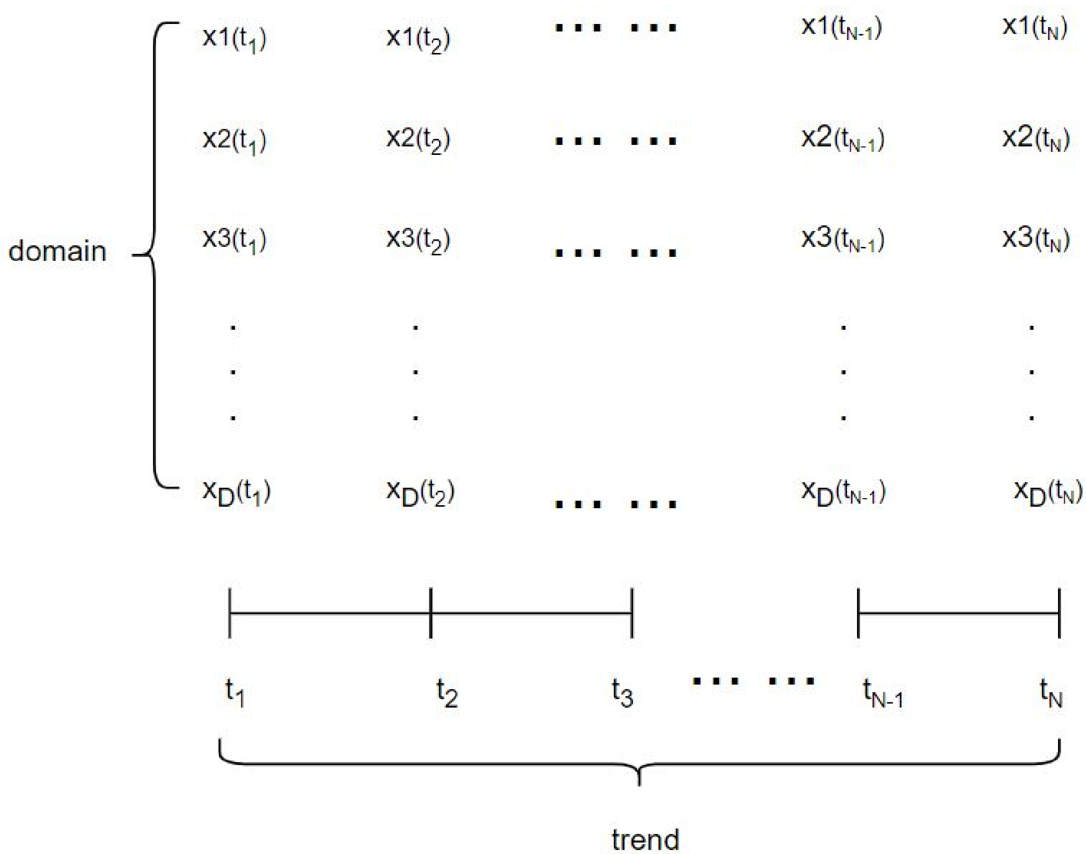
There is the reduced stochastic model(RSM):

$$
\begin{aligned}
dx_i(t) = &\sum_{j,k=i-1}^{i+1} B^{xxx}_{i,j,k} x_j x_k dt + \sum_{j=i-1}^{i+1} M^{xx}_{i,j} x_j dt + \sum_{j,k,l=i-2}^{i+2} C_{i,j,k,l} x_j x_k x_l dt \\
&+ \sigma^{(1)} dW^{(1)}_i + \sigma^{(2)}_i x_i dW^{(2)}_i
\end{aligned}
$$

Remove all {i-2}, {i+2} item in the above RSM. Assume the following reduced reduced stochastic model(RRSM) exist.

$$
\begin{aligned}
dx_i &= \sum_{j=i-1}^{i+1} A_j x_j dt + \sum_{j,k=i-1}^{i+1} B_{j,k} x_j x_k dt + \sum_{j,k,l=i-1}^{i+1} C_{j,k,l} x_j x_k x_l dt + \sigma_1 dW_1 + \sigma_2 x_i dW_2 \\
&= (A_1 x_{i-1} + A_2 x_i + A_3 x_{i+1}) dt \\
&+ (B_1 x_{i-1}^2 + B_2 x_{i-1} x_i + B_3 x_{i-1} x_{i+1} + B_4 x_i^2 + B_5 x_i x_{i+1} + B_6 x_{i+1}^2) dt \\
&+ (C_1 x_{i-1}^3 + C_2 x_{i-1}^2 x_i + C_3 x_{i-1}^2 x_{i+1} + C_4 x_{i-1} x_i^2 + C_5 x_{i-1} x_i x_{i+1} + C_6 x_{i-1} x_{i+1}^2 + C_7 x_i^3 + C_8 x_i^2 x_{i+1} + C_9 x_i x_{i+1}^2 + C_{10} x_{i+1}^3) dt \\
&+ \sigma_1 dW1 + \sigma_2 x_i dW2
\end{aligned}
$$

For each x_i, considering N discrete observations $\{x_i(1), \ldots, x_i(N)\}$ of the sample path $\{x_i(t)\}$ at discrete time $t = k * \Delta t, k = 1, \ldots, N$. Totally, we have observed time series data with the size of $N * D$.



Firstly, the Euler scheme produces the discretization.

$$\begin{aligned}
x_i(t+1)-x_i(t) = &(A_1x_{i-1}(t)+A_2x_i(t)+A_3x_{i+1}(t))\Delta t\\
&+(B_1x_{i-1}^2(t)+B_2x_{i-1}(t)x_i(t)+B_3x_{i-1}(t)x_{i+1}(t)+B_4x_i^2(t)+B_5x_i(t)x_{i+1}(t)+B_6x_{i+1}^2(t))\Delta t\\
&+(C_1x_{i-1}(t)^3+C_2x_{i-1}^2(t)x_i(t)+C_3x_{i-1}^2(t)x_{i+1}(t)+C_4x_{i-1}(t)x_i^2(t)+C_5x_{i-1}(t)x_ivx_{i+1}(t)\\
&+C_6x_{i-1}(t)x_{i+1}^2(t)+C_7x_i^3(t)+C_8x_i^2(t)x_{i+1}(t)+C_9x_i(t)x_{i+1}^2(t)+C_{10}x_{i+1}^3(t))\Delta t\\
&+\sigma_1\Delta W1+\sigma_2x_i(t)\Delta W2
\end{aligned}$$

Since $\sigma_1\Delta W1+\sigma_2x_i(t)\Delta W2\sim N(0,(\sigma_1^2+\sigma_2^2x_i(t)^2)\Delta t)$,we have the transition probability density:

$$\begin{aligned}
&p_\theta(x_i(t+1)|x_{i-1}(t),x_i(t),x_{i+1}(t))\\
&=\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2x_i(t)^2)\Delta t}}exp\{-\frac{[x_i(t+1)-x_i(t)-(\sum_{j=i-1}^{i+1}A_jx_j(t)+\sum_{j,k=i-1}^{i+1}B_{j,k}x_j(t)x_k(t)+\sum_{j,k,l=i-1}^{i+1}C_{j,k,l}x_j(t)x_k(t)x_l(t))\Delta t]^2}{2(\sigma_1^2+\sigma_2^2x_i^2(t))\Delta t}
\end{aligned}$$

The negative log-likelihood function of N observations:

$$\begin{aligned}
&l=-\ln L(\theta|x_1,\dots,x_N)\\
&=\sum_{t=1}^N log\,p_\theta(x_i(t+1)|x_{i-1}(t),x_i(t),x_{i+1}(t))\cdot p_0\\
&=\sum_{t=1}^N\{\frac{[x_i(t+1)-x_i(t)-(\sum_{j=i-1}^{i+1}A_jx_j(t)+\sum_{j,k=i-1}^{i+1}B_{j,k}x_j(t)x_k(t)+\sum_{j,k,l=i-1}^{i+1}C_{j,k,l}x_j(t)x_k(t)x_l(t))\Delta t]^2}{2(\sigma_1^2+\sigma_2^2x_i^2(t))\Delta t}+\frac{1}{2}log(2\pi(\sigma_1^2+\sigma_2^2x_i^2(t))\Delta t)\}\\
&=\sum_{t=1}^N\{\frac{1}{2(\sigma_1^2+\sigma_2^2x_i^2(t))\Delta t}[x_i(t+1)-x_i(t)-(A_1x_{i-1}(t)+A_2x_i(t)+A_3x_{i+1}(t)\\
&+B_1x_{i-1}^2(t)+B_2x_{i-1}(t)x_i(t)+B_3x_{i-1}(t)x_{i+1}(t)+B_4x_i^2(t)+B_5x_i(t)x_{i+1}(t)+B_6x_{i+1}^2(t)\\
&+C_1x_{i-1}^3(t)+C_2x_{i-1}^2(t)x_i(t)+C_3x_{i-1}^2(t)x_{i+1}(t)+C_4x_{i-1}(t)x_i^2(t)+C_5x_{i-1}(t)x_i(t)x_{i+1}(t)\\
&+C_6x_{i-1}(t)x_{i+1}^2(t)+C_7x_i^3(t)+C_8x_i^2(t)x_{i+1}(t)+C_9x_i(t)x_{i+1}^2(t)+C_{10}x_{i+1}^3(t))\Delta t]^2\\
&+\frac{1}{2}log(2\pi(\sigma_1^2+\sigma_2^2x_i^2(t))\Delta t)\}
\end{aligned}$$

The Maximum Likelihood Estimate :

$$\hat\theta=argmax(L(\theta|X_1,\dots,X_N)=argmin(l)$$

For simple donation, at time t, let's mark

$$\begin{aligned}
&\theta=[A_1,A_2,A_3,B_1,B_2,B_3,B_4,B_5,B_6,C_1,C_2,C_3,C_4,C_5,C_6,C_7,C_8,C_9,C_{10}]\\
&Y_i(t)=[x_{i-1},x_i,x_{i+1},x_{i-1}^2,x_{i-1}x_i,x_{i-1}x_{i+1},x_i^2,x_ix_{i+1},x_{i+1}^2,x_{i-1}^3,x_{i-1}^2x_i,x_{i-1}^2x_{i+1},x_{i-1}x_i^2,x_{i-1}x_ix_{i+1},x_{i-1}x_{i+1}^2,x_i^3,x_i^2x_{i+1},x_ix_{i+1}^2,x_{i+1}^3]'\\
&p_i(t)=x_i(t+1)-x_i(t)-[\sum_{j=i-1}^{i+1}A_jx_j(t)+\sum_{j,k=i-1}^{i+1}B_{j,k}x_j(t)x_k(t)+\sum_{j,k,l=i-1}^{i+1}C_{j,k,l}x_j(t)x_k(t)x_l(t)]\Delta t\\
&=x_i(t+1)-x_i(t)-\theta Y_i(t)\Delta t\\
&q_i(t)=\sigma_1^2+\sigma_2^2x_i^2(t)
\end{aligned}$$

Where θ is the ODE parameters and $[\sigma_1, \sigma_2]$is the last 2 parameters. There are totally 21 unknown parameters. We rewrite the negative likelihood function, where p is only in term of first 19 parameters θ, and q is only in term of the last 2 parameters $[\sigma_1, \sigma_2]$. It makes the analysis of the Gaussian matrix and Hessian matrix clearly.

$$l = \sum_{t=1}^{N} \{ \frac{p_i^2}{2q_i\Delta t} + \frac{1}{2} log(2\pi q_i \Delta t) \}$$

Differentiating with respect to each parameters, we get the first differential vector or gradient G:

$$G = \begin{bmatrix} \frac{\partial(l)}{\partial\theta(1)} \\ \frac{\partial(l)}{\partial\theta(2)} \\ \cdots \\ \frac{\partial(l)}{\partial\theta(19)} \\ \frac{\partial(l)}{\partial\sigma_1} \\ \frac{\partial(l)}{\partial\sigma_2} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{N}\{-\frac{p_i(t)}{q_i(t)}x_{i-1}(t)\} \\ \sum_{t=1}^{N}\{-\frac{p_i(t)}{q_i(t)}x_i(t)\} \\ \cdots \\ \sum_{t=1}^{N}\{-\frac{p_i(t)}{q_i(t)}x_{i+1}^3(t)\} \\ \sum_{t=1}^{N}\{[-\frac{p_i^2(t)}{q_i^2(t)\Delta t} + \frac{1}{q_i(t)}]\sigma_1\} \\ \sum_{t=1}^{N}\{[-\frac{p_i^2(t)}{q_i^2(t)*\Delta t} + \frac{1}{q_i(t)}]\sigma_2 x_i^2(t)\} \end{bmatrix}$$

Where $\frac{\partial(l)}{\partial\theta(j)} = \sum_{t=1}^{N}\{-\frac{p_i(t)}{q_i(t)}Y_{i,j}(t)\}$. $j = 1, \ldots, 19$

The Hessian or second derivative:

$$H = \begin{bmatrix} \frac{\partial^2(l)}{\partial^2\theta(1)} & \cdots & \frac{\partial^2(l)}{\partial\theta(1)\partial\theta(19)} & \frac{\partial^2(l)}{\partial\theta(1)\partial\sigma_1} & \frac{\partial^2(l)}{\partial\theta(1)\partial\sigma_2} \\ * & \cdots & \frac{\partial^2(l)}{\partial\theta(2)\partial\theta(19)} & \frac{\partial^2(l)}{\partial\theta(2)\partial\sigma_1} & \frac{\partial^2(l)}{\partial\theta(2)\partial\sigma_2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ * & * & * & \frac{\partial^2(l)}{\partial\theta(19)\partial\sigma_1} & \frac{\partial^2(l)}{\partial\theta(19)\partial\sigma_2} \\ * & * & * & \frac{\partial^2(l)}{\partial^2\sigma_1} & \frac{\partial^2(l)}{\partial\sigma_1\partial\sigma_2} \\ * & * & * & * & \frac{\partial^2(l)}{\partial^2\sigma_2} \end{bmatrix}$$

$$= \begin{bmatrix} \sum\frac{\Delta t}{q_i}x_{i-1}^2 & \cdots & \sum\frac{\Delta t}{q_i}x_{i-1}x_{i+1}^3 & \sum\frac{2p_i\sigma_1}{q_i^2}x_{i-1} & \sum\frac{2p_i\sigma_2x_i^2}{q_i^2}x_{i-1} \\ * & \cdots & \sum\frac{\Delta t}{q_i}x_i x_{i+1}^3 & \sum\frac{2p_i\sigma_1}{q_i^2}x_i & \sum\frac{2p_i\sigma_2x_i^2}{q_i^2}x_i \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ * & * & \sum\frac{\Delta t}{q_i}x_{i+1}^6 & \sum\frac{2p_i\sigma_1}{q_i^2}x_{i+1}^3 & \sum\frac{2p_i\sigma_2x_i^2}{q_i^2}x_{i+1}^3 \\ * & * & * & \sum\{[-\frac{p_i^2(t)}{q_i^2(t)\Delta t} + \frac{1}{q_i(t)}] + [\frac{2p_i^2}{q_i^3(t)} - \frac{1}{q_i^2(t)}]2\sigma_1^2\} & \sum[\frac{2p_i^2}{q_i^3(t)\Delta t} - \frac{1}{q_i^2(t)}]2\sigma_1\sigma_2x_i^2 \\ * & * & * & * & \sum x_i^2(t)\{[-\frac{p_i^2(t)}{q_i^2(t)\Delta t} + \frac{1}{q_i(t)}] + [\frac{2p_i^2}{q_i^3(t)} - \frac{1}{q_i^2(t)}]2\sigma_2^2x_i^2(t)\} \end{bmatrix}$$

H is symmetric. Where $\frac{\partial^2(l)}{\partial\theta(j)\partial\theta(k)} = \sum_{t=1}^{N}\frac{\Delta t}{q_i(t)}Y_{i,j}(t)Y_{i,k}(t)$. $j, k = 1, \ldots, 19$

Until now, we have the function need to minimize , the gradient and the Hessian matrix of function, which is necessary information for optimization algorithm. Basically, we use function "fminunc" with "trust-region" algorithm to minimize negative log-likelihood function to get our maximum likelihood estimation. This algorithm is based on the interior-reflective Newton method. Each iteration involves the approximate solution of a large linear system using the method of preconditioned conjugate gradients (PCG).

Initiation

A large error in the initial estimate can contribute to non-convergence of the algorithm. To overcome this problem, we try to have a nice guess of parameters.

- close to true parameters. for example 1.1*true parameters.

  This is for checking if the "trust-region" algorithm gives the right optimal result without initial problem.

- small random numbers, for example $N(0,1)$

- some guess with assumption

For the third option of initiation, we assume $\sigma_2 = 0$, then we have thee estimate for θ and σ_1

$$\hat{\theta} = [\sum_{t=1}^{N} Y_i(t) \cdot Y_i^{'}(t)]^{-1} \cdot [\sum_{t=1}^{N} (x_i(t+1) - x_i(t)) \cdot Y_i(t)]/\Delta t$$

In $\hat{\theta}$, there are all parameters we need for p_i, so that p_i update to $\hat{p_i}$

$$\hat{\sigma_1} = \pm\sqrt{\frac{\sum_{t=1}^{N} \hat{p_i}^2(t)}{N\Delta t}}$$

Now we apply second assumption and the guess value,

$$\begin{aligned}
l = \sum_{t=1}^{N} \{ & \frac{[x_i(t+1) - x_i(t) - (\sum_{j=i-1}^{i+1} A_j x_j(t) + \sum_{j,k=i-1}^{i+1} B_{j,k} x_j(t) x_k(t) + \sum_{j,k,l=i-1}^{i+1} C_{j,k,l} x_j(t) x_k(t) x_l(t))\Delta t]^2}{2(\sigma_1^2 + \sigma_2^2 x_i^2(t))\Delta t} \\
& + \frac{1}{2} log(2\pi(\sigma_1^2 + \sigma_2^2 x_i^2(t))\Delta t)\} \\
& \approx \frac{\sum_{t=1}^{N} [x_i(t+1) - x_i(t) - (\sum_{j=i-1}^{i+1} A_j x_j(t) + \sum_{j,k=i-1}^{i+1} B_{j,k} x_j(t) x_k(t) + \sum_{j,k,l=i-1}^{i+1} C_{j,k,l} x_j(t) x_k(t) x_l(t))\Delta t]^2}{2(\sigma_1^2 + \sigma_2^2 \sum_{i=1}^{N} x_i^2(t))^2 \Delta t} \\
& + \frac{1}{2} log(2\pi(\sigma_1^2 + \sigma_2^2 \sum_{i=1}^{N} x_i^2(t))\Delta t)
\end{aligned}$$

We have the guess for σ_2

$$\hat{\sigma_2} = \pm\sqrt{\frac{\sum_{t=1}^{N} p_i\hat{(t)}^2 - \hat{\sigma_1}^2 \Delta t}{\sum_{t=1}^{N} x_i^2(t)\Delta t}}$$