Limits of accuracy for parameter estimation and localisation in Single-Molecule Microscopy via sequential Monte Carlo methods*

Alix Marie d'Avigneau[†‡], Sumeetpal S. Singh[†], and Raimund J. Ober[‡]

Abstract. Assessing the quality of parameter estimates for models describing the motion of single molecules in cellular environments is an important problem in fluorescence microscopy. In this work, we consider the fundamental data model, where molecules emit photons at random time instances and these photons arrive at random locations on the detector according to complex point spread functions (PSFs). The randomness and non-Gaussian PSF of the detection process, and the random trajectory of the molecule, makes inference challenging. Moreover, the presence of other closely spaced molecules causes further uncertainty in the origin of the measurements, which impacts the statistical precision of the estimates. We quantify the limits of accuracy of model parameter estimates and separation distance between closely spaced molecules (known as the resolution problem) by computing the Cramér-Rao lower bound (CRLB), or equivalently the inverse of the Fisher information matrix (FIM), for the variance of estimates. Results on the CRLB obtained from the fundamental model are crucial, in that they provide a lower bound for more practical scenarios. While analytic expressions for the FIM can be derived for static and deterministically moving molecules, the analytical tools to evaluate the FIM for molecules whose trajectories follow stochastic differential equations (SDEs) are still for the most part missing. We address this by presenting a general sequential Monte Carlo (SMC) based methodology for both parameter inference and computing the desired accuracy limits for non-static molecules and a non-Gaussian fundamental detection model. For the first time, we are able to estimate the FIM for stochastically moving molecules observed through the Airy and Born and Wolf detection models. This is achieved by estimating the score and observed information matrix via SMC. We summarise the outcome of our numerical work by delineating the qualitative behaviours for the accuracy limits as functions of various experimental settings like collected photon count, molecule diffusion, etc. We also verify that we can recover known results from the static molecule case.

Key words. Single-molecule microscopy, Fluorescence microscopy, Particle filtering, Particle smoothing, Sequential Monte Carlo (SMC), Fisher information matrix, Stochastic differential equations (SDEs)

AMS subject classifications. 65C05, 92C55

1. Introduction.

1.1. Motivation. In recent years, *single-molecule microscopy* has become a powerful tool in cell biology [54, 53]. It has allowed significant insight to be gained into the behaviour of single molecules in cellular environments using *fluorescence microscopy*. Single-molecule fluorescence microscopy (see [55, 39] for reviews) consists of using a suitable fluorophore to label the molecule(s) of interest, exciting said fluorophore with a specific light source and capturing the fluorescence or photons emitted by the molecule(s) through an optical micro-

*Submitted to the editors 04/06/2021.

Funding: This work was funded in part by Wellcome Trust.

[†]Signal Processing and Communications Group, Department of Engineering, University of Cambridge, Cambridge, UK (agem2@cam.ac.uk, sss40@cam.ac.uk).

[‡]Center for Cancer Immunology, Faculty of Medicine, University of Southampton, Southampton, UK (a.marie-davigneau@soton.ac.uk, r.ober@soton.ac.uk).

scope system onto a detector during a fixed acquisition time. Many biological applications rely on being able to accurately track moving molecules (or localise them in the static case) and also estimate their model parameters. Molecule location estimates, which are themselves useful, are also used to estimate the separation distance between two closely spaced molecules [54, 38], which is needed to quantify the microscopy technique's resolution (discussed below). By model parameters, we mean the drift and diffusion coefficients that describe the motion of randomly moving molecules, but also more generally other parameters for any assumed statistical elements/model for the image acquisition pipeline (see Example 2.1, Example 4.2, section 5). In addition to solving these estimation problems by devising appropriate numerical techniques to compute them, it is also essential to quantify their accuracy, and tools from statistical estimation theory such as the *Cramér-Rao lower bound* (CRLB) [13, 51, 25, 14] are popular in the microscopy literature [40, 10, 48]. Not only is the CRLB able to quantify the accuracy of the estimates, it can also provide the qualitative relationship between estimation accuracy and various experimental settings, such as the average number of photons captured by the detector, the speed of one or multiple diffusing molecules, or the distance between molecules, which is particularly important in experimental design. For example, one might aim to evaluate how an increase in the speed (or diffusion) of a stochastically diffusing molecule might reduce the accuracy of estimates for its mean location, and whether this loss in accuracy can be mitigated by increasing the mean number of photons captured by the detector.

1.2. Methods for assessing the accuracy of parameter estimates. In the past, in the context of the resolution problem, *Rayleigh's criterion* [4] has been used to define the minimum distance between two point sources such that they can be distinguished in the image. However, a drawback of employing Rayleigh's criterion is that it ignores the statistical aspect of the separation distance estimation problem. For example, it doesn't account for the fact that each new observation (taking the form of a captured emitted photon) brings new information on the separation distance. In contrast, in estimation theory, the CRLB establishes a lower bound on the variance of unbiased estimates, and is therefore often used as a benchmark for the quality of a given estimator. As a result, the CRLB plays an important part in experimental design for single-molecule microscopy [40, 49]. For example, in [48, 50], the authors present an improved microscope resolution measure in the form of the square root of the CRLB for the separation distance between two molecules, which is referred to as the *limit of accuracy* with which the separation distance between the two objects can be estimated based on the observed data. A particular advantage of this new resolution measure is that it predicts that increasing the photon count makes it possible to estimate a separation distance between two molecules that is shorter than Rayleigh's criterion. In the context of localisation and estimation of parameters for models describing the motion of a single molecule, we also quantify the limits of accuracy for these model parameter estimates by computing the CRLB.

Evaluating these limits of accuracy is a challenging task. In this paper, we consider the *fundamental data model* [40, 49], which is crucial in that it provides more easily computed lower bounds for the limits of accuracy of more realistic practical models, where factors such as pixelisation and readout noise come into play and make inference more challenging [58]. Indeed, the limits of accuracy derived for the fundamental model are often known as the *fundamental limits of accuracy*. In this model, the detection process of the emitted fluorescence

already presents its own challenges, as it is intrinsically random both in time and location. While many methods [7, 6, 8] have assumed that the arrival times of the photons on the detector were uniformly distributed, [40, 49] suggest that the arrival times of photons follow a Poisson process. As for the arrival location of these photons on the detector, a wide range of measurement models exist − corresponding to the various types of detector. The typical measurement model used for an in-focus source is the *Airy profile* [59, 10]. If the molecule is out of focus, 3D models are generally used instead, such as the *Born and Wolf model* [4]. Often, these models make parameter inference difficult, and researchers have often opted for a Gaussian approximation to these models, such as in [2, 52, 37]. However, [59] argue that in practice, assuming Gaussian distributed photon locations on the detector is not an accurate approximation of the underlying model.

While it is important to be able to accurately study the behaviours and interactions of single molecules within a cell, it is especially challenging when those molecules have stochastic trajectories. The motion of an object in a cellular environment is affected by a multitude of deterministic, as well as random factors [5], and in many applications [59, 6], the trajectories of single molecules are modelled by *stochastic differential equations* (SDEs) [43]. The CRLB is obtained by taking the inverse of the *Fisher information matrix* (FIM), and analytical expressions for the FIM, and thus the limit of accuracy (given by the square root of the CRLB) for the location of an in-focus *static* (or unmoving) molecule have been derived in [40, 10]. Similar results for an out-of-focus static molecule are available in [41], and analytical expressions have also been derived in the context of molecules with deterministic linear or circular trajectories in [60]. As for the resolution problem, it is addressed in [48, 49] in a static molecule context and in [36] for two dynamic molecules with deterministic trajectories. However, when molecules have stochastic trajectories, the analytical tools to obtain the CRLB and tackle many of these problems are still for the most part missing. In this paper, we propose a numerical approach to address these problems.

In the context of stochastically moving molecules, [59] developed a method to obtain the FIM for a molecule whose trajectory is described by a linear SDE. For a 2D Gaussian approximation of the photon detection process, the authors take advantage of the Kalman filter formulae to obtain an analytical form for the FIM for a specific set of photon detection times. However, if the Airy profile is used instead, the computational cost of performing numerical integration becomes prohibitive for more than a single photon. Among other things, we build on [59] and provide effective methodological advances which enable the estimation of the FIM for the hyperparameters of models with Airy and Born and Wolf distributed photon locations.

1.3. Contributions. In this paper, we develop an effective and general numerical framework to obtain *sequential Monte Carlo* (SMC) approximations of expectations of interest, including for stochastically moving molecules. The ability to approximate these expectations is important for estimating the *score* and *observed information matrix* (OIM) for the hyperparameters of interest, and can also be employed to obtain maximum likelihood (ML) estimates of said hyperparameters. Access to the score and/or OIM is vital in order to be able to estimate the FIM. To achieve this, the observation interval is first discretised and the problem reformulated as a discrete-time state space model, which takes into account the random arrival times of photons on the detector in the form of missing observations. Then, a

particle filter is employed in conjunction with *forward smoothing* methods [16, 46] to obtain particle approximations of the expectations of interest. Our work complements [1], in which the authors similarly employed time discretisation of the observation interval, but they did not attempt to estimate the CRLB for hyperparameters. With our approach, we are for the first time able to obtain the limits of accuracy for parameters of a single molecule whose trajectory follows an SDE, thus providing new insights beyond existing results for molecules that are static or following a deterministic trajectory. Our SMC-based methodology is also more general than the Kalman filter-based approach of [59], and has no systemic limitations (i.e. variance in estimates of the limits of accuracy can always be reduced by increasing the number of Monte Carlo samples). We are also able to generalise results for the optical microscope resolution problem from considering the separation distance between two static molecules to that between two stochastically diffusing molecules.

The numerical experiments in this paper consist first of applying the methodology to estimate the limit of accuracy for a single stochastically moving molecule with 2D Gaussian, Airy, and Born and Wolf photon detection models by using estimates of the score and OIM obtained by forward smoothing. This is repeated for various expected mean photon counts to verify that for molecules with stochastic trajectories, the limit of accuracy exhibits an inverse square root decay with respect to mean photon count, i.e. the uncertainty of the hyperparameter estimates decreases as the expected number of photons increases. This has already been proven for static molecules [42, 41, 48]. The methodology is also applied in the context of the optical microscope resolution problem to obtain estimates of the limit of accuracy for the mean separation distance between two closely spaced diffusing molecules. Thanks to our numerical approach, insights can be obtained into the generalisation to diffusing molecules of results proven in [50] on this resolution problem for two static molecules. For instance, in [40], it was shown that the limit of accuracy for the location of a static molecule has a linear relationship with the standard deviation of the photon detection profile. From our numerical results, we show that when molecules are diffusing, the appropriate relationship behaves qualitatively with the diffusion coefficient standard deviation in a similar way, i.e. it can be translated into additional observation uncertainty. The qualitative relationships observed through our numerical experiments for stochastically moving molecules are summarised in Table 1.

For reference, a glossary of the mathematical symbols employed throughout this paper is available in Table 2. This paper is structured as follows. In section 2, the model is presented, including the molecule trajectory, described by a stochastic differential equation (SDE), and the photon detection time and location processes. In section 3, the model is formulated as a discrete-time state space model with a discretised observation interval. Then, section 4 establishes the main parameter inference aims and methods, which consist of particle filtering and smoothing of additive functionals in order to estimate the score and OIM for hyperparameters, and methods to estimate the FIM from the score and OIM. Numerical experiments are run in section 5 to first estimate the limit of accuracy for the drift and diffusion coefficients of the SDE for all photon detection profiles and then estimate the limit of accuracy for the separation distance between two dynamic molecules. Finally, section 6 provides concluding remarks.

2. Model specification. For the purpose of this paper, a basic optical system is considered, also known in [59, 10] as the *fundamental data model*. See Figure 1 for an overview of the optical system. Under the fundamental model, we assume that the photons are observed under ideal conditions, in which the detector $\mathcal{Y} = \mathbb{R}^2$ is non-pixelated. This model does not describe image data obtained from actual microscopy experiments the way more realistic, or *practical* models do. However, the fundamental model is crucial, in that it offers an obtainable lower bound to the CRLB of parameters of the more realistic practical model, which is much more difficult to obtain. In this section, the various aspects of the model are described. These include the true molecule trajectory, occurring in the object space, the photon detection locations in the image space, and the times at which photons arrive on the detector.

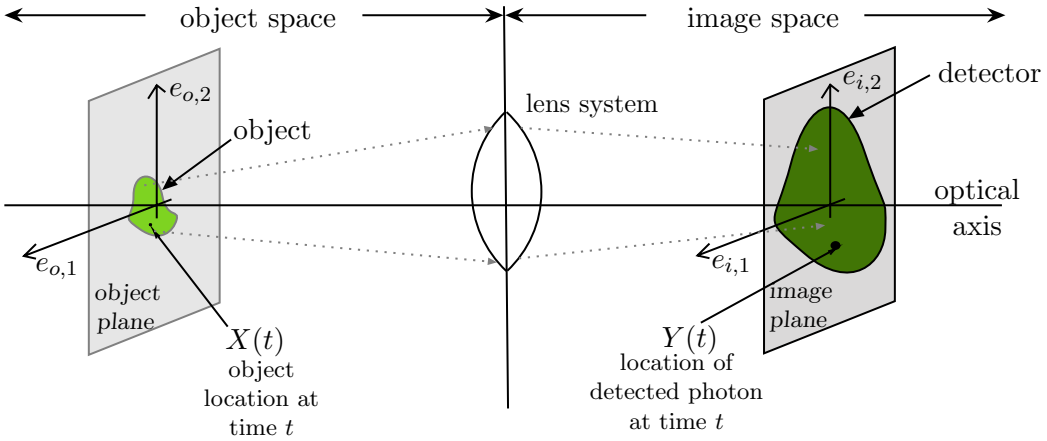


Figure 1: *Illustration of an optical microscope. At time $t \geq t_0$, the molecule is located at $X(t)$ in the object space and might be moving along the object plane. If the molecule is out of focus, it will instead move along a plane parallel to the object plane but displaced along the optical axis. The molecule emits photons through the lens system into the image space and its image is acquired on the planar detector $\mathcal{Y}$ located on the image plane. The location of the detected photons at time t is denoted by $Y(t)$.*

2.1. Molecule trajectory. For notational simplicity, let $X_t := X(t) \in \mathbb{R}^d$ denote the true, d-dimensional location of the molecule at time t. Given hyperparameters θ, let $f^{\theta}_{s,t}(x_t|x_s)$ denote the probability density function of X_t given the previous location X_s. Assume that the molecule trajectory $(X_t)_{t_0 \leq t \leq T}$ follows a linear *stochastic differential equation* (SDE)

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}B_t, \tag{2.1}$$

where $b(t, X_t) := b_0 + b_1(t)X_t$ and $\sigma(t, X_t) := \sigma(t)$ represent the *drift* and *diffusion* coefficients, respectively, b_0 is the zero order drift coefficient, and $(\mathrm{d}B_t)_{t_0 \leq t \leq T}$ is a Wiener process with $\mathbb{E}\left[\mathrm{d}B_t \mathrm{d}B_t^{\intercal}\right] = \mathbb{I}_{d \times d}$. According to [31, 24] the solution to the SDE in (2.1) at discrete time points $t_0 < t_1 < \ldots$ is given by

$$X_{t_{i+1}} = \Phi(t_i, t_{i+1})X_{t_i} + a(t_i, t_{i+1}) + W_g(t_i, t_{i+1}), \tag{2.2}$$

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

$$\frac{\mathrm{d}\Phi(s,t)}{\mathrm{d}t} = b_1(t)\Phi(s,t), \tag{2.3}$$
$$\Phi(t,t) = \mathbb{I}_{d\times d}, \qquad \Phi(s,t)\Phi(t,u) = \Phi(s,u),$$

the vector $a(t_i, t_{i+1}) \in \mathbb{R}^d$ is given by

$$a(t_i,t_{i+1}) = \int_{t_i}^{t_{i+1}} b_0\Phi(t_i,t)\mathrm{d}t,$$

and finally the process $\left(W_g(t_i,t_{i+1}) = \int_{t_i}^{t_{i+1}} \Phi(t_i,t)\sigma(t)\mathrm{d}B_t\right)_{i=1}^{\infty}$ is a white noise sequence with mean zero and covariance

$$R(t_i,t_{i+1}) = \int_{t_i}^{t_{i+1}} \Phi(t_i,t)\sigma(t)\sigma^\intercal(t)\Phi^\intercal(t_i,t)\mathrm{d}t. \tag{2.4}$$

Therefore, the transition density $f^{\theta}_{t_{i+1},t_i}(x'|x)$ can be expressed as a Gaussian with mean $\mu(x,t_i,t_{i+1}) = \Phi(t_i,t_{i+1})x + a(t_i,t_{i+1})$ and covariance $R(t_i,t_{i+1})$:

$$X_{t_{i+1}}|(X_{t_i} = x) \sim \mathcal{N}\left(\mu(x,t_i,t_{i+1}), R(t_i,t_{i+1})\right). \tag{2.5}$$

Example 2.1. Let the trajectory of a molecule be given by the following SDE

$$\mathrm{d}X_t = b\mathbb{I}_{d\times d}X_t\mathrm{d}t + \sqrt{2}\sigma\mathrm{d}B_t, \tag{2.6}$$

where in the drift term $b \in \mathbb{R}$, in the diffusion term $\sigma > 0$, and $(\mathrm{d}B_t)_{t_0\leq t\leq T}$ is a Wiener process and let $\theta = (\sigma^2, b)$. Assuming the time points $t_0, t_1, \ldots$ are equidistant, i.e. $t_{i+1} - t_i = \Delta$ for all $i = 0, 1, \ldots$, let the fundamental matrix $\Phi_\Delta := \varphi^\theta_\Delta \mathbb{I}_{d\times d}$ where $\varphi^\theta_\Delta \in \mathbb{R}$ and the covariance matrix $R_\Delta := r^\theta_\Delta \mathbb{I}_{d\times d}$ where $r^\theta_\Delta > 0$. Then, by solving (2.3) and plugging the result into (2.4), we obtain

$$\varphi^\theta_\Delta = \begin{cases} e^{\Delta b}, & \text{if } b \neq 0, \\ 1, & \text{if } b = 0, \end{cases} \qquad \text{and} \qquad r^\theta_\Delta = \begin{cases} \frac{\sigma^2}{b}\left(e^{2\Delta b} - 1\right), & \text{if } b \neq 0, \\ 2\sigma^2\Delta, & \text{if } b = 0. \end{cases}$$

The initial distribution $X_{t_0} \sim \mathcal{N}(x_0, P_0)$ has covariance matrix $P_0 = p_0^2\mathbb{I}_{d\times d}$ where $p_0 \in \mathbb{R}$.

In a 2D setting (i.e. $d = 2$), let the drift $b = -10\ \mathrm{s}^{-1}$, the diffusion $\sigma^2 = 1\ \mu\mathrm{m}^2/\mathrm{s}$ and the initial covariance $p_0^2 = 10^{-2}\ \mu\mathrm{m}^2$ and mean $x_0 = (4.4, 4.4)^\intercal\ \mu\mathrm{m}$. Note that for the purpose of this example, the initial covariance matrix is diagonal, but there is no restriction to employing a more general, non-diagonal initial covariance matrix. By simulating the molecule trajectory for the time interval $[0, 0.1]$ seconds, we obtain the trajectory in Figure 2.

2.2. Photon detection locations. The true molecule trajectory cannot be observed directly. Instead, a fluorescence microscope is used: the molecule of interest is labelled using a suitable fluorophore, magnified through a lens system and the photons it emits arrive on a detector $\mathcal{Y} := \mathbb{R}^2$ for a fixed time period (see Figure 1). The arrival location of a photon on the

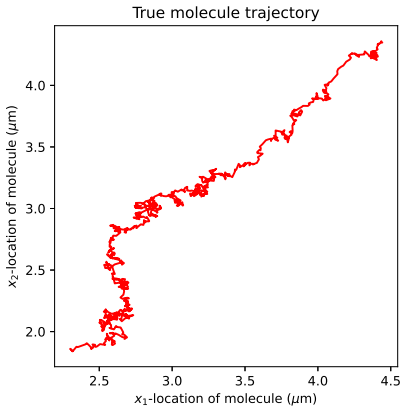


Figure 2: *Trajectory of a molecule in the object space with stochastic trajectory described in (2.6) and with diffusion and drift coefficients $\sigma^2 = 1\ \mu m^2/s$ and $b = -10\ s^{-1}$, respectively. The molecule moves during an interval of $[0, 0.1]$ seconds and its initial location is Gaussian distributed with mean $x_0 = (4.4, 4.4)^\intercal\ \mu m$ and covariance $P_0 = 10^{-2}\mathbb{I}_{2\times 2}\ \mu m^2$.*

detector is random, and using the typical approximation of the optical microscope from [27], it can be described as follows. Let $Y \in \mathcal{Y}$ denote the observed location of a detected photon. For an object located at $(x_{0,1}, x_{0,2}, z_0) \in \mathbb{R}^3$ in the object space, its *photon distribution profile* [49] is given by the density

$$g_\theta(y|x) := \frac{1}{|M|} q_{z_0}\left(M^{-1}y - (x_{0,1}, x_{0,2})^\intercal\right), \quad y \in \mathbb{R}^2, \tag{2.7}$$

where $M \in \mathbb{R}^{2\times 2}$ is an invertible *lateral magnification matrix* and the *image function* $q_{z_0} : \mathbb{R}^2 \to \mathbb{R}$ describes the image of an object in the detector space when that object is located at $(0, 0, z_0)$ in the object space. Note that the subscript θ is used in the left-hand side of (2.7) to include dependence on hyperparameters. Depending on the model considered and inference aims, the hyperparameter(s) of interest can be $(x_{0,1}, x_{0,2})$ if the object is static and/or z_0 if an out-of-focus molecule is considered.

Three types of image functions are considered. First of all, according to optical diffraction theory from [4], an in-focus point source (i.e. when $z_0 = 0$) will typically generate an image that follows the Airy profile, given by

$$q(x_1, x_2) = \frac{J_1^2\left(\frac{2\pi n_\alpha}{\lambda_e}\sqrt{x_1^2 + x_2^2}\right)}{\pi(x_1^2 + x_2^2)}, \quad (x_1, x_2) \in \mathbb{R}^2, \tag{2.8}$$

where n_α is the numerical aperture of the objective lens, λ_e is the emission wavelength of the molecule and $J_1(\cdot)$ represents the first order Bessel function of the first kind.

Often, to simplify the problem, the 2D Gaussian approximation to the Airy profile has been used instead (see [11, 57, 61, 56]):

$$q(x_1, x_2) = \frac{1}{2\pi\sigma_a^2}\exp\left[-\frac{x_1^2 + x_2^2}{2\sigma_a^2}\right], \quad (x_1, x_2) \in \mathbb{R}^2. \tag{2.9}$$

If the point source of interest is out of focus, then a 3D Born and Wolf model [4] is used instead:

(2.10)
$$q_{z_0}(x_1,x_2) = \frac{4\pi n_\alpha^2}{\lambda_e^2}\left|\int_0^1 J_0\left(\frac{2\pi n_\alpha}{\lambda_e}\sqrt{x_1^2+x_2^2}\rho\right)\exp\left(\frac{j\pi n_\alpha^2 z_0}{n_o\lambda_e}\rho^2\right)\rho d\rho\right|^2, \quad (x_1,x_2)\in\mathbb{R}^2,$$

where $z_0 \in \mathbb{R}$ is the location of the object on the optical axis, n_o is the refractive index of the objective lens immersion medium and $J_0(\cdot)$ is the zero-th order Bessel function of the first kind. Note that the Airy profile is a special case of the Born and Wolf model. Indeed, if the object is in focus, then $z_0 = 0$ on the optical axis and (2.8) and (2.10) coincide.

2.3. Photon detection times. Just like the photon detection locations, the times at which the photons arrive on the detector $\mathcal{Y}$ are random. More specifically, in [59, 10], the arrival of the photons on the detector, or *photon detection process*, can be modelled as a Poisson process. Let $N(t)$ be the number of photons detected at time $t \geq t_0$ for initial time $t_0 \in \mathbb{R}$ and let $\lambda(t)$ be the *photon detection rate*, representing the rate at which the photons emitted by the object hit the detector at any given time t. For example, the detection rate of an object that has high photostability will simply be constant, while an exponentially decaying $\lambda(t)$ can indicate that the object image is photobleaching, or fading over time. The arrival times of the photons on the detector $\mathcal{Y}$ are denoted $t_1, t_2, \ldots$ where t_i denotes the arrival time of the i-th photon.

2.4. The observed data. Let $n_p = N(T) - N(t_0)$ be the number of photons detected in the interval $[t_0, T]$. We have now established the two aspects of the data that can be observed in a basic optical system during this interval, namely the detection times $t_1, t_2, \ldots, t_{n_p}$ of photons and the location of those detected photons $Y_{t_1}, Y_{t_2}, \ldots, Y_{t_{n_p}}$ on the detector $\mathcal{Y}$. Assume that, conditionally on the current object location X_{t_i}, the location of the i-th detected photon Y_{t_i} at time t_i is independent of the previous locations and time points of the detected photons, i.e. for $x_{t_i} \in \mathcal{X}$,

(2.11)
$$p_\theta(y_{t_i}|x_{t_i}, y_{t_{i-1}}, \ldots, y_{t_0}) = p_\theta(y_{t_i}|x_{t_i}) =: g_\theta(y_{t_i}|x_{t_i}), \quad y_{t_i} \in \mathcal{Y},$$

where the density g_θ is the photon distribution profile from (2.7). This is a reasonable assumption, as at any given time, processes such as photon emission and image formation only depend on the state of the emitting fluorescent molecule at that time, and not on any prior event.

Example 2.2. Let the trajectory of a molecule be given by the SDE in Example 2.1 and simulated using the same parameters and for the same time interval. Let $\mathcal{Y}$ be a non-pixelated detector. Then, let the photon detection rate be constant such that the mean number of photons is 500, and the photon distribution profile be given by (2.7), where the magnification matrix $M = m\mathbb{I}_{2\times 2}$ with $m = 100$. The image functions for the Airy, 2D Gaussian and Born and Wolf profiles are given by (2.8), (2.9) and (2.10) respectively, where $n_\alpha = 1.4$, $\lambda_e = 0.52$ μm, $n_o = 1.515$, $\sigma_a^2 = 49 \times 10^{-4}$ μm^2 and $z_0 = 1$ μm. By simulating the detected photon locations based on the same molecule trajectory and according to these three models, we obtain the observed photon trajectories in Figure 3. Note that the parameters of the Airy and

2D Gaussian profiles have been chosen so that the Gaussian profile approximates the Airy profile.

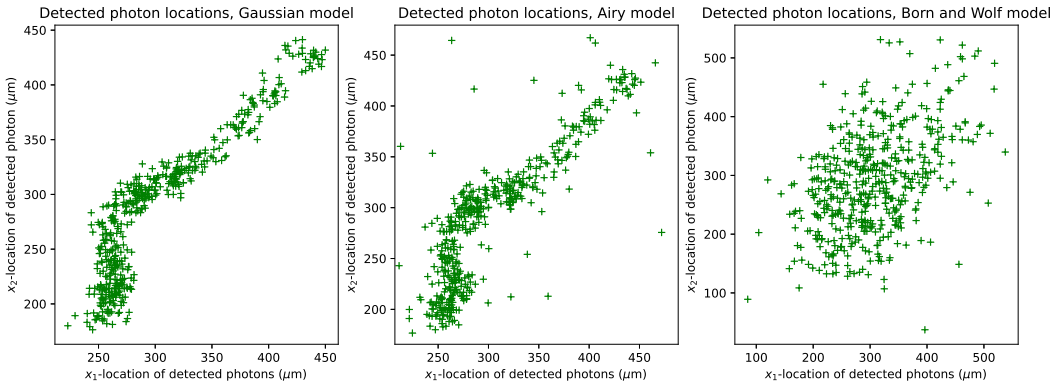


Figure 3: *Detected photon locations of a moving molecule with stochastic trajectory for the 2D Gaussian (left), Airy (middle) profiles and Born and Wolf model (right).*

3. The model as a state space model.

It is possible to reformulate this model as a state space model that takes into account the random arrival times of photons. This is achieved by discretising the time interval during which photons are recorded.

3.1. Reformulation.

For simplicity, we assume for the rest of this paper (unless stated otherwise) that the photon detection rate is constant, i.e. $\lambda(t) = \lambda \in [0, 1]$ for all $t \geq t_0$. First of all, let $X_t = (x_{t,1}, x_{t,2}) \in \mathcal{X}$ where $\mathcal{X} := \mathbb{R}^2$ denotes the state of the molecule at time $t \geq t_0$, which includes its location $x_{t,1:2}$ on the object plane. The location of the object on the optical axis is assumed to be constant and equal to the initial location parameter, i.e. z_0 for all $t \geq t_0$. The probability of recording an observation, i.e. detecting a photon in the small interval $(t, t + h]$ is

$$\mathbb{P}\left[N(t+h) - N(t) > 0\right] = \lambda h + o(t), \quad \lambda \in [0, 1],\, t \geq t_0.$$

Let t_i denote the arrival time of the i-th photon on a detector $\mathcal{Y}$ for $i = 1, 2, \ldots$ and $Y_{t_i} \in \mathcal{Y}$ be the location of the captured photon on the detector. Assume the location of a detected photon is distributed according to the probability density function

$$Y_{t_i} | \left(X_{t_i} = x\right) \sim g_\theta(\cdot|x), \quad i = 1, 2, \ldots,$$

where g_θ is the photon distribution profile given in (2.7). The recorded data in the time interval $[t_0, T]$, $0 \leq t_0 < T$ comprises of n_p observations with arrival times $t_0 < t_1 < \ldots < t_{n_p} \leq T$ and photon locations $y_{t_1}, \ldots, y_{t_{n_p}}$. The inference objective is to estimate the trajectory of the molecule $(X_t)_{t_0 \leq t \leq T}$ given data (t_i, y_{t_i}), $i = 1, \ldots, n_p$. As seen in subsection 2.1, the molecule evolves according to the probability density function

$$X_{t_{i+1}} | \left(X_{t_i} = x\right) \sim f^\theta_{t_i, t_{i+1}}(\,\cdot\,|x), \quad i = 1, 2, \ldots, n_p,$$

where θ denotes the model parameters and $f^{\theta}_{s,t}$ for $t > s \geq t_0$ is the homogeneous continuous-time Markov transition density given by the the Gaussian distribution in (2.5) for $d = 2$.

3.1.1. Non-constant photon detection rate. If the photon detection rate $\lambda(t)$ is not assumed to be constant, then we redefine the state of an object at time $t \geq t_0$ as $X_t = (x_{t,1}, x_{t,2}, \lambda_t) \in \mathcal{X}$ where $\mathcal{X} := \mathbb{R}^2 \times [0, 1]$. The state at time t now includes the location of the molecule $(x_{t,1}, x_{t,2})$ as well as the probability λ_t of detecting a photon it emits. The Markov transition density $p_\theta(x'|x)$ can be defined as follows

$$p_\theta(x_{t_{i+1}}|x_{t_i}) = f^{\theta}_{t_i,t_{i+1}}(x_{t_{i+1},1:2}|x_{t_i,1:2}) l_\theta(\lambda_{t_{i+1}}|\lambda_{t_i}), \quad x_{t_{i+1}}, x_{t_i} \in \mathcal{X},$$

where t_i and t_{i+1} denote the arrival times of the i-th and $(i+1)$-th photons, respectively, $f^{\theta}_{t_i,t_{i+1}}$ is the Markov transition density for the object location defined above and l_θ is the Markov transition density for the photon detection rate.

3.2. Time discretisation. Let $(t_1, y_{t_1}), \ldots, (t_{n_p}, y_{t_{n_p}})$ be a realisation of the photon arrival times and locations observed in the time interval $[t_0, T]$. Setting $t_0 := 0$ for convenience, we adopt a discrete time formulation where $[0, T]$ is divided into segments of length Δ. Let $x_k \in \mathcal{X}$ denote the state of the molecule at time $t = (k-1)\Delta$ where $k = 1, \ldots, n$ for $n := \lceil T/\Delta \rceil$. We assume the discretisation is fine enough so that an interval $(k\Delta, k\Delta + \Delta]$ contains at most one arrival time t_i. Then, for $k = 1, \ldots, n$, let

$$y_k = \begin{cases} \emptyset, & \text{if} \quad t_i \notin (k\Delta - \Delta, k\Delta], \quad \forall i = 0, 1, \ldots, n_p, \\ y_{t_i}, & \text{if} \quad t_i \in (k\Delta - \Delta, k\Delta], \end{cases}$$

where $y_{t_i} \in \mathcal{Y}$ denotes the location of the i-th detected photon on the detector $\mathcal{Y}$. The vector y_k is assigned $\emptyset$ to indicate the absence of an observation in the corresponding interval. See section SM1 for details on why the time discretisation is a valid approximation of the Poisson process. If $x = (x_1, x_2, \lambda) \in \mathcal{X}$, let

$$G^{\theta}_k(x) = \begin{cases} 1 - \Delta\lambda, & \text{if } y_k = \emptyset, \\ \lambda g_\theta(y_{t_i}|x_{1:2}), & \text{if } y_k = y_{t_i}, \end{cases} \tag{3.1}$$

where g_θ is the photon distribution profile (2.7), then $G^{\theta}_k(x)$ is the so called potential function. The potential $G^{\theta}_k(x)$ plays the role of the likelihood in Bayesian estimation problems. In the above context, the expression for $y_k = \emptyset$ corresponds to the probability of no photon being observed during that time interval. When a photon is observed in the interval, with observation time t_i and observation location y_{t_i} on the detector, the expression for G^{θ}_k is the product of the probability $\Delta\lambda$ of receiving one photon, with the uniform probability density $1/\Delta$ for the arrival time t_i in that interval and the density of the location of the observation given that the molecule is situated at $x_{1:2}$ in the object space (the Δ terms then cancel out).

For $k = 1, \ldots, n$, the probability density function of X_{k+1} given the previous state X_k is $f^{\theta}_{\Delta}(x_{k+1}|x_k) := f^{\theta}_{k\Delta,k\Delta+\Delta}(x_{k+1}|x_k)$ from (2.5), thus transforming (2.2) into

$$X_{k+1} = \Phi_\Delta X_k + a_\Delta + W_x, \quad W_x \sim \mathcal{N}(0, R_\Delta),$$

where $\Phi_\Delta = \Phi(k\Delta, k\Delta + \Delta)$ is now constant and similarly for a_Δ and R_Δ.

To summarise, $(X_k)_{k=1}^\infty$ and $(Y_k)_{k=1}^\infty$ are $\mathcal{X}$- and $\mathcal{Y} \cup \emptyset$-valued stochastic processes where the molecule trajectory in the object space $(X_k)_{k=1}^\infty$ corresponds to the unobserved latent Markov process with Markov transition density $f_\Delta^\theta(x'|x)$ and initial density $\nu_\theta(x)$, and the photon detection locations (or lack of) $(Y_k)_{k=1}^\infty$ represent the observed process with conditional density or potential function $G_k^\theta(x)$, i.e.

$$X_1 \sim \nu_\theta(\cdot), \quad X_{k+1} | \, (X_k = x) \sim f_\Delta^\theta(\cdot|x), \tag{3.2}$$

$$Y_k | \, (X_k = x) \sim G_k^\theta(x), \quad k = 1, 2, \ldots. \tag{3.3}$$

Note that if the object is static, so that the drift and diffusion coefficient in (2.1) are zero, the model simplifies from a state space model to a basic inference problem with independent observations. The observed process is still described by (3.3) but the location of the object x_0 becomes part of the hyperparameters.

4. Parameter inference.

4.1. Inference aim. Now that we have formulated the problem in (3.2) and (3.3) as a state space model, the first aim is going to be to estimate the posterior probability density function of $X_{1:n} := \{X_1, \ldots, X_n\}$, $n \in \mathbb{N}$, given the observations $Y_{1:n}$, also known as the *joint smoothing distribution*, which is given by

$$p_\theta(x_{1:n}|y_{1:n}) = \frac{p_\theta(x_{1:n}, y_{1:n})}{p_\theta(y_{1:n})}, \tag{4.1}$$

where the numerator represents the *joint density*

$$p_\theta(x_{1:n}, y_{1:n}) = \nu_\theta(x_1) \prod_{k=2}^n f_\Delta^\theta(x_k|x_{k-1}) \prod_{k=1}^n G_k^\theta(x_k), \tag{4.2}$$

where $\nu_\theta(x_1)$ is the initial distribution of X_1, and the denominator represents the *marginal likelihood* of the observed data

$$p_\theta(y_{1:n}) = \int_{\mathcal{X}^n} p_\theta(x_{1:n}, y_{1:n}) \mathrm{d}x_{1:n}. \tag{4.3}$$

Estimating $p_\theta(x_{1:n}|y_{1:n})$ is what allows the molecule to be tracked and is done using a particle filter. The second aim is to obtain particle approximations of smoothed additive functionals, which in turn will allow for estimation of the score and OIM for of the hyperparameters θ, as well as other applications such as ML estimation of said hyperparameters via gradient ascent and Expectation-Maximization (EM). Finally, the third aim is to use the estimates of the score and OIM of the hyperparameters to obtain an approximation of their FIM.

4.2. Tracking the molecule using a particle filter. The particle approximation of the marginal posterior of $X_1, \ldots, X_n$ defined in (4.1) is given by

$$\hat{p}(x_{1:n}|y_{1:n}) = \sum_{i=1}^N \omega_n^{(i)} \delta_{X_{1:n}^{(i)}}(x_{1:n}),$$

where $X_{1:n}^{(1:N)}$ are the particles, $\omega_n^{(1:N)}$ their corresponding normalised *importance weights*, i.e. $\sum_{i=1}^N \omega_n^{(i)} = 1$ and $\delta_{v_0}(v)$ denotes the dirac delta mass located at v_0. To obtain this particle approximation, we employ sequential Monte Carlo (SMC) methods in the form of a *particle filter* (see [9, 18, 19, 12] for comprehensive reviews of SMC methods). There is flexibility in the specific choice of particle filter, but the general form they take follows three key steps, namely *resample*→*propagate*→*weight*. For $k = 2, \ldots, n$, the *resampling* step avoids weight degeneracy [21, 34] and consists of drawing indices $\iota_{k-1}^{(1:N)}$ with probabilities corresponding to the normalised weights $\omega_{k-1}^{(1:N)}$, then, depending on the resampling algorithm considered, resetting the weights accordingly, e.g. $\omega_{k-1}^{(1:N)} := \frac{1}{N}$. The propagation and weighting steps consist of advancing the (resampled) particle population $(X_{k-1}^{\iota_{k-1}^{(1:N)}}, \omega_{k-1}^{(1:N)})$ forward in time via the *proposal density* $\eta_k(x_k|x_{k-1})$ (*propagate*) and updating the importance weights (*weight*) as follows (see section SM2 for more details.):

$$\omega_k^{(i)} = \frac{\omega_{k-1}^{(i)} \tilde{w}\left(X_{k-1}^{(i)}, X_k^{(i)}\right)}{\sum_{j=1}^N \omega_{k-1}^{(j)} \tilde{w}\left(X_{k-1}^{(j)}, X_k^{(j)}\right)},$$

where $\tilde{w}(x_{k-1}, x_k)$ is known as the *incremental weight* and is given by

$$\tilde{w}(x_{k-1}, x_k) = \frac{G_k^\theta(x_k) f_\Delta^\theta(x_k|x_{k-1})}{\eta_k(x_k|x_{k-1})}.$$

The proposal density is user-defined. For example, if $\eta_t(x_k|x_{k-1}) = f_\Delta^\theta(x_k|x_{k-1})$, the particle filter becomes the well-known *bootstrap filter*, introduced in [28] and the computation of the incremental weights simplifies to $\tilde{w}(x_k) = G_t^\theta(x_k)$. A generic particle filter is summarised in Algorithm SM2.1.

Given weighted particle sample $(X_{k-1}^{(1:N)}, \omega_{k-1}^{(1:N)})$ at step k, we denote an iteration of running the particle filter (steps 4-6 of Algorithm SM2.1) as

$$\left(X_k^{(1:N)}, \omega_k^{(1:N)}\right) := \mathsf{PF}_\Delta\left(X_{k-1}^{(1:N)}, \omega_{k-1}^{(1:N)}\right).$$

For this particular problem, we must also take into account the missing observations introduced by the time discretisation. Since a lack of observation does not bring any new information, it suffices to only run the particle filter at segments which contain an observation. A typical iteration of this approach is summarised in Algorithm 4.1. The interval counter is initialised at $c_0 := 1$ and counts the number of discrete intervals since (and including) the last observation. An example of particle filtering for stochastically moving molecules observed through the 2D Gaussian, Airy and Born and Wolf models is available in Example SM2.1.

4.3. Particle approximations of expectations of additive functionals. The second inference aim is to obtain estimates of the score and observed information matrix (OIM) for the hyperparameters θ. To achieve these aims, we make use of smoothed additive functionals. Assume that there exists a real-valued function S_k^θ, $k \geq 0$ such that it is an *additive functional*

Algorithm 4.1 Particle filter for SDE with missing observations

Input: weighted particle sample $\left(X_{k-1}^{(1:N)}, \omega_{k-1}^{(1:N)}\right)$ and interval counter c_{k-1} at step $k-1$.

1: **if** $y_k = \emptyset$ **then**

2: $c_k := c_{k-1} + 1$

3: Do not run the particle filter

$$\left(X_k^{(1:N)}, \omega_k^{(1:N)}\right) := \left(X_{k-1}^{(1:N)}, \omega_{k-1}^{(1:N)}\right).$$

4: **else**

5: Run the particle filter with updated interval length, i.e.

$$\left(X_k^{(1:N)}, \omega_k^{(1:N)}\right) := \mathsf{PF}_{c_k\Delta}\left(X_{k-1}^{(1:N)}, \omega_{k-1}^{(1:N)}\right).$$

6: $c_k := 1$

7: **end if**

Output: updated particle sample $\left(X_k^{(1:N)}, \omega_k^{(1:N)}\right)$.

given by

$$S_k^\theta(x_{1:k}) = \sum_{j=1}^{k} s_j^\theta(x_{j-1}, x_j), \tag{4.4}$$

where $s_1^\theta(x_0, x_1) := s_1^\theta(x_1)$ and $\left\{s_k^\theta\right\}_{k\geq 0}$ is a sequence of *sufficient statistics* which may depend on the value of the observations $y_{0:k}$. The main aim is to compute the posterior or *smoothing expectation*, given by

$$\mathcal{S}_k(\theta) := \mathbb{E}_\theta\left[S_k^\theta\left(X_{1:k}\right)|y_{1:k}\right] = \int_{\mathcal{X}} S_k^\theta\left(x_{1:k}\right) p_\theta(x_{1:k}|y_{1:k}) \mathrm{d}x_{1:k}. \tag{4.5}$$

If the model in question is linear and Gaussian or the state space $\mathcal{X}$ is finite, then the expectation $\mathcal{S}_k(\theta)$ can be computed exactly by recursion. However, this is not the case if the Airy or Born and Wolf profiles are used to describe photon distribution. In this case, SMC methods can again be employed to approximate the expectation as follows

$$\hat{\mathcal{S}}_k(\theta) := \sum_{i=1}^{N} \omega_k^{(i)} S_k^\theta\left(X_{1:k}^{(i)}\right),$$

where the weighted sample $(X_{1:k}^{(1:N)}, \omega_k^{(1:N)})$ is a particle approximation of the joint smoothing distribution $p_\theta(x_{1:k}|y_{1:k})$ obtained using a particle filter.

A simple way of estimating the smoothing expectation $\mathcal{S}_n(\theta)$ for a set of n observations $y_{1:n}$ is to run the desired particle filter in a 'forward pass' through the whole data to obtain the particle approximation $(X_n^{(1:N)}, \omega_n^{(1:N)})$ at the final step n, followed then by a 'backward

smoothing' pass through the data, starting from the latest sample y_n. This is the case of algorithms such as the fixed-lag smoother by [33, 44, 45], forward-filtering backward smoothing (FFBSm) by [20, 30, 32] and forward-filtering backward simulation (FFBSi) by [26]. However, if one wishes to avoid multiple passes through the data, it is also possible to take advantage of the form of the additive functional in (4.4) to estimate $\mathcal{S}_k(\theta)$ in an online or 'forward-only' fashion, as proposed in [16] and further developed in [46]. Introducing the *auxiliary function*

$$T_k^\theta(x_k) := \int_{\mathcal{X}^{k-1}} S_k^\theta(x_{1:k}) p_\theta(x_{1:k-1}|y_{1:k-1}, x_k) \mathrm{d}x_{1:k-1},$$

the following recursion is then created:

$$T_k^\theta(x_k) = \int_{\mathcal{X}} \left[T_{k-1}^\theta(x_{k-1}) + s_k^\theta(x_{k-1}, x_k) \right] p_\theta(x_{k-1}|y_{1:k-1}, x_k) \mathrm{d}x_{k-1}, \tag{4.6}$$

where $T_0^\theta := 0$ and its particle approximation given the weighted sample $(X_{1:k}^{(1:N)}, \omega_k^{(1:N)})$ and previous state particle approximation $\hat{T}_{k-1}^\theta(X_{k-1}^{(1:N)})$ is given by

$$\hat{T}_k^\theta\left(X_k^{(i)}\right) = \sum_{j=1}^N \Psi_k^\theta(i,j) \left[\hat{T}_{k-1}^\theta\left(X_{k-1}^{(j)}\right) + s_k^\theta\left(X_{k-1}^{(j)}, X_k^{(i)}\right) \right] \tag{4.7}$$

for all $i \in \{1, \dots, N\}$, and where

$$\Psi_k^\theta(i,j) := \frac{\omega_{k-1}^{(j)} f_\Delta^\theta\left(X_k^{(i)}|X_{k-1}^{(j)}\right)}{\sum_{j=1}^N \omega_{k-1}^{(j)} f_\Delta^\theta\left(X_k^{(i)}|X_{k-1}^{(j)}\right)}. \tag{4.8}$$

Finally, using the recursion on the auxiliary function T_k^θ, the smoothing expectation in (4.5) can be rewritten as

$$\mathcal{S}_k(\theta) = \int_{\mathcal{X}} T_k^\theta(x_k) p_\theta(x_k|y_{1:k}) \mathrm{d}x_k, \tag{4.9}$$

and its particle approximation is

$$\hat{\mathcal{S}}_k(\theta) = \sum_{i=1}^N \omega_k^{(i)} \hat{T}_k^\theta\left(X_k^{(i)}\right). \tag{4.10}$$

This algorithm is known as *Forward smoothing SMC* (SMC-FS) and is summarised in the context of our experiments in Algorithm 4.2.

4.4. Estimation of the score and observed information matrix (OIM). The score and OIM have important applications to ML estimation, e.g. see [35, 47]. They can also be instrumental in assessing the performance of such an estimator, either directly, as argued by [23], or as tools to estimate the FIM when the latter cannot be computed exactly, as we will see in this section. We aim to compute, recursively in time, the score vector $\mathcal{G}_k(\theta) := \nabla \log p_\theta(y_{1:k})$ and OIM $\mathcal{H}_k(\theta) := -\nabla^2 \log p_\theta(y_{1:k})$ where $p_\theta(y_{1:k})$ denotes the marginal likelihood at step $1 \leq k \leq n$ defined in (4.3), ∇ denotes the gradient and ∇^2 the Hessian.

Algorithm 4.2 Forward smoothing SMC (SMC-FS)

Where (i) *or* (j) *appears, the operation is performed for all* $i, j \in \{1, \ldots, N\}$.
At $k = 1$,

Initialise the particle filter to obtain the weighted particle sample $\left(X_1^{(1:N)}, \omega_1^{(1:N)}\right)$.
Initialise the interval counter $c_0 := 1$.
Set $\hat{T}_1^\theta\left(X_1^{(i)}\right) := 0$.
for $k = 2, \ldots, n$ **do**
  if $y_k = \emptyset$ **then**
    $c_k := c_{k-1} + 1$
  else
    Use the particle filter to update the weighted particle sample, i.e.

$$\left(X_k^{(1:N)}, \omega_k^{(1:N)}\right) := \mathsf{PF}_{c_k\Delta}\left(X_{k-1}^{(1:N)}, \omega_{k-1}^{(1:N)}\right).$$

    Evaluate

$$\Psi_k^\theta(i,j) := \frac{\omega_{k-1}^{(j)} f_{c_k\Delta}^\theta\left(X_k^{(i)} | X_{k-1}^{(j)}\right)}{\sum_{j=1}^N \omega_{k-1}^{(j)} f_{c_k\Delta}^\theta\left(X_k^{(i)} | X_{k-1}^{(j)}\right)}.$$

    Update the auxiliary function estimate

$$\hat{T}_k^\theta\left(X_k^{(i)}\right) = \sum_{j=1}^N \Psi_k^\theta(i,j)\left[\hat{T}_{k-1}^\theta\left(X_{k-1}^{(j)}\right) + s_k^\theta\left(X_{k-1}^{(j)}, X_k^{(i)}\right)\right].$$

    Update the smoothing expectation estimate

$$\hat{\mathcal{S}}_k(\theta) = \sum_{i=1}^N \omega_k^{(i)} \hat{T}_k^\theta\left(X_k^{(i)}\right).$$

    Reset the interval counter $c_k := 1$.
  end if
end for
Output: smoothing expectation estimate $\hat{\mathcal{S}}_n$.

4.4.1. Establishing the sufficient statistics. The key to obtaining the particle approximation (4.10) of a smoothing expectation (4.9) of interest is to establish the relevant additive functionals and sufficient statistics. First of all, assume that the regularity conditions allowing for differentiation and integration to be switched around in expressions are satisfied. Let us

establish the Fisher and Louis identities for the score and OIM, respectively, from [9, 18]:

$$\mathcal{G}_k(\theta) = \int_{\mathcal{X}} \nabla \log p_\theta(x_k, y_{1:k}) p_\theta(x_k|y_{1:k}) \mathrm{d}x_k, \tag{4.11}$$

$$\mathcal{H}_k(\theta) = \nabla \log p_\theta(y_{1:k}) \nabla \log p_\theta(y_{1:k})^\intercal - \frac{\nabla^2 p_\theta(y_{1:k})}{p_\theta(y_{1:k})},$$

where

$$\frac{\nabla^2 p_\theta(y_{1:k})}{p_\theta(y_{1:k})} = \int_{\mathcal{X}} \nabla \log p_\theta(x_k, y_{1:k}) \nabla \log p_\theta(x_k, y_{1:k})^\intercal p_\theta(x_k|y_{1:k}) \mathrm{d}x_k$$

$$+ \int_{\mathcal{X}} \nabla^2 \log p_\theta(x_k, y_{1:k}) p_\theta(x_k|y_{1:k}) \mathrm{d}x_k, \tag{4.12}$$

and note that (4.11) and (4.12) can be rewritten as

$$\nabla \log p_\theta(y_{1:k}) = \mathbb{E}\left[\alpha_k^\theta(X_k)|y_{1:k}\right], \tag{4.13}$$

$$\frac{\nabla^2 p_\theta(y_{1:k})}{p_\theta(y_{1:k})} = \mathbb{E}\left[\alpha_k^\theta(X_k)\alpha_k^\theta(X_k)^\intercal|y_{1:k}\right] + \mathbb{E}\left[\beta_k^\theta(X_k)|y_{1:k}\right], \tag{4.14}$$

where the expectations here are with respect to the density $p(x_k|y_{1:k})$, and correspond to the smoothing expectations in (4.9), with the functions $\alpha_k^\theta(x_k) := \nabla \log p_\theta(x_k, y_{1:k})$ and $\beta_k^\theta := \nabla^2 \log p_\theta(x_k, y_{1:k})$ acting as the auxiliary functions of interest. A recursion for α_k^θ and β_k^θ is straightforward to obtain, more details in [47]. For α_k^θ and β_k^θ, (4.6) becomes

$$\alpha_k^\theta(x_k) = \int_{\mathcal{X}} \left[\alpha_{k-1}^\theta(x_{k-1}) + s_k^\alpha(x_{k-1}, x_k)\right] p_\theta(x_{k-1}|y_{1:k-1}, x_k) \mathrm{d}x_{k-1},$$

$$\beta_k^\theta(x_k) = \int_{\mathcal{X}} \left[\beta_{k-1}^\theta(x_{k-1}) + s_k^\beta(x_{k-1}, x_k)\right] p_\theta(x_{k-1}|y_{1:k-1}, x_k) \mathrm{d}x_{k-1} - \alpha_k^\theta(x_k)\alpha_k^\theta(x_k)^\intercal,$$

where the sufficient statistics are given by

$$s_k^\alpha(x_{k-1}, x_k) := \nabla \log G_k^\theta(x_k) + \nabla \log f_\Delta^\theta(x_k|x_{k-1}), \tag{4.15}$$

$$s_k^\beta(x_{k-1}, x_k) := \left[\alpha_{k-1}^\theta(x_{k-1}) + s_k^\alpha(x_{k-1}, x_k)\right]\left[\alpha_{k-1}^\theta(x_{k-1}) + s_k^\alpha(x_{k-1}, x_k)\right]^\intercal$$

$$+ \nabla^2 \log G_k^\theta(x_k) + \nabla^2 \log f_\Delta^\theta(x_k|x_{k-1}). \tag{4.16}$$

Finally, to approximate the score and OIM, adapt the particle approximation in (4.7) to the recursions in (4.15) and (4.16) to obtain the score estimate, given by a weighted sum (4.10) approximating the smoothing expectation (4.13), i.e.

$$\hat{\mathcal{G}}_k(\theta) = \sum_{i=1}^N \omega_k^{(i)} \hat{\alpha}_k^\theta\left(X_k^{(i)}\right)$$

and the OIM estimate

$$\hat{\mathcal{H}}_k(\theta) = \hat{\mathcal{G}}_k(\theta)\hat{\mathcal{G}}_k(\theta)^\intercal - \sum_{i=1}^N \omega_k^{(i)} \left[\hat{\alpha}_k^\theta\left(X_k^{(i)}\right)\hat{\alpha}_k^\theta\left(X_k^{(i)}\right)^\intercal + \hat{\beta}_k^\theta\left(X_k^{(i)}\right)\right],$$

where the weighted sum is the particle approximation (4.10) of the smoothed expectation in (4.14). In Example 4.1, we apply this framework to a possible application of the single-molecule tracking model. We focus for now on the case where the photon distribution is described by the Airy or 2D Gaussian profile.

Example 4.1. Let the trajectory of a molecule be given by the following SDE

$$\mathrm{d}X_t = b\mathbb{I}_{2\times 2}X_t\mathrm{d}t + \sqrt{2}\sigma\mathrm{d}B_t,$$

where in the drift term, $b \neq 0$, in the diffusion term, $\sigma > 0$, and $(\mathrm{d}B_t)_{t_0\leq t\leq T}$ is a Wiener process. Let the photon detection process be described by the Airy or 2D Gaussian profile. Then, the parameters of interest are $\theta = (\sigma^2, b)$. Recall from subsection 3.2 and Example 2.1 that the solution to the SDE can be written as

$$X_k = e^{\Delta b}X_{k-1} + W_x, \quad W_x \sim \mathcal{N}\left(0, \frac{\sigma^2}{b}\left(e^{2\Delta b} - 1\right)\mathbb{I}_{2\times 2}\right), \tag{4.17}$$

and since the potential function G_k does not depend on θ in this case, it can be dropped from (4.15) and (4.16) and the components of the sufficient statistic $s_k^\alpha(x_{k-1}, x_k)$ for the additive functional α_k^θ are

$$\frac{\partial}{\partial\sigma^2}\log f_\Delta^\theta(x_k|x_{k-1}) = -\frac{1}{\sigma^2} + \frac{b\left\|x_k - e^{\Delta b}x_{k-1}\right\|^2}{2\sigma^4\left(e^{2\Delta b}-1\right)},$$

$$\begin{aligned}\frac{\partial}{\partial b}\log f_\Delta^\theta(x_k|x_{k-1}) = \frac{1}{b} - \frac{2\Delta e^{2\Delta b}}{(e^{2\Delta b}-1)} - \frac{\left\|x_k - e^{\Delta b}x_{k-1}\right\|^2}{2\sigma^2(e^{2\Delta b}-1)} \\ + \frac{\Delta b e^{\Delta b}(x_k - e^{\Delta b}x_{k-1})^\intercal x_{k-1}}{\sigma^2(e^{2\Delta b}-1)} + \frac{\left\|x_k - e^{\Delta b}x_{k-1}\right\|^2\Delta b e^{2\Delta b}}{\sigma^2(e^{2\Delta b}-1)^2}.\end{aligned}$$

The components of the sufficient statistic $s_k^\beta(x_{k-1}, x_k)$ for β_k^θ are given in section SM3. Note that these derivatives can be evaluated for any value of Δ, and it is therefore possible to adapt them in order to only compute sufficient statistics when an observation is recorded as in Algorithm 4.1. This is reflected in Algorithm 4.2.

4.5. Estimating the Fisher information matrix (FIM). The *Fisher information matrix* (FIM) is widely used in estimation problems as an indicator of the performance of a given estimator. Indeed, it is a key element of the Cramér-Rao inequality, or *Cramér-Rao Lower Bound* (CRLB) derived by [13, 51, 25, 14], which states that for an unbiased estimate $\hat{\theta}$ of the parameter θ, its covariance has lower bound

$$\mathrm{Cov}(\hat{\theta}) \succeq \mathcal{I}_n(\theta)^{-1},$$

where given matrices A and B, the inequality $A \succeq B$ indicates that $A - B$ is a positive semi-definite matrix, and $\mathcal{I}_n(\theta)$ denotes the FIM in a random sample $Y_1, \ldots, Y_n$ of size n [15], defined as

$$\mathcal{I}_n(\theta) = \mathbb{E}_\theta\left[\nabla\log p_\theta(Y_{1:n})\nabla\log p_\theta(Y_{1:n})^\intercal\right] \tag{4.18}$$

$$= \mathbb{E}_\theta\left[-\nabla^2\log p_\theta(Y_{1:n})\right], \tag{4.19}$$

where the second equality is proven in [22]. When the expectations in (4.18) and (4.19) are intractable – which is the case when the Airy profile is used to describe the photon detection locations in the single-molecule tracking model – there are several ways one can go about estimating the FIM.

4.5.1. Estimating the FIM for a single large sample using the OIM. Firstly, note that from (4.19), the relationship between the FIM and OIM is simply

$$\mathcal{I}_n(\theta) = \mathbb{E}_\theta \left[\mathcal{H}_n(\theta) \right], \tag{4.20}$$

where $\mathcal{H}_n(\theta) = -\nabla^2 \log p_\theta(y_{1:n})$ denotes the OIM. Then, for a general state space model, in [3], it was proven that under mild assumptions,

$$\frac{1}{n}\mathcal{H}_n(\theta) \to \mathcal{I}(\theta) \quad \text{as } n \to \infty,$$

where $\mathcal{I}(\theta)$ is the asymptotic FIM. See [29] for the corresponding result for multiple targets. So for a large enough sample size n, i.e. if the interval during which the molecule(s) of interest are observed is long enough, the OIM and FIM can be used interchangeably, i.e. for $n \gg 1$,

$$\mathcal{H}_n(\theta) \approx \mathcal{I}_n(\theta). \tag{4.21}$$

See Figure 4 for an illustration. Therefore, the first way of estimating the *asymptotic* FIM in the single-molecule tracking model is simply to obtain the OIM for a large sample size. For more details on the OIM as an estimate of the FIM, see [17].

4.5.2. Estimating the FIM using the mean outer product of the score. If the molecule(s) of interest are only observed for a short interval, then the size n of the sample of interest is not large enough to estimate the FIM using the OIM. It is then also possible to instead obtain a particle approximation of the expectation in (4.18) using the score as follows: generate D datasets $y_{1:n}^{(1:D)}$ of (smaller) size n where $y_{1:n}^{(d)} := \{y_1^{(d)}, \dots, y_n^{(d)}\}$, and according to the same parameters θ. The outer product of the score can then be used in the estimate of the FIM as follows:

$$\hat{\mathcal{I}}_n(\theta) = \frac{1}{D}\sum_{j=1}^{D} \mathcal{G}_n^{(d)}(\theta)\mathcal{G}_n^{(d)}(\theta)^\intercal, \tag{4.22}$$

where for $d = 1, \dots, D$, the vector $\mathcal{G}_n^{(d)}(\theta) := \nabla \log p_\theta(y_{1:n}^{(d)})$ is the score for the d-th dataset of size n. An advantage of this approach is that the OIM need not be computed.

4.5.3. Estimating the FIM using the mean OIM. When multiple datasets are available, the OIM can also similarly be averaged over D datasets to estimate the FIM as follows:

$$\hat{\mathcal{I}}_n(\theta) = \frac{1}{D}\sum_{j=1}^{D} \mathcal{H}_n^{(d)}(\theta). \tag{4.23}$$

This third approach is the Monte Carlo estimator of the expectation in (4.20), and can be seen as averaging the first estimation method in (4.21).

Now that the various methods for estimating the FIM have been established, it can be used in an experimental design setting to plan experiments with the aim of returning the most accurate parameter estimates. See section SM4 for details on how ML estimates can similarly be obtained via EM and gradient ascent methods with the use of smoothed additive functionals and SMC-FS.

Example 4.2. To verify these approaches to estimate the FIM, consider the straightforward special case of estimating the FIM for the location $x_0 = (x_{0,1}, x_{0,2})$ parameters of a static molecule emitting photons at a constant rate. In [40, 10], the analytical expression for the FIM is derived for the Airy profile, and its diagonal components given observations $y_{1:n}$ are given by

$$\mathcal{I}_n^{\text{Airy}}(x_{0,1}) = \mathcal{I}_n^{\text{Airy}}(x_{0,2}) = N_{phot}\alpha^2,$$

where $\alpha = \frac{2\pi n_a}{\lambda_e}$, N_{phot} denotes the expected photon count, and $\mathcal{I}^{\text{Airy}}(x_{0,i})$ denotes the (i,i)-th element of the FIM, corresponding to parameter component x_i, for the Airy profile. As mentioned in subsection 3.2, having a static molecule simplifies the model. Since we have independent data, the true values of score $\mathcal{G}$ and OIM $\mathcal{H}$ can be derived as follows. Given a set of observations $y_{1:n}$ distributed according to the Airy profile,

$$\mathcal{G}_n^{\text{Airy}}(x_0) = \sum_{k=1}^{n} \gamma_k (M^{-1}y_k - x_0)\mathbb{1}_{y_k \neq \emptyset},$$

$$\mathcal{H}_n^{\text{Airy}}(x_0) = \sum_{k=1}^{n} \left(\chi_k (M^{-1}y_k - x_0)(M^{-1}y_k - x_0)^\intercal + \gamma_k \mathbb{I}_{2\times 2}\right) \mathbb{1}_{y_k \neq \emptyset},$$

where

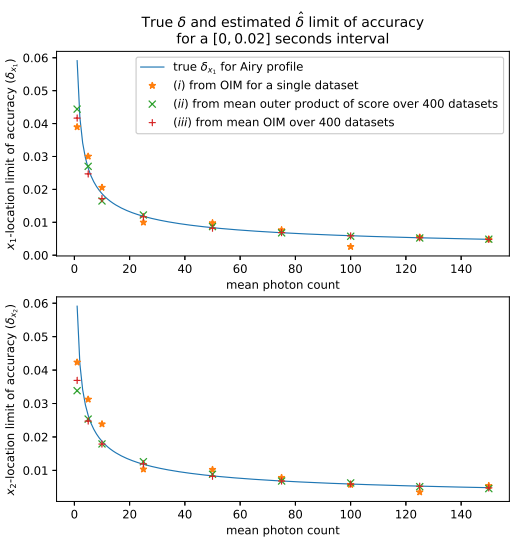
$$\gamma_k = \frac{2\alpha}{r}\frac{J_2(\alpha r_k)}{J_1(\alpha r_k)}, \qquad \chi_k = -\frac{2\alpha^2}{r_k^2}\left[\frac{J_3(\alpha r_k)}{J_1(\alpha r_k)} - \frac{J_2^2(\alpha r_k)}{J_1^2(\alpha r_k)}\right],$$

and $r_k = \sqrt{(M^{-1}y_k - x_0)^\intercal (M^{-1}y_k - x_0)}$. See supplementary material SM5 for the full derivation.

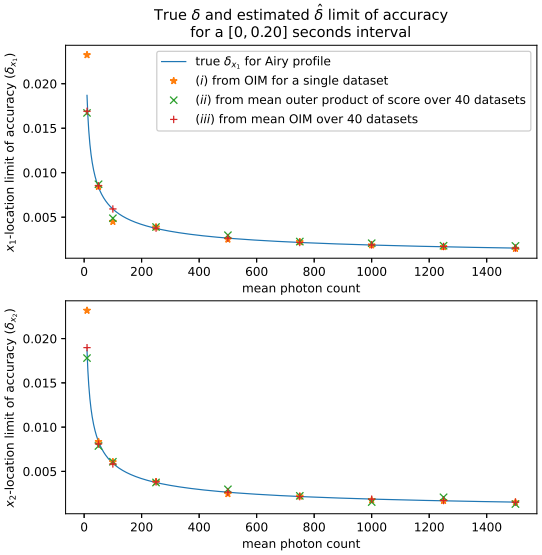
Using the same settings as in Example 2.2, we simulate $D_l = 40$ 'large' datasets according to the Airy profile consisting of observations obtained during the interval $[0, 0.2]$ seconds. We also simulate $D_s = 400$ 'short' datasets consisting of observations obtained during the shorter interval $[0, 0.02]$ seconds. The score and OIM are obtained for all datasets and the FIM for the large and short datasets is estimated in three ways: *(i)* using the OIM returned from a single dataset selected at random (4.21), *(ii)* using the mean outer product of the score (4.22) over all datasets and *(iii)* using the mean OIM across all datasets (4.23). Finally, the square root of the CRLB, also known as the (fundamental) *limit of accuracy* and defined as

$$\delta_\vartheta = \sqrt{CRLB_\vartheta}$$

for parameter ϑ is obtained. This is repeated for various expected photon counts in order to compare the evolution of the estimated limit of accuracy as the expected number of photons



(a) Many short datasets, Airy profile



(b) Few long datasets, Airy profile

Figure 4: *True and estimated limit of accuracy for mean photon counts ranging from* ***(a)*** *1 to 150* ***(b)*** *10 to 1500. The limit of accuracy is estimated for the location parameters* (x_1, x_2) *of a static in-focus molecule. The estimates are obtained by taking the square root of the inverse of the FIM, obtained for* ***(a)*** *400 'short'* ***(b)*** *40 'long' simulated datasets using approaches (i)* $\star$*, (ii)* $\times$ *and (iii)* $+$ *for comparison purposes. To generate each dataset, the photon detection times are simulated according to a Poisson process with constant rate corresponding to the expected mean photon count for* ***(a)*** $[0, 0.02]$ ***(b)*** $[0, 0.2]$ *seconds and the intervals are discretised. The photon detection locations are generated according to the Airy profile, with parameters as in Example 2.2. The true limit of accuracy (blue solid line) is also computed as it is available analytically [42]. Estimates of the limit of accuracy based on a single dataset (approach (i)) are more accurate when the dataset is long, while taking the mean outer product of the score over all datasets (approach (ii)) yields more accurate estimates for a large number of short datasets. Approach (iii) provides a good balance between the two. In general, estimates of the limit of accuracy are relatively poor for very low mean photon counts but quickly improve as it increases.*

increases to the true limit of accuracy obtained using the true FIM. In Figure 4, it is apparent that, apart from very low photon counts, all approaches are able to return accurate estimates of the limit of accuracy. Comparing Figure 4a and Figure 4b, it also becomes apparent that for long datasets, approach *(i)* is slightly more accurate than *(ii)*, and the opposite is true for short datasets. In both cases, approach *(iii)* is the most accurate. Similar results can be obtained for the 2D Gaussian profile and Born and Wolf model, as analytical expressions for the FIM are also available for a static object [42, 41].

5. Numerical experiments. In this section, we apply the particle smoother known as SMC-FS to estimate the FIM, and thus the limit of accuracy, for various parameters in the context of one or multiple moving molecules with stochastic trajectories. Experiments are

first run with photon detection locations described by the Gaussian and Airy profiles, and then the Born and Wolf model, where an additional hyperparameter, namely the optical axis location, must be considered as well. The methodology is then applied to the optical microscope resolution problem, where the limit of accuracy for the mean separation distance between two closely spaced diffusing molecules is assessed.

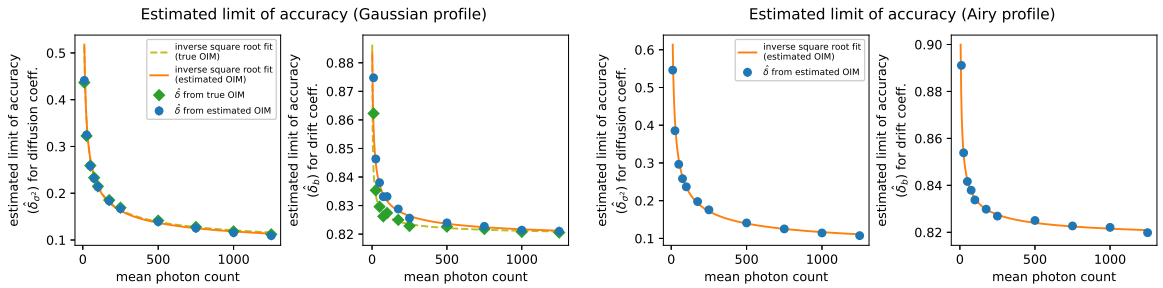
Unless stated otherwise, the FIM for any given settings is estimated according to (4.23), i.e. by generating several datasets according to the same settings, estimating the OIM for each dataset using the SMC-FS algorithm (Algorithm 4.2) and averaging the estimated OIM over all generated datasets. The particle filter employed in the SMC-FS algorithm is the bootstrap filter. A large number of datasets is needed to minimise Monte Carlo error in FIM estimates, so to speed up computations we adopt a distributed computing approach: the datasets and repeat runs of the SMC-FS algorithm to estimate the OIMs are divided evenly among 60 to 64 CPUs and run in parallel. We note that for our methodology, access to a large number of CPUs is beneficial to both the accuracy of estimates and the speed at which they can be obtained. The wall clock speed of the SMC-FS algorithm is also affected by the mean photon count N_{phot} considered. Indeed, as described in Algorithm 4.2, the filtering and smoothing steps only occur in segments where a photon is observed, so the expected complexity of a full run of the SMC-FS algorithm is $\mathcal{O}(N_{phot}N^2)$ where N is the size of the SMC particle population (generally $N = 500$).

5.1. Limit of accuracy of drift and diffusion coefficients for the Gaussian and Airy profiles. Consider a molecule with trajectory described by the SDE in Example 2.1. In [59], the authors took advantage of the Kalman filter formulae to evaluate the FIM for the diffusion (σ^2) and drift (b) coefficients. However, it was only possible to obtain an analytic solution for a particular set of detection times $t_1, t_2, \ldots$ and for the 2D Gaussian photon distribution profile. Otherwise, the computational cost of performing numerical integration was too high for more than one photon.

In our particle filtering framework, it is also possible to take advantage of the Kalman filter formulae when considering the 2D Gaussian model in order to obtain an accurate approximation of the true score and OIM by numerical differentiation, and for any detection times schedule. An estimate of the FIM is therefore obtained by evaluating the true OIM for 3000 datasets and taking their mean, as described in subsection 4.5. The molecule trajectories are simulated for $[0, 0.2]$ seconds, with diffusion coefficient $\sigma^2 = 1\ \mu\text{m}^2/\text{s}$, drift coefficient $b = -10\ \text{s}^{-1}$, and initial location Gaussian distributed with mean $x_0 = (5.5, 5.5)^\intercal\ \mu\text{m}$ and covariance $P_0 = 10^{-2}\mathbb{I}_{2\times 2}\ \mu\text{m}^2$. The observations for the first experiment are generated according to the 2D Gaussian profile (2.9) with parameters as in Example 2.2. It is not possible to employ the Kalman filter formulae for the Airy and Born and Wolf profiles, and we must resort to using the SMC-FS algorithm instead. First of all, to evaluate the performance of the SMC-FS algorithm, the algorithm is employed using $N = 500$ particles to estimate the score and OIM for the same 3000 2D Gaussian profile datasets, and we similarly take the mean OIM over all datasets to estimate the FIM. Next, we move on to the Airy profile, for which it was too computationally costly in [59] to obtain the FIM for more than a single photon. We estimate the OIM for the diffusion and drift coefficients using the SMC-FS algorithm with $N = 500$ particles for 2040 datasets, where the molecule trajectories are simulated using the

same parameters as for the 2D Gaussian profile, and the observations are generated according to the Airy profile (2.8) with parameters as in Example 2.2. This is repeated for various mean photon counts ranging from 10 to 1250. Then, the limit of accuracy estimate, denoted $\hat{\delta}_\vartheta$ for hyperparameter ϑ, is computed, and the results are displayed in Figure 5.

Both Figure 5a and Figure 5b display an inverse square root decay of the limit of accuracy with respect to the mean photon count. This is consistent with the results for a static molecule from Example 4.2, and means that the quality of diffusion and drift estimates improves as the mean photon count increases. In addition to that, comparing the limit of accuracy obtained from the estimated and true OIM for the 2D Gaussian profile in Figure 5a indicates that the SMC-FS algorithm is able to return accurate estimates of the score and FIM for a stochastically moving molecule. Indeed, apart from a very slight discrepancy for very low photon counts for the drift coefficient, the estimates of the limit of accuracy are almost indistinguishable.



(a) *2D Gaussian profile* ***(b)*** *Airy profile*

Figure 5: *Evolution of the estimated limit accuracy for mean photon counts ranging from 10 to 1250. The limit of accuracy is estimated for the diffusion (σ^2) and drift (b) coefficients for an in-focus molecule with stochastic trajectory. The estimates are obtained by taking the square root of the inverse of the FIM, obtained by estimating the OIM using the SMC-FS algorithm with 500 particles for* ***(a)*** *3000 and* ***(b)*** *2040 simulated datasets. To generate each dataset, the molecule's trajectory was simulated according to the SDE in Example 2.1 for the interval $[0, 0.2]$ seconds, with $\sigma^2 = 1\ \mu m^2/s$, $b = -10\ s^{-1}$, and initial location Gaussian distributed with mean $x_0 = (5.5, 5.5)^\intercal\ \mu m$ and covariance $P_0 = 10^{-2}\mathbb{I}_{2\times 2}\ \mu m^2$. The observations are generated according to the* ***(a)*** *2D Gaussian and* ***(b)*** *Airy profiles, with parameters as in Example 2.2. For the* ***(a)*** *2D Gaussian profile, the limit of accuracy is also estimated by using the true OIM obtained using numerical differentiation applied to the Kalman filter. An inverse square root curve (orange and green dashed) is fitted to the resulting estimated limits of accuracy for comparison.*

5.2. Limit of accuracy of drift, diffusion and optical axis location for the Born and Wolf model. When the molecule is out of focus, which means the photon detection locations are distributed according to the Born and Wolf model (2.10), the FIM components for the diffusion and drift coefficients can be obtained as for the Airy and Gaussian profiles. However, a new hyperparameter must be considered, namely the optical axis location, denoted z_0. While previously, differentiating the log potential function was not needed, the vector of hyperparameters is now $\theta = (\sigma^2, b, z_0)$, and $G_k^\theta(x_k)$ depends on z_0 for $k = 1, \ldots, n$.

While it requires numerical integration, differentiating $\log q_{z_0}(x_1, x_2)$ for a given $x = (x_1, x_2) \in \mathbb{R}^2$ with respect to z_0 is not impossible. For notational simplicity, let $\alpha := \frac{2\pi n_\alpha}{\lambda_e}$, $r := \sqrt{x_1^2 + x_2^2}$ and $W := \frac{\pi n_\alpha^2}{n_o \lambda_e}$ and rewrite (2.10) as

$$q_{z_0}(x_1, x_2) = \frac{\alpha^2}{\pi}\left(U_{z_0}^2 + V_{z_0}^2\right),$$

where

$$U_{z_0} := \int_0^1 J_0(\alpha r \rho)\cos\left(W z_0 \rho^2\right)\rho d\rho, \qquad V_{z_0} := \int_0^1 J_0(\alpha r \rho)\sin\left(W z_0 \rho^2\right)\rho d\rho.$$

The first derivative was derived in [41] and is given by

$$\frac{\partial \log q_{z_0}(x_1, x_2)}{\partial z_0} = 2\frac{U_{z_0}\dot{U}_{z_0} + V_{z_0}\dot{V}_{z_0}}{U_{z_0}^2 + V_{z_0}^2},$$

where

$$\dot{U}_{z_0} := \frac{\partial U_{z_0}}{\partial z_0} = \int_0^1 J_0(\alpha r \rho)\cos\left(W z_0 \rho^2\right) W \rho^3 d\rho,$$
$$\dot{V}_{z_0} := \frac{\partial V_{z_0}}{\partial z_0} = -\int_0^1 J_0(\alpha r \rho)\sin\left(W z_0 \rho^2\right) W \rho^3 d\rho.$$

The second derivative with respect to z_0 is given by

$$\frac{\partial^2 \log q_{z_0}(x_1, x_2)}{\partial z_0^2} = 2\frac{U_{z_0}\ddot{U}_{z_0} + \dot{U}_{z_0}^2 + V_{z_0}\ddot{V}_{z_0} + \dot{V}_{z_0}^2}{U_{z_0}^2 + V_{z_0}^2} - \left(\frac{\partial \log q_{z_0}(x_1, x_2)}{\partial z_0}\right)^2,$$

where

$$\ddot{U}_{z_0} := \frac{\partial^2 U_{z_0}}{\partial z_0^2} = -\int_0^1 J_0(\alpha r \rho)\cos\left(W z_0 \rho^2\right) W^2 \rho^5 d\rho,$$
$$\ddot{V}_{z_0} := \frac{\partial^2 V_{z_0}}{\partial z_0^2} = -\int_0^1 J_0(\alpha r \rho)\sin\left(W z_0 \rho^2\right) W^2 \rho^5 d\rho.$$

The potential function only depends on z_0, so any cross terms in the FIM and OIM between z_0 and either σ^2 or b will be zero.

The OIM is estimated for the diffusion (σ^2), drift (b) coefficients and optical axis location (z_0) using the SMC-FS algorithm with 500 particles for 2040 datasets, where the molecule trajectories are simulated using the same parameters as for the 2D Gaussian and Airy profiles, and the observations are generated according to the Born and Wolf model with parameters as in Example 2.2 (i.e. $z_0 = 1$ μm). Then, the limit of accuracy for mean photon counts ranging from 10 to 1250 is computed, and the results are displayed in Figure 6. Once again, there is an inverse square root decay of the limit of accuracy with respect to the mean photon count for all hyperparameters considered.

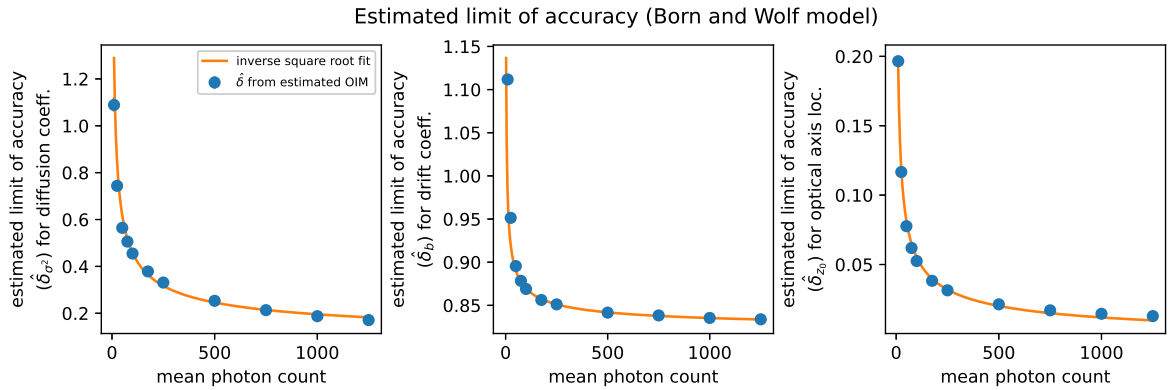


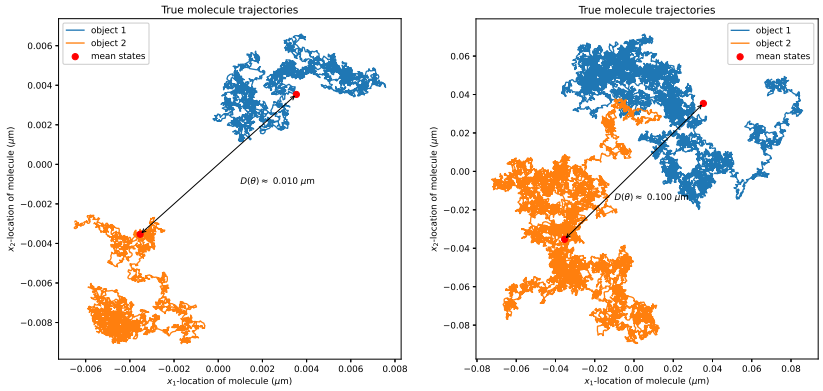
Figure 6: *Evolution of the estimated limit accuracy for mean photon counts ranging from 10 to 1250. The limit of accuracy is estimated for the diffusion (σ^2), drift (b) coefficients and optical axis location (z_0) for an out-of-focus molecule with stochastic trajectory. The estimates are obtained by taking the square root of the inverse of the FIM, obtained by estimating the OIM using the SMC-FS algorithm with 500 particles for 2040 simulated datasets. To generate each dataset, the molecule trajectories are simulated according to the SDE in Example 2.1 for the interval $[0, 0.2]$ seconds, with $\sigma^2 = 1\ \mu m^2/s$, $b = -10\ s^{-1}$, and initial location Gaussian distributed with mean $x_0 = (5.5, 5.5)^\intercal\ \mu m$ and covariance $P_0 = 10^{-2}\mathbb{I}_{2\times 2}\ \mu m^2$. The observations are generated according to the Born and Wolf model with parameters as in Example 2.2, where $z_0 = 1\ \mu m$. An inverse square root curve (orange) is fitted to the resulting estimated limits of accuracy for comparison.*

5.3. Limit of accuracy of the separation distance between two molecules for the Airy profile. Being able to estimate the distance of separation between two closely spaced molecules is an important aspect of single-molecule microscopy. In the past, Rayleigh's criterion [4] has been used to define the minimum distance between two point sources such that they can be distinguished in the image. However, [48] treated the separation distance problem as a statistical estimation task and derived the CRLB (or inverse of the FIM) for the mean square error of the separation distance estimate. It was shown that Rayleigh's minimum distance can be surpassed by capturing more photons, e.g. by observing the molecules for a longer period. So far, the limit of accuracy has only been derived for static molecules. In this experiment, we apply our methodology to estimate the limit of accuracy for the locations and separation distance between two molecules that are not static, but diffusing independently at their respective stationary distributions, as illustrated in Figure 7.

Let $X_t = (X_{t,1}, X_{t,2})^\intercal$ be the cartesian coordinates of a moving molecule with stationary distribution $\mathcal{N}(x_0, \sigma^2\mathbb{I}_{2\times 2})$ for all t, where x_0 is referred to as the *mean state*. The continuous time dynamics are given by

$$\mathrm{d}X_t = (x_0 - X_t)\mathrm{d}t + \sqrt{2}\sigma \mathrm{d}B_t. \tag{5.1}$$

From subsection 2.1, it is straightforward to establish the solution to this SDE, which yields the conditional pdf $f_\Delta^{x_0}$ of X_{k+1} at the $(k+1)$-th discrete segment, given $X_k = x$ at the k-th



(a) *(b)*

Figure 7: *Examples of two molecules diffusing independently at a mean separation distance of* ***(a)*** $0.01\ \mu m$ *with diffusion coefficient* $\sigma^2 = 10^{-4}\ \mu m^2/s$ ***(b)*** $0.1\ \mu m$ *with* $\sigma^2 = 10^{-3}\ \mu m^2/s$. *For an Airy distributed photon detection profile with* $n_\alpha = 1.4$ *and* $\lambda_e = 0.52\ \mu m$, *Rayleigh's resolution limit is* $\approx 0.227\ \mu m$. *Increasing the value of the diffusion coefficient* σ^2 *will often lead to the molecule trajectories overlapping.*

segment, as

$$X_{k+1}|(X_k = x) = \Phi_\Delta x + a_\Delta + W_x, \quad W_x \sim \mathcal{N}(0, R_\Delta),$$

where $\Phi_\Delta = e^{-\Delta}$, $a_\Delta = x_0(1 - e^{-\Delta})$ and $R_\Delta = \sigma^2(1 - e^{-2\Delta})\mathbb{I}_{2\times 2}$.

In this experiment, consider two independently diffusing molecules whose states are (X_t, V_t), where X_t is the state of the first molecule and V_t is the state of the second. Assume that the initial state of each molecule is the same as its corresponding mean state, i.e. $(x_0, v_0) =: \theta = (\theta_1, \theta_2, \theta_3, \theta_4)^\intercal$, and is non-random but unknown and to be estimated. The conditional probability density function of (X_{k+1}, V_{k+1}) given $(X_k, V_k) = (x_k, v_k)$ is $f_\Delta^{x_0}(x_{k+1}|x_k) f_\Delta^{v_0}(v_{k+1}|v_k)$ owing to their independent motions.

Let $\hat{\theta} = \left(\hat{\theta}_1(Y_{1:n}), \hat{\theta}_2(Y_{1:n}), \hat{\theta}_3(Y_{1:n}), \hat{\theta}_4(Y_{1:n})\right)^\intercal$ denote an estimate of θ given observations $Y_{1:n}$. Recall that the FIM, denoted $\mathcal{I}_n(\theta)$, is given by

$$\mathcal{I}_n(\theta) = \mathbb{E}\left[\nabla \log p_\theta(Y_{1:n})\ \nabla \log p_\theta(Y_{1:n})^T\right].$$

For any scalar-valued function $D(\theta) \in \mathbb{R}$, we can estimate $D(\theta)$ using $D(\hat{\theta})$ where $\hat{\theta}$ is the estimate of θ. Assuming the estimate is unbiased, we have the following CRLB for the function D,

$$\mathbb{E}\left[\left(D(\hat{\theta}) - D(\theta)\right)^2\right] \geq \nabla D(\theta)^\intercal \mathcal{I}_n(\theta)^{-1} \nabla D(\theta), \tag{5.2}$$

where $\nabla D(\theta) := (\partial D/\partial\theta_1, \ldots, \partial D/\partial\theta_4)^\intercal$. For example, to estimate the separation between

the two molecules we have $D(\theta) = \sqrt{(\theta_1 - \theta_3)^2 + (\theta_2 - \theta_4)^2}$, and as a result

$$\nabla D(\theta) = \frac{1}{D(\theta)} \begin{pmatrix} \theta_1 - \theta_3 \\ \theta_2 - \theta_4 \\ -(\theta_1 - \theta_3) \\ -(\theta_2 - \theta_4) \end{pmatrix}.$$

This experiment is essentially the dynamic version of the experiments on estimating the separation of two static molecules by [50]. The key difference here is that the molecules are diffusing. The observations $Y_{1:n}$ are generated as in [50], i.e. according to the following mixture

$$G_k(x_k, v_k) = \begin{cases} 1 - \Delta\lambda_\theta, & \text{if } y_k = \emptyset, \\ \lambda_x g(y_k|x_k) + \lambda_v g(y_k|v_k), & \text{otherwise,} \end{cases} \tag{5.3}$$

where g is the photon distribution profile given in (2.7) and $\lambda_\theta = \lambda_x + \lambda_v$. The measurement model considered in this experiment is the Airy profile (2.8), but it is straightforward to also apply the methodology to the 2D Gaussian profile and Born and Wolf model.

In the first part of the experiment, we analytically replicate results similar to those in [48, 50] for two static molecules, then observe how introducing diffusion affects the progression of the limit of accuracy $\delta_{D(\theta)}$ for the separation distance (obtained using (5.2)), as this separation distance between the two molecules increases. We set $\lambda_x = \lambda_v = \lambda$ for simplicity. Evaluating $\delta_{D(\theta)}$ analytically for the static case is performed as in [48], with a mean photon count, denoted N_{phot}, of 3000. For the dynamic case, the molecules are observed during an interval of $[0, 1]$ seconds with the same mean photon count, and for diffusion coefficients σ^2 varying from 5×10^{-3} to 10^{-4} μm^2/s. The parameters of the Airy profile are unchanged (i.e. $n_\alpha = 1.4$, $\lambda_e = 0.52$ μm), as is the lateral magnification matrix ($M = 100\mathbb{I}_{2\times 2}$). The estimate of the limit of accuracy is obtained by estimating the OIM for the mean locations x_0 and v_0 via the SMC-FS algorithm for 640 to 1024 datasets then applying (5.2). The resulting estimated limits of accuracy $\hat{\delta}_{D(\theta)}$ are given in Figure 8a. The second part of the experiment involves similarly estimating the limits of accuracy $\delta_{D(\theta)}$ for various separation distances, but this time the diffusion coefficient remains fixed, i.e. $\sigma^2 = 10^{-4}$ μm^2/s, and the mean photon count N_{phot} is set to vary between 100 and 4500. The resulting estimated limits of accuracy are given in Figure 8b.

As the separation distance $D(\theta)$ gets closer to zero, the limit of accuracy increases, indicating that estimates would become less accurate. Additionally, an inverse square root curve was fit to each set of estimated limits of accuracy in Figure 8a and Figure 8b. This is consistent with results in [48] that showed an inverse square root relationship between separation distance and $\delta^{static}_{D(\theta)}$ for two static molecules, and indicates that these results can be generalised to dynamic molecules. Additionally, in [40], it is suggested that the limit of accuracy for the location of a static molecule, known as *localisation accuracy* and denoted δ^{loc}, is of the form $\frac{\sigma_a}{\sqrt{N_{phot}}}$ where N_{phot} is the mean photon count and σ_a the standard deviation of the photon detection profile. The interpretation for this is that the quality of location estimates of a single static molecule deteriorates as the measurement uncertainty σ_a increases. Now in [50],

it is proven that the limit of accuracy for the separation distance between two molecules $\delta^{static}_{D(\theta)}$ and the localisation accuracy for each of these molecules are related as follows:

$$H^{sta}_{N_{phot}} := \lim_{D(\theta)\to\infty} \delta^{static}_{D(\theta)} = \sqrt{\left(\delta^{sta,loc}_{x_0}\right)^2 + \left(\delta^{sta,loc}_{v_0}\right)^2}, \tag{5.4}$$

where $\delta^{sta,loc}_{x_0}$ and $\delta^{sta,loc}_{v_0}$ denote the localisation accuracy for the first and second (static) molecule observed independently with cumulative mean photon count N_{phot}, respectively. Even though the separation distance goes to infinity, its limit of accuracy $\delta_{D(\theta)}$ remains finite. This means that as $D(\theta) \to \infty$, evaluating the limit of accuracy for the separation distance between two (static) molecules becomes equivalent to two independent localisation accuracy problems. It also means that $\delta^{static}_{D(\theta)}$ is similarly affected by measurement uncertainty σ_a as are the localisation accuracies for the two molecules.

In this experiment, the introduction of diffusion negatively affects the improvement in estimation accuracy as the mean distance of separation between the two molecules increases. This is evidenced in Figure 8a by the more and more slowly decaying limits of accuracy as the value of σ^2 increases, and in Figure 9a by the linearly increasing trend in $\hat{\delta}_{D(\theta)}$ for all values of $D(\theta)$ as σ increases. As a result, the diffusion coefficient in the dynamic model can be translated into additional observation uncertainty which affects $\delta_{D(\theta)}$ in a way reminiscent of how σ_a affects $\delta^{static}_{D(\theta)}$. More generally, from our numerical results, we observe the relationship for our dynamic application behaves qualitatively as

$$\sqrt{\frac{\sigma_a^2 + \sigma^2}{N_{phot}}},$$

where, as above, σ_a is the standard deviation of the photon detection process, also known as measurement uncertainty.

We now investigate the relationship between $\delta_{D(\theta)}$ and the dynamic equivalent to the localisation accuracy, namely the limit of accuracy for the mean locations x_0 and v_0 of each individual, stochastically moving molecule, denoted $\delta^{sto,loc}_{x_0}$ and $\delta^{sto,loc}_{v_0}$, respectively. The limits $\delta^{sto,loc}_{x_0}$ and $\delta^{sto,loc}_{v_0}$ can be estimated independently by repeatedly taking the mean estimated OIM for x_0 and v_0 based on two separate sets of 640 simulated datasets (one for each molecule) for mean photon counts ranging from 50 to 2250 (half of N_{phot} each, given we have $\lambda_x = \lambda_v = \lambda$ under current settings). The distance

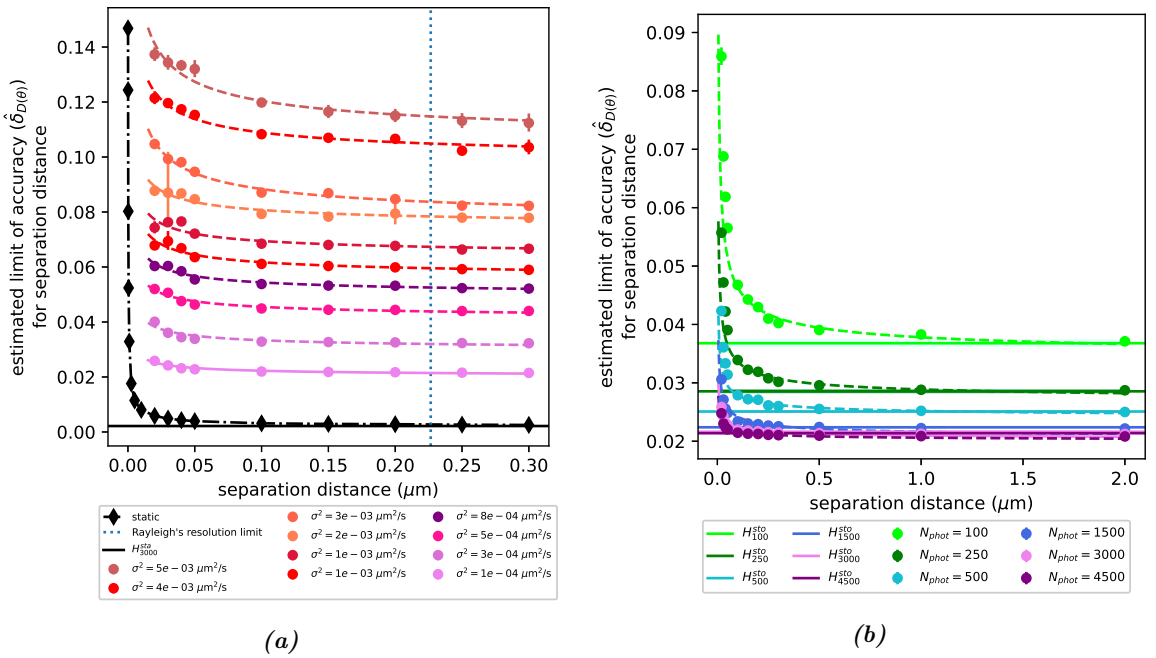
$$H^{sto}_{N_{phot}} := \sqrt{\left(\delta^{sto,loc}_{x_0}\right)^2 + \left(\delta^{sto,loc}_{v_0}\right)^2}$$

between the limits of accuracy $\delta^{sto,loc}_{x_0}$ and $\delta^{sto,loc}_{v_0}$ of each individual object with various (cumulative) mean photon counts N_{phot} is illustrated as horizontal lines in Figure 8b, which appear to act as asymptotes, thus indicating that the relationship in (5.4) can be generalised to stochastically moving molecules. While the introduction of diffusion leads to less accurate estimates, Figure 8b displays a stronger decay in the limit of accuracy as the mean photon count N_{phot} increases, thus indicating that increasing the mean photon count N_{phot} improves those estimates, as was the case for static molecules in [48]. This is reinforced in Figure 9b,

765 which also suggests that the relationship between $\delta_{D(\theta)}$ and N_{phot} is an inverse square root.
766 This is also a generalisation to the dynamic case of results in [48] which showed an inverse
767 square root relationship between $\delta^{static}_{D(\theta)}$ and N_{phot} for two static molecules.
768 In summary, this experiment employs the numerical framework developed in this paper for
769 estimating the FIM of parameters of dynamic molecules using SMC in order to gain insights
770 into generalising results from [49, 50] about the effects of separation distance, measurement
771 uncertainty and mean photon count to a context in which the two molecules considered follow
772 a SDE rather than being static. These effects, as well at that of the measurement uncertainty,
773 can all be observed by applying our methodology and are summarised in Table 1. We also
774 summarise in Table 1 the results on the limits of accuracy for the drift and diffusion coefficients
775 of a single stochastically moving molecule observed via the 2D Gaussian, Airy profiles and
776 the Born and Wolf model from subsection 5.1 and subsection 5.2. Note that the limits of
777 accuracy for the mean locations of each molecule, denoted $\delta_\theta := (\delta_{\theta_1}, \delta_{\theta_2}, \delta_{\theta_3}, \delta_{\theta_4})^\intercal$, can also
778 be estimated as part of our methodology (as their FIM is required for (5.2)) and return
779 similar relationships with separation distance, mean photon count, diffusion coefficient and
780 measurement uncertainty as $\delta_{D(\theta)}$ (not reported here).
781 In this section, results on the relationship between the limits of accuracy for various
782 parameters and the mean photon count N_{phot} have been extended from a single static [40, 10,
783 41] or deterministically moving molecule [60] to a molecule whose trajectory follows an SDE.
784 Additionally, insights have been gained into generalising results for the optical microscope
785 resolution problem, which considers the separation distance between two static molecules
786 [48, 49], to two stochastically diffusing molecules. The qualitative relationships observed and
787 summarised in Table 1 are important in an experimental design context, as they provide
788 information on how the accuracy of parameter estimates is affected by various experimental
789 setups. For example, the $\mathcal{O}(N_{phot}^{-1/2})$ relationship between limits of accuracy and mean photon
790 count indicates that quadrupling the number of photons can help halve the standard deviation
of parameter estimates.

Limit of accuracy δ $\delta_\vartheta = \text{std}(\hat{\vartheta})$	**Qualitative Dependence** Parameter	Relationship	**Reference**
$\delta_{D(\theta)}, \delta_\theta$	$D(\theta)$ separation distance	$\mathcal{O}\left(D(\theta)^{-1/2}\right)$	Figure 8
$\delta_{D(\theta)}, \delta_\theta$	σ^2 diffusion coefficient	$\mathcal{O}(\sigma)$	Figure 9a
$\delta_{D(\theta)}, \delta_\theta$	σ_a^2 measurement uncertainty	$\mathcal{O}(\sigma_a)$	[48, 40]
$\delta_{D(\theta)}, \delta_\theta$	N_{phot} mean photon count	$\mathcal{O}\left(N_{phot}^{-1/2}\right)$	Figure 9b
$\delta_{\sigma^2}, \delta_b, \delta_{z_0}$	N_{phot} mean photon count	$\mathcal{O}\left(N_{phot}^{-1/2}\right)$	Figure 5, Figure 6

Table 1: *Summary of the qualitative relationships between the limits of accuracy (or standard deviation of parameter estimates)* $\delta_\theta := (\delta_{\theta_1}, \delta_{\theta_2}, \delta_{\theta_3}, \delta_{\theta_4})^\intercal$ *and* $\delta_{D(\theta)}$ *for the mean locations* $\theta = (x_0, v_0) = (\theta_1, \theta_2, \theta_3, \theta_4)^\intercal$ *and separation distance* $D(\theta)$*, respectively, of two stochastically diffusing molecules observed simultaneously. Also included in the table is the relationship between mean photon count and the limits of accuracy for the hyperparameters of the SDE trajectory (drift* b *and diffusion* σ^2 *coefficients) and photon detection process (optical axis location* z_0*) of a single molecule. Note that when we increase the mean photon count* N_{phot}*, the observation interval length remains fixed.*

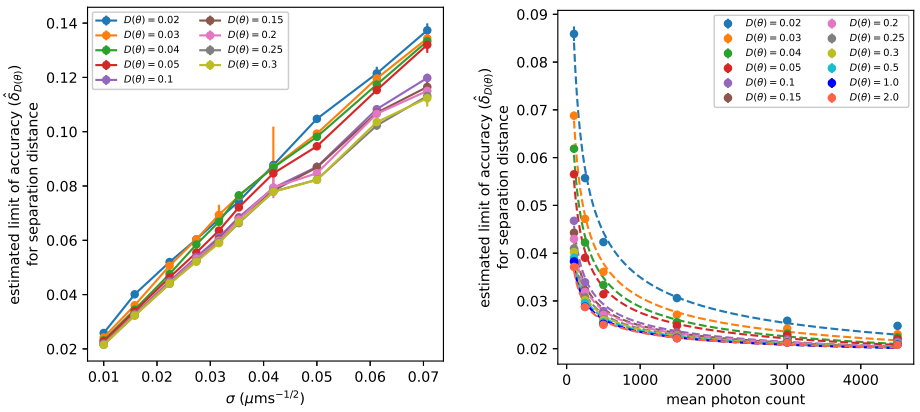


(a) *(b)*

Figure 8: *Comparison of the evolution of the estimated limit accuracy for separation distances ranging from* 20×10^{-3} *to* 2 μm *for various* ***(a)*** *diffusion coefficient* (σ^2) *values* ***(b)*** *mean photon counts* (N_{phot}). *The limit of accuracy for the separation distance* $\delta_{D(\theta)}$, *where* $\theta = (x_0, v_0) = (\theta_1, \theta_2, \theta_3, \theta_4)$, *is estimated using the square root of the CRLB obtained using* (5.2) *(in the dynamic case) and evaluated using analytical results from [48] (in the static case). The estimates of* $\mathcal{I}_n(\theta)$ *in* (5.2) *are obtained by running the SMC-FS algorithm with 500 particles for 640 to 1024 simulated datasets. For the dynamic case, the molecule trajectories are initialised at their respective mean locations* x_0 *and* v_0 *and each is propagated according to its corresponding SDE* (5.1) *during an interval of* $[0, 1]$ *seconds with* ***(a)*** *fixed and mean photon count* $N_{phot} = 3000$ ***(b)*** *fixed diffusion coefficient* $\sigma^2 = 10^{-4}$ $\mu m^2/s$. *The observations are generated according to a mixture of Airy profiles* (5.3) *with parameters as in Example* 2.2. *This is repeated for* ***(a)*** σ^2 *varying from* 10^{-3} *to* 10^{-4} $\mu m^2/s$ ***(b)*** N_{phot} *varying from* 100 *to* 4500. *Finally, an inverse square root curve is fitted to each of the resulting sets of estimated limits of accuracy for comparison purposes. Note that the pink set of estimates and their corresponding solid fitted curve in* ***(a)*** *coincide with those in* ***(b)***. *In* ***(b)***, *the horizontal lines correspond to the equivalent mean photon counts and represent the distances* $H^{sto}_{N_{phot}}$ *between the limits of accuracy* $\hat{\delta}^{sto,loc}_{x_0}$ *and* $\hat{\delta}^{sto,loc}_{v_0}$ *for the mean locations* x_0 *and* v_0 *of each individual object, estimated independently for each molecule using the SMC-FS algorithm. Note that any variation in estimates for low separation distances is due to Monte Carlo error, and can be reduced by increasing the number of simulated datasets.*

6. Conclusion.

In this paper, we introduced an SMC approach to performing parameter inference when tracking a molecule with stochastic trajectory for a fixed time interval. The three main aspects of this fundamental model in single-molecule microscopy were the true location of the molecule in the object space, which follows a linear SDE, the Poisson distributed arrival process of the photons it emits on the detector in the image space, and the arrival location of those photons on the detector, which follows either a 2D Gaussian, Airy profile,



(a) *(b)*

Figure 9: *Evolution of the estimated limit of accuracy for the separation distance $\delta_{D(\theta)}$ (obtained using (5.2)) between two stochastically moving molecules observed simultaneously for* ***(a)*** *σ ranging from $\sqrt{10^{-4}}$ to $\sqrt{5 \times 10^{-3}}$ μm $s^{-1/2}$* ***(b)*** *N_{phot} ranging from 100 to 4500. Estimates are obtained through the same algorithm and parameters as in* ***(a)*** *Figure 8a* ***(b)*** *Figure 8b, with separation distances ranging from 20×10^{-3} to 2 μm. In* ***(b)****, inverse square root curves are fitted to the resulting estimates $\hat{\delta}_{D(\theta)}$ for comparison.*

or Born and Wolf model.

First of all, we discretised the time interval in order to formulate the problem as a discrete-time state space model, in which all states are equally spaced in time, but a number of observations are marked as missing. From this, SMC methods were applied for parameter inference. A general forward smoothing algorithm was employed to estimate the score and OIM of the data regardless of the distribution of the photon locations. For the first time, this allowed for the estimation of the FIM and hence the limit of accuracy (square root of the CRLB), which could not be done before for the Airy profile and Born and Wolf model, and could only be achieved analytically for a specific set of photon detection times for the 2D Gaussian profile. The methodology was subsequently applied to characterise the precision limits for estimating the separation distance between two moving molecules, thus providing new insights into results for the static case from [50]. The outcome of our numerical work was summarised in Table 1, which sums up the qualitative behaviours of the limits of accuracy as functions of the mean photon count, separation distance, diffusion coefficient and measurement uncertainty.

Although for the first time a method has been described to estimate the limit of accuracy for the hyperparameters of dynamic single molecules with non-uniform observation times and complex measurement models, such as the Airy profile or Born and Wolf model, there is scope to use the techniques developed here to provide a wider range of more computationally efficient approaches. Indeed, an advantage of the straightforward state space model formulation of the problem is access to the vast range of filtering and smoothing algorithms available. While

we employed forward smoothing, any kind of particle smoothing algorithm would be suitable, and indeed, the SMC-FS algorithm of [16] employed for forward smoothing, even though it mitigates issues related to path degeneracy, is of $\mathcal{O}(N^2)$ complexity. For example, the PaRIS algorithm of [46] can reduce the complexity of the algorithm to linear.

REFERENCES

[1] T. T. Ashley and S. B. Andersson, *Method for simultaneous localization and parameter estimation in particle tracking experiments*, Physical Review E, 92 (2015), p. 052707.

[2] A. J. Berglund, *Statistics of camera-based single-particle tracking*, Physical Review E, 82 (2010), p. 011917.

[3] P. J. Bickel, Y. Ritov, T. Ryden, et al., *Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models*, The Annals of Statistics, 26 (1998), pp. 1614–1635.

[4] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*, Elsevier, 2013.

[5] V. Briane, C. Kervrann, and M. Vimond, *Statistical analysis of particle trajectories in living cells*, Physical Review E, 97 (2018), p. 062121.

[6] C. P. Calderon, *Motion blur filtering: a statistical approach for extracting confinement forces and diffusivity from a single blurred trajectory*, Physical Review E, 93 (2016), p. 053303.

[7] C. P. Calderon and K. Bloom, *Inferring latent states and refining force estimates via hierarchical dirichlet process modeling in single particle tracking experiments*, PloS one, 10 (2015).

[8] C. P. Calderon, M. A. Thompson, J. M. Casolari, R. C. Paffenroth, and W. Moerner, *Quantifying transient 3D dynamical phenomena of single mRNA particles in live yeast cell measurements*, The Journal of Physical Chemistry B, 117 (2013), pp. 15701–15713.

[9] O. Cappé, E. Moulines, and T. Rydén, *Inference in hidden Markov models*, Springer Science & Business Media, 2006.

[10] J. Chao, E. S. Ward, and R. J. Ober, *Fisher information theory for parameter estimation in single molecule microscopy: tutorial*, JOSA A, 33 (2016), pp. B36–B57.

[11] M. K. Cheezum, W. F. Walker, and W. H. Guilford, *Quantitative comparison of algorithms for tracking single fluorescent particles*, Biophysical journal, 81 (2001), pp. 2378–2388.

[12] N. Chopin and O. Papaspiliopoulos, *An introduction to sequential Monte Carlo*, Springer, 2020.

[13] H. Cramér, *Mathematical methods of statistics*, vol. 43, Princeton university press, 1999.

[14] G. Darmois, *Sur les limites de la dispersion de certaines estimations*, Revue de l'Institut International de Statistique, (1945), pp. 9–15.

[15] M. H. DeGroot and M. J. Schervish, *Probability and statistics*, Pearson Education, 2012.

[16] P. Del Moral, A. Doucet, and S. Singh, *Forward smoothing using sequential Monte Carlo*, arXiv preprint arXiv:1012.5390, (2010).

[17] R. Douc, E. Moulines, T. Rydén, et al., *Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime*, The Annals of statistics, 32 (2004), pp. 2254–2304.

[18] R. Douc, E. Moulines, and D. Stoffer, *Nonlinear time series: Theory, methods and applications with R examples*, CRC press, 2014.

[19] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer Science & Business Media, 2001.

[20] A. Doucet, S. Godsill, and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Statistics and computing, 10 (2000), pp. 197–208.

[21] A. Doucet and A. M. Johansen, *A tutorial on particle filtering and smoothing: Fifteen years later*, Handbook of nonlinear filtering, 12 (2009), p. 3.

[22] J. Duchi, *Lecture notes for statistics 311/electrical engineering 377*, URL: https://stanford.edu/class/stats311/Lectures/full_notes. pdf, 2 (2016), p. 23.

[23] B. Efron and D. V. Hinkley, *Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information*, Biometrika, 65 (1978), pp. 457–483.

[24] L. C. Evans, *An introduction to stochastic differential equations*, vol. 82, American Mathematical Soc.,

$[t_0, T]$	observation interval, $t \in [t_0, T]$
$\mathcal{X}$	object plane
$X(t)$, X_t	object location at time t in the object space
$f^{\theta}_{s,t}(x_t \vert x_s)$	probability density of X_t given previous location X_s, relabelled $f^{\theta}_{\Delta}(x_{k+1} \vert x_k)$ after discretisation, see subsection 2.1
$\mathcal{Y}$	detector (in the image space)
$Y(t)$, Y_t	location of detected photon at time t in the image space
$g_\theta(y \vert x)$	photon distribution profile for an object located at x, see (2.7)
$q_{z_0}(x_1, x_2)$	image function, see (2.7)
$N(t)$	number of photons detected at time t
$\lambda(t), \lambda$	photon detection rate
$b(t, X_t)$, b	drift coefficient, see (2.1), (2.6)
$\sigma(t, X_t)$, σ^2	diffusion coefficient, see (2.1), (2.6)
$\mathrm{d}B_t$	Wiener process
θ	vector of model parameters (hyperparameters)
z_0	optical axis location ($z_0 = 0$ if object is in focus)
M	lateral magnification matrix
$t_0 < t_1 < \ldots < t_{n_p}$	arrival times of n_p photons on the detector
Δ	discretised segment length, see subsection 3.2
$G^{\theta}_k(x)$	potential function for an object located at x at k-th discrete segment, see (3.1)
$\mathcal{I}_n(\theta)$	Fisher information matrix (FIM) for vector θ, see (4.18), (4.19)
CRLB_ϑ	Cramér-Rao lower bound for parameter ϑ
δ_ϑ	limit of accuracy for parameter ϑ
$e_{o,1:2}, e_{i,1:2}$	unit vectors in the object (o) and image (i) plane, see Figure 1
$\mathbb{E}[\cdot], \mathrm{Cov}[\cdot]$	expectation, covariance
$\mathbb{P}[\cdot]$	probability
$\mathbb{I}_{d\times d}$	identity matrix of dimensions $d \times d$
$J_a(\cdot)$	Bessel function of the first kind of order a
$\mathcal{N}(\mu, \Sigma)$	multivariate Gaussian distribution with mean μ and covariance Σ
∇, ∇^2	gradient, Hessian
σ_a^2	measurement uncertainty, see (2.9)
n_α	numerical aperture of the objective lens, see (2.8), (2.10)
λ_e	emission wavelength, see (2.8), (2.10)
n_o	refractive index of the objective lens immersion medium, see (2.10)
$\delta_{v_0}(v)$	Dirac delta mass located at v_0
(i)	particle index (superscript)
$\omega_k^{(i)}$	i-th particle normalised importance weight, see subsection 4.2
$p_\theta(x_{1:n} \vert y_{1:n})$	joint smoothing distribution, see (4.1)
$S^{\theta}_k(x_{1:k})$	additive functional, see (4.4)
$s^{\theta}_j(x_{j-1}, x_j)$	sufficient statistic, see (4.4)
$\mathcal{S}_k(\theta)$	smoothing expectation for parameter vector θ, see (4.5), (4.9)
$T^{\theta}_k(x_k)$	auxiliary function, see (4.6)
$\Vert\cdot\Vert^2$	vector inner product, i.e. $\Vert x\Vert^2 = x^\intercal x$
$o(\cdot)$	little-o notation
$\mathcal{O}(\cdot)$	big O notation

Table 2: *Glossary of mathematical symbols.*

2012.
[25] M. Fréchet, *Sur l'extension de certaines évaluations statistiques au cas de petits échantillons*, Revue de l'Institut International de Statistique, (1943), pp. 182–205.
[26] S. J. Godsill, A. Doucet, and M. West, *Monte Carlo smoothing for nonlinear time series*, Journal of the american statistical association, 99 (2004), pp. 156–168.
[27] J. W. Goodman, *Introduction to Fourier optics*, Roberts and Company Publishers, 2005.
[28] N. J. Gordon, D. J. Salmond, and A. F. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, in IEE proceedings F (radar and signal processing), vol. 140, IET, 1993, pp. 107–113, https://doi.org/10.1049/ip-f-2.1993.0015.
[29] J. Houssineau, S. S. Singh, and A. Jasra, *Identification of multiobject dynamical systems: Consistency and fisher information*, SIAM Journal on Control and Optimization, 57 (2019), pp. 2603–2627.
[30] M. Hürzeler and H. R. Künsch, *Monte Carlo approximations for general state-space models*, Journal of Computational and graphical Statistics, 7 (1998), pp. 175–193.
[31] A. H. Jazwinski, *Stochastic processes and filtering theory*, Courier Corporation, 2007.
[32] G. Kitagawa, *Monte Carlo filter and smoother for non-Gaussian nonlinear state space models*, Journal of computational and graphical statistics, 5 (1996), pp. 1–25.
[33] G. Kitagawa and S. Sato, *Monte Carlo smoothing and self-organising state-space model*, in Sequential Monte Carlo methods in practice, Springer, 2001, pp. 177–195.
[34] A. Kong, J. S. Liu, and W. H. Wong, *Sequential imputations and Bayesian missing data problems*, Journal of the American statistical association, 89 (1994), pp. 278–288.
[35] F. Le Gland and L. Mevel, *Recursive identification in hidden Markov models*, in Proceedings of the 36th Conference on Decision and Control, San Diego 1997, vol. 4, 1997, pp. 3468–3473.
[36] Z. Lin, Y. Wong, and R. J. Ober, *Limit of the accuracy of parameter estimation for two molecules moving in close proximity*, in 2015 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2015, pp. 441–444.
[37] X. Michalet and A. J. Berglund, *Optimal diffusion coefficient estimation in single-particle tracking*, Physical Review E, 85 (2012), p. 061916.
[38] X. Michalet, A. N. Kapanidis, T. Laurence, F. Pinaud, S. Doose, M. Pflughoefft, and S. Weiss, *The power and prospects of fluorescence microscopies and spectroscopies*, Annual review of biophysics and biomolecular structure, 32 (2003), pp. 161–182.
[39] W. Moerner and D. P. Fromm, *Methods of single-molecule fluorescence spectroscopy and microscopy*, Review of Scientific instruments, 74 (2003), pp. 3597–3619.
[40] R. J. Ober, S. Ram, and E. S. Ward, *Localization accuracy in single-molecule microscopy*, Biophysical journal, 86 (2004), pp. 1185–1200.
[41] R. J. Ober, E. S. Ward, and J. Chao, *Quantitative Bioimaging: An Introduction to Biology, Instrumentation, Experimentss and Data Analysis for Scientists and Engineers*, CRC Publishing, 2020, ch. Localizing Objects and Single Molecules in Three Dimensions, pp. 377–401.
[42] R. J. Ober, E. S. Ward, and J. Chao, *Quantitative Bioimaging: An Introduction to Biology, Instrumentation, Experimentss and Data Analysis for Scientists and Engineers*, CRC Publishing, 2020, ch. Localizing Objects and Single Molecules in Two Dimensions, pp. 337–375.
[43] B. Oksendal, *Stochastic differential equations: an introduction with applications*, Springer Science & Business Media, 2013.
[44] J. Olsson, O. Cappé, R. Douc, E. Moulines, et al., *Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models*, Bernoulli, 14 (2008), pp. 155–179.
[45] J. Olsson, J. Ströjby, et al., *Particle-based likelihood inference in partially observed diffusion processes using generalised poisson estimators*, Electronic Journal of Statistics, 5 (2011), pp. 1090–1122.
[46] J. Olsson, J. Westerborn, et al., *Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm*, Bernoulli, 23 (2017), pp. 1951–1996.
[47] G. Poyiadjis, A. Doucet, and S. S. Singh, *Particle approximations of the score and observed information matrix in state space models with application to parameter estimation*, Biometrika, 98 (2011), pp. 65–80.
[48] S. Ram, E. S. Ward, and R. J. Ober, *Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy*, Proceedings of the National Academy of Sciences, 103 (2006), pp. 4457–4462.

[49] S. Ram, E. S. Ward, and R. J. Ober, *A stochastic analysis of performance limits for optical microscopes*, Multidimensional Systems and Signal Processing, 17 (2006), pp. 27–57.
[50] S. Ram, E. S. Ward, and R. J. Ober, *A stochastic analysis of distance estimation approaches in single molecule microscopy: quantifying the resolution limits of photon-limited imaging systems*, Multidimensional systems and signal processing, 24 (2013), pp. 503–542.
[51] C. R. Rao, *Information and the accuracy attainable in the estimation of statistical parameters*, in Breakthroughs in statistics, Springer, 1992, pp. 235–247.
[52] P. K. Relich, M. J. Olah, P. J. Cutler, and K. A. Lidke, *Estimation of the diffusion constant from intermittent trajectories with variable position uncertainties*, Physical Review E, 93 (2016), p. 042401.
[53] M. J. Saxton, *Single-particle tracking: the distribution of diffusion coefficients.*, Biophysical journal, 72 (1997), p. 1744.
[54] M. J. Saxton and K. Jacobson, *Single-particle tracking: applications to membrane dynamics*, Annual review of biophysics and biomolecular structure, 26 (1997), pp. 373–399.
[55] S. Shashkova and M. C. Leake, *Single-molecule fluorescence microscopy review: shedding new light on old problems*, Bioscience reports, 37 (2017), p. BSR20170031.
[56] S. Stallinga and B. Rieger, *Accuracy of the Gaussian point spread function model in 2D localization microscopy*, Optics express, 18 (2010), pp. 24461–24476.
[57] R. E. Thompson, D. R. Larson, and W. W. Webb, *Precise nanometer localization analysis for individual fluorescent probes*, Biophysical journal, 82 (2002), pp. 2775–2783.
[58] M. R. Vahid, B. Hanzon, and R. J. Ober, *Effect of Pixelation on the Parameter Estimation of Single Molecule Trajectories*, IEEE Transactions on Computational Imaging, 7 (2020), pp. 98–113.
[59] M. R. Vahid, B. Hanzon, and R. J. Ober, *Fisher information matrix for single molecules with stochastic trajectories*, SIAM Journal on Imaging Sciences, 13 (2020), pp. 234–264, https://doi.org/10.1137/19M1242562.
[60] Y. Wong, Z. Lin, and R. J. Ober, *Limit of the accuracy of parameter estimation for moving single molecules imaged by fluorescence microscopy*, IEEE Transactions on Signal Processing, 59 (2010), pp. 895–911.
[61] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, *Gaussian approximations of fluorescence microscope point-spread function models*, Applied optics, 46 (2007), pp. 1819–1829.