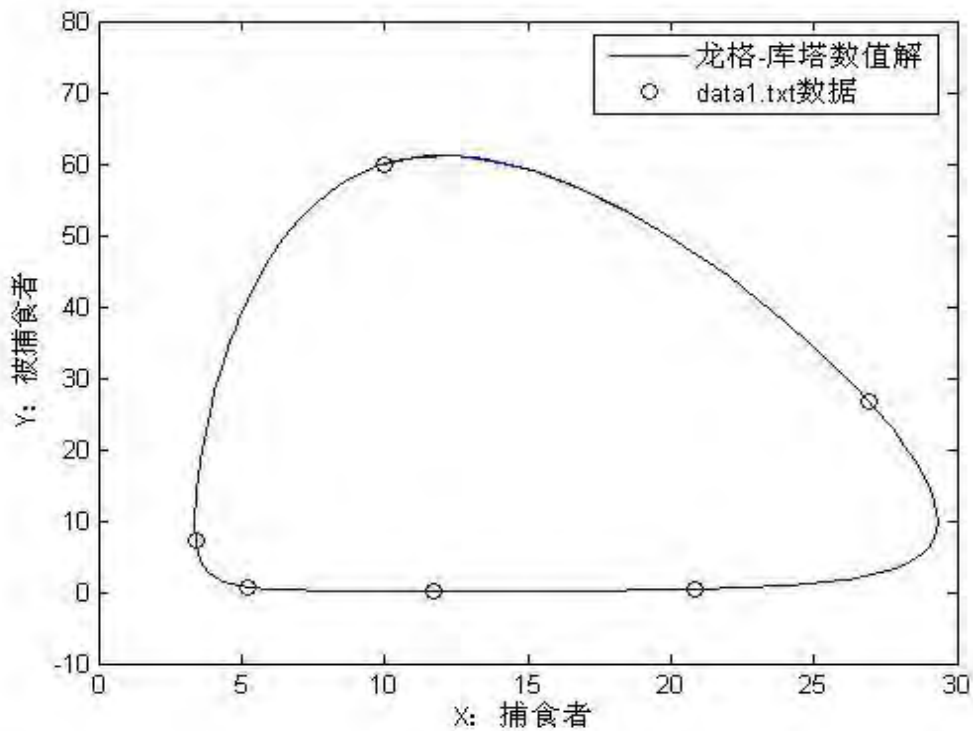


图 2 龙格-库塔数值解与 DATA1 数据对照

## 2, 模型的建立

我们的目标是使用最少数目的观测数据组，即

目标： $n = \min n$  (n为观测数据组的个数)

约束：

$$\begin{cases} \frac{\alpha_1}{C} \ln y_i + \frac{\alpha_2}{C} y_i - \frac{\alpha_3}{C} \ln x_i - \frac{\alpha_4}{C} x_i = 1 \\ \alpha'_i = \frac{\alpha_i}{C} (1 \leq i \leq 4) \\ Ct = C't' \\ q = (\frac{|x'' - x|}{x}, \frac{|y'' - y|}{y}) \end{cases}$$

$t$  为 DATA1 中无误差观测间隔, 为已知,  $t'$  为新系统观测间隔, 这个新观测间隔由后面介绍的搜索算法可以获得,  $C' = -1$ , 所以  $C$  可求,  $\alpha'_i$  由 DATA1 中任意 4 组数据解出, 由  $C$  已求出, 则  $\alpha_i$  可求, 验证  $\alpha_i$  的精度是用求出的  $\alpha_i$  再建立新模型  $v\_new$ , 采用 DATA1 观测间隔得到观测值  $x'', y''$ ,  $q$  为该观测值与 DATA1 观测值的相对误差, 这在结果验证中详细介绍。

### 3, 搜索算法

我们采用一种搜索算法寻找与 DATA1 中无误差数据极为接近的高精度  $x''$ ,  $y''$ , 以此确定新的观测间隔  $t'$ 。

由于观测数据无误差, 我们可以选择 DATA1 中任意一组数据作为微分方程的初值, 根据四阶龙格-库塔算法求解得到微分方程的数值解。不失一般性, 我们选取第一组数据作为初值, 即  $\alpha_5 = 10, \alpha_6 = 60$ 。

以 DATA1 中的  $t_0 = 0, \alpha_5 = 10, \alpha_6 = 60$  为起点进行龙格-库塔数值解的搜索。

**搜索算法:**

第一步: 预估生态系统的周期  $T' = \frac{2\pi}{\sqrt{|\alpha_1 \alpha_3|}}$ ;

第二步: 在预估周期附近搜索真实周期  $T$ ;

第二步: 以  $T/1000$  为步长, 生成 1000 组数据, 搜索其中与目标点距离最近的点; 在此点前后两点构成的区间内重复上述搜索。随搜索深度上升, 获得点的精度随之上升, 我们在计算中考虑到精度和效率, 选择搜索深度为三, 即精确到  $4T/10^9$

### 4, 问题的求解

为了计算  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ , 首先在龙格-库塔数值解基础上搜索出第一组  $x$ ,  $y$  值, 得到观测时间, 然后继续搜索第二组  $x, y$  值, 得到对应的观测时间, 这时两观测时间间隔可求, 如表 4,

表 4 新系统观测间隔

$t'_1$	9.78811239999959	$t'_3$	9.78798039999948
$t'_2$	9.78813999999955	$t'_4$	9.78811159999959
$t'_5$	9.78790879999960		

记表 4 中各个观测间隔为  $t'_i (1 \leq i \leq 5)$ ，它为后一观测时间减去前一观测时间的值。

在已知数据 DATA1 中，各个观测值与观测间隔如表 5 所示，

表 5 DATA1 中无误差的观测数据及观测间隔

$t$	$x$	$y$
none	10.	60.
0.1	11.750840650304518	0.1374480266382216
0.1	3.4133367257849176	7.108705996120129
0.100000000000000004	20.80921881438798	0.4251595082899424
0.09999999999999996	5.231982481945481	0.7182384931413827
0.1	26.925221604818102	26.706603701525214

记表 5 中各个观测间隔为  $t_i (1 \leq i \leq 5)$ ，它为后一观测时间减去前一观测时间的值。

由定理一，我们从理论上知道  $Ct = C't' = -t'$ ，结合表 4，表 5，

$$C = -\frac{\sum_{i=1}^5 \left( \frac{t'_i}{t_i} \right)}{5} = -97.88050639999564$$

又因为  $\alpha_1 = C\alpha'_1$ ， $\alpha_2 = C\alpha'_2$ ， $\alpha_3 = C\alpha'_3$ ， $\alpha_4 = C\alpha'_4$ ，结合表 2， $\alpha_k (1 \leq k \leq 4)$  的值即可确定，因为我们的值与无误差观测值有极小差距，我们的  $\alpha_i$  用  $\alpha''_i$  表示。

表 6  $\alpha''_i$  (即  $\alpha_i$ ) 的值

$\alpha''_1$	$\alpha''_2$	$\alpha''_3$	$\alpha''_4$
-14.16483646104345	1.41647414045314	84.99139811968647	-7.08266991315792

对于  $\alpha_5$ ， $\alpha_6$  来说， $x(t_0) = \alpha_5$ ， $y(t_0) = \alpha_6$ ，由观测数据  $x(t_j)$ ， $y(t_j)$  已知，而  $t_0$  未知，所以， $\alpha_5$ ， $\alpha_6$  可以是观测数据中任意一组  $x(t_j)$ ， $y(t_j)$  的值。我们取  $t_0 = 0$ ，由 DATA1 知道， $x(t_0) = 10$ ， $y(t_0) = 60$ ，所以  $\alpha_5 = 10$ ， $\alpha_6 = 60$ 。

由此，我们通过构建一个新的观测系统找到了  $C$  与观测间隔  $t$  的对应关系，并且除了确定  $\alpha'_1$ ， $\alpha'_2$ ， $\alpha'_3$ ， $\alpha'_4$  需要的 DATA1 中的 4 组数据之外，没有再需要 DATA1 中的其它数据，因此，至少需要 4 组数据就可以确定  $\alpha_k (1 \leq k \leq 6)$  的值。

## 5, 结果验证

为了验证我们的结果, 对于得到的表 6 中的  $\alpha_i''$  的值, 加上  $\alpha_5=10$ ,  $\alpha_6=60$ , 可以再重新构建一个 volterra 模型 (模型 v\_new), 将 DATA1 中的观测时间代入模型 v\_new, 得到  $x''$ ,  $y''$

表 7 模型 v\_new 中取与 DATA1 相同观测间隔对应的 x, y 值

$t$	$x''$	$y''$
NAN	10	60
0.1	11.75093304697659	0.13745135207061
0.1	3.41337698875765	7.10823455966832
0.100000000000000004	20.80928895137538	0.42516779854981
0.099999999999999996	5.23202553781729	0.71817659827730
0.1	26.92525664851230	26.70655846347881

定义相对误差  $q = (\frac{|x''-x|}{x}, \frac{|y''-y|}{y})$ , 其中 x, y 为 DATA1 中无误差的观测数据, 得到相对误差表如表 8 所示。

表 8  $x''$ ,  $y''$  值与 DATA1 的相对误差表 (单位:  $1.0 \times 10^{-4}$ )

$q_x$	$q_y$
0.04997209503443	0.01160449505851
0.07862984004704	0.24194107885441
0.11795781068136	-0.66318181124025
0.03370476711753	0.19499175488534
0.08229360851517	-0.86175921617448
0.01301519248942	-0.01693889904760

由此可见我们得到的  $x''$ ,  $y''$  值具有 5 个数量级以上的精度, 它与无误差观测值是极为接近的。所以我们得到的  $\alpha_k (1 \leq k \leq 6)$  的值是高精度的, 如果需要我们加大搜索深度, 提高数据精度。

**问题三**, 在观测资料有误差 (时间变量不含有误差) 的情况下, 确定参数  $\alpha_k (1 \leq k \leq 6)$  在某种意义下的最优解, 并与仿真结果比较, 进而改进模型。

### 1, 误差数据预处理:

当观测数据有误差时, 我们可以进行如下数据预处理以部分去除数据中误差的干扰。

S1: 根据全体数据对, 按照极小范数最小二乘求解生态系统参数

$$\alpha'_1\text{-all}, \alpha'_2\text{-all}, \alpha'_3\text{-all}, \alpha'_4\text{-all}.$$

S2: 因为只要四组数据就可以求解方程组

$$\begin{bmatrix} \ln y_1 & y_1 & -\ln x_1 & -x_1 \\ \ln y_2 & y_2 & -\ln x_2 & -x_2 \\ \ln y_3 & y_3 & -\ln x_3 & -x_3 \\ \ln y_4 & y_4 & -\ln x_4 & -x_4 \end{bmatrix} \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \alpha'_3 \\ \alpha'_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

从而得到一组系统参数  $\alpha'_1, \alpha'_2, \alpha'_3, \alpha'_4$ 。但是由于误差的影响, 使得这些

参数与总体参数  $\alpha'_{1\_all}, \alpha'_{2\_all}, \alpha'_{3\_all}, \alpha'_{4\_all}$  的偏差程度不同。我们可以

设定偏差门限, 从而去除含有较大误差的数据组。

S3: 对于剩余的数据对, 选择一组作为初始值采用前面提到的搜索算法搜索相邻的下一组值, 得到观测时间间隔;

S4: 计算上一步计算得到的观测时间间隔的中位数, 根据预先设定比例门限, 去除时间间隔不属于中位数某一定义的邻域的数据对。

经过以上步骤, 在一定程度上去除了含有较大误差的数据组, 减少了数据量, 对于后续的优化过程具有重要作用。

## 2, 模型的建立及求解

### (1) DATA2 系统

由第二问求解过程, 我们知道  $\alpha'_k (1 \leq k \leq 4)$  的求解只与数据点有关。我们可以利用第一步数据预处理后的全体数据方程做极小范数最小二乘解, 得到  $\alpha'_k (1 \leq k \leq 4)$ 。

由定理一参数 C 与系统观测间隔成反比, 而 DATA2 系统有给定的观测间隔, 所以与第二问类似, 建立  $C' = -1$  的虚拟系统。假设 DATA2 中观测时间相邻的两点在同一周期内 (第一点为初始值), 则让新系统采用与 DATA2 系统相同的初始值, 经过与第二问类似的龙格-库塔计算及搜索, 搜索的目标是 x, y 在新旧系统对对应值的均方差最小, DATA2 系统的相邻两点的第二点反映到新系统内就得到定理一所示的新系统对应于 DATA2 系统的观测时间, 从而得到新系统观测间隔, 即:

表 9 新系统与 DATA2 系统的对应关系

	DATA2 系统			新系统 $C' = -1$		
第一个观测点	$t_0$	观测间隔 $\Delta t$	观测值 x, y	$t'_0$	观测间隔 $\Delta t'$	搜索值
第二个观测点	$t_1$			$t'_1$		$x_{sou}$ $y_{sou}$

根据定理一, 经过 DATA2 中 150 组数据的运算, C 可求 ( $C = -\frac{\sum_{i=1}^{150} \left( \frac{\Delta t'_i}{\Delta t_i} \right)}{150}$ ),

对这 150 组数据采用最小二乘解,  $\alpha'_k (1 \leq k \leq 4)$  可求。

DATA2 中 150 组数据点和龙格-库塔数值解的相轨图如图 3，

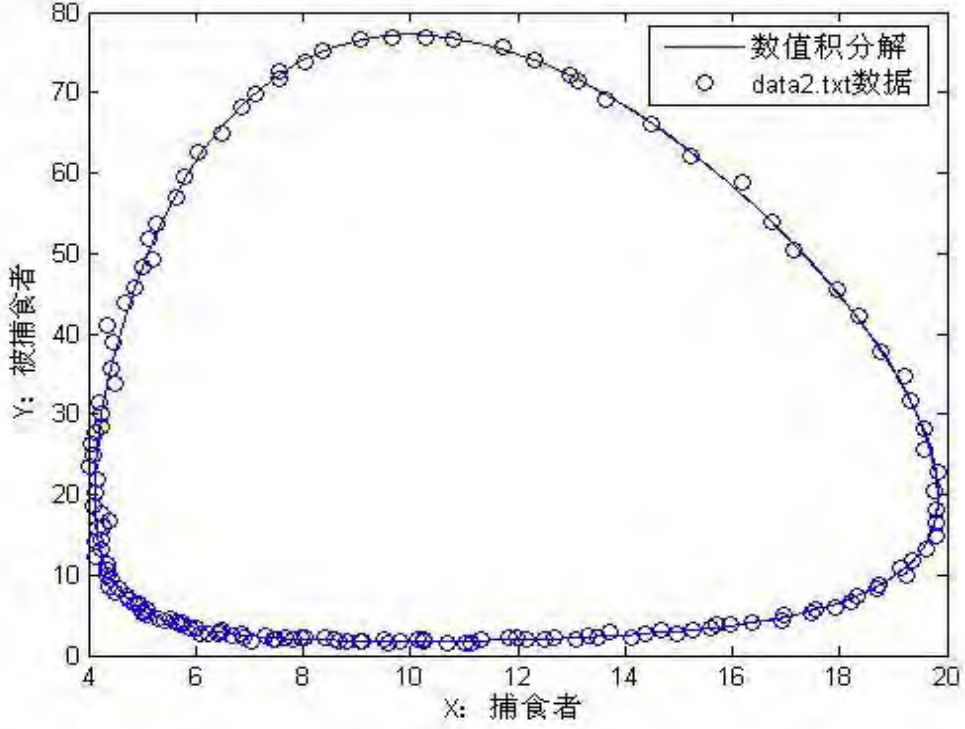


图 3 DATA2 中 150 组数据点和 DATA2 系统的龙格-库塔数值解  
我们的目标是让  $x, y$  的均方误差最小，由此建立目标规划模型

目标：

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] / N}$$

约束：

$$\left\{ \begin{array}{l} \ln y_i \alpha'_1 + y_i \alpha'_2 - \ln x_i \alpha'_3 - x_i \alpha'_4 = 1 \quad (1 \leq n \leq 150) \\ \alpha'_k = \frac{\alpha_k}{C} \quad (k=1, \dots, 4) \\ C = -\frac{\sum_{i=1}^{150} \left( \frac{\Delta t'_i}{t_i} \right)}{150} \\ \Delta t'_n = g(x(t_n), y(t_n), x(t_{n+1}), y(t_{n+1})) \text{ 搜索得到的观测间隔,} \\ \quad \quad \quad t_n \text{ 为 DATA2 给定的观测间隔} \\ \mathbf{y} = \begin{pmatrix} \hat{x}_i \\ \hat{y}_i \end{pmatrix}, \mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \mathbf{F}_n = \begin{pmatrix} f_{1, n} \\ f_{2, n} \end{pmatrix}, \mathbf{y}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \\ \frac{d\mathbf{y}}{dt} = \mathbf{F}(x, y), \mathbf{y}(t_0) = \mathbf{y}_0, \text{ 该初值取 DATA2 中第一点的值} \\ \mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{F}_n \end{array} \right.$$

通过这个模型，我们可以直接求出满足它的  $C$  值， $C = -214.331054077302$ ，  
这个值是没有经过优化的，DATA2 系统可直接解出它的极小范数最小二乘解：

$$\alpha_1' = 0.14110584537284, \quad \alpha_2' = -0.00707342152192$$

$$\alpha_3' = -0.71664088731583, \quad \alpha_4' = 0.07169148047485$$

根据  $\alpha_k = C\alpha_k'$ ,  $\alpha_k (1 \leq k \leq 6)$  最优解如表 7 所示, (我们将  $\alpha_k$  记为  $\alpha_k''$ ), 我们定 DATA2 中的初始点  $x(t_0) = \alpha_5$ ,  $y(t_0) = \alpha_6$

表 10 DATA2 系统的最优解

$\alpha_1'' = -30.243364575230,$	$\alpha_2'' = 1.516053890726,$
$\alpha_3'' = 153.598396773296,$	$\alpha_4'' = -15.365710578537$
$\alpha_5 = x(t_0) = 12.9622856330353,$	$\alpha_6 = y(t_0) = 72.1230158334478$

## (2) DATA3 系统

DATA3 中对  $n=1500$  建立与之相同的目标规划模型, 也直接求出满足它的 C 值,  $C = -2443.11748423156$ , 这个值是没有经过优化的, DATA3 系统也可直接解出它的极小范数最小二乘解:

$$\alpha_1' = 0.12093924694628, \quad \alpha_2' = -0.00616119838760$$

$$\alpha_3' = -0.73882548606458, \quad \alpha_4' = 0.07346474303084$$

根据  $\alpha_k = C\alpha_k'$ ,  $\alpha_k (1 \leq k \leq 6)$  最优解如表 8 所示, (我们将  $\alpha_k$  记为  $\alpha_k''$ ), 我们同样定 DATA3 中的初始点  $x(t_0) = \alpha_5$ ,  $y(t_0) = \alpha_6$

表 11 DATA3 系统的最优解

$\alpha_1'' = -295.468788744257$	$\alpha_2'' = 15.0525315045727$
$\alpha_3'' = 1805.03746280025$	$\alpha_4'' = -179.482998173218$
$\alpha_5 = x(t_0) = 12.8230578534927$	$\alpha_6 = y(t_0) = 73.4072806908665$

## 3, 所求模型与仿真结果比较

DATA2 中, 对求出的最优  $\alpha_k (1 \leq k \leq 6)$  建立生态系统, 把无误差的观测时间代入该系统, 可以发现在无误差观测时间下  $x$ ,  $y$  的值与 DATA2 中  $x$ ,  $y$  的值有较大误差, 目标  $P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]} / N = 7.37696260317394$ , 同样,



对 DATA3，对求出的最优  $\alpha_k (1 \leq k \leq 6)$  建立生态系统，把无误差的观测时间代入该系统，可以发现在无误差观测时间下  $x, y$  的值与 DATA3 中  $x, y$  的值的均方差  $P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]} / N = 33.47612623110470$ ，也非常大，所以表 7，表 8 中的值是非常不精确的，下面设计改良模型的方法。

#### 4，模型的改良

##### (1) DATA3 系统

首先将 DATA3 中的初始点  $x(t_0) = \alpha_5, y(t_0) = \alpha_6$  固定，修正新系统的 C，取区间  $[C_1 = 80\%C, C_2 = 120\%C]$ ，采用二分法进行搜索，当二分法的搜索精度保证 C 精确到  $10^{-4}$  时就停止搜索，找到满足该搜索目标的  $C''$  作为改良的 C 值，这时的均方差

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]} / N = 1.19206591140184,$$

$$C'' = -2425.449031106557$$

确定  $C''$  值，反过来再修正  $\alpha_5, \alpha_6$ ，这里就包含了迭代的思想，由于 C 一定，且  $\alpha'_k (1 \leq k \leq 4)$  不变，则  $\alpha_k (1 \leq k \leq 4)$  一定，我们认为  $\alpha_5, \alpha_6$  的修正值在 DATA3 的初始值附近，给定  $\alpha_5, \alpha_6$  的  $\pm 20\%$ （同 C）为区间，对于各个不同的  $\alpha_5, \alpha_6$  ——  $\alpha_k (1 \leq k \leq 4)$  与  $\alpha_5, \alpha_6$  构成不同系统，寻找使  $\min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]} / N$  最小的系统，这通过对  $x(t_0), y(t_0)$  调用 MATLAB 语句做最小二乘曲线拟合来实现，这个系统对应的初值就是  $\alpha_5, \alpha_6$  的修正值。

$$\alpha_5 = 12.93604597845705, \quad \alpha_6 = 73.00417942616193$$

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]} / N = 1.18427007187202$$

综上所述：

未优化的均方差

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]} / N = 33.47612623110470$$

首先优化 C 值，优化后

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] / N} = 1.19206591140184,$$

$$C'' = -2425.449031106557$$

然后优化初值，优化后

$$\alpha_5 = 12.93604597845705, \quad \alpha_6 = 73.00417942616193$$

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] / N} = 1.18427007187202$$

P 值在第一次优化之后的显著下降，在第二次优化之后进一步下降，从而可以知道这种优化方法是非常好的。

至此，我们可以建立改良模型如下：

目标：

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] / N}$$

约束：

$$\left\{ \begin{array}{l} \begin{bmatrix} \ln y_1 & y_1 & -\ln x_1 & -x_1 \\ \ln y_2 & y_2 & -\ln x_2 & -x_2 \\ \vdots & \vdots & \vdots & \vdots \\ \ln y_n & y_n & -\ln x_n & -x_n \end{bmatrix} \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \alpha'_3 \\ \alpha'_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, (1 \leq n \leq 150) \\ \alpha'_k = \frac{\alpha_k}{C} (k = 1, \dots, 4) \\ C = -\frac{\sum_{i=1}^{150} \left( \frac{\Delta t'_i}{t_i} \right)}{150} \\ C'' \in [80\% C, 120\% C] \\ \alpha''_5 \in [80\% \alpha_5, 120\% \alpha_5] \\ \alpha''_6 \in [80\% \alpha_6, 120\% \alpha_6] \\ \Delta t'_n = g(x(t_n), y(t_n), x(t_{n+1}), y(t_{n+1})) \text{ 搜索得到的观测间隔,} \\ t_n \text{ 为 DATA2 给定的观测间隔} \\ \mathbf{y} = \begin{pmatrix} \hat{x}_i \\ \hat{y}_i \end{pmatrix}, \mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \mathbf{F}_n = \begin{pmatrix} f_{1, n} \\ f_{2, n} \end{pmatrix}, \mathbf{y}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \\ \frac{d\mathbf{y}}{dt} = \mathbf{F}(x, y), \mathbf{y}(t_0) = \mathbf{y}_0, \text{ 该初值取 DATA2 中第一点的值} \\ \mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{F}_n \end{array} \right.$$

## (2) DATA2 系统

DATA2 系统采用与 DATA3 系统相同的优化方法，计算得

$$C'' = -214.331054077302$$

$$\alpha_5 = 12.9622856330353, \quad \alpha_6 = 72.1230158334478$$

$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] / N} = 0.25392540697850$$

**问题四**，假设连观测资料的时间变量也含有误差，确定参数  $\alpha_k (1 \leq k \leq 6)$  在某种意义下的最优解。

当观测资料的时间变量也含有误差时，使得观测时间的间隔含有更大的误差。如上所述，求解  $\alpha'_k (1 \leq k \leq 4)$  时，相当于求解相轨线方程，这与时间变量没有关系，因此这四个参数的求解精度仅受观测数据误差的影响。同问题三的求解过程，我们在得到  $\alpha'_k (1 \leq k \leq 4)$  时，选取  $C' = -1$  的新的观测系统，则这是一个除初值外均确定的系统。设  $\Delta t'_n = g(x(t_n), y(t_n), x(t_{n+1}), y(t_{n+1}))$  为搜索得到的时间间隔，该时间间隔与这两组数据有关，搜索结果为时间间隔  $\Delta t'_n$  与  $t_{n+1} - t_n$  之间的对

应关系，如果时间没有误差，我们可以直接计算  $C = \frac{\Delta t'_n}{t_{n+1} - t_n}$ ，但是由于时间误差

的存在，直接计算将会把时间上的误差以倒数形式传递到  $C$  上，当  $t_{n+1} - t_n$  较小时即使很小的时间误差都会对结果造成巨大的影响。因此我们对以上常数计算方法进行改进，提出了一下新的计算方法以降低误差的影响。

新的常数  $C$  计算方法为  $C = \frac{\sum \Delta t'_n}{\sum (t_{n+1} - t_n)} = \frac{\sum \Delta t'_n}{t_{n+1} - t_0}$ 。即选取连续若干组数据，

每次均是从前一组搜索相邻下一组数据，从上式可以看出经过这样的处理后，分母上是一段比较长的时间间隔，两个时间的误差对间隔的影响将会大大降低，从而使得计算得到的常数  $C$  更加稳定。

实际处理中还可以多次分组，对已经比较稳定的常数  $C$  进行最小二乘拟合以进一步降低误差水平，提高数据精度。

通过以上方法确定了常数  $C$ ，下面考虑优化初值，方法同问题三求解过程中的循环优化方法。

因为时间，观测值均含有误差，因此我们进行参数优化高精度求解的目标不应该仅仅包含观测数据的均方误差，还应该包含时间的均方误差，即均方误差和

为  $\sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (t_{ni} - \hat{t}_{ni})^2] / N}$ ， $t_{ni}$  为 DATA4 系统时间观测值， $\hat{t}_{ni}$  为

观测时间的估计值。

根据以上分析，我们建立时间含有误差下的单目标非线性规划模型

目标：
$$P = \min \sqrt{\sum_{i=1}^N [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (t_{ni} - \hat{t}_{ni})^2]} / N$$

约束：

$$\begin{cases} (\ln y_i)\alpha'_1 + y_i\alpha'_2 - (\ln x_i)\alpha'_3 - x_i\alpha'_4 = 1 (i=1, \dots, M) \\ \alpha'_k = \frac{\alpha_k}{C} (k=1, \dots, 4) \\ C = \frac{\sum \Delta t'_n}{\sum (t_{n+1} - t_n)} = \frac{\sum \Delta t'_n}{t_{n+1} - t_0} \\ \Delta t'_n = g(x(t_n), y(t_n), x(t_{n+1}), y(t_{n+1})) \\ \mathbf{y} = \begin{pmatrix} \hat{x}_i \\ \hat{y}_i \end{pmatrix}, \mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \mathbf{F}_n = \begin{pmatrix} f_{1, n} \\ f_{2, n} \end{pmatrix}, \mathbf{y}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \\ \frac{d\mathbf{y}}{dt} = \mathbf{F}(x, y), \mathbf{y}(t_0) = \mathbf{y}_0 \\ \mathbf{y}_{n+1} = \mathbf{y}_n + h\mathbf{F}_n \end{cases}$$

其中  $t_{ni}$  为 DATA4 系统时间观测值， $\hat{t}_{ni}$  为新系统时间观测值

**时间含有误差情形下时间估计算法：**

S1：对于给定的数据组 DATA4，进行数据预处理；

S2：利用预处理结果数据计算  $\alpha'_k (1 \leq k \leq 4)$ ；

S3：预估数据周期 T；

S4：搜索时间间隔，并估计  $C = \frac{\sum \Delta t'_n}{\sum (t_{n+1} - t_n)} = \frac{\sum \Delta t'_n}{t_{n+1} - t_0}$ ；

S5：对 C 和初值进行优化；

S5：代入参数和初始值计算均方误差；

DATA4 系统可直接解出它的极小范数最小二乘解：((

$$\alpha'_1 = 0.14110584537284, \alpha'_2 = -0.00707342152192$$

$$\alpha'_3 = -0.71664088731583, \alpha'_4 = 0.07169148047485))$$

通过这个模型，我们可以直接求出满足它的 C 值。经过优化获得

$$C'' = -2441.374050383611$$

因为计算时间问题，我们没有进行初始值优化。最终以第一组参数为初始值获得的仿真数据中，两变量和时间的均方差分别为

$$P_{xy} = 0.54933666827330, P = 0.00711664554174$$

综合为

$$P=0.77691196210573$$

该值为取第一组参数为初始值条件下最优。

## 五、模型精度分析及评价

### 模型精度分析：

对于问题一，我们利用全部六组无误差观测数据，根据极小范数最小二乘求解  $\alpha'_k (1 \leq k \leq 4)$ ，在  $\alpha_2$  已知时简单的比例运算得到模型参数。此过程误差只能来自矩阵求逆过程和矩阵乘法运算的截断误差，提高计算过程中数据精度就能够提高模型参数的精度， $\alpha_5, \alpha_6$  直接来自真实的观测数据精度得到自然保证；

对于问题二，利用四组数据得到  $\alpha'_k (1 \leq k \leq 4)$ ，其精度依赖于数据运算中的数据精度。在求解 C 的过程中，我们采用了在微分方程数值解中搜索观测数据从而确定时间间隔的方法。其精度受到微分方程数值解中步长  $h$  和搜索算法的影响，减小步长  $h$ ，提高搜索深度均能够提高 C 的精度，从而提高  $\alpha_k (1 \leq k \leq 4)$  的精度，

$\alpha_5, \alpha_6$  直接来自真实的观测数据精度得到自然保证；

对于问题三，利用多组误差观测数据根据极小范数最小二乘得到  $\alpha'_k (1 \leq k \leq 4)$ ，其精度依赖于数据运算中的数据精度和观测数据中的误差，C 的精度受到微分方程数值解中步长  $h$  和搜索算法的影响。在模型的改进中，我们增加了对于初值和常数 C 的优化过程，由优化过程的最小均方误差目标，我们知道优化过程可以提高参数的准确度和精度；

对于问题四，当观测时间存在误差时， $\alpha'_k (1 \leq k \leq 4)$  的精度依赖于数据运算中的数据精度和观测数据中的误差。在时间间隔搜索过程，我们采用了预估周期校正算法，其数据精度可以通过搜索深度和步长进一步提高。以上我们对各个问题中模型参数的精度进行了定性分析，进一步我们可以通过函数的泰勒级数展开和误差传递进行定量分析。

### 模型评价

如上分析该模型最大的优点是保证了参数估计的高精度，该模型的缺点是，建立的生态系统参数高精度估计模型，对于小幅度噪声具有一定的适应性，但是当噪声强度较大时，该模型得到的结果精度显著下降。

## 六、模型的推广

### 一．生态过程周期的多解性。

在第二，三，四问的求解过程中，我们均采用了固定初始点搜索相邻点的搜索算法，在算法中我们均假设相邻两个数据来自真实生态过程的一个周期，此种假设下我们根据观测数据得到的生态过程周期是所有可能周期的最大值。这里我

们给出生态过程周期的一般表达式。

设该生态系统的真实演变过程周期为  $T$ ，而获取的数据其实是对真实演变过程的采样。这里就涉及了数据处理中一般的采样问题，根据奈奎斯特采样定理，当采样频率大于信号中最高频率时，可以根据采样数据恢复原始数据，进而从采样数据的周期得到信号的真实周期。但一般的数据观测过程不一定满足采样定理，因此同样的观测数据对应着多组真实的周期。对于原始演变过程不同采样频率的采样可能得到相同的采样数据，这是一个一对多的关系，因此不能从采样数据得到准确的演变周期，只能得到最大演变周期，前提是每两个相邻数据均属于一个演变周期，另外可以得到演变周期的一般表达式。如下两组数据均是对于同一过程的观测，初始时刻相同，观测间隔不同。

$$\begin{aligned} & x(t_0), x(t_0 + \Delta T), x(t_0 + 2\Delta T), \dots, x(t_0 + n\Delta T) \\ & x(t_0), x(t_0 + \Delta T'), x(t_0 + 2\Delta T'), \dots, x(t_0 + n\Delta T') \end{aligned}$$

对于以上两个观测，当  $\Delta T' = \Delta T + NT$  时，两组观测结果完全相同，其中  $N$  为整数。

根据观测过程我们有如下关系式：

$$\frac{\Delta T}{\Delta T'} = \frac{\Delta T}{\Delta T + NT} = \frac{T_s}{T_s'}, \quad T_s = \frac{\Delta T}{\Delta T + NT} T_s', \quad T = \frac{\Delta T(T_s' - T_s)}{NT_s} T_s',$$

其中  $T_s, T_s'$  为两种观测情形下得到的观测数据周期， $T_s'$  为生态系统最大演变周期， $T$  为生态系统的一般周期表达式。

## 二. 第三、四问模型优化过程中初始值 $\alpha_5, \alpha_6$ 和系统常数 $C$ 的循环优化过程

在第三问模型优化改进中，我们提出了首先利用观测数据选取初值优化系统常数  $C$  使得总的均方差达到极小，然后利用优化的系统常数  $C$  进一步优化初始参数  $\alpha_5, \alpha_6$  使得总的均方差达到新的极小。

这一过程可以继续循环进行，即反复优化初始参数  $\alpha_5, \alpha_6$  和系统常数  $C$ ，最后当两次优化结果的差异小于设定的高精度门限时，停止循环过程，此时得到了优化  $\alpha_5, \alpha_6$  和  $C$  的最优结果。由于每一次优化过程均要按照龙格-库塔方法求解微分方程组的数值解，故时间复杂度较大。

## 参考文献

- [1] 吕学富，飞行器飞行力学，西北工业大学出版社，1994 年 10 月
- [2] 姜启源，数学模型（第二版），高等教育出版社，1992 年 4 月
- [3] 徐仲，张凯院，陆全，冷国伟，矩阵论简明教程，科学出版社，2005 年 6 月
- [4] 张肇炽，徐仲，代数与几何基础，高等教育出版社，2001 年 6 月

- [5] 王沫然, MATLAB 与科学计算 (第二版), 电子工业出版社, 2003 年 9 月
- [6] 崔祜涛, 史雪岩, 崔平远, 栾恩杰, 绕飞慢自旋小天体的航天器运动分析, 航空学报, 第25卷第1期, 2004年1月
- [7] 封建湖, 车刚明, 聂玉峰, 数值分析原理, 科学出版社, 2001 年 9 月
- [8] 王慧忠, 张新全, Lotka-Volterra 数学模型在草地管理中的应用, 草业学报, 第 15 卷, 第 1 期, 2006 年 2 月