There is a one dimensional SDE:

$$dX_t = -aX_t^3dt - gX_tdt + sdW1 + \sigma X_tdW2$$

where

$a, g > 0$;

W1 and W2 are independent Wiener process;

$\theta = [a, g, s, \sigma]'$ is the parameter vector .

Considering N discrete observations $\{X_1, \ldots, X_N\}$ of the sample path $\{X_t\}$ at discrete time $t = k * \Delta t$,$k = 1, \ldots, N$. The method of maximum likelihood finds the values of the model parameter, θ , that maximize the likelihood function,$L(\theta; X_t)$. Intuitively, this selects the parameter values that make the data most probable.

Firstly, the Euler scheme produces the discretization

$$X_{t+\Delta t} - X_t = -aX_t^3\Delta t - gX_t\Delta t + s\Delta W1 + \sigma X_t\Delta W2$$

Since $s\Delta W1 + \sigma X_t\Delta W2 \sim N(0, (s^2 + \sigma^2X_t^2)\Delta t)$, we have

$$X_{t+\Delta t}|X_t \sim N(X_t - aX_t^3\Delta t - gX_t\Delta t, (s^2 + \sigma^2X_t^2)\Delta t)$$

The transition probability density:

$$p_\theta(X_{t+\Delta t}|X_t) = \frac{1}{\sqrt{2\pi(s^2 + \sigma^2X_t^2)\Delta t}} exp\{-\frac{(X_{t+\Delta t} - X_t + aX_t^3\Delta t + gX_t\Delta t)^2}{2(s^2 + \sigma^2X_t^2)\Delta t}\}$$

In practice, it is often convenient to work with the natural logarithm of the likelihood function, called the log-likelihood. An MLE is the same regardless of whether we maximize the likelihood or the log-likelihood, because log is a strictly increasing function. Corresponding to general optimization tool, negative log-likelihood turns to a minimizing problem.

The negative log-likelihood function of N observations:

$$\begin{aligned} l &= -logL(\theta|X_1, \ldots, X_N) \\ &= -\sum_{i=1}^{N} log\, p_\theta(X_i|X_{i-1}) \\ &= \sum_{i=1}^{N}\{\frac{[X_{i+1} - X_i + aX_i^3\Delta t + gX_i\Delta t]^2}{2(s^2 + \sigma^2X_i^2)\Delta t} + \frac{1}{2}log(2\pi(s^2 + \sigma^2X_i^2)\Delta t)\} \end{aligned}$$

The Maximum Likelihood Estimate :

$$\hat{\theta} = argmax(L(\theta|X_1, \ldots, X_N)) = argmin(l(\theta|X_1, \ldots, X_N))$$

Differentiating with respect to a,g,s,σ, we get the first differential vector or gradient G:

$$G = \begin{bmatrix} \frac{\partial(l)}{\partial a} \\ \frac{\partial(l)}{\partial g} \\ \frac{\partial(l)}{\partial s} \\ \frac{\partial(l)}{\partial \sigma} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N}\{\frac{X_i^3(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)}{s^2+\sigma^2X_i^2}\} \\ \sum_{i=1}^{N}\{\frac{X_i(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)}{s^2+\sigma^2X_i^2}\} \\ \sum_{i=1}^{N}\{-\frac{s(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)^2}{(s^2+\sigma^2X_i^2)^2\Delta t} + \frac{s}{s^2+\sigma^2X_i^2}\} \\ \sum_{i=1}^{N}\{-\frac{\sigma X_i^2(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)^2}{(s^2+\sigma^2X_i^2)^2\Delta t} + \frac{\sigma X_i^2}{s^2+\sigma^2X_i^2}\} \end{bmatrix}$$

Staring with easy case. If $\sigma = 0$, maximizing the log-likelihood function.

$$\begin{cases} \frac{\partial l}{\partial a} = \frac{1}{s^2}\cdot\sum_{i=1}^{N}[X(i+1)-X(i)+aX(i)^3\Delta t+gX(i)\Delta t]\cdot X(i)^3 = 0 \\ \frac{\partial l}{\partial g} = \frac{1}{s^2}\cdot\sum_{i=1}^{N}[X(i+1)-X(i)+aX(i)^3\Delta t+gX(i)\Delta t]\cdot X(i) = 0 \\ \frac{\partial l}{\partial s} = \frac{1}{s^3\Delta t}\cdot\sum_{i=1}^{N}[X(i+1)-X(i)+aX(i)^3\Delta t+gX(i)\Delta t]^2 - \frac{N}{s} = 0 \end{cases}$$

Assume $\sigma = 0$, analytical solution for MLE of this SDE:

$$\boxed{\begin{aligned} \begin{bmatrix} \hat{a} \\ \hat{g} \end{bmatrix} &= \begin{bmatrix} \sum_{i=1}^{N} X_i^6, & \sum_{i=1}^{N} X_i^4 \\ \sum_{i=1}^{N} X_i^4, & \sum_{i=1}^{N} X_i^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{i=1}^{N}[X_{i+1}-X_i]\cdot X_i^3 \\ \sum_{i=1}^{N}[X_{i+1}-X_i]\cdot X_i \end{bmatrix} /\Delta t \\ \hat{s} &= \pm\sqrt{\frac{\sum_{i=1}^{N}[X(i+1)-X(i)+\hat{a}X(i)^3\Delta t+\hat{g}X(i)\Delta t]^2}{N\Delta t}} \end{aligned}}$$

If $\sigma \neq 0$, maximizing the log-likelihood function analytically is difficult. It can only be found via numerical global optimization.

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 θ_0, the first derivative close to zero.

Firstly, we assume

$$\begin{aligned} l &= \sum_{i=1}^{N}\{\frac{[X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t]^2}{2(s^2+\sigma^2X_i^2)\Delta t} + \frac{1}{2}log(2\pi(s^2+\sigma^2X_i^2)\Delta t)\} \\ &\approx \frac{\sum_{i=1}^{N}[X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t]^2}{2(s^2+\sigma^2\sum_{i=1}^{N}X_i^2)\Delta t} + \frac{1}{2}log(2\pi(s^2+\sigma^2\sum_{i=1}^{N}X_i^2)\Delta t) \end{aligned}$$

Then we have the estimate for a,g

$$\begin{bmatrix} \hat{a_0} \\ \hat{g_0} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} X_i^6, & \sum_{i=1}^{N} X_i^4 \\ \sum_{i=1}^{N} X_i^4, & \sum_{i=1}^{N} X_i^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{i=1}^{N}[X_{i+1}-X_i]\cdot X_i^3 \\ \sum_{i=1}^{N}[X_{i+1}-X_i]\cdot X_i \end{bmatrix} /\Delta t$$

Which is as same as the special case $\sigma = 0$.

IF we use of MLE of s when $\sigma = 0$,we have

$$\hat{s_0} = \pm\sqrt{\frac{\sum_{i=1}^{N}[X_{i+1} - X_i + \hat{a_0}X_i^3\Delta t + \hat{g_0}X_i)\Delta t]^2}{N\Delta t}}$$

After we have the guess for a,g,s, plug in the approximation assumption,

$$\hat{\sigma_0} = \pm\sqrt{\frac{\sum_{i=1}^{N}(X_{i+1} - X_i + \hat{a_0}X_i^3\Delta t + \hat{g_0}X_i\Delta t)^2 - \hat{s_0}^2\Delta t}{\sum_{i=1}^{N}X_i^2\Delta t}}$$

In next step, We use the optimization tool (function @fminunc with trust region)in Matlab to minimize l.

In "true region" formula , we need to provide G and the Hessian function or second derivative:

$$H = l^{''}(\theta)$$

$$= \sum_{i=1}^{N}\begin{bmatrix} \frac{X_i^6\Delta t}{s^2+\sigma^2X_i^2} & \frac{X_i^4\Delta t}{s^2+\sigma^2X_i^2} & \frac{-2sX_i^3(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)}{(s^2+\sigma^2X_i^2)^2} & \frac{-2\sigma X_i^5(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)}{(s^2+\sigma^2X_i^2)^2} \\ * & \frac{X_i^2\Delta t}{s^2+\sigma^2X_i^2} & \frac{-2sX_i(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)}{(s^2+\sigma^2X_i^2)^2} & \frac{-2\sigma X_i^3(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)}{(s^2+\sigma^2X_i^2)^2} \\ * & * & \frac{(3s^2-\sigma^2X_i^2)(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)^2}{(s^2+\sigma^2X_i^2)^3\Delta t} + \frac{-s^2+\sigma^2X_i^2}{(s^2+\sigma^2X_i^2)^2} & \frac{4s\sigma X_i^2(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)^2}{(s^2+\sigma^2X_i^2)^3\Delta t} + \frac{-2s\sigma X_i^2}{(s^2+\sigma^2X_i^2)^2} \\ * & * & * & \frac{X_i^2(-s^2+3\sigma^2X_i^2)(X_{i+1}-X_i+aX_i^3\Delta t+gX_i\Delta t)^2}{(s^2+\sigma^2X_i^2)^3\Delta t} + \frac{X_i^2(s^2-\sigma^2X_i^2)}{(s^2+\sigma^2X_i^2)^2} \end{bmatrix}$$

H is symmetric.

Guess+optimization tool			
	true value	mean of MLE	Relative Error
a	1	1.0043	0.43%
g	1	0.9957	0.43%
s	1	1.0000	0.00%
σ	0.5	-0.4999	0.02%
a	1	1.0011	0.11%
g	1	1.0014	0.14%
s	1	0.9992	0.08%
σ	1	1.0026	0.26%
a	1	0.9981	0.19%
g	1	1.0082	0.82%
s	1	1.0000	0.00%
σ	2	2.0193	0.97%