The problem is to estimate $\mu(.)$ and $\sigma(.)^2$ for a 1d diffusion $dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$ from a single path. As a warm up example, consider estimating $\theta = (\mu, \sigma)$ for arithmetic Brownian motion $X_t = \mu t + \sigma W_t$ with n observations at equidistant intervals δ_n with $\delta_n \to 0$. Then the log likelihood of the increments ΔX_i of X is

$$\ell_n(\mu, \sigma) \quad = \quad const. - n \log \sigma \; - \sum_{i=1}^{n} \frac{(\Delta X_i - \mu\delta_n)^2}{2\sigma^2 \delta_n}$$

where *const.* is independent of (μ, σ). Then the **Fisher information** $-\mathbb{E}(\frac{\partial^2 \ell_n}{\partial\theta_i \partial\theta_j})$ is given by

$$I_n(\mu, \sigma) \quad = \quad \begin{bmatrix} \frac{T}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

where $T = n\delta_n$, so (by **Cramer-Rao**) the covariance of any unbiased estimator for $\theta = (\mu, \sigma)$ is $\geq I_n(\mu, \sigma)^{-1}$. In particular, the variance of any unbiased estimator for μ is $\geq I_{1,1}(\mu, \sigma)^{-1} = \frac{\sigma^2}{T}$. Hence we need $T \to \infty$ (i.e. the observation window tending to ∞) to get a consistent estimator for μ, as also discussed earlier this week (note the condition that $\delta_n \to 0$ and $T \to \infty$ also appears in **Theorem 4.1** in [WM24]). [1]

More generally, if σ is known and $\hat{\mu}$ is *any* estimator for μ with **bias** $b(\mu)$, then

$$\text{MSE}_\mu(\hat{\mu}) \quad := \quad \mathbb{E}_\mu((\hat{\mu} - \mu)^2) \quad \geq \quad \frac{(1 + b'(\mu))^2}{I(\mu)} \; + b(\mu)^2 \quad = \quad \frac{\sigma^2}{T}(1 + b'(\mu))^2 \; + b(\mu)^2$$

so for T fixed, the right hand side cannot be zero if $b(\mu) \neq 0$.

Note **Theorem 4.1** in the cited article [WM24] also requires X to be **ergodic** (see Eq 3.2 in [WM24] for definition) which is obviously not the case for the process in **q4** of the draft project since X there is not mean-reverting.

SINDy method discussed in the draft and [WM24]

The simplest case is when the "dictionary" of functions only consists of the constant function 1. Then (for an ergodic 1d diffusion) Eq 4.1 in [WM24] is $\sum_{m=0}^{N-1}(\frac{\Delta X_m}{\Delta t} - v)^2$. Minimizing in v leads to the first-order optimality condition:

$$2\sum_{m=0}^{N-1}(\frac{\Delta X_m}{\Delta t} - \hat{v}) \quad = \quad 0$$

so we recover the obvious unbiased estimate: $\hat{\mu} = \frac{1}{N}\sum_{m=0}^{N-1} \frac{\Delta X_m}{\Delta t} = (X_T - X_0)/T$. Similarly, to estimate the diffusion coefficient, we consider

$$\frac{d}{dv}\sum_{m=0}^{N-1}(\frac{(\Delta X_m)^2}{\Delta t} - v)^2 \,|_{v=\hat{v}} \quad = \quad -2\sum_{m=0}^{N-1}(\frac{(\Delta X_m)^2}{\Delta t} - \hat{v}) \quad = \quad 0$$

which leads to $\hat{\sigma}^2 N\Delta t = \sum_{m=0}^{N-1}(\Delta X_m)^2$, i.e. usual estimation method for $\hat{\sigma}^2$ using realized variance (see also more involved analysis on page 7 in [WM24]).

Again if X is arithmetic Brownian motion: $X_t = \mu t + \sigma W_t$, the $\frac{\Delta X_i}{\Delta t}$'s are i.i.d. If $\Delta t = 1$, then $\mathbb{E}(\hat{v}) = \mu$, $\text{Var}(\hat{v}) = \frac{\sigma^2}{N}$ and (from the SLLN) $\hat{v} \to \mu$ as $N \to \infty$.

But if $\Delta t = T/N$ for some T **fixed**, $v = (X_T - X_0)/T$, which has variance $\frac{\sigma^2}{T}$ which clearly does not vanish as $N \to \infty$ (recall Theorem 4.1 in [WM24] requires $T \to \infty$ and $\Delta t \to 0$).

If we now include the next term v_1, the task becomes more or less the same as the first part of Project 3: just performing linear regression on

$$\frac{\Delta X_t}{\Delta t} \quad \text{vs} \quad v_0 + v_1 X_t$$

(in P3 the equation the corresponding linear regression is

$$\Delta\xi_t \quad \text{vs} \quad -a(\xi_{t-1} - \bar{\xi})$$

where they are fitting the parameters for a discrete-time OU process (AR(1) process), see page 74 in [Kut04] for the case of a general AR(d) process. For P3, their unit of time is days with $\Delta t = 1$ day, and their ξ process is ergodic and T is very large ($\approx$ 35yrs$\times$252 = 8820 days).

[1] One can also replace W by an fBM and compute a 3x3 matrix $I_n(\mu, \sigma, H)$ although one runs into issues with a singular matrix in the high frequency regime because the increment sizes are of order $\sigma^2\Delta^{2H}$, so taking logs gives $\log(\sigma^2) + 2H\log(\Delta)$ so H dominates when $\Delta \ll 1$, see [Kaw13].

Remark 0.1 ChatGPT generated this code in Python:
`https://colab.research.google.com/drive/19fwbrMJu4WOMNNG7DuFuoGVTb8qbjXVn?usp=sharing` for the SINDy problem for a standard OU process with a dictionary of constant and linear functions, which seems to work well with a large T and small δ_n, but as expected (from the arguments above) it doesn't work well with a fixed time horizon T. Note I didn't use any **regularization** so not sure if this is **sparse** in the sense you mean, and this only took about 5 minutes to write and clean up and the code is rather short, so this task maybe needs to be more substantial in the project.

Continuous observation - the Trajectory Fitting Estimator (TFE)

For a discrete-time autoregressive (AR) process, the TFE method is essentially the first task of Project 3 as discussed above.

We now discuss the analog of this approach for diffusion processes with continuous observation. Following pages 5 and 74 in [Kut04], let X be an ergodic 1d diffusion:

$$X_t \quad = \quad X_0 + \int_0^t S(\theta, X_s)ds + \int_0^t \sigma(X_s)dW_s$$

for $\theta \in \Theta \subset \mathbb{R}^d$ as in [WM24], and assume $\sigma(.)$ is known now since it can easily be estimated with realized variance if X is observed continuously. Set

$$\hat{X}_t(\theta) \quad = \quad X_0 + \int_0^t S(\theta, X_s)ds \quad \Rightarrow \quad \nabla_\theta \hat{X}_t(\theta) \quad = \quad \int_0^t \nabla_\theta S(\theta, X_s)ds\,. \tag{1}$$

Then the TFE estimator for θ is $\theta_T^\star = \arg\inf_{\theta\in\Theta}\int_0^T (X_t - \hat{X}_t(\theta))^2 dt$, which leads to the first order conditions:

$$\nabla_\theta \int_0^T (X_t - \hat{X}_t(\theta))^2 dt\,|_{\theta=\theta_T^\star} \quad = \quad 2\int_0^T (X_t - \hat{X}_t(\theta_T^\star))\,\nabla_\theta \hat{X}_t(\theta_T^\star)dt \quad = \quad 0 \tag{2}$$

which gives us n non-linear equations for the n unknowns $\theta_T^\star$ if Θ has dimension n.

Remark 0.2 See discussion on the **identifiability condition** for the TFE on page 75 in [Kut04], although there may be a typo in that formula.

TFE Explicit Example: the OU process

Example 1.4.3 in [Kut04]: For the OU case $dX_t = (b - aX_t)dt + \sigma dW_t$ with b, σ known, $\nabla_\theta S(\theta, x) = \frac{\partial}{\partial a}S(a, x) = -x$, so (1) becomes

$$\partial_a \hat{X}_t(a) \quad = \quad -\int_0^t X_s ds \quad = \quad -Y_t\,.$$

Hence (2) simplifies to

$$\int_0^T (X_t - X_0 - bt + aY_t)\cdot -Y_t dt \quad = \quad 0$$

where $Y_t = \int_0^t X_s ds$, which we then re-arrange to obtain the TFE estimate for a as

$$a_T^\star \quad = \quad -\frac{\int_0^T (X_t - X_0 - bt)Y_t dt}{\int_0^T Y_t^2 dt}$$

(the **MLE** for a which is obtained by maximizing the Girsanov factor with respect to a reference θ_0 involves a similar formula). Page 76 in [Kut04] shows that $a_T^\star$ is consistent and $\sqrt{T}(a_T^\star - a)$ is **asymptotically Normal** as $T \to \infty$, although the MLE for a has lower asymptotic variance, so in general $\hat{a}_T$ is not **asymptotically efficient**.

Remark 0.3 See page 5 in [Kut04] for discussion on what happens when the observed diffusion process does not belong to the prescribed parametric family.

References

[Kaw13] Kawai, R., "Fisher Information for Fractional Brownian Motion Under High-Frequency Discrete Sampling", *Communications in Statistics – Theory and Methods*, vol. 42, no. 9, pp. 1628–1636, 2013.

[Kut04] Y.A. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2004, `https://annas-archive.org/md5/7897b3238b6838d92775e61516b033ed` Slow Partner Server #1

[WM24] M. Wanner and I. Mezic, "On higher order drift and diffusion estimates for stochastic SINDy", *SIAM Journal on Applied Dynamical Systems*, 23 (2024), pp. 1504–1539