# SUPPLEMENTARY MATERIALS: FISHER INFORMATION MATRIX FOR SINGLE MOLECULES WITH STOCHASTIC TRAJECTORIES *

MILAD R. VAHID ¶†‡, BERNARD HANZON§, AND RAIMUND J. OBER †¶‖

**SM1. Proof of Theorem 2.2.** 1. According to [SM6, SM7] and Lemma SM14.1 (see Section SM14.1), the probability $P\Big(\mathcal{D}_{[t]} = \emptyset, N(t) = 0\Big)$ is given by

$$P\Big(\mathcal{D}_{[t]} = \emptyset, N(t) = 0\Big) = P\Big(N(t) = 0\Big) = e^{-\int_{t_0}^{t} \Lambda(\tau) d\tau}.$$

Also, the probability density function $p_{[t]}$ of $\mathcal{D}_{[t]}$ and $N(t)$ is given by

(SM1.1)
$$p_{[t]}\Big(d_K, K\Big) = p_{\mathcal{U}_K|\mathcal{T}_K,N(t)}\Big(r_1, \cdots, r_K|\tau_1, \cdots, \tau_K, K\Big) p_{\mathcal{T}_K|N(t)}\Big(\tau_1, \cdots, \tau_K|K\Big) P\Big(N(t) = K\Big),$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ and $K = 1, 2, \cdots$. According to Lemma SM14.1 (see Section SM14.1),

$$P\Big(N(t) = K\Big) = \frac{1}{K!} e^{-\int_{t_0}^{t} \Lambda(\tau) d\tau} \left(\int_{t_0}^{t} \Lambda(\tau) d\tau\right)^K, \quad K = 0, 1, \cdots, \tag{SM1.2}$$

and

$$p_{\mathcal{T}_K|N(t)}\Big(\tau_1, \cdots, \tau_K|K\Big) = \frac{K! \prod_{k=1}^{K} \Lambda(\tau_k)}{\Big(\int_{t_0}^{t} \Lambda(\tau) d\tau\Big)^K}, \quad K = 1, 2, \cdots, \tag{SM1.3}$$

and using the assumption of Definition 2.1 (it is assumed that $U_l, l = 1, \cdots, K$, is only dependent of the previous and current time points $T_1, \cdots, T_l$, and is independent of the future time points $T_{l+1}, \cdots, T_K$ and the total number $N(t)$ of the detected photons),

$$\begin{aligned}
&p_{\mathcal{U}_K|\mathcal{T}_K,N(t)}\Big(r_1, \cdots, r_K|\tau_1, \cdots, \tau_K, K\Big) \\
&\quad = p_{\mathcal{U}_K|\mathcal{T}_K}\Big(r_1, \cdots, r_K|\tau_1, \cdots, \tau_K\Big) \\
&\quad = p_{U_K|\mathcal{T}_K,\mathcal{D}_{K-1}}\Big(r_K|\tau_K, d_{K-1}\Big) p_{U_{K-1}|\mathcal{T}_{K-1},\mathcal{T}_K,\mathcal{D}_{K-2}}\Big(r_{K-1}|\tau_{K-1}, \tau_K, d_{K-2}\Big) \\
&\qquad \times \cdots \times p_{U_1|\mathcal{T}_K}\Big(r_1|\tau_1, \cdots, \tau_K\Big) \\
&\quad = p_{U_K|T_K,\mathcal{D}_{K-1}}\Big(r_K|\tau_K, d_{K-1}\Big) p_{U_{K-1}|T_{K-1},\mathcal{D}_{K-2}}\Big(r_{K-1}|\tau_{K-1}, d_{K-2}\Big) \\
&\qquad \times \cdots \times p_{U_1|T_1}\Big(r_1|\tau_1\Big) \\
&\quad = \prod_{l=1}^{K} p_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big), \qquad \text{(SM1.4)}
\end{aligned}$$

*Submitted to the editors February 2, 2019.
**Funding:** This work was supported in part by the National Institutes of Health (R01 GM085575).

†Department of Biomedical Engineering, Texas A&M University, College Station, TX 77843, USA (milad.rafiee@tamu.edu, raimund.ober@tamu.edu).
‡Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA (miladrv@stanford.edu).
§Edgeworth Centre for Financial Mathematics, School of Mathematical Sciences, University College Cork, Ireland (b.hanzon@ucc.ie).
¶Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, TX 77843, USA (milad.rafiee@tamu.edu, raimund.ober@tamu.edu).
‖Center for Cancer Immunology, Faculty of Medicine, University of Southampton, Southampton, UK (r.ober@soton.ac.uk).

where $p_{U_1|T_1,\mathcal{D}_0}\Big(r_1|\tau_1,d_0\Big) := p_{U_1|T_1}\Big(r_1|\tau_1\Big)$. By substituting Eqs. (SM1.2)-(SM1.4) into Eq. (SM1.1), we have

$$p_{[t]}\Big(d_K,K\Big) = e^{-\int_{t_0}^{t}\Lambda(\tau)d\tau}\prod_{k=1}^{K}\Lambda(\tau_k)\left[\prod_{l=1}^{K}p_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l,d_{l-1}\Big)\right].$$

2. The probability density function $p_L$ of $\mathcal{D}_L$ is given by

$$\begin{aligned} p_L\Big(d_L\Big) &= p_{\mathcal{U}_L|\mathcal{T}_L}\Big(r_1,\cdots,r_L|\tau_1,\cdots,\tau_L\Big)p_{\mathcal{T}_L}\Big(\tau_1,\cdots,\tau_L\Big) \\ &= \left[\prod_{l=1}^{L}p_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l,d_{l-1}\Big)\right]p_{\mathcal{T}_L}\Big(\tau_1,\cdots,\tau_L\Big), \end{aligned} \tag{SM1.5}$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$ and whereby Lemma SM14.1 (see Section SM14.1),

$$p_{\mathcal{T}_L}\Big(\tau_1,\cdots,\tau_L\Big) = e^{-\int_{t_0}^{\tau_L}\Lambda(\tau)d\tau}\prod_{k=1}^{L}\Lambda(\tau_k). \tag{SM1.6}$$

By substituting Eq. (SM1.6) into Eq. (SM1.5), we have

$$p_L\Big(d_L\Big) = e^{-\int_{t_0}^{\tau_L}\Lambda(\tau)d\tau}\prod_{k=1}^{L}\Lambda(\tau_k)\left[\prod_{l=1}^{L}p_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l,d_{l-1}\Big)\right],$$

and it completes the proof.

**SM2. Proof of Corollary 2.4.** The conditional probability density function $p_{U_l|T_l,\mathcal{D}_{l-1}}$ in Eqs. (2.1) and (2.2) of Theorem 2.2 can be written as, for $x := (x_0,y_0,z_0) \in \mathbb{R}^3$,

$$\begin{aligned} &p_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l,d_{l-1}\Big) \\ &= \int_{\mathbb{R}^3} p_{U_l,X(T_l)|T_l,\mathcal{D}_{l-1}}\Big(r_l,x|\tau_l,d_{l-1}\Big)dx \\ &= \int_{\mathbb{R}^3} p_{U_l|X(T_l),T_l,\mathcal{D}_{l-1}}\Big(r_l|x,\tau_l,d_{l-1}\Big)p_{X(T_l)|T_l,\mathcal{D}_{l-1}}\Big(x|\tau_l,d_{l-1}\Big)dx \\ &= \int_{\mathbb{R}^3} f_x\left(r_l\right)p_{pr_l}\Big(x|\tau_l,d_{l-1}\Big)dx \\ &= \frac{1}{|\det(M)|}\int_{\mathbb{R}^3} q_{z_0}\Big(M^{-1}r_l-(x_0,y_0)\Big)p_{pr_l}\Big(x|\tau_l,d_{l-1}\Big)dx, \end{aligned} \tag{SM2.1}$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ for $\mathcal{G}_{[t]}$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$ for $\mathcal{G}_L$), $p_{pr_l} := p_{X(T_l)|T_l,\mathcal{D}_{l-1}}, l = 1,2,\cdots,$ denotes the distribution of the prediction of the object location, and $p_{pr_1}\Big(x|\tau_1,d_0\Big) := p_{pr_1}\Big(x|\tau_1\Big)$, in which we have used the assumption of Definition 2.3.

**SM3. Proof of Theorem 3.2.** Let $\mathcal{G}_{[t]}\Big(\Big(\tilde{X},H,\tilde{W},Z\Big),\left(\mathcal{U}_{[t]},\mathcal{T}_{[t]}\right),\tilde{\Phi},\mathcal{C},\Theta\Big)$ (or $\mathcal{G}_L\Big(\Big(\tilde{X},H,\tilde{W},Z\Big),\left(\mathcal{U}_L,\mathcal{T}_L\right),\tilde{\Phi},\mathcal{C},\Theta\Big)$) be an image detection process with expanded state space $\tilde{X}$ for a time interval $[t_0,t]$ (or for a fixed number $L$ of photons). Let $\mathcal{D}_k := \left(\mathcal{U}_k,\mathcal{T}_k\right), k = 0,1,\cdots,$ and

$$\tilde{p}_{pr_l}\Big(\bar{x}|\tau_l,d_{l-1}\Big) := p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\Big(\bar{x}|\tau_l,d_{l-1}\Big), \quad \bar{x} \in \mathbb{R}^k,$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$), be the distribution of the prediction of the object location, and $\tilde{p}_{pr_1}\Big(\bar{x}|\tau_1, d_0\Big) := \tilde{p}_{pr_1}\Big(\bar{x}|\tau_1\Big)$.

1. If $H = \begin{bmatrix} I_{3\times 3} & 0_{3\times(k-3)} \end{bmatrix}$, then, for $x = H\bar{x}, x := (x_1, x_2, x_3)^T \in \mathbb{R}^3, \bar{x} := (\bar{x}_1, \cdots, \bar{x}_k)^T \in \mathbb{R}^k$, we have $\bar{x} = (x_1, x_2, x_3, \bar{x}_4, \cdots, \bar{x}_k)^T$ and the probability density function $p_{pr_l} := p_{X(T_l)|T_l,\mathcal{D}_{l-1}}$ is given by

$$p_{pr_l}\Big(x|\tau_l, d_{l-1}\Big) = \int_{\mathbb{R}^{k-3}} \tilde{p}_{pr_l}\Big(\bar{x}|\tau_l, d_{l-1}\Big) d\bar{x}_4 \cdots d\bar{x}_k. \tag{SM3.1}$$

Next, we extend the results obtained in the previous part to more general matrices $H$. Assume that there exists non-singular matrix $H_1 \in \mathbb{R}^{3\times 3}$ and matrix $H_2 \in \mathbb{R}^{3\times(k-3)}$ such that $H = \begin{bmatrix} H_1 & H_2 \end{bmatrix}$. Let $\dot{X}(\tau) = S\tilde{X}(\tau), \tau \geq t_0$, where

$$S := \begin{bmatrix} H_1 & H_2 \\ 0_{(k-3)\times 3} & I_{(k-3)\times(k-3)} \end{bmatrix} \in \mathbb{R}^{k\times k}.$$

Hence, $S$ is a non-singular matrix with determinant $\det(S) = \det(H_1) \neq 0$ and $S^{-1}$ exists. Then, for $\dot{x} = S\bar{x}, \bar{x} := (\bar{x}_1, \cdots, \bar{x}_k)^T \in \mathbb{R}^k, \dot{x} := (\dot{x}_1, \cdots, \dot{x}_k)^T \in \mathbb{R}^k$, we have

$$\dot{x} = S\bar{x} = (H\bar{x}, \bar{x}_4, \cdots, \bar{x}_k)^T = (x_1, x_2, x_3, \dot{x}_4, \cdots, \dot{x}_k)^T,$$

and, according to Eq. (SM3.1),

$$\begin{aligned} p_{pr_l}\Big(x|\tau_l, d_{l-1}\Big) &= \int_{\mathbb{R}^{k-3}} \dot{p}_{pr_l}\Big(\dot{x}|\tau_l, d_{l-1}\Big) d\dot{x}_4 \cdots d\dot{x}_k \\ &= \int_{\mathbb{R}^{k-3}} \tilde{p}_{pr_l}\Big(S^{-1}\bar{x}|\tau_l, d_{l-1}\Big) |H_1|^{-1} d\bar{x}_4 \cdots d\bar{x}_k, \end{aligned}$$

where $\dot{p}_{pr_l} := p_{\dot{X}(T_l)|T_l,\mathcal{D}_{l-1}}$ and $S^{-1}$ is given by

$$S^{-1} = \begin{bmatrix} H_1^{-1} & -H_1^{-1}H_2 \\ 0_{(k-3)\times 3} & I_{(k-3)\times(k-3)} \end{bmatrix}.$$

2. Then, $\tilde{p}_{pr_l}$ can be calculated through the following steps:
Step 1. For $l = 0$, Eq. (3.5) becomes

$$\tilde{X}(T_1) = \tilde{\phi}(t_0, T_1)\tilde{X}(t_0) + \tilde{W}(t_0, T_1).$$

Then, by conditioning the both sides of the above equation on $T_1 = \tau_1$, the conditional probability density function $p_{\tilde{X}(T_1)|T_1}$ is given by

$$p_{\tilde{X}(T_1)|T_1}\Big(\bar{x}|\tau_1\Big) = \Big(p_{\tilde{\phi}(t_0,\tau_1)\tilde{X}(t_0)} * p_{\tilde{W}(t_0,\tau_1)}\Big)(\bar{x}),$$

where $\bar{x} \in \mathbb{R}^k$, and $*$ denotes the convolution operator. Then,

$$\begin{aligned} \tilde{p}_{pr_1}\Big(\bar{x}|\tau_1\Big) &:= p_{\tilde{X}(T_1)|T_1}\Big(\bar{x}|\tau_1\Big) \\ &= \int_{\mathbb{R}^k} p_{\tilde{\phi}(t_0,\tau_1)\tilde{X}(t_0)}\Big(\bar{x}_o\Big) p_{\tilde{W}(t_0,\tau_1)}\Big(\bar{x} - \bar{x}_o\Big) d\bar{x}_o \\ &= \frac{1}{\Big|\det\Big(\tilde{\phi}(t_0,\tau_1)\Big)\Big|} \int_{\mathbb{R}^k} p_{\tilde{X}(t_0)}\Big(\tilde{\phi}^{-1}(t_0,\tau_1)\bar{x}_o\Big) p_{\tilde{W}(t_0,\tau_1)}\Big(\bar{x} - \bar{x}_o\Big) d\bar{x}_o. \end{aligned}$$

Step $2l$. For $l = 1, 2, \cdots$, let

$$A_l := \left\{\tilde{X}(T_l) = \bar{x}\right\}, \quad B_l := \{U_l = r_l\}, \quad \text{and} \qquad C_l := \{T_l = \tau_l\} \cap \{\mathcal{D}_{l-1} = d_{l-1}\}.$$

Then, according to Bayes' rule, we have the relation between the conditional probability densities of $A_l, B_l$, and $C_l$ as follows

$$p\Big(A_l|B_l, C_l\Big) = \frac{p\Big(B_l|A_l, C_l\Big)p\Big(A_l|C_l\Big)}{p\Big(B_l|C_l\Big)},$$

i.e.,

$$\begin{aligned}
p_{\tilde{X}(T_l)|\mathcal{D}_l}\Big(\bar{x}|d_l\Big) &= \frac{p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\Big(r_l|\bar{x},\tau_l,d_{l-1}\Big)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\Big(\bar{x}|\tau_l,d_{l-1}\Big)}{p_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l,d_{l-1}\Big)} \\
&= \frac{p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\Big(r_l|\bar{x},\tau_l,d_{l-1}\Big)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\Big(\bar{x}|\tau_l,d_{l-1}\Big)}{\int_{\mathbb{R}^k} p_{U_l,\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\Big(r_l,\bar{x}_o|\tau_l,d_{l-1}\Big)d\bar{x}_o} \\
&= \frac{p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\Big(r_l|\bar{x},\tau_l,d_{l-1}\Big)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\Big(\bar{x}|\tau_l,d_{l-1}\Big)}{\int_{\mathbb{R}^k} p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\Big(r_l|\bar{x}_o,\tau_l,d_{l-1}\Big)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\Big(\bar{x}_o|\tau_l,d_{l-1}\Big)d\bar{x}_o}. \qquad \text{(SM3.2)}
\end{aligned}$$

Since initial location of the object, observation noise, and process noise are mutually independent, according to Eq. (3.6) and Theorem 2.7 of [SM3], we have

$$p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\Big(r_l|\bar{x},\tau_l,d_{l-1}\Big) = p_{U_l|\tilde{X}(\tau_l)}\Big(r_l|\bar{x}\Big) = p_{Z(H\bar{x})}\left(r_l\right), \quad \bar{x} \in \mathbb{R}^k. \qquad \text{(SM3.3)}$$

Therefore, by substituting Eq. (SM3.3) into Eq. (SM3.2) (note that we calculated $\tilde{p}_{pr_l}$ in the previous step),

$$\begin{aligned}
p_{fi_l}\Big(\bar{x}|d_l\Big) &:= p_{\tilde{X}(T_l)|\mathcal{D}_l}\Big(\bar{x}|d_l\Big) \\
&= \frac{p_{Z(H\bar{x})}\left(r_l\right)\tilde{p}_{pr_l}\Big(\bar{x}|\tau_l,d_{l-1}\Big)}{\int_{\mathbb{R}^k} p_{Z(H\bar{x}_o)}\left(r_l\right)\tilde{p}_{pr_l}\Big(\bar{x}_o|\tau_l,d_{l-1}\Big)d\bar{x}_o}.
\end{aligned}$$

Step $2l+1$. By conditioning the both sides of Eq. (3.5) on $T_{l+1} = \tau_{l+1}$ and $\mathcal{D}_l = d_l$, we have , for $l = 1, 2, \cdots$,

$$\begin{aligned}
&p_{\tilde{X}(T_{l+1})|T_{l+1},\mathcal{D}_l}\Big(\bar{x}|\tau_{l+1},d_l\Big) \\
&\qquad = p_{\tilde{\phi}(T_l,T_{l+1})\tilde{X}(T_l)|T_{l+1},\mathcal{D}_l}\Big(\bar{x}|\tau_{l+1},d_l\Big) * p_{\tilde{W}(T_l,T_{l+1})|T_{l+1},\mathcal{D}_l}\Big(\bar{x}|\tau_{l+1},d_l\Big),
\end{aligned}$$

which, according to the independence of $\tilde{W}(T_l, T_{l+1})$ and $\mathcal{U}_l, \mathcal{T}_{l-1}$, becomes

$$\begin{aligned}
&p_{\tilde{X}(T_{l+1})|T_{l+1},\mathcal{D}_l}\Big(\bar{x}|\tau_{l+1},d_l\Big) \\
&\quad = p_{\tilde{\phi}(\tau_l,\tau_{l+1})\tilde{X}(T_l)|\mathcal{D}_l}\Big(\bar{x}|d_l\Big) * p_{\tilde{W}(\tau_l,\tau_{l+1})}\Big(\bar{x}\Big) \\
&\quad = \int_{\mathbb{R}^k} p_{\tilde{\phi}(\tau_l,\tau_{l+1})\tilde{X}(T_l)|\mathcal{D}_l}\Big(\bar{x}_o|d_l\Big)p_{\tilde{W}(\tau_l,\tau_{l+1})}\Big(\bar{x}-\bar{x}_o\Big)d\bar{x}_o \\
&\quad = \frac{1}{\Big|\det\Big(\tilde{\phi}(\tau_l,\tau_{l+1})\Big)\Big|}\int_{\mathbb{R}^k} p_{\tilde{X}(T_l)|\mathcal{D}_l}\Big(\tilde{\phi}^{-1}(\tau_l,\tau_{l+1})\bar{x}_o|d_l\Big)p_{\tilde{W}(\tau_l,\tau_{l+1})}\Big(\bar{x}-\bar{x}_o\Big)d\bar{x}_o,
\end{aligned}$$

or equivalently (note that we calculated $\tilde{p}_{fi_l}$ in the previous step),

(SM3.4)
$$\tilde{p}_{pr_{l+1}}\left(\bar{x}|\tau_{l+1},d_l\right)=\frac{1}{\left|\det\left(\tilde{\phi}(\tau_l,\tau_{l+1})\right)\right|}\int_{\mathbb{R}^k}\tilde{p}_{fi_l}\left(\tilde{\phi}^{-1}(\tau_l,\tau_{l+1})\bar{x}_o|d_l\right)p_{\tilde{W}(\tau_l,\tau_{l+1})}\left(\bar{x}-\bar{x}_o\right)d\bar{x}_o.$$

3.1. See Theorem 7.2 of [SM3].
3.2. Setting $C:=M'H$, Eq. (3.7) becomes

$$U_l=Z(H\tilde{X}(\tau_l))=C\tilde{X}(\tau_l)+Z_{g,l},\quad l=1,2,\cdots,\tag{SM3.5}$$

where

$$p_{Z_{g,l}}(r):=\frac{1}{2\pi\left[\det(\Sigma_g)\right]^{1/2}}\exp\left(-\frac{1}{2}r^T\Sigma_g^{-1}r\right),\quad r\in\mathcal{C}.$$

Since $Z_{g,l}$ is independent of $\mathcal{D}_{l-1}$, $T_l$ and $\tilde{X}$, then, according to Eq. (SM3.5), $Z(H\tilde{X}(\tau_l))$ is the sum of two independent Gaussian random variables and its probability density function is given by

$$\begin{aligned}p_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)&=p_{Z(H\tilde{X}(\tau_l))|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)\\&=\frac{1}{2\pi\left[\det\left(R_l\right)\right]^{1/2}}\exp\left(-\frac{1}{2}(r_l-\hat{r}_l)^TR_l^{-1}(r_l-\hat{r}_l)\right),\end{aligned}\tag{SM3.6}$$

where $R_l:=CP_l^{l-1}C^T+\Sigma_g$ and $\hat{r}_l:=C\hat{x}_l^{l-1}$.

**SM4. Example of maximum likelihood estimation.** Let the 3D (or 2D) motion of an object be given by the following continuous-time stochastic differential equation

$$dX(\tau)=(V+FX(\tau))\,d\tau+\sqrt{2D}dB(\tau),\qquad\tau\geq t_0,\tag{SM4.1}$$

where $V\in\mathbb{R}^3$ (or $\mathbb{R}^2$), $F\in\mathbb{R}^{3\times3}$ (or $\mathbb{R}^{2\times2}$) and $D>0$ denote the zero order drift, first order drift and diffusion coefficients, respectively, and $\{B(\tau),\tau\geq t_0\}$ is a 3 (or 2)-vector Brownian motion process with
$E\left\{dB(\tau)dB(\tau)^T\right\}=I_{3\times3}$ (or $I_{2\times2}$).

As an example, next, for the case that the first order drift is a diagonal matrix, i.e., $FI_{2\times2},F\in\mathbb{R}$, and there is no zero order drift, we derive expressions for the likelihood function and its derivative with respect to the drift and diffusion coefficients for 2D trajectories. Let $X(t_0)$ be Gaussian distributed with mean $x_0\in\mathbb{R}^2$ and diagonal covariance matrix $P_0:=\rho_0I_{2\times2},\rho_0>0$, which is assumed to be independent of $B(\tau)$. Assume that the photon detection rate $\Lambda$, the magnification matrix $M=mI_{2\times2},m>0$, and the covariance matrix $\Sigma_g=vI_{2\times2},v>0$, of the measurement noise are independent of the parameter vector $\theta\in\Theta$. Also, for the corresponding discrete system at time points $\tau_0:=t_0\leq\tau_1<\tau_2<\cdots<\tau_K<\cdots$, let the transition matrix be given by $\phi(\tau_{l-1},\tau_l):=\phi^s(\tau_{l-1},\tau_l)I_{2\times2},\phi^s(\tau_{l-1},\tau_l)\in\mathbb{R}$, and the process noise covariance matrix be given by $Q_g(\tau_{l-1},\tau_l):=q^s(\tau_{l-1},\tau_l)I_{2\times2},q^s(\tau_{l-1},\tau_l)>0$. Then, the covariances of the states, which can be calculated through the Kalman filter formulae recursively, are also scalar matrices, i.e., can be defined as $P_{\theta,l}^{l-1}:=\rho_{\theta,l}^{l-1}I_{2\times2},\rho_{\theta,l}^{l-1}>0,l=1,2,\cdots$. Also, let $\hat{x}_{\theta,l}^{l-1}$ denote the means of the states.

Then, the maximum likelihood estimate $\hat{\theta}_{mle}$ of $\theta = (\theta_1, \cdots, \theta_n)$ is the solution of the following equation, according to Eq. (SM11.6), for the acquired data denoted by $d_K \in \mathcal{C}^K \times \mathbb{R}^K_{[\infty]}, K = 1, 2, \cdots,$

$$\frac{\partial \log \mathcal{L}(\theta|d_K)}{\partial \theta_i} = \sum_{l=1}^{K} -\frac{1}{2}\text{trace}\left[\left(\frac{m^2 d\rho_{\theta,l}^{l-1,i}}{m^2\rho_{\theta,l}^{l-1}+v}\right)\left(I_{2\times 2} - \frac{\left(r_l - m\hat{x}_{\theta,l}^{l-1}\right)\left(r_l - m\hat{x}_{\theta,l}^{l-1}\right)^T}{m^2\rho_{\theta,l}^{l-1}+v}\right)\right] - \frac{m d\hat{x}_{\theta,l}^{l-1,i}\left(r_l - m\hat{x}_{\theta,l}^{l-1}\right)}{m^2\rho_{\theta,l}^{l-1}+v}$$

$$= \sum_{l=1}^{K} -\frac{1}{2}\left(\frac{m^2 d\rho_{\theta,l}^{l-1,i}}{m^2\rho_{\theta,l}^{l-1}+v}\right)\left(2 - \frac{\left\|r_l - m\hat{x}_{\theta,l}^{l-1}\right\|^2}{m^2\rho_{\theta,l}^{l-1}+v}\right) - \frac{m d\hat{x}_{\theta,l}^{l-1,i}\left(r_l - m\hat{x}_{\theta,l}^{l-1}\right)}{m^2\rho_{\theta,l}^{l-1}+v} = 0, \tag{SM4.2}$$

where $i = 1, \cdots, n$, $\|.\|$ denotes the Euclidean norm, and for $l = 1, 2, \cdots$, $\hat{X}_{\theta,l}^{(i)} := \begin{bmatrix}\hat{x}_{\theta,l}^{l-1} & d\hat{x}_{\theta,l}^{l-1,i}\end{bmatrix}^T$, $d\hat{x}_{\theta,l}^{l-1,i} := \frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_i}$, $P_{\theta,l}^{(i)} := \begin{bmatrix}\rho_{\theta,l}^{l-1} & d\rho_{\theta,l}^{l-1,i}\end{bmatrix}^T$, and $d\rho_{\theta,l}^{l-1,i} := \frac{\partial \rho_{\theta,l}^{l-1}}{\partial \theta_i}$ can be calculated through the following recursive formulas, by combining the Kalman filtering equations (Eqs. (3.11) and (3.12)) and their derivatives, and using Lemma SM14.2 (see Section SM14.2),

$$\begin{aligned}\hat{X}_{\theta,l+1}^{(i)} &= A_{\theta,l+1}^{(i)}\hat{X}_{\theta,l}^{(i)} + B_{\theta,l+1}^{(i)}\left(r_l - m\hat{x}_{\theta,l}^{l-1}\right),\\ P_{\theta,l+1}^{(i)} &= C_{\theta,l+1}^{(i)}P_{\theta,l}^{(i)} + \begin{bmatrix} q_\theta^s(\tau_l,\tau_{l+1}) \\ \frac{\partial q_\theta^s(\tau_l,\tau_{l+1})}{\partial\theta_i}\end{bmatrix},\end{aligned} \tag{SM4.3}$$

where $\hat{X}_{\theta,1}^{(i)} = \begin{bmatrix}\phi_\theta(t_0,\tau_1)x_{\theta,0} \\ \frac{\partial\phi_\theta(t_0,\tau_1)}{\partial\theta_i}x_{\theta,0} + \phi_\theta(t_0,\tau_1)\frac{\partial x_{\theta,0}}{\partial\theta_i}\end{bmatrix}$, $P_{\theta,1}^{(i)} = \begin{bmatrix}\left(\phi_\theta^s(t_0,\tau_1)\right)^2\rho_0 + q_\theta^s(t_0,\tau_1) \\ \frac{\partial\left(\phi_\theta^s(t_0,\tau_1)\right)^2}{\partial\theta_i}\rho_0 + \frac{\partial q_\theta^s(t_0,\tau_1)}{\partial\theta_i}\end{bmatrix}$, and

$$\begin{aligned}A_{\theta,l+1}^{(i)} &:= \begin{bmatrix}\phi_\theta(\tau_l,\tau_{l+1}) & 0_{2\times2} \\ \frac{\partial\phi_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i} & \phi_\theta(\tau_l,\tau_{l+1})\left(I_{2\times2} - K_{\theta,l}M\right)\end{bmatrix},\\ B_{\theta,l+1}^{(i)} &:= \begin{bmatrix}\phi_\theta(\tau_l,\tau_{l+1})K_{\theta,l} \\ \phi_\theta(\tau_l,\tau_{l+1})\frac{\partial K_{\theta,l}}{\partial\theta_i} + \frac{\partial\phi_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i}K_{\theta,l}\end{bmatrix},\\ C_{\theta,l+1}^{(i)} &:= \begin{bmatrix}\left(1 - mk_{\theta,l}^s\right)\left(\phi_\theta^s(\tau_l,\tau_{l+1})\right)^2 & 0 \\ \left(1 - mk_{\theta,l}^s\right)\frac{\partial\left(\phi_\theta^s(\tau_l,\tau_{l+1})\right)^2}{\partial\theta_i} - m\frac{\partial k_{\theta,l}^s}{\partial\theta_i}\left(\phi_\theta^s(\tau_l,\tau_{l+1})\right)^2 & \left(1 - mk_{\theta,l}^s\right)\left(\phi_\theta^s(\tau_l,\tau_{l+1})\right)^2\end{bmatrix},\end{aligned} \tag{SM4.4}$$

where the Kalman gain and its derivative are given by

$$K_{\theta,l} = k_{\theta,l}^s I_{2\times2}, \quad k_{\theta,l}^s := \frac{m\rho_{\theta,l}^{l-1}}{m^2\rho_{\theta,l}^{l-1}+v}, \quad \frac{\partial k_{\theta,l}^s}{\partial\theta_i} = \frac{mvd\rho_{\theta,l}^{l-1,i}}{\left(m^2\rho_{\theta,l}^{l-1}+v\right)^2}. \tag{SM4.5}$$

1. If $F \neq 0$, then, for $l = 1, 2, \cdots$,

$$\phi^s(\tau_{l-1},\tau_l) = e^{F(\tau_l - \tau_{l-1})}, \quad q^s(\tau_{l-1},\tau_l) = \frac{D}{F}\left(e^{2F(\tau_l-\tau_{l-1})} - 1\right).$$

(a) If the only unknown parameter is the first order drift coefficient $F$, i.e., $\theta = F$, then, for $\Delta\tau_{l+1} := \tau_{l+1} - \tau_l$,

$$A_{\theta,l+1} = \begin{bmatrix} e^{F\Delta\tau_{l+1}} I_{2\times 2} & 0_{2\times 2} \\ \Delta\tau_{l+1} e^{F\Delta\tau_{l+1}} I_{2\times 2} & e^{F\Delta\tau_{l+1}} \left(I_{2\times 2} - mK_{\theta,l}\right) \end{bmatrix},$$

$$B_{\theta,l+1} = \begin{bmatrix} e^{F\Delta\tau_{l+1}} K_{\theta,l} \\ e^{F\Delta\tau_{l+1}} \left(\frac{\partial K_{\theta,l}}{\partial\theta} + \Delta\tau_{l+1} K_{\theta,l}\right) \end{bmatrix},$$

$$C_{\theta,l+1} = \begin{bmatrix} \left(1 - mk^s_{\theta,l}\right) e^{2F\Delta\tau_{l+1}} & 0 \\ \left(1 - mk^s_{\theta,l}\right) 2\Delta\tau_{l+1} e^{2F\Delta\tau_{l+1}} - m\frac{\partial k^s_{\theta,l}}{\partial\theta} e^{2F\Delta\tau_{l+1}} & \left(1 - mk^s_{\theta,l}\right) e^{2F\Delta\tau_{l+1}} \end{bmatrix},$$

and

$$\hat{X}_{\theta,1} = \begin{bmatrix} e^{F\Delta\tau_1} x_0 \\ \Delta\tau_1 e^{F\Delta\tau_1} x_0 \end{bmatrix}, \quad P_{\theta,1} = \begin{bmatrix} e^{2F\Delta\tau_1}\rho_0 + \frac{D}{F}\left(e^{2F\Delta\tau_1} - 1\right) \\ 2\Delta\tau_1 e^{2F\Delta\tau_1}\rho_0 + \frac{D}{F} e^{2F\Delta\tau_1}\left(-\frac{1}{F} + 2\Delta\tau_1\right) + \frac{D}{F^2} \end{bmatrix}. \tag{SM4.6}$$

(b) If the only unknown parameter is the diffusion coefficient $D$, i.e., $\theta = D$, then,

$$A_{\theta,l+1} = \begin{bmatrix} e^{F\Delta\tau_{l+1}} I_{2\times 2} & 0_{2\times 2} \\ 0_{2\times 2} & e^{F\Delta\tau_{l+1}} \left(I_{2\times 2} - mK_{\theta,l}\right) \end{bmatrix}, \quad B_{\theta,l+1} = \begin{bmatrix} e^{F\Delta\tau_{l+1}} K_{\theta,l} \\ e^{F\Delta\tau_{l+1}} \frac{\partial K_{\theta,l}}{\partial\theta} \end{bmatrix},$$

$$C_{\theta,l+1} = \begin{bmatrix} \left(1 - mk^s_{\theta,l}\right) e^{2F\Delta\tau_{l+1}} & 0 \\ -m\frac{\partial k^s_{\theta,l}}{\partial\theta} e^{2F\Delta\tau_{l+1}} & \left(1 - mk^s_{\theta,l}\right) e^{2F\Delta\tau_{l+1}} \end{bmatrix},$$

and

$$\hat{X}_{\theta,1} = \begin{bmatrix} e^{F\Delta\tau_1} x_0 \\ 0_{2\times 1} \end{bmatrix}, \quad P_{\theta,1} = \begin{bmatrix} e^{2F\Delta\tau_1}\rho_0 + \frac{D}{F}\left(e^{2F\Delta\tau_1} - 1\right) \\ 2\Delta\tau_1 e^{2F\Delta\tau_1}\rho_0 + \frac{1}{F}\left(e^{2F\Delta\tau_1} - 1\right) \end{bmatrix}. \tag{SM4.7}$$

2. If $F = 0$, then, for $l = 1, 2, \cdots$,

$$\phi^s(\tau_{l-1}, \tau_l) = 1, \quad q^s(\tau_{l-1}, \tau_l) = 2D\left(\tau_l - \tau_{l-1}\right).$$

If the only unknown parameter is the diffusion coefficient $D$, i.e., $\theta = D$, then,

$$A_{\theta,l+1} = \begin{bmatrix} I_{2\times 2} & 0_{2\times 2} \\ 0_{2\times 2} & I_{2\times 2} - mK_{\theta,l} \end{bmatrix}, \quad B_{\theta,l+1} = \begin{bmatrix} K_{\theta,l} \\ \frac{\partial K_{\theta,l}}{\partial\theta} \end{bmatrix}, \quad C_{\theta,l+1} = \begin{bmatrix} 1 - mk^s_{\theta,l} & 0 \\ -m\frac{\partial k^s_{\theta,l}}{\partial\theta} & 1 - mk^s_{\theta,l} \end{bmatrix},$$

and

$$\hat{X}_{\theta,1} = \begin{bmatrix} x_0 \\ 0_{2\times 1} \end{bmatrix}, \quad P_{\theta,1} = \begin{bmatrix} \rho_0 + 2D\Delta\tau_1 \\ 2\Delta\tau_1 \end{bmatrix}. \tag{SM4.8}$$

**SM5. More estimation results for Gaussian measurements.**

**SM6. Sequential Monte Carlo method.** Here, for the acquired data denoted by $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$), $l = 1, 2, \cdots$, we approximate the distribution $p_{pr_{l+1}}\left(x_{l+1}|\tau_{l+1}, d_l\right)$ through the sequential Monte Carlo method provided in [SM5]. Note that

$$\begin{aligned} p_{pr_{l+1}}\left(x_{l+1}|\tau_{l+1}, d_l\right) &= p_{X(T_{l+1})|T_{l+1},\mathcal{D}_l}\left(x_{l+1}|\tau_{l+1}, d_l\right) \\ &= \int_{\mathbb{R}^3} p_{X(T_{l+1}),X(T_l)|T_{l+1},\mathcal{D}_l}\left(x_{l+1}, x|\tau_{l+1}, d_l\right) dx \\ &= \int_{\mathbb{R}^3} p_{X(T_{l+1})|X(T_l),T_{l+1},\mathcal{D}_l}\left(x_{l+1}|x, \tau_{l+1}, d_l\right) p_{X(T_l)|T_{l+1},\mathcal{D}_l}\left(x|\tau_{l+1}, d_l\right) dx \\ &= \int_{\mathbb{R}^3} p_{X(T_{l+1})|X(T_l),T_{l+1},\mathcal{D}_l}\left(x_{l+1}|x, \tau_{l+1}, d_l\right) p_{fi_l}\left(x|d_l\right) dx, \end{aligned} \tag{SM6.1}$$

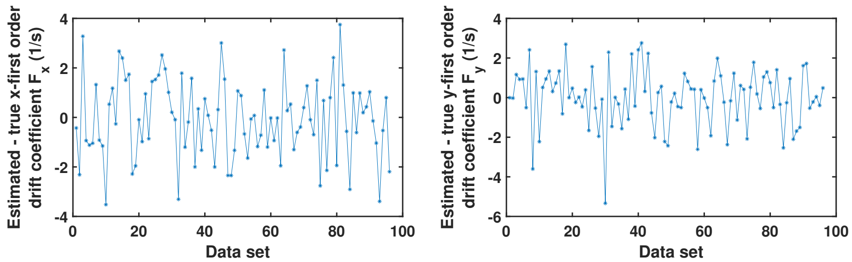


FIG. SM1. *Analysis of the error of the first order drift matrix estimates produced by the maximum likelihood estimation method for the Gaussian measurement model. Differences between the estimates of the first order drift coefficients in x and y directions and their true values for the data sets of 2D trajectory of an in-focus molecule in the object space simulated using Eq. (SM4.1) where the time points are drawn from a Poisson process with mean 500 in the time interval* $[0, 100]$ *ms with the first order drift matrix* $F = [F_{ij}], i, j = 1, 2, F_{12} = F_{21} = 0, F_{11} = F_x = -10/s$, $F_{22} = F_y = -5/s$ *and the diffusion coefficient* $D = 1\ \mu m^2/s$. *Also, we assume that there is no zero order drift and the initial location of the molecule is Gaussian distributed with mean* $x_0 = (4.4, 4.4)^T$ $\mu m$ *and covariance* $P_0 = 10 I_{2\times 2}\ nm^2$.

where $p_{fi_l}\left(x|d_l\right) := p_{X(T_l)|\mathcal{D}_l}\left(x|d_l\right), x \in \mathbb{R}^3$, and for the linear stochastic system with state $X(\tau) \in \mathbb{R}^3, \tau \geq t_0$, zero-mean Gaussian process noise with covariance matrix $Q_g(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3\times 3}, Q_g(\tau_l, \tau_{l+1}) > 0$, and state-transition matrix $\phi(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3\times 3}$, we have, for $x_l \in \mathbb{R}^3$,

$$\begin{aligned} &p_{X(T_{l+1})|X(T_l),T_{l+1},\mathcal{D}_l}\left(x_{l+1}|x_l, \tau_{l+1}, d_l\right) \\ &\quad = p_{X(T_{l+1})|X(T_l)}\left(x_{l+1}|x_l\right) \\ &\quad = \frac{1}{(2\pi)^{3/2}\left[\det\left(Q_g(\tau_l, \tau_{l+1})\right)\right]^{1/2}} \\ &\qquad \times \exp\left(-\frac{1}{2}(x_{l+1} - \phi(\tau_l, \tau_{l+1})x_l)^T Q_g^{-1}(\tau_l, \tau_{l+1})(x_{l+1} - \phi(\tau_l, \tau_{l+1})x_l)\right). \end{aligned} \tag{SM6.2}$$

The distribution $p_{fi_l}$ of the filtered object location can be approximated as [SM5]

$$p_{fi_l}\left(x_l|d_l\right) \approx \sum_{i=1}^{N} w_l^i(r_l)\delta\left(x_l - \dot{x}_l^i\right), \tag{SM6.3}$$

where $\delta$ is the Dirac delta function, and the samples $\dot{x}_l^i$ and their corresponding weights $w_l^i(r_l), i = 1, \cdots, N$, are given through the following sequential Monte Carlo algorithm. Finally, by substituting Eqs. (SM6.2) and (SM6.3) into Eq. (SM6.1), the distribution $p_{pr_{l+1}}$ can be approximated as

$$\begin{aligned} p_{pr_{l+1}}\left(x_{l+1}|\tau_{l+1}, d_l\right) &\approx \sum_{i=1}^{N} w_l^i(r_l) p_{X(T_{l+1})|X(T_l)}\left(x_{l+1}|\dot{x}_l^i\right) \\ &= \sum_{i=1}^{N} \frac{w_l^i(r_l)}{(2\pi)^{3/2}\left[\det\left(Q_g(\tau_l, \tau_{l+1})\right)\right]^{1/2}} \exp\Bigg( -\frac{1}{2}(x_{l+1} - \phi(\tau_l, \tau_{l+1})\dot{x}_l^i)^T \\ &\qquad \times Q_g^{-1}(\tau_l, \tau_{l+1})(x_{l+1} - \phi(\tau_l, \tau_{l+1})\dot{x}_l^i)\Bigg). \end{aligned}$$

Sequential Monte Carlo (particle filter) algorithm: [SM5]

1. Draw initial samples $\{x_0^i\}_{i=1}^N$ according to $p_{X(t_0)}(x_0)$, i.e., $x_0^i \sim p_{X(t_0)}(x_0), i = 1, \cdots, N$, and set $l = 1$.

2. Draw independent and identically distributed samples $\{\dot{x}_l^i\}_{i=1}^N$ according to $p_{X(T_l)|X(T_{l-1})}\left(\dot{x}_l|x_{l-1}^i\right)$, i.e., $\dot{x}_l^i \sim p_{X(T_l)|X(T_{l-1})}\left(\dot{x}_l|x_{l-1}^i\right), i = 1, \cdots, N$.

3. Compute the weights sequence $\{w_l^i(r_l)\}_{i=1}^N$ as

$$w_l^i(r_l) = \frac{f_{\dot{x}_l^i}(r_l)}{\sum_{i=1}^N f_{\dot{x}_l^i}(r_l)}, \quad i = 1, \cdots, N.$$

4. Resample new particles $x_l^j, j = 1, \cdots, N$, from the set $\{\dot{x}_l^i\}_{i=1}^N$ according to the importance weights $w_l^i(r_l)$, i.e., according to

$$P\left(x_l^j = \dot{x}_l^i\right) = w_l^i(r_l), \quad i = 1, \cdots, N,$$

where $P\left(x_l^j = \dot{x}_l^i\right)$ denotes the probability of $x_l^j = \dot{x}_l^i$.

5. Increment $l \mapsto l + 1$ and return to step 2.

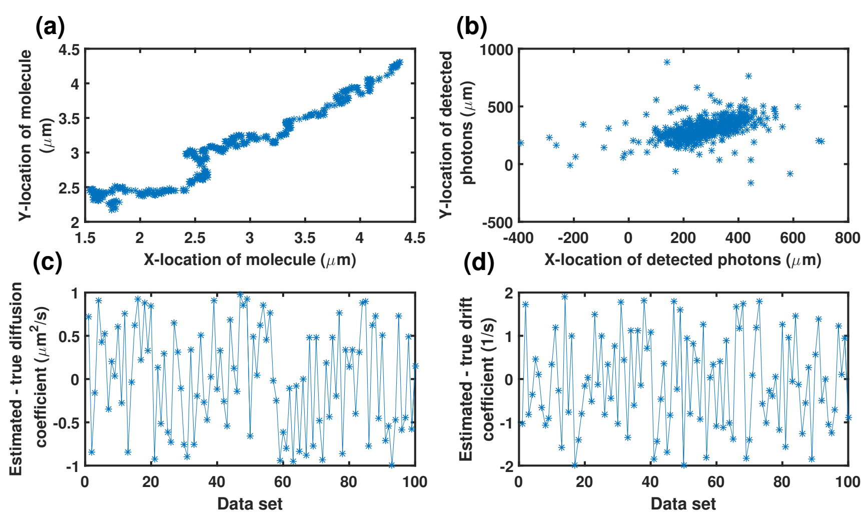


FIG. SM2. *Analysis of the error of diffusion coefficient and first order drift coefficient estimates produced by the maximum likelihood estimation method for the Airy measurement model. (a) The two-dimensional single molecule trajectory simulated in Fig. 3(a). (b) Detected locations of the photons emitted from the molecule trajectory of part (a) in the image space which are simulated using Eq. (3.6) with the Airy profile (Eq. (2.4)) and* $\alpha := \frac{2\pi n_a}{\lambda} = 2.59$. *(c) Differences between the diffusion coefficient estimates and the true diffusion coefficient value for 100 data sets, each containing a trajectory of a molecule simulated using Eqs. (SM4.1) and (3.6) with the Airy profile, and the parameters given in parts (a) and (b). (d) Differences between the first order drift coefficient estimates and its true value for the data sets of part (c).*

**SM7. Estimation results for Airy measurements.** Here, we analyze the error of the diffusion and first order drift coefficient estimates for simulated data sets with the Airy measurement profile, with the same standard deviation as the Born and Wolf and Gaussian data presented in Figs. 3-8, and obtain similar results (see Figs.

SM2 and SM3). We also show the differences between the means of the distributions of the prediction of the molecule locations and the true locations of the molecule in Fig. SM5 (see Section SM14.6).

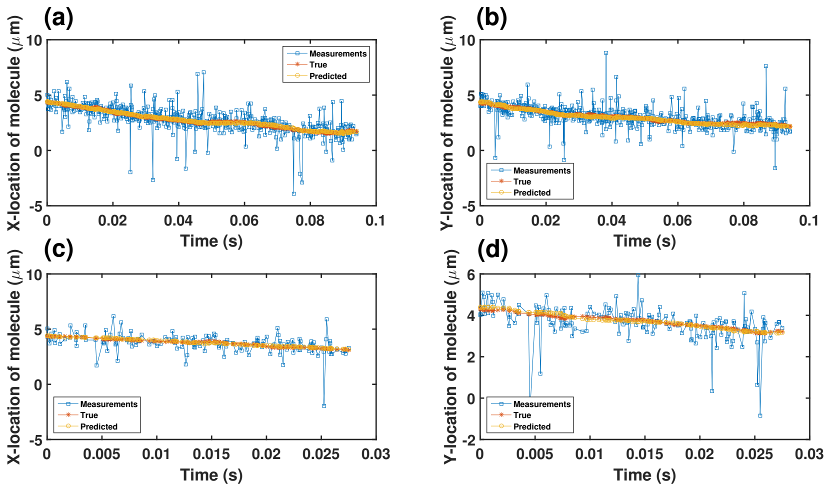


Fig. SM3. *Predicted locations of the molecule for the Airy measurement model. (a) and (b) Means of the distributions of the prediction of the molecule x- and y-locations and the true x- and y-locations of the molecule for the same data set as in Figs. SM2(a) and SM2(b). The measurements transformed from the image space to the object space are also shown. (c) and (d) Means of the distributions of the prediction of the molecule x- and y-locations and the true x- and y-locations of the molecule over the time interval* $[0, 27.5]$ *ms.*

**SM8. Proof of Theorem 5.2.** Let $\mathcal{G}_{[t]}\Big(\left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \mathcal{C}, \Theta\Big)$ and $\mathcal{G}_L\Big(\left(\mathcal{U}_L, \mathcal{T}_L\right), \mathcal{C}, \Theta\Big)$ be image detection processes for a time interval $[t_0, t]$ and for a fixed number $L$ of photons, respectively. Let $\mathcal{D}_{[t]} := \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \mathcal{D}_k := \left(\mathcal{U}_k, \mathcal{T}_k\right), k = 0, 1, \cdots$. Assume that the conditional probability density functions $p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}, l = 1, 2, \cdots$, of $U_l$, given $T_l$ and $\mathcal{D}_{l-1}$, satisfy the following regularity conditions, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$,

(a) $\dfrac{\partial p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big)}{\partial \theta_i}$ exists for $i = 1, \cdots, n$,

(b) $\displaystyle\int_{\mathcal{C}} \left| \frac{\partial p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r|\tau_l, d_{l-1}\Big)}{\partial \theta_i} \right| dr < \infty$ for $i = 1, \cdots, n$,

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ for $\mathcal{G}_{[t]}$, $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$ for $\mathcal{G}_L$, and $p^{\theta}\Big(r_1|\tau_1, d_0\Big) := p^{\theta}\Big(r_1|\tau_1\Big)$.

1.1. Then, the Fisher information matrix $I_{[t]}(\theta)$ of $\mathcal{G}_{[t]}$ is given by

$$I_{[t]}(\theta) = E\left[\left(\frac{\partial \log \mathcal{L}(\theta|d_K)}{\partial \theta}\right)^T \left(\frac{\partial \log \mathcal{L}(\theta|d_K)}{\partial \theta}\right)\right], \tag{SM8.1}$$

where $d_K \in \mathcal{C}^K \times \mathbb{R}^K_{[t]}$, $K = 1, 2, \cdots$, and $\mathcal{L}$ denotes the likelihood function. By substituting the expression of the likelihood function $\mathcal{L}_{[t]}$ of $\mathcal{G}_{[t]}$ (Eq. (4.1)) into Eq.

(SM8.1), according to [SM6, SM7], we have

$$
\begin{aligned}
I_{[t]}(\theta) &= P_\theta\Big(N(t)=0\Big)\left(\frac{\partial \log P_\theta\Big(N(t)=0\Big)}{\partial\theta}\right)^T\left(\frac{\partial \log P_\theta\Big(N(t)=0\Big)}{\partial\theta}\right)\\
&\quad+\sum_{K=1}^{\infty}\int_{t_0}^{t}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2}\int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p_{[t]}^{\theta}\Big(d_K,K\Big)\left(\frac{\partial \log p_{[t]}^{\theta}\Big(d_K,K\Big)}{\partial\theta}\right)^T\left(\frac{\partial \log p_{[t]}^{\theta}\Big(d_K,K\Big)}{\partial\theta}\right)\\
&\qquad\times dr_1\cdots dr_K d\tau_1 d\tau_2\cdots d\tau_K\\
&= \frac{1}{P_\theta\Big(N(t)=0\Big)}\left(\frac{\partial P_\theta\Big(N(t)=0\Big)}{\partial\theta}\right)^T\left(\frac{\partial P_\theta\Big(N(t)=0\Big)}{\partial\theta}\right)\\
&\quad+\sum_{K=1}^{\infty}\int_{t_0}^{t}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2}\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}\frac{1}{p_{[t]}^{\theta}\Big(d_K,K\Big)}\left(\frac{\partial p_{[t]}^{\theta}\Big(d_K,K\Big)}{\partial\theta}\right)^T\left(\frac{\partial p_{[t]}^{\theta}\Big(d_K,K\Big)}{\partial\theta}\right)\\
&\qquad\times dr_1\cdots dr_K d\tau_1 d\tau_2\cdots d\tau_K,
\end{aligned}
\tag{SM8.2}
$$

where $P_\theta\Big(N(t)=0\Big)$ is the probability of $N(t)=0$ and $p_{[t]}^{\theta}$ denotes the probability density function of $\mathcal{D}_{[t]}$ and $N(t)$.

1.2. Assume that the photon detection rate $\Lambda$ is independent of $\theta$. By substituting Eqs. (SM1.1)-(SM1.3) into Eq. (SM8.2), we have

$$
\begin{aligned}
I_{[t]}(\theta) &= \sum_{K=1}^{\infty} P\Big(N(t)=K\Big)\int_{t_0}^{t}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2}\Bigg[\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}\frac{1}{p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_1,\cdots,r_K|\tau_1,\cdots,\tau_K\Big)}\\
&\quad\times\left(\frac{\partial p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_1,\cdots,r_K|\tau_1,\cdots,\tau_K\Big)}{\partial\theta}\right)^T\left(\frac{\partial p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_1,\cdots,r_K|\tau_1,\cdots,\tau_K\Big)}{\partial\theta}\right)dr_1\cdots dr_K\Bigg]\\
&\quad\times p_{\mathcal{T}_K|N(t)}\Big(\tau_1,\cdots,\tau_K|K\Big)d\tau_1 d\tau_2\cdots d\tau_K\\
&= e^{-\int_{t_0}^{t}\Lambda(\tau)d\tau}\sum_{K=1}^{\infty}\left\{\int_{t_0}^{t}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2} I_{\tau_1,\cdots,\tau_K}(\theta)\prod_{k=1}^{K}\Lambda(\tau_k)d\tau_1 d\tau_2\cdots d\tau_K\right\},
\end{aligned}
\tag{SM8.3}
$$

where, for $t_0\le\tau_1<\cdots<\tau_K\le t$, $I_{\tau_1,\cdots,\tau_K}$ is given by

$$
\begin{aligned}
&I_{\tau_1,\cdots,\tau_K}(\theta)\\
&= E_{\mathcal{U}_K|\mathcal{T}_K=\tau_{1:K}}\left\{\left(\frac{\partial \log p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)}{\partial\theta}\right)^T\left(\frac{\partial \log p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)}{\partial\theta}\right)\right\}\\
&= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)\left(\frac{\partial \log p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)}{\partial\theta}\right)^T\left(\frac{\partial \log p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)}{\partial\theta}\right)\\
&\qquad\times dr_K\cdots dr_1\\
&= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}}\frac{1}{p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)}\left(\frac{\partial p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)}{\partial\theta}\right)^T\left(\frac{\partial p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)}{\partial\theta}\right)\\
&\qquad\times dr_K\cdots dr_1,
\end{aligned}
\tag{SM8.4}
$$

where $r_{1:K}:=(r_1,\cdots,r_K)\,,\tau_{1:K}:=(\tau_1,\cdots,\tau_K)\,,K=1,2,\cdots$. Since $p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_{1:K}|\tau_{1:K}\Big)=\prod_{l=1}^{K}p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\Big(r_l|\tau_l,d_{l-1}\Big)$, according to Lemma 1 (chain rule for the Fisher information matrix) of [SM8], we have

$$I_{\tau_1,\cdots,\tau_K}(\theta) = \begin{cases} \sum_{l=1}^{K} I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta), & t_0 \le \tau_1 < \cdots < \tau_K \le t, \\ 0, & \text{otherwise}, \end{cases}$$

where the Fisher information matrix $I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}$ of $U_l, l = 1, \cdots, K$, calculated with respect to the conditional probability density function $p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}$ at fixed time points $\mathcal{T}_l = \tau_{1:l}$, is given by

$$\begin{aligned} I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta) &= E_{\mathcal{U}_l|\mathcal{T}_l=\tau_{1:l}}\left\{\left(\frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right)^T\left(\frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right)\right\} \\ &= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p^{\theta}_{\mathcal{U}_l|\mathcal{T}_l}\left(r_{1:l}|\tau_{1:l}\right)\left(\frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right)^T \\ &\quad\times\left(\frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right) dr_l\cdots dr_1 \\ &= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right)\Bigg[\int_{\mathcal{C}}\frac{1}{p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)} \\ &\quad\times\left(\frac{\partial p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right)^T\left(\frac{\partial p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right) dr_l\Bigg] dr_{l-1}\cdots dr_1. \end{aligned} \tag{SM8.5}$$

2.1. Moreover, by substituting the expression for the likelihood function $\mathcal{L}_L$ of $\mathcal{G}_L$ (Eq. (4.2)) into Eq. (SM8.1), the Fisher information matrix $I_L(\theta)$ of $\mathcal{G}_L$ can be obtained as

$$\begin{aligned} I_L(\theta) &= \int_{t_0}^{\infty}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2}\int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p^{\theta}_L\left(d_L\right)\left(\frac{\partial \log p^{\theta}_L\left(d_L\right)}{\partial\theta}\right)^T\left(\frac{\partial \log p^{\theta}_L\left(d_L\right)}{\partial\theta}\right) dr_1\cdots dr_L \\ &\quad\times d\tau_1 d\tau_2\cdots d\tau_L \\ &= \int_{t_0}^{\infty}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2}\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}\frac{1}{p^{\theta}_L\left(d_L\right)}\left(\frac{\partial p^{\theta}_L\left(d_L\right)}{\partial\theta}\right)^T\left(\frac{\partial p^{\theta}_L\left(d_L\right)}{\partial\theta}\right) dr_1\cdots dr_L d\tau_1 d\tau_2\cdots d\tau_L, \end{aligned} \tag{SM8.6}$$

where $d_L \in \mathcal{C}^L \times \mathbb{R}^L_{[\infty]}$, and $p^{\theta}_L$ denotes the probability density function of $\mathcal{D}_L$.

2.2. Assume that the photon detection rate $\Lambda$ is independent of $\theta$. By substituting Eq. (2.2) into Eq. (SM8.6), we have, according to Eq. (SM1.6) and using the similar procedure used in the previous part,

$$\begin{aligned} I_L(\theta) &= \int_{t_0}^{\infty}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2}\Bigg[\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}\frac{1}{p^{\theta}_{\mathcal{U}_L|\mathcal{T}_L}\left(r_{1:L}|\tau_{1:L}\right)}\left(\frac{\partial p^{\theta}_{\mathcal{U}_L|\mathcal{T}_L}\left(r_{1:L}|\tau_{1:L}\right)}{\partial\theta}\right)^T \\ &\quad\times\left(\frac{\partial p^{\theta}_{\mathcal{U}_L|\mathcal{T}_L}\left(r_{1:L}|\tau_{1:L}\right)}{\partial\theta}\right) dr_1\cdots dr_L\Bigg] p_{\mathcal{T}_L}\left(\tau_{1:L}\right) d\tau_1 d\tau_2\cdots d\tau_L \\ &= \int_{t_0}^{\infty}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2} I_{\tau_1,\cdots,\tau_L}(\theta) e^{-\int_{t_0}^{\tau_L}\Lambda(\tau)d\tau}\prod_{k=1}^{L}\Lambda(\tau_k) d\tau_1 d\tau_2\cdots d\tau_L. \end{aligned}$$

**SM9. Proof of Corollary 5.4.** Let $\mathcal{G}_{[t]}\left(X, \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), q, \mathcal{C}, \Theta\right)$ (or $\mathcal{G}_L\left(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta\right)$) be an image detection process driven by the stochastic trajectory $X$ and image function $q$ for a time interval $[t_0, t]$ (or a fixed number $L$ of photons). Assume that the photon detection rate $\Lambda$ is independent of $\theta$. The Fisher information matrix $I_{\tau_1,\cdots,\tau_K}$ in Eq. (5.2) (or Eq. (5.6)) of Theorem 5.2 is given by

$$I_{\tau_1,\cdots,\tau_K}(\theta)=\begin{cases}\sum_{l=1}^{K} I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta), & t_0\le\tau_1<\cdots<\tau_K\le t,\\ 0, & \text{otherwise,}\end{cases}$$

where, for $r_{1:l}:=(r_1,\cdots,r_l)\,,\tau_{1:l}:=(\tau_1,\cdots,\tau_l)\,,l=1,\cdots,K$,

$$\begin{aligned}I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta)&=\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right)\left[\int_{\mathcal{C}}\frac{1}{p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}\right.\\ &\left.\times\left(\frac{\partial p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}{\partial\theta}\right)^T\left(\frac{\partial p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}{\partial\theta}\right)dr_l\right]dr_{l-1}\cdots dr_1,\end{aligned}\tag{SM9.1}$$

and $I^{\tau_1}_{U_1|T_1}$ is given by Eq. (5.5). According to Eq. (SM2.1), we can express the conditional probability density functions $p^{\theta}_{U_l|\mathcal{D}_{l-1},T_l}$ in terms of the image profile $f_x, x\in\mathbb{R}^3$, as

$$p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)=\int_{\mathbb{R}^3}f^{\theta}_{x_o}\left(r_l\right)p^{\theta}_{pr_l}\left(x_o|\tau_l,d_{l-1}\right)dx_o,\tag{SM9.2}$$

where $p^{\theta}_{pr_l}:=p^{\theta}_{X(T_l)|T_l,\mathcal{D}_{l-1}}$ denotes the distribution of the prediction of the object location, $p^{\theta}_{pr_1}\left(x_o|\tau_1,d_0\right):=p^{\theta}_{pr_1}\left(x_o|\tau_1\right)$, and $x_o\in\mathbb{R}^3$ denotes a running variable in the object space. By substituting Eq. (SM9.2) into Eq. (SM9.1), we have, for $dp^{\theta}_{pr_l}:=\frac{\partial p^{\theta}_{pr_l}}{\partial\theta}$ and $df^{\theta}_x:=\frac{\partial f^{\theta}_x}{\partial\theta}$,

$$\begin{aligned}&I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta)\\ &=\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right)\left\{\int_{\mathcal{C}}\frac{1}{p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}\right.\\ &\quad\times\left(\int_{\mathbb{R}^3}\left[\left(df^{\theta}_{x_1}\left(r_l\right)\right)^T p^{\theta}_{pr_l}\left(x_1|\tau_l,d_{l-1}\right)+f^{\theta}_{x_1}\left(r_l\right)\left(dp^{\theta}_{pr_l}\left(x_1|\tau_l,d_{l-1}\right)\right)^T\right]dx_1\right)\\ &\quad\left.\times\left(\int_{\mathbb{R}^3}\left[df^{\theta}_{x_2}\left(r_l\right)p^{\theta}_{pr_l}\left(x_2|\tau_l,d_{l-1}\right)+f^{\theta}_{x_2}\left(r_l\right)dp^{\theta}_{pr_l}\left(x_2|\tau_l,d_{l-1}\right)\right]dx_2\right)dr_l\right\}dr_{l-1}\cdots dr_1\\ &=\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right)\left\{\int_{\mathcal{C}}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\frac{1}{p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}\right.\\ &\quad\times\left[\left(df^{\theta}_{x_1}\left(r_l\right)\right)^T\quad\left(dp^{\theta}_{pr_l}\left(x_1|\tau_l,d_{l-1}\right)\right)^T\right]\begin{bmatrix}p^{\theta}_{pr_l}\left(x_1|\tau_l,d_{l-1}\right)\\ f^{\theta}_{x_1}\left(r_l\right)\end{bmatrix}\begin{bmatrix}p^{\theta}_{pr_l}\left(x_2|\tau_l,d_{l-1}\right)\\ f^{\theta}_{x_2}\left(r_l\right)\end{bmatrix}^T\begin{bmatrix}df^{\theta}_{x_2}\left(r_l\right)\\ dp^{\theta}_{pr_l}\left(x_2|\tau_l,d_{l-1}\right)\end{bmatrix}\\ &\quad\left.\times\,dx_1dx_2dr_l\right\}dr_{l-1}\cdots dr_1\\ &=\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right)\left[\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\left(\int_{\mathcal{C}}\frac{F^{\theta}_l\left(x_1,d_l\right)\left[F^{\theta}_l\left(x_2,d_l\right)\right]^T}{p^{\theta}_{U_l|T_l,D_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}dr_l\right)dx_1dx_2\right]\\ &\quad\times dr_{l-1}\cdots dr_1,\end{aligned}\tag{SM9.3}$$

where for $l = 1, 2, \cdots$,

$$F_l^\theta\Big(x, d_l\Big) := \Big[\Big(df_x^\theta\,(r_l)\Big)^T \quad \Big(dp_{pr_l}^\theta\Big(x|\tau_l, d_{l-1}\Big)\Big)^T\Big] \begin{bmatrix} p_{pr_l}^\theta\Big(x|\tau_l, d_{l-1}\Big) \\ f_x^\theta\,(r_l) \end{bmatrix}, \quad x \in \mathbb{R}^3,$$

$$p_{U_l|T_l,\mathcal{D}_{l-1}}^\theta\Big(r_l|\tau_l, d_{l-1}\Big) = \int_{\mathbb{R}^3} f_{x_o}^\theta\,(r_l)\, p_{pr_l}^\theta\Big(x_o|\tau_l, d_{l-1}\Big) dx_o,$$

$$p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^\theta\Big(r_{1:l-1}|\tau_{1:l-1}\Big) = \prod_{i=1}^{l-1} \int_{\mathbb{R}^3} f_{x_o}^\theta\,(r_i)\, p_{pr_i}^\theta\Big(x_o|\tau_i, d_{i-1}\Big) dx_o,$$

with $I_{U_1|T_1}^{\tau_1}$ given by

$$I_{U_1|T_1}^{\tau_1}(\theta) = \int_{\mathcal{C}} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{1}{p_{U_1|T_1}^\theta\Big(r|\tau_1\Big)} \Big[\Big(df_{x_1}^\theta\,(r)\Big)^T \quad \Big(dp_{pr_1}^\theta\Big(x_1|\tau_1\Big)\Big)^T\Big] \begin{bmatrix} p_{pr_1}^\theta\Big(x_1|\tau_1\Big) \\ f_{x_1}^\theta\,(r) \end{bmatrix}$$
$$\times \begin{bmatrix} p_{pr_1}^\theta\Big(x_2|\tau_1\Big) \\ f_{x_2}^\theta\,(r) \end{bmatrix}^T \begin{bmatrix} df_{x_2}^\theta\,(r) \\ dp_{pr_1}^\theta\Big(x_2|\tau_1\Big) \end{bmatrix} dx_1 dx_2 dr.$$

**SM10. Proof of Corollary 5.6.** For $t_0 \le \tau_1 < \cdots < \tau_K$, let $\mathcal{G}_{\tau_1,\cdots,\tau_K}\Big(\left(\mathcal{U}_K, \mathcal{T}_K\right), \mathcal{C}, \Theta\Big)$ be an image detection process at fixed time points $\tau_1, \cdots, \tau_K$. Assume that

$$p_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) = p_{U_l|T_l}\Big(r_l|\tau_l\Big), \quad d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}, \quad l = 1, 2, \cdots.$$

1. According to Eq. (2.8), we have, for $r_l \in \mathbb{R}^2, l = 1, 2, \cdots$,

$$p_{\mathcal{U}_K|\mathcal{T}_K}\Big(r_1, \cdots, r_K|\tau_1, \cdots, \tau_K\Big) = \prod_{l=1}^{K} p_{U_l|T_l}^\theta\Big(r_l|\tau_l\Big). \tag{SM10.1}$$

By substituting Eq. (SM10.1) into Eq. (5.4), we have

$$I_{\tau_1,\cdots,\tau_K}(\theta) = \begin{cases} \sum_{l=1}^K I_{U_l|T_l}^{\tau_l}(\theta), & t_0 \le \tau_1 < \cdots < \tau_K, \\ 0, & \text{otherwise}, \end{cases}$$

where for $l = 1, \cdots, K$,

$$I_{U_l|T_l}^{\tau_l}(\theta) = \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} p_{U_1|T_1}^\theta\Big(r_1|\tau_1\Big) dr_1\right) \cdots \left(\int_{\mathbb{R}^2} p_{U_{l-1}|T_{l-1}}^\theta\Big(r_{l-1}|\tau_{l-1}\Big) dr_{l-1}\right)$$
$$\times \frac{1}{p_{U_l|T_l}^\theta\Big(r_l|\tau_l\Big)} \left(\frac{\partial p_{U_l|T_l}^\theta\Big(r_l|\tau_l\Big)}{\partial\theta}\right)^T \left(\frac{\partial p_{U_l|T_l}^\theta\Big(r_l|\tau_l\Big)}{\partial\theta}\right) dr_l$$
$$= \int_{\mathbb{R}^2} \frac{1}{p_{U_l|T_l}^\theta\Big(r|\tau_l\Big)} \left(\frac{\partial p_{U_l|T_l}^\theta\Big(r|\tau_l\Big)}{\partial\theta}\right)^T \left(\frac{\partial p_{U_l|T_l}^\theta\Big(r|\tau_l\Big)}{\partial\theta}\right) dr. \tag{SM10.2}$$

2.1. For an object with deterministic trajectory $X_\tau(\theta) := (x_\tau(\theta), y_\tau(\theta)) \in \mathbb{R}^2, \tau \ge t_0$, assume that there exists an image function $q$: $\mathbb{R}^2 \mapsto \mathbb{R}$, which is assumed to be independent of the parameter vector $\theta$, such that for $r = (x, y) \in \mathbb{R}^2, t_0 \le \tau \le t$, and a magnification factor $M > 1$,

$$p_{U_l|T_l}^\theta\Big(r|\tau\Big) = f_{X_\tau(\theta)}\,(r) = \frac{1}{M^2} q\bigg(\frac{x}{M} - x_\tau(\theta), \frac{y}{M} - y_\tau(\theta)\bigg). \tag{SM10.3}$$

Then, by substituting Eq. (SM10.3) into Eq. (SM10.2), $I_{\tau_l} := I^{\tau_l}_{U_l|T_l}$ is obtained as, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$,

$$
\begin{aligned}
I_{\tau_l}(\theta) &= \frac{1}{M^2}\int_{\mathbb{R}^2} \frac{1}{q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)} \left(\frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)}{\partial\theta}\right)^T \\
&\qquad \times \left(\frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)}{\partial\theta}\right) dxdy \\
&= \frac{1}{M^2}\int_{\mathbb{R}^2} \frac{1}{q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)} \begin{bmatrix} \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)}{\partial\theta_1} \\ \vdots \\ \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)}{\partial\theta_n} \end{bmatrix} \begin{bmatrix} \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)}{\partial\theta_1} \\ \vdots \\ \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right)}{\partial\theta_n} \end{bmatrix}^T dxdy.
\end{aligned}
\tag{SM10.4}
$$

For each $(x, y) \in \mathbb{R}^2$, let $h_{x,y} = (h_x, h_y)\colon \mathbb{R}^2 \mapsto \mathbb{R}^2$, such that, for $\theta \in \Theta$, $(x_{\tau_l}(\theta), y_{\tau_l}(\theta)) \in \mathbb{R}^2$,

$$h_x(x_{\tau_l}(\theta), y_{\tau_l}(\theta)) = \frac{x}{M} - x_{\tau_l}(\theta), \quad h_y(x_{\tau_l}(\theta), y_{\tau_l}(\theta)) = \frac{y}{M} - y_{\tau_l}(\theta).$$

Then, for $d_{\tau_l} = (x_{\tau_l}, y_{\tau_l})\colon \Theta \mapsto \mathbb{R}^2$, the composite function $(q \circ h_{x,y} \circ d_{\tau_l})(\theta)$ is given by

$$(q \circ h_{x,y} \circ d_{\tau_l})(\theta) = q(h_{x,y}(d_{\tau_l}(\theta))) = q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right),$$

and therefore, using the formal definition of partial derivatives, we can rewrite Eq. (SM10.4) as

$$
\begin{aligned}
I_{\tau_l}(\theta) &= \frac{1}{M^2}\int_{\mathbb{R}^2} \frac{1}{(q \circ h_{x,y} \circ d_{\tau_l})(\theta)} \begin{bmatrix} (D_1(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \\ \vdots \\ (D_n(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \end{bmatrix} \begin{bmatrix} (D_1(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \\ \vdots \\ (D_n(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \end{bmatrix}^T \\
&\qquad \times dxdy.
\end{aligned}
\tag{SM10.5}
$$

Assume that $d_{\tau_l}$ is continuously differentiable on all of $\Theta$, and $h_{x,y}$ is differentiable at $d_{\tau_l}(\theta)$. Also, suppose that $q$ is differentiable at $h_{x,y}\left(d_{\tau_l}(\theta)\right)$ . Then, according to Theorem SM14.3 (see Section SM14.3), for $i =, 1 \cdots, n$,

$$
\begin{aligned}
(D_i(q \circ h_{x,y} \circ d_{\tau_l}))(\theta) &= (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_1 h_x)(d_{\tau_l}(\theta))(D_i x_{\tau_l})(\theta) \\
&\quad + (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 h_x)(d_{\tau_l}(\theta))(D_i y_{\tau_l})(\theta) \\
&\quad + (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_1 h_y)(d_{\tau_l}(\theta))(D_i x_{\tau_l})(\theta) \\
&\quad + (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 h_y)(d_{\tau_l}(\theta))(D_i y_{\tau_l})(\theta) \\
&= -(D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_i x_{\tau_l})(\theta) - (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_i y_{\tau_l})(\theta) \\
&= -\begin{bmatrix}(D_i x_{\tau_l})(\theta) & (D_i y_{\tau_l})(\theta)\end{bmatrix} \begin{bmatrix}(D_1 q)(h_{x,y}(d_{\tau_l}(\theta))) \\ (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))\end{bmatrix}.
\end{aligned}
\tag{SM10.6}
$$

By substituting Eq. (SM10.6) into Eq. (SM10.5), we have, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$,

$$
\begin{aligned}
I_{\tau_l}(\theta) &= \frac{1}{M^2} V_\theta^T(\tau_l) \left[\int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} \begin{bmatrix}(D_1 q)(h_{x,y}(d_{\tau_l}(\theta))) \\ (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))\end{bmatrix} \begin{bmatrix}(D_1 q)(h_{x,y}(d_{\tau_l}(\theta))) \\ (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))\end{bmatrix}^T dxdy\right] V_\theta(\tau_l) \\
&= \frac{1}{M^2} V_\theta^T(\tau_l) \begin{bmatrix} \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} \left[(D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))\right]^2 dxdy & \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 q)(h_{x,y}(d_{\tau_l}(\theta))) dxdy \\ \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 q)(h_{x,y}(d_{\tau_l}(\theta))) dxdy & \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} \left[(D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))\right]^2 dxdy \end{bmatrix} V_\theta(\tau_l),
\end{aligned}
\tag{SM10.7}
$$

where

$$V_\theta(\tau_l) := \begin{bmatrix} (D_1 x_{\tau_l})(\theta) & \cdots & (D_n x_{\tau_l})(\theta) \\ (D_1 y_{\tau_l})(\theta) & \cdots & (D_n y_{\tau_l})(\theta) \end{bmatrix} \in \mathbb{R}^{2\times n}.$$

Let $w_1 \colon \mathbb{R}^2 \mapsto \mathbb{R}$, such that

$$w_1(u,v) = \frac{1}{q(u,v)} \left[(D_1 q)(u,v)\right]^2, \quad (u,v) \in \mathbb{R}^2,$$

be an integrable function. Also, for each $\theta = (\theta_1, \cdots, \theta_n) \in \Theta, (x_{\tau_l}(\theta), y_{\tau_l}(\theta)) \in \mathbb{R}^2$, let $g_{\theta,\tau_l} = (g^1_{\theta,\tau_l}, g^2_{\theta,\tau_l}) \colon \mathbb{R}^2 \mapsto \mathbb{R}^2$, such that

$$g_{\theta,\tau_l}(x,y) = (g^1_{\theta,\tau_l}(x,y), g^2_{\theta,\tau_l}(x,y)) = \left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right) = (u,v).$$

Then, we have for the Jacobian $J(g_{\theta,\tau_l})$ of $g_{\theta,\tau_l}$,

$$J(g_{\theta,\tau_l}) = \begin{bmatrix} \frac{\partial g^1_{\theta,\tau_l}(x,y)}{\partial x} & \frac{\partial g^1_{\theta,\tau_l}(x,y)}{\partial y} \\ \frac{\partial g^2_{\theta,\tau_l}(x,y)}{\partial x} & \frac{\partial g^2_{\theta,\tau_l}(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{1}{M} & 0 \\ 0 & \frac{1}{M} \end{bmatrix},$$

and the modulus of its determinant is given by

$$\left|\det \begin{pmatrix} \frac{1}{M} & 0 \\ 0 & \frac{1}{M} \end{pmatrix}\right| = \left|\frac{1}{M^2}\right| = \frac{1}{M^2}.$$

Then, according to Theorem SM14.4 (see Section SM14.4),

$$\begin{aligned} \int_{\mathbb{R}^2} w_1(u,v)dudv &= \frac{1}{M^2}\int_{\mathbb{R}^2} w_1\left(g_{\theta,\tau_l}(x,y)\right) dxdy \\ &= \frac{1}{M^2}\int_{\mathbb{R}^2} w_1\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right) dxdy. \end{aligned} \tag{SM10.8}$$

Also, let $w_2, w_3 \colon \mathbb{R}^2 \mapsto \mathbb{R}$, such that

$$w_2(u,v) = \frac{1}{q(u,v)}(D_1 q)(u,v)(D_2 q)(u,v), \quad (u,v) \in \mathbb{R}^2,$$

and

$$w_3(u,v) = \frac{1}{q(u,v)} \left[(D_2 q)(u,v)\right]^2, \quad (u,v) \in \mathbb{R}^2,$$

be integrable functions. Similarly, according to Theorem SM14.4 (see Section SM14.4),

$$\int_{\mathbb{R}^2} w_i(u,v)dudv = \frac{1}{M^2}\int_{\mathbb{R}^2} w_i\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta)\right) dxdy, \quad i = 2,3. \tag{SM10.9}$$

Then, by substituting Eqs. (SM10.8) and (SM10.9) into Eq. (SM10.7),

$$\begin{aligned} I_{\tau_l}(\theta) &= V^T_\theta(\tau_l) \begin{bmatrix} \int_{\mathbb{R}^2} \frac{1}{q(u,v)} \left[(D_1 q)(u,v)\right]^2 dudv & \int_{\mathbb{R}^2} \frac{1}{q(u,v)} (D_1 q)(u,v)(D_2 q)(u,v)dudv \\ \int_{\mathbb{R}^2} \frac{1}{q(u,v)} (D_1 q)(u,v)(D_2 q)(u,v)dudv & \int_{\mathbb{R}^2} \frac{1}{q(u,v)} \left[(D_2 q)(u,v)\right]^2 dudv \end{bmatrix} V_\theta(\tau_l) \\ &= V^T_\theta(\tau_l) \left[\int_{\mathbb{R}^2} \frac{1}{q(u,v)} \begin{bmatrix} (D_1 q)(u,v) \\ (D_2 q)(u,v) \end{bmatrix} \begin{bmatrix} (D_1 q)(u,v) \\ (D_2 q)(u,v) \end{bmatrix}^T dudv\right] V_\theta(\tau_l). \end{aligned}$$

2.2. The results follow by using the similar procedure used in the previous part.

**SM11. Proof of Corollary 5.7.** According to Theorem 5.2, the Fisher information matrix $I_{\tau_1,\cdots,\tau_K}$ in Eq. (5.2) (or Eq. (5.6)) can be calculated as

$$I_{\tau_1,\cdots,\tau_K}(\theta) = \begin{cases} \sum_{l=1}^{K} I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta), & t_0 \leq \tau_1 < \cdots < \tau_K \leq t, \\ 0, & \text{otherwise}, \end{cases}$$

where

$$I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta) = \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right) \Bigg[ \int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) \times \left( \frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta} \right)^T \left( \frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta} \right) dr_l \Bigg] dr_{l-1} \cdots dr_1,$$

with $d_l \in \mathcal{C}^l \in \mathbb{R}^l_{[t]}$ (or $d_l \in \mathcal{C}^l \in \mathbb{R}^l_{[\infty]}$), and $r_{1:l} := (r_1, \cdots, r_l), \tau_{1:l} := (\tau_1, \cdots, \tau_l), l = 1, 2, \cdots$. Under the certain regularity conditions, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta, i, j = 1, \cdots, n$, the $i, j^{th}$ entry $\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}\right]_{i,j}$ of $I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}$ can be calculated as

$$\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta)\right]_{i,j} = \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right) \times \left[ -\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) \frac{\partial^2 \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta_i \theta_j} dr_l \right] dr_{l-1} \cdots dr_1. \tag{SM11.1}$$

According to Eq. (SM3.6),

$$p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) = \frac{1}{2\pi \left[\det\left(R_{\theta,l}\right)\right]^{1/2}} \exp\left( -\frac{1}{2} e^T_{\theta,l} R^{-1}_{\theta,l} e_{\theta,l} \right), \tag{SM11.2}$$

where $e_{\theta,l} := r_l - C\hat{x}^{l-1}_{\theta,l}, R_l := CP^{l-1}_{\theta,l}C^T + \Sigma_g$, and for $l = 0, 1, \cdots$,

$$\begin{aligned} \hat{x}^l_{\theta,l+1} &= \tilde{\phi}_\theta(\tau_l, \tau_{l+1})\hat{x}^l_{\theta,l} + \tilde{a}_\theta(\tau_l, \tau_{l+1}), \\ P^l_{\theta,l+1} &= \tilde{\phi}_\theta(\tau_l, \tau_{l+1})P^l_{\theta,l}\tilde{\phi}^T_\theta(\tau_l, \tau_{l+1}) + \tilde{Q}_\theta(\tau_l, \tau_{l+1}), \end{aligned} \tag{SM11.3}$$

and for $l = 1, 2, \cdots$,

$$\begin{aligned} \hat{x}^l_{\theta,l} &= \hat{x}^{l-1}_{\theta,l} + K_{\theta,l}(r_l - C\hat{x}^{l-1}_{\theta,l}), \\ P^l_{\theta,l} &= P^{l-1}_{\theta,l} - K_{\theta,l}CP^{l-1}_{\theta,l}, \\ K_{\theta,l} &= P^{l-1}_{\theta,l}C^T\left(CP^{l-1}_{\theta,l}C^T + \Sigma_g\right)^{-1}, \end{aligned} \tag{SM11.4}$$

where $\hat{x}^0_{\theta,0} := \bar{x}_{\theta,0}, P^0_{\theta,0} := \tilde{P}_{\theta,0}$. In order to calculate $\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}\right]_{i,j}, i, j = 1, \cdots, n$, in Eq. (SM11.1), we first calculate, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$ and $i = 1, \cdots, n$, the derivative of $\log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}$ with respect to $\theta_i$ as below

$$\begin{aligned} \frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta_i} = &-\frac{1}{2}\text{trace}\left(R_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}\right) \\ &-\frac{1}{2}\left( \frac{\partial e^T_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} e_{\theta,l} - e^T_{\theta,l} R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} e_{\theta,l} + e^T_{\theta,l} R^{-1}_{\theta,l} \frac{\partial e_{\theta,l}}{\partial \theta_i} \right). \end{aligned} \tag{SM11.5}$$

Since the covariance matrix $R_{\theta,l}$ is symmetric, then, $\frac{\partial e^T_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} e_{\theta,l} = e^T_{\theta,l} R^{-1}_{\theta,l} \frac{\partial e_{\theta,l}}{\partial \theta_i}$, and therefore, according to Eq. (SM11.5),(note that $\text{trace}\left(e^T_{\theta,l} R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} e_{\theta,l}\right) =$

$e^T_{\theta,l} R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} e_{\theta,l}),$

$$\begin{aligned}
\frac{\partial \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta_i} &= -\frac{1}{2}\text{trace}\left(R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}\right) + \frac{1}{2}\text{trace}\left(e^T_{\theta,l}R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}e_{\theta,l}\right)\\
&\quad -\frac{1}{2}\left(\frac{\partial e^T_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}e_{\theta,l} + e^T_{\theta,l}R^{-1}_{\theta,l}\frac{\partial e_{\theta,l}}{\partial \theta_i}\right)\\
&= -\frac{1}{2}\text{trace}\left(R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i} - R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}e_{\theta,l}e^T_{\theta,l}\right) - \frac{\partial e^T_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}e_{\theta,l}\\
&= -\frac{1}{2}\text{trace}\left[\left(R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}\right)(I - R^{-1}_{\theta,l}e_{\theta,l}e^T_{\theta,l})\right] - \frac{\partial e^T_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}e_{\theta,l},
\end{aligned} \tag{SM11.6}$$

where $I$ denotes the identity matrix with the corresponding size. Differentiating Eq. (SM11.6) with respect to $\theta_j$, gives [SM2]

$$\begin{aligned}
&\frac{\partial^2 \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta_i \partial \theta_j} =\\
&\quad -\frac{1}{2}\text{trace}\left[\left(\frac{\partial R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}}{\partial \theta_j}\right)\left(I - R^{-1}_{\theta,l}e_{\theta,l}e^T_{\theta,l}\right)\right] - \frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_j}R^{-1}_{\theta,l}e_{\theta,l}e^T_{\theta,l}\right]\\
&\quad + \frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}\left(\frac{\partial e_{\theta,l}}{\partial \theta_j}e^T_{\theta,l} + e_{\theta,l}\frac{\partial e^T_{\theta,l}}{\partial \theta_j}\right)\right] - \frac{\partial^2 e_{\theta,l}}{\partial \theta_i \theta_j}R^{-1}_{\theta,l}e_{\theta,l} - \frac{\partial e^T_{\theta,l}}{\partial \theta_i}\frac{\partial R^{-1}_{\theta,l}}{\partial \theta_j}e_{\theta,l}\\
&\quad - \frac{\partial e^T_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}\frac{\partial e_{\theta,l}}{\partial \theta_j}.
\end{aligned}$$

Therefore, the inner integral in Eq. (SM11.1) can be calculated as

$$\begin{aligned}
&\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial^2 \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta_i \theta_j}dr_l =\\
&\quad -\frac{1}{2}\text{trace}\Bigg[\left(\frac{\partial R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}}{\partial \theta_j}\right)\underbrace{\left(I - R^{-1}_{\theta,l}\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)e_{\theta,l}e^T_{\theta,l}dr_l\right)}_{Term_1}\Bigg]\\
&\quad -\frac{1}{2}\text{trace}\Bigg[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_j}R^{-1}_{\theta,l}\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)e_{\theta,l}e^T_{\theta,l}dr_l\Bigg]\\
&\quad +\frac{1}{2}\text{trace}\Bigg[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}\underbrace{\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\left(\frac{\partial e_{\theta,l}}{\partial \theta_j}e^T_{\theta,l} + e_{\theta,l}\frac{\partial e^T_{\theta,l}}{\partial \theta_j}\right)dr_l}_{Term_2}\Bigg]\\
&\quad -\underbrace{\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial^2 e^T_{\theta,l}}{\partial \theta_i \theta_j}R^{-1}_{\theta,l}e_{\theta,l}dr_l}_{Term_3} + \underbrace{\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial e^T_{\theta,l}}{\partial \theta_i}\frac{\partial R^{-1}_{\theta,l}}{\partial \theta_j}e_{\theta,l}dr_l}_{Term_4}\\
&\quad -\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial e^T_{\theta,l}}{\partial \theta_i}R^{-1}_{\theta,l}\frac{\partial e_{\theta,l}}{\partial \theta_j}dr_l.
\end{aligned} \tag{SM11.7}$$

Note that for $j = 1, \cdots, n$,

$$\begin{aligned}
&\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) e_{\theta,l} \frac{\partial e^T_{\theta,l}}{\partial\theta_j} dr_l \\
&\quad= \int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big)\Big(r_l - C\hat{x}^{l-1}_{\theta,l}\Big) \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_j} C^T dr_l \\
&\quad= \left[\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big)\Big(r_l - C\hat{x}^{l-1}_{\theta,l}\Big) dr_l\right] \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_j} C^T \\
&\quad= \left[\int_{\mathcal{C}} r_l p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) dr_l - C\hat{x}^{l-1}_{\theta,l}\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) dr_l\right] \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_j} C^T \\
&\quad= \left[C\hat{x}^{l-1}_{\theta,l} - C\hat{x}^{l-1}_{\theta,l}\right] \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_j} C^T = 0.
\end{aligned}$$

Similarly, $\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) \frac{\partial e_{\theta,l}}{\partial\theta_j} e^T_{\theta,l} dr_l = 0$, and therefore, $Term_2$, $Term_3$, and $Term_4$ in Eq. (SM11.7) are equal to zero. Then, noting that

$$\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) e_{\theta,l} e^T_{\theta,l} dr_l = R_{\theta,l},$$

we have $Term_1 = 0$, and Eq. (SM11.7) becomes

$$\begin{aligned}
&\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) \frac{\partial^2 \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big)}{\partial\theta_i\theta_j} dr_l \\
&\quad= -\frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i} R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_j}\right] \\
&\qquad - \left[\int_{\mathbb{R}^2} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\Big(r_l|\tau_l, d_{l-1}\Big) dr_l\right]\left(\frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_i} C^T R^{-1}_{\theta,l} C \frac{\partial\hat{x}^{l-1}_{\theta,l}}{\partial\theta_j}\right) \\
&\quad= -\frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i} R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_j}\right] - \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_i} C^T R^{-1}_{\theta,l} C \frac{\partial\hat{x}^{l-1}_{\theta,l}}{\partial\theta_j}. \qquad \text{(SM11.8)}
\end{aligned}$$

By substituting the above equation in Eq. (SM11.1), we have

$$\begin{aligned}
\left[I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta)\right]_{i,j} &= \frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i} R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_j}\right] \\
&\quad + \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p^{\theta}_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\Big(r_{1:l-1}|\tau_{1:l-1}\Big) \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_i} C^T R^{-1}_{\theta,l} C \frac{\partial\hat{x}^{l-1}_{\theta,l}}{\partial\theta_j} dr_{l-1}\cdots dr_1 \\
&= \frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i} R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_j}\right] + E\left[\frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_i} C^T R^{-1}_{\theta,l} C \frac{\partial\hat{x}^{l-1}_{\theta,l}}{\partial\theta_j}\right] \\
&= \frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i} R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_j}\right] + \text{trace}\left\{R^{-1}_{\theta,l} E\left[C\frac{\partial\hat{x}^{l-1}_{\theta,l}}{\partial\theta_j}\frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_i} C^T\right]\right\} \\
&= \frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i} R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_j}\right] + \text{trace}\left\{R^{-1}_{\theta,l} C E\left[\frac{\partial\hat{x}^{l-1}_{\theta,l}}{\partial\theta_j}\frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial\theta_i}\right] C^T\right\}. \qquad \text{(SM11.9)}
\end{aligned}$$

According to Eqs. (SM11.3) and (SM11.4),

$$\hat{x}^{l}_{\theta,l+1} = \tilde{\phi}_\theta(\tau_l, \tau_{l+1})\left(\hat{x}^{l-1}_{\theta,l} + K_{\theta,l}(r_l - C\hat{x}^{l-1}_{\theta,l})\right), \quad l = 1, 2, \cdots. \qquad \text{(SM11.10)}$$

Then, according to Lemma SM14.2 (see Section SM14.2), by differentiating Eq. (SM11.10) with respect to $\theta_i, i = 1, \cdots, n$, after some straightforward calculations, for $X^{(i)}_{\theta,l} := \begin{bmatrix} \hat{x}^{l-1}_{\theta,l} \\ \frac{\partial \hat{x}^{l-1}_{\theta,l}}{\partial \theta_i} \end{bmatrix}$, we have the following recursive formulation:

$$X^{(i)}_{\theta,l+1} = A^{(i)}_{\theta,l+1} X^{(i)}_{\theta,l} + B^{(i)}_{\theta,l+1} e_{\theta,l}, \quad \theta = (\theta_1, \cdots, \theta_n) \in \Theta, \quad i = 1, \cdots, n, \tag{SM11.11}$$

and

$$A^{(i)}_{\theta,l+1} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) & 0_{k\times k} \\ \frac{\partial \tilde{\phi}_\theta(\tau_l, \tau_{l+1})}{\partial \theta_i} & \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) \left( \tilde{I}_{k\times k} - K_{\theta,l} C \right) \end{bmatrix}, \quad B^{(i)}_{\theta,l+1} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) K_{\theta,l} \\ \frac{\partial \left( \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) K_{\theta,l} \right)}{\partial \theta_i} \end{bmatrix}.$$

According to Lemma SM14.5 (see Section SM14.5) and using Eq. (SM11.11), we have, for $l = 1, 2, \cdots$,

$$\begin{aligned}
&E\left\{ X^{(j)}_{\theta,l+1} \left( X^{(i)}_{\theta,l+1} \right)^T \right\} \\
&= E\Big\{ A^{(j)}_{\theta,l+1} X^{(j)}_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \left( A^{(i)}_{\theta,l+1} \right)^T + A^{(j)}_{\theta,l+1} X^{(j)}_{\theta,l} e^T_{\theta,l} \left( B^{(i)}_{\theta,l+1} \right)^T \\
&\qquad + B^{(j)}_{\theta,l+1} e_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \left( A^{(i)}_{\theta,l+1} \right)^T + B^{(j)}_{\theta,l+1} e_{\theta,l} e^T_{\theta,l} \left( B^{(j)}_{\theta,l+1} \right)^T \Big\} \\
&= A^{(j)}_{\theta,l+1} E\left\{ X^{(j)}_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \right\} \left( A^{(i)}_{\theta,l+1} \right)^T + A^{(j)}_{\theta,l+1} \underbrace{E\left\{ X^{(j)}_{\theta,l} e^T_{\theta,l} \right\}}_{0} \left( B^{(i)}_{\theta,l+1} \right)^T \\
&\qquad + B^{(j)}_{\theta,l+1} \underbrace{E\left\{ e_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \right\}}_{0} \left( A^{(i)}_{\theta,l+1} \right)^T + B^{(j)}_{\theta,l+1} E\left\{ e_{\theta,l} e^T_{\theta,l} \right\} \left( B^{(j)}_{\theta,l+1} \right)^T \\
&= A^{(j)}_{\theta,l+1} E\left\{ X^{(j)}_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \right\} \left( A^{(i)}_{\theta,l+1} \right)^T \\
&\qquad + B^{(j)}_{\theta,l+1} \left( CE\left[ (X_\theta(\tau_l) - \hat{x}^{l-1}_{\theta,l})(X_\theta(\tau_l) - \hat{x}^{l-1}_{\theta,l})^T \right] C^T + \Sigma_g \right) \left( B^{(j)}_{\theta,l+1} \right)^T \\
&= A^{(j)}_{\theta,l+1} E\left\{ X^{(j)}_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \right\} \left( A^{(i)}_{\theta,l+1} \right)^T + B^{(j)}_{\theta,l+1} R_{\theta,l} \left( B^{(j)}_{\theta,l+1} \right)^T. \qquad \text{(SM11.12)}
\end{aligned}$$

Finally, by rewriting the Fisher information expression (Eq. (SM11.9)) as (let $\tilde{C} := \begin{bmatrix} 0_{2\times k} & C \end{bmatrix}$, where $0_{2\times k}$ denotes the $2 \times k$ zero matrix)

$$\left[ I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta) \right]_{i,j} = \frac{1}{2}\text{trace}\left[ R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_j} \right] + \text{trace}\left\{ R^{-1}_{\theta,l} \tilde{C} E\left[ X^{(j)}_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \right] \tilde{C}^T \right\}, \tag{SM11.13}$$

and substituting Eq. (SM11.12) into Eq. (SM11.13), we have

$$\left[ I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta) \right]_{i,j} = \frac{1}{2}\text{trace}\left[ R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_j} \right] + \text{trace}\left\{ R^{-1}_{\theta,l} \tilde{C} S^{(ji)}_{\theta,l} \tilde{C}^T \right\},$$

where $S^{(ji)}_{\theta,l} := E\left\{ X^{(j)}_{\theta,l} \left( X^{(i)}_{\theta,l} \right)^T \right\}$, $l = 1, 2, \cdots$, can be calculated recursively as

$$\begin{aligned}
&S^{(ji)}_{\theta,l+1} - A^{(j)}_{\theta,l+1} S^{(ji)}_{\theta,l} \left( A^{(i)}_{\theta,l+1} \right)^T = B^{(j)}_{\theta,l+1} R_{\theta,l} \left( B^{(j)}_{\theta,l+1} \right)^T, \quad l = 2, 3, \cdots, \\
&S^{(ji)}_{\theta,1} = \begin{bmatrix} \tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta,0} + \tilde{a}_\theta(\tau_0, \tau_1) \\ \frac{\partial \left( \tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta,0} + \tilde{a}_\theta(\tau_0, \tau_1) \right)}{\partial \theta_j} \end{bmatrix} \begin{bmatrix} \left( \tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta,0} + \tilde{a}_\theta(\tau_0, \tau_1) \right) \\ \left( \frac{\partial \left( \tilde{\phi}_\theta(\tau_0, \tau_1) \bar{x}_{\theta,0} + \tilde{a}_\theta(\tau_0, \tau_1) \right)}{\partial \theta_i} \right) \end{bmatrix}^T, \qquad \text{(SM11.14)}
\end{aligned}$$

and it completes the proof.

**SM12. Example for Fisher information calculation.** For the data model described in the example provided in Section SM4, the Fisher information matrix is given by Eqs. (5.13) and (5.14), where $S_{\theta,l}, l = 1, 2, \cdots$, is given recursively by, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$ and $i, j = 1, \cdots, n$,

$$S_{\theta,l}^{(ji)} - A_{\theta,l}^{(j)} S_{\theta,l-1}^{(ji)} \left(A_{\theta,l}^{(i)}\right)^T = B_{\theta,l}^{(j)} R_{\theta,l-1} \left(B_{\theta,l}^{(j)}\right)^T, \quad l = 2, 3, \cdots,$$

and

$$S_{\theta,1}^{(ji)} = \begin{bmatrix} \phi_\theta(\tau_0, \tau_1) x_{\theta,0} \\ \frac{\partial(\phi_\theta(\tau_0,\tau_1)x_{\theta,0})}{\partial\theta_j} \end{bmatrix} \begin{bmatrix} (\phi_\theta(\tau_0, \tau_1) x_{\theta,0})^T & \left(\frac{\partial(\phi_\theta(\tau_0,\tau_1)x_{\theta,0})}{\partial\theta_i}\right)^T \end{bmatrix},$$

with coefficient matrices given through Eqs. (SM4.3)-(SM4.5). In Section SM4, we calculated these coefficient matrices for the first order drift and diffusion coefficients estimation problem in different scenarios.

**SM13. Computation of general Fisher information matrix.** We calculate the Fisher information matrix numerically, for the case that we have one photon, through the following algorithm (here, it is assumed that $\theta = D$, where $D$ is the diffusion coefficient).

1. For $a, b \in \mathbb{R}, a < b$, let $x_i := a + ih, y_i := a + ih, i = 0, \cdots, n$, and $h := \frac{b-a}{n}$. Approximate $p_{X(\tau_1)}$ as

$$\begin{aligned} p_{X(\tau_1)}(x_1) &= \int_{\mathbb{R}^2} p_{X(\tau_1)|X(t_0)}(x_1|x)\, p_{X(t_0)}(x)\, dx \\ &\approx h^2 \sum_{i=0}^{n} \sum_{j=0}^{n} p_{X(\tau_1)|X(t_0)}(x_1|(x_i, y_j))\, p_{X(t_0)}(x_i, y_j), \quad x_1 \in \mathbb{R}^2. \end{aligned}$$

2. Approximate $dp_{X(\tau_1)} := \frac{\partial p_{X(\tau_1)}}{\partial D}$ as

$$\begin{aligned} dp_{X(\tau_1)}(x_1) &= \int_{\mathbb{R}^2} dp_{X(\tau_1)|X(t_0)}(x_1|x)\, p_{X(t_0)}(x)\, dx \\ &\approx h^2 \sum_{i=0}^{n} \sum_{j=0}^{n} dp_{X(\tau_1)|X(t_0)}(x_1|(x_i, y_j))\, p_{X(t_0)}(x_i, y_j), \quad x_1 \in \mathbb{R}^2. \qquad \text{(SM13.1)} \end{aligned}$$

3. Approximate $p_{U_1|T_1}$ as

$$\begin{aligned} p_{U_1|T_1}(r|\tau_1) &= \int_{\mathbb{R}^2} p_{X(\tau_1)}(x)\, p_{U_1|X(\tau_1)}(r|x)\, dx \\ &= \int_{\mathbb{R}^2} p_{X(\tau_1)}(x)\, f_x(r)\, dx \\ &= \frac{1}{|\det(M)|} \int_{\mathbb{R}^2} p_{X(\tau_1)}(x)\, q\left(M^{-1}r - x\right) dx \\ &\approx \frac{h^2}{|\det(M)|} \sum_{i=0}^{n} \sum_{j=0}^{n} p_{X(\tau_1)}(x_i, y_j)\, q\left(M^{-1}r - (x_i, y_j)\right), \quad r \in \mathcal{C}. \end{aligned}$$

4. Approximate $dp_{U_1|T_1} := \frac{\partial p_{U_1|T_1}}{\partial D}$ as

$$\begin{aligned} dp_{U_1|T_1}(r|\tau_1) &= \int_{\mathbb{R}^2} dp_{X(\tau_1)}(x)\, f_x(r)\, dx \\ &= \frac{1}{|\det(M)|}\int_{\mathbb{R}^2} dp_{X(\tau_1)}(x)\, q\left(M^{-1}r - x\right) dx \\ &\approx \frac{h^2}{|\det(M)|}\sum_{i=0}^{n}\sum_{j=0}^{n} dp_{X(\tau_1)}(x_i, y_j)\, q\left(M^{-1}r - (x_i, y_j)\right), \quad r \in \mathcal{C}. \end{aligned}$$

5. Let $r_{x_i} = Mx_i, r_{y_i} = My_i, i = 0, \cdots, n$, and $h_r = Mh$. Approximate the Fisher information matrix $I(D)$ of diffusion coefficient $D$ as

$$\begin{aligned} I(D) &= \int_{\mathcal{C}} \frac{1}{p_{U_1|T_1}(r|\tau_1)} dp^2_{U_1|T_1}(r|\tau_1)\, dr \\ &\approx h_r^2 \sum_{i=0}^{n}\sum_{j=0}^{n} \frac{1}{p_{U_1|T_1}\left((r_{x_i}, r_{y_j})|\tau_1\right)} dp^2_{U_1|T_1}\left((r_{x_i}, r_{y_j})|\tau_1\right). \end{aligned}$$

## SM14. Appendix.

### SM14.1. Joint probability distribution of arrival time points for a Poisson process.

LEMMA SM14.1. *For $t_0 \in \mathbb{R}$, let $\{N(\tau), \tau \geq t_0\}$ be a Poisson process with intensity function $\Lambda(\tau), \tau \geq t_0$. Let $\mathcal{T}_l := (T_1, \cdots, T_l)^T, l = 1, \cdots, N(\tau), \tau \geq t_0$, where the 1D random variable $T_l$ describes the $l^{th}$ arrival time points of $\{N(\tau), \tau \geq t_0\}$.*

*1. Then, $N(\tau), \tau \geq t_0$, is Poisson distributed with mean $\int_{t_0}^{\tau} \Lambda(\psi) d\psi$, i.e., for $L = 0, 1, \cdots$, the probability $P\Big(N(\tau) = L\Big)$ is given by*

$$P(N(\tau) = L) = \frac{1}{L!}\left(\int_{t_0}^{\tau} \Lambda(\psi) d\psi\right)^L e^{-\int_{t_0}^{\tau} \Lambda(\psi) d\psi}, \quad \tau \geq t_0.$$

*2. For $t_0 \leq \tau_1 < \cdots < \tau_L, L = 1, 2, \cdots$, the probability density function $p_{\mathcal{T}_L}$ of $\mathcal{T}_L$ is given by*

$$p_{\mathcal{T}_L}\Big(\tau_1, \cdots, \tau_L\Big) = \left(\prod_{l=1}^{L} \Lambda(\tau_l)\right) e^{-\int_{t_0}^{\tau_L} \Lambda(\tau) d\tau}.$$

*3. For $t_0 \leq \tau_1 < \cdots < \tau_L \leq t, L = 1, 2, \cdots$, the conditional probability density function $p_{\mathcal{T}_L|N(t)}$ is given by*

$$p_{\mathcal{T}_L|N(t)}(\tau_1, \cdots, \tau_L|L) = \frac{L!\left(\prod_{l=1}^{L} \Lambda(\tau_l)\right)}{\left(\int_{t_0}^{t} \Lambda(\tau) d\tau\right)^L}.$$

*Proof.* See Section 2 of [SM6]. ∎

### SM14.2. Derivative of state estimates.

LEMMA SM14.2. *Let $\Theta$ denote a parameter space that is an open subset of $\mathbb{R}^n$, and let $\tau_1 \in \mathbb{R}$. For $\theta = (\theta_1, \cdots, \theta_n) \in \Theta, r_l \in \mathcal{C}, l = 1, 2, \cdots$, and $\tau_1 < \tau_2 < \cdots$, let*

$$\hat{x}^l_{\theta,l+1} = \tilde{\phi}_\theta(\tau_l, \tau_{l+1})\left(\hat{x}^{l-1}_{\theta,l} + K_{\theta,l}(r_l - C\hat{x}^{l-1}_{\theta,l})\right), \quad \hat{x}^l_{\theta,l+1} \in \mathbb{R}^k, \tag{SM14.1}$$

*where* $\tilde{\phi}_\theta(\tau_l, \tau_{l+1}) \in \mathbb{R}^{k\times k}, C \in \mathbb{R}^{2\times k}, K_{\theta,l} \in \mathbb{R}^{k\times 2}$*, and their derivatives with respect to* $\theta_i, i = 1, \cdots, n$*, exist. Let* $X^{(i)}_{\theta,l} := \begin{bmatrix} \hat{x}^{l-1}_{\theta,l} \\ \frac{\partial \hat{x}^{l-1}_{\theta,l}}{\partial\theta_i} \end{bmatrix}$ *and* $e_{\theta,l} := r_l - C\hat{x}^{l-1}_{\theta,l}$*. Then,*

$$X^{(i)}_{\theta,l+1} = A^{(i)}_{\theta,l+1}X^{(i)}_{\theta,l} + B^{(i)}_{\theta,l+1}e_{\theta,l}, \quad l = 1, 2, \cdots,$$

*where*

$$A^{(i)}_{\theta,l+1} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) & 0_{k\times k} \\ \frac{\partial\tilde{\phi}_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i} & \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) \left(I_{k\times k} - K_{\theta,l}M\right) \end{bmatrix},$$

$$B^{(i)}_{\theta,l+1} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_l, \tau_{l+1})K_{\theta,l} \\ \tilde{\phi}_\theta(\tau_l, \tau_{l+1})\frac{\partial K_{\theta,l}}{\partial\theta_i} + \frac{\partial\tilde{\phi}_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i}K_{\theta,l} \end{bmatrix}.$$

*Proof.* By differentiating Eq. (SM14.1) (Kalman state estimate update formula) with respect to $\theta_i, i = 1, \cdots, n$, we have, for $l = 1, 2, \cdots$,

$$\begin{aligned} \frac{\partial \hat{x}^{l}_{\theta,l+1}}{\partial\theta_i} &= \begin{bmatrix} \frac{\partial\tilde{\phi}_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i} & \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) \left(I_{k\times k} - K_{\theta,l}C\right) \end{bmatrix} \begin{bmatrix} \hat{x}^{l-1}_{\theta,l} \\ \frac{\partial \hat{x}^{l-1}_{\theta,l}}{\partial\theta_i} \end{bmatrix} \\ &+ \left( \tilde{\phi}_\theta(\tau_l, \tau_{l+1})\frac{\partial K_{\theta,l}}{\partial\theta_i} + \frac{\partial\tilde{\phi}_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i}K_{\theta,l} \right) e_{\theta,l}, \end{aligned} \tag{SM14.2}$$

Then, by combining Eqs. (SM14.1) and (SM14.2), for $X^{(i)}_{\theta,l} = \begin{bmatrix} \hat{x}^{l-1}_{\theta,l} \\ \frac{\partial \hat{x}^{l-1}_{\theta,l}}{\partial\theta_i} \end{bmatrix}$, we have the following recursive formulation

$$X^{(i)}_{\theta,l+1} = A^{(i)}_{\theta,l+1}X^{(i)}_{\theta,l} + B^{(i)}_{\theta,l+1}e_{\theta,l},$$

where

$$A^{(i)}_{\theta,l+1} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) & 0_{k\times k} \\ \frac{\partial\tilde{\phi}_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i} & \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) \left(I_{k\times k} - K_{\theta,l}C\right) \end{bmatrix},$$

$$B^{(i)}_{\theta,l+1} := \begin{bmatrix} \tilde{\phi}_\theta(\tau_l, \tau_{l+1})K_{\theta,l} \\ \tilde{\phi}_\theta(\tau_l, \tau_{l+1})\frac{\partial K_{\theta,l}}{\partial\theta_i} + \frac{\partial\tilde{\phi}_\theta(\tau_l,\tau_{l+1})}{\partial\theta_i}K_{\theta,l} \end{bmatrix}.$$

□

### SM14.3. Chain rule.

THEOREM SM14.3. *Let* $S$ *be an open set in* $\mathbb{R}^K$ *and let* $c$ *be a point of* $S$*. Let* $d = (d_1, \cdots, d_M)$ *be a function mapping* $S$ *into an open set* $H$ *in* $\mathbb{R}^M$*, i.e.,* $d$: $S \mapsto H$*, that is differentiable at* $c$*. Let* $h = (h_1, \cdots, h_N)$ *be a function mapping* $H$ *into an open set* $Q$ *in* $\mathbb{R}^N$*, i.e.,* $h$: $H \mapsto Q$*, that is differentiable at* $d(c)$*. Let* $q$ *be a real-valued function defined on* $Q$ *that is differentiable at* $h(d(c))$*. Then,*

$$(D_k(q \circ h \circ d))(c) = \sum_{i=1}^{N}\sum_{j=1}^{M}(D_iq)(h(d(c)))(D_jh_i)(d(c))(D_kd_j)(c), \quad k = 1, \cdots, K.$$

*Proof.* See the proof of Corollary 8.4.3 of [SM1]. □

### SM14.4. Integral transformation theorem.

THEOREM SM14.4. *Let* $g = (g_1, g_2, \cdots, g_n)$: $B \subseteq \mathbb{R}^n \mapsto \mathbb{R}^n$ *be an injective and continuously differentiable function. Let* $w$: $\mathbb{R}^n \mapsto \mathbb{R}$ *be an integral function and* $A \subseteq \mathbb{R}^n$*, then the* integral transformation theorem *is given by*

$$\begin{aligned} \int_{g(A)} w(y_1, y_2, \cdots, y_n)dy_1dy_2\cdots dy_n = \int_A & w(g(x_1, x_2, \cdots, x_n)) \\ & \times \left|\det(J(g)(x_1, x_2, \cdots, x_n))\right| dx_1dx_2\cdots dx_n, \end{aligned}$$

*where the* Jacobian matrix *is given by*

$$J(g) := \begin{bmatrix} \frac{\partial g_1(x_1,x_2,\cdots,x_n)}{\partial x_1} & \frac{\partial g_1(x_1,x_2,\cdots,x_n)}{\partial x_2} & \cdots & \frac{\partial g_1(x_1,x_2,\cdots,x_n)}{\partial x_n} \\ \frac{\partial g_2(x_1,x_2,\cdots,x_n)}{\partial x_1} & \frac{\partial g_2(x_1,x_2,\cdots,x_n)}{\partial x_2} & \cdots & \frac{\partial g_2(x_1,x_2,\cdots,x_n)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n(x_1,x_2,\cdots,x_n)}{\partial x_1} & \frac{\partial g_n(x_1,x_2,\cdots,x_n)}{\partial x_2} & \cdots & \frac{\partial g_n(x_1,x_2,\cdots,x_n)}{\partial x_n} \end{bmatrix}.$$

*Proof.* See Section 10.3 of [SM1]. ∎

**SM14.5. Innovation representation of the state space model.**

LEMMA SM14.5. *Let* $\mathcal{G}^g_{[t]}\left(\left(\tilde{X}, H, \tilde{W}_g, Z_g\right), \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \tilde{\Phi}, M', \mathcal{C}, \Theta\right)$ *(or* $\mathcal{G}^g_L\left(\left(\tilde{X}, H, \tilde{W}_g, Z_g\right), \left(\mathcal{U}_L, \mathcal{T}_L\right), \tilde{\Phi}, M', \mathcal{C}, \Theta\right)$*) be an image detection process with expanded state space* $\tilde{X}$ *and Gaussian process and measurement noise models for a time interval* $[t_0, t]$ *(or for a fixed number* $L$ *of photons). Let* $C := M'H$*. Assume that* $C$ *and* $Z_g$ *are independent of the parameter vector* $\theta \in \Theta$*. For* $\theta = (\theta_1, \cdots, \theta_n)$ *and* $\hat{x}^{l-1}_{\theta,l} := E\left[\tilde{X}_\theta(\tau_l)|r_{l-1}, \cdots, r_1\right]$*, let* $X^{(i)}_{\theta,l} := \begin{bmatrix} \hat{x}^{l-1}_{\theta,l} \\ \frac{\partial \hat{x}^{l-1}_{\theta,l}}{\partial \theta_i} \end{bmatrix}, i = 1, \cdots, n$*, be the extended state vector and* $e_{\theta,l} := r_l - C\hat{x}^{l-1}_{\theta,l}$ *be the prediction error. Then,* $E\left[e_{\theta,l} X^{(i)}_{\theta,l}\right] = 0, i = 1, \cdots, n$*.*

*Proof.* Since the measurement noise $Z_g$, the process noise $\tilde{W}_g$, and the initial condition of the state vector $\tilde{X}$ are independent, the prediction error $e_{\theta,l}$ and the extended state vector $X^{(i)}_{\theta,l}, i = 1, \cdots, n$, are independent (see the proof of Theorem 5 of [SM4]), and we have

$$\begin{aligned} E\left[X^{(i)}_{\theta,l} e_{\theta,l}\right] &= E\left[X^{(i)}_{\theta,l}\right] E\left[e_{\theta,l}\right] \\ &= E\left[X^{(i)}_{\theta,l}\right] E\left[C(\tilde{X}_\theta(\tau_l) - \hat{x}^{l-1}_{\theta,l}) + Z_{g,l}\right] \\ &= E\left[X^{(i)}_{\theta,l}\right] \left\{C\left(E\left[\tilde{X}_\theta(\tau_l)\right] - E\left[E\left[\tilde{X}_\theta(\tau_l)|r_{l-1}, \cdots, r_1\right]\right]\right) + E\left[Z_{g,l}\right]\right\}. \end{aligned}$$

According to the law of total expectation, $E\left[E\left[\tilde{X}_\theta(\tau_l)|r_{l-1}, \cdots, r_1\right]\right] = E\left[\tilde{X}_\theta(\tau_l)\right]$, and therefore, we have

$$E\left[X^{(i)}_{\theta,l} e_{\theta,l}\right] = E\left[X^{(i)}_{\theta,l}\right] \left\{C\left(E\left[\tilde{X}_\theta(\tau_l)\right] - E\left[\tilde{X}_\theta(\tau_l)\right]\right) + 0\right\} = 0.$$ ∎

**SM14.6. Analysis of the error of the predicted locations of the molecule.** In this section, the errors between the means of the distributions of the prediction of the molecule locations and the true locations of the molecule for Born and Wolf, Airy and Gaussian measurements are shown in Figs. SM4, SM5 and SM6, respectively.

REFERENCES

[SM1] S. A. DOUGLASS, *Introduction to mathematical analysis*, Addison-Wesley Publishing Company, USA, (1996).

[SM2] A. C. HARVEY, *Forecasting, structural time series models and the Kalman filter*, Cambridge Univ. Press, UK, (1990).

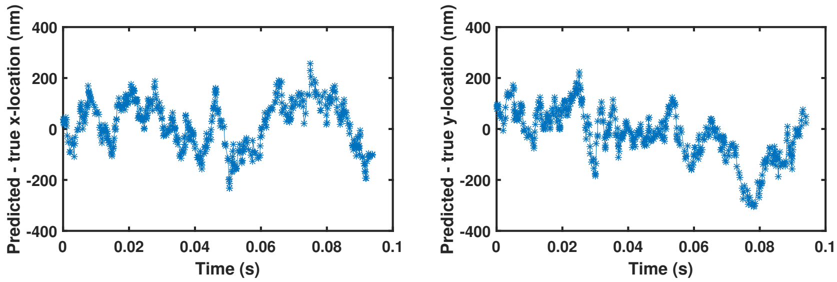


Fig. SM4. *Analysis of the error of the predicted locations of the molecule for the Born and Wolf measurement model. Shown in the left and right plots are the differences between the means of the distributions of the prediction of the molecule $x$-locations and the true $x$-values, and the means of the distributions of the prediction of the molecule $y$-locations and the true $y$-values, respectively, for the data sets of Fig. 5.*

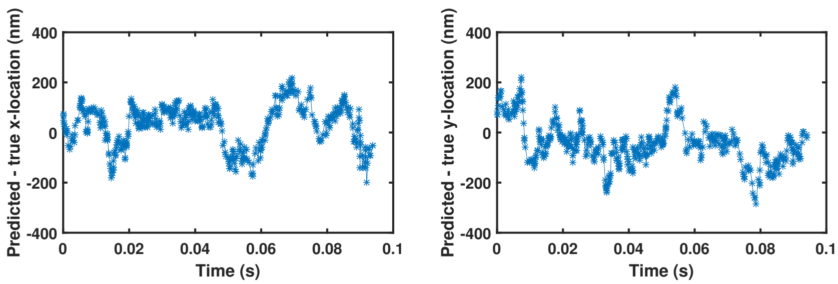


Fig. SM5. *Analysis of the error of the predicted locations of the molecule for the Airy measurement model. Shown in the left and right plots are the differences between the means of the distributions of the prediction of the molecule $x$-locations and the true $x$-values, and the means of the distributions of the prediction of the molecule $y$-locations and the true $y$-values, respectively, for the data sets of Fig. SM3.*

[SM3] A. H. Jazwinski, *Stochastic processes and filtering theory*, Acad. Press, New York, USA, (1970).
[SM4] R. L. M. Peeters and B. Hanzon, *Symbolic computation of Fisher information matrices for parametrized state-space systemss*, Automatica, 35 (1999), pp. 1059–1071.
[SM5] T. B. Schön, A. Wills, and B. Ninness, *System identification of nonlinear state-space models*, Automatica, 47 (2011), pp. 39–49.
[SM6] D. L. Snyder and M. I. Miller, *Random point processes in time and space*, Springer Verlag, New York, USA, (1991).
[SM7] R. L. Streit, *Poisson point processes*, Springer, Boston, USA, (2010).
[SM8] R. Zamir, *A proof of the Fisher information inequality via a data processing argument*, IEEE Trans. Inform. Theory, 44 (1998), pp. 1246–1250.

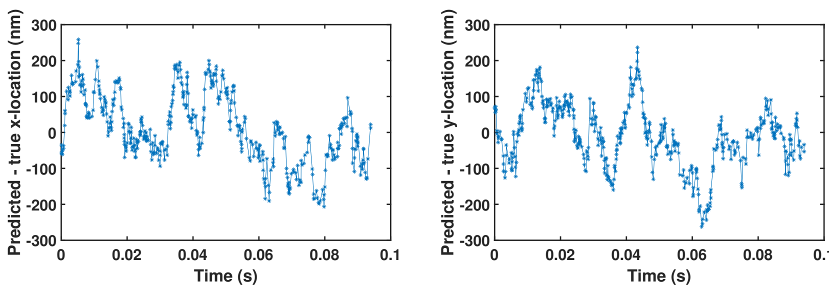


FIG. SM6. *Analysis of the error of the predicted locations of the molecule for the Gaussian measurement noise case. Shown in the left and right plots are the differences between the means of the distributions of the prediction of the molecule x-locations and the true x-values, and the means of the distributions of the prediction of the molecule y-locations and the true y-values, respectively, for the data sets of Fig. 8.*