De-biasing particle filtering for a continuous time hidden Markov model with a Cox process observation model

Department of Engineering, University of Cambridge

NIASRA, School of Mathematics and Applied Statistics, University of Wollongong

ENSAE, Institut Polytechnique de Paris

Supplementary Material

S1 Bridge density for linear Gaussian diffusion

Consider the following stochastic differential equation (SDE),

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{S1.1}$$

where X_t is an n-dimensional diffusion process, W_t is an m-dimensional standard Brownian motion for $m \leq n$. For $b(t, X_t) := b_0 + b_1(t)X_t$ and $\sigma(t, X_t) := \sigma(t) \in \mathbb{R}^{n\times m}$, the solution to (S1.1) at discrete time points $t_0 < t_1 < \dots$ is given by Jazwinski (2007); Evans (2012)

$$X_{t_{i+1}} = \Phi(t_i, t_{i+1})X_{t_i} + a(t_i, t_{i+1}) + \int_{t_i}^{t_{i+1}} \Phi(t_i, t)\sigma(t)dW_t \tag{S1.2}$$

where the fundamental matrix function $\Phi \in \mathbb{R}^{n\times n}$ satisfies the following for all $s, t, u \geq t_0$

$$\frac{d\Phi(s,t)}{dt} = b_1(t)\Phi(s,t), \quad \Phi(t,t) = \mathbb{I}_{n\times n}, \quad \Phi(s,t)\Phi(t,u) = \Phi(s,u),$$

the vector $a(t_i, t_{i+1}) \in \mathbb{R}^n$ is given by $a(t_i, t_{i+1}) = \int_{t_i}^{t_{i+1}} b_0\Phi(t_i, t)dt$. Therefore the transition density $f_{t_i,t_{i+1}}(x'|x)$ can be expressed as a Gaussian as follows,

$$\begin{aligned}
f_{t_i,t_{i+1}}(x'|x) &:= \frac{\exp\left(-\frac{1}{2}\left(x' - \mu(x, t_i, t_{i+1})\right)^\top R^{-1}(t_i, t_{i+1})\left(x' - \mu(x, t_i, t_{i+1})\right)\right)}{\sqrt{|2\pi R^{-1}(t_i, t_{i+1})|}}, \\
&\propto \frac{\exp\left(-\frac{1}{2}\left(x - \hat{\mu}(x', t_i, t_{i+1})\right)^\top \hat{R}^{-1}(t_i, t_{i+1})\left(x - \hat{\mu}(x', t_i, t_{i+1})\right)\right)}{\sqrt{|2\pi \hat{R}^{-1}(t_i, t_{i+1})|}} \\
&=: \hat{f}_{t_i,t_{i+1}}(x|x'),
\end{aligned}$$

where

$$\begin{aligned}
\mu(x, t_i, t_{i+1}) &:= \Phi(t_i, t_{i+1})x + a(t_i, t_{i+1}), \\
\hat{\mu}(x', t_i, t_{i+1}) &:= \Phi^{-1}(t_i, t_{i+1})x' - \Phi^{-1}(t_i, t_{i+1})a, \\
R(t_i, t_{i+1}) &:= \int_{t_i}^{t_{i+1}} \Phi(t_i, t)\sigma(t)\sigma^\top(t)\Phi^\top(t_i, t)dt, \\
\hat{R}(t_i, t_{i+1}) &:= \int_{t_i}^{t_{i+1}} \Phi^{-1}(t, t_{i+1})\sigma(t)\sigma^\top(t)\Phi^{-\top}(t, t_{i+1})dt.
\end{aligned}$$

Assume $s < \tau < t$, then in addition to sampling $p(x_\tau|x_s)$ exactly, one can also sample $X_\tau \sim p(x_\tau|x_s, x_t)$ exactly where

$$\begin{aligned}
&p(x_\tau|x_s, x_t) \propto f_{s,\tau}(x_\tau|x_s)f_{\tau,t}(x_t|x_\tau) \propto f_{s,\tau}(x_\tau|x_s)\hat{f}_{\tau,t}(x_\tau|x_t) \\
&\propto \mathrm{N}\left(x_\tau; \left(R^{-1}(s,\tau) + \hat{R}^{-1}(\tau,t)\right)^{-1} \left(R^{-1}(s,\tau)\mu(x_s, s, \tau) + \hat{R}^{-1}(\tau,t)\hat{\mu}(x_t, \tau, t)\right),\right. \\
&\quad \left.\left(R^{-1}(s,\tau) + \hat{R}^{-1}(\tau,t)\right)^{-1}\right).
\end{aligned}$$

S2 Proof of Lemma 1

The following propositions will be used in the final proof.

Proposition 1. *Let X be the Brownian motion which starts at $X_0 = x_0$,* *then the following equality holds for any $a > 0$:*

$$\Pr\Big(\sup_{0\le s\le\Delta} X_s - X_0 \ge a | X_\Delta = x_\Delta\Big) = \begin{cases} \exp\Big\{-\frac{2a}{\Delta}[a-(x_\Delta - x_0)]\Big\}, & a > x_\Delta - x_0 \\ 1, & a \le x_\Delta - x_0 \end{cases}$$

Proof. Define τ_a as the hitting time of a as follows,

$$\tau_a = \inf\{s \in [0,\Delta] \,|X_s - X_0 = a\}$$

A hitting time is also a stopping time. Then by applying the reflection principle (please refer to Theorem 2.19 of Mörters and Peres (2010)), the process $\{X^* : t \ge 0\}$, called Brownian motion $\{X_t : t \ge 0\}$ reflected at τ_a, defined by

$$\begin{aligned} X_t^* &= X_t \mathbb{I}_{t\le\tau_a} + (2X_{\tau_a} - X_t)\mathbb{I}_{t>\tau_a} \\ &= X_t \mathbb{I}_{t\le\tau_a} + (2a + 2x_0 - X_t)\mathbb{I}_{t>\tau_a} \end{aligned}$$

is also a Brownian motion. Thus,

$$\begin{aligned} &\Pr\Big(\tau_a \le \Delta, X_\Delta \in [x_\Delta, x_\Delta + dx]\Big) \\ =&\Pr\Big(X_\Delta^* \in [2a + 2x_0 - x_\Delta - dx, 2a + 2x_0 - x_\Delta]\Big) \\ =&\Pr\Big(X_\Delta^* - X_0 \in [2a + x_0 - x_\Delta - dx, 2a + x_0 - x_\Delta]\Big) \\ =&\frac{dx}{\sqrt{2\pi\Delta}} \exp\Big\{-\frac{[2a-(x_\Delta - x_0)]^2}{2\Delta}\Big\}. \end{aligned}$$

Note that,

$$\begin{aligned}
&\Pr\Big(X_\Delta \in [x_\Delta, x_\Delta + dx]\Big)\\
=&\Pr\Big(X_\Delta - X_0 \in [x_\Delta - x_0, x_\Delta - x_0 + dx]\Big)\\
=&\frac{dx}{\sqrt{2\pi\Delta}} \exp\Big\{ - \frac{(x_\Delta - x_0)^2}{2\Delta}\Big\}.
\end{aligned}$$

Division between two equations above concludes the proof. □

Proposition 2. *Let X be the Brownian motion which starts at $X_0 = x_0$, then the following equality holds for any $a > 0$:*

$$\begin{aligned}
&\Pr\left(\inf_{0\le s\le\Delta} X_s - X_0 \le -a | X_\Delta = x_\Delta\right)\\
=&\begin{cases}\exp\left\{-\frac{2a}{\Delta}\left[a + (x_\Delta - x_0)\right]\right\}, & a > -(x_\Delta - x_0)\\ 1, & a \le -(x_\Delta - x_0).\end{cases}
\end{aligned}$$

Proof. A similar approach as in Proof S2 but define $\tau_{-a} = \inf\{s \in [0,\Delta] | X_s - X_0 = -a\}$ and apply the reflection principle by defining the Brownian motion $\{X^* : t \ge 0\}$, the Brownian motion $\{X_t : t \ge 0\}$ reflected at τ_{-a}, formally defined by

$$X_t^* = X_t \mathbb{I}_{t\le -\tau_a} + (-2a + 2x_0 - X_t)\mathbb{I}_{t\ge -\tau_a}.$$

□

Proposition 3. *Let X be defined as in Proposition 1, then the following* *inequality holds:*

$$\Pr\left(\sup_{0\le s\le\Delta}|X_s-X_0|\ge a|X_\Delta=x_\Delta\right)$$
$$\le\begin{cases}2\exp\Big\{-\frac{2a}{\Delta}[a-|x_\Delta-x_0|]\Big\}, & a>|x_\Delta-x_0|\\ 1, & a<|x_\Delta-x_0|\end{cases}$$

Proof.

$$\begin{aligned}&\Pr\Big(\sup_{0\le s\le\Delta}|X_s-X_0|\ge a|X_\Delta=x_\Delta\Big)\\ \le&\Pr\Big(\sup_{0\le s\le\Delta}X_s-X_0\ge a|X_\Delta=x_\Delta\Big)+\Pr\Big(\inf_{0\le s\le\Delta}X_s-X_0\le -a|X_\Delta=x_\Delta\Big)\\ =&\begin{cases}\exp\left\{-\frac{2a}{\Delta}\left[a-(x_\Delta-x_0)\right]\right\}+\exp\left\{-\frac{2a}{\Delta}\left[a+(x_\Delta-x_0)\right]\right\} & a>|x_\Delta-x_0|\\ 1, & a\le|x_\Delta-x_0|\end{cases}\\ \le&\begin{cases}2\exp\left\{-\frac{2a}{\Delta}\left[a-|x_\Delta-x_0|\right]\right\} & a>|x_\Delta-x_0|\\ 1, & a\le|x_\Delta-x_0|\end{cases}\end{aligned}$$

□

Proof of Lemma 1

Proof.

$$
\begin{aligned}
&\Pr\left(E_1 > 0 \middle| \kappa = k > 0\right) \\
=&\mathbb{E}\left\{\mathbb{I}\left[\left(\prod_{j=1}^{k}\left(1+\frac{\Delta}{\eta}\left(\lambda\left(X_0\right)\right)-\lambda\left(X_{\tau_j}\right)\right)\right)>0\right]\middle|\kappa = k, \tau_1, \ldots, \tau_k\right\} \\
\geq&\mathbb{E}\left\{\mathbb{I}\left[\max_{j\in\{1,\ldots,k\}}\frac{\Delta}{\eta}|\lambda\left(X_0\right)-\lambda\left(X_{\tau_j}\right)|<1\right]\middle|\kappa = k, \tau_1, \ldots, \tau_k\right\} \\
=&\Pr\left(\max_{j\in\{1,\ldots,k\}}|\lambda\left(X_0\right)-\lambda\left(X_{\tau_j}\right)|\leq\frac{\eta}{\Delta}\middle|\kappa = k, \tau_1, \ldots, \tau_k\right).
\end{aligned}
$$

We can obtain an upperbound for $\Pr(E_1 < 0|\kappa = k)$ by

$$
\begin{aligned}
\Pr\left(E_1 < 0\middle|\kappa = k\right) &\leq \Pr\left(\max_{j\in\{1,\ldots,k\}}|\lambda\left(X_0\right)-\lambda\left(X_{\tau_j}\right)|\geq\frac{\eta}{\Delta}\middle|\kappa = k, \tau_1, \ldots, \tau_k\right) \\
&\leq \Pr\left(\max_{j\in\{1,\ldots,k\}}|X_0 - X_{\tau_j}|\geq\frac{\eta}{\Delta l}\middle|\kappa = k, \tau_1, \ldots, \tau_k\right)
\end{aligned}
$$

where we assume $\lambda(\cdot)$ is an $l-$Lipschitz function.

$$
\Pr\left(\max_{j\in\{1,\ldots,k\}}|X_0 - X_{\tau_j}|\geq\frac{\eta}{\Delta l}\middle|\kappa = k, \tau_1, \ldots, \tau_k\right) \leq \Pr\left(\sup_{0\leq s\leq\Delta}|X_0 - X_s|\geq\frac{\eta}{\Delta l}\right).
$$

Applying Proposition 3, we have

$$
\Pr\left(E_1 < 0|\kappa > 0\right) \leq \begin{cases} 2\exp\left\{-\frac{\frac{2\eta}{\Delta l}\left(\frac{\eta}{\Delta l}-|x_\Delta - x_0|\right)}{\Delta}\right\}, & \frac{\eta}{\Delta l}\geq|x_\Delta - x_0| \\ 1, & \frac{\eta}{\Delta l}\leq|x_\Delta - x_0| \end{cases} \tag{S2.3}
$$

□

S3 Expectation of the probability bound

This section establishes the unqualified bound ((4.8) in the manuscript).

The goal is to determine the following expectation for $Y = |X_\Delta - X_0|$ where $X_\Delta - X_0 \sim \mathcal{N}(0, \Delta)$, (thus Y is a half-normal random variable):

$$
\begin{aligned}
&\mathbb{E}\left\{\mathbb{I}\left[E<0\right]|\kappa>0\right\} \\
=&\mathbb{E}\left\{\mathbb{I}\left[E<0\right]\times\mathbb{I}\left[Y<\frac{\eta}{\Delta l}\right]+\mathbb{I}\left[E<0\right]\times\mathbb{I}\left[Y\geq\frac{\eta}{\Delta l}\right]|\kappa>0\right\} \\
\leq&\mathbb{E}\left\{2\exp\left(-\frac{\frac{2\eta}{\Delta l}\left(\frac{\eta}{\Delta l}-Y\right)}{\Delta}\right)\mathbb{I}\left[Y<\frac{\eta}{\Delta l}\right]\right\}+\Pr\left(Y\geq\frac{\eta}{\Delta l}\right) \\
=&\int_0^\infty 2\exp\left(-\frac{\frac{2\eta}{\Delta l}\left(\frac{\eta}{\Delta l}-y\right)}{\Delta}\right)\mathbb{I}\left[y<\frac{\eta}{\Delta l}\right]\times\frac{\sqrt{2}}{\sqrt{\pi\Delta}}\exp\left(-\frac{y^2}{2\Delta}\right)dy+\Pr\left(Y\geq\frac{\eta}{\Delta l}\right) \\
=&\int_0^{\frac{\eta}{\Delta l}}\frac{2\sqrt{2}}{\sqrt{\pi\Delta}}\exp\left(-\frac{2\eta^2}{\Delta^3 l^2}\right)\times\exp\left(-\frac{\left(y-\frac{2\eta}{\Delta l}\right)^2-\frac{4\eta^2}{\Delta^2 l^2}}{2\Delta}\right)dy+\Pr\left(Y\geq\frac{\eta}{\Delta l}\right) \\
=&\int_0^{\frac{\eta}{\Delta l}}\frac{2\sqrt{2}}{\sqrt{\pi\Delta}}\exp\left(-\frac{\left(y-\frac{2\eta}{\Delta l}\right)^2}{2\Delta}\right)dy+\Pr\left(|X_\Delta-X_0|\geq\frac{\eta}{\Delta l}\right) \\
=&4\left[\Phi\left(\frac{2\eta}{\Delta^{\frac{3}{2}}l}\right)-\Phi\left(\frac{\eta}{\Delta^{\frac{3}{2}}l}\right)\right]+2\times\left(1-\Phi\left(\frac{\eta}{\Delta^{\frac{3}{2}}l}\right)\right) \\
=&2+4\Phi\left(\frac{2\eta}{\Delta^{\frac{3}{2}}l}\right)-6\Phi\left(\frac{\eta}{\Delta^{\frac{3}{2}}l}\right).
\end{aligned}
$$

S4 Proof of Lemma 2

Proof.

$$\begin{aligned}
E_i &= \exp\left(-\Delta\lambda(X_{(i-1)\Delta})\right)\prod_{j=1}^{\kappa_i}\left(1+\frac{\lambda(X_{(i-1)\Delta})-\lambda(X_{\tau_j})}{l}\right) \\
&\leq \prod_{j=1}^{\kappa_i}\left(1+\left|X_{(i-1)\Delta}-X_{\tau_j}\right|\right) \\
&\leq \prod_{j=1}^{\kappa_i}\left(1+\max_{(i-1)\Delta\leq s\leq i\Delta}\left|X_s-X_{(i-1)\Delta}\right|\right) \\
&= \left(1+\max_{0\leq s\leq\Delta}|B_s|\right)^{\kappa_i} =: F_i. \qquad \text{(S4.4)}
\end{aligned}$$

We truncate the Poisson estimate as $E_i^+ = E_i\mathbb{I}_{A_i^c}$ and bound $\mathbb{I}_A\prod_{i=1}^m E_i$ as follows.

$$\left|\mathbb{E}\left\{\mathbb{I}_A\prod_{i=1}^m E_i\right\}\right| \leq \mathbb{E}\left\{\prod_{i=1}^m E_i^2\right\}^{\frac{1}{2}}\mathbb{E}\left\{\mathbb{I}_A\right\}^{\frac{1}{2}}. \qquad \text{(S4.5)}$$

The term $\mathbb{E}\{\mathbb{I}_A\}^{\frac{1}{2}}$ can be bounded using the union bound

$$\mathbb{E}\{\mathbb{I}_A\} \leq \sum_{i=1}^m \mathbb{E}\{\mathbb{I}_{A_i}\} = m\mathbb{E}\{\mathbb{I}_{A_i}\} = m\times 2\exp\left(-\frac{1}{2\Delta}\right) \qquad \text{(S4.6)}$$

The other term can be proved to be finite, i.e. $\mathbb{E}\{\prod_{i=1}^m E_i^2\} < \infty$. Since the increment of Brownian motion X is independent of each other, and using the inequality (S4.4), one can show

$$\mathbb{E}\left\{\prod_{i=1}^m E_i^2\right\} \leq \mathbb{E}\left\{\prod_{i=1}^m F_i^2\right\} = \mathbb{E}\left\{F_i^2\right\}^m, \qquad \forall \quad i.$$

It remains therefore to bound $\mathbb{E}\left\{F_i^2\right\}$.

$$
\begin{aligned}
\mathbb{E}\left\{F_i^2\right\} &= \mathbb{E}\left\{\left(1+\max_{0\le s\le\Delta}|B_s|\right)^{2\kappa_i}\right\} \\
&= \mathbb{E}\left\{\sum_{k=0}^{\infty}\frac{(\Delta l)^k e^{-\Delta l}}{k!}\left(1+\max_{0\le s\le\Delta}|B_s|\right)^{2k}\right\} \\
&= \exp(-\Delta l)\mathbb{E}\left\{\exp\left(\Delta l(1+\max_{0\le s\le\Delta}|B_s|)^2\right)\right\} \\
&\le \exp(-\Delta l)\mathbb{E}\left\{\exp\left(\Delta l(2+2\times\max_{0\le s\le\Delta}B_s^2)\right)\right\} \\
&= \exp(\Delta l)\mathbb{E}\left\{\exp\left(2\Delta l\times\max_{0\le s\le\Delta}B_s^2\right)\right\} \\
&= \exp(\Delta l)\times\int_0^{\infty}\Pr\left(\exp\left(2\Delta l\times\max_{0\le s\le\Delta}B_s^2\right)>w\right)dw \\
&= \exp(\Delta l)\times\left[1+\int_1^{\infty}\Pr\left(\exp\left(2\Delta l\times\max_{0\le s\le\Delta}B_s^2\right)>w\right)dw\right] \\
&= \exp(\Delta l)\times\left[1+\int_1^{\infty}\Pr\left(\max_{0\le s\le\Delta}|B_s|>\sqrt{\frac{\log(w)}{2\Delta l}}\right)dw\right] \\
&\le \exp(\Delta l)\times\left[1+\int_1^{\infty}2\exp\left(-\frac{\log(w)}{4\Delta^2 l}\right)dw\right] \\
&= \exp(\Delta l)\times\left[1+\int_1^{\infty}2w^{-\frac{1}{4\Delta^2 l}}dw\right] \\
&= \exp(\Delta l)\times\left(\frac{1+4\Delta^2 l}{1-4\Delta^2 l}\right)
\end{aligned}
$$

where in the fourth last line we apply the inequality for running maximum of Brownian motion which starts at zero, i.e. $\Pr\left(\max_{0\le s\le\Delta}|B_s|>a\right)\le 2\exp\left(-\frac{a^2}{2\Delta}\right)$ for any positive number a.

Therefore,

$$\mathbb{E}\left\{E_1^2 \cdots E_m^2\right\} \leq \exp\left(Tl\right) \times \left(\frac{1+4\Delta^2 l}{1-4\Delta^2 l}\right)^m. \tag{S4.7}$$

Plugging (S4.6) and (S4.7) into (S4.5) concludes the proof. □

S5 Experiments for Wald's identity

In this section, we wish to numerically show that the Wald estimate is biased, i.e. $\mathbb{E}^\theta(\hat{L}(\theta))/L(\theta)$ changes as θ changes where $K = \inf\{k > 0 : E^{(1)} + \ldots + E^{(k)} > 0\}$, $\hat{L}(\theta) = \sum_{i=1}^K E^{(i)}$ and $L(\theta) = \mathbb{E}^\theta\left(G^\theta(X_0)\right)$.

Here is an example of which we know the true solution to. The dynamics that describes how one dimensional process X evolves is given by

$$dX_t = bdt + dW_t$$

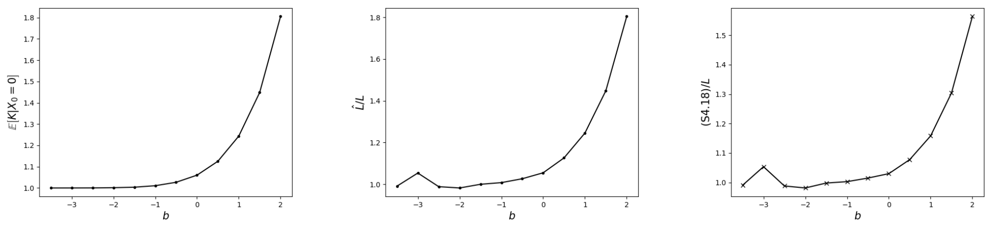
where b is some constant and W is a one dimensional Brownian motion. Hence

$$X_t | X_{t'} = x \sim \mathcal{N}\left(x + b \times (t - t'), t - t'\right).$$

Thus, $\theta = b$ in this case. We can exactly calculate $L(b)$ for $\lambda(x) = x + 10$,

$$L(b) = \mathbb{E}\left(\exp\left(-\int_0^T \lambda(X_t)dt\right) \Big| X_0 = 0\right) = \exp\left(-10T - \frac{b}{2}T^2 + \frac{1}{6}T^3\right)$$

where the expectation is taken with respect to Brownian motion $X | X_0 = 0$. Note that $\int_0^T (10 + X_0 + bt + W_t - W_0)dt \sim \mathcal{N}\left(10T + \frac{b}{2}T^2, \frac{1}{3}T^3\right)$ for $X_0 = W_0 = 0$.



(a) (b) (c)

Figure 1: (a) Plot of $\mathbb{E}^\theta(K|X_0 = 0)$, (b) $\hat{L}'/L$ and (c) estimate (S5.8)$/L$ versus $\theta = b$.

We obtain $N = 10^6$ independent samples $\hat{L}(b)$ for every value of b:

$$\hat{L}'(b) = \frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{K_j} E_j^{(i)} \approx \mathbb{E}^\theta(\hat{L}(b))$$

where $K_j = \inf\{k > 0 : E_j^{(1)} + \ldots + E_j^{(k)} > 0\}$. Each $E_j^{(i)}$ is an independent sample where $E_j^{(i)} \leftarrow \mathrm{PE}(T, 0, T, 0)$. Figure 1a shows that as b increases, the number of draws to make Wald estimate positive increases. For this example E is

$$E = \exp\left(-T(X_0 + 10)\right)\prod_{i=1}^{\kappa}\left[1 + [X_0 - X_{\tau_i}]\right]$$

where $\kappa \sim \mathcal{P}o(T)$ and $\tau_1, \ldots, \tau_\kappa \sim \mathcal{U}(0, T)$ are i.i.d. samples. Larger b (i.e. larger drift dragging the particle towards positive direction) increases the chances of meeting negative Poisson estimate. In Figure 1b, we notice a clear trend that the empirical ratio, $\hat{L}'(b)/L(b)$, increases with b. Finally we plot

$$\frac{1}{N}\sum_{j=1}^{N}\frac{1}{K_j}\sum_{i=1}^{K_j} E_j^{(i)}. \tag{S5.8}$$

S6 Exact computation of likelihood function

This section is to determine the following likelihood function.

$$\mathbb{E}\left\{\exp\left[-\int_{\tau}^{T}(\alpha X_t+\beta)\,dt\right]\Big|X_\tau=x_0,X_T=x_1\right\}$$

The procedure can be splitted into 4 steps.

1. As $X_t|X_T$ is a Gaussian process, the Lebesgue integral is Gaussian random variable, see Folland (1999): approximate the given integral as Riemann sums and each Riemann sum is Gaussian and hence the limit will also be Gaussian.

$$\left(\int_{\tau}^{T}\alpha X_t+\beta dt\Big|X_\tau=x_0,X_T=x_1\right)\sim\mathcal{N}\left(\alpha\mu+\beta\left(T-\tau\right),\alpha^2\sigma^2\right)$$

2. Calculate mean μ:

$$\begin{aligned}\mathbb{E}\left[\int_{\tau}^{T}X_t dt\Big|X_\tau=x_0,X_T=x_1\right]&=\int_{\tau}^{T}\mathbb{E}\left[X_t\Big|X_\tau=x_0,X_T=x_1\right]dt\\&=\int_{\tau}^{T}x_0+\frac{t-\tau}{T-\tau}\left(x_1-x_0\right)dt\\&=\frac{1}{2}\left(T-\tau\right)\left(x_0+x_1\right)\end{aligned}$$

3. Calculate variance σ^2:

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_\tau^T X_t dt\right)\left(\int_\tau^T X_t dt\right)\Big| X_\tau = x_0, X_T = x_1\right] - \mu^2 \\
=&\mathbb{E}\left[\int_{[\tau,T]^2} X_u X_v dudv \Big| X_\tau = x_0, X_T = x_1\right] - \mu^2 \\
=&\int_{[\tau,T]^2} \mathbb{E}\left[X_u X_v \Big| X_\tau = x_0, X_T = x_1\right] dudv - \mu^2 \\
=&\int_{[\tau,T]^2} \operatorname{cov}(X_u, X_v) + \mathbb{E}\left[X_u \Big| X_\tau = x_0, X_T = x_1\right] \times \\
&\mathbb{E}\left[X_v \Big| X_\tau = x_0, X_T = x_1\right] dudv - \mu^2 \\
=&\int_{[\tau,T]^2} \frac{(u \wedge v - \tau)(T - u \vee v)}{T - \tau} + \left(x_0 + \frac{u-\tau}{T-\tau}(x_1 - x_0)\right) \times \\
&\left(x_0 + \frac{v-\tau}{T-\tau}(x_1 - x_0)\right) dudv - \mu^2 \\
=&\int_\tau^T \int_\tau^v \frac{(u-\tau)(T-v)}{T-\tau} dudv + \int_\tau^T \int_v^T \frac{(v-\tau)(T-u)}{T-\tau} dudv \\
=&\frac{(T-\tau)^3}{12}
\end{aligned}$$

4. Calculate the likelihood:

$$\begin{aligned}
&\mathbb{E}\left\{\exp\left[-\int_\tau^T (\alpha X_t + \beta) dt\right] \Big| X_\tau = x_0, X_T = x_1\right\} \\
=&\exp\left[-\frac{\alpha}{2}(T-\tau)(x_0 + x_1) - \beta(T-\tau) + \frac{\alpha^2 (T-\tau)^3}{24}\right]
\end{aligned}$$

Therefore, the exact likelihood for $t_{1:n_p}$ and $y_{t_1:t_{n_p}}$, where n_p is the number of observations, is

$$\mathcal{L} = \mathbb{E}\left\{\left[\prod_{i=1}^{n_p}(X_{t_i}+10)\, g^{\theta}(y_{t_i}|X_{t_i}) \right.\right.$$
$$\left. \times \exp\left(-\frac{t_i - t_{i-1}}{2}\left(X_{t_{i-1}} + X_{t_i}\right) - 10(t_i - t_{i-1}) + \frac{(t_i - t_{i-1})^3}{24}\right)\right]$$
$$\left. \times \exp\left(-\frac{T - t_{n_p}}{2}\left(X_{t_{n_p}} + X_T\right) - 10(T - t_{n_p}) + \frac{(T - t_{n_p})^3}{24}\right)\right\}.$$

To find the ground truth for values of $n_p > 2$, we use Algorithm 3 described in the manuscript with line 8 using the exact evaluation (given by (S6.1)). This allows the computation of Monte Carlo estimate described in Section 5.1 of manuscript.

$$E_k^{(i)} = \exp\left[-\frac{1}{2}\left(t_k^{\Delta} - t_{k-1}^{\Delta}\right)\left(X_{k-1}^{\Delta} + X_k^{\Delta}\right) - 10\left(t_k^{\Delta} - t_{k-1}^{\Delta}\right) + \frac{\left(t_k^{\Delta} - t_{k-1}^{\Delta}\right)^3}{24}\right]. \tag{S6.1}$$

S7 No Observation Case and Two Observations Case

The exact likelihood for no observation received within $[0,T]$ is

$$\begin{aligned}
&\mathbb{E}\left\{\exp\left(-\int_0^T \lambda(X_s)ds\right)\right\}\\
=&\mathbb{E}\left\{\mathbb{E}\left[\exp\left(-\int_0^T \lambda(X_s)ds\right)|X_0=0,X_T\right]\right\}\\
=&\mathbb{E}\left\{\exp\left(-\frac{TX_T}{2}-10T+\frac{T^3}{24}\right)\right\}\\
=&\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi T}}\exp\left(-10T+\frac{T^3}{24}+\frac{T^3}{8}\right)\exp\left(-\frac{(x_T-\frac{T^2}{2})^2}{2T}\right)dx_T\\
=&\exp\left(-10T+\frac{T^3}{6}\right)
\end{aligned}$$

and the exact likelihood for two observations received within $[0,T]$ is

$$\begin{aligned}
\mathcal{L}=&\mathbb{E}\left\{\left[\prod_{i=1}^{2}\lambda\left(X_{t_i}\right)g^{\theta}\left(y_{t_i}|X_{t_i}\right)\exp\left(-\int_{t_{i-1}}^{t_i}\lambda(X_s)ds\right)\right]\exp\left(-\int_{t_2}^{T}\lambda\left(X_s\right)ds\right)\right\}\\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(v_1+10\right)\left(v_1+v_2+10\right)\\
&\times\frac{1}{2\pi\sigma_y^2}\exp\left(-\frac{(y_{t_1}-v_1)^2+(y_{t_2}-v_1-v_2)^2}{2\sigma_y^2}\right)\\
&\times\exp\left(-\frac{t_1v_1}{2}-\frac{t_2-t_1}{2}(2v_1+v_2)-\frac{T-t_2}{2}(2v_1+2v_2+v_3)-10T\right)\\
&\times\exp\left(\frac{t_1^3+(t_2-t_1)^3+(T-t_2)^3}{24}\right)\\
&\times\frac{1}{\sqrt{2\pi t_1}}\times\frac{1}{\sqrt{2\pi(t_2-t_1)}}\times\frac{1}{\sqrt{2\pi(T-t_2)}}\\
&\times\exp\left(-\frac{v_1^2}{2t_1}-\frac{v_2^2}{2(t_2-t_1)}-\frac{v_3^2}{2(T-t_2)}\right)dv_3dv_2dv_1
\end{aligned}$$

The first integral with respect to v_3 is

$$\int_{-\infty}^{\infty} \exp\left(-\frac{T-t_2}{2}v_3 - \frac{v_3^2}{2(T-t_2)}\right) dv_3 = \sqrt{2\pi(T-t_2)}\exp\left(\frac{1}{8}(T-t_2)^3\right).$$

The second integral with respect to v_2 is

$$\begin{aligned}
&\int_{-\infty}^{\infty} (v_1+v_2+10)\exp\left(-\frac{(y_{t_2}-v_1-v_2)^2}{2\sigma_y^2}\right) \times \exp\left(-\frac{t_2-t_1}{2}v_2 - \frac{T-t_2}{2}\times 2v_2\right)\\
&\times \exp\left(-\frac{v_2^2}{2(t_2-t_1)}\right) dv_2\\
=&\int_{-\infty}^{\infty} (v_1+v_2+10)\exp\left(-\left(\frac{1}{2\sigma_y^2}+\frac{1}{2(t_2-t_1)}\right)v_2^2 - \left(\frac{-y_{t_2}+v_1}{\sigma_y^2} - \frac{t_1+t_2-2T}{2}\right)v_2\right)\\
&\times \exp\left(-\frac{y_{t_2}^2-2v_1y_{t_2}+v_1^2}{2\sigma_y^2}\right) dv_2\\
=&\int_{-\infty}^{\infty} (v_1+v_2+10)\exp\left(-\frac{y_{t_2}^2-2v_1y_{t_2}+v_1^2}{2\sigma_y^2}+\frac{\mu_2^2}{2\sigma_2^2}\right)\exp\left(-\frac{(v_2-\mu_2)^2}{2\sigma_2^2}\right) dv_2\\
=&\sqrt{2\pi\sigma_2^2}\,(\mu_2+v_1+10)\exp\left(-\frac{y_{t_2}^2-2v_1y_{t_2}+v_1^2}{2\sigma_y^2}+\frac{\mu_2^2}{2\sigma_2^2}\right)
\end{aligned}$$

where

$$\sigma_2^2 = \left(\frac{1}{\sigma_y^2}+\frac{1}{t_2-t_1}\right)^{-1}, \qquad \mu_2 = \sigma_2^2\left(\frac{y_{t_2}-v_1}{\sigma_y^2}+\frac{t_1+t_2-2T}{2}\right) = \sigma_2^2(av_1+b)$$

and for

$$a = -\frac{1}{\sigma_y^2}, \qquad b = \frac{y_{t_2}}{\sigma_y^2}+\frac{t_1+t_2-2T}{2}.$$

The third integral with respect to v_1 is as follows,

$$\begin{aligned}
&\int_{-\infty}^{\infty} (v_1+10)(\mu_2+v_1+10)\exp\left(-\frac{(y_{t_1}-v_1)^2+y_{t_2}^2-2v_1y_{t_2}+v_1^2}{2\sigma_y^2}\right)\\
&\times\exp\left(\frac{\sigma_2^2}{2}\left(a^2v_1^2+2abv_1+b^2\right)\right)\\
&\times\exp\left(-\frac{t_1v_1}{2}-(t_2-t_1)v_1-(T-t_2)v_1\right)\exp\left(-\frac{v_1^2}{2t_1}\right)dv_1\\
=&\int_{-\infty}^{\infty}\left[(\sigma_2^2a+1)v_1^2+\left(10(\sigma_2^2a+1)+\sigma_2^2b+20\right)v_1+10\sigma_2^2b+100\right]\\
&\times\exp\left(-\left(\frac{1}{\sigma_y^2}-\frac{a^2\sigma_2^2}{2}+\frac{1}{2t_1}\right)v_1^2+\left(\frac{2y_{t_1}+2y_{t_2}}{2\sigma_y^2}+ab\sigma_2^2+\frac{t_1-2T}{2}\right)v_1\right)\\
&\times\exp\left(-\frac{y_{t_1}^2+y_{t_2}^2}{2\sigma_y^2}+\frac{b^2\sigma_2^2}{2}\right)dv_1\\
=&\sqrt{2\pi\sigma_1^2}\left[(\sigma_2^2a+1)(\mu_1^2+\sigma_1^2)+\left(10(\sigma_2^2a+1)+\sigma_2^2b+20\right)\mu_1+10\sigma_2^2b+100\right]\\
&\times\exp\left(-\frac{y_{t_1}^2+y_{t_2}^2}{2\sigma_y^2}+\frac{b^2\sigma_2^2}{2}+\frac{\mu_1^2}{2\sigma_1^2}\right)
\end{aligned}$$

where $\sigma_1^2=\left(\frac{2}{\sigma_y^2}-a^2\sigma_2^2+\frac{1}{t_1}\right)^{-1}$ and $\mu_1=\sigma_1^2\left(\frac{y_{t_1}+y_{t_2}}{\sigma_y^2}+ab\sigma_2^2+\frac{t_1-2T}{2}\right)$

Therefore,

$$\begin{aligned}
\mathcal{L}=&\frac{1}{2\pi\sigma_y^2}\times\frac{\sigma_1\sigma_2}{\sqrt{t_1(t_2-t_1)}}\left[\left(\sigma_2^2a+1\right)\left(\mu_1^2+\sigma_1^2\right)+\left(10\left(\sigma_2^2a+1\right)+\sigma_2^2b+20\right)\mu_1\right.\\
&\left.+10\sigma_2^2b+100\right]\times\exp\left(-\frac{y_{t_1}^2+y_{t_2}^2}{2\sigma_y^2}+\frac{b^2\sigma_2^2}{2}+\frac{\mu_1^2}{2\sigma_1^2}\right)\\
&\times\exp\left(-10T+\frac{t_1^3+(t_2-t_1)^3+(T-t_2)^3}{24}+\frac{1}{8}(T-t_2)^3\right).
\end{aligned}$$

The exact likelihood are used to compute the relative MSE for Section 5.1 of manuscript and Section S8 of Supplementary Material.

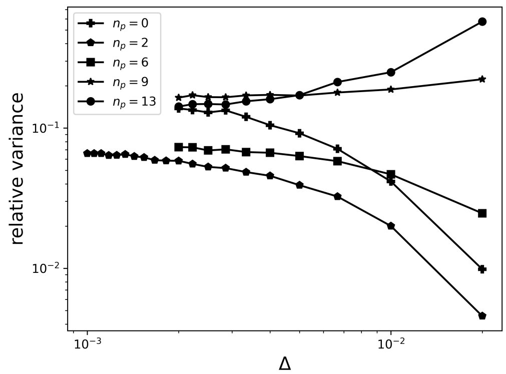


Figure 2: Plot of relative variance versus Δ for different values of n_p and fixed $N = 100$ in log scale.

S8 Empirical relationship between relative variance and Δ

Figure 2 reports the relationship between relative variance and Δ for different n_p values and fixed $N = 100$. For $n_p = 0$ and $n_p = 2$ cases, the exact likelihood is computed using solutions calculated in Section S7, for other larger values of n_p, the Monte Carlo estimate $\mathcal{L}_{\mathrm{MC}}$ is used in relative variance computation. Results show that the relationship between relative variance and Δ can be highly n_p-dependent. As n_p of problem increases, the rate of change in relative variance becomes less positive when Δ approaches zero. A more general trend that applies to all values of n_p is that the relative variance eventually becomes constant as Δ goes to zero.

S9 Born and Wolf observation model

In this section, we plot the point spread function of Born and Wolf observation model at different de-focus levels. For $(x_1, x_2) \in \mathbb{R}^2$,

$$q_{x_3}(x_1, x_2) = \frac{4\pi n_\alpha^2}{\lambda_e^2} \left| \int_0^1 J_0(\frac{2\pi n_\alpha}{\lambda_e} \sqrt{x_1^2 + x_2^2}\rho) \exp(\frac{j\pi n_\alpha^2 x_3}{n_0 \lambda_e}\rho^2)\rho d\rho \right|^2, \quad \text{(S9.2)}$$

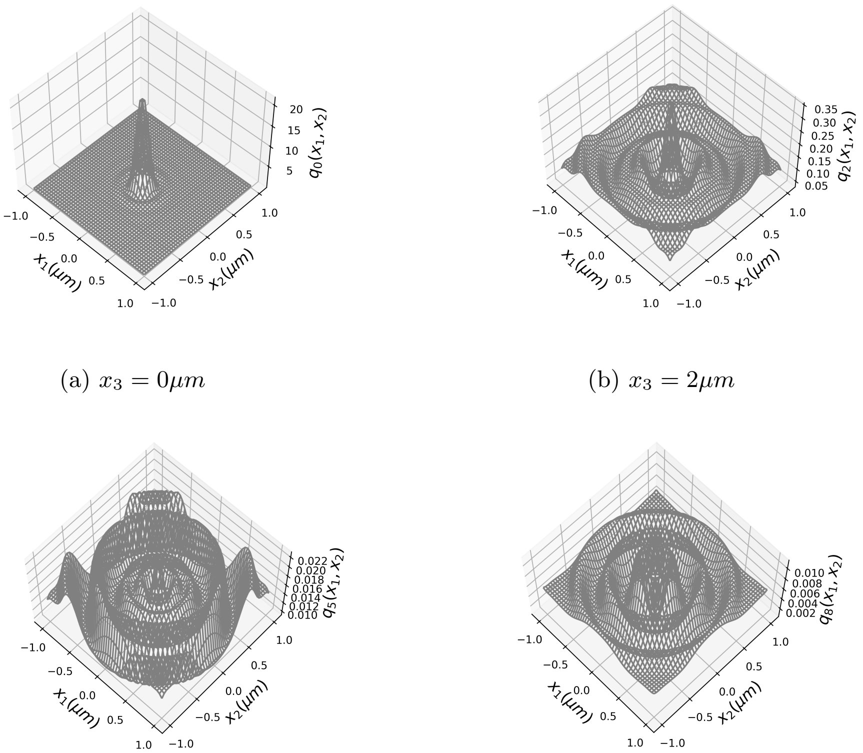
where n_0 is the refractive index of the objective lens immersion medium and n_α is the numerical aperture of the objective lens. λ_e is the emission wavelength of the molecule. $J_0(\cdot)$ and $J_1(\cdot)$ represents the zero-th order and the first order Bessel function of the first kind, respectively.

S10 Thinning algorithm for creating data

This section describes the thinning algorithm we use to generate observation data. Please refer Algorithm 1 for details.

S11 Additional experiments

In Figure 4a, we plot the true trajectory of a molecule, and simulate using parameters $\{\theta = (1.0, 1.0, 1.0)^\top, \mu = (0.5, 0.5, 6.0)^\top, p_0 = 0.01\}$ for time interval [0, 5.0]. Other parameters remain the same as in the manuscript. Figure 4c shows the filtered (x_1, x_2) mean locations of molecules, which deviate from their right positions. Figure 4d shows the filtered mean of x_3



(a) $x_3 = 0\mu m$

(b) $x_3 = 2\mu m$

(c) $x_3 = 5\mu m$

(d) $x_3 = 8\mu m$

Figure 3: Born and Wolf point spread function at different defocus levels. Mesh representations are shown for (S9.2) at different defocuses x_3, computed with wavelength $\lambda_e = 0.52\mu m$, numerical aperture $n_\alpha = 1.4$, refractive index of the objective lens immersion medium $n_0 = 1.515$. The x_3 values shown correspond to point source positions (a) $x_3 = 0\mu m$ (in focus), (b) $2\mu m$, (c) $5\mu m$ and (d) $8\mu m$.

and regions of ± 1 standard deviation together with the true state of X_3 at the observation times. In comparison to the Figure 6c and 6d of the manuscript, Figure 4c and 4d of Supplementary Material shows that higher

Algorithm 1: Thinning Algorithm for simulating the observation times with intensity function $\lambda(X_t)$ on $[0, T]$

```
Input: λ_max = λ_0, T
Generate N ~ Po(λ_max T);
Generate t_1, t_2, ..., t_N ~ U(0, T);
Sort t_1, t_2, ..., t_N and relabel them so that t_1 < t_2 < ... < t_N;
Generate X_0 ~ ν(x) and set τ = 0;
for i ∈ {1 : N} do
    Propagate X_{t_i} from previous τ, i.e. X_{t_i} ~ f^θ_{t_i − τ}(x_{t_i} | X_τ);
    Generate U ~ U(0, 1);
    if U ≤ λ(X_{t_i}) / λ_max then
        Keep t_i as a real observation time and set τ = t_i;
        Generate y_{t_i} which is a realisation of Y_{t_i} ~ g^θ(y | X_{t_i}) with Born and
          Wolf point spread function.
Output: all pairs of (t_i, y_{t_i})
```

values of x_3 degrade the estimation quality of particle filtering algorithm on the state of molecule and this is due to the exponential function structure of Born and Wolf image function which generates photons that are detected very far from the true molecule position.

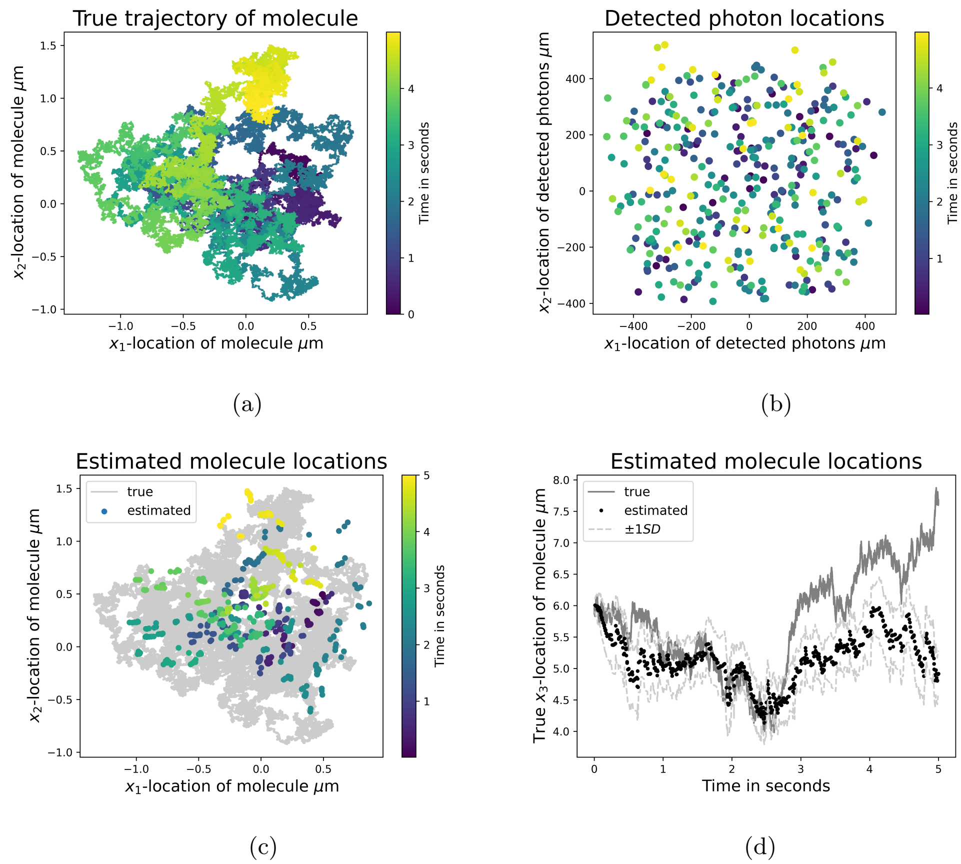


Figure 4: (a) True trajectory of a molecule; (b) observed photon locations; (c) estimated (x_1, x_2) molecule locations and (d) true x_3 molecule locations and estimated location.

References

Evans, L. C. (2012). *An introduction to stochastic differential equations*, Volume 82. American Mathematical Soc.

Folland, G. B. (1999). *Real analysis: modern techniques and their applications*, Volume 40. John Wiley & Sons.

Jazwinski, A. H. (2007). *Stochastic processes and filtering theory.* Courier Corporation.

Mörters, P. and Y. Peres (2010). *Brownian motion*, Volume 30. Cambridge University Press.