PARAMETER ESTIMATION METHODS FOR SINGLE MOLECULE DATA ANALYSIS

A Dissertation

by

MILAD RAFIEE VAHID

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Raimund J. Ober
Committee Members,	E. Sally Ward
	Alvin T. Yeh
	Vladislav V. Yakovlev
Head of Department,	Mike McShane

December 2019

Major Subject: Biomedical Engineering

ABSTRACT

The advent of single molecule microscopy enabled the study of subcellular processes at the single molecule level, which was not possible before through conventional microscopy. This work makes contributions to parameter estimation in two important applications of single molecule microscopy: single molecule super-resolution microscopy and single molecule tracking.

In single molecule super-resolution microscopy, images of subsets of stochastically photoactivated fluorophores are acquired over a sequence of frames. In order to construct a high-resolution image of the structures labeled by these fluorophores, the locations of the fluorophores are accurately estimated in each frame. In this study, we develop a novel state space-based fluorophore localization method from multi-emitter super-resolution images. An important property of our proposed method is its capability to distinguish very closely spaced molecules and estimate their locations accurately. Moreover, in contrast to most of the available methods that need prior knowledge about the number of emitting fluorophores in the image, our method determines this number using a procedure that utilizes a least-squares criterion.

Study of intercellular and intracellular trafficking processes of single molecules plays an important role in biological studies. Most of available single molecule trajectory parameter estimation methods are limited to Brownian motion and Gaussian measurement models.

Another contribution of this research is the development of a stochastic framework to calculate the maximum likelihood estimates of the biophysical parameters of biomolecular interactions, e.g., drift and diffusion coefficients, where the motion of an object of interest, e.g., single molecule, is modeled more generally by stochastic differential equations, and the measurements are the detected photons emitted by the moving fluorescently labeled object using both ideal unpixelated and practical pixelated detectors, such as charge-coupled device (CCD) and electron multiplying CCD (EMCCD) cameras.

More importantly, we develop a general framework to calculate the Cramér-Rao lower bound (CRLB), given by the inverse of the Fisher information matrix, for the estimation of unknown

parameters and use it as a benchmark in the evaluation of the standard deviation of the estimates.

DEDICATION

To my mother, my father and my wife.

ACKNOWLEDGMENTS

I would like to thank my supervisor, Dr. Raimund J. Ober, for the guidance, encouragement and advice he has provided throughout my time as his student. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. I would also like to extend my gratitude to Dr. E. Sally Ward for her support during my Ph.D. work. I would like to thank my committee members, Dr. Alvin T. Yeh and Dr. Vladislav V. Yakovlev, for taking the time to serve on my committee. I would also like to thank Dr. Bernard Hanzon for his valuable comments on the mathematical parts of my projects.

I would like to express my appreciation to all current and former members of the laboratory of Drs. E. Sally Ward and Raimund J. Ober. I have been very lucky to have a chance to work with a group of smart and dedicated researchers. I would especially like to acknowledge Dr. Jerry Chao for his many helpful discussions.

Finally, I would like to express my heartfelt gratitude to Shiva, my wife, for her continued support, encouragement and patience in all of the ups and downs of my research. Also, I would like to thank my parents and sisters for their patience and their belief in me. Their unconditional and unwavering support has allowed me to focus my time and energy on my Ph.D. research.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

Funding Sources

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE   Page

LIST OF TABLES

1. INTRODUCTION

The advent of optical microscopy has revolutionized biological investigations by providing a powerful tool for the observation of cellular structures at a high magnification that enables the revelation of many details of the structures that are not visible with the naked eye. Among all microscopy modalities, fluorescence microscopy has been of particular interest in biological research. In contrast to basic optical microscopy methods, in which light passes through a sample of interest and is captured by a detector, in fluorescence microscopy, the sample is labeled by fluorescent markers and illuminated by an excitation light with a specific wavelength. As a result, the fluorescent molecules emit light with a higher wavelength, which further will be captured by a detector and form an image of the sample. It allows biologists to observe specific molecules within their cellular context, which is not possible through other microscopy modalities.

During the recent past decades, the performance of fluorescence microscopy methods has been improved significantly due to the advances in technological developments. Examples include the developments of highly sensitive detectors, e.g., charge-coupled devices (CCDs), electron multiplying CCD (EMCCD) and scientific complementary metal-oxide semiconductors (sCMOS) cameras, high speed computers used to control imaging experiments and microscopy image analysis algorithms. Moreover, many fluorescent markers, e.g., fluorescent proteins (FPs), organic dyes and quantum dots, have been developed to label protein molecules in living cells. These advances enable the study of cellular processes at the single molecule level.

1.1 Single molecule microscopy

Traditionally, microscopy studies were bulk studies, i.e., the information from such studies came from several closely spaced molecules. Therefore, these studies reflected the behavior of ensembles of molecules as opposed to individual ones [1]. Single molecule microscopy, the Nobel prize-winning technique, has revolutionized the field of microscopy by providing quantitative information on the behavior of individual molecules in cellular environments, which were not

available before through bulk studies [2, 3, 4, 5, 6, 7]. Understanding the intracellular dynamics through different molecular pathways has been greatly benefited from the analysis of data acquired using single molecule microscopy techniques [8, 9]. There are many challenges that need to be addressed in the application of single molecule microscopy techniques. Here, we study two specific applications of these techniques: single molecule super-resolution microscopy and single molecule tracking.

Single molecule super-resolution microscopy enables imaging at sub-diffraction limit resolution by producing images of subsets of stochastically photoactivated fluorophores over a sequence of frames. In each frame of the sequence, the fluorophores are accurately localized, and the estimated locations are used to construct a high-resolution image of the cellular structures labeled by the fluorophores.

Another important application in the realm of single molecule data analysis is single molecule tracking, which plays an important role in the study of dynamic processes in living cells. Protein transport within and across cells is vital in maintaining normal cellular function. In biological studies, single molecule tracking methods have been used to study intracellular trafficking processes, which is very important to understanding the molecular basis of different types of diseases caused by the defects in these processes. In the following, we review the state-of-the-art methods proposed in the fields of single molecule super-resolution microscopy and single molecule tracking.

1.2   Single molecule super-resolution microscopy

Single molecule super-resolution methods have been successful at achieving sub-diffraction-limit resolution based on two key innovations: photoactivable fluorophores and powerful fluorophore localization algorithms [10]. In these methods, a fluorescently labeled cellular structure is imaged over a sequence of frames. In each frame, only a small number of stochastically photoactivated fluorophores are detected. In order to construct a high-resolution image of the cellular structure, the locations of individual emitting fluorophores are estimated with sub-pixel precision from each frame and used to re-render the structure. Many fluorophore localization methods are available, and they typically comprise the following separate steps: a detection step that identifies

2

the regions in the image that contain emitting fluorophores, and an estimation step that determines the locations of these fluorophores accurately. In recent years, several methods have been developed that use fitting-based algorithms to solve the estimation problem. The basis of most of these methods is to fit a point spread function (PSF) model to the acquired data and estimate the parameters of the model by minimizing the difference between the data and the model through an iterative approach. For example, in [11], a method was proposed that uses the maximum likelihood estimator to localize multiple emitters simultaneously within a two-dimensional (2D) fitting subregion. Similarly, the DAOSTORM algorithm [12] fits multiple PSFs in a recursive approach by analyzing pixel clusters in the residual image. In this algorithm, the fluorophore locations are determined by minimizing a least-squares criterion.

Fitting-based algorithms are not the only approaches used to solve the estimation problem for multi-emitter images. Many other localization methods have been developed that use non-fitting algorithms for the estimation problem. These methods are preferable when accurate PSF and noise models are not available. As an example, the QuickPALM software uses the simple centroid method [13, 14]. In this method, the fluorophore location is estimated as the average photon location, or centroid. However, the image background causes a systematic deviation in centroid-based methods. To solve the background bias problem in centroid-based methods, the virtual window center of mass (VWCM) method has been demonstrated to be a good background-corrected centroid estimator [15]. Although centroid methods are fast and computationally simple, their accuracy is not comparable to that of good fitting-based methods. Another important class of non-fitting algorithms has been developed based on sparse support recovery methods [16]. The compressive-sensing-based method CSSTORM [17], structured sparse model and Bayesian information criterion (SSC-BIC) [18], and fast localization algorithm based on a continuous space formulation (FALCON) [19] are well-known examples of such algorithms. Among them, CSSTORM has been shown to achieve accurate localization for emitter densities as high as 10 emitters/μm^2 [20]. In this method, a large-scale convex optimization problem needs to be solved in an iterative approach [21]. Huang et al. [20] have proposed a non-fitting algorithm by transferring the

molecule localization problem to the frequency domain. Their proposed algorithm is based on a 2D spectrum-estimation method called matrix enhancement and matrix pencil (MEMP) [22]. MEMP can provide a significant speed advantage over CSSTORM while retaining the same level of accuracy (it is 100 times faster than CSSTORM with l1-homotopy [20]). Huang et al., however, assume that the PSF can be approximated by a Gaussian function, which can be problematic in practice due to the fact that the Gaussian model is often not an accurate analytical PSF.

1.3 Single molecule tracking

The ability to track objects of interest, e.g., subcellular organelles and molecules, in cellular environments plays an important role in studying biological systems. In particular, single molecule tracking, which enables following subcellular processes at the single molecule level, has become a vital tool in cell biology [23, 24, 25, 26]. In biological studies, single molecule tracking methods have been used to study the intracellular trafficking of fluorescently labeled antibodies, e.g., prostate-specific membrane antigen (PSMA) antibodies [27, 28], by analyzing the velocity and path of the fluorescent molecules.

In general, the motion of an object in cellular environments is subject to different types of forces, e.g., deterministic forces due to the environment and random forces due to random collisions with other objects [29, 30]. It has been shown that the motion of a moving object in such environments can be modeled by stochastic differential equations (SDEs) [31]. In particular, in many biological applications, solutions of linear SDEs are good fits to experimental single molecule trajectories [32, 33, 34]. In a basic fluorescence microscope, a fluorescently labeled object of interest is imaged by a detector which detects the photons emitted by the object during the acquisition time. Since the detection process of the emitted photons is inherently a random phenomenon, the acquired measurements are stochastic in nature. These measurements, according to the optical diffraction theory, can be modeled by different distributions. For example, a typical distribution for an in-focus molecule is an Airy profile [35], whereas, classical Born and Wolf profiles [36] are used instead for out-of-focus molecules. In some cases, it is possible and computationally beneficial to approximate these complex profiles with simple Gaussian models [37].

4

In many dynamical systems, the time points of the measurements are assumed to be equidistant. However, the time points of detection of the photons correspond to the arrival times of a Poisson process [6, 7]. This gives rise to the non-uniform sampling of the continuous-time stochastic process that describes the motion of the object. Since the parameters of the motion model of the object are highly time-dependent, this randomized non-uniform sampling causes significant fluctuations in the values of these parameters.

In recent years, many methods have been developed to analyze the trajectories of a molecule in cellular environments. In most of these methods, the model for the motion of the molecule is assumed to be limited to a Brownian motion (pure diffusion) model described only by the diffusion coefficient, and only a small number of the available methods consider more general motion models. The methods developed to analyze pure diffusion models are mostly based on the mean square displacement approach [38], in which the diffusion coefficient is estimated by a linear regression of the mean square displacement of the Gaussian distributed observed locations of the molecule as a function of the time lag [39, 40, 41]. Mean square displacement-based methods are not the only approaches used to estimate the diffusion coefficient from a set of measurements. For example, Relich et al. [42] have proposed a method for the maximum likelihood estimation of the diffusion coefficient, with an information-based confidence interval, from Gaussian measurements. In all of these methods, the motion of a molecule is assumed as a pure diffusion model, and the measurements are modeled by independent and identically distributed Gaussian random variables [26].

However, in general, the motion of a molecule is not limited to the pure diffusion model, and the diffusion coefficient is only one of the parameters that play a role in the motion of the molecule. Also, the Gaussian assumption for the measurements is problematic in practice due to the fact that the Gaussian model is often not an accurate analytical model. In [43], Ashley and Andersson have proposed a simultaneous localization and parameter estimation algorithm for more complex motion models, such as confined [31] and tethered motions [44], which employs the expectation maximization algorithm, in conjunction with sequential Monte Carlo methods [45].

For the general object tracking problem, in [46, 47], a sequential Monte Carlo method has been developed for the parameter estimation from nonlinear non-Gaussian state-space models. Briane et al. [30] have developed a method for classifying the object trajectories in living cells into three types of diffusion: Brownian motion, subdiffusion (diffusion in a closed domain or a crowded area) and superdiffusion (diffusion in a specific direction). In [32, 33, 34], the motion of a moving object has been described more generally by a linear SDE, and the parameters of the model have been estimated using a maximum likelihood estimation method. However, they do not consider the randomness of the time points at which the measurements occur. Their proposed framework also does not allow for non-Gaussian measurements.

1.4 Overview of the thesis

Estimation of the parameters of stochastic models that describe the image formation of single molecules and their trajectories in cellular environments from microscopy data is a critical task in the analysis of single molecule data. In summary, We propose (1) a novel location estimation method from super-resolution single molecule microscopy image data, and (2) a stochastic framework in which we calculate the maximum likelihood estimates of the biophysical parameters of the molecular interactions, e.g., diffusion and drift coefficients, from single molecule trajectory data. More importantly, we develop a general framework to calculate the Cramér-Rao lower bound (CRLB), given by the inverse of the Fisher information matrix, for the estimation of unknown parameters and use it as a benchmark in the evaluation of the standard deviation of the estimates.

The organization of the dissertation is as follows. Chapter 2 is devoted to introducing a state space algorithm for single molecule localization in a non-iterative manner. Generally, a single molecule fluorescence image contains multiple peaks of intensity that correspond to emitting fluorophores. Our solution is to model such an image by the frequency response of a multi-order system, as the locations of the poles of such a system determine the peak locations in the frequency domain. To realize this localization algorithm, we take advantage of the balanced state space realization algorithm used in [48, 49, 50] for the reduction of noise in fluorescence microscopy images. This realization algorithm is based on the singular value decomposition (SVD)

of a Hankel matrix. To associate the peak locations in the image with the poles of the underlying system, we apply this realization algorithm to the inverse Fourier transform of the image rather than to the image itself. In our algorithm, the number of emitting fluorophores, which correspond to the most significant peaks in the image, is ultimately determined using a procedure that utilizes a least-squares criterion. Our algorithm also allows us to derive a theoretical reconstruction of the image. A reconstructed image is an image that looks similar to the original image, but is specified analytically in terms of the state space parameters of the system calculated using our proposed localization algorithm.

In Chapter 3, we introduce a general state space system to model the motion of a fluorescently labeled molecule in cellular environments. We assume that we have an ideal and unpixelated detector and the measurements are the locations and time points of the detected photons emitted by the molecule, referred to as the fundamental data model. Our model allows for Poisson time points and arbitrary distributed, e.g, Airy and Born and Wolf, locations. We then develop a stochastic framework in which we calculate the maximum likelihood estimates of the parameters of the model that describes the motion of the molecule in cellular environments.

Chapter 4 discusses the performance of the proposed estimation method in Chapter 3, in terms of the standard deviation of the estimates. For the purpose of evaluating the performance of our proposed estimation method, we develop a general framework to calculate the Fisher information matrix of the unknown parameters of the general motion model. There are some cases in which Gaussian approximations of measurements are very useful due to, for example, the ability to use computationally efficient algorithms in linear systems or the Kalman filter formulae. In particular, for Gaussian measurements, we calculate the Fisher information matrix by taking advantage of its relationship with the Kalman filter formula through a computationally efficient algorithm. To the best of our knowledge, even for Gaussian measurements, there currently exists no systematic methodology to evaluate the standard deviations of the estimates using the CRLB for the general motion model considered here.

Chapter 5 deals with the effect of pixelation of the detector on the parameter estimation of

single molecule trajectories and Fisher information matrix. In Chapters 3 and 4, we focused on the fundamental microscopy data model, in which the image of a molecule is acquired by an unpixelated detector. However, in practice, pixelated detectors, e.g., CCD cameras, are commonly used for acquiring the image of the fluorescently labeled molecule. In this chapter, we extend our previous results obtained from unpixelated detectors to pixelated detectors.

Finally, a summary of the thesis and conclusions are provided in Chapter 6.

2.   A STATE SPACE APPROACH TO SINGLE MOLECULE LOCALIZATION IN FLUORESCENCE MICROSCOPY*

2.1   Introduction

Single molecule super-resolution microscopy enables imaging at sub-diffraction-limit resolution by producing images of subsets of stochastically photoactivated fluorophores over a sequence of frames. In each frame of the sequence, the fluorophores are accurately localized, and the estimated locations are used to construct a high-resolution image of the cellular structures labeled by the fluorophores. Many methods have been developed for localizing fluorophores from the images. The majority of these methods comprise two separate steps: detection and estimation. In the detection step, fluorophores are identified. In the estimation step, the locations of the identified fluorophores are estimated through an iterative approach.

In this chapter, we propose a non-iterative state space-based localization method which combines the detection and estimation steps. Generally, a single molecule fluorescence image contains multiple peaks of intensity that correspond to emitting fluorophores. Our solution is to model such an image by the frequency response of a multi-order system, as the locations of the poles of such a system determine the peak locations in the frequency domain. To realize this localization algorithm, we take advantage of the balanced state space realization algorithm used in [48, 49, 50] for the reduction of noise in fluorescence microscopy images. This realization algorithm is based on the singular value decomposition of a Hankel matrix. To associate the peak locations in the image with the poles of the underlying system, we apply this realization algorithm to the inverse Fourier transform of the image rather than to the image itself. In our algorithm, the number of emitting fluorophores, which correspond to the most significant peaks in the image, is ultimately determined using a procedure that utilizes a least-squares criterion. Our algorithm also allows us to derive a theoretical reconstruction of the image. A reconstructed image is an image that looks

similar to the original image, but is specified analytically in terms of the state space parameters of the system calculated using our proposed localization algorithm.

Note that the realization algorithm from [50] was developed for the reduction of noise in a three-dimensional (3D) data set comprising a z-stack of microscopy images. Here, we apply a 2D version of that realization algorithm because the super-resolution microscopy data sets for which we develop our localization algorithm consist of 2D images that are analyzed independently of one another. Our localization algorithm, however, can be extended to the 3D localization of fluorescent emitters from a z-stack by simply applying the 3D version of the realization algorithm as presented in [50]. In the following, we first show the existence of minimal and asymptotically stable systems that realize a 2D image in the frequency domain.

2.2 System identification using frequency measurements

In this section, we show the existence of minimal and asymptotically stable systems that realize a finite 2D sequence in the frequency domain. We begin by demonstrating the existence of minimal and asymptotically stable systems for finite one-dimensional (1D) sequences in Lemma 1, using a subspace-based method similar to that described in [51], and then extend the result to two dimensions in Theorem 1. The basis of Lemma 1 is given by Proposition 1, which states that a finite 1D data set can be expressed as the impulse response of a minimal and asymptotically stable system [50]. Note that the stability of subspace methods was analyzed previously by Maciejowski in [52], where similar results were reported.

Proposition 1. *For positive integer N, let $X(n) \in \mathbb{C}^{p \times m}, p, m \in \mathbb{N}, n = 1, 2, ..., N$, be a 1D matrix-valued sequence. Then, there exists a minimal and asymptotically stable system (A, B, C), such that*

$$X(n) = CA^{n-1}B, \quad n = 1, 2, ..., N. \tag{2.1}$$

Proposition 1 enables us to write the following lemma, which shows the existence of a minimal and asymptotically stable system that realizes a finite 1D sequence in the frequency domain.

Lemma 1. *Let $\tilde{X}(k) \in \mathbb{R}, k = 1, 2, ..., N$, be a finite 1D sequence. For $n = 1, 2, ..., N$, let*

10

$X(n) := (IDFT(\tilde{X}))(n) = \frac{1}{N}\sum_{k=1}^{N}\tilde{X}(k)e^{i2\pi kn/N}$ *be the inverse discrete Fourier transform (inverse DFT, or IDFT) of* $\tilde{X}$. *Then, there exists a minimal and asymptotically stable system* (A, B, C), *such that*

$$X(n) = CA^{n-1}B, \quad n = 1, 2, ..., N. \tag{2.2}$$

Moreover,

$$\tilde{X}(k) = \tilde{C}(e^{i2\pi k/N}I - \tilde{A})^{-1}\tilde{B}, \quad k = 1, 2, ..., N, \tag{2.3}$$

where $\tilde{A} := A, \tilde{B} := (I - A^N)B, \tilde{C} = C$. *If* $A^N = 0$, *then* $(\tilde{A}, \tilde{B}, \tilde{C}) = (A, B, C)$.

Proof. Let $\tilde{X}(k) \in \mathbb{R}, k = 1, 2, ..., N$, be a finite 1D sequence. Let

$$X(n) := (IDFT(\tilde{X}))(n) = \frac{1}{N}\sum_{k=1}^{N}\tilde{X}(k)e^{i2\pi kn/N}, \quad n = 1, 2, ..., N, \tag{2.4}$$

be the IDFT of $\tilde{X}$. Then, according to Proposition 1, there exists a minimal and asymptotically stable system (A, B, C), such that

$$X(n) = CA^{n-1}B, \quad n = 1, 2, ..., N. \tag{2.5}$$

According to Eqs. (2.4) and (2.5), we then have, for $k = 1, 2, ..., N$,

$$
\begin{aligned}
\tilde{X}(k) &= (DFT(X))(k) \\
&= \sum_{n=1}^{N} X(n)e^{-i2\pi kn/N} \\
&= CBe^{-i2\pi k/N} + CABe^{-i4\pi k/N} + \cdots + CA^{N-1}Be^{-i2\pi kN/N} \\
&= Ce^{-i2\pi k/N}\left(I + Ae^{-i2\pi k/N} + \cdots + A^{N-1}e^{-i2\pi k(N-1)/N}\right)B \\
&= Ce^{-i2\pi k/N}\left[\sum_{n=0}^{N-1}\left(Ae^{-i2\pi kn/N}\right)^n\right]B.
\end{aligned}
\tag{2.6}
$$

For a square matrix $T \in \mathbb{C}^{m \times m}, m \in \mathbb{N}$, where the number 1 is not an eigenvalue of T, we have

11

the identity $\sum_{n=0}^{N-1} T^n = (I - T)^{-1}(I - T^N)$. Then, since the realization (A, B, C) is asymptotically stable, i.e., $|\lambda(A)| < 1$ holds for any eigenvalue $\lambda(A)$ of A, the number 1 is not an eigenvalue of $Ae^{-i2\pi k/N}, k = 1, ..., N$ (or equivalently, $I - Ae^{-i2\pi k/N}, k = 1, ..., N$, is invertible), and $\sum_{n=0}^{N-1} \left(Ae^{-i2\pi k/N}\right)^n = (I - Ae^{-i2\pi k/N})^{-1}(I - A^N)$. Substituting this expression into Eq. (2.6), we have, for $k = 1, 2, ..., N$,

$$
\begin{aligned}
\tilde{X}(k) &= Ce^{-i2\pi k/N}(I - Ae^{-i2\pi k/N})^{-1}(I - A^N)B \\
&= C(e^{i2\pi k/N}I - A)^{-1}(I - A^N)B \\
&= \tilde{C}(e^{i2\pi k/N}I - \tilde{A})^{-1}\tilde{B},
\end{aligned}
\tag{2.7}
$$

where $\tilde{A} := A, \tilde{B} := (I - A^N)B, \tilde{C} = C$. If $A^N = 0$, then $(\tilde{A}, \tilde{B}, \tilde{C}) = (A, B, C)$. $\qquad\square$

In the following theorem, we extend the results obtained for 1D sequences to 2D sequences.

Theorem 1. *Let $\tilde{X}(k_1, k_2) \in \mathbb{R}, k_i = 1, 2, ..., N_i, i = 1, 2$, be a finite 2D sequence. For $n_i = 1, 2, ..., N_i, i = 1, 2$, let*

$$
X(n_1, n_2) := (IDFT_{2D}(\tilde{X}))(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1=1}^{N_1} \sum_{k_2=1}^{N_2} \tilde{X}(k_1, k_2)e^{i2\pi(k_1 n_1/N_1 + k_2 n_2/N_2)}, \tag{2.8}
$$

be the inverse 2D DFT of $\tilde{X}$. Then, there exist minimal and asymptotically stable systems $(A_i, B_i, C_i), i = 1, 2$, such that

$$
X(n_1, n_2) = X_1(n_1)X_2(n_2), \quad n_i = 1, 2, ..., N_i, \quad i = 1, 2, \tag{2.9}
$$

where, for $i = 1, 2$,

$$
X_i(n_i) := C_i A_i^{n_i - 1} B_i, \quad n_i = 1, 2, ..., N_i. \tag{2.10}
$$

Moreover,

$$
\tilde{X}(k_1, k_2) = \tilde{X}_1(k_1)\tilde{X}_2(k_2), \quad k_i = 1, 2, ..., N_i, \quad i = 1, 2, \tag{2.11}
$$

where, for $k_j = 1, 2, ..., N_j, j = 1, 2,$

$$\tilde{X}_j(k_j) := \tilde{C}_j(e^{i2\pi k_j/N_j}I - \tilde{A}_j)^{-1}\tilde{B}_j, \tag{2.12}$$

where $\tilde{A}_j := A_j, \tilde{B}_j := (I - A_j^{N_j})B_j, \tilde{C}_j := C_j.$ *For* $j = 1, 2,$ *if* $A_j^{N_j} = 0,$ *then* $(\tilde{A}_j, \tilde{B}_j, \tilde{C}_j) = (A_j, B_j, C_j).$

Proof. Let $\tilde{X}(k_1, k_2) \in \mathbb{R}, k_i = 1, 2, ..., N_i, i = 1, 2,$ be a finite 2D sequence. For $n_i = 1, ..., N_i, i = 1, 2,$ let

$$X(n_1, n_2) := (IDFT_{2D}(\tilde{X}))(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1=1}^{N_1} \sum_{k_2=1}^{N_2} \tilde{X}(k_1, k_2) e^{i2\pi(k_1 n_1/N_1 + k_2 n_2/N_2)}, \tag{2.13}$$

be the inverse 2D DFT of $\tilde{X}$. Arrange the entries of X to form a matrix Q as

$$Q := \begin{bmatrix} X(1,1) & X(1,2) & \cdots & X(1,N_2) \\ X(2,1) & X(2,2) & \cdots & X(2,N_2) \\ \vdots & \vdots & \ddots & \vdots \\ X(N_1,1) & X(N_1,2) & \cdots & X(N_1,N_2) \end{bmatrix}. \tag{2.14}$$

Decompose Q via SVD as $Q = U\Sigma V$, where for $r \in \mathbb{N}, U \in \mathbb{C}^{N_1 \times r}, \Sigma \in \mathbb{C}^{r \times r}$ and $V \in \mathbb{C}^{r \times N_2}$. For $n_i = 1, 2, ..., N_i, i = 1, 2,$ define $X_1(n_1) \in \mathbb{C}^{1 \times r}$ and $X_2(n_2) \in \mathbb{C}^{r \times 1}$, such that

$$\begin{bmatrix} X_1(1) \\ X_1(2) \\ \vdots \\ X_1(N_1) \end{bmatrix} := U\Sigma^{1/2}, \quad \begin{bmatrix} X_2(1) & X_2(2) & \cdots & X_2(N_2) \end{bmatrix} := \Sigma^{1/2}V. \tag{2.15}$$

Then

$$X(n_1, n_2) = X_1(n_1)X_2(n_2), \quad n_i = 1, 2, ..., N_i, \quad i = 1, 2. \tag{2.16}$$

Moreover, according to Proposition 1, there exist minimal and asymptotically stable systems $(A_i, B_i, C_i), i = 1, 2$, such that, for $i = 1, 2$,

$$X_i(n_i) = C_i A_i^{n_i-1} B_i, \quad n_i = 1, 2, ..., N_i. \tag{2.17}$$

According to Eqs. (2.13) and (2.16),

$$\begin{aligned}
\tilde{X}(k_1, k_2) &= (DFT_{2D}(X))(k_1, k_2) \\
&= \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} X(n_1, n_2) e^{-i2\pi(k_1 n_1/N_1 + k_2 n_2/N_2)} \\
&= \left(\sum_{n_1=1}^{N_1} X_1(n_1) e^{-i2\pi k_1 n_1/N_1} \right) \left(\sum_{n_2=1}^{N_2} X_2(n_2) e^{-i2\pi k_2 n_2/N_2} \right) \\
&= \tilde{X}_1(k_1) \tilde{X}_2(k_2), \quad k_i = 1, 2, ..., N_i, \quad i = 1, 2,
\end{aligned} \tag{2.18}$$

where $\tilde{X}_i(k_i) := (DFT(X_i))(k_i), k_i = 1, 2, ..., N_i, i = 1, 2$. Then, according to Lemma 1, for $k_j = 1, 2, ,, ..., N_j, j = 1, 2$,

$$\tilde{X}_j(k_j) := \tilde{C}_j (e^{i2\pi k_j/N_j} I - \tilde{A}_j)^{-1} \tilde{B}_j, \tag{2.19}$$

where $\tilde{A}_j := A_j, \tilde{B}_j := (I - A_j^{N_j}) B_j, \tilde{C}_j := C_j$. For $j = 1, 2$, if $A_j^{N_j} = 0$, then $(\tilde{A}_j, \tilde{B}_j, \tilde{C}_j) = (A_j, B_j, C_j)$. $\square$

2.3 Location estimation

So far, we have shown the existence of minimal and asymptotically stable systems (A_i, B_i, C_i), $i = 1, 2$, that realize a finite 2D sequence $\tilde{X}(k_1, k_2) \in \mathbb{R}, k_i = 1, 2, ..., N_i, i = 1, 2$, in the frequency domain. Here, we summarize our overall localization approach. Given that $\tilde{X}$ is a single molecule image with multiple peaks of intensity, we determine the locations of the molecules by calculating the pole locations of $(A_i, B_i, C_i), i = 1, 2$.

In Theorem 1, we have shown that there exist minimal and asymptotically stable systems

$(A_i, B_i, C_i), i = 1, 2$, such that

$$\tilde{X}(k_1, k_2) = \tilde{X}_1(k_1)\tilde{X}_2(k_2), \quad k_i = 1, 2, ..., N_i, \quad i = 1, 2, \tag{2.20}$$

where, for $k_j = 1, 2, ..., N_j, j = 1, 2$,

$$\tilde{X}_j(k_j) := \tilde{C}_j(e^{i2\pi k_j/N_j}I - \tilde{A}_j)^{-1}\tilde{B}_j, \tag{2.21}$$

where $\tilde{A}_j := A_j$, $\tilde{B}_j := (I - A_j^{N_j})B_j$, and $\tilde{C}_j := C_j$. For $j = 1, 2$, if $A_j^{N_j} = 0$, then $(\tilde{A}_j, \tilde{B}_j, \tilde{C}_j) = (A_j, B_j, C_j)$. If we diagonalize $A_i, i = 1, 2$, then the diagonal elements of the resulting matrix $\bar{A}_i$ give the poles of the system. In the following, we use matrix diagonalization in Eq. (2.20) to express $\tilde{X}$ in terms of the poles of the system.

For $s_1, s_2 \in \mathbb{N}$ and $A_i \in \mathbb{C}^{s_i \times s_i}, i = 1, 2$, which are diagonalizable, i.e., for $t_i = 1, 2, ..., s_i, i = 1, 2$, and some invertible $T_i \in \mathbb{C}^{s_i \times s_i}$, we have the diagonal matrix $\bar{A}_i := T_iA_iT_i^{-1} = \text{diag}(a_1^i, \cdots, a_{s_i}^i), a_{t_i}^i \in \mathbb{C}$, then with $\bar{B}_i := T_iB_i = \left[b_1^i, \cdots, b_{s_i}^i\right]^T, \bar{C}_i := C_iT_i^{-1} = \left[c_1^i, \cdots, c_{s_i}^i\right], i = 1, 2$, where $b_{t_1}^1 \in \mathbb{C}^{1 \times r}, b_{t_2}^2 \in \mathbb{C}, c_{t_1}^1 \in \mathbb{C}, c_{t_2}^2 \in \mathbb{C}^{r \times 1}, t_i = 1, 2, ..., s_i, i = 1, 2$, for $k_j = 1, 2, ..., N_j, j = 1, 2$, we can write $\tilde{X}$ in terms of the poles of the system as

$$\tilde{X}(k_1, k_2) = \prod_{j=1}^2 \bar{C}_j(e^{i2\pi k_j/N_j}I - \bar{A}_j)^{-1}\bar{B}_j$$

$$= \bar{C}_1 \begin{bmatrix} \frac{b_1^1 c_1^2}{(e^{i2\pi k_1/N_1}-a_1^1)(e^{i2\pi k_2/N_2}-a_1^2)} & \frac{b_1^1 c_2^2}{(e^{i2\pi k_1/N_1}-a_1^1)(e^{i2\pi k_2/N_2}-a_2^2)} & \cdots & \frac{b_1^1 c_{s_2}^2}{(e^{i2\pi k_1/N_1}-a_1^1)(e^{i2\pi k_2/N_2}-a_{s_2}^2)} \\ \frac{b_2^1 c_1^2}{(e^{i2\pi k_1/N_1}-a_2^1)(e^{i2\pi k_2/N_2}-a_1^2)} & \frac{b_2^1 c_2^2}{(e^{i2\pi k_1/N_1}-a_2^1)(e^{i2\pi k_2/N_2}-a_2^2)} & \cdots & \frac{b_2^1 c_{s_2}^2}{(e^{i2\pi k_1/N_1}-a_2^1)(e^{i2\pi k_2/N_2}-a_{s_2}^2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b_{s_1}^1 c_1^2}{(e^{i2\pi k_1/N_1}-a_{s_1}^1)(e^{i2\pi k_2/N_2}-a_1^2)} & \frac{b_{s_1}^1 c_2^2}{(e^{i2\pi k_1/N_1}-a_{s_1}^1)(e^{i2\pi k_2/N_2}-a_2^2)} & \cdots & \frac{b_{s_1}^1 c_{s_2}^2}{(e^{i2\pi k_1/N_1}-a_{s_1}^1)(e^{i2\pi k_2/N_2}-a_{s_2}^2)} \end{bmatrix} \bar{B}_2$$

$$= \sum_{l=1}^{s_1}\sum_{j=1}^{s_2} \frac{c_l^1 b_l^1 c_j^2 b_j^2}{(e^{i2\pi k_1/N_1} - a_l^1)(e^{i2\pi k_2/N_2} - a_j^2)}. \tag{2.22}$$

Equation (2.22) provides an analytical expression for the reconstructed image, in which the poles of $\tilde{X}$ occur at $(a_{t_1}^1, a_{t_2}^2), t_i = 1, ..., s_i, i = 1, 2$. (Note that for $t_i = 1, ..., s_i, i = 1, 2$, if the row $b_{t_1}^1$ of $\bar{B}_1$ and the column $c_{t_2}^2$ of $\bar{C}_2$ are perpendicular, i.e., $b_{t_1}^1 c_{t_2}^2 = 0$, then the residual of the pole $(a_{t_1}^1, a_{t_2}^2)$ is equal to zero, and $(a_{t_1}^1, a_{t_2}^2)$ is not associated with a peak in the corresponding 2D

15

image.)

We next obtain the locations of the molecules in the object space in terms of the phase of the calculated poles. Let the 2D sequence $\tilde{X}$ denote the pixel intensities of our $N_1 \times N_2$ image with pixel width Δx and pixel height Δy, obtained by sampling the image at the center of each pixel. Assume $a_{t_j}^j = \left| a_{t_j}^j \right| e^{i w_{t_j}^j}, 0 \leq w_{t_j}^j \leq 2\pi, t_j = 1, ..., s_j, j = 1, 2$. Then, by linearly mapping a $2\pi \times 2\pi$ square region in the frequency domain to the region with area $N_1 \times N_2$ pixels in the image space (between the center of the first pixel and the center of the last pixel) and converting from image space units to object space units, the set containing the peak locations (i.e., the molecule locations) in the object space is given by $\left\{ (x_{t_2}, y_{t_1}) : c_{t_1}^1 b_{t_1}^1 c_{t_2}^2 b_{t_2}^2 \neq 0, t_i = 1, ..., s_i, i = 1, 2 \right\}$, where

$$x_{t_2} := \frac{\Delta x w_{t_2}^2 N_1}{2M\pi} + \frac{\Delta x}{2M}, \quad y_{t_1} := \frac{\Delta y w_{t_1}^1 N_2}{2M\pi} + \frac{\Delta y}{2M}, \tag{2.23}$$

and $M > 0$ denotes the lateral magnification of the microscope system.

2.4 Algorithm

We now explain our proposed approach in more detail. In Section 2.3, we have calculated the poles of a 2D single molecule image $\tilde{X}$ in terms of the elements of minimal and asymptotically stable systems $(A_i, B_i, C_i), i = 1, 2$. Here, using the balanced state space realization algorithm introduced by Maciejowski [52], we propose a step-by-step algorithm to calculate systems $(A_i, B_i, C_i), i = 1, 2$, that realize $\tilde{X}$, and to determine the locations of the single molecules using the realization.

Algorithm 1. *Let $\tilde{X}(k_1, k_2) \in \mathbb{R}, k_i = 1, 2, ..., N_i, i = 1, 2$, represent the acquired image data.*

I. Subtract an estimated background level $\hat{\beta}$, e.g., the average of the data points near the boundary of the image data $\tilde{X}$, from the image data $\tilde{X}$, and define the background-subtracted image $\tilde{X}_{bs}$ as

$$\tilde{X}_{bs}(k_1, k_2) := \tilde{X}(k_1, k_2) - \hat{\beta}, \quad k_i = 1, 2, ..., N_i, \quad i = 1, 2. \tag{2.24}$$

II. Let X be the 2D IDFT of $\tilde{X}_{bs}$, i.e.,

$$X(n_1, n_2) := (IDFT_{2D}(\tilde{X}_{bs}))(n_1, n_2), \quad n_i = 1, 2, ..., N_i, \quad i = 1, 2. \tag{2.25}$$

III. Arrange the entries of X to form a matrix Q as

$$Q := \begin{bmatrix} X(1,1) & X(1,2) & \cdots & X(1, N_2) \\ X(2,1) & X(2,2) & \cdots & X(2, N_2) \\ \vdots & \vdots & \ddots & \vdots \\ X(N_1, 1) & X(N_1, 2) & \cdots & X(N_1, N_2) \end{bmatrix}. \tag{2.26}$$

Decompose Q via SVD as $Q = U\Sigma V$. Let the positive integer $r \leq K, K = \min(N_1, N_2)$, denote the number of retained singular values (see Section 2.4.1). Partition $\Sigma = diag(\hat{\Sigma}, \hat{\hat{\Sigma}}), \hat{\Sigma} \in \mathbb{C}^{r \times r}, U = \begin{bmatrix} \hat{U} & \hat{\hat{U}} \end{bmatrix}, \hat{U} \in \mathbb{C}^{N_1 \times r}, and V = \begin{bmatrix} \hat{V} & \hat{\hat{V}} \end{bmatrix}^T, \hat{V} \in \mathbb{C}^{r \times N_2}. For $n_i = 1, 2, ..., N_i, i = 1, 2$, define $X_1^r(n_1) \in \mathbb{C}^{1 \times r}$ and $X_2^r(n_2) \in \mathbb{C}^{r \times 1}$, such that

$$\begin{bmatrix} X_1^r(1) \\ X_1^r(2) \\ \vdots \\ X_1^r(N_1) \end{bmatrix} := \hat{U}\hat{\Sigma}^{1/2}, \quad \begin{bmatrix} X_2^r(1) & X_2^r(2) & \cdots & X_2^r(N_2) \end{bmatrix} := \hat{\Sigma}^{1/2}\hat{V}. \tag{2.27}$$

IV. Construct the Hankel matrices $H_1 \in \mathbb{C}^{(N_1+1) \times (N_1+1)r}, H_2 \in \mathbb{C}^{(N_2+1)r \times (N_2+1)}$ as

$$H_i := \begin{bmatrix} X_i^r(1) & X_i^r(2) & \cdots & X_i^r(N_i - 1) & X_i^r(N_i) & 0 \\ X_i^r(2) & X_i^r(3) & \cdots & X_i^r(N_i) & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ X_i^r(N_i) & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}, \quad i = 1, 2, \tag{2.28}$$

where 0 denotes a block of zeros of the corresponding size. For $i = 1, 2$, decompose H_i via

SVD as $H_i = U_i \Sigma_i V_i$. Let the positive integers $s_i \leq N_i, i = 1, 2$, denote the numbers of retained singular values in the respective SVDs (see Section 2.4.2). For $i = 1, 2$, partition $\Sigma_i = diag(\hat{\Sigma}_i, \hat{\hat{\Sigma}}_i), \hat{\Sigma}_i \in \mathbb{C}^{s_i \times s_i}, U_i = \begin{bmatrix} \hat{U}_i & \hat{\hat{U}}_i \end{bmatrix}, \hat{U}_1 \in \mathbb{C}^{(N_1+1) \times s_1}, \hat{U}_2 \in \mathbb{C}^{(N_2+1)r \times s_2}$, and $V_i = \begin{bmatrix} \hat{V}_i & \hat{\hat{V}}_i \end{bmatrix}^T, \hat{V}_1 \in \mathbb{C}^{s_1 \times (N_1+1)r}, \hat{V}_2 \in \mathbb{C}^{s_2 \times (N_2+1)}$, conformally. Let $C_1^{r;s_1} \in \mathbb{C}^{1 \times s_1}$ and $C_2^{r;s_2} \in \mathbb{C}^{r \times s_2}$ be the first row of $\hat{U}_1 \hat{\Sigma}_1^{1/2}$ and the first r rows of $\hat{U}_2 \hat{\Sigma}_2^{1/2}$, respectively. Also, let $B_1^{r;s_1} \in \mathbb{C}^{s_1 \times r}$ and $B_2^{r;s_2} \in \mathbb{C}^{s_2 \times 1}$ be the first r columns of $\hat{\Sigma}_1^{1/2} \hat{V}_1$ and the first column of $\hat{\Sigma}_2^{1/2} \hat{V}_2$, respectively. Assuming

$$\hat{U}_i = \begin{bmatrix} \bar{U}_1^i \\ \vdots \\ \bar{U}_{N_i}^i \\ \bar{U}_{N_i+1}^i \end{bmatrix}, \quad i = 1, 2, \tag{2.29}$$

where $\bar{U}_{n_1}^1 \in \mathbb{C}^{1 \times s_1}, \bar{U}_{n_2}^2 \in \mathbb{C}^{r \times s_2}, n_i = 1, ..., N_i + 1, i = 1, 2$, define

$$\hat{U}_i^{\uparrow} := \begin{bmatrix} \bar{U}_2^i \\ \vdots \\ \bar{U}_{N_i+1}^i \end{bmatrix}, \quad \hat{U}_i^{\downarrow} := \begin{bmatrix} \bar{U}_1^i \\ \vdots \\ \bar{U}_{N_i}^i \end{bmatrix}, \quad i = 1, 2. \tag{2.30}$$

Then, let $A_i^{r;s_i} = \hat{\Sigma}_i^{-1/2} \hat{U}_i^{\downarrow *} \hat{U}_i^{\uparrow} \hat{\Sigma}_i^{1/2} \in \mathbb{C}^{s_i \times s_i}, i = 1, 2$.

V. Diagonalize $A_j^{r;s_j} \in \mathbb{C}^{s_j \times s_j}, j = 1, 2$, i.e., for $t_j = 1, 2, ..., s_j, j = 1, 2$, and some invertible $T_j \in \mathbb{C}^{s_j \times s_j}$, let $\bar{A}_j^{r;s_j} := T_j A_j^{r;s_j} T_j^{-1} = diag(a_1^j, \cdots, a_{s_j}^j), a_{t_j}^j = \left| a_{t_j}^j \right| e^{iw_{t_j}^j} \in \mathbb{C}, 0 \leq w_{t_j}^j \leq 2\pi$, be a corresponding diagonal matrix for $A_j^{r;s_j}$. Also, let $\bar{B}_j^{r;s_j} := T_j B_j^{r;s_j} = \begin{bmatrix} b_1^j, \cdots, b_{s_j}^j \end{bmatrix}^T, \bar{C}_j^{r;s_j} := C_j^{r;s_j} T_j^{-1} = \begin{bmatrix} c_1^j, \cdots, c_{s_j}^j \end{bmatrix}, j = 1, 2$, where $b_{t_1}^1 \in \mathbb{C}^{1 \times r}, b_{t_2}^2 \in \mathbb{C}, c_{t_1}^1 \in \mathbb{C}, c_{t_2}^2 \in \mathbb{C}^{r \times 1}, t_j = 1, 2, ..., s_j, j = 1, 2$.

Note that in theory, there is a possibility that $A_1^{r;s_1}$ and/or $A_2^{r;s_2}$ are not diagonalizable. In practice, however, because the diagonalization is numerically computed, $A_1^{r;s_1}$ and/or $A_2^{r;s_2}$ can be expected to be diagonalizable. In the unlikely scenario where $A_1^{r;s_1}$ and/or $A_2^{r;s_2}$ are not diagonal-

izable, very small perturbations of the data can be introduced to alter slightly their eigenvalues and make them diagonalizable. A perturbation can be achieved, for example, by simply adding a very small value to a pixel of the image.

VI. For $h = \min(s_1, s_2)$, calculate, in the object space, the estimated peak locations (x_k, y_k), $x_k \in \{\hat{x}_1, ..., \hat{x}_{s_2}\}, y_k \in \{\hat{y}_1, ..., \hat{y}_{s_1}\}, k = 1, ..., h$, where

$$\hat{x}_{t_2} := \frac{\Delta x w_{t_2}^2 N_1}{2M\pi} + \frac{\Delta x}{2M}, \quad \hat{y}_{t_1} := \frac{\Delta y w_{t_1}^1 N_2}{2M\pi} + \frac{\Delta y}{2M}, \quad t_i = 1, 2, ..., s_i, \quad i = 1, 2, \quad (2.31)$$

where Δx and Δy are the width and height of each pixel of the image, respectively, and $M > 0$ denotes the lateral magnification of the microscope system.

The proposed algorithm crucially depends on SVD. Most of the singular values resulting from an SVD are relatively small and are considered to correspond to noise [50]. Here, an important question is how many singular values are associated with noise and should be discarded in each SVD? In the following subsections, we describe the determination of the number of retained singular values in the three SVDs of the algorithm, and importantly, the number of single molecules in the given image. In addition, we give a description of the maximum likelihood estimator with which we will demonstrate the use of the results of the algorithm as the initial conditions for an estimation routine.

2.4.1   Determination of the number of retained singular values in the first SVD

Let $\sigma_1 \geq ... \geq \sigma_K \geq 0, K = \min(N_1, N_2)$, denote the singular values in the first SVD. For $r = 1, ..., K$, let $E_r := \sum_{i=1}^{r} \sigma_i^2$ be the energy of the sequence $\sigma_i, i = 1, ..., r$. Estimate the optimal number of retained singular values r in the first SVD as

$$\hat{r} = \min_{r=1,...,K} \left\{ r : \frac{E_r}{E_K} > \tau \right\}, \quad (2.32)$$

where $E_K := \sum_{i=1}^{K} \sigma_i^2$ is the energy of the sequence of all singular values and $\tau \in \mathbb{R}$ denotes a threshold value typically chosen in the range [0.8, 0.9] [20]. In Section 2.6.1.3, we examine the

effect of different threshold values on the detection rate of the algorithm.

2.4.2 Determination of the number of retained singular values in the second and third SVDs, and the number of single molecules in the image

Let $\sigma_1^i \geq ... \geq \sigma_{N_i}^i \geq 0, i = 1, 2$, be the singular values in the second and third SVDs, respectively. For $l_i = 1, ..., N_i, i = 1, 2$, let

$$\hat{l}_i = \min_{l_i=1,...,N_i} \left\{ l_i : \frac{E_{l_i}}{E_{N_i}} > \tau_i \right\}, \tag{2.33}$$

where $E_{l_i} := \sum_{j=1}^{l_i} (\sigma_j^i)^2$ and $E_{N_i} := \sum_{j=1}^{N_i} (\sigma_j^i)^2$ are the energies of the sequences $\sigma_1^i, ..., \sigma_{l_i}^i$ and $\sigma_1^i, ..., \sigma_{N_i}^i$, respectively, and $\tau_i \in \mathbb{R}$ denotes a threshold value which is again typically chosen in the range [0.8, 0.9] (see Section 2.6.1.3). The estimates $\hat{l}_i, i = 1, 2$, thus denote the number of singular values that remain after discarding those that are considered to obviously correspond to noise.

We next try to reduce further the number of singular values to retain using an optimization approach that minimizes the difference between the original image and the reconstructed image obtained by the estimated locations of the peaks of the image. For $s_i = 1, ..., \hat{l}_i, i = 1, 2$, let $\tilde{X}^{r;s_1,s_2}(k_1, k_2) = \sum_{l=1}^{s_1} \sum_{j=1}^{s_2} \frac{c_l^1 b_l^1 c_j^2 b_j^2}{(e^{i2\pi k_1/N_1} - a_l^1)(e^{i2\pi k_2/N_2} - a_j^2)}, k_t = 1, ..., N_t, t = 1, 2$, be the estimated data calculated via the algorithm by retaining r singular values in the first SVD and s_1 and s_2 singular values in the second and third SVDs, respectively. In other words, $\tilde{X}^{r;s_1,s_2}$ is the reconstructed image of Eq. (2.22) after discarding the singular values corresponding to noise. Denoting the poles of $\tilde{X}^{r;s_1,s_2}$ by $(\bar{a}_k^1, \bar{a}_k^2), \bar{a}_k^t \in \left\{ a_1^t, \cdots, a_{s_t}^t \right\}, \bar{a}_k^t := |\bar{a}_k^t| e^{i\bar{w}_k^t}, 0 \leq \bar{w}_k^t \leq 2\pi, k = 1, ..., s_1 s_2, t = 1, 2$, and their corresponding product of coefficients in the numerator by $p_k \in \mathbb{C}, k = 1, ..., s_1 s_2$, assume the peak magnitudes to be $|p_1| \geq \cdots \geq |p_{s_1 s_2}| \geq 0$.

Let $h = \min(s_1, s_2)$ denote the number of single molecules, assuming that we retain s_1 and s_2 singular values in the second and third SVDs, respectively. In the following, we estimate the optimal number of single molecules. Let $\hat{\theta}^h := (\hat{\theta}_1, ..., \hat{\theta}_h) \in \mathbb{R}^{2h}, \hat{\theta}_n := (\hat{x}_n, \hat{y}_n) \in \mathbb{R}^2, n =$

$1, ..., h$, such that

$$\hat{x}_n := \frac{\Delta x}{2M} \frac{\bar{w}_n^2 N_1}{\pi} + \frac{\Delta x}{2M}, \quad \hat{y}_n := \frac{\Delta y}{2M} \frac{\bar{w}_n^1 N_2}{\pi} + \frac{\Delta y}{2M}, \tag{2.34}$$

are the estimated locations of the h peaks with the largest magnitudes. In general, we consider all possible h-combinations of the poles of $\tilde{X}^{r;s_1,s_2}$, but in most cases, the single molecules are associated with the peaks with the largest magnitudes. Let $\{z_1, ..., z_{N_{pix}}\}$ denote our acquired data, where N_{pix} denotes the number of pixels in the image. Then, the estimated number of single molecules $\hat{h}$ is given by

$$\hat{h} = \underset{h=\min(s_1,s_2),\, s_i=1,...,\hat{l}_i,\, i=1,2}{\arg\min} \left(\sum_{k=1}^{N_{pix}} (z_k - \mu_{\hat{\theta}^h}(k))^2 \right), \tag{2.35}$$

where, in the case that the single molecule image is modeled with a 2D PSF, we have, for $k = 1, ..., N_{pix}$, the mean number of photons detected in the kth pixel given by[7]

$$\mu_{\hat{\theta}^h}(k) := \sum_{n=1}^{h} \frac{N_{p,n}}{M^2} \int_{C_k} q\left(\frac{x}{M} - \hat{x}_n, \frac{y}{M} - \hat{y}_n\right) dx\, dy, \quad \hat{\theta}^h \in \mathbb{R}^{2h}, \quad h = \min(s_1, s_2), \tag{2.36}$$

where $N_{p,n}$ is the expected number of photons due to the nth molecule that impact the detector plane during the image exposure, $C_k \subset \mathbb{R}^2$ denotes the region in the detector plane occupied by the kth pixel, and q is the 2D PSF of the optical system. If the PSF is the Airy profile, then q is given by

$$q(x, y) := \frac{J_1^2\left(\frac{2\pi n_a}{\lambda}\sqrt{x^2 + y^2}\right)}{\pi(x^2 + y^2)}, \quad (x, y) \in \mathbb{R}^2, \tag{2.37}$$

where n_a denotes the numerical aperture of the objective lens, λ denotes the emission wavelength of the molecule, and J_1 denotes the first order Bessel function of the first kind.

2.4.3 Fitting single molecule images using the maximum likelihood estimator

The molecule locations estimated with the proposed algorithm can be used as the initial conditions in any estimation routine. In this chapter, we demonstrate this using the maximum likelihood estimation routine [37, 6]. In the following, we briefly explain the basis of the maximum likelihood

estimation.

Let Θ denote the parameter space that is an open subset of $\mathbb{R}^n$. The maximum likelihood estimate $\hat{\theta}_{mle}$ of $\theta \in \Theta$, for data incorporating Gaussian readout noise, is given by

$$\hat{\theta}_{mle} = \operatorname*{arg\,min}_{\theta \in \Theta} \left(-L(\theta | z_1, ..., z_{N_{pix}}) \right), \tag{2.38}$$

where $\left\{ z_1, ..., z_{N_{pix}} \right\}$ denotes an image with N_{pix} pixels and L is the log-likelihood function given by [7]

$$L(\theta | z_1, ..., z_{N_{pix}}) = \sum_{k=1}^{N_{pix}} \log \left(\frac{1}{\sqrt{2\pi}\sigma_k} \sum_{l=0}^{\infty} \left(\frac{[\mu_\theta(k) + \beta_k]^l \, e^{-[\mu_\theta(k)+\beta_k]}}{l!} e^{-\frac{1}{2}\left(\frac{z_k - l - \eta_k}{\sigma_k}\right)^2} \right) \right). \tag{2.39}$$

In Eq. (2.39), in the case of one molecule, $\mu_\theta(k)$ is the mean photon count in the kth pixel due to the molecule and is given by

$$\mu_\theta(k) := \frac{N_p}{M^2} \int_{C_k} q\left(\frac{x}{M} - x_0, \frac{y}{M} - y_0 \right) dx dy, \quad k = 1, ..., N_{pix}, \tag{2.40}$$

where $\theta = (x_0, y_0) \in \mathbb{R}^2$ denotes the location of the molecule in the object space, N_p is the expected number of photons from the molecule that are detected over the detector plane, and q is the Airy profile given by Eq. (2.37). Also, β_k is the background level in the kth pixel, and η_k and σ_k denote the mean and standard deviation of the Gaussian readout noise in the kth pixel, respectively.

2.5 Methods

2.5.1 Simulation parameters

To analyze the performance of the proposed algorithm, we simulated different data sets using parameters commonly used in single molecule experiments. Some data sets comprise repeat images of one molecule, and some comprise repeat images of more than one molecule. Also, some data sets are such that each image contains a different set of molecules whose locations are randomly chosen based on uniform distributions that place the molecules within different spatial

intervals inside the image. Regardless of the data set, the image of a molecule was generated with the Airy profile of Eq. (2.37) with a numerical aperture of $n_a = 1.4$ and an emission wavelength of $\lambda = 485$ nm. Furthermore, a lateral magnification of $M = 100$, a detector pixel size of 6.5 μm $\times$ 6.5 μm, and a zero-mean Gaussian readout noise with standard deviation $\sigma = 6\ e^-$ per pixel, were assumed. Also, we assumed that all simulated images are background-subtracted.

2.5.2 Imaging experiments

2.5.2.1 Sample preparation

High-performance Zeiss coverslips (#1.5) were prepared as follows: coverslips were sonicated with the following solutions in succession (each for 20 minutes): 50% HPLC-grade ethanol, 1mM HCl with 50% HPLC-grade ethanol, 1M KOH with 50% HPLC-grade ethanol, and 50% HPLC-grade ethanol. The cleaned coverslips were attached to MatTek dishes. 200 μl of Poly-L-lysine (PLL) solution (Sigma-Aldrich) were added to the glass bottom area of the dishes for 10 minutes at room temperature. PPL was removed and 250-pM Alexa Fluor 647 fluorescent dye (Invitrogen) in 200 μl of phosphate-buffered saline (PBS) was applied for 10 minutes at room temperature. The sample was then washed with PBS twice at room temperature followed by the addition of 1 ml of PBS to the sample.

2.5.2.2 Microscopy setup

Custom laser excitation optics were installed for a Zeiss Axio Observer.A1 microscope. The laser optics were configured with 635-nm and 405-nm diode lasers (OptoEngine) for the excitation and photoactivation, respectively, of Alexa Fluor 647. The excitation light was reflected using a dichroic filter (Di01-R405/488/561/635-25x36; Semrock) and focused on the back focal plane of a 63$\times$, 1.46 NA Zeiss objective lens. The emission light from the Alexa Fluor 647 dye was collected by the objective lens and filtered using a single bandpass filter (FF01-676/29-25; Semrock). The images were recorded using an electron-multiplying charge-coupled device camera (iXon DU897-BV; Andor) in conventional readout mode. The camera pixel size was 16 μm $\times$ 16 μm. All components, including lasers, shutters and cameras, were controlled and synchronized

using custom software written in the C programming language.

2.5.2.3 *Super-resolution imaging*

We first removed PBS from the single molecule sample prepared in Section 2.5.2.1 and added the imaging buffer (50-mM beta-mercaptoethylamine (MEA), 0.5-mg/ml glucose oxidase, 40-μg/ml catalase in PBS, pH 7.4, with 10% glucose). The sample was sealed with a coverslip and then positioned on the sample stage of the microscope for 5 to 10 minutes for temperature equilibration and the oxygen scavenging process. Images were subsequently acquired at a rate of 20 frames per second. The sample was illuminated with the 635-nm and 405-nm diode lasers alternately with photoactivation by the 405-nm laser every third frame. The frames with 405-nm laser illumination were excluded from data analysis.

2.6 Results and discussion

In this section, we present and discuss the results of the proposed algorithm when applied to both simulated and experimental images of single molecules.

2.6.1 Results for simulated data

Using simulated data sets, we first examine the performance of the algorithm in terms of the detection rate. We then analyze the bias and accuracy of the algorithm. The bias is assessed by the average of the deviations of the estimated molecule locations from the ground truth. The accuracy is assessed by looking at the square root of the average of the squared deviations from the ground truth. For repeat images of the same molecules, however, we look instead at the standard deviation of the estimates. In particular, for data sets containing repeat images of one molecule, we compare the standard deviation of the estimates with the limit of the localization accuracy given by the square root of the CRLB. Besides these analyses, we examine the effect on the detection rate when different threshold values are used in the first, second, and third SVDs of the algorithm.

2.6.1.1 One molecule

Here, to evaluate the detection rate of the algorithm, we first simulated data sets in which each image contains one molecule, whose location was randomly chosen based on a uniform distribution that places it within the image. For a given data set, the mean photon count is the same for the molecule in every image. Different data sets differ by this mean photon count, which ranges from 500 to 4500. For each mean photon count, we simulated 200 images. To establish statistical measures of the detection rate, we needed to pair the molecules localized by the algorithm with the molecules from the ground truth. For this purpose, we used the Hungarian algorithm with a search area of radius 100 nm [53]. Then, we categorized the localized molecules which were successfully paired with ground truth molecules as true positives. The ground truth molecules that were not paired with any localized molecule and the localized molecules which were not paired with any ground truth molecule were categorized as false negatives and false positives, respectively. Denoting the number of true positives by TP, the number of false negatives by FN, and the number of false positives by FP, we define the precision (PRE) and recall (REC) measures as [53]

$$PRE := \frac{TP}{FP + TP}, \quad REC := \frac{TP}{FN + TP}. \tag{2.41}$$

Figure 2.1 shows the results of the measures of the detection rate for data sets consisting of images containing one molecule each. It can be seen that for all mean photon counts considered, there are no false negatives and the recall is 1. Also, the figure shows that by increasing the mean photon count, the precision increases. However, it is important to note that even when the mean number of photons is as low as 500, a relatively large number of detected molecules are true positives (about 86%).

We next examine the bias of the algorithm for a data set in which each frame contains one molecule whose location in the image is chosen randomly. For this purpose, we simulated 1000 15×15-pixel images, each containing one molecule with a mean photon count of 1500 photons. In each image, the location of the molecule was drawn from a uniform probability distribution

Figure 2.1: Analysis of the detection rate of the algorithm, applied to data sets in which each image contains one molecule, whose location in the image is chosen randomly according to a uniform probability distribution. For a given data set, the same mean photon count is used to simulate the molecule in each image. Different data sets differ by this mean photon count. For each mean photon count, 200 images of size 30×30 pixels were simulated using the parameters given in Section 2.5.1. The Hungarian algorithm with a search area of radius 100 nm is used to pair the localized molecules with the ground truth molecules. Reprinted with permission from [54].
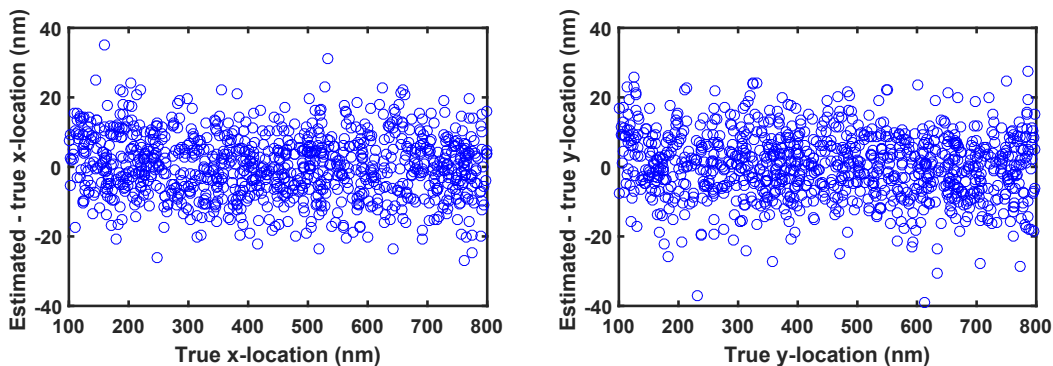
Figure 2.2: Analysis of the error of location estimates obtained from a data set in which each frame contains one molecule whose location in the image is chosen randomly. Shown in the left and right plots are the differences between the x-estimates and the true x-values, and the differences between the y-estimates and the true y-values, respectively, for the true positives obtained with the algorithm. The data set consists of 1000 15×15-pixel images, each of a molecule with a mean photon count of 1500 photons whose location is randomly chosen from a uniform probability distribution that places the molecule between the 2nd and 14th pixel in both the x and y dimensions. The images were simulated using the parameters given in Section 2.5.1. Reprinted with permission from [54].

26

that places it between the 2nd and 14th pixel in both the x and y dimensions. (We assumed that no molecule was located near the edges of the 15×15-pixel image.) As shown in Fig. 2.2, the deviations of both the x and y location estimates from the ground truth are, overall, centered around 0 nm. Therefore, the results suggest that, in the case where there is only one molecule per image, there is no systematic bias associated with the algorithm in this case (the average of x and y deviations are 0.321 nm and 0.335 nm, respectively). Also, the square root of the average of the squares of the x and y deviations are 9.123 nm and 9.467 nm, respectively, which are close to the standard deviations of the estimated locations obtained for a data set consisting of repeat images of one molecule with the same mean photon count of 1500 photons (analysis of data sets with repeat images is presented next). This suggests that the variation of the deviations about the ground truth is reasonable.

To examine further the bias of the algorithm, we simulated data sets containing repeat images of one molecule. The data sets differ by the mean photon count of the molecule, which we assume does not vary from frame to frame in a given data set. This mean photon count ranges from 500 to 4500 for the different data sets. For each data set, we simulated 1000 repeat images. Figure 2.3 shows, as a function of the mean photon count, the differences between the averages of the x- and y-estimates for the correctly detected (i.e., true positive) molecules and the corresponding true x- and y-coordinates. Similar to the case of data sets with non-repeat images [Fig. 2.2], the evenness of the spread of the estimated bias about 0 nm for both coordinates suggests that when there is only one molecule per image, there is no systematic bias associated with our proposed algorithm.

We next evaluate the performance of the algorithm in terms of the standard deviation of the estimates for the sets of repeat images. For nine of the data sets from Fig. 2.3, we calculated the standard deviations of the x-estimates and y-estimates for the correctly detected (i.e., true positive) molecules. The percentage differences between the standard deviations and the CRLB-based limits of the x-localization accuracy and y-localization accuracy [7] are shown in Fig. 2.4. The percentage difference is the absolute difference between the standard deviation of the estimates and the corresponding limit of accuracy, expressed as a percentage of the limit of accuracy. As

Figure 2.3: Analysis of the average of location estimates obtained from repeat images of one molecule. Shown in the left and right plots are the difference between the average of the x-estimates and the true x-value, and the difference between the average of the y-estimates and the true y-value, respectively, for data sets that differ by the mean photon count assumed for the molecule per image. For each mean photon count, the data set consists of 1000 images of size 15×15 pixels, simulated using the parameters given in Section 2.5.1. Reprinted with permission from [54].
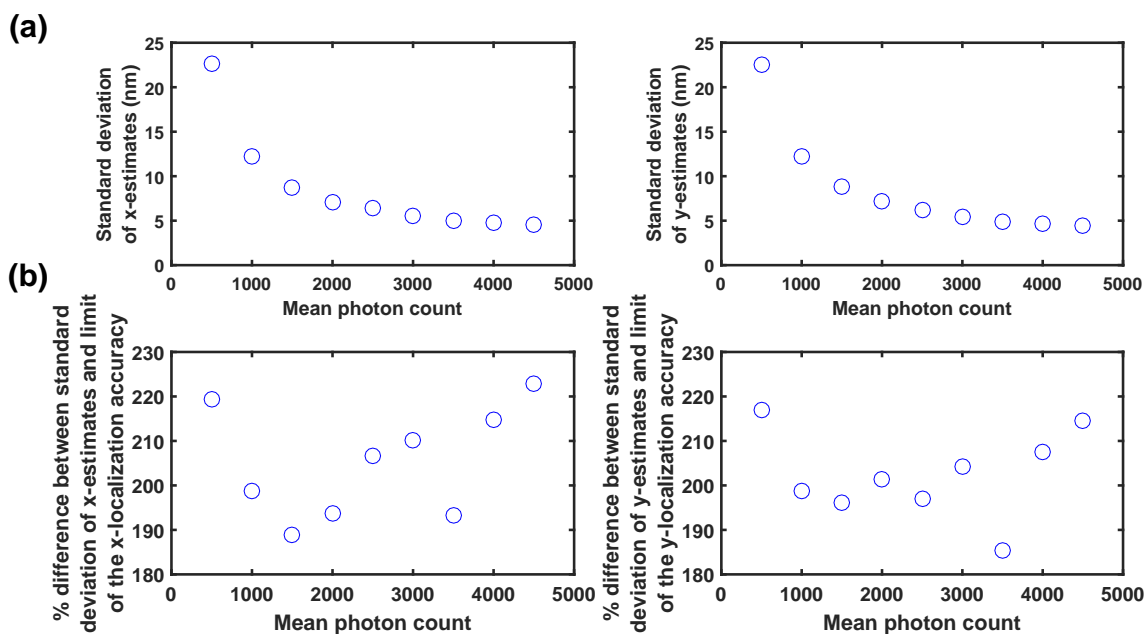
Figure 2.4: Analysis of the standard deviation of location estimates obtained from repeat images of one molecule. (a) The standard deviations of the x- and y-estimates for nine of the data sets from Fig. 2.3. (b) The percentage difference between the standard deviation of the x-estimates and the limit of the x-localization accuracy, and the percentage difference between the standard deviation of the y-estimates and the limit of the y-localization accuracy. The percentage difference is the absolute difference between the standard deviation of the estimates and the corresponding limit of accuracy, expressed as a percentage of the limit of accuracy. Reprinted with permission from [54].

28
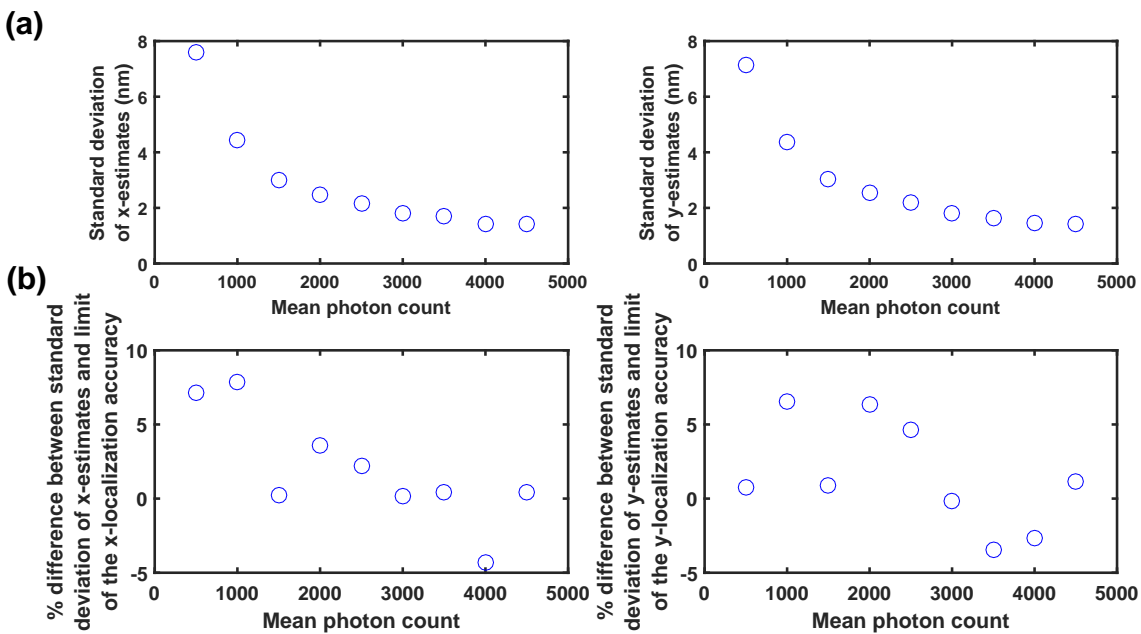
Figure 2.5: Analysis of the standard deviation of location estimates produced by the maximum likelihood estimator when the location estimates obtained with the algorithm are used as the initial conditions. (a) The standard deviations of the maximum likelihood x- and y-estimates for the same data sets as in Fig. 2.4, which comprise repeat images of one molecule. (b) The percentage difference between the standard deviation of the x-estimates and the limit of the x-localization accuracy, and the percentage difference between the standard deviation of the y-estimates and the limit of the y-localization accuracy. Reprinted with permission from [54].
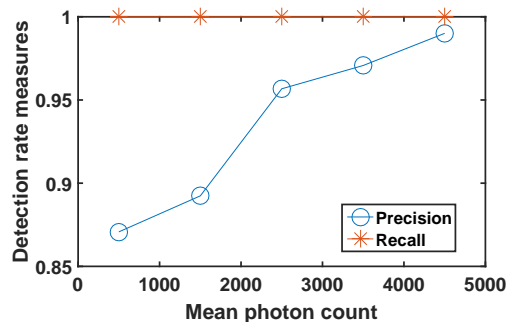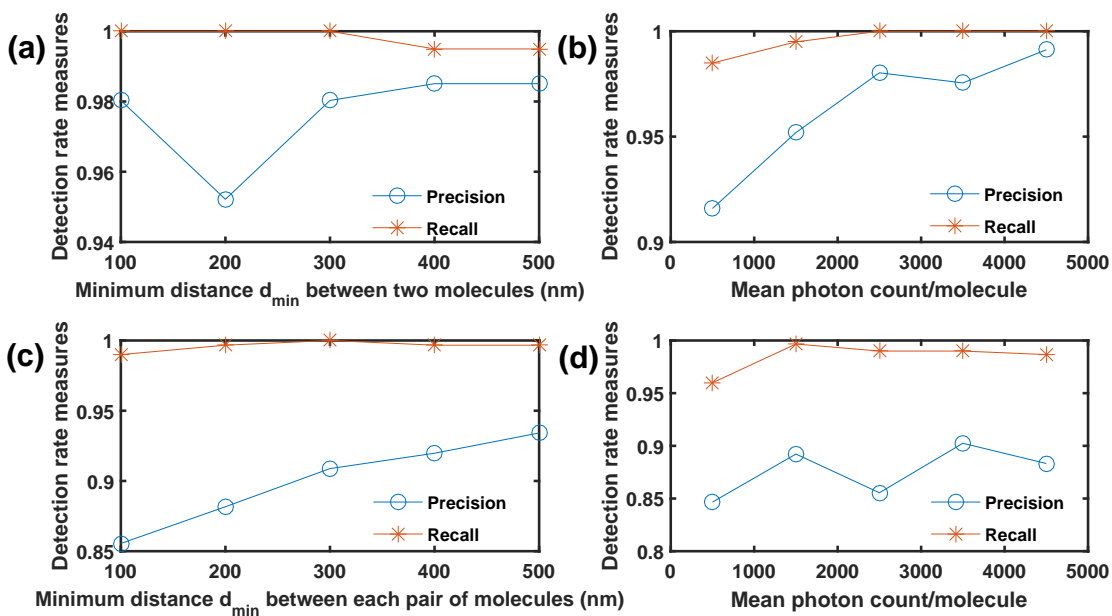
Figure 2.6: Analysis of the detection rate of the algorithm, applied to data sets in which each image contains multiple molecules whose locations in the image are chosen randomly. For a given data set, the mean photon count is the same for each molecule in every frame. The location of each molecule is drawn from a uniform distribution that places it inside the image, with the constraint that the distance between each pair of molecules is not less than the minimum distance d_{min}. For each data set, we simulated 200 images of size 30×30 pixels using the parameters given in Section 2.5.1. For data sets in which there are two molecules per image, the precision and recall measures are shown as a function of d_{min} in (a), where the mean photon count is 2500 photons/molecule, and as a function of the mean photon count in (b), where $d_{min} = 100$ nm. For data sets in which there are three molecules per image, the precision and recall measures are shown as a function of d_{min} in (c), where the mean photon count is 2500 photons/molecule, and as a function of the mean photon count in (d), where $d_{min} = 100$ nm. The Hungarian algorithm with a search area of radius 100 nm is used to pair the localized molecules with the ground truth molecules. Reprinted with permission from [54].

30

shown in Fig. 2.4, when the mean number of photons increases, the standard deviation of the estimates decreases. Also, as can be seen in the second row of Fig. 2.4, the differences between the standard deviations of the estimates and their respective limits of the localization accuracy are around twice (i.e., around 200% of) the limits of accuracy. This difference likely arises from the fact that our algorithm approximates an Airy profile with the frequency response of a first-order system, and there is a difference between the shape of the peak of an Airy profile and that of the first-order system in the frequency domain. In Appendix A, we applied our algorithm to images simulated using the frequency response of a first-order system rather than an Airy profile, and in that case, the standard deviations of the x- and y-estimates came close to their respective limits of accuracy.

Also, we used the location estimates obtained with the algorithm from a set of repeat images as initial conditions for the maximum likelihood estimation of the location of the molecule from those same images. This maximum likelihood estimator fits an Airy photon distribution profile to the image data, and the equations that describe how the maximum likelihood estimates are calculated are given in Section 2.4.3 [Eqs. (2.38) and (2.39)]. We calculated the standard deviations of the resulting x-estimates and y-estimates, and the percentage differences between them and the limits of the x-localization accuracy and y-localization accuracy [Fig. 2.5]. We only considered those estimates for which the estimated locations were within the image. As can be seen in Fig. 2.5, the standard deviations are substantially smaller compared to those obtained with the algorithm [Fig. 2.4], and come close to the limits of accuracy, consistent with the results in [37] and [6].

2.6.1.2 *Multiple molecules*

So far, we have evaluated the performance of the algorithm in the case where we have only one molecule in any given image. Here, we analyze the results obtained when the algorithm was used to simultaneously localize molecules from images that contain multiple closely spaced molecules. As before, we first analyze the detection rate of the algorithm. For this purpose, we simulated data sets containing images of either two or three molecules. For each image, the location of each molecule is randomly chosen from a uniform probability distribution that places the molecule
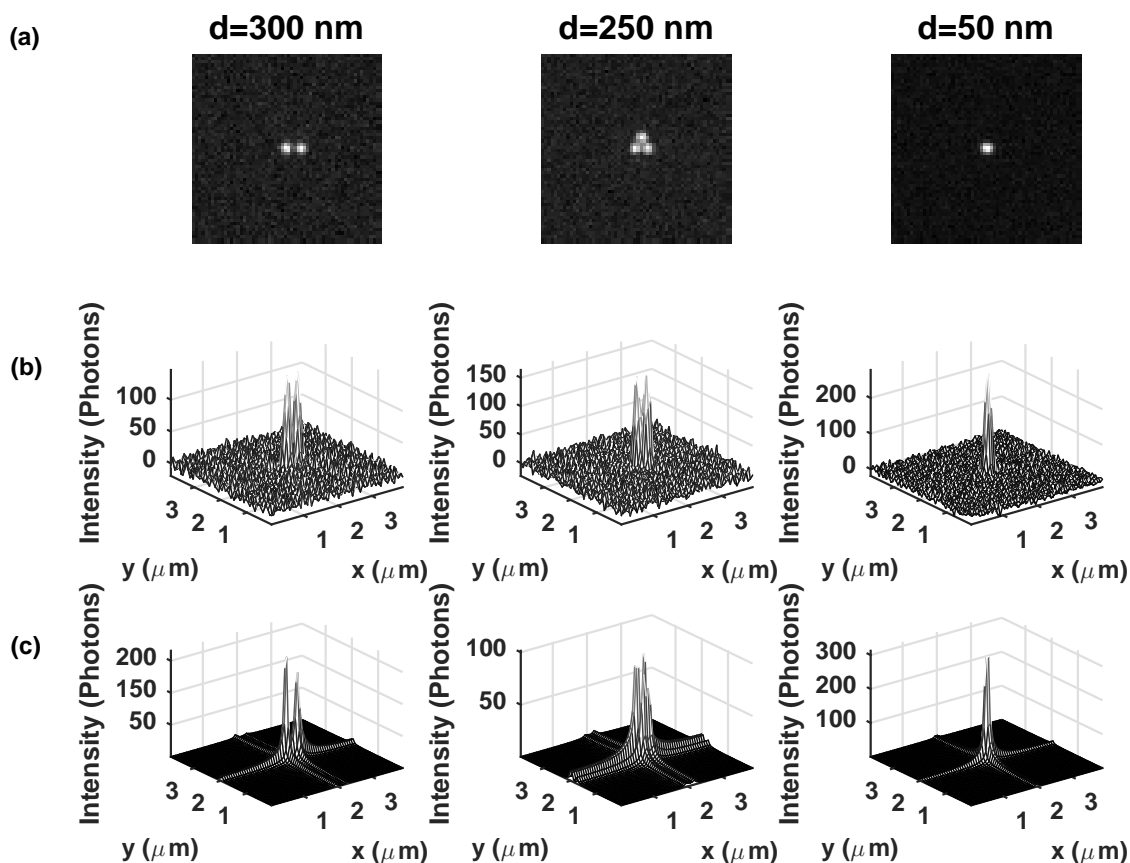
Figure 2.7: Reconstruction of images containing two or three closely spaced molecules. (a) Images of size 60×60 pixels of 2, 3, and 2 closely spaced point sources separated from one another by a distance d of 300 nm, 250 nm, and 50 nm, respectively. The images are simulated using the parameters given in Section 2.5.1. (b) Mesh plots of the images shown in (a). (c) Mesh plots of the magnitude of the reconstructed image (algorithm result), showing the detection of 2, 3, and 2 single molecules in the image. Reprinted with permission from [54].

inside the image, with the constraint that the distance between each pair of molecules is not less than a minimum distance d_{min}. In one case, all data sets are simulated with the same minimum distance of $d_{min} = 100$ nm, but differ by the mean photon count per molecule, which ranges from 500 to 4500. More specifically, each data set comprises 200 images, where each image contains molecules with the same given mean photon count. In another case, the mean photon count is the same for all data sets at 2500 photons/molecule, but the data sets differ by d_{min}, which ranges from 100 nm to 500 nm. Here, each data set comprises 200 images, where each image contains molecules separated by the same given minimum distance d_{min}. Figure 2.6 shows the precision and recall measures for the different data sets (similar to the one-molecule case, the Hungarian algorithm with a search area of radius 100 nm was used to pair the localized molecules with the ground truth molecules). Specifically, the figure shows that the recall for all data sets is more than 95%. The precision is likewise quite good, as even when the mean number of photons is as low as 500 photons/molecule, or the minimum distance between each pair of molecules is as small as 100 nm, a relatively high percentage (around 85%) of the detected molecules are true positives. We also analyzed the detection rate for data sets with more molecules per image (5, 7, and 9 molecules per image), and obtained similar results. To give examples of the reconstructed image calculated from our algorithm, we simulated images of size 60×60 pixels containing two or three closely spaced molecules separated by a distance d of 50 nm, 250 nm, and 300 nm, with a mean photon count of 1500 photons/molecule. We then reconstructed each image by applying our algorithm. As shown in Fig. 2.7, in all cases we were able to distinguish the closely spaced molecules from each other.

We next analyze the bias and accuracy of the algorithm when applied to data sets consisting of repeat images of multiple molecules. Unlike the one-molecule case, bias is observed here. We first characterize this bias by demonstrating its dependence on the distances between the molecules relative to the image size. In the following, we focus on data sets comprising images of two molecules, though we also analyzed data sets with three and five molecules and obtained similar results. We simulated data sets comprising 15×15-pixel, 20×20-pixel, and 40×40-pixel images
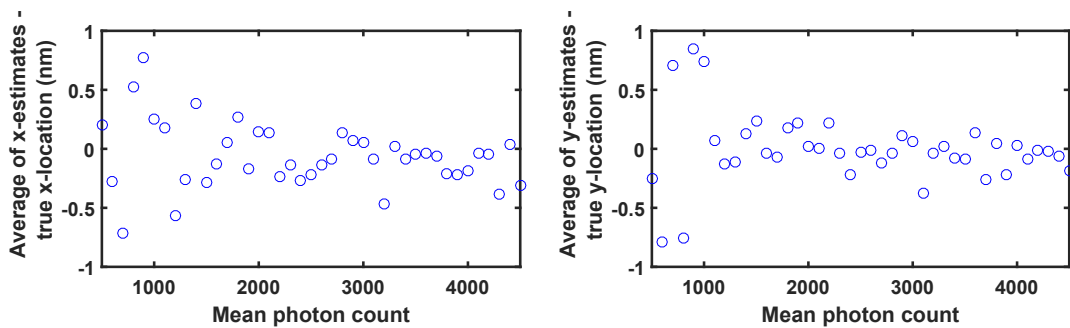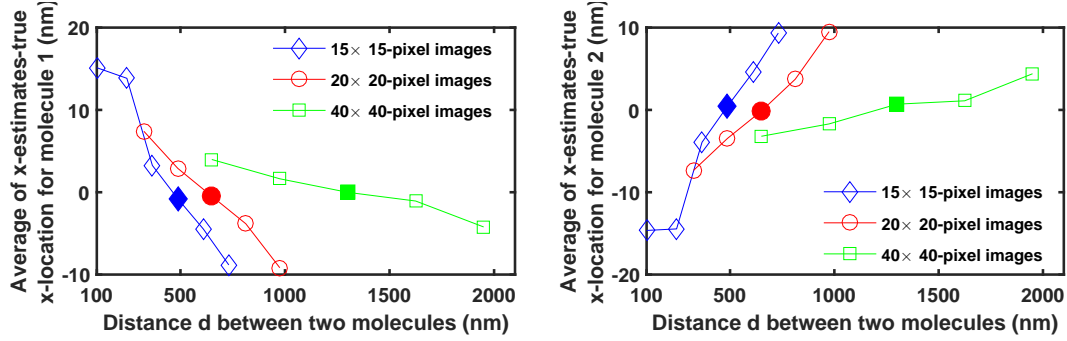
Figure 2.8: Analysis of the average of the location estimates obtained from sets of repeat images of two molecules. Shown in the left and right plots are the differences between the average of the x-estimates and the true x-value for the first and second molecules, respectively, for data sets comprising 15×15-pixel, 20×20-pixel, and 40×40-pixel images. For each image size, distances d between the two molecules are chosen around half of the side length of the square region occupied by the image in the object space. For a given data set, we simulated 500 images with a mean photon count of 2500 photons/molecule and the parameters given in Section 2.5.1. The results for $d = s/2$, where $s = 65N$ nm is the side length of the square region occupied by an $N \times N$-pixel image in the object space, are shown with filled symbols. Reprinted with permission from [54].

in order to evaluate different combinations of the distance d between the two molecules and the image size. For a given data set, we simulated 500 images with a mean photon count of 2500 photons/molecule, a pixel size of 6.5 μm $\times$ 6.5 μm, and a lateral magnification of 100. For these settings, the area occupied by an $N \times N$-pixel image in the object space is an $s \times s$ square region, where $s = 65N$ nm (e.g., for N=20, $s = 1300$ nm). For each data set, the difference between the average of the estimated x-locations for the correctly detected molecules and the corresponding true x-coordinate is plotted in Fig. 2.8. For each image size considered, the figure shows that as d approaches $s/2$ (e.g., for N=20, $s/2 = 650$ nm), i.e., as the difference between the phases of the poles of the second-order system resulting from the algorithm approaches the maximum of π rad on the unit circle, the effect of the poles on each other decreases and the estimated bias for the location of each molecule approaches 0 nm. The results for $d = s/2$ are shown with filled symbols in Fig. 2.8. Also, we calculated the results for the y-estimates and obtained similar results.

Having characterized the nature of the bias, we next analyze, as we did in the case of one molecule, the bias and accuracy of the algorithm as a function of the mean photon count per
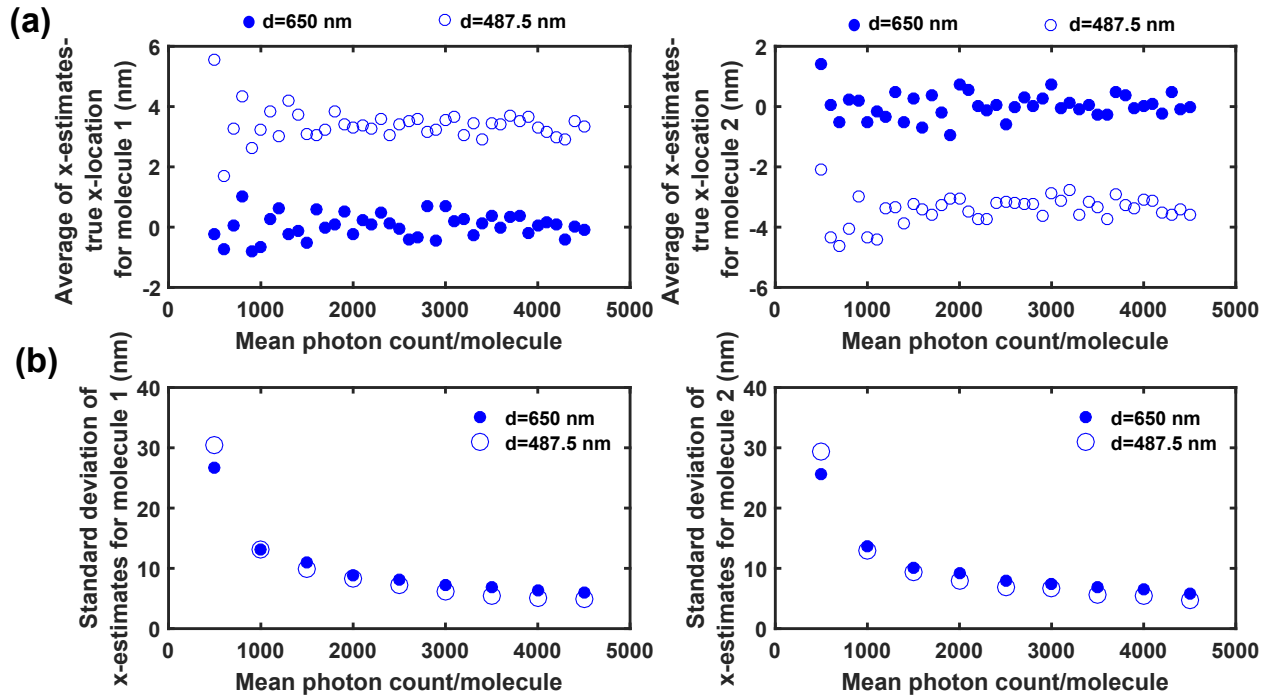
Figure 2.9: Analysis of the average and standard deviation of location estimates obtained from sets of repeat images of two molecules as a function of the mean photon count per molecule. Two scenarios are considered - one in which the distance d between the two molecules is 650 nm, and one in which d is 487.5 nm. For each scenario, the data sets differ by the mean photon count per molecule. For each mean photon count, the data set consists of 500 repeat images of size 20×20 pixels, simulated using the parameters given in Section 2.5.1. (a) Differences between the average of the estimated x-locations and the corresponding true x-coordinates for the two molecules. (b) The standard deviations of the estimated x-locations for the two molecules. Reprinted with permission from [54].

molecule. For this purpose, we simulated data sets which contain repeat images of two molecules. These data sets again differ by the mean photon count per molecule, which we assume does not vary from frame to frame. This mean photon count ranges from 500 to 4500 for the different data sets. We simulated 500 20×20-pixel images per data set. In one case, the distance d between the two molecules is 650 nm (which corresponds to the filled circle in Fig. 2.8), and in another case, $d = 487.5$ nm (which corresponds to the first open circle to the left of the filled circle in Fig. 2.8). Looking at the two distances allows us to verify the effect of different distances between the molecules relative to the image size. To assess the bias of the algorithm, for each molecule in a given data set, we calculated the difference between the average of the estimated x-locations and the corresponding true x-coordinate. As can be seen in Fig. 2.9, when $d = 650$ nm, the estimated bias is around 0 nm for both molecules. On the other hand, when $d = 487.5$ nm, the estimated bias levels are around 3.5 nm and -3.5 nm for molecules 1 and 2, respectively. These bias results are consistent with the illustration of bias in Fig. 2.8. Note that we also analyzed the y-estimates and obtained similar results.

For each distance d, we calculated the standard deviations of the estimated x-locations for nine of the data sets. As shown in Fig. 2.9, for both distances, as the mean number of photons per molecule increases, the standard deviation of the estimates decreases. Also, even when the mean photon count is as low as 500 photons/molecule, the plots show that our algorithm can still localize the molecules with relatively high accuracy (the standard deviations of the x-estimates are around 30 nm for both molecules when $d = 487.5$ nm, and around 27 nm when $d = 650$ nm). Similar results were obtained for the y-estimates.

2.6.1.3 *Analysis of the effect of threshold values on the detection rate of the algorithm*

In Sections 2.4.1 and 2.4.2, typical threshold values in the range [0.8, 0.9] are suggested for the first, second, and third SVDs of the algorithm. Here, we carry out a more in-depth analysis on the effect of the threshold values on the detection rate of the algorithm. The results of our analysis show that in the case that we have a relatively large number of photons per molecule, the detection rate of the algorithm is not very sensitive to the threshold values. Provided that a relatively high

Table 2.1: Detection rate of the algorithm as a function of the threshold values used for the retention of singular values in the first, second and third SVDs.

Data set	Mean photon count per molecule	Threshold for the 1st SVD	Threshold for the 2nd and 3rd SVDs	Recall	Precision
1	500	0.7	0.7	1	0.857
		0.7	0.8	1	0.850
		0.7	0.9	1	0.850
		0.8	0.7	1	0.786
		0.8	0.8	1	0.768
		0.8	0.9	1	0.768
		0.9	0.7	0.994	0.755
		0.9	0.8	0.994	0.722
		0.9	0.9	0.994	0.722
2	1000	0.7	0.7	0.994	0.933
		0.7	0.8	0.994	0.933
		0.7	0.9	0.994	0.933
		0.8	0.7	0.994	0.933
		0.8	0.8	0.994	0.933
		0.8	0.9	0.994	0.933
		0.9	0.7	0.994	0.933
		0.9	0.8	0.994	0.933
		0.9	0.9	0.994	0.933
3	2500	0.7	0.7	0.994	0.994
		0.7	0.8	0.994	0.994
		0.7	0.9	0.994	0.994
		0.8	0.7	0.994	0.994
		0.8	0.8	0.994	0.994
		0.8	0.9	0.994	0.994
		0.9	0.7	0.994	0.994
		0.9	0.8	0.994	0.994
		0.9	0.9	0.994	0.994

Results are shown for three simulated data sets that differ by the mean photon count per molecule per image. Each data set consists of 100 images in which there are two molecules per image. The location of each molecule is randomly chosen from a uniform distribution that places the molecule inside the image, and is subject to the constraint that the distance between the two molecules is not less than 400 nm. Reprinted with permission from [54].
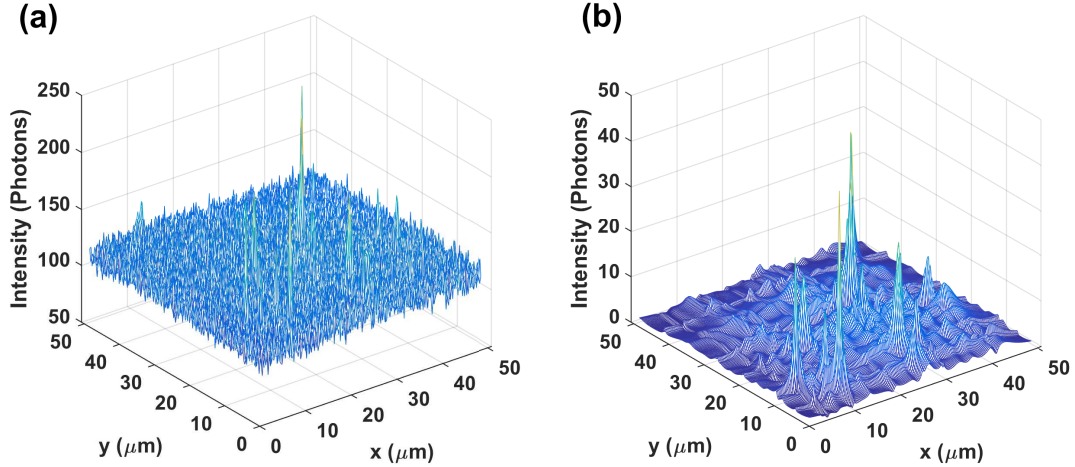
Figure 2.10: Result of the algorithm applied to an experimental super-resolution image. (a) Image of individual Alexa Fluor 647 molecules acquired using the microscopy setup described in Section 2.5.2. The pixel size and image size are 16 μm × 16 μm and 192 × 192 pixels, respectively. (b) The magnitude of the reconstructed image obtained with the algorithm. Reprinted with permission from [54].

threshold value (e.g., in the range [0.7, 0.9]) is used to ensure that singular values corresponding to signal are not discarded, it appears that the optimization procedure of Section 2.4.2 is able to remove the singular values corresponding to noise and yield the correct result. On the other hand, in the case of a low number of photons per molecule, the differences between the singular values that correspond to noise and the singular values associated with signal are often small, and there is no straightforward guideline to choose the threshold values.

In our analysis, we consider three simulated data sets in which there are two molecules per image. The three data sets differ by the mean photon count per molecule per image, which we chose to be 500, 1000, and 2500. As can be seen in Table 2.1, when the mean photon count is 1000 or 2500 per molecule, the recall and precision remain unchanged for threshold values of 0.7, 0.8, and 0.9. However, in the case of the low mean photon count of 500 per molecule, use of a high threshold value, such as 0.8 or 0.9, for the first SVD, leads to the retention of more noise singular values and results in a nontrivial reduction of the precision.

(a) **(b)** **(c)**

(d) **(e)** **(f)**
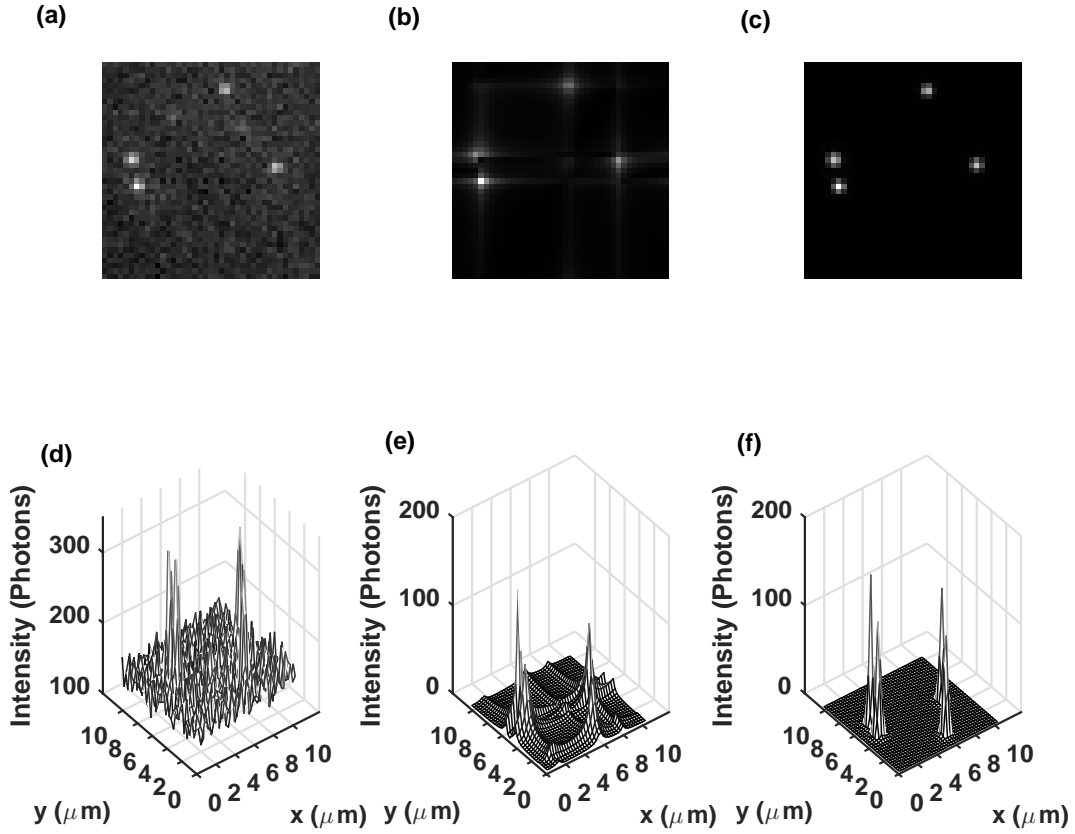
Figure 2.11: Results of the algorithm applied to an ROI from an experimental super-resolution image. (a) A 41×41-pixel ROI of the super-resolution image shown in Fig. 2.10. (b) The magnitude of the reconstructed image (algorithm result). (c) The image reconstructed using Eq. (2.36), in which the single molecule locations estimated using our algorithm are used in the computation of the Airy profile q in Eq. (2.37), and in which the mean photon counts $N_{p,n}$ and the parameter $\alpha := \frac{2\pi n_a}{\lambda}$ are separately estimated with a maximum likelihood estimator. (d), (e), and (f) show the mesh plots of the images in (a), (b), and (c), respectively. Reprinted with permission from [54].

2.6.2 Results for experimental data

Here, we present the results obtained by applying the proposed localization algorithm to experimental single molecule image data acquired as described in Section 2.5.2. Both the acquired image and the reconstructed image are shown in Fig. 2.10, demonstrating that we were able to recover the locations of the significant peaks in the original image that are associated with the locations of individual Alexa Fluor 647 dye molecules.

We next applied the algorithm to a relatively small 41×41-pixel region of interest (ROI) [Figs. 2.11(a) and 2.11(d)] in the acquired image so that a better visual comparison can be made between the reconstructed image obtained with the algorithm and the actual image. In addition to the image reconstructed in terms of the parameters of the multi-order system [Figs. 2.11(b) and 2.11(e)], an image is reconstructed using Eq. (2.36), in which the single molecule locations estimated using our algorithm are used in the computation of the Airy profile q [Figs. 2.11(c) and 2.11(f)]. For this purpose, we separately estimated the mean photon counts $N_{p,n}$ in Eq. (2.36) and the parameter $\alpha := \frac{2\pi n_a}{\lambda}$ in Eq. (2.37) using a maximum likelihood estimator. The reconstruction using the Airy profile provides a better visual comparison with the actual image by showing peaks with more comparable intensities.

3. PARAMETER ESTIMATION OF SINGLE MOLECULE TRAJECTORIES*

3.1 Introduction

Tracking of objects in cellular environments has become a vital tool in molecular cell biology. A particularly important example is single molecule tracking which enables the study of the motion of a molecule in cellular environments by locating the molecule over time. As mentioned in the introduction chapter, the performance of available single molecule trajectory data analysis methods are limited by several limiting factors. For example, most available methods only focus on Gaussian measurements that occur at equidistant time points. Here, in order to address these limitations, we consider a dynamical system where the motion of an object is modeled by stochastic differential equations, and measurements are the detected photons emitted by the moving fluorescently labeled object, which occur at discrete time points, corresponding to the arrival times of a Poisson process. The measurements are distributed according to the optical diffraction theory, and therefore, they would be modeled by different distributions, e.g., an Airy profile for an in-focus and a Born and Wolf profile for an out-of-focus molecule with respect to the detector. For some special circumstances, Gaussian image models have been proposed. Although in most of the available approaches, the motion of the molecule is limited to Brownian motion (pure diffusion) model, the dynamical system considered here allows for more general motion models.

In the following, we present the statistical description of the acquired data, and derive a general formula for the likelihood function of the described data model. Here, we use the following notation

$$\mathcal{C}^l \times \mathbb{R}^l_{[t]} := \left\{ (r_1, \cdots, r_l, \tau_1, \cdots, \tau_l) \, | r_1, \cdots, r_l \in \mathcal{C}, t_0 \leq \tau_1 < \cdots < \tau_l \leq t \right\}, \qquad (3.1)$$

where $\mathcal{C} := \mathbb{R}^2$, $t_0 \in \mathbb{R}$, and $l = 1, 2, \cdots$. If there is no bound on τ_l, we denote the set in Eq. (3.1)

41

by $\mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$.

3.2 Fundamental data model

A basic setup of an optical system considered here is shown in Fig. 3.1, where an object is in the object space and its image is captured by a planar detector in the image space. In the fundamental data model, we assume that the microscopy image data is acquired under ideal conditions. It assumes the use of an image detector that has an unpixelated photon detection area. The detection of a photon is intrinsically random in terms of both the time and the location on the detector at which the photon is detected. In general, the temporal part of the detection of the emitted photons can be modeled as a counting process $\{N(\tau), \tau \geq t_0\}$. Here, we assume that $\{N(\tau), \tau \geq t_0\}$ is a Poisson process referred to as the *photon detection process* that is characterized by the intensity function $\Lambda(\tau), \tau \geq t_0$, referred to as *the photon detection rate*. The spatial component of the photon detection process is specified by random variables, referred to as *the photon location variables*, that describe the locations at which photons emitted by the object of interest are detected.
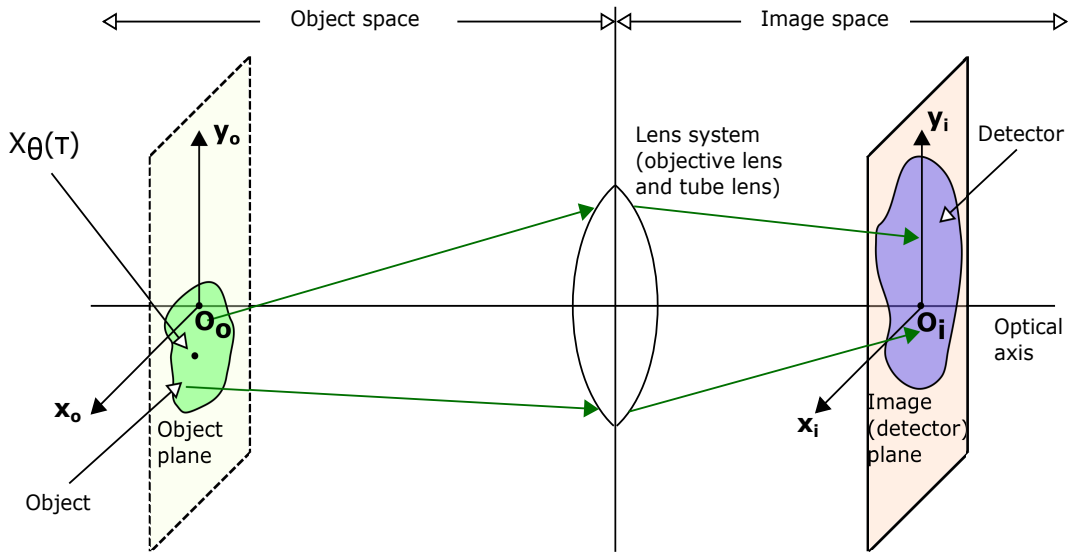
Figure 3.1: Schematic of an optical microscope. An object located in the object (focal) plane is imaged by an optical lens system and the image of the object is acquired by the planar detector in the image space. A 2D random variable $X_\theta(\tau), \tau \geq t_0$, describes the location of the object in the object plane at time τ.

42

In the following definition, we define a spatio-temporal process referred to as the *image detection process*, which models the acquired data, for two different acquisition methods, one when the time interval over which photons are detected is given and the other when the total number of detected photons is given. For a fixed acquisition time, due to the stochastic nature of photon emission, the total number of detected photons varies for every image, while in the other case, the number of detected photons remains the same.

Definition 1. *Let $\mathcal{C} := \mathbb{R}^2$ denote a non-pixelated detector. Let $\mathbb{R}^n, n = 1, 2, \cdots$, be the n-dimensional full parameter space. Let the parameter space Θ describe an open subset of $\mathbb{R}^n$ containing the true parameters. Elements in Θ are described by a parameter vector $\theta \in \Theta$. Let the 1D random variables $T_1, T_2, \cdots$, describe the time points of detection of the photons that impact the detector $\mathcal{C}$, which are arrival time points associated with a Poisson process with intensity function $\Lambda(\tau), \tau \geq t_0, t_0 \in \mathbb{R}$. Let $U_1, U_2, \cdots$, be 2D random variables that describe the locations of detection of the photons that impact the detector $\mathcal{C}$. For $l = 1, 2, \cdots$, let $\mathcal{U}_l := (U_1, \cdots, U_l), \mathcal{U}_0 = \emptyset$, and $\mathcal{T}_l := (T_1, \cdots, T_l), \mathcal{T}_0 = \emptyset$. Assume that the current location of the detected photon, given the current and previous time points, is independent of the future time points, i.e., for $r \in \mathcal{C}$ and $t_0 \leq \tau_1 < \tau_2 < \cdots$,*

$$p_{U_l | \mathcal{T}_k}\left(r | \tau_1, \cdots, \tau_k\right) = p_{U_l | \mathcal{T}_l}\left(r | \tau_1, \cdots, \tau_l\right), \quad \text{for all } k, l = 1, 2, \cdots, \quad k \geq l,$$

where, for random vectors X and Y, the conditional probability density function of X, given Y, is denoted by $p_{X|Y}$.

1. For a fixed acquisition time interval $[t_0, t]$, an image *detection process $\mathcal{G}_{[t]}\left(\left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \mathcal{C}, \Theta\right)$ for a time interval $[t_0, t]$ is defined as a spatio-temporal process whose temporal part $\mathcal{T}_{[t]}$ and spatial part $\mathcal{U}_{[t]}$ describe the time points and the locations of detection of the photons that impact the detector $\mathcal{C}$ in the time interval $[t_0, t]$, respectively, i.e., for $\omega \in \Omega$, where Ω is the sample space,*

$$U_{S_t(\omega)} = T_{S_t(\omega)} = \emptyset, \quad S_t(\omega) = 0,$$

43

and

$$\mathcal{T}_{[t]}(\omega) := \left(T_1(\omega), \cdots, T_{S_t(\omega)}(\omega)\right), \quad \mathcal{U}_{[t]}(\omega) := \left(U_1(\omega), \cdots, U_{S_t(\omega)}(\omega)\right), \quad S_t(w) > 0,$$

where $t_0 \leq T_1(\omega) < \cdots < T_{S_t(\omega)}(\omega) \leq t$, and S_t is a discrete 1D random variable that takes its values in the non-negative integers such that $T_{S_t(\omega)}(\omega) \leq t, T_{S_t(\omega)+1}(\omega) > t, S_t(w) > 0$.

2. Given a fixed number $L = 1, 2, \cdots$, of photons, an image detection process $\mathcal{G}_L\left(\left(\mathcal{U}_L, \mathcal{T}_L\right), \mathcal{C}, \Theta\right)$ *for a fixed number L of photons* is defined as a spatio-temporal process whose temporal and spatial parts describe the time points and the locations of detection of the L photons that impact the detector $\mathcal{C}$, respectively. Moreover, given $\mathcal{T}_L = (\tau_1, \cdots, \tau_L), t_0 \leq \tau_1 < \tau_2 < \cdots < \tau_L$, $\mathcal{G}_{\tau_1, \cdots, \tau_L}\left(\left(\mathcal{U}_L, \mathcal{T}_L\right), \mathcal{C}, \Theta\right)$ *is referred to as the* image detection process at fixed time points $\tau_1, \cdots, \tau_L$.

In Theorem 2, we state expressions for the probability/probability density functions of image detection processes for a fixed time interval and for a fixed number of photons in terms of the conditional distributions of the locations of the detected photons, given the previous locations and the current and previous time points of the detected photons. We further show that each of these conditional distributions can be expressed in terms of a scaled and shifted version of the image of the object and the distribution of the prediction of the object location, given the previous locations and time points of the detected photons. We drop the parameter vector $\theta \in \Theta$, when it is clear from the context.

Theorem 2. *Let $\mathcal{G}_{[t]}\left(\left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \mathcal{C}, \Theta\right)$ and $\mathcal{G}_L\left(\left(\mathcal{U}_L, \mathcal{T}_L\right), \mathcal{C}, \Theta\right)$ be image detection processes for a time interval $[t_0, t]$ and for a fixed number L of photons, respectively. Let $\mathcal{D}_{[t]} := \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \mathcal{D}_k := \left(\mathcal{U}_k, \mathcal{T}_k\right), k = 0, 1, \cdots$.*

1. Then, the probability of $\mathcal{D}_{[t]} = \emptyset$ and $N(t) = 0$ is given by

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

44

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

$$p_{[t]}\left(d_K, K\right) = e^{-\int_{t_0}^{t} \Lambda(\tau)d\tau} \prod_{k=1}^{K} \Lambda(\tau_k) \left[\prod_{l=1}^{K} p_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\right], \tag{3.2}$$

where $d_K \in \mathcal{C}^K \times \mathbb{R}_{[t]}^K$, $K = 1, 2, \cdots$, and $p_{U_l|T_l,\mathcal{D}_{l-1}}$ denotes the conditional probability density function of U_l, given $T_l, \mathcal{D}_{l-1}$, with $p_{U_1|T_1,\mathcal{D}_0}\left(r_1|\tau_1, d_0\right) := p_{U_1|T_1}\left(r_1|\tau_1\right)$.

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

$$p_L\left(d_L\right) = e^{-\int_{t_0}^{\tau_L} \Lambda(\tau)d\tau} \prod_{k=1}^{L} \Lambda(\tau_k) \left[\prod_{l=1}^{L} p_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\right], \quad d_L \in \mathcal{C}^L \times \mathbb{R}_{[\infty]}^L. \tag{3.3}$$

Proof. 1. According to [55, 56] and Lemma 4 (see Section B.6), the probability $P\left(\mathcal{D}_{[t]} = \emptyset, N(t) = 0\right)$ is given by

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

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

$$p_{[t]}\left(d_K, K\right) = p_{\mathcal{U}_K|\mathcal{T}_K,N(t)}\left(r_1, \cdots, r_K|\tau_1, \cdots, \tau_K, K\right) p_{\mathcal{T}_K|N(t)}\left(\tau_1, \cdots, \tau_K|K\right) P\left(N(t) = K\right), \tag{3.4}$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$ and $K = 1, 2, \cdots$. According to Lemma 4 (see Section B.6),

$$P\left(N(t) = K\right) = \frac{1}{K!} e^{-\int_{t_0}^{t} \Lambda(\tau)d\tau} \left(\int_{t_0}^{t} \Lambda(\tau)d\tau\right)^K, \quad K = 0, 1, \cdots, \tag{3.5}$$

and

$$p_{\mathcal{T}_K|N(t)}\left(\tau_1, \cdots, \tau_K|K\right) = \frac{K! \prod_{k=1}^{K} \Lambda(\tau_k)}{\left(\int_{t_0}^{t} \Lambda(\tau)d\tau\right)^K}, \quad K = 1, 2, \cdots, \tag{3.6}$$

and using the assumption of Definition 1 (it is assumed that $U_l, l = 1, \cdots, K$, is only dependent of the previous and current time points $T_1, \cdots, T_l$, and is independent of the future time points

$T_{l+1}, \cdots, T_K$ and the total number $N(t)$ of the detected photons),

$$p_{\mathcal{U}_K|\mathcal{T}_K, N(t)}\Big(r_1, \cdots, r_K | \tau_1, \cdots, \tau_K, K\Big)$$

$$= p_{\mathcal{U}_K|\mathcal{T}_K}\Big(r_1, \cdots, r_K | \tau_1, \cdots, \tau_K\Big)$$

$$= p_{U_K|T_K, \mathcal{D}_{K-1}}\Big(r_K | \tau_K, d_{K-1}\Big) p_{U_{K-1}|T_{K-1}, T_K, \mathcal{D}_{K-2}}\Big(r_{K-1} | \tau_{K-1}, \tau_K, d_{K-2}\Big)$$

$$\times \cdots \times p_{U_1|\mathcal{T}_K}\Big(r_1 | \tau_1, \cdots, \tau_K\Big)$$

$$= p_{U_K|T_K, \mathcal{D}_{K-1}}\Big(r_K | \tau_K, d_{K-1}\Big) p_{U_{K-1}|T_{K-1}, \mathcal{D}_{K-2}}\Big(r_{K-1} | \tau_{K-1}, d_{K-2}\Big)$$

$$\times \cdots \times p_{U_1|T_1}\Big(r_1 | \tau_1\Big)$$

$$= \prod_{l=1}^{K} p_{U_l|T_l, \mathcal{D}_{l-1}}\Big(r_l | \tau_l, d_{l-1}\Big), \tag{3.7}$$

where $p_{U_1|T_1, \mathcal{D}_0}\Big(r_1 | \tau_1, d_0\Big) := p_{U_1|T_1}\Big(r_1 | \tau_1\Big)$. By substituting Eqs. (3.5)-(3.7) into Eq. (3.4), we have

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

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

$$p_L\Big(d_L\Big) = p_{\mathcal{U}_L|\mathcal{T}_L}\Big(r_1, \cdots, r_L | \tau_1, \cdots, \tau_L\Big) p_{\mathcal{T}_L}\Big(\tau_1, \cdots, \tau_L\Big)$$

$$= \left[\prod_{l=1}^{L} p_{U_l|T_l, \mathcal{D}_{l-1}}\Big(r_l | \tau_l, d_{l-1}\Big)\right] p_{\mathcal{T}_L}\Big(\tau_1, \cdots, \tau_L\Big), \tag{3.8}$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$ and whereby Lemma 4 (see Section B.6),

$$p_{\mathcal{T}_L}\Big(\tau_1, \cdots, \tau_L\Big) = e^{-\int_{t_0}^{\tau_L} \Lambda(\tau)d\tau} \prod_{k=1}^{L} \Lambda(\tau_k). \tag{3.9}$$

By substituting Eq. (3.9) into Eq. (3.8), we have

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

46

and it completes the proof. □

Note that, as can be seen in the above theorem, the probability density function of an image detection process for a time interval $[t_0, t]$ depends on the integral of the photon detection rate $\Lambda(\tau), \tau \geq t_0$, over the time interval $[t_0, t]$, and the probability density function of an image detection process for a fixed number L of photons depends on the integral of the photon detection rate over the time interval $[t_0, \tau_L]$, where τ_L denotes the time point of the Lth (last) detected photon.

The probability density function of the location at which a photon emitted by the object of interest is detected, is referred to as the *image profile* of the object. So far we have made no assumptions about the specific functional form of the image profile of the object. In many practical cases, the image profile can be described as a scaled and shifted version of the image function. In such cases, an image function describes the image of an object on the detector plane at unit lateral magnification. Also, in general, the trajectory of the object can be described by a random process. In the following definition, we define image detection processes driven by a stochastic trajectory of the object and the image function for a fixed time interval and for a fixed number of photons.

Definition 2. *Let* $\mathcal{G}_{[t]}\Big(\big(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\big), \mathcal{C}, \Theta\Big)$ *and* $\mathcal{G}_L\Big(\big(\mathcal{U}_L, \mathcal{T}_L\big), \mathcal{C}, \Theta\Big)$ *be image detection processes for a time interval* $[t_0, t]$ *and for a fixed number* L *of photons, respectively. Let* $X(\tau), \tau \geq t_0$*, denote a 3D random process that describes the 3D stochastic trajectory of the object. Also, let* $\{f_x\}_{x \in \mathbb{R}^3}$ *defined on the detector* $\mathcal{C}$*, be a family of image profiles of an object located at* $x \in \mathbb{R}^3$ *in the object space. Assume that the current location of the detected photon, given the current location of the object, is independent of the previous locations and time points of the detected photons, i.e., for all* $x \in \mathbb{R}^3$,

$$p_{U_l|X(T_l),T_l,\mathcal{D}_{l-1}}\Big(r_l|x, \tau_l, d_{l-1}\Big) = p_{U_l|X(\tau_l)}\Big(r_l|x\Big) := f_x(r_l), \quad r_l \in \mathcal{C},$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ *for* $\mathcal{G}_{[t]}$*,* $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$ *for* $\mathcal{G}_L$*,* $p_{U_l|X(T_l),T_l,\mathcal{D}_{l-1}}$ *is the conditional probability density function of* U_l*, given* $X(T_l), T_l, \mathcal{D}_{l-1}$*, and* $p_{U_l|X(\tau_l)}$ *denotes the conditional probability density function of* U_l*, given* $X(\tau_l)$*. Assume that there exists a function* $q_{z_0} \colon \mathbb{R}^2 \mapsto \mathbb{R}, z_0 \in \mathbb{R}$*, such that*

for an invertible matrix $M \in \mathbb{R}^{2\times2}$ and $x := (x_0, y_0, z_0) \in \mathbb{R}^3$,

$$f_x(r) := \frac{1}{|\det(M)|} q_{z_0}\left(M^{-1}r - (x_0, y_0)^T\right), \quad r \in \mathcal{C}. \tag{3.10}$$

In the above equation, q_{z_0}, which is referred to as the image function, *is a function that describes, at unit lateral magnification, the image of the object in the detector plane when the object is located at $(0, 0, z_0)$ in the object space.*

Image detection processes $\mathcal{G}_{[t]}\left(X, (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), q, \mathcal{C}, \Theta\right)$ and $\mathcal{G}_L\left(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta\right)$ driven by the stochastic trajectory X and image function q for a time interval $[t_0, t]$ and for a fixed number L of photons *are defined as the spatio-temporal processes $\mathcal{G}_{[t]}$ and $\mathcal{G}_L$, respectively.*

In the classical case of a measurement error, the image function q_{z_0} is defined as a function of $\left(r - M(x_0, y_0)^T\right)$, which is the deviation between two locations in the image space. Here, however, in order to be consistent with our previous framework developed for a static object, q_{z_0} is defined as a function of $\left(M^{-1}r - (x_0, y_0)^T\right)$, which is the difference between two points in the object space.

We next illustrate specific image functions that describe the image of a point source. According to the optical diffraction theory, when a point source is in-focus with respect to the detector, the intensity distribution of the image of the point source is described by an Airy profile given by [7] (see Fig. 3.2(a))

$$q(x, y) = \frac{J_1^2\left(\frac{2\pi n_a}{\lambda}\sqrt{x^2 + y^2}\right)}{\pi(x^2 + y^2)}, \quad (x, y) \in \mathbb{R}^2, \tag{3.11}$$

where n_a denotes the numerical aperture of the objective lens, λ denotes the emission wavelength of the molecule, and J_1 denotes the first order Bessel function of the first kind. The 2D Gaussian profile, on the other hand, which has been widely used to approximate the Airy profile, is given by

$$q(x, y) = \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2}\left(\frac{x^2+y^2}{\sigma^2}\right)}, \quad (x, y) \in \mathbb{R}^2, \tag{3.12}$$

where $\sigma > 0$.

For an out-of-focus point source, the image function can be obtained by the classical Born and Wolf model given by [36]

$$q_{z_0}(x,y) = \frac{4\pi n_a^2}{\lambda^2} \left| \int_0^1 J_0\left(\frac{2\pi n_a}{\lambda}\sqrt{x^2+y^2}\rho\right) e^{\frac{j\pi n_a^2 z_0}{n_o\lambda}\rho^2} \rho d\rho \right|^2, \quad (x,y) \in \mathbb{R}^2, \tag{3.13}$$

where J_0 is the zeroth-order Bessel function of the first kind, n_o is the refractive index of the objective lens immersion medium, and $z_0 \in \mathbb{R}$ is the z-location of the point source on the optical axis in the object space. When the point source is in-focus with respect to the detector, i.e., it lies in the object plane, then $z_0 = 0$ and Eqs. (3.11) and (3.13) are equivalent.
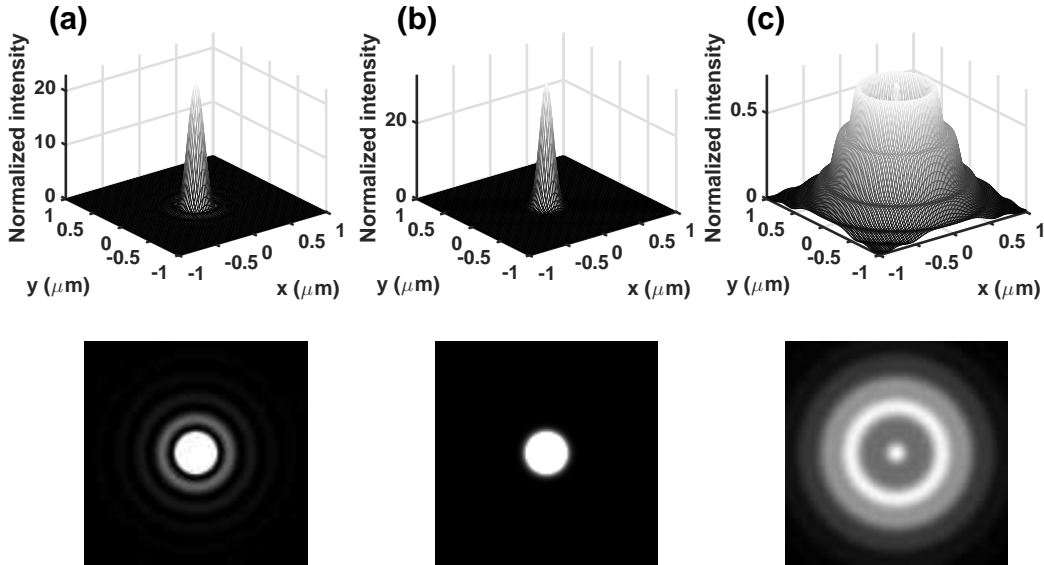
Figure 3.2: Image function examples. (a) Airy and (b) symmetric Gaussian profiles, which describe the images of an in-focus point source, simulated by Eqs. (3.11) and (3.12), respectively, with the parameters given in Section 3.4.1. (c) Born and Wolf profile simulated by Eq. (3.13) with the out-of-focus level $z_0 = 1$ μm, and the parameters given in Section 3.4.1.

We calculate $p_{U_l|T_l,\mathcal{D}_{l-1}}, l = 1, 2, \cdots$, for more general cases. In the following corollary to

Theorem 2, by describing these conditional probability density functions in terms of the image function, we derive expressions for the probability density functions of the image detection processes driven by the stochastic trajectory X and image function q for a time interval $[t_0, t]$ and for a fixed number L of photons.

Corollary 1. *Let $\mathcal{G}_{[t]}\left(X, \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), q, \mathcal{C}, \Theta\right)$ (or $\mathcal{G}_L\left(X, \left(\mathcal{U}_L, \mathcal{T}_L\right), q, \mathcal{C}, \Theta\right)$) be an image detection process driven by the stochastic trajectory X and image function q for a time interval $[t_0, t]$ (or for a fixed number L of photons). Then, the conditional probability density function $p_{U_l|T_l, \mathcal{D}_{l-1}}, l = 1, 2, \cdots$, in Eq. (3.2) (or in Eq. (3.3)) of Theorem 2 is given by, for $x := (x_0, y_0, z_0) \in \mathbb{R}^3$,*

$$
\begin{aligned}
p_{U_l|T_l, \mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) &= \int_{\mathbb{R}^3} f_x\left(r_l\right) p_{pr_l}\left(x|\tau_l, d_{l-1}\right) dx \\
&= \frac{1}{|\det(M)|} \int_{\mathbb{R}^3} q_{z_0}\left(M^{-1}r_l - (x_0, y_0)\right) p_{pr_l}\left(x|\tau_l, d_{l-1}\right) dx,
\end{aligned} \qquad (3.14)
$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$), $p_{pr_l} := p_{X(T_l)|T_l, \mathcal{D}_{l-1}}$ denotes the distribution of the prediction of the object location, $p_{pr_1}\left(x|\tau_1, d_0\right) := p_{pr_1}\left(x|\tau_1\right)$, and $f_x, x \in \mathbb{R}^3$, is the image profile of an object located at x in the object space.

Proof. The conditional probability density function $p_{U_l|T_l, \mathcal{D}_{l-1}}$ in Eqs. (3.2) and (3.3) of Theorem 2 can be written as, for $x := (x_0, y_0, z_0) \in \mathbb{R}^3$,

$$
\begin{aligned}
p_{U_l|T_l, \mathcal{D}_{l-1}}&\left(r_l|\tau_l, d_{l-1}\right) \\
&= \int_{\mathbb{R}^3} p_{U_l, X(T_l)|T_l, \mathcal{D}_{l-1}}\left(r_l, x|\tau_l, d_{l-1}\right) dx \\
&= \int_{\mathbb{R}^3} p_{U_l|X(T_l), T_l, \mathcal{D}_{l-1}}\left(r_l|x, \tau_l, d_{l-1}\right) p_{X(T_l)|T_l, \mathcal{D}_{l-1}}\left(x|\tau_l, d_{l-1}\right) dx \\
&= \int_{\mathbb{R}^3} f_x\left(r_l\right) p_{pr_l}\left(x|\tau_l, d_{l-1}\right) dx \\
&= \frac{1}{|\det(M)|} \int_{\mathbb{R}^3} q_{z_0}\left(M^{-1}r_l - (x_0, y_0)\right) p_{pr_l}\left(x|\tau_l, d_{l-1}\right) dx,
\end{aligned} \qquad (3.15)
$$

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

As can be seen in the above corollary, the expression of the probability density function of the image detection process depends on the distribution $p_{pr_l}, l = 1, 2, \cdots$, of the prediction of the object location, given the previous locations of the detected photons and the current and previous time points. In the following section, we introduce linear stochastic systems and calculate $p_{pr_l}, l = 1, 2, \cdots$, for them.

In Theorem 2, we expressed the probability density functions of image detection processes in terms of conditional probability densities $p_{U_l|T_l,\mathcal{D}_{l-1}}, l = 1, 2, \cdots$, of the locations of the detected photons, given the previous locations and the current and previous time points of the detected photons. In particular, for an object with a deterministic trajectory or a static object, the conditional probability densities $p_{U_l|T_l,\mathcal{D}_{l-1}}, l = 1, 2, \cdots$, are given as follows. For an object with deterministic trajectory $X(\tau) \in \mathbb{R}^3, \tau \geq t_0$, we have

$$p_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) = p_{U_l|T_l}(r_l|\tau_l) := f_{X(\tau_l)}(r_l). \tag{3.16}$$

Also, for a static object with position $X_0 \in \mathbb{R}^3$, we have

$$p_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) = p_{U_l}(r_l) := f_{X_0}(r_l). \tag{3.17}$$

3.3 Linear stochastic systems

In general, the motion of an object in cellular environments is subject to different types of forces, e.g., deterministic forces due to the environment and random forces due to random collisions with other objects [29, 30]. The 3D random variable $X(\tau)$ denotes the location of the object at time $\tau \geq t_0$. Then, the motion of the object is assumed to be modeled through a general state space system with state $\tilde{X}(\tau) \in \mathbb{R}^k, \tau \geq t_0$, as

$$\tilde{X}(\tau_{l+1}) = \tilde{\phi}(\tau_l, \tau_{l+1})\tilde{X}(\tau_l) + \tilde{W}(\tau_l, \tau_{l+1}), \quad \tau_0 := t_0 \leq \tau_1 < \cdots < \tau_{l+1} < \cdots, \tag{3.18}$$

where we assume that there exists a function $g \colon \mathbb{R}^k \mapsto \mathbb{R}^3$ such that $X(\tau) = g\left(\tilde{X}(\tau)\right), \tau \geq t_0$, $\tilde{\phi}(\tau_l, \tau_{l+1}) \in \mathbb{R}^{k \times k}$ is a state transition matrix, and $\left\{\tilde{W}(\tau_l, \tau_{l+1}), \; l = 1, 2, \cdots\right\}$ is a sequence of k-dimensional random variables with probability density functions $p_{\tilde{W}(\tau_l, \tau_{l+1})}$.

The general system of discrete evolution equations described by Eq. (3.18) can arise, for example, from stochastic differential equations [31]. In particular, in many biological applications, solutions of linear stochastic differential equations are good fits to experimental single-molecule trajectories [31]. As an example, we assume that the motion of the object of interest, e.g., a single molecule, is described by the following linear vector stochastic differential equation [57, 34]

$$dX(\tau) = (V + F(\tau)X(\tau))\, d\tau + G(\tau)dB(\tau), \qquad \tau \geq t_0, \tag{3.19}$$

where the 3D random process $X(\tau)$ describes the location of the object at time $\tau \geq t_0$, $F \in \mathbb{R}^{3 \times 3}$ and $G \in \mathbb{R}^{3 \times r}$ are continuous matrix time-functions, $V \in \mathbb{R}^3$ is a zero order drift, and $\{B(\tau) \in \mathbb{R}^r, \; \tau \geq t_0\}$ is a random process [58].

Here, we assume that $\{B(\tau) \in \mathbb{R}^r, \tau \geq t_0\}$ is an r-vector Brownian motion (Wiener) process with $E\left\{dB(\tau)dB(\tau)^T\right\} = I_{r \times r}, \tau \geq t_0$, where $I_{r \times r}$ is the $r \times r$ identity matrix [32, 33, 34]. Then, the solution of Eq. (3.19) at discrete time points $\tau_0 := t_0 \leq \tau_1 < \cdots < \tau_{l+1} < \cdots$ is given by [59]

$$X(\tau_{l+1}) = \phi(\tau_l, \tau_{l+1})X(\tau_l) + a(\tau_l, \tau_{l+1}) + W_g(\tau_l, \tau_{l+1}), \tag{3.20}$$

where the continuous matrix time-function $\phi \in \mathbb{R}^{3 \times 3}$ is given by

$$\frac{d\phi(t, \tau)}{dt} = F(t)\phi(t, \tau), \quad \phi(\tau, \tau) = I_{3 \times 3}, \quad \text{for all } t, \tau \geq t_0,$$

$$\phi(t, \tau)\phi(\tau, \psi) = \phi(t, \psi), \quad \text{for all } t, \tau, \psi \geq t_0,$$

and the vector $a(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3 \times 1}$ is given by

$$a(\tau_l, \tau_{l+1}) := \int_{\tau_l}^{\tau_{l+1}} \phi(\tau, \tau_{l+1}) V d\tau.$$

Also, in this case, $\left\{ W_g(\tau_l, \tau_{l+1}) := \int_{\tau_l}^{\tau_{l+1}} \phi(\tau, \tau_{l+1}) G(\tau) dB(\tau), l = 1, 2, \cdots \right\}$ is a zero mean white Gaussian sequence with covariance $Q_g(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3 \times 3}$ given by

$$Q_g(\tau_l, \tau_{l+1}) = \int_{\tau_l}^{\tau_{l+1}} \phi(\tau, \tau_{l+1}) G(\tau) G^T(\tau) \phi^T(\tau, \tau_{l+1}) d\tau.$$

By letting $\tilde{X}(\tau) := \begin{bmatrix} X(\tau) \\ 1 \end{bmatrix} \in \mathbb{R}^4$, $X(\tau) = g\left(\tilde{X}(\tau)\right) := \begin{bmatrix} I_{3 \times 3} & 0_{3 \times 1} \end{bmatrix} \tilde{X}(\tau), \tau \geq t_0$, where $0_{3 \times 1}$ is the 3×1 zero matrix, we obtain expressions of the form of Eq. (3.18), where the continuous matrix time-function $\tilde{\phi} \in \mathbb{R}^{4 \times 4}$ is given by

$$\tilde{\phi}(\tau_l, \tau_{l+1}) = \begin{bmatrix} \phi(\tau_l, \tau_{l+1}) & a(\tau_l, \tau_{l+1}) \\ 0_{1 \times 3} & 1 \end{bmatrix},$$

and $\tilde{W}(\tau_l, \tau_{l+1}) = \begin{bmatrix} W_g(\tau_l, \tau_{l+1}) \\ 0 \end{bmatrix} \in \mathbb{R}^4$.

As an another example, for pure diffusion motion, when V and $F(\tau), \tau \geq 0$, in Eq. (3.19) are equal to zero, the discrete motion model is given by

$$X(\tau_{l+1}) = X(\tau_l) + W_g(\tau_l, \tau_{l+1}), \quad \tau_0 := t_0 \leq \tau_1 < \cdots < \tau_{l+1} < \cdots . \tag{3.21}$$

Setting $\tilde{X}(\tau) := X(\tau), \tau \geq t_0$, with g the identity function, $\tilde{\phi}(\tau_l, \tau_{l+1}) = \phi(\tau_l, \tau_{l+1}) = I_{3 \times 3}$, and $\tilde{W}(\tau_l, \tau_{l+1}) = W_g(\tau_l, \tau_{l+1})$, we again obtain expressions of the form of Eq. (3.18).

The above discussion motivates us to model the motion of the object, in the following definition, by Eq. (3.18) with, in general, an arbitrary distributed process noise $\tilde{W}$. In particular, we also consider the special case of Gaussian distributed process noise $\tilde{W}_g$, separately.

Definition 3. *Let* $\mathcal{G}_{[t]}\left(X,\left(\mathcal{U}_{[t]},\mathcal{T}_{[t]}\right),q,\mathcal{C},\Theta\right)$ *and* $\mathcal{G}_L\left(X,\left(\mathcal{U}_L,\mathcal{T}_L\right),q,\mathcal{C},\Theta\right)$ *be image detection processes driven by a stochastic trajectory* X *and image function* q *for a fixed time interval* $[t_0,t]$ *and for a fixed number* L *of photons. Let* $p_{X(t_0)}$ *be the probability density function of the initial location* $X(t_0)$ *of the object. We assume that*

a. the motion of the object is modeled through a general state space system with state $\tilde{X}(\tau) \in \mathbb{R}^k, \tau \geq t_0$, *as*

$$\tilde{X}(\tau_{l+1}) = \tilde{\phi}(\tau_l,\tau_{l+1})\tilde{X}(\tau_l) + \tilde{W}(\tau_l,\tau_{l+1}), \quad \tau_0 := t_0 \leq \tau_1 < \cdots < \tau_{l+1} < \cdots, \tag{3.22}$$

where we assume that there exists a function $g\colon \mathbb{R}^k \mapsto \mathbb{R}^3$ *such that* $X(\tau) = g\left(\tilde{X}(\tau)\right), \tau \geq t_0$, $\tilde{\phi}(\tau_l,\tau_{l+1}) \in \tilde{\Phi}$, *where* $\tilde{\Phi} = \left\{\tilde{\phi}(\tau,\psi)\right\}_{\psi > \tau \geq t_0}$ *is a family of* $k \times k$ *invertible real-valued state-transition matrices, and* $\left\{\tilde{W}(\tau_l,\tau_{l+1}), l = 0,1,2,\cdots\right\}$ *is a process noise sequence of independent* k-*dimensional random variables with probability density functions* $p_{\tilde{W}(\tau_l,\tau_{l+1})}$.

b. We assume that

$$U_l = Z\left(X(\tau_l)\right), \quad l = 1,2,\cdots, \tag{3.23}$$

where $\left\{Z\left(X(\tau_l)\right), l = 1,2,\cdots\right\}$ *is a measurement sequence of independent 2D random variables with probability density functions* $p_{Z(X(\tau_l))} = f_{X(\tau_l)}$, *where* Z *is a random function that maps the object space into the image space,* $f_{X(\tau_l)}$ *is the image profile of an object located at* $X(\tau_l)$ *defined in Definition 2 and*

c. We assume that the sequences $\left\{\tilde{W}(\tau_l,\tau_{l+1}), l = 0,1,\cdots\right\}$, $\left\{Z\left(X(\tau_l)\right), l = 1,2,\cdots\right\}$, *and* $\tilde{X}(t_0)$ *are independent of one another.*

The image detection process $\mathcal{G}_{[t]}\left(X,\left(\mathcal{U}_{[t]},\mathcal{T}_{[t]}\right),q,\mathcal{C},\Theta\right)$ *(or* $\mathcal{G}_L\left(X,\left(\mathcal{U}_L,\mathcal{T}_L\right),q,\mathcal{C},\Theta\right)$) *with the additional properties (a)-(c) is called an* image detection process with expanded state space $\tilde{X}$ *for a time interval* $[t_0,t]$ *(or for a fixed number* L *of photons), and is denoted by* $\mathcal{G}_{[t]}\left(\left(\tilde{X},g,\tilde{W},Z\right),\left(\mathcal{U}_{[t]},\mathcal{T}_{[t]}\right),\tilde{\Phi},\mathcal{C},\Theta\right)$ *(or* $\mathcal{G}_L\left(\left(\tilde{X},g,\tilde{W},Z\right),\left(\mathcal{U}_L,\mathcal{T}_L\right),\tilde{\Phi},\mathcal{C},\Theta\right)$).

We further assume that

α. $\left\{\tilde{W}_g(\tau_l, \tau_{l+1}) := \tilde{W}(\tau_l, \tau_{l+1}), l = 0, 1, \cdots\right\}$ is a white Gaussian sequence with mean zero and covariance matrix $\tilde{Q}_g(\tau_l, \tau_{l+1}) \in \mathbb{R}^{k \times k}, \tilde{Q}_g(\tau_l, \tau_{l+1}) > 0$,

β.

$$Z(X(\tau_l)) = M'X(\tau_l) + Z_{g,l} \quad l = 1, 2, \cdots, \tag{3.24}$$

where $M' := \begin{bmatrix} M & 0_{2\times 1} \end{bmatrix} \in \mathbb{R}^{2\times 3}$, in which $M \in \mathbb{R}^{2\times 2}$ is an invertible magnification matrix used in the definition of the image function (Eq. (3.10)), where $0_{2\times 1}$ is the 2×1 zero matrix, and $\{Z_{g,l}, l = 1, 2, \cdots\}$ is a measurement noise sequence of independent 2D Gaussian random variables with mean zero and the same covariance matrix $\Sigma_g \in \mathbb{R}^{2\times 2}, \Sigma_g > 0$.

γ. We assume that the initial state $\tilde{X}(t_0)$ is Gaussian distributed with mean $\tilde{x}_0 \in \mathbb{R}^k$ and covariance matrix $\tilde{P}_0 \in \mathbb{R}^{k \times k}, \tilde{P}_0 > 0$.

If, in addition, an expanded image detection process has the properties (α)-(γ), it is called an image detection process with expanded state space $\tilde{X}$ and Gaussian process and measurement noise models, and is denoted by $\mathcal{G}^g_{[t]}\left(\left(\tilde{X}, g, \tilde{W}_g, Z_g\right), \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$ (or $\mathcal{G}^g_L\left(\left(\tilde{X}, g, \tilde{W}_g, Z_g\right), \left(\mathcal{U}_L, \mathcal{T}_L\right), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$) for a time interval $[t_0, t]$ (or for a fixed number L of photons).

In Corollary 1, we calculated the probability density function of the image detection process in terms of the image function q and the distribution $p_{pr_l}, l = 1, 2, \cdots$, of the prediction of the object location, given the previous locations of the detected photons and the current and previous time points. In the following theorem, for a linear stochastic system and Gaussian process and measurement noise, we calculate these distributions using the Kalman filter formulae. Also, for a more general Markov motion model described by a first order system with arbitrary distributed process and measurement noise, we calculate these distributions recursively.

Theorem 3. Let $\mathcal{G}_{[t]}\left(\left(\tilde{X}, g, \tilde{W}, Z\right), \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \tilde{\Phi}, \mathcal{C}, \Theta \right)$ (or $\mathcal{G}_L\left(\left(\tilde{X}, g, \tilde{W}, Z\right), \left(\mathcal{U}_L, \mathcal{T}_L\right), \tilde{\Phi}, \mathcal{C}, \Theta \right)$) be an image detection process with expanded state space $\tilde{X}$ for a time interval $[t_0, t]$ (or for a

fixed number L of photons). Let $\mathcal{D}_k := (\mathcal{U}_k, \mathcal{T}_k)$, $k = 0, 1, \cdots$, and

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

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

1. Then, $\tilde{p}_{pr_l}$, $l = 0, 1, 2, \cdots$, can be calculated through the following recursive formula

$$\tilde{p}_{pr_{l+1}}\left(\tilde{x}|\tau_{l+1}, d_l\right) = \frac{1}{|\det\left(\phi(\tau_l, \tau_{l+1})\right)|} \int_{\mathbb{R}^k} \tilde{p}_{fi_l}\left(\phi^{-1}(\tau_l, \tau_{l+1})\tilde{x}_o|d_l\right) p_{\tilde{W}(\tau_l, \tau_{l+1})}\left(\tilde{x} - \tilde{x}_o\right) d\tilde{x}_o, \quad (3.25)$$

where $d_0 = \emptyset$, and the distribution $\tilde{p}_{fi_l}\left(\tilde{x}|d_l\right) := p_{\tilde{X}(T_l)|\mathcal{D}_l}\left(\tilde{x}|d_l\right)$ of the filtered object location is given by

$$\tilde{p}_{fi_l}\left(\tilde{x}|d_l\right) = \frac{p_{Z(g(\tilde{x}))}\left(r_l\right) \tilde{p}_{pr_l}\left(\tilde{x}|\tau_l, d_{l-1}\right)}{\int_{\mathbb{R}^k} p_{Z(g(\tilde{x}_o))}\left(r_l\right) \tilde{p}_{pr_l}\left(\tilde{x}_o|\tau_l, d_{l-1}\right) d\tilde{x}_o}. \quad (3.26)$$

2.1. Let $\mathcal{G}^g_{[t]}\left(\left(\tilde{X}, g, \tilde{W}_g, Z_g\right), \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), \tilde{\Phi}, M', \mathcal{C}, \Theta\right)$ (or $\mathcal{G}^g_L\left(\left(\tilde{X}, g, \tilde{W}_g, Z_g\right), \left(\mathcal{U}_L, \mathcal{T}_L\right), \tilde{\Phi}, M', \mathcal{C}, \Theta\right)$) be an image detection process with expanded state space $\tilde{X}$ and Gaussian process and measurement noise models for a time interval $[t_0, t]$ (or for a fixed number L of photons). Assume that

$$X(\tau) = g(\tilde{X}(\tau)) := G\tilde{X}(\tau), \quad \tau \geq t_0,$$

where $G \in \mathbb{R}^{3 \times k}$, and let $C := M'G$. Then, for $l = 0, 1, \cdots$,

$$\tilde{p}_{pr_{l+1}}\left(\tilde{x}|d_l, \tau_{l+1}\right) = \frac{1}{2\pi \left[\det(P^l_{l+1})\right]^{1/2}} \exp\left(-\frac{1}{2}(\tilde{x} - \hat{x}^l_{l+1})^T \left(P^l_{l+1}\right)^{-1} (\tilde{x} - \hat{x}^l_{l+1})\right), \quad \tilde{x} \in \mathbb{R}^k,$$

$$(3.27)$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$), $\hat{x}^0_1 = \tilde{\phi}(\tau_0, \tau_1)\tilde{x}_0$, $P^0_1 = \tilde{\phi}(\tau_0, \tau_1)\tilde{P}_0\tilde{\phi}^T(\tau_0, \tau_1) + \tilde{Q}_g(\tau_0, \tau_1)$,

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

$$\hat{x}^l_{l+1} = \tilde{\phi}(\tau_l, \tau_{l+1})\hat{x}^l_l,$$

$$P^l_{l+1} = \tilde{\phi}(\tau_l, \tau_{l+1})P^l_l\tilde{\phi}^T(\tau_l, \tau_{l+1}) + \tilde{Q}_g(\tau_l, \tau_{l+1}), \quad (3.28)$$

with

$$K_l = P^{l-1}_l C^T \left(C P^{l-1}_l C^T + \Sigma_g \right)^{-1},$$

$$\hat{x}^l_l = \hat{x}^{l-1}_l + K_l(r_l - C\hat{x}^{l-1}_l),$$

$$P^l_l = P^{l-1}_l - K_l C P^{l-1}_l. \quad (3.29)$$

2.2. Moreover, the conditional probability density function $p_{U_l|T_l, \mathcal{D}_{l-1}}$ is given by

$$p_{U_l|T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1} \right) = \frac{1}{2\pi \left[\det(R_l) \right]^{1/2}} \exp\left(-\frac{1}{2}(r_l - \hat{r}_l)^T R_l^{-1}(r_l - \hat{r}_l) \right), \quad (3.30)$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$), $R_l := C P^{l-1}_l C^T + \Sigma_g$ and $\hat{r}_l := C\hat{x}^{l-1}_l$.

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

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

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

1. Then, $\tilde{p}_{pr_l}$ can be calculated through the following steps:

Step 1. For $l = 0$, Eq. (3.22) becomes

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

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

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

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

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

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

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

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

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

i.e.,

$$
\begin{aligned}
p_{\tilde{X}(T_l)|\mathcal{D}_l}\left(\tilde{x}|d_l\right) &= \frac{p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\left(r_l|\tilde{x},\tau_l,d_{l-1}\right)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\left(\tilde{x}|\tau_l,d_{l-1}\right)}{p_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)} \\
&= \frac{p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\left(r_l|\tilde{x},\tau_l,d_{l-1}\right)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\left(\tilde{x}|\tau_l,d_{l-1}\right)}{\int_{\mathbb{R}^k} p_{U_l,\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\left(r_l,\tilde{x}_o|\tau_l,d_{l-1}\right)d\tilde{x}_o} \\
&= \frac{p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\left(r_l|\tilde{x},\tau_l,d_{l-1}\right)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\left(\tilde{x}|\tau_l,d_{l-1}\right)}{\int_{\mathbb{R}^k} p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\left(r_l|\tilde{x}_o,\tau_l,d_{l-1}\right)p_{\tilde{X}(T_l)|T_l,\mathcal{D}_{l-1}}\left(\tilde{x}_o|\tau_l,d_{l-1}\right)d\tilde{x}_o}.
\end{aligned} \tag{3.31}
$$

Since initial location of the object, observation noise, and process noise are mutually independent, according to Eq. (3.23) and Theorem 2.7 of [59], we have

$$
p_{U_l|\tilde{X}(T_l),T_l,\mathcal{D}_{l-1}}\left(r_l|\tilde{x},\tau_l,d_{l-1}\right) = p_{U_l|\tilde{X}(\tau_l)}\left(r_l|\tilde{x}\right) = p_{Z(\tilde{x})}\left(r_l\right), \quad \tilde{x}\in\mathbb{R}^k. \tag{3.32}
$$

Therefore, by substituting Eq. (3.32) into Eq. (3.31) (note that we calculated $\tilde{p}_{pr_l}$ in the previous step),

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

Step $2l + 1$. By conditioning the both sides of Eq. (3.22) on $T_{l+1} = \tau_{l+1}$ and $\mathcal{D}_l = d_l$, we have , for $l = 1, 2, \cdots$,

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

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

$$p_{\tilde{X}(T_{l+1})|T_{l+1}, \mathcal{D}_l}\left(\tilde{x}|\tau_{l+1}, d_l\right)$$

$$= p_{\tilde{\phi}(\tau_l, \tau_{l+1})\tilde{X}(T_l)|\mathcal{D}_l}\left(\tilde{x}|d_l\right) * p_{\tilde{W}(\tau_l, \tau_{l+1})}\left(\tilde{x}\right)$$

$$= \int_{\mathbb{R}^k} p_{\tilde{\phi}(\tau_l, \tau_{l+1})\tilde{X}(T_l)|\mathcal{D}_l}\left(\tilde{x}_o|d_l\right) p_{\tilde{W}(\tau_l, \tau_{l+1})}\left(\tilde{x} - \tilde{x}_o\right) d\tilde{x}_o$$

$$= \frac{1}{\left|\det\left(\tilde{\phi}(\tau_l, \tau_{l+1})\right)\right|} \int_{\mathbb{R}^k} p_{\tilde{X}(T_l)|\mathcal{D}_l}\left(\tilde{\phi}^{-1}(\tau_l, \tau_{l+1})\tilde{x}_o|d_l\right) p_{\tilde{W}(\tau_l, \tau_{l+1})}\left(\tilde{x} - \tilde{x}_o\right) d\tilde{x}_o,$$

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

$$\tilde{p}_{pr_{l+1}}\left(\tilde{x}|\tau_{l+1}, d_l\right) = \frac{1}{\left|\det\left(\tilde{\phi}(\tau_l, \tau_{l+1})\right)\right|} \int_{\mathbb{R}^k} \tilde{p}_{fi_l}\left(\tilde{\phi}^{-1}(\tau_l, \tau_{l+1})\tilde{x}_o|d_l\right) p_{\tilde{W}(\tau_l, \tau_{l+1})}\left(\tilde{x} - \tilde{x}_o\right) d\tilde{x}_o. \tag{3.33}$$

2.1. See Theorem 7.2 of [59].

2.2. Setting $C := M'G$, Eq. (3.24) becomes

$$U_l = Z(\tilde{X}(\tau_l)) = C\tilde{X}(\tau_l) + Z_{g,l}, \quad l = 1, 2, \cdots, \tag{3.34}$$

where

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

Since $Z_{g,l}$ is independent of $\mathcal{D}_{l-1}$, T_l and $\tilde{X}$, then, according to Eq. (3.34), $Z(\tilde{X}(\tau_l))$ is the sum of two independent Gaussian random variables and its probability density function is given by

$$p_{U_l|T_l, \mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) = p_{Z(\tilde{X}(\tau_l))|T_l, \mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)$$

$$= \frac{1}{2\pi \left[\det\left(R_l\right)\right]^{1/2}} \exp\left(-\frac{1}{2}(r_l - \hat{r}_l)^T R_l^{-1}(r_l - \hat{r}_l)\right), \tag{3.35}$$

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

3.4   Maximum likelihood estimation

The main purpose of the presented materials in the previous section is to provide a mathematical framework to estimate the parameters of interest, such as the parameters of the model that describes the motion of a moving object with stochastic trajectories, from the acquired data. Here, we use the maximum likelihood estimation approach as follows. For a general parameter estimation problem, denoting the acquired data by $\bar{d} \in \mathbb{R}^m, m = 1, 2, \cdots$, the maximum likelihood estimate $\hat{\theta}_{mle}$ of $\theta \in \Theta$, if it exists, is given by

$$\hat{\theta}_{mle} = \operatorname*{argmin}_{\theta \in \Theta} \Big(-\log \mathcal{L}(\theta|\bar{d}) \Big),$$

where $\mathcal{L}$ denotes the likelihood function. In our specific problem, the acquired data for the fixed time interval $[t_0, t]$ acquisition case is denoted by $\bar{d}_K \in \mathcal{C}^K \times \mathbb{R}^K_{[t]}, K = 0, 1, \cdots$. Then, the likelihood function $\mathcal{L}_{[t]}$ of $\mathcal{G}_{[t]}\Big(\big(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\big), \mathcal{C}, \Theta\Big)$ is given by, according to Theorem 2 (see also [55, 56]), for $\theta \in \Theta$,

$$\mathcal{L}_{[t]}(\theta|\bar{d}_K) = \begin{cases} e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}, & K = 0, \\ e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau} \prod_{k=1}^K \Lambda_\theta(\bar{\tau}_k) \left[\prod_{l=1}^K p^\theta_{U_l|T_l, \mathcal{D}_{l-1}}\Big(\bar{r}_l|\bar{\tau}_l, \bar{d}_{l-1}\Big) \right], & K = 1, 2, \cdots, \end{cases} \tag{3.36}$$

and the likelihood function $\mathcal{L}_L$ of $\mathcal{G}_L\Big(\big(\mathcal{U}_L, \mathcal{T}_L\big), \mathcal{C}, \Theta\Big)$ is given by

$$\mathcal{L}_L(\theta|\bar{d}_L) = p^\theta_L(\bar{d}_L) = e^{-\int_{t_0}^{\bar{\tau}_L} \Lambda_\theta(\tau)d\tau} \prod_{k=1}^L \Lambda_\theta(\bar{\tau}_k) \left[\prod_{l=1}^L p^\theta_{U_l|T_l, \mathcal{D}_{l-1}}\Big(\bar{r}_l|\bar{\tau}_l, \bar{d}_{l-1}\Big) \right], \tag{3.37}$$

where $\bar{d}_L \in \mathcal{C}^L \times \mathbb{R}^L_{[\infty]}, L = 1, 2, \cdots$.

In appendix Section B.1, we provide an example to illustrate our results for the specific case that the motion model is described by a simple linear stochastic differential equation and the parameter vector contains the drift and diffusion coefficients.

In the following, we present and discuss the results of the proposed maximum likelihood estimation method when applied to simulated data sets containing linear stochastic trajectories of a single molecule.

3.4.1 Simulated parameters

To analyze the performance of the proposed maximum likelihood estimation method, we simulated different data sets using parameters commonly used in single molecule experiments. Unless otherwise stated, the images of in-focus and out-of-focus molecules were generated with Airy and Born and Wolf profiles (Eqs. (3.11) and (3.13)), respectively, where $n_a = 1.4, \lambda = 520$ nm, $n_o = 1.515$, and $z_0 = 1$ μm. For the Gaussian measurement case, the image of a molecule was generated with a zero-mean Gaussian measurement noise with the probability density function given by Eq. (3.12), where $\sigma = 70$ nm, which is related to the corresponding Airy profile.

Furthermore, a measurement (magnification) matrix $M = 100I_{2\times2}$ was assumed to map the object space to the image space.

3.4.2 Estimation results

Using simulated data sets, we first examine the performance of the maximum likelihood estimation method used to estimate the parameters of the linear motion model of a moving molecule in terms of the bias of the method. The bias is assessed by the average of the deviations of the estimates from the true value. For this purpose, we simulated 100 data sets, each containing a trajectory of an out-of-focus molecule simulated using Eqs. (B.1) and (3.23), with the Born and Wolf profile (Eq. (3.13)) and the parameters given in Section 3.4.1, with a mean photon count of 500 photons in the time interval $[0, 100]$ ms, where the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1$ μm^2/s. In Figs. 3.3(a) and 3.3(b), an example of a molecule trajectory in the object space and its image in the image space are shown. For these data sets, we calculated the maximum likelihood estimates of the diffusion and the first order drift coefficients, separately. For this purpose, we needed to obtain the distributions of the prediction in the likelihood function expressions (Eqs. (3.36) and (3.37)) through Eqs. (3.25) and (3.26), which in general is a computationally expensive problem. We approximated the distributions of the prediction using a sequential Monte Carlo algorithm proposed in [45]. The overall approach is explained in appendix Section B.2 in detail. In Figs. 3.3(c) and 3.3(d), the differences between the maximum likelihood estimates

(a) Y-location of molecule (μm) vs X-location of molecule (μm)

(b) Y-location of detected photons (μm) vs X-location of detected photons (μm)

(c) Estimated − true diffusion coefficient (μm^2/s) vs Data set

(d) Estimated − true drift coefficient (1/s) vs Data set

Figure 3.3: Analysis of the error of diffusion coefficient and drift coefficient estimates produced by the maximum likelihood estimation method for the Born and Wolf measurement model. (a) A typical two-dimensional single molecule trajectory in the object space simulated using Eq. (B.1) where the time points are drawn from a Poisson process with mean 500 in the time interval $[0, 100]$ ms with the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1\ \mu$m^2/s. Also, we assume that the initial location of the molecule is Gaussian distributed with mean $x_0 = (4.4, 4.4)^T$ μm and covariance $P_0 = 10I_{2\times2}$ nm^2. (b) Detected locations of the photons emitted from the molecule trajectory of part (a) in the image space, which are simulated using Eq. (3.23) with the Born and Wolf profile (Eq. (3.13)) and the parameters given in Section 3.4.1. (c) Differences between the diffusion coefficient estimates and the true diffusion coefficient value for 100 data sets, each containing a trajectory of a molecule simulated using Eqs. (B.1) and (3.23) with the Born and Wolf profile, and the parameters given in parts (a) and (b). (d) Differences between the first order drift coefficient estimates and the true first order drift coefficient value for the data sets of part (c).

of the diffusion and the first order drift coefficients and the true values are plotted. As can be seen, the deviations of the estimates from the ground truth are, overall, centered around 0 nm, which suggests that there is no systematic bias associated with our proposed method (the average of the diffusion coefficient deviations and the first order drift coefficient deviations are -0.0319 μm^2/s and 0.0307/s, respectively).

Figure 3.4: Predicted locations of the molecule for the Born and Wolf measurement model. (a) and (b) Means of the distributions of the prediction of the molecule x- and y-locations, where the first order drift coefficient is unknown, and the true x- and y-locations of the molecule for the same data set as in Figs. 3.3(a) and 3.3(b). The measurements transformed from the image space to the object space are also shown. (c) and (d) Means of the distributions of the prediction of the molecule x- and y-locations and the true x- and y-locations of the molecule over the time interval $[0, 27.5]$ ms.

We further investigate the distribution $p_{pr_l}, l = 1, 2, \cdots$, of the prediction of the molecule location, given previous observations, for the molecule trajectory shown in Figs. 3.3(a) and 3.3(b). The means of the distributions of the prediction of the molecule x- and y-locations, where the first order drift coefficient is unknown, and the true x- and y-locations are shown in Fig. 3.4(a) and 3.4(b). We also show the measurements transformed from the image space to the object space.

For a better visual comparison, the means of the distributions of the prediction of the molecule locations and the true locations for x- and y-coordinates are also shown over a shorter time interval in Figs. 3.4(c) and 3.4(d). As can be seen, the predicted locations are able to track the true locations of the molecule for both x- and y-coordinates. We also show the differences between the means of the distributions of the prediction of the molecule locations and the true locations of the molecule in Fig. B.3 (see Section B.11 in Appendix B). We also applied the proposed method to trajectory data of an in-focus molecule simulated using an Airy profile, with the same standard deviation as the Born and Wolf data, and obtained similar results (see Figs. B.1, B.2 and B.4 in appendix Section B.3).

As mentioned, in some applications, it is useful to approximate the point spread function of an optical system with a Gaussian profile. We analyzed the error of the estimates for simulated data sets with Gaussian measurement noise, with the same standard deviation as the Born and Wolf data, and obtained similar results (see Figs. 3.5, 3.6 and B.5). In order to calculate the predicted locations of the molecule for Gaussian measurements, we took advantage of the relationship between the likelihood function and Kalman filter formulae (see Theorem 3). It improved the computational efficiency significantly.
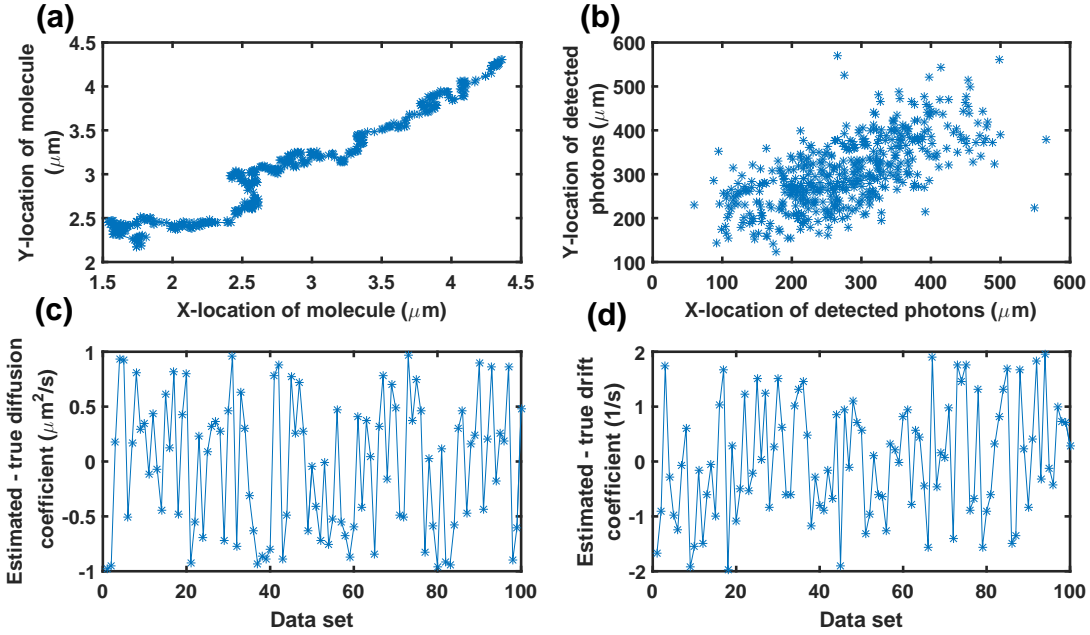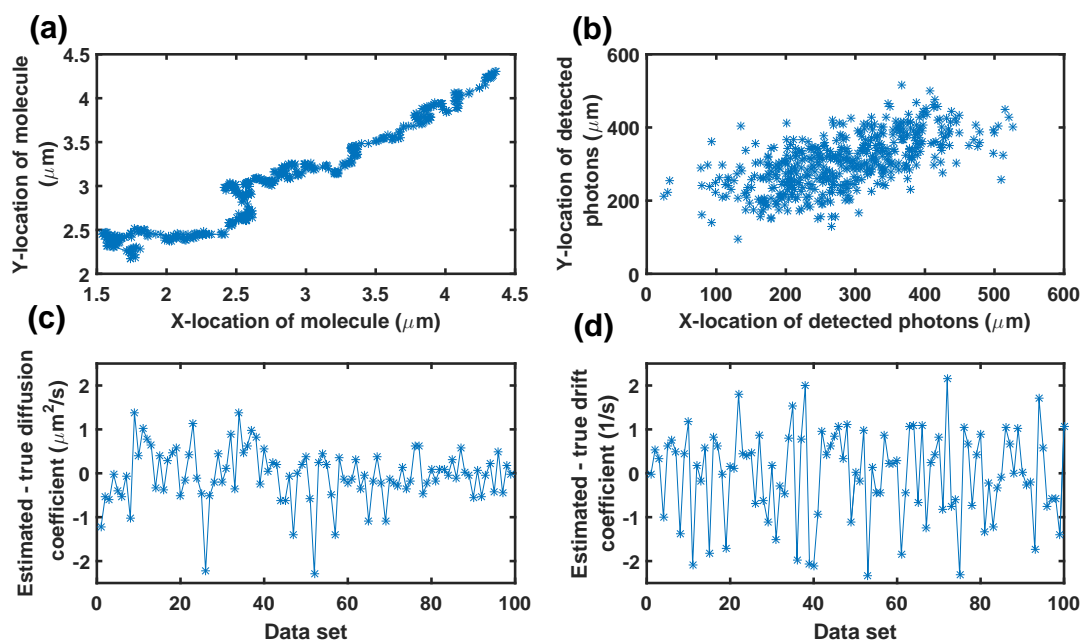
Figure 3.5: Analysis of the error of diffusion coefficient and drift coefficient estimates produced by the maximum likelihood estimation method for the Gaussian measurement noise case. (a) The two-dimensional single molecule trajectory simulated in Fig. 3.3(a). (b) Detected locations of the photons emitted from the molecule trajectory of part (a) in the image space, which are simulated using Eq. (3.24) with the Gaussian measurement noise (Eq. (3.12)) and $\sigma = 0.51$ μm. (c) Differences between the diffusion coefficient estimates and the true diffusion coefficient value for 100 data sets, each containing a trajectory of a molecule simulated using Eqs. (B.1) and (3.24) with the Gaussian profile, and the parameters given in parts (a) and (b). (d) Differences between the first order drift coefficient estimates and the true first order drift coefficient value for the data sets of part (c).
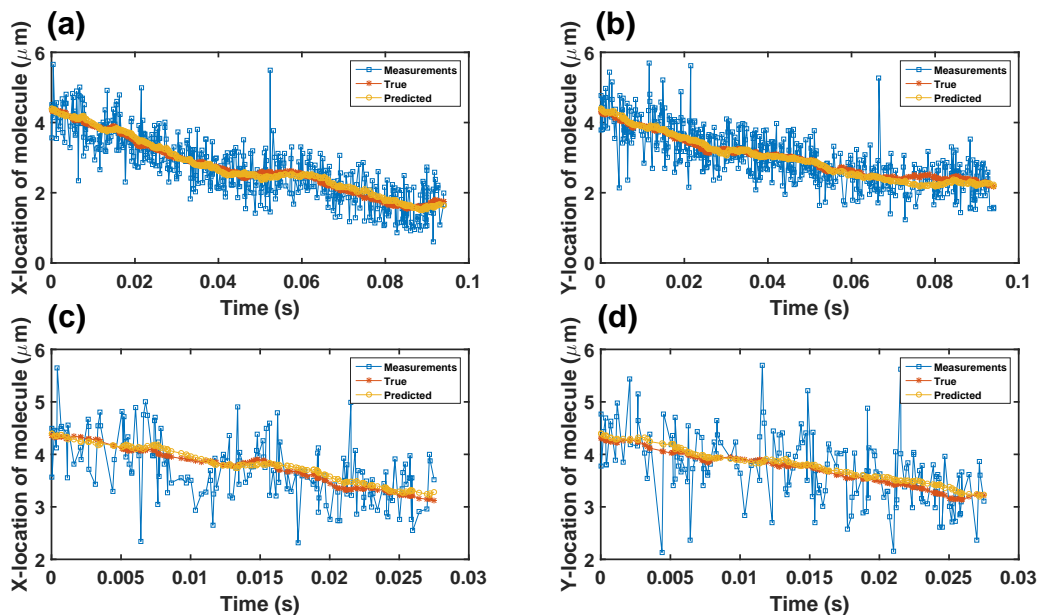
66
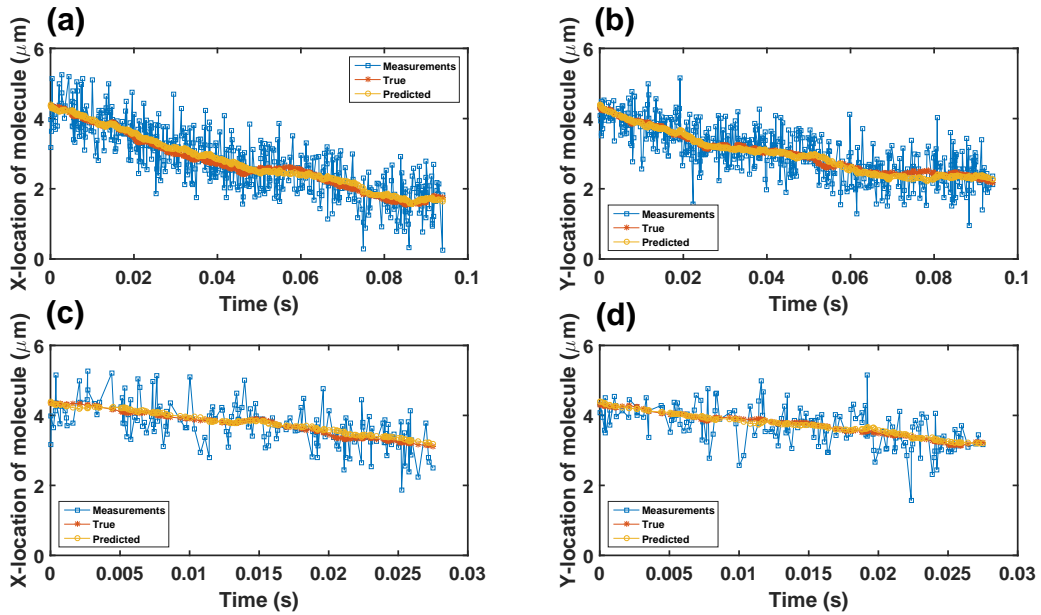
Figure 3.6: Predicted locations of the molecule for the Gaussian measurement noise case. (a) and (b) Means of the distributions of the prediction of the molecule x- and y-locations, where the first order drift coefficient is unknown, and the true x- and y-locations of the molecule for the same data set as in Figs. 3.5(a) and 3.5(b). The measurements transformed from the image space to the object space are also shown. (c) and (d) Means of the distributions of the prediction of the molecule x- and y-locations and the true x- and y-locations of the molecule over the time interval $[0, 27.5]$ ms.

4. FISHER INFORMATION MATRIX AND CRLB[*]

4.1 Introduction

In any estimation problem, the performance of the estimator can be evaluated by calculating their standard deviations from the true parameter values. According to a well-known result from estimation theory, assuming that the estimator is unbiased, its standard deviation is then at best equal to the square root of the CRLB, which is given by the inverse of the Fisher information matrix [35, 7, 6]. In other words, according to the Cramér-Rao inequality, the covariance matrix of any unbiased estimator $\hat{\theta}$ of an unknown vector parameter θ is bounded from below by the inverse of the Fisher information matrix $I(\theta)$, i.e., $\text{Cov}(\hat{\theta}) \geq I^{-1}(\theta)$. Therefore, a benchmark on the standard deviation of estimates can be obtained by the square root of the inverse of the Fisher information matrix. Note that the Fisher information matrix only depends on the statistical nature of the acquired data and is independent of the applied estimation technique. In the following, we introduce a mathematical framework to calculate general expressions for the CRLB and Fisher information matrix relating to the parameter estimation problem.

4.2 Fisher information matrix

Since this concept is very important when we have fixed time points, as we defined image detection processes and their probability density functions at fixed time points in Section 3.2, here, we first introduce a notation for the Fisher information matrix of these processes in Definition 4, and use it to calculate the Fisher information matrix of image detection processes for the fixed time interval and for the fixed number of photons in Theorem 4.

Definition 4. *For $t_0 \leq \tau_1 < \cdots < \tau_K$, let $\mathcal{G}_{\tau_1,\cdots,\tau_K}\left(\left(\mathcal{U}_K, \mathcal{T}_K\right), \mathcal{C}, \Theta \right)$ be an image detection process at fixed time points $\tau_1, \cdots, \tau_K$. We introduce the following notation for the Fisher information*

matrix of $\mathcal{G}_{\tau_1,\cdots,\tau_K}\Big((\mathcal{U}_K,\mathcal{T}_K),\mathcal{C},\Theta\Big)$ as, for a row parameter vector $\theta \in \Theta$,

$$I_{\tau_1,\cdots,\tau_K}(\theta) := E_{\mathcal{U}_K|\mathcal{T}_K=\tau_{1:K}}\left\{\left(\frac{\partial \log p^\theta_{\mathcal{U}_K|\mathcal{T}_K}\left(r_{1:K}|\tau_{1:K}\right)}{\partial \theta}\right)^T \left(\frac{\partial \log p^\theta_{\mathcal{U}_K|\mathcal{T}_K}\left(r_{1:K}|\tau_{1:K}\right)}{\partial \theta}\right)\right\}$$

$$= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p^\theta_{\mathcal{U}_K|\mathcal{T}_K}\left(r_{1:K}|\tau_{1:K}\right)\left(\frac{\partial \log p^\theta_{\mathcal{U}_K|\mathcal{T}_K}\left(r_{1:K}|\tau_{1:K}\right)}{\partial \theta}\right)^T$$

$$\times \left(\frac{\partial \log p^\theta_{\mathcal{U}_K|\mathcal{T}_K}\left(r_{1:K}|\tau_{1:K}\right)}{\partial \theta}\right) dr_1\cdots dr_K,$$

for $t_0 \leq \tau_1 < \cdots < \tau_K$, and $I_{\tau_1,\cdots,\tau_K}(\theta) = 0$, otherwise, where $r_{1:K} := (r_1,\cdots,r_K)$

, $r_1,\cdots,r_k \in \mathcal{C}$, $\tau_{1:K} := (\tau_1,\cdots,\tau_K)$, $K = 1,2,\cdots$, and $E_{\mathcal{U}_K|\mathcal{T}_K=\tau_{1:K}}$ denotes the expected value

with respect to the conditional probability density function $p^\theta_{\mathcal{U}_K|\mathcal{T}_K}$ of $\mathcal{U}_K$, given $\mathcal{T}_K = \tau_{1:K}$.

Theorem 4. *Let* $\mathcal{G}_{[t]}\Big((\mathcal{U}_{[t]},\mathcal{T}_{[t]}),\mathcal{C},\Theta\Big)$ *and* $\mathcal{G}_L\Big((\mathcal{U}_L,\mathcal{T}_L),\mathcal{C},\Theta\Big)$ *be image detection processes for*

a time interval $[t_0,t]$ *and for a fixed number* L *of photons, respectively. Let* $\mathcal{D}_{[t]} := \big(\mathcal{U}_{[t]},\mathcal{T}_{[t]}\big)$, $\mathcal{D}_k :=$

$(\mathcal{U}_k,\mathcal{T}_k)$, $k = 0,1,\cdots$. *Assume that the conditional probability density functions* $p^\theta_{U_l|T_l,\mathcal{D}_{l-1}}$, $l =$

$1,2,\cdots$, *of* U_l, *given* T_l *and* $\mathcal{D}_{l-1}$, *satisfy the following regularity conditions, for* $\theta = (\theta_1,\cdots,\theta_n) \in$

Θ,

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

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

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

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

$$I_{[t]}(\theta) = \frac{1}{P_\theta\big(N(t)=0\big)}\left(\frac{\partial P_\theta\big(N(t)=0\big)}{\partial \theta}\right)^T \left(\frac{\partial P_\theta\big(N(t)=0\big)}{\partial \theta}\right)$$

$$+ \sum_{K=1}^{\infty}\int_{t_0}^{t}\int_{t_0}^{\tau_K}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2}\left[\int_{\mathcal{C}}\cdots\int_{\mathcal{C}}\frac{1}{p^\theta_{[t]}\big(d_K,K\big)}\left(\frac{\partial p^\theta_{[t]}\big(d_K,K\big)}{\partial \theta}\right)^T \left(\frac{\partial p^\theta_{[t]}\big(d_K,K\big)}{\partial \theta}\right)\right.$$

$$\left. \times\, dr_1\cdots dr_K\right]d\tau_1 d\tau_2\cdots d\tau_{K-1}d\tau_K, \tag{4.1}$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$, *and* $p^\theta_{[t]}$ *denotes the probability density function of* $\mathcal{D}_{[t]}$ *and* $N(t)$.

69

1.2. Assume that the photon detection rate Λ is independent of θ. Then, $I_{[t]}$ can be calculated as

$$I_{[t]}(\theta) = e^{-\int_{t_0}^t \Lambda(\tau)d\tau} \sum_{K=1}^{\infty} \left\{ \int_{t_0}^t \int_{t_0}^{\tau_K} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} I_{\tau_1,\cdots,\tau_K}(\theta) \prod_{k=1}^K \Lambda(\tau_k) \right.$$
$$\left. \times d\tau_1 d\tau_2 \cdots d\tau_{K-1} d\tau_K \right\}, \tag{4.2}$$

where the Fisher information matrix $I_{\tau_1,\cdots,\tau_K}$ of the image detection process at fixed time points $\tau_1, \cdots, \tau_K$ $\mathcal{G}_{\tau_1,\cdots,\tau_K}\left((\mathcal{U}_L, \mathcal{T}_L), \mathcal{C}, \Theta \right)$ is given by

$$I_{\tau_1,\cdots,\tau_K}(\theta) = \begin{cases} \sum_{l=1}^K I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta), & t_0 \leq \tau_1 < \cdots < \tau_K \leq t, \\ 0, & otherwise, \end{cases} \tag{4.3}$$

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

$$I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta) = E_{\mathcal{U}_l|T_l=\tau_{1:l}} \left\{ \left(\frac{\partial \log p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta} \right)^T \left(\frac{\partial \log p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta} \right) \right\}$$
$$= \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_{\mathcal{U}_{l-1}|T_{l-1}}^{\theta}\left(r_{1:l-1}|\tau_{1:l-1}\right) \left[\int_{\mathcal{C}} \frac{1}{p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l, d_{l-1}\right)} \right.$$
$$\left. \times \left(\frac{\partial p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta} \right)^T \left(\frac{\partial p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l, d_{l-1}\right)}{\partial \theta} \right) dr_l \right] dr_{l-1} \cdots dr_1, \tag{4.4}$$

with $r_{1:l} := (r_1, \cdots, r_l)$, $\tau_{1:l} := (\tau_1, \cdots, \tau_l)$, and $I_{U_1|T_1}^{\tau_1}$ given by

$$I_{U_1|T_1}^{\tau_1}(\theta) = \int_{\mathcal{C}} \frac{1}{p_{U_1|T_1}^{\theta}\left(r|\tau_1\right)} \left(\frac{\partial p_{U_1|T_1}^{\theta}\left(r|\tau_1\right)}{\partial \theta} \right)^T \left(\frac{\partial p_{U_1|T_1}^{\theta}\left(r|\tau_1\right)}{\partial \theta} \right) dr. \tag{4.5}$$

2.1. The Fisher information matrix I_L of $\mathcal{G}_L$ is given by

$$I_L(\theta) = \int_{t_0}^{\infty} \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \left[\int_{\mathcal{C}} \cdots \int_{\mathcal{C}} \frac{1}{p_L^{\theta}(d_L)} \left(\frac{\partial p_L^{\theta}(d_L)}{\partial \theta} \right)^T \left(\frac{\partial p_L^{\theta}(d_L)}{\partial \theta} \right) dr_1 \cdots dr_L \right]$$

$$\times \, d\tau_1 d\tau_2 \cdots d\tau_{L-1} d\tau_L,$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$, and p_L^{θ} denotes the probability density function of $\mathcal{D}_L$.

2.2. Assume that the photon detection rate Λ is independent of θ. Then, I_L can be obtained as

$$I_L(\theta) = \int_{t_0}^{\infty} \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} I_{\tau_1,\cdots,\tau_L}(\theta) e^{- \int_{t_0}^{\tau_L} \Lambda(\tau) d\tau} \prod_{k=1}^{L} \Lambda(\tau_k)$$

$$\times \, d\tau_1 d\tau_2 \cdots d\tau_{L-1} d\tau_L. \tag{4.6}$$

Remark 1. *Note that for $K = 1$, the time integral of Eq. (4.2) is calculated over the interval $[t_0, t]$, i.e., $\int_{t_0}^{t} I_{\tau_1}(\theta) \Lambda(\tau_1) d\tau_1$.*

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

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

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

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

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

$$I_{[t]}(\theta) = E\left[\left(\frac{\partial \log \mathcal{L}(\theta | d_K)}{\partial \theta} \right)^T \left(\frac{\partial \log \mathcal{L}(\theta | d_K)}{\partial \theta} \right) \right], \tag{4.7}$$

where $d_K \in \mathcal{C}^K \times \mathbb{R}^K_{[t]}$, $K = 1, 2, \cdots$, and $\mathcal{L}$ denotes the likelihood function. By substituting the

expression of the likelihood function $\mathcal{L}_{[t]}$ of $\mathcal{G}_{[t]}$ (Eq. (3.36)) into Eq. (4.7), according to [55, 56], we have

$$
\begin{aligned}
I_{[t]}(\theta) &= P_\theta\Big(N(t) = 0\Big) \left(\frac{\partial \log P_\theta\Big(N(t) = 0\Big)}{\partial \theta}\right)^T \left(\frac{\partial \log P_\theta\Big(N(t) = 0\Big)}{\partial \theta}\right) \\
&\quad + \sum_{K=1}^{\infty} \int_{t_0}^{t} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_{[t]}^{\theta}\Big(d_K, K\Big) \left(\frac{\partial \log p_{[t]}^{\theta}\Big(d_K, K\Big)}{\partial \theta}\right)^T \left(\frac{\partial \log p_{[t]}^{\theta}\Big(d_K, K\Big)}{\partial \theta}\right) \\
&\qquad \times dr_1 \cdots dr_K d\tau_1 d\tau_2 \cdots d\tau_K \\
&= \frac{1}{P_\theta\Big(N(t) = 0\Big)} \left(\frac{\partial P_\theta\Big(N(t) = 0\Big)}{\partial \theta}\right)^T \left(\frac{\partial P_\theta\Big(N(t) = 0\Big)}{\partial \theta}\right) \\
&\quad + \sum_{K=1}^{\infty} \int_{t_0}^{t} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} \frac{1}{p_{[t]}^{\theta}\Big(d_K, K\Big)} \left(\frac{\partial p_{[t]}^{\theta}\Big(d_K, K\Big)}{\partial \theta}\right)^T \left(\frac{\partial p_{[t]}^{\theta}\Big(d_K, K\Big)}{\partial \theta}\right) \\
&\qquad \times dr_1 \cdots dr_K d\tau_1 d\tau_2 \cdots d\tau_K,
\end{aligned}
\tag{4.8}
$$

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

1.2. Assume that the photon detection rate Λ is independent of θ. By substituting Eqs. (3.4)-(3.6) into Eq. (4.8), we have

$$
\begin{aligned}
I_{[t]}(\theta) &= \sum_{K=1}^{\infty} P\Big(N(t) = K\Big) \int_{t_0}^{t} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \left[\int_{\mathcal{C}} \cdots \int_{\mathcal{C}} \frac{1}{p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_1, \cdots, r_K | \tau_1, \cdots, \tau_K\Big)}\right. \\
&\quad \times \left(\frac{\partial p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_1, \cdots, r_K | \tau_1, \cdots, \tau_K\Big)}{\partial \theta}\right)^T \left(\frac{\partial p_{\mathcal{U}_K|\mathcal{T}_K}^{\theta}\Big(r_1, \cdots, r_K | \tau_1, \cdots, \tau_K\Big)}{\partial \theta}\right) dr_1 \cdots dr_K \Bigg] \\
&\quad \times p_{\mathcal{T}_K|N(t)}\Big(\tau_1, \cdots, \tau_K | K\Big) d\tau_1 d\tau_2 \cdots d\tau_K \\
&= e^{-\int_{t_0}^{t} \Lambda(\tau) d\tau} \sum_{K=1}^{\infty} \left\{\int_{t_0}^{t} \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} I_{\tau_1, \cdots, \tau_K}(\theta) \prod_{k=1}^{K} \Lambda(\tau_k) d\tau_1 d\tau_2 \cdots d\tau_K \right\},
\end{aligned}
\tag{4.9}
$$

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

$$I_{\tau_1, \cdots, \tau_K}(\theta)$$

$$= E_{\mathcal{U}_K | \mathcal{T}_K = \tau_{1:K}} \left\{ \left(\frac{\partial \log p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right)}{\partial \theta} \right)^T \left(\frac{\partial \log p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right)}{\partial \theta} \right) \right\}$$

$$= \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right) \left(\frac{\partial \log p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right)}{\partial \theta} \right)^T \left(\frac{\partial \log p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right)}{\partial \theta} \right)$$

$$\times dr_K \cdots dr_1$$

$$= \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} \frac{1}{p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right)} \left(\frac{\partial p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right)}{\partial \theta} \right)^T \left(\frac{\partial p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right)}{\partial \theta} \right)$$

$$\times dr_K \cdots dr_1, \tag{4.10}$$

where $r_{1:K} := (r_1, \cdots, r_K)$, $\tau_{1:K} := (\tau_1, \cdots, \tau_K)$, $K = 1, 2, \cdots$. Since $p^\theta_{\mathcal{U}_K | \mathcal{T}_K}\left(r_{1:K} | \tau_{1:K}\right) = \prod_{l=1}^{K} p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)$, according to Lemma 1 (chain rule for the Fisher information matrix) of [60], we have

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

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

$$I^{\tau_1, \cdots, \tau_l}_{U_l | T_l, \mathcal{D}_{l-1}}(\theta) = E_{\mathcal{U}_l | \mathcal{T}_l = \tau_{1:l}} \left\{ \left(\frac{\partial \log p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)}{\partial \theta} \right)^T \left(\frac{\partial \log p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)}{\partial \theta} \right) \right\}$$

$$= \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p^\theta_{\mathcal{U}_l | \mathcal{T}_l}\left(r_{1:l} | \tau_{1:l}\right) \left(\frac{\partial \log p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)}{\partial \theta} \right)^T$$

$$\times \left(\frac{\partial \log p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)}{\partial \theta} \right) dr_l \cdots dr_1$$

$$= \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p^\theta_{\mathcal{U}_{l-1} | \mathcal{T}_{l-1}}\left(r_{1:l-1} | \tau_{1:l-1}\right) \left[\int_{\mathcal{C}} \frac{1}{p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)} \right.$$

$$\left. \times \left(\frac{\partial p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)}{\partial \theta} \right)^T \left(\frac{\partial p^\theta_{U_l | T_l, \mathcal{D}_{l-1}}\left(r_l | \tau_l, d_{l-1}\right)}{\partial \theta} \right) dr_l \right] dr_{l-1} \cdots dr_1. \tag{4.11}$$

73

2.1. Moreover, by substituting the expression for the likelihood function $\mathcal{L}_L$ of $\mathcal{G}_L$ (Eq. (3.37)) into Eq. (4.7), the Fisher information matrix $I_L(\theta)$ of $\mathcal{G}_L$ can be obtained as

$$
\begin{aligned}
I_L(\theta) &= \int_{t_0}^\infty \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_L^\theta\left(d_L\right) \left(\frac{\partial \log p_L^\theta\left(d_L\right)}{\partial \theta}\right)^T \left(\frac{\partial \log p_L^\theta\left(d_L\right)}{\partial \theta}\right) dr_1 \cdots dr_L \\
&\quad \times d\tau_1 d\tau_2 \cdots d\tau_L \\
&= \int_{t_0}^\infty \cdots \int_{t_0}^{\tau_3} \int_{t_0}^{\tau_2} \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} \frac{1}{p_L^\theta\left(d_L\right)} \left(\frac{\partial p_L^\theta\left(d_L\right)}{\partial \theta}\right)^T \left(\frac{\partial p_L^\theta\left(d_L\right)}{\partial \theta}\right) dr_1 \cdots dr_L d\tau_1 d\tau_2 \cdots d\tau_L, \quad (4.12)
\end{aligned}
$$

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

2.2. Assume that the photon detection rate Λ is independent of θ. By substituting Eq. (3.3) into Eq. (4.12), we have, according to Eq. (3.9) and using the similar procedure used in the previous part,

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

$\square$

We next derive expressions for the Fisher information matrices of the image detection processes driven by the stochastic trajectory X and image function q for a time interval $[t_0, t]$ and for a fixed number L of photons in the following corollary to Theorem 4.

Corollary 2. *Let $\mathcal{G}_{[t]}\left(X, \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), q, \mathcal{C}, \Theta\right)$ (or $\mathcal{G}_L\left(X, (\mathcal{U}_L, \mathcal{T}_L), q, \mathcal{C}, \Theta\right)$) be an image detection process driven by the stochastic trajectory X and image function q for a time interval $[t_0, t]$ (or for a fixed number L of photons). Let, for a row parameter vector $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$, the*

n-dimensional vector F_l^θ be given by

$$F_l^\theta\left(x, d_l\right) := \underbrace{\left[\left(df_x^\theta\left(r_l\right)\right)^T \quad \left(dp_{pr_l}^\theta\left(x|\tau_l, d_{l-1}\right)\right)^T\right]}_{\text{Block row vector}} \begin{bmatrix} p_{pr_l}^\theta\left(x|\tau_l, d_{l-1}\right) \\ f_x^\theta\left(r_l\right) \end{bmatrix}, \quad x \in \mathbb{R}^3, \quad (4.13)$$

where $d_l \in \mathcal{C}^l \times \mathbb{R}_{[t]}^l$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l$), $r_{1:l} := (r_1, \cdots, r_l)$, $\tau_{1:l} := (\tau_1, \cdots, \tau_l)$, $p_{pr_l}^\theta := p_{X(T_l)|T_l, \mathcal{D}_{l-1}}^\theta$, $p_{pr_1}^\theta\left(x|\tau_1, d_0\right) := p_{pr_1}^\theta\left(x|\tau_1\right)$, denotes the distribution of the prediction of the object location, and $dp_{pr_l}^\theta := \frac{\partial p_{pr_l}^\theta}{\partial \theta}$, $df_x^\theta := \frac{\partial f_x^\theta}{\partial \theta}$. Assume that the photon detection rate Λ is independent of θ. Then, $I_{\tau_1, \cdots, \tau_K}$ in Eq. (4.2) (or Eq. (4.6)) of Theorem 4 is given by

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

where

$$I_{U_l|T_l, \mathcal{D}_{l-1}}^{\tau_1, \cdots, \tau_l}(\theta) = \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^\theta\left(r_{1:l-1}|\tau_{1:l-1}\right)$$
$$\times \left[\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\int_{\mathcal{C}} \frac{F_l^\theta\left(x_1, d_l\right)\left[F_l^\theta\left(x_2, d_l\right)\right]^T}{p_{U_l|T_l, D_{l-1}}^\theta\left(r_l|\tau_l, d_{l-1}\right)} dr_l\right) dx_1 dx_2\right] dr_{l-1} \cdots dr_1, \quad (4.14)$$

and

$$p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^\theta\left(r_{1:l-1}|\tau_{1:l-1}\right) = \prod_{i=1}^{l-1} \int_{\mathbb{R}^3} f_{x_o}^\theta\left(r_i\right) p_{pr_l}^\theta\left(x_o|\tau_i, d_{i-1}\right) dx_o, \quad (4.15)$$

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

$$I_{U_1|T_1}^{\tau_1}(\theta) = \int_{\mathcal{C}} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{1}{p_{U_1|T_1}^\theta\left(r|\tau_1\right)} \left[\left(df_{x_1}^\theta\left(r\right)\right)^T \quad \left(dp_{pr_1}^\theta\left(x_1|\tau_1\right)\right)^T\right] \begin{bmatrix} p_{pr_1}^\theta\left(x_1|\tau_1\right) \\ f_{x_1}^\theta\left(r\right) \end{bmatrix}$$
$$\times \begin{bmatrix} p_{pr_1}^\theta\left(x_2|\tau_1\right) \\ f_{x_2}^\theta\left(r\right) \end{bmatrix}^T \begin{bmatrix} df_{x_2}^\theta\left(r\right) \\ dp_{pr_1}^\theta\left(x_2|\tau_1\right) \end{bmatrix} dx_1 dx_2 dr. \quad (4.16)$$

Remark 2. *Note that if the image function q is independent of the parameter vector θ, then,*

$$F_l^\theta\left(x, d_l\right) = f_x\left(r_l\right)\left(dp_{pr_l}^\theta\left(x|\tau_l, d_{l-1}\right)\right)^T, \quad x \in \mathbb{R}^3,$$

and the expression for $I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}$ can be simplified as

$$
\begin{aligned}
&I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta) \\
&= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^\theta\left(r_{1:l-1}|\tau_{1:l-1}\right)\left[\int_{\mathcal{C}} \frac{1}{p_{U_l|T_l,\mathcal{D}_{l-1}}^\theta\left(r_l|\tau_l, d_{l-1}\right)}\right. \\
&\quad \times \left(\frac{\partial}{\partial\theta}\int_{\mathbb{R}^3} f_{x_o}\left(r_l\right) p_{pr_l}^\theta\left(x_o|\tau_l, d_{l-1}\right)dx_o\right)^T \left(\frac{\partial}{\partial\theta}\int_{\mathbb{R}^3} f_{x_o}\left(r_l\right) p_{pr_l}^\theta\left(x_o|\tau_l, d_{l-1}\right)dx_o\right)dr_l\bigg] \\
&\quad \times dr_{l-1}\cdots dr_1 \\
&= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^\theta\left(r_{1:l-1}|\tau_{1:l-1}\right)\left\{\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\left[\int_{\mathcal{C}} \frac{1}{p_{U_l|T_l,\mathcal{D}_{l-1}}^\theta\left(r_l|\tau_l, d_{l-1}\right)}\right.\right. \\
&\quad \times f_{x_1}\left(r_l\right) f_{x_2}\left(r_l\right) \left(\frac{\partial p_{pr_l}^\theta\left(x_1|\tau_l, d_{l-1}\right)}{\partial\theta}\right)^T \left(\frac{\partial p_{pr_l}^\theta\left(x_2|\tau_l, d_{l-1}\right)}{\partial\theta}\right) dr_l\bigg] dx_1 dx_2\bigg\} dr_{l-1}\cdots dr_1. \quad (4.17)
\end{aligned}
$$

Proof. Let $\mathcal{G}_{[t]}\left(X,\left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]}\right), q, \mathcal{C}, \Theta\right)$ (or $\mathcal{G}_L\left(X,\left(\mathcal{U}_L, \ \mathcal{T}_L\right), q, \mathcal{C}, \Theta\right)$) be an image detection process driven by the stochastic trajectory X and image function q for a time interval $[t_0, t]$ (or a fixed number L of photons). Assume that the photon detection rate Λ is independent of θ. The Fisher information matrix $I_{\tau_1,\cdots,\tau_K}$ in Eq. (4.2) (or Eq. (4.6)) of Theorem 4 is given by

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

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

$$
\begin{aligned}
I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta) &= \int_{\mathcal{C}}\cdots\int_{\mathcal{C}} p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^\theta\left(r_{1:l-1}|\tau_{1:l-1}\right)\left[\int_{\mathcal{C}} \frac{1}{p_{U_l|T_l,\mathcal{D}_{l-1}}^\theta\left(r_l|\tau_l, d_{l-1}\right)}\right. \\
&\quad \times \left(\frac{\partial p_{U_l|T_l,\mathcal{D}_{l-1}}^\theta\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right)^T \left(\frac{\partial p_{U_l|T_l,\mathcal{D}_{l-1}}^\theta\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta}\right) dr_l\bigg] dr_{l-1}\cdots dr_1, \quad (4.18)
\end{aligned}
$$

and $I_{U_1|T_1}^{\tau_1}$ is given by Eq. (4.5). According to Eq. (3.15), we can express the conditional proba-

bility density functions $p^\theta_{U_l|\mathcal{D}_{l-1},T_l}$ in terms of the image profile $f_x, x \in \mathbb{R}^3$, as

$$p^\theta_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right) = \int_{\mathbb{R}^3} f^\theta_{x_o}(r_l)\, p^\theta_{pr_l}\left(x_o|\tau_l, d_{l-1}\right) dx_o, \tag{4.19}$$

where $p^\theta_{pr_l} := p^\theta_{X(T_l)|T_l,\mathcal{D}_{l-1}}$ denotes the distribution of the prediction of the object location, $p^\theta_{pr_1}\left(x_o|\tau_1, d_0\right) := p^\theta_{pr_1}\left(x_o|\tau_1\right)$, and $x_o \in \mathbb{R}^3$ denotes a running variable in the object space. By substituting Eq. (4.19) into Eq. (4.18), we have, for $dp^\theta_{pr_l} := \frac{\partial p^\theta_{pr_l}}{\partial \theta}$ and $df^\theta_x := \frac{\partial f^\theta_x}{\partial \theta}$,

$$I^{\tau_1,\cdots,\tau_l}_{U_l|T_l,\mathcal{D}_{l-1}}(\theta)$$

$$= \int_\mathcal{C} \cdots \int_\mathcal{C} p^\theta_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right)\Bigg\{ \int_\mathcal{C} \frac{1}{p^\theta_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}$$

$$\times \left(\int_{\mathbb{R}^3} \left[\left(df^\theta_{x_1}(r_l)\right)^T p^\theta_{pr_l}\left(x_1|\tau_l,d_{l-1}\right) + f^\theta_{x_1}(r_l)\left(dp^\theta_{pr_l}\left(x_1|\tau_l,d_{l-1}\right)\right)^T\right] dx_1 \right)$$

$$\times \left(\int_{\mathbb{R}^3} \left[df^\theta_{x_2}(r_l)\, p^\theta_{pr_l}\left(x_2|\tau_l,d_{l-1}\right) + f^\theta_{x_2}(r_l)\, dp^\theta_{pr_l}\left(x_2|\tau_l,d_{l-1}\right)\right] dx_2 \right) dr_l \Bigg\} dr_{l-1}\cdots dr_1$$

$$= \int_\mathcal{C} \cdots \int_\mathcal{C} p^\theta_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right)\Bigg\{ \int_\mathcal{C} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{1}{p^\theta_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}$$

$$\times \left[\left(df^\theta_{x_1}(r_l)\right)^T \quad \left(dp^\theta_{pr_l}\left(x_1|\tau_l,d_{l-1}\right)\right)^T\right] \begin{bmatrix} p^\theta_{pr_l}\left(x_1|\tau_l,d_{l-1}\right) \\ f^\theta_{x_1}(r_l) \end{bmatrix} \begin{bmatrix} p^\theta_{pr_l}\left(x_2|\tau_l,d_{l-1}\right) \\ f^\theta_{x_2}(r_l) \end{bmatrix}^T \begin{bmatrix} df^\theta_{x_2}(r_l) \\ dp^\theta_{pr_l}\left(x_2|\tau_l,d_{l-1}\right) \end{bmatrix}$$

$$\times dx_1 dx_2 dr_l \Bigg\} dr_{l-1}\cdots dr_1$$

$$= \int_\mathcal{C} \cdots \int_\mathcal{C} p^\theta_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right) \left[\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \left(\int_\mathcal{C} \frac{F^\theta_l\left(x_1,d_l\right)\left[F^\theta_l\left(x_2,d_l\right)\right]^T}{p^\theta_{U_l|T_l,D_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)} dr_l \right) dx_1 dx_2 \right]$$

$$\times dr_{l-1}\cdots dr_1, \tag{4.20}$$

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

$$F^\theta_l\left(x,d_l\right) := \left[\left(df^\theta_x(r_l)\right)^T \quad \left(dp^\theta_{pr_l}\left(x|\tau_l,d_{l-1}\right)\right)^T\right] \begin{bmatrix} p^\theta_{pr_l}\left(x|\tau_l,d_{l-1}\right) \\ f^\theta_x(r_l) \end{bmatrix}, \quad x \in \mathbb{R}^3,$$

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

$$p^\theta_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}\left(r_{1:l-1}|\tau_{1:l-1}\right) = \prod_{i=1}^{l-1} \int_{\mathbb{R}^3} f^\theta_{x_o}(r_i)\, p^\theta_{pr_l}\left(x_o|\tau_i,d_{i-1}\right) dx_o,$$

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

$$
I_{U_1|T_1}^{\tau_1}(\theta) = \int_{\mathcal{C}} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{1}{p_{U_1|T_1}^{\theta}(r|\tau_1)} \left[(df_{x_1}^{\theta}(r))^T \quad \left(dp_{pr_1}^{\theta}(x_1|\tau_1) \right)^T \right] \begin{bmatrix} p_{pr_1}^{\theta}(x_1|\tau_1) \\ f_{x_1}^{\theta}(r) \end{bmatrix}
$$

$$
\times \begin{bmatrix} p_{pr_1}^{\theta}(x_2|\tau_1) \\ f_{x_2}^{\theta}(r) \end{bmatrix}^T \begin{bmatrix} df_{x_2}^{\theta}(r) \\ dp_{pr_1}^{\theta}(x_2|\tau_1) \end{bmatrix} dx_1 dx_2 dr.
$$

$\square$

As mentioned in Section 3.2, for special cases of an object with a deterministic trajectory and a static object, the probability density function of the image detection process $\mathcal{G}_{\tau_1,\cdots,\tau_K}$ at fixed time points $t_0 \leq \tau_1 < \cdots < \tau_K$ is simplified as given by Eqs. (3.16) and (3.17), respectively. We next in Corollary 3 to Theorem 4 calculate the Fisher information matrix for these special cases, and show that the obtained results are consistent with the results presented in [6, 61, 54, 62].

Corollary 3. *For* $t_0 \leq \tau_1 < \cdots < \tau_K$, *let* $\mathcal{G}_{\tau_1,\cdots,\tau_K}\left((\mathcal{U}_K, \mathcal{T}_K), \mathcal{C}, \Theta \right)$ *be an image detection process at fixed time points* $\tau_1, \cdots, \tau_K$. *Assume that* $p_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1} \right) = p_{U_l|T_l}\left(r_l|\tau_l \right), d_l \in \mathcal{C}^l \times \mathbb{R}_{[\infty]}^l, l = 1, 2, \cdots$.

1. Then, the Fisher information matrix $I_{\tau_1,\cdots,\tau_K}$ *of* $\mathcal{G}_{\tau_1,\cdots,\tau_K}\left((\mathcal{U}_K, \mathcal{T}_K), \mathcal{C}, \Theta \right)$ *is given by*

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

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

$$
I_{U_l|T_l}^{\tau_l}(\theta) = \int_{\mathbb{R}^2} \frac{1}{p_{U_l|T_l}^{\theta}(r|\tau_l)} \left(\frac{\partial p_{U_l|T_l}^{\theta}(r|\tau_l)}{\partial \theta} \right)^T \left(\frac{\partial p_{U_l|T_l}^{\theta}(r|\tau_l)}{\partial \theta} \right) dr.
$$

2.1. For an object with deterministic trajectory $X_\tau(\theta) := (x_\tau(\theta), y_\tau(\theta)) \in \mathbb{R}^2, \tau \geq t_0$, *assume that there exists an image function* $q \colon \mathbb{R}^2 \mapsto \mathbb{R}$, *which describes the image of an object on the*

detector plane at unit lateral magnification and it is assumed to be independent of the parameter

vector $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$, such that

$$p^\theta_{U_l|T_l}\left(r|\tau\right) := \frac{1}{M^2} q\left(\frac{x}{M} - x_\tau(\theta), \frac{y}{M} - y_\tau(\theta)\right),$$

where $r = (x, y) \in \mathbb{R}^2$, $t_0 \leq \tau \leq t$, and $M > 1$ is a magnification factor. Let $D_1 q$ and $D_2 q$ be the

partial derivatives of q with respect to the x- and y-coordinates, respectively. Also, let $D_j x_\tau$ and

$D_j y_\tau, j = 1, \cdots, n$, denote the partial derivatives of x_τ and y_τ with respect to the jth parameter

coordinate, respectively. Then, for $t_0 \leq \tau_1 < \cdots < \tau_K$,

$$I_{\tau_1,\cdots,\tau_K}(\theta) = \sum_{l=1}^{K} I_{\tau_l}(\theta),$$

where

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

and

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

2.2. For a static object with position $X_0(\theta) = (x_0(\theta), y_0(\theta)) \in \mathbb{R}^2$, we have, for $t_0 \leq \tau_1 < \cdots < \tau_K$,

$$I_{\tau_1,\cdots,\tau_K}(\theta) = I(\theta) = K\tilde{I}(\theta),$$

where

$$\tilde{I}(\theta) = V_\theta^T \left(\int_{\mathbb{R}^2} \frac{1}{q(u,v)} \begin{bmatrix} (D_1 q)(u,v) \\ (D_2 q)(u,v) \end{bmatrix} \begin{bmatrix} (D_1 q)(u,v) \\ (D_2 q)(u,v) \end{bmatrix}^T dudv \right) V_\theta,$$

and for $\theta = (\theta_1, \cdots, \theta_n) \in \theta$,

$$V_\theta := \begin{bmatrix} (D_1 x_0)(\theta) & \cdots & (D_n x_0)(\theta) \\ (D_1 y_0)(\theta) & \cdots & (D_n y_0)(\theta) \end{bmatrix} \in \mathbb{R}^{2 \times n}.$$

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

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

1. According to Eq. (3.16), we have, for $r_l \in \mathbb{R}^2, l = 1, 2, \cdots$,

$$p_{\mathcal{U}_K|\mathcal{T}_K}\left(r_1, \cdots, r_K | \tau_1, \cdots, \tau_K \right) = \prod_{l=1}^K p_{U_l|T_l}^\theta\left(r_l | \tau_l \right). \tag{4.21}$$

By substituting Eq. (4.21) into Eq. (4.4), we have

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

80

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

$$
\begin{aligned}
I_{U_l|T_l}^{\tau_l}(\theta) &= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} p_{U_1|T_1}^{\theta}\left(r_1|\tau_1\right) dr_1 \right) \cdots \left(\int_{\mathbb{R}^2} p_{U_{l-1}|T_{l-1}}^{\theta}\left(r_{l-1}|\tau_{l-1}\right) dr_{l-1} \right) \\
&\qquad \times \frac{1}{p_{U_l|T_l}^{\theta}\left(r_l|\tau_l\right)} \left(\frac{\partial p_{U_l|T_l}^{\theta}\left(r_l|\tau_l\right)}{\partial \theta} \right)^T \left(\frac{\partial p_{U_l|T_l}^{\theta}\left(r_l|\tau_l\right)}{\partial \theta} \right) dr_l \\
&= \int_{\mathbb{R}^2} \frac{1}{p_{U_l|T_l}^{\theta}\left(r|\tau_l\right)} \left(\frac{\partial p_{U_l|T_l}^{\theta}\left(r|\tau_l\right)}{\partial \theta} \right)^T \left(\frac{\partial p_{U_l|T_l}^{\theta}\left(r|\tau_l\right)}{\partial \theta} \right) dr.
\end{aligned}
\tag{4.22}
$$

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

$$
p_{U_l|T_l}^{\theta}\left(r|\tau\right) = f_{X_\tau(\theta)}(r) = \frac{1}{M^2} q\left(\frac{x}{M} - x_\tau(\theta), \frac{y}{M} - y_\tau(\theta) \right).
\tag{4.23}
$$

Then, by substituting Eq. (4.23) into Eq. (4.22), $I_{\tau_l} := I_{U_l|T_l}^{\tau_l}$ is obtained as, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$,

$$
\begin{aligned}
I_{\tau_l}(\theta) &= \frac{1}{M^2} \int_{\mathbb{R}^2} \frac{1}{q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)} \left(\frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)}{\partial \theta} \right)^T \\
&\qquad\qquad \times \left(\frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)}{\partial \theta} \right) dxdy \\
&= \frac{1}{M^2} \int_{\mathbb{R}^2} \frac{1}{q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)} \begin{bmatrix} \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)}{\partial \theta_1} \\ \vdots \\ \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)}{\partial \theta_n} \end{bmatrix} \begin{bmatrix} \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)}{\partial \theta_1} \\ \vdots \\ \frac{\partial q\left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right)}{\partial \theta_n} \end{bmatrix}^T dxdy.
\end{aligned}
\tag{4.24}
$$

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

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

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

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

and therefore, using the formal definition of partial derivatives, we can rewrite Eq. (4.24) as

$$I_{\tau_l}(\theta) = \frac{1}{M^2}\int_{\mathbb{R}^2} \frac{1}{(q \circ h_{x,y} \circ d_{\tau_l})(\theta)} \begin{bmatrix} (D_1(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \\ \vdots \\ (D_n(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \end{bmatrix} \begin{bmatrix} (D_1(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \\ \vdots \\ (D_n(q \circ h_{x,y} \circ d_{\tau_l}))(\theta_1, \cdots, \theta_n) \end{bmatrix}^T dxdy. \quad (4.25)$$

Assume that d_{τ_l} is continuously differentiable on all of Θ, and $h_{x,y}$ is differentiable at $d_{\tau_l}(\theta)$. Also, suppose that q is differentiable at $h_{x,y}\left(d_{\tau_l}(\theta)\right)$. Then, according to Theorem 8 (see Section B.8), for $i =, 1 \cdots, n$,

$$(D_i(q \circ h_{x,y} \circ d_{\tau_l}))(\theta) = (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_1 h_x)(d_{\tau_l}(\theta))(D_i x_{\tau_l})(\theta)$$
$$+ (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 h_x)(d_{\tau_l}(\theta))(D_i y_{\tau_l})(\theta)$$
$$+ (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_1 h_y)(d_{\tau_l}(\theta))(D_i x_{\tau_l})(\theta)$$
$$+ (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 h_y)(d_{\tau_l}(\theta))(D_i y_{\tau_l})(\theta)$$
$$= -(D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_i x_{\tau_l})(\theta) - (D_2 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_i y_{\tau_l})(\theta)$$
$$= -\begin{bmatrix} (D_i x_{\tau_l})(\theta) & (D_i y_{\tau_l})(\theta) \end{bmatrix} \begin{bmatrix} (D_1 q)(h_{x,y}(d_{\tau_l}(\theta))) \\ (D_2 q)(h_{x,y}(d_{\tau_l}(\theta))) \end{bmatrix}. \quad (4.26)$$

By substituting Eq. (4.26) into Eq. (4.25), we have, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$,

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

$$
= \frac{1}{M^2} V_\theta^T(\tau_l) \left[\begin{array}{l} \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} \left[(D_1 q)(h_{x,y}(d_{\tau_l}(\theta))) \right]^2 dxdy \\[2ex] \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 q)(h_{x,y}(d_{\tau_l}(\theta))) dxdy \end{array} \right.
$$

$$
\left. \begin{array}{l} \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} (D_1 q)(h_{x,y}(d_{\tau_l}(\theta)))(D_2 q)(h_{x,y}(d_{\tau_l}(\theta))) dxdy \\[2ex] \int_{\mathbb{R}^2} \frac{1}{q(h_{x,y}(d_{\tau_l}(\theta)))} \left[(D_2 q)(h_{x,y}(d_{\tau_l}(\theta))) \right]^2 dxdy \end{array} \right] V_\theta(\tau_l), \tag{4.27}
$$

where

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

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

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

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

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

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

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

and the modulus of its determinant is given by

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

Then, according to Theorem 9 (see Section B.9),

$$\begin{aligned}
\int_{\mathbb{R}^2} w_1(u,v) du dv &= \frac{1}{M^2} \int_{\mathbb{R}^2} w_1 \left(g_{\theta,\tau_l}(x,y) \right) dx dy \\
&= \frac{1}{M^2} \int_{\mathbb{R}^2} w_1 \left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right) dx dy.
\end{aligned} \tag{4.28}$$

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

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

and

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

be integrable functions. Similarly, according to Theorem 9 (see Section B.9),

$$\int_{\mathbb{R}^2} w_i(u,v) du dv = \frac{1}{M^2} \int_{\mathbb{R}^2} w_i \left(\frac{x}{M} - x_{\tau_l}(\theta), \frac{y}{M} - y_{\tau_l}(\theta) \right) dx dy, \quad i = 2, 3. \tag{4.29}$$

Then, by substituting Eqs. (4.28) and (4.29) into Eq. (4.27),

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

2.2. The results follow by using the similar procedure used in the previous part. $\qquad \square$

The material presented in Theorem 4 and Corollary 2 provides a mathematical framework to calculate the Fisher information matrix of image detection processes for a fixed time interval and for a fixed number of photons for a moving object with a general stochastic motion model. As mentioned before, in many biological applications, the motion of a small object in subcellular environments can be modeled by a linear stochastic differential equation. The solution of this linear stochastic differential equation can be modeled by a first order system driven by Gaussian noise. In Corollary 4 to Theorem 4, we obtain recursive expressions for the Fisher information matrices for both image detection processes for a fixed time interval and a fixed number of photons, in case that the dynamical system is described by a first order system with Gaussian process and measurement noise.

Corollary 4. *Let* $\mathcal{G}^g_{[t]}\Big(\big(\tilde{X}, g, \tilde{W}_g, Z_g \big), (\mathcal{U}_{[t]}, \mathcal{T}_{[t]}), \tilde{\Phi}, M', \mathcal{C}, \Theta \Big)$ *(or* $\mathcal{G}^g_L\Big(\big(\tilde{X}, g, \tilde{W}_g, Z_g \big),$ $(\mathcal{U}_L, \mathcal{T}_L), \tilde{\Phi}, M', \mathcal{C}, \Theta \Big)$*) be an image detection process with expanded state space* $\tilde{X}$ *and Gaussian process and measurement noise models for a time interval* $[t_0, t]$ *(or for a fixed number* L *of photons). Assume that*

$$X(\tau) = g(\tilde{X}(\tau)) := G\tilde{X}(\tau), \quad \tau \geq t_0,$$

where $G \in \mathbb{R}^{3 \times k}$*, and let* $C := M'G$*. Assume that the photon detection rate* Λ*,* C *and* Z_g *are independent of* θ*. Let*

$$S^{(ji)}_{\theta,l} - A^{(j)}_{\theta,l} S^{(ji)}_{\theta,l-1} \left(A^{(i)}_{\theta,l} \right)^T = B^{(j)}_{\theta,l} R_{\theta,l-1} \left(B^{(j)}_{\theta,l} \right)^T, \quad l = 2, 3, \cdots,$$

$$S^{(ji)}_{\theta,1} = \begin{bmatrix} \tilde{\phi}_\theta(\tau_0, \tau_1)\tilde{x}_{\theta,0} \\ \frac{\partial(\tilde{\phi}_\theta(\tau_0, \tau_1)\tilde{x}_{\theta,0})}{\partial\theta_j} \end{bmatrix} \left[\left(\tilde{\phi}_\theta(\tau_0, \tau_1)\tilde{x}_{\theta,0} \right)^T \quad \left(\frac{\partial(\tilde{\phi}_\theta(\tau_0, \tau_1)\tilde{x}_{\theta,0})}{\partial\theta_i} \right)^T \right], \qquad (4.30)$$

where

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

and $R_{\theta,l} := CP^{l-1}_{\theta,l}C^T + \Sigma_g$*,* $K_{\theta,l} := P^{l-1}_{\theta,l}C^T \left(CP^{l-1}_{\theta,l}C^T + \Sigma_g \right)^{-1}, l = 1, 2, \cdots,$ *where* $P^{l-1}_{\theta,l}$ *is*

85

obtained through Eqs. (3.28) and (3.29).

Then, the Fisher information matrix $I_{\tau_1,\cdots,\tau_K}$ in Eq. (4.2) (or Eq. (4.6)) of Theorem 4 can be calculated as

$$
I_{\tau_1,\cdots,\tau_K}(\theta) = \begin{cases} \sum_{l=1}^{K} I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta), & t_0 \leq \tau_1 < \cdots < \tau_K \leq t, \\ \\ 0, & \textit{otherwise,} \end{cases}
\tag{4.31}
$$

where, for $\theta = (\theta_1,\cdots,\theta_n) \in \Theta$ and $l = 1,\cdots,K$, the i,j^{th}, $i,j = 1,\cdots,n$, entry $\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}\right]_{i,j}$ of $I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}$ can be calculated as

$$
\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta)\right]_{i,j} = \frac{1}{2}trace\left[R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_j}\right] + trace\left\{R_{\theta,l}^{-1}\tilde{C}S_{\theta,l}^{(ji)}\tilde{C}^T\right\},
\tag{4.32}
$$

with $\tilde{C} := \begin{bmatrix} 0_{2\times k} & C \end{bmatrix}$.

Proof. According to Theorem 4, the Fisher information matrix $I_{\tau_1,\cdots,\tau_K}$ in Eq. (4.2) (or Eq. (4.6)) can be calculated as

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

where

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

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

$$
\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta)\right]_{i,j} = \int_{\mathcal{C}} \cdots \int_{\mathcal{C}} p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^{\theta}\left(r_{1:l-1}|\tau_{1:l-1}\right)
$$
$$
\times \left[-\int_{\mathcal{C}} p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l,d_{l-1}\right)\frac{\partial^2 \log p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l,d_{l-1}\right)}{\partial \theta_i \theta_j} dr_l\right] dr_{l-1}\cdots dr_1.
\tag{4.33}
$$

According to Eq. (3.35),

$$p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l,d_{l-1}\right) = \frac{1}{2\pi\left[\det\left(R_{\theta,l}\right)\right]^{1/2}}\exp\left(-\frac{1}{2}e_{\theta,l}^T R_{\theta,l}^{-1}e_{\theta,l}\right),\tag{4.34}$$

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

$$\hat{x}_{\theta,l+1}^l = \tilde{\phi}_\theta(\tau_l,\tau_{l+1})\hat{x}_{\theta,l}^l,$$
$$P_{\theta,l+1}^l = \tilde{\phi}_\theta(\tau_l,\tau_{l+1})P_{\theta,l}^l\tilde{\phi}_\theta^T(\tau_l,\tau_{l+1}) + \tilde{Q}_\theta(\tau_l,\tau_{l+1}),\tag{4.35}$$

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

$$\hat{x}_{\theta,l}^l = \hat{x}_{\theta,l}^{l-1} + K_{\theta,l}(r_l - C\hat{x}_{\theta,l}^{l-1}),$$
$$P_{\theta,l}^l = P_{\theta,l}^{l-1} - K_{\theta,l}CP_{\theta,l}^{l-1},$$
$$K_{\theta,l} = P_{\theta,l}^{l-1}C^T\left(CP_{\theta,l}^{l-1}C^T + \Sigma_g\right)^{-1},\tag{4.36}$$

where $\hat{x}_{\theta,0}^0 := \tilde{x}_{\theta,0}$, $P_{\theta,0}^0 := \tilde{P}_{\theta,0}$. In order to calculate $\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}\right]_{i,j}$, $i,j = 1, \cdots, n$, in Eq. (4.33), we first calculate, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$ and $i = 1, \cdots, n$, the derivative of $\log p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}$ with respect to θ_i as below

$$\frac{\partial \log p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l,d_{l-1}\right)}{\partial \theta_i} = -\frac{1}{2}\text{trace}\left(R_{\theta,l}\frac{\partial R_{\theta,l}}{\partial \theta_i}\right)$$
$$-\frac{1}{2}\left(\frac{\partial e_{\theta,l}^T}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l} - e_{\theta,l}^T R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l} + e_{\theta,l}^T R_{\theta,l}^{-1}\frac{\partial e_{\theta,l}}{\partial \theta_i}\right).\tag{4.37}$$

Since the covariance matrix $R_{\theta,l}$ is symmetric, then, $\frac{\partial e_{\theta,l}^T}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l} = e_{\theta,l}^T R_{\theta,l}^{-1}\frac{\partial e_{\theta,l}}{\partial \theta_i}$, and therefore, according to Eq. (4.37),(note that trace $\left(e_{\theta,l}^T R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l}\right) = e_{\theta,l}^T R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l}$),

$$\frac{\partial \log p_{U_l|T_l,\mathcal{D}_{l-1}}^{\theta}\left(r_l|\tau_l,d_{l-1}\right)}{\partial \theta_i} = -\frac{1}{2}\text{trace}\left(R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}\right) + \frac{1}{2}\text{trace}\left(e_{\theta,l}^T R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l}\right)$$
$$-\frac{1}{2}\left(\frac{\partial e_{\theta,l}^T}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l} + e_{\theta,l}^T R_{\theta,l}^{-1}\frac{\partial e_{\theta,l}}{\partial \theta_i}\right)$$
$$= -\frac{1}{2}\text{trace}\left(R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i} - R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l}e_{\theta,l}^T\right) - \frac{\partial e_{\theta,l}^T}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l}$$
$$= -\frac{1}{2}\text{trace}\left[\left(R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}\right)(I - R_{\theta,l}^{-1}e_{\theta,l}e_{\theta,l}^T)\right] - \frac{\partial e_{\theta,l}^T}{\partial \theta_i}R_{\theta,l}^{-1}e_{\theta,l},\tag{4.38}$$

where I denotes the identity matrix with the corresponding size. Differentiating Eq. (4.38) with respect to θ_j, gives [63]

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

Therefore, the inner integral in Eq. (4.33) can be calculated as

$$
\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial^2 \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)}{\partial\theta_i\theta_j}dr_l =
$$
$$
-\frac{1}{2}\text{trace}\left[\left(\frac{\partial R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i}}{\partial\theta_j}\right)\underbrace{\left(I - R^{-1}_{\theta,l}\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)e_{\theta,l}e^{T}_{\theta,l}dr_l\right)}_{Term_1}\right]
$$
$$
-\frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i}R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_j}R^{-1}_{\theta,l}\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)e_{\theta,l}e^{T}_{\theta,l}dr_l\right]
$$
$$
+\frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l}\frac{\partial R_{\theta,l}}{\partial\theta_i}R^{-1}_{\theta,l}\underbrace{\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\left(\frac{\partial e_{\theta,l}}{\partial\theta_j}e^{T}_{\theta,l} + e_{\theta,l}\frac{\partial e^{T}_{\theta,l}}{\partial\theta_j}\right)dr_l}_{Term_2}\right]
$$
$$
-\underbrace{\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial^2 e^{T}_{\theta,l}}{\partial\theta_i\theta_j}R^{-1}_{\theta,l}e_{\theta,l}dr_l}_{Term_3} + \underbrace{\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial e^{T}_{\theta,l}}{\partial\theta_i}\frac{\partial R^{-1}_{\theta,l}}{\partial\theta_j}e_{\theta,l}dr_l}_{Term_4}
$$
$$
-\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l, d_{l-1}\right)\frac{\partial e^{T}_{\theta,l}}{\partial\theta_i}R^{-1}_{\theta,l}\frac{\partial e_{\theta,l}}{\partial\theta_j}dr_l. \tag{4.39}
$$

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

$$\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right) e_{\theta,l} \frac{\partial e^T_{\theta,l}}{\partial \theta_j} dr_l$$

$$= \int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)\left(r_l - C\hat{x}^{l-1}_{\theta,l}\right) \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial \theta_j} C^T dr_l$$

$$= \left[\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)\left(r_l - C\hat{x}^{l-1}_{\theta,l}\right) dr_l\right] \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial \theta_j} C^T$$

$$= \left[\int_{\mathcal{C}} r_l p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right) dr_l - C\hat{x}^{l-1}_{\theta,l} \int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right) dr_l\right] \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial \theta_j} C^T$$

$$= \left[C\hat{x}^{l-1}_{\theta,l} - C\hat{x}^{l-1}_{\theta,l}\right] \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial \theta_j} C^T = 0.$$

Similarly, $\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right) \frac{\partial e_{\theta,l}}{\partial \theta_j} e^T_{\theta,l} dr_l = 0$, and therefore, $Term_2$, $Term_3$, and $Term_4$ in Eq. (4.39) are equal to zero. Then, noting that

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

we have $Term_1 = 0$, and Eq. (4.39) becomes

$$\int_{\mathcal{C}} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right) \frac{\partial^2 \log p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right)}{\partial \theta_i \theta_j} dr_l$$

$$= -\frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_j}\right]$$

$$- \left[\int_{\mathbb{R}^2} p^{\theta}_{U_l|T_l,\mathcal{D}_{l-1}}\left(r_l|\tau_l,d_{l-1}\right) dr_l\right] \left(\frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial \theta_i} C^T R^{-1}_{\theta,l} C \frac{\partial \hat{x}^{l-1}_{\theta,l}}{\partial \theta_j}\right)$$

$$= -\frac{1}{2}\text{trace}\left[R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_i} R^{-1}_{\theta,l} \frac{\partial R_{\theta,l}}{\partial \theta_j}\right] - \frac{\partial\left(\hat{x}^{l-1}_{\theta,l}\right)^T}{\partial \theta_i} C^T R^{-1}_{\theta,l} C \frac{\partial \hat{x}^{l-1}_{\theta,l}}{\partial \theta_j}. \tag{4.40}$$

By substituting the above equation in Eq. (4.33), we have

$$
\begin{aligned}
\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta)\right]_{i,j} &= \frac{1}{2}\text{trace}\left[R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_j}\right] \\
&\quad + \int_{\mathcal{C}}\cdots\int_{\mathcal{C}}p_{\mathcal{U}_{l-1}|\mathcal{T}_{l-1}}^{\theta}\left(r_{1:l-1}|\tau_{1:l-1}\right)\frac{\partial\left(\hat{x}_{\theta,l}^{l-1}\right)^T}{\partial \theta_i}C^TR_{\theta,l}^{-1}C\frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_j}dr_{l-1}\cdots dr_1 \\
&= \frac{1}{2}\text{trace}\left[R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_j}\right] + E\left[\frac{\partial\left(\hat{x}_{\theta,l}^{l-1}\right)^T}{\partial \theta_i}C^TR_{\theta,l}^{-1}C\frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_j}\right] \\
&= \frac{1}{2}\text{trace}\left[R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_j}\right] + \text{trace}\left\{R_{\theta,l}^{-1}E\left[C\frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_j}\frac{\partial\left(\hat{x}_{\theta,l}^{l-1}\right)^T}{\partial \theta_i}C^T\right]\right\} \\
&= \frac{1}{2}\text{trace}\left[R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_i}R_{\theta,l}^{-1}\frac{\partial R_{\theta,l}}{\partial \theta_j}\right] + \text{trace}\left\{R_{\theta,l}^{-1}CE\left[\frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_j}\frac{\partial\left(\hat{x}_{\theta,l}^{l-1}\right)^T}{\partial \theta_i}\right]C^T\right\}.
\end{aligned} \tag{4.41}
$$

According to Eqs. (4.35) and (4.36),

$$
\hat{x}_{\theta,l+1}^l = \tilde{\phi}_\theta(\tau_l, \tau_{l+1})\left(\hat{x}_{\theta,l}^{l-1} + K_{\theta,l}(r_l - C\hat{x}_{\theta,l}^{l-1})\right), \quad l = 1, 2, \cdots. \tag{4.42}
$$

Then, according to Lemma 5 (see Section B.7), by differentiating Eq. (4.42) with respect to θ_i, $i = 1, \cdots, n$, after some straightforward calculations, for $X_{\theta,l}^{(i)} := \begin{bmatrix} \hat{x}_{\theta,l}^{l-1} \\ \frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_i} \end{bmatrix}$, we have the following recursive formulation:

$$
X_{\theta,l+1}^{(i)} = A_{\theta,l+1}^{(i)}X_{\theta,l}^{(i)} + B_{\theta,l+1}^{(i)}e_{\theta,l}, \quad \theta = (\theta_1, \cdots, \theta_n) \in \Theta, \quad i = 1, \cdots, n, \tag{4.43}
$$

and

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

According to Lemma 6 (see Section B.10) and using Eq. (4.43), we have, for $l = 1, 2, \cdots,$

$$E\left\{ X_{\theta,l+1}^{(j)} \left(X_{\theta,l+1}^{(i)} \right)^T \right\}$$

$$= E\left\{ A_{\theta,l+1}^{(j)} X_{\theta,l}^{(j)} \left(X_{\theta,l}^{(i)} \right)^T \left(A_{\theta,l+1}^{(i)} \right)^T + A_{\theta,l+1}^{(j)} X_{\theta,l}^{(j)} e_{\theta,l}^T \left(B_{\theta,l+1}^{(i)} \right)^T \right.$$

$$\left. + B_{\theta,l+1}^{(j)} e_{\theta,l} \left(X_{\theta,l}^{(i)} \right)^T \left(A_{\theta,l+1}^{(i)} \right)^T + B_{\theta,l+1}^{(j)} e_{\theta,l} e_{\theta,l}^T \left(B_{\theta,l+1}^{(j)} \right)^T \right\}$$

$$= A_{\theta,l+1}^{(j)} E\left\{ X_{\theta,l}^{(j)} \left(X_{\theta,l}^{(i)} \right)^T \right\} \left(A_{\theta,l+1}^{(i)} \right)^T + A_{\theta,l+1}^{(j)} \underbrace{E\left\{ X_{\theta,l}^{(j)} e_{\theta,l}^T \right\}}_{0} \left(B_{\theta,l+1}^{(i)} \right)^T$$

$$+ B_{\theta,l+1}^{(j)} \underbrace{E\left\{ e_{\theta,l} \left(X_{\theta,l}^{(i)} \right)^T \right\}}_{0} \left(A_{\theta,l+1}^{(i)} \right)^T + B_{\theta,l+1}^{(j)} E\left\{ e_{\theta,l} e_{\theta,l}^T \right\} \left(B_{\theta,l+1}^{(j)} \right)^T$$

$$= A_{\theta,l+1}^{(j)} E\left\{ X_{\theta,l}^{(j)} \left(X_{\theta,l}^{(i)} \right)^T \right\} \left(A_{\theta,l+1}^{(i)} \right)^T$$

$$+ B_{\theta,l+1}^{(j)} \left(CE\left[(X_\theta(\tau_l) - \hat{x}_{\theta,l}^{l-1})(X_\theta(\tau_l) - \hat{x}_{\theta,l}^{l-1})^T \right] C^T + \Sigma_g \right) \left(B_{\theta,l+1}^{(j)} \right)^T$$

$$= A_{\theta,l+1}^{(j)} E\left\{ X_{\theta,l}^{(j)} \left(X_{\theta,l}^{(i)} \right)^T \right\} \left(A_{\theta,l+1}^{(i)} \right)^T + B_{\theta,l+1}^{(j)} R_{\theta,l} \left(B_{\theta,l+1}^{(j)} \right)^T. \qquad (4.44)$$

Finally, by rewriting the Fisher information expression (Eq. (4.41)) as (let $\tilde{C} := \begin{bmatrix} 0_{2 \times k} & C \end{bmatrix}$, where $0_{2 \times k}$ denotes the $2 \times k$ zero matrix)

$$\left[I_{U_l|T_l,\mathcal{D}_{l-1}}^{\tau_1,\cdots,\tau_l}(\theta) \right]_{i,j} = \frac{1}{2}\text{trace}\left[R_{\theta,l}^{-1} \frac{\partial R_{\theta,l}}{\partial \theta_i} R_{\theta,l}^{-1} \frac{\partial R_{\theta,l}}{\partial \theta_j} \right] + \text{trace}\left\{ R_{\theta,l}^{-1} \tilde{C} E\left[X_{\theta,l}^{(j)} \left(X_{\theta,l}^{(i)} \right)^T \right] \tilde{C}^T \right\}, \quad (4.45)$$

and substituting Eq. (4.44) into Eq. (4.45), we have

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

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

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

$$S_{\theta,1}^{(ji)} = \begin{bmatrix} \tilde{\phi}_\theta(t_0, \tau_1)\tilde{x}_{\theta,0} \\ \frac{\partial (\tilde{\phi}_\theta(t_0,\tau_1)\tilde{x}_{\theta,0})}{\partial \theta_j} \end{bmatrix} \left[\left(\tilde{\phi}_\theta(t_0, \tau_1)\tilde{x}_{\theta,0} \right)^T \quad \left(\frac{\partial (\tilde{\phi}_\theta(t_0,\tau_1)\tilde{x}_{\theta,0})}{\partial \theta_i} \right)^T \right], \qquad (4.46)$$

and it completes the proof. $\qquad \square$

In Section B.4, we provide an example to illustrate our results for calculating the Fisher information matrix for the specific case of a linear trajectory described in the example provided in Section B.1 of Appendix B.

4.2.1 CRLB and standard deviation of estimates for different photon counts

We next evaluate the performance of our proposed maximum likelihood estimation method in terms of the standard deviation of the estimates. For this purpose, we simulated data sets of the detected photons emitted from a molecule, referred to as the images of a molecule, with a stochastic trajectory, which differ by the mean photon count, i.e., the mean number of detected photons during the exposure time interval, assumed for each trajectory. This mean photon count ranges from 250 to 1250. For each mean photon count, the data set consists of 100 repeat images simulated using the Gaussian profile (Eq. (3.24)) with the parameters given in Section 3.4.1. For these data sets, we calculated the maximum likelihood estimates of the diffusion and the first order drift coefficients, separately. Also, for the given data set and time points, we obtained the square roots of the CRLBs for the diffusion and the first order drift coefficient by calculating the square roots of the inverse of their corresponding Fisher information matrices at the fixed time points. It can be seen in Fig. 4.1(a) that the standard deviations of the estimates are close to the square roots of their corresponding CRLBs, and when the mean number of photons increases, the standard deviation of the estimates decreases. Also, the percentage differences between the standard deviations and the square roots of the CRLBs are shown in Fig. 4.1(b). The percentage difference is the difference between the standard deviation of the estimates and the square root of the corresponding CRLB, expressed as a percentage of the square root of the corresponding CRLB. As can be seen in Fig. 4.1(b), these percentage differences are at most around 10%.

4.2.2 Fisher information matrix for non-Gaussian measurement noise

So far, for computational purposes and taking advantage of the Kalman filter formulation, we have focused on computing the Fisher information matrix and CRLB only for Gaussian measurements. Although the Gaussian assumption is very useful in some applications, there are many
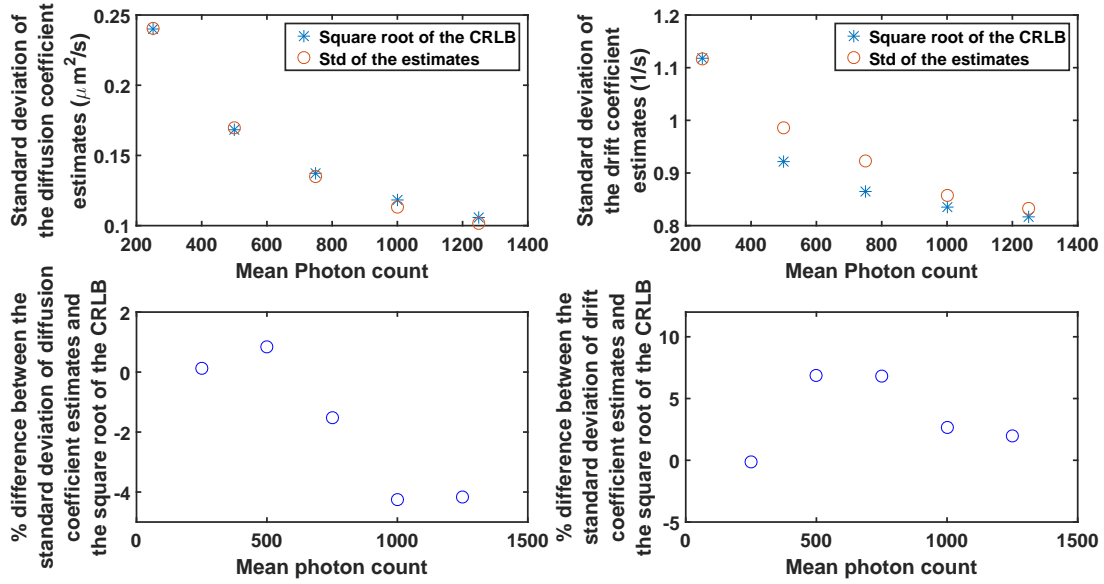
Figure 4.1: Analysis of the standard deviation of diffusion coefficient and drift coefficient estimates produced by the maximum likelihood estimation method for the Gaussian measurement noise case. Shown in the first row are the standard deviations of the diffusion coefficient and the first order drift coefficient estimates versus the square roots of their corresponding CRLBs for simulated data sets. The simulated data sets are the detected photons emitted from a molecule, referred to as the images of a molecule, with a stochastic trajectory, which differ by the mean photon count assumed for each trajectory. For each mean photon count, the data set consists of 100 repeat images. For a given data set, the time points of the detected photons are drawn from a Poisson process and are the same for all trajectories. All trajectories are simulated in the object space using Eq. (B.1) with the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1$ μm^2/s. Also, we assume that the initial location of the molecule is Gaussian distributed with mean $x_0 = (5, 5)^T$ μm and covariance $P_0 = 10 I_{2 \times 2}$ nm^2. Detected locations of the photons emitted from the molecule in the image space are simulated using Eq. (3.24) with the parameters given in Section 3.4.1. Shown in the second row are the percentage differences between the standard deviation of the diffusion coefficient and the first order drift coefficient estimates and the square roots of their corresponding CRLBs.

cases for which this assumption can be problematic in practice due to the fact that the Gaussian model is often not a suitable approximation for an analytical image profile. As mentioned earlier, from the optical diffraction theory, a typical point spread function for an in-focus molecule is given by the Airy profile. Also, for the out-of-focus scenario, the image function is given by a classical model of Born and Wolf [36].

Here, we computed the Fisher information matrix of both drift and diffusion coefficients for the Airy measurements case and compared the results with the Fisher information matrix obtained for the case that the Airy profile is approximated by a 2D Gaussian profile. The typical approximation of the Airy profile with $\alpha := 2\pi n_a/\lambda$ by a 2D Gaussian profile with standard deviation σ yields a value of $\sigma = 1.323/\alpha$ [6]. We only focused on the one photon case, since computing the integrals of the Fisher information expression for the Airy profile case numerically requires a large number of samples and it is computationally expensive (see Section B.5 in Appendix B for the detailed computational procedure). As shown in Fig. 4.2, the difference between the Fisher information matrices of these two different profiles can be significant.
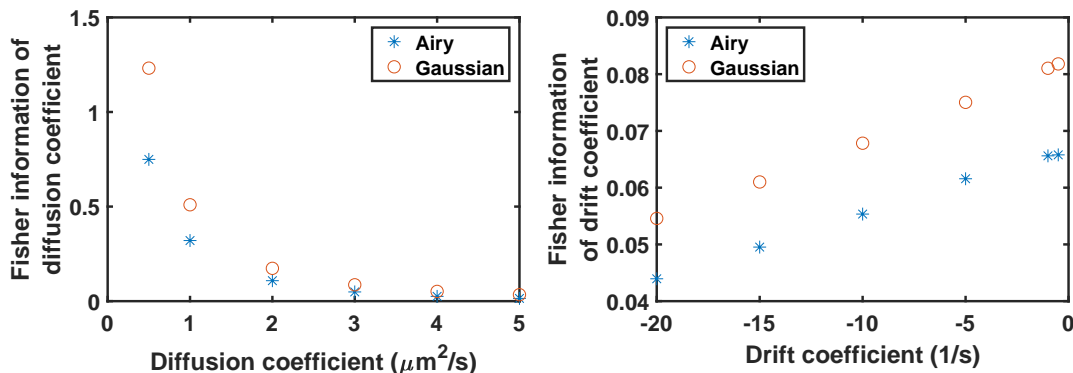

Figure 4.2: Fisher information matrix for Airy measurement noise versus Gaussian measurement noise. Fisher information matrix of diffusion and first order drift coefficients for the Airy measurement noise with parameter $\alpha = 2\pi n_a/\lambda$ given in Section 3.4.1 and by a 2D Gaussian profile with standard deviation $\sigma = 1.323/\alpha$, in case we have one photon with an arrival time of $\tau_1 = 20$ ms.

4.2.3 CRLB and Fisher information matrix for different sets of time points

To examine further the CRLB on parameter estimation for a moving single molecule with a stochastic trajectory, we calculated the square root of the CRLB for the simulated trajectories with the same parameters as in Fig. 4.1, and different time points drawn from a Poisson process with a mean value which ranges from 250 to 1250. As can be seen in Fig. 4.3, the square root of the CRLB for the drift coefficient highly depends on the time points of the observations, and does not necessarily improve by increasing the number of observations. On the other hand, the square root of the CRLB for the diffusion coefficient does not depend on the time points of the observations significantly, and always improves by increasing the number of observations.
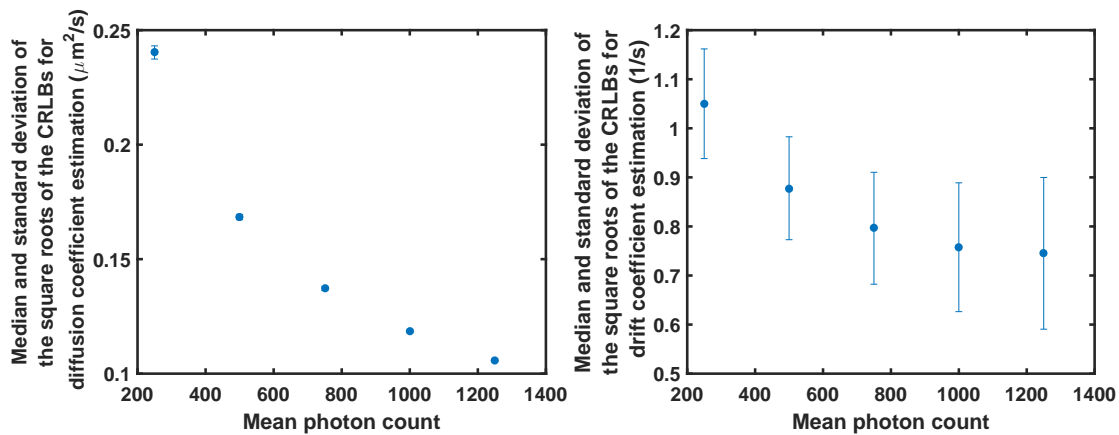

Figure 4.3: Analysis of the square root of the CRLB of the diffusion coefficient and drift coefficient estimates for different sets of Poisson distributed time points. Medians and standard deviations of the square roots of the CRLBs of the diffusion coefficient and the first order drift coefficient estimates are shown by the circles and error bars, respectively, for the simulated trajectories with the same parameters as in Fig. 4.1, and different time points drawn from a Poisson process with the same mean value, which ranges from 250 to 1250.

We also show the Fisher information matrices (and Fisher information matrix increments, i.e., the amount of information that we get by detecting a new photon) for Poisson distributed time points and for equally distributed time points in Fig. 4.4. For this purpose, we simulated two data

sets of single molecule trajectories with Gaussian measurements, first containing a trajectory of a molecule simulated using Eqs. (B.1), where the time points are drawn from a Poisson process with mean 250 in the time interval $[0, 50]$ ms, and second containing 250 equally spaced time points in the time interval $[0, 50]$ ms. We then calculated the Fisher information matrix increments and Fisher information matrix (sum of the increments) on the diffusion coefficient estimation for both data sets. As can be seen, the Fisher information matrix increments are the same for the equally spaced time points. However, for different realizations of Poisson time points, Fisher information matrix increments are different from each other.
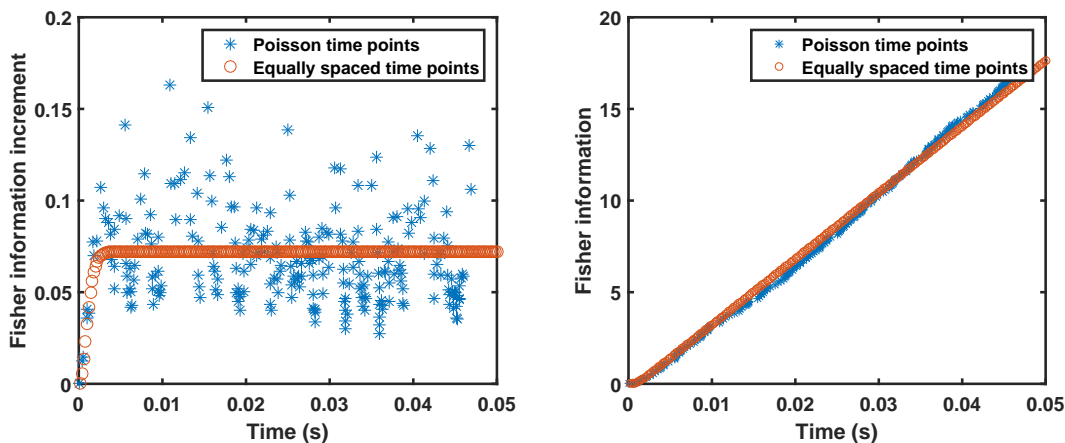

Figure 4.4: Fisher information analysis of single molecule trajectories simulated using Poisson distributed and equally spaced time points. Shown in the left are the Fisher information matrix increments on the diffusion coefficient estimation for data sets of two trajectories, first containing a trajectory of a molecule simulated using Eqs. (B.1), where the time points are drawn from a Poisson process with mean 250 in the time interval $[0, 50]$ ms, and second containing 250 equally spaced time points in the time interval $[0, 50]$ ms, with the parameters given in Fig. (4.1). Shown in the right is the Fisher information matrix (sum of the increments) for both trajectories.

5. EFFECT OF PIXELATION ON THE PARAMETER ESTIMATION OF SINGLE MOLECULE TRAJECTORIES AND FISHER INFORMATION MATRIX

5.1 Introduction

In Chapters 3 and 4, we studied the parameter estimation of single molecule trajectories for the fundamental data model, in which we had an ideal unpixelated detector and the measurements were the time points and locations of the detected photons. In practice, pixelated detectors are used in fluorescence microscopy, and therefore, the time points and exact locations of detection of the photons are not available anymore. In this case, the only information that we have is the pixel areas in which the photons impact the detector. Hence, the parameter estimation of single molecule trajectories form pixelated images are not a trivial problem.

The majority of available methods model the effect of pixelation by using an additive noise in the fundamental data model. However, in general, this approximation does not describe the underlying stochastic model precisely. For example, in [15, 39, 40], the effect of pixelation is encapsulated in a Gaussian additive random variable, referred to as the localization uncertainty. Therefore, there is a need for a systematic approach to analyze the effect of pixelation.

In this chapter, we introduce the practical data model, in which the measurements are the numbers of the photons detected in each pixel of a pixelated detector. Then, since the number of detected photons in each pixel can be described in terms of the time points and locations of the photons on the detector plane, we extend the results obtained for the fundamental data model to the practical data model.

5.2 Practical data model

In the practical data model, the data acquired by a pixelated detector are the number of detected photons at each pixel (Fig. 5.1). Let the pixelated detector $\mathcal{C}^p$ be defined as a collection $\{C_1, \cdots, C_K\}$ of open and disjoint subsets of $\mathcal{C}_s$, a region within $\mathbb{R}^2$ corresponding to the photon detection area of the detector, such that $\bigcup_{k=1}^{K} C_k = \mathcal{C}_s$. If $\mathcal{C}_s = \mathbb{R}^2$, the model is referred to as the
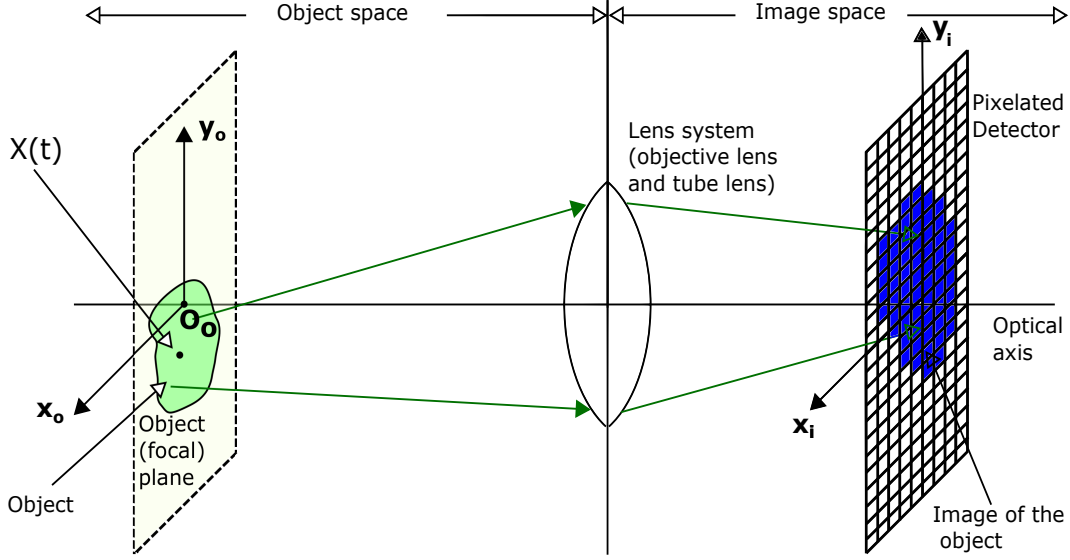
Figure 5.1: Schematic of an optical microscope. An object located in the object (focal) plane is imaged by an optical lens system and the image of the object is acquired by a pixelated detector in the image space. A 2D random variable $X(t), t \geq t_0$, describes the location of the object in the object plane at time t.

full practical data model. We use the random variable $S_k, k = 1, \cdots, K$, to describe the number of photons in the pixel C_k that result from the detection of photons from the object of interest. We also introduce the following notation. For $L, K = 1, 2, \cdots$, let A_K^L be a set of L-dimensional vectors $(a_1, \cdots, a_L), a_1, \cdots, a_L = 1, \cdots, K$. For a vector $v \in A_K^L$, let $\|v\|_{=k}, k = 1, \cdots, K$, denote the number of the elements of v which are equal to k. For example, for $v = (1, 1, 2) \in A_3^3$, we have $\|v\|_{=1} = 2, \|v\|_{=2} = 1$, and $\|v\|_{=3} = 0$. For $z_1, \cdots, z_K = 0, 1, \cdots$, and $\sum_{k=1}^{K} z_k = L$, let

$$A_K^L (z_1, \cdots, z_K) := \left\{ v \in A_K^L \mid \|v\|_{=k} = z_k, k = 1, \cdots, K \right\}. \tag{5.1}$$

For example,

$$A_2^3 := \{(1,1,1),(1,1,2),(1,2,1),(2,1,1),(2,2,1),(2,1,2),(1,2,2),(2,2,2)\},$$

$$A_2^3(1,2) := \{(2,2,1),(2,1,2),(1,2,2)\},$$

$$A_2^3(2,1) := \{(1,1,2),(1,2,1),(2,1,1)\},$$

$$A_2^3(0,3) := \{(2,2,2)\},$$

$$A_2^3(3,0) := \{(1,1,1)\}.$$

Note that the size $\left|A_K^L(z_1,\cdots,z_K)\right|$ of the set $A_K^L(z_1,\cdots,z_K)$ is equal to $\frac{L!}{z_1!\cdots z_K!}$. In Theorem 5, we calculate the joint probability of $S_1,\cdots,S_K$. But, we first calculate the joint probability density function of the locations of the photons detected by an ideal unpixelated detector in Lemma 2. In this chapter, the location of the photon emitted by the object, at time $\tau \geq t_0$, on the image plane is described by $U(X(\tau))$, where U is a random function that maps the object space into the image space. Also, we assume that only the locations of the photons emitted by the object can be detected by the unpixelated detector and the temporal information is not available. This further helps us to derive the formulae for the practical data model directly from the fundamental data model formulae.

Lemma 2. *The conditional probability density function* $p_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}$ *of* $U(X(T_1)),\cdots,$ *$U(X(T_L))$, given $N(t)$, can be calculated as*

$$p_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}(r_1,\cdots,r_L|L)$$
$$= \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L} \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} f_{x_1}(r_1)\cdots f_{x_L}(r_L) \left(\int_{t_0}^t \int_{t_0}^{\tau_L}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2} p_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L)\right.$$
$$\left.\times \prod_{i=1}^L \Lambda(\tau_i)d\tau_1 d\tau_2\cdots d\tau_{L-1}d\tau_L\right)dx_1\cdots dx_L, \tag{5.2}$$

where $r_1,\cdots,r_L \in \mathbb{R}^2$, and $p_{X(\tau_1),\cdots,X(\tau_L)}, t_0 \leq \tau_1 < \cdots < \tau_L \leq t$, is the joint probability density function of $X(\tau_1),\cdots,X(\tau_L)$. If $\{X(\tau_1),\cdots,X(\tau_L)\}$ is a Markov sequence, then, for

$x_1, \cdots, x_L \in \mathbb{R}^2$,

$$p_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L) = p_{X(\tau_L)|X(\tau_{L-1})}(x_L|x_{L-1})\cdots p_{X(\tau_2)|X(\tau_1)}(x_2|x_1)\, p_{X(\tau_1)}(x_1),$$

where $p_{X(\tau_l)|X(\tau_{l-1})}, l = 2,\cdots, L$, is the conditional probability density function of $X(\tau_l)$, given $X(\tau_{l-1})$, and $p_{X(\tau_1)}$ is the probability density function of $X(\tau_1)$.

Proof. We have

$p_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}(r_1,\cdots,r_L|L)$

$\displaystyle = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} p_{U(X(T_1)),\cdots,U(X(T_L)),X(T_1),\cdots,X(T_L)|N(t)}(r_{1:L}, x_{1:L}|L)\, dx_L\cdots dx_1$

$\displaystyle = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} p_{U(X(T_1)),\cdots,U(X(T_L))|X(T_1),\cdots,X(T_L),N(t)}(r_{1:L}|x_{1:L}, L)\, p_{X(T_1),\cdots,X(T_L)|N(t)}(x_{1:L}|L)\, dx_L\cdots dx_1$

$\displaystyle = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} p_{U(X(T_1))|X(T_1)}(r_1|x_1)\cdots p_{U(X(T_L))|X(T_L)}(r_L|x_L)$

$\displaystyle \times \left(\int_{t_0}^{t}\int_{t_0}^{\tau_L}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2} p_{X(T_1),\cdots,X(T_L),T_1,\cdots,T_L|N(t)}(x_{1:L},\tau_{1:L}|L)\, d\tau_1 d\tau_2\cdots d\tau_{L-1}d\tau_L \right) dx_L\cdots dx_1$

$\displaystyle = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} p_{U(x_1)}(r_1)\cdots p_{U(x_L)}(r_L) \left(\int_{t_0}^{t}\int_{t_0}^{\tau_L}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2} p_{X(T_1),\cdots,X(T_L)|T_1,\cdots,T_L,N(t)}(x_{1:L}|\tau_{1:L}, L) \right.$

$\displaystyle \left. \times\, p_{T_1,\cdots,T_L|N(t)}(\tau_{1:L}|L)\, d\tau_1 d\tau_2\cdots d\tau_{L-1}d\tau_L \right) dx_L\cdots dx_1$

$\displaystyle = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} f_{x_1}(r_1)\cdots f_{x_L}(r_L) \left(\int_{t_0}^{t}\int_{t_0}^{\tau_L}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2} p_{X(\tau_1),\cdots,X(\tau_L)}(x_{1:L}) \frac{L!\prod_{l=1}^{L}\Lambda(\tau_l)}{\left(\int_{t_0}^{t}\Lambda(\psi)d\psi\right)^L} \right.$

$\displaystyle \left. \times\, d\tau_1 d\tau_2\cdots d\tau_{L-1}d\tau_L \right) dx_L\cdots dx_1$

$\displaystyle = \frac{L!}{\left(\int_{t_0}^{t}\Lambda(\psi)d\psi\right)^L} \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} f_{x_1}(r_1)\cdots f_{x_L}(r_L) \left(\int_{t_0}^{t}\int_{t_0}^{\tau_L}\cdots\int_{t_0}^{\tau_3}\int_{t_0}^{\tau_2} p_{X(\tau_1),\cdots,X(\tau_L)}(x_{1:L}) \right.$

$\displaystyle \left. \times \prod_{l=1}^{L}\Lambda(\tau_l)d\tau_1 d\tau_2\cdots d\tau_{L-1}d\tau_L \right) dx_L\cdots dx_1,$

where $x_{1:L} := (x_1,\cdots,x_L)$, $r_{1:L} := (r_1,\cdots,r_L)$, $\tau_{1:L} := (\tau_1,\cdots,\tau_L)$, and $p_{X(\tau_1),\cdots,X(\tau_L)}, t_0 \leq \tau_1 < \cdots < \tau_L \leq t$, is the joint probability density function of $X(\tau_1),\cdots,X(\tau_L)$. $\qquad\square$

In the following theorem, we calculate the probability $Pr\left[S_1 = z_1,\cdots,S_K = z_K\right]$ in terms of $p_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}$ calculated in the above lemma.

Theorem 5. *1. In the practical data model, for $z_1, \cdots, z_K = 0, 1, \cdots,$ and $\sum_{k=1}^{K} z_k = L$, the probability $Pr[S_1 = z_1, \cdots, S_K = z_K]$ is given by*

$$Pr[S_1 = z_1, \cdots, S_K = z_K] = \sum_{z=0}^{\infty} \left(\sum_{v := (v_1, \cdots, v_{L+z}) \in A_{K+1}^{L+z}(z_1, \cdots, z_K, z)} \right.$$
$$\left. \int_{C_{v_1}} \cdots \int_{C_{v_{L+z}}} p_{U(X(T_1)), \cdots, U(X(T_{L+z}))|N(t)} (r_1, \cdots, r_{L+z}|L+z) \, dr_{L+z} \cdots dr_1 \right),$$

$$(5.3)$$

where $p_{U(X(T_1)), \cdots, U(X(T_{L+z}))|N(t)}, z = 0, 1, \cdots,$ is the conditional probability density function of $U(X(T_1)), \cdots, U(X(T_{L+z}))$, given $N(t)$.

2. In the full practical data model, we have

$$Pr[S_1 = z_1, \cdots, S_K = z_K] = \sum_{v := (v_1, \cdots, v_L) \in A_K^L(z_1, \cdots, z_K)}$$
$$\int_{C_{v_1}} \cdots \int_{C_{v_L}} p_{U(X(T_1)), \cdots, U(X(T_L))|N(t)} (r_1, \cdots, r_L|L) \, dr_L \cdots dr_1. \qquad (5.4)$$

Proof. 1. Let the random variable S_{K+1} describe the number of photons in the complement pixel $C_{K+1} := \mathbb{R}^2 - \bigcup_{k=1}^{K} C_k$ that result from the detection of the photons emitted from the object of interest. Then, according to the definitions of $S_1, \cdots, S_{K+1}$, we have, for $z_1, \cdots, z_K = 0, 1, \cdots,$ and $L = \sum_{k=1}^{K} z_k$,

$$Pr[S_1 = z_1, \cdots, S_K = z_K]$$
$$= \sum_{z=0}^{\infty} Pr[S_1 = z_1, \cdots, S_K = z_K, S_{K+1} = z]$$
$$= \sum_{z=0}^{\infty} Pr\left[\bigcup_{v := (v_1, \cdots, v_{L+z}) \in A_{K+1}^{L+z}(z_1, \cdots, z_K, z)} \left\{ \bigcap_{l=1}^{L+z} (U(X(T_l)) \in C_{v_l}) \, | N(t) = L + z \right\} \right]. \qquad (5.5)$$

Since the events $\left\{ \bigcap_{l=1}^{L+z} (U(X(T_l)) \in C_{v_l}) \right\}$ are mutually exclusive, we have

$$Pr\left[\bigcup_{v := (v_1, \cdots, v_{L+z}) \in A_{K+1}^{L+z}(z_1, \cdots, z_K, z)} \left\{ \bigcap_{l=1}^{L+z} (U(X(T_l)) \in C_{v_l}) \, | N(t) = L + z \right\} \right]$$
$$= \sum_{v := (v_1, \cdots, v_{L+z}) \in A_{K+1}^{L+z}(z_1, \cdots, z_K, z)} Pr\left[\bigcap_{l=1}^{L+z} (U(X(T_l)) \in C_{v_l}) \, | N(t) = L + z \right],$$

101

and therefore,

$$Pr\left[S_1 = z_1, \cdots, S_K = z_K\right]$$

$$= \sum_{z=0}^{\infty} \left(\sum_{v:=(v_1,\cdots,v_{L+z}) \in A_{K+1}^{L+z}(z_1,\cdots,z_K,z)} Pr\left[\bigcap_{l=1}^{L+z} \left(U(X(T_l)) \in C_{v_l}\right) | N(t) = L + z\right] \right)$$

$$= \sum_{z=0}^{\infty} \left(\sum_{v:=(v_1,\cdots,v_{L+z}) \in A_{K+1}^{L+z}(z_1,\cdots,z_K,z)} \int_{C_{v_1}} \cdots \int_{C_{v_{L+z}}} p_{U(X(T_1)),\cdots,U(X(T_{L+z}))|N(t)}\left(r_1,\cdots,r_{L+z}|L+z\right) \right.$$

$$\left. \times dr_{L+z} \cdots dr_1 \right). \tag{5.6}$$

2. The result follows using the similar approach used in part 1. $\qquad\square$

5.3 Maximum likelihood estimation

Let Θ denote the parameter space that is an open subset of $\mathbb{R}^n$. The maximum likelihood estimate $\hat{\theta}_{mle}$ of $\theta \in \Theta$ is given by

$$\hat{\theta}_{mle} = \operatorname*{argmin}_{\theta \in \Theta} \left(-\log \mathcal{L}_p(\theta | z_1, \cdots, z_K) \right),$$

where the likelihood function $\mathcal{L}_p$ for the practical data model is given by

$$\mathcal{L}_p(\theta | z_1, \cdots, z_K) = Pr^\theta \left[S_1 = z_1, \cdots, S_K = z_K\right],$$

in which $\{z_1, \cdots, z_K\}$, $z_1, \cdots, z_K = 0, 1, \cdots, L = \sum_{k=1}^{K} z_k$, denotes an image with K pixels. In the above equation, the probability $Pr^\theta \left[S_1 = z_1, \cdots, S_K = z_K\right]$ is given by Eqs. (5.3) and (5.4) for the practical and full practical data models, respectively.

In general, computing the integrals of the likelihood function is not a trivial task. Here, based on the Monte Carlo approach provided in [45], we develop an algorithm to approximate these integrals. The basis of our algorithm is the law of large numbers which can be stated as follows. Let $X(\tau_1), \cdots, X(\tau_L)$ be 2D random variables that describe the locations of the object at time points

$\tau_0 := t_0 \leq \tau_1 < \cdots < \tau_L \leq t$. Let $p_{X(\tau_1),\cdots,X(\tau_L)}$ be the joint distribution of $X(\tau_1),\cdots,X(\tau_L)$. For $h\colon \mathbb{R}^{2\times L} \mapsto \mathbb{R}, v := (v_1,\cdots,v_L) \in A_K^L(z_1,\cdots,z_K), z_1,\cdots,z_K = 0,1,\cdots$, and $\sum_{k=1}^K z_k = L$, let

$$h(x_1,\cdots,x_L) = \prod_{l=1}^L I_{C_{v_l}}(x_l), \quad x_1,\cdots,x_L \in \mathbb{R}^2, \tag{5.7}$$

where, for an invertible magnification matrix $M \in \mathbb{R}^{2\times 2}$,

$$I_{C_{v_l}}(x_l) = \int_{C_{v_l}} f_{x_l}(r)dr = \frac{1}{|\det(M)|} \int_{C_{v_l}} q\left(M^{-1}r - x_l\right) dr. \tag{5.8}$$

Also, let

$$E\{h(X(\tau_1),\cdots,X(\tau_L))\} = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} h(x_1,\cdots,x_L)\, p_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L)\, dx_L\cdots dx_1,$$

be the expected value of $h(X(\tau_1),\cdots,X(\tau_L))$ with respect to $p_{X(\tau_1),\cdots,X(\tau_L)}$. Then, according to the law of large numbers,

$$\lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^M \left(\prod_{l=1}^L I_{C_{v_l}}(x_l^m)\right) = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} \left(\prod_{l=1}^L I_{C_{v_l}}(x_l)\right) p_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L)\, dx_L\cdots dx_1,$$

where $\{X^m := (x_1^m,\cdots,x_L^m)\}_{m=1}^M, x_l^m \in \mathbb{R}^2, l = 1,\cdots,L, m = 1,\cdots,M$, is a sequence of independent and identically distributed samples drawn from the distribution $p_{X(\tau_1),\cdots,X(\tau_L)}$. In case that $\{X(\tau_l)\}_{l=1}^L$ is a Markov sequence, i.e.,

$$p_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L) = p_{X(\tau_1)}(x_1)\prod_{l=2}^L p_{X(\tau_l)|X(\tau_{l-1})}(x_l|x_{l-1}), \quad x_1,\cdots,x_L \in \mathbb{R}^2,$$

we draw $X^m, m = 1,\cdots,M$, through the following Monte Carlo algorithm:

Algorithm 2 (Monte Carlo method). *Step 1. Draw independent and identically distributed (i.i.d.) samples $\{x_1^i\}_{i=1}^M$ according to $p_{X(\tau_1)}(x), x \in \mathbb{R}^2$, i.e., $x_1^i \sim p_{X(\tau_1)}(x), i = 1,\cdots,M$.*

Step 2. Draw i.i.d. samples $\{x_2^i\}_{i=1}^M$ according to $p_{X(\tau_2)|X(\tau_1)}\left(x|x_1^i\right), x \in \mathbb{R}^2$, i.e., $x_2^i \sim$

103

$p_{X(\tau_2)|X(\tau_1)}\left(x|x_1^i\right), i = 1, \cdots, M.$

$\vdots$

    Step L. Draw i.i.d. samples $\{x_L^i\}_{i=1}^M$ according to $p_{X(\tau_L)|X(\tau_{L-1})}\left(x|x_{L-1}^i\right), x \in \mathbb{R}^2$, i.e., $x_L^i \sim$

$p_{X(\tau_L)|X(\tau_{L-1})}$

$\left(x|x_{L-1}^i\right), i = 1, \cdots, M.$

    Step $L+1$. For $v := (v_1, \cdots, v_L) \in A_K^L(z_1, \cdots, z_K), z_1, \cdots, z_K = 0, 1, \cdots,$ and $\sum_{k=1}^K z_k = $

L, approximate the probability $Pr\left[\bigcap_{l=1}^L (X(\tau_l) \in C_{v_l})\right]$ as

$$Pr\left[\bigcap_{l=1}^L (X(\tau_l) \in C_{v_l})\right] = \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \left(\prod_{l=1}^L I_{C_{v_l}}(x_l)\right) p_{X(\tau_1),\cdots,X(\tau_L)}(x_1, \cdots, x_L) \, dx_L \cdots dx_1$$

$$\approx \frac{1}{M} \sum_{m=1}^M \left(\prod_{l=1}^L I_{C_{v_l}}(x_l^m)\right).$$

In the following example, we assess the performance of the above algorithm in the computation of the likelihood function for a simple scenario.

Example 1. *Assume that we have a typical two-dimensional single molecule trajectory $X(\tau)$ in the object space, where the time point $\tau = 0.01$ ms is fixed, with the first order drift coefficient $F = 10$/s and the diffusion coefficient $D = 1$ μ^2/s. Also, we assume that the initial location of the molecule is known and given by $x_0 = (2.4, 2.4)^T$ μm. In the fundamental data model, detected locations of the photons emitted from the molecule in the image space are simulated using a zero-mean Gaussian profile with covariance matrix $\Sigma = 0.01 I_{2\times2}$ μ^2m. In the practical data model, a 60×60 pixelated detector with square pixels of a width of $W = 16$ μm is used to acquire the pixelated image of the molecule trajectory. Assume that the photon emitted from the object hits the pixel C_1 centered at $\left(c_x^1, c_y^1\right) = (230.75, 237.25)^T$ μm at the image space. Then, using Algorithm 2 we have*

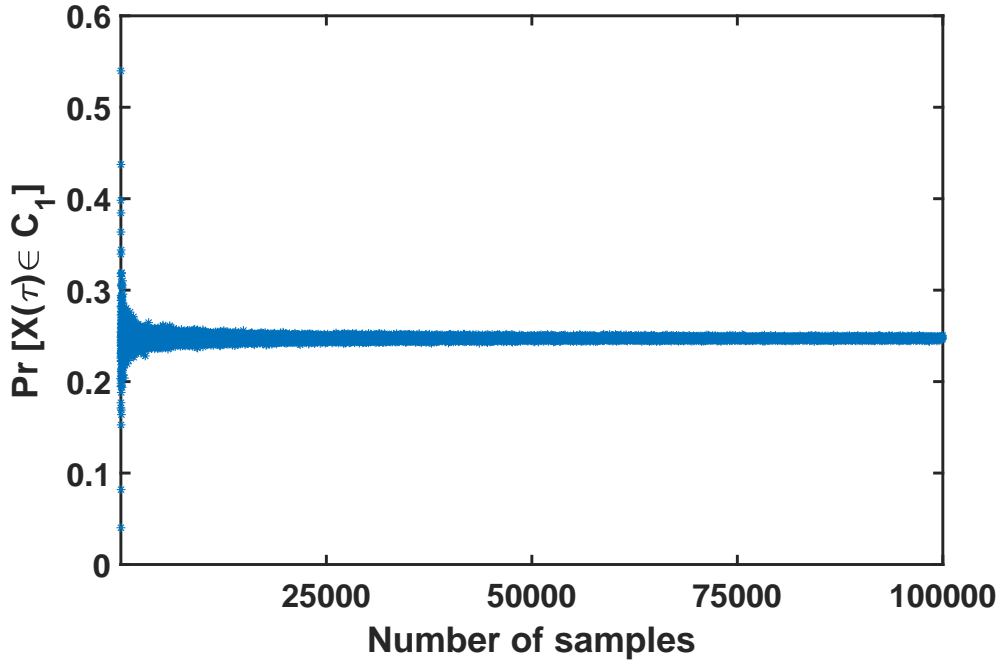$$Pr\left[(X(\tau) \in C_1\right] = \int_{\mathbb{R}^2} I_{C_1}(x) p_{X(\tau)}(x) dx \approx \frac{1}{M} \sum_{m=1}^M I_{C_1}(x^m),$$

Figure 5.2: Convergence of the Monte Carlo method. The probabilities $Pr\left[(X(\tau) \in C_1\right]$ for different number M of Monte Carlo samples, where $X(\tau)$ is a two-dimensional single molecule trajectory, are shown in which the time point $\tau = 0.01$ ms is fixed, with the first order drift coefficient $F = 10$/s and the diffusion coefficient $D = 1 \ \mu^2$/s. Also, assume that the initial location of the molecule is known and given by $x_0 = (2.4, 2.4)^T \ \mu$m. Detected locations of the photons emitted from the molecule in the image space are simulated using a zero-mean Gaussian model with covariance matrix $\Sigma = 0.01 I_{2\times 2} \ \mu^2$m. A 60×60 pixelated detector with square pixels of a width of $W = 16 \ \mu$m is used to acquire the pixelated image of the molecule trajectory. Assume that the photon emitted from the object hits the pixel C_1 centered at $\left(c_x^1, c_y^1\right) = (230.75, 237.25)^T$ μm at the image space.

where, for an invertible magnification matrix $M \in \mathbb{R}^{2 \times 2}$,

$$I_{C_1}(x) = \frac{1}{|\det(M)|} \int_{C_1} q\left(M^{-1}r - x\right) dr$$

$$= \frac{1}{|\det(M)|} \int_{c_x^1 - \frac{W}{2}}^{c_x^1 + \frac{W}{2}} \int_{c_y^1 - \frac{W}{2}}^{c_y^1 + \frac{W}{2}} q\left(M^{-1}(r_x, r_y) - x\right) dr_y dr_x, \quad x \in \mathbb{R}^2, \qquad (5.9)$$

and $\{x^m\}_{m=1}^M, x^m \in \mathbb{R}^2, m = 1, \cdots, M$, is a sequence of independent and identically distributed samples drawn from the distribution $p_{X(\tau)}$ using Algorithm 2. In Fig. 5.2, we have shown the probabilities $Pr\left[(X(\tau) \in C_1\right]$ computed for different number M of Monte Carlo samples. As can be seen in Fig. 5.3, the standard deviation of the probabilities decreases by increasing the number of samples, which suggests the convergence of these probabilities.
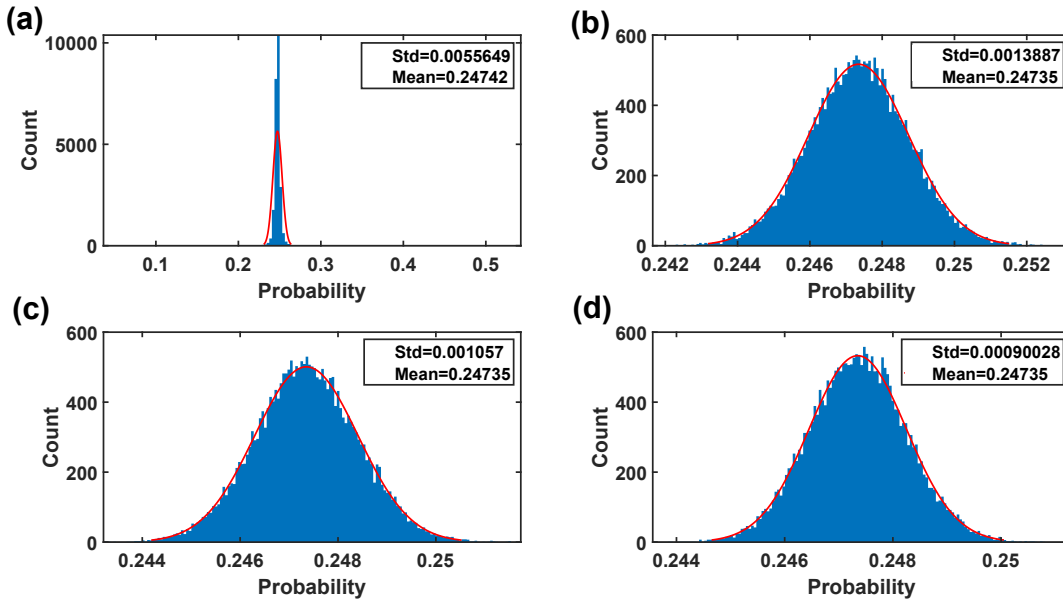

Figure 5.3: Histograms, means and standard deviations of the probabilities computed using the Monte Carlo method. Gaussian models fitted to the histograms of the probabilities computed using the Monte Carlo method are shown for (a) first, (b) second, (c) third and (d) fourth quarters of the data sets of Fig. 5.2.

We next examine the performance of our proposed parameter estimation method. For this purpose, we simulated pixelated images of single molecule trajectories. These trajectories were simulated using Eq. (B.1) with four time points, where the time points were drawn from a Poisson process, and the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1.5$ μm^2/s. Also, we assumed that the initial location of the molecule was fixed at $(2.4, 2.4)^T$ μm. The locations of the photons emitted from the molecule trajectories, in the image space, were simulated using Eq. (3.23) with the Gaussian measurement noise (Eq. (3.12)) and $\sigma = 0.1$ μm. We assumed that these photons were detected using a pixelated detector of pixel size and image size of 6.5×6.5 μm and 60×60 pixels, respectively. We then estimated all parameters of the trajectories, e.g., initial location of the molecule, drift and diffusion coefficients, together using Algorithm 2, where the number of Monte Carlo samples at each step is equal to 2500. The errors (estimate - true value) of the estimation are shown in Figs. 5.4 and 5.5. As can be seen in these figures, the spreads of the errors are around zero and there is no systematic bias associated with the estimates.

We also applied the algorithm to the pixelated images of single molecule trajectories simulated using an Airy point spread functions with $\alpha = \frac{2\pi n_a}{\lambda} = 13.23$, which corresponds to a Gaussian profile with $\sigma = 0.1$ μm. The parameters of the molecule trajectories were the same as the parameters of the data set of Fig. 5.4. As can be seen in Figs. 5.6 and 5.7, we have obtained the similar results as the Gaussian case.

We further evaluate the performance of the proposed method in terms of the standard deviation of the estimates. In order to do this, we simulated the pixelated images of a stationary object using a pixelated detector of pixel size and image size of 6.5×6.5 μm and 60×60 pixels, respectively, assuming that three photons were detected by the detector. The locations of the photons in the image space were simulated using Eq. (3.23) with the Gaussian measurement noise (Eq. (3.12)) and $\sigma = 0.1$ μm. We then estimated the location of the molecule using Algorithm 2, where the number of Monte Carlo samples at each step is equal to 10000. The errors of the location estimates are shown in Fig. 5.8. As before, the errors are spreading around zero and no systematic bias can be
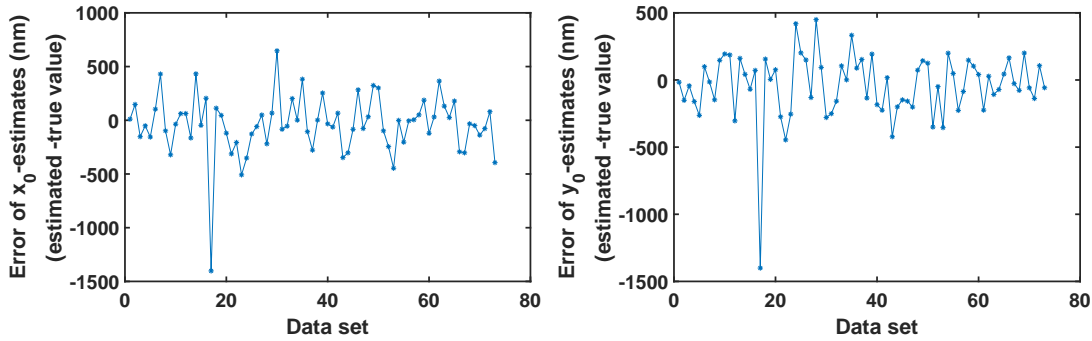
Figure 5.4: Analysis of the error of initial location estimates from pixelated images of single molecule trajectories for the Gaussian measurement noise case. Differences between the estimates of the initial x_0- and y_0-location of the molecule and their true values from the images of the molecule trajectories simulated using Eqs. (B.1) with four time points, where the time points are drawn from a Poisson process, and the first order drift coefficient $F = -10$/s and the diffusion coefficient $D = 1.5$ μm^2/s. The initial location of the molecule is fixed at $X_0 := (x_0, y_0) = (2.4, 2.4)^T$ μm. The locations of the photons emitted from the molecule trajectories, in the image space, are simulated using Eq. (3.23) with the Gaussian measurement noise (Eq. (3.12)) and $\sigma = 0.1$ μm. These photons are detected using a pixelated detector of pixel size and image size of 6.5×6.5 μm and 60×60 pixels, respectively.
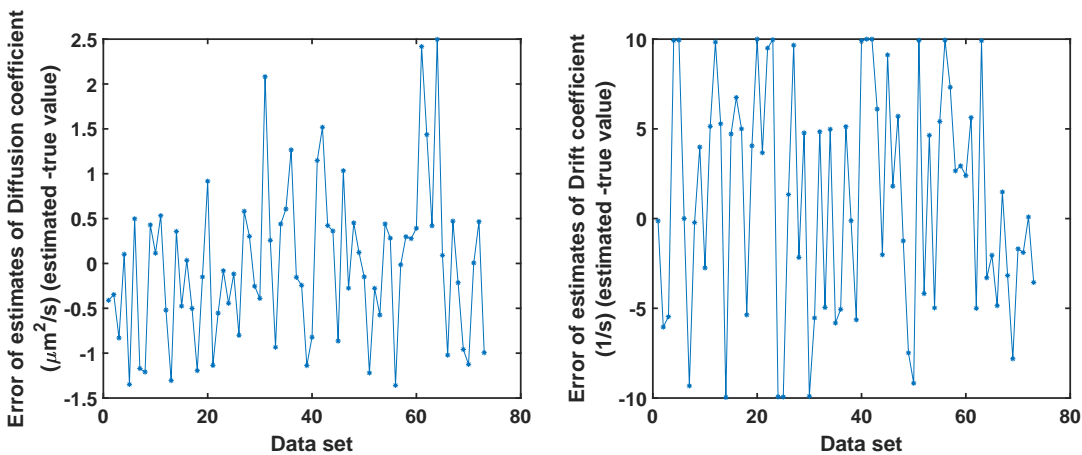

Figure 5.5: Analysis of the error of diffusion coefficient and drift coefficient estimates from pixelated images of single molecule trajectories for the Gaussian measurement noise case. Differences between the diffusion (first order drift) coefficient estimates and the true diffusion (first order drift) coefficient value for data sets of Fig. 5.4.
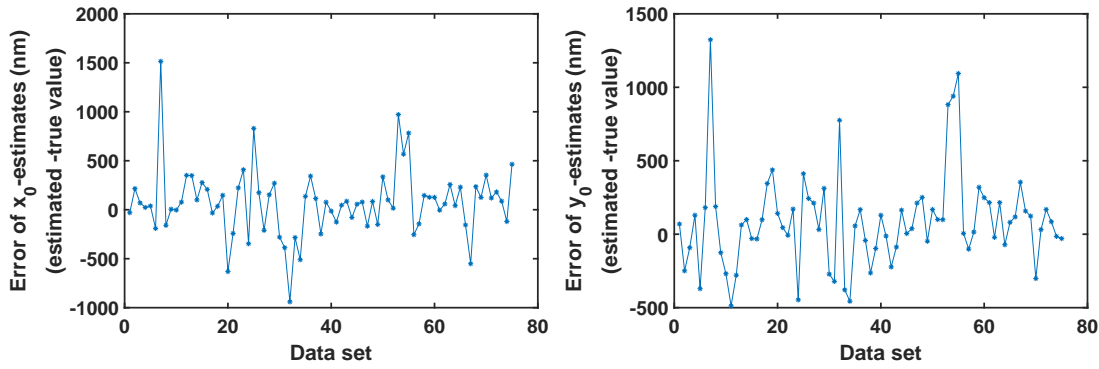
Figure 5.6: Analysis of the error of initial location estimates from pixelated images of single molecule trajectories for the Airy measurement noise case. Differences between the estimates of the initial x_0- and y_0-location of the molecule and their true values from the images of the molecule trajectories simulated using the parameters of the data set of Fig. 5.4. The locations of the photons emitted from the molecule trajectories, in the image space, are simulated using an Airy model with $\alpha = \frac{2\pi n_a}{\lambda} = 13.23$. These photons are detected using a pixelated detector of pixel size and image size of $6.5 \times 6.5\ \mu$m and 60×60 pixels, respectively.
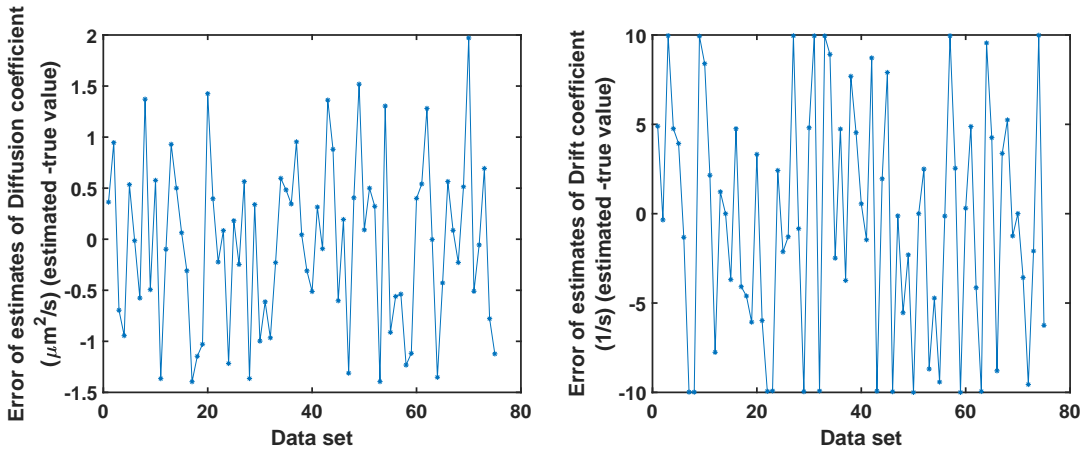

Figure 5.7: Analysis of the error of diffusion coefficient and drift coefficient estimates from pixelated images of single molecule trajectories for the Airy measurement noise case. Differences between the diffusion (first order drift) coefficient estimates and the true diffusion (first order drift) coefficient value for data sets of Fig. 5.6.

seen. We also calculated the standard deviations of the estimates. These standard deviations, which are computed as 57.4 nm and 59.6 nm for the x_0- and y_0-locations of the molecule, respectively, are close to the localization accuracy, i.e., the square root of the CRLB, which is given as 58.37 nm for both x- and y-directions, reported in [6].

Here, we only consider a small number of photons, since, in general, the computation of the likelihood function (Eq. (5.4)) is expensive. It is mostly because of the large number of the members of the set $A_K^L(z_1, \cdots, z_K)$, which is equal to $\frac{L!}{z_1! \cdots z_K!}$, when L increases. In the future, using more advanced computational methods, the formulae provided in this chapter can be implemented more efficiently.
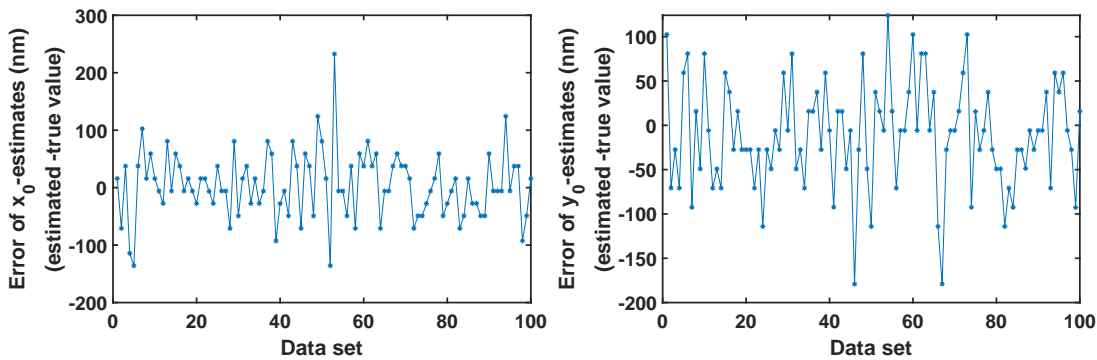

Figure 5.8: Analysis of the error of location estimates from pixelated images of a stationary molecule for the Gaussian measurement noise case. Differences between the estimates of the initial x_0- and y_0-location of the molecule and their true values from the simulated images of a stationary molecule using a pixelated detector of pixel size and image size of 6.5×6.5 μm and 60×60 pixels, respectively, assuming that three photons are detected by the detector. The locations of the photons in the image space are simulated using Eq. (3.23) with the Gaussian measurement noise (Eq. (3.12)) and $\sigma = 0.1$ μm.

5.3.1 Fisher information matrix

In the previous chapter, we provided a recursive formulation for the Fisher information matrix. Here, we calculate a new expression for the Fisher information matrix for the fundamental data model. Next, we use this new result to derive an expression for the Fisher information matrix for

the practical data model. We first, in Definition 5, introduce a notation for the Fisher information matrix of the fundamental data model given the number of photons.

Definition 5. *Let the parameter space Θ describe an open subset of $\mathbb{R}^n$ containing the true parameters. For $L = 1, 2, \cdots$, and a row parameter vector $\theta \in \Theta$, we introduce the Fisher information matrix of the fundamental data model given $N(t) = L$, as*

$$
I^f_{N(t)=L}(\theta) := E_{p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)=L}} \left\{ \left(\frac{\partial \log p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}\, (r_{1:L}|L)}{\partial \theta} \right)^T \right.
$$

$$
\left. \times \left(\frac{\partial \log p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}\, (r_{1:L}|L)}{\partial \theta} \right) \right\}
$$

$$
= \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}\, (r_{1:L}|L) \left(\frac{\partial \log p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)|N(t)}\, (r_{1:L}|L)}{\partial \theta} \right)^T
$$

$$
\times \left(\frac{\partial \log p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}\, (r_{1:L}|L)}{\partial \theta} \right) dr_1 \cdots dr_L, \tag{5.10}
$$

where $r_{1:L} := (r_1, \cdots, r_L), r_1, \cdots, r_L \in \mathbb{R}^2$, $E_{p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)=L}}$ is the expected value with respect to the probability $p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)=L}$, and $p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}$ is the conditional probability density function of $U(X(T_1)), \cdots, U(X(T_L))$, given $N(t)$.

In the following theorem, we calculate the Fisher information matrix of the fundamental data model defined in the above definition. In the rest of this paper, we only focus on the estimation of the parameters of the motion model of the object, such as the initial location of the object, drift and diffusion coefficients, i.e., we assume that Λ and f_x are independent of θ.

Theorem 6. *For a row parameter vector $\theta \in \Theta$, the Fisher information matrix $I^f_{N(t)=L}(\theta), L = 1, 2, \cdots$, of the fundamental data model given $N(t) = L$, can be calculated as*

$$
I^f_{N(t)=L}(\theta) = \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi \right)^L} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \frac{I^f_\theta\, (r_{1:L}, r_{1:L})}{p^\theta_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}\, (r_1, \cdots, r_L|L)} dr_1 \cdots dr_L, \tag{5.11}
$$

where, for $r_{1:L} := (r_1, \cdots, r_L)$, $r'_{1:L} := (r'_1, \cdots, r'_L)$, *and* $r_1, \cdots, r_L, r'_1, \cdots, r'_L \in \mathbb{R}^2$,

$$I_\theta^f(r_{1:L}, r'_{1:L}) := \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \left\{ \int_{t_0}^t \int_{t_0}^{\tau'_L} \cdots \int_{t_0}^{\tau'_2} \int_{t_0}^t \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_2} dF_\theta^T(r_{1:L}, x_{1:L}, \tau_{1:L}) \right.$$

$$\left. \times dF_\theta(r'_{1:L}, x'_{1:L}, \tau'_{1:L}) \, d\tau_1 \cdots d\tau_{L-1} d\tau_L d\tau'_1 \cdots d\tau'_{L-1} d\tau'_L \right\} dx_1 \cdots dx_L dx'_1 \cdots dx'_L,$$

$$p_{U(X(T_1)), \cdots, U(X(T_L))|N(t)}^\theta(r_{1:L}|L) = \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \left(\int_{t_0}^t \int_{t_0}^{\tau''_L} \cdots \int_{t_0}^{\tau''_2} F_\theta(r_{1:L}, x''_{1:L}, \tau''_{1:L}) \, d\tau''_1 \cdots d\tau''_{L-1} d\tau''_L \right)$$

$$\times dx''_1 \cdots dx''_L,$$

$$F_\theta(r_{1:L}, x_{1:L}, \tau_{1:L}) := \left(\prod_{i=1}^L f_{x_i}(r_i) \Lambda(\tau_i) \right) p_{X(\tau_1), \cdots, X(\tau_L)}^\theta(x_1, \cdots, x_L),$$

$$dF_\theta(r_{1:L}, x_{1:L}, \tau_{1:L}) := \frac{\partial}{\partial \theta} F_\theta(r_{1:L}, x_{1:L}, \tau_{1:L}) = \left(\prod_{i=1}^L f_{x_i}(r_i) \Lambda(\tau_i) \right) dp_{X(\tau_1), \cdots, X(\tau_L)}^\theta(x_1, \cdots, x_L), \qquad (5.12)$$

in which $x_{1:L} := (x_1, \cdots, x_L)$, $x'_{1:L} := (x'_1, \cdots, x'_L)$, $x_1, \cdots, x_L, x'_1, \cdots, x'_L \in \mathbb{R}^2$, $\tau_{1:L} :=$
$(\tau_1, \cdots, \tau_L)$, $\tau'_{1:L} := (\tau'_1, \cdots, \tau'_L)$, $t_0 \leq \tau_1 < \cdots < \tau_L \leq t$, $t_0 \leq \tau'_1 < \cdots < \tau'_L \leq t$, *and*
$dp_{X(\tau_1), \cdots, X(\tau_L)}^\theta(x_1, \cdots, x_L) := \frac{\partial p_{X(\tau_1), \cdots, X(\tau_L)}^\theta(x_1, \cdots, x_L)}{\partial \theta}$.

Proof. For a row parameter vector $\theta \in \Theta$, the Fisher information matrix $I_{N(t)=1}(\theta)$, given $N(t) = 1$, can be calculated as

$$I_{N(t)=1}^f(\theta) = \frac{1}{\int_{t_0}^t \Lambda(\psi) d\psi}$$

$$\times \int_{\mathbb{R}^2} \left[\frac{\int_{\mathbb{R}^2} \int_{\mathbb{R}^2} f_x(r) f_{x'}(r) \left\{ \int_{t_0}^t \int_{t_0}^t \left(dp_{X(\tau_1)}^\theta(x) \right)^T dp_{X(\tau_2)}^\theta(x') \Lambda(\tau_1) \Lambda(\tau_2) d\tau_1 d\tau_2 \right\} dx dx'}{\int_{\mathbb{R}^2} f_x''(r) \left(\int_{t_0}^t p_{X(\tau)}^\theta(x'') \Lambda(\tau) d\tau \right) dx''} \right] dr,$$

where $dp_{X(\tau)}^\theta(x) := \frac{\partial p_{X(\tau)}^\theta(x)}{\partial \theta}$, $x \in \mathbb{R}^2$, $\tau \geq t_0$. In general, the Fisher information matrix $I_{N(t)=L}^f(\theta)$, given $N(t) = L$, can be calculated as, according to Eq. (5.10) of Definition 5,

$$I_{N(t)=L}^f(\theta) := \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \frac{1}{p_{U(X(T_1)), \cdots, U(X(T_L))|N(t)}^\theta(r_{1:L}|L)} \left(\frac{\partial p_{U(X(T_1)), \cdots, U(X(T_L))|N(t)}^\theta(r_{1:L}|L)}{\partial \theta} \right)^T$$

$$\times \left(\frac{\partial p_{U(X(T_1)), \cdots, U(X(T_L))|N(t)}^\theta(r_{1:L}|L)}{\partial \theta} \right) dr_1 \cdots dr_L, \qquad (5.13)$$

where, for $r_{1:L} := (r_1, \cdots, r_L)$, $r_1, \cdots, r_L \in \mathbb{R}^2$,

$$p^\theta_{U(X(T_1)), \cdots, U(X(T_L))|N(t)}(r_1, \cdots, r_L|L) = \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \prod_{i=1}^L f_{x_i}(r_i)$$

$$\times \left(\int_{t_0}^t \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_2} p^\theta_{X(\tau_1), \cdots, X(\tau_L)}(x_1, \cdots, x_L) \prod_{j=1}^L \Lambda(\tau_j)d\tau_1 \cdots d\tau_{L-1}d\tau_L\right) dx_1 \cdots dx_L,$$

$$(5.14)$$

and $p^\theta_{X(\tau_1), \cdots, X(\tau_L)}$, $t_0 \le \tau_1 < \cdots < \tau_L \le t_0$, is the joint probability density function of $X(\tau_1), \cdots, X(\tau_L)$. By substituting Eq. (5.14) into Eq. (5.13), we have

$$I^f_{N(t)=L}(\theta) = \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \left\{ \int_{t_0}^t \int_{t_0}^{\tau'_L} \cdots \int_{t_0}^{\tau'_2} \int_{t_0}^t \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_2} \right.$$

$$\times \left(\int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \frac{\prod_{i=1}^L f_{x_i}(r_i) \prod_{j=1}^L f_{x'_j}(r_j)}{\int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \prod_{k=1}^L f_{x''_k}(r_k) \left(\int_{t_0}^t \int_{t_0}^{\tau''_L} \cdots \int_{t_0}^{\tau''_2} p^\theta_{X(\tau''_1), \cdots, X(\tau''_L)}(x''_1, \cdots, x''_L) \prod_{k=1}^L \Lambda(\tau''_k)d\tau''_1 \cdots d\tau''_{L-1}d\tau''_L \right) dx''_1 \cdots dx''_L} \right.$$

$$\times dr_1 \cdots dr_L \bigg) \left(dp^\theta_{X(\tau_1), \cdots, X(\tau_L)}(x_1, \cdots, x_L) \right)^T dp^\theta_{X(\tau'_1), \cdots, X(\tau'_L)}(x'_1, \cdots, x'_L) \prod_{i=1}^L \Lambda(\tau_i) \prod_{j=1}^L \Lambda(\tau'_j)d\tau_1 \cdots d\tau_{L-1}d\tau_L d\tau'_1 \cdots d\tau'_{L-1}d\tau'_L \bigg\}$$

$$\times dx_1 \cdots dx_L dx'_1 \cdots dx'_L$$

$$= \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \prod_{i=1}^L f_{x_i}(r_i) \prod_{j=1}^L f_{x'_j}(r_j) \bigg\{ \int_{t_0}^t \int_{t_0}^{\tau'_L} \cdots \int_{t_0}^{\tau'_2} \int_{t_0}^t \int_{t_0}^{\tau_L} \cdots \int_{t_0}^{\tau_2} \right.$$

$$\times \frac{\left(dp^\theta_{X(\tau_1), \cdots, X(\tau_L)}(x_1, \cdots, x_L) \right)^T dp^\theta_{X(\tau'_1), \cdots, X(\tau'_L)}(x'_1, \cdots, x'_L) \prod_{i=1}^L \Lambda(\tau_i) \prod_{j=1}^L \Lambda(\tau'_j)d\tau_1 \cdots d\tau_{L-1}d\tau_L d\tau'_1 \cdots d\tau'_{L-1}d\tau'_L \bigg\}}{\int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} \prod_{k=1}^L f_{x''_k}(r_k) \left(\int_{t_0}^t \int_{t_0}^{\tau''_L} \cdots \int_{t_0}^{\tau''_2} p^\theta_{X(\tau''_1), \cdots, X(\tau''_L)}(x''_1, \cdots, x''_L) \prod_{k=1}^L \Lambda(\tau''_k)d\tau''_1 \cdots d\tau''_{L-1}d\tau''_L \right) dx''_1 \cdots dx''_L}$$

$$\times dx_1 \cdots dx_L dx'_1 \cdots dx'_L \bigg) dr_1 \cdots dr_L,$$

$$(5.15)$$

where $dp^\theta_{X(\tau_1), \cdots, X(\tau_L)}(x_1, \cdots, x_L) := \frac{\partial p^\theta_{X(\tau_1), \cdots, X(\tau_L)}(x_1, \cdots, x_L)}{\partial \theta}$, $x_1, \cdots, x_L \in \mathbb{R}^2$, $t_0 \le \tau_1 < \cdots < \tau_L \le t$. $\square$

In the following lemma, for an object's motion modeled by a linear stochastic system, we calculate the derivatives $dp^\theta_{X(\tau_1), \cdots, X(\tau_L)}$ in the Fisher information matrix (Eq. (5.15)) derived in the above theorem.

Lemma 3. *Let the parameter space Θ describe an open subset of $\mathbb{R}^n$ containing the true parameters. For $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$, let*

$$X(\tau_{l+1}) = \phi^\theta(\tau_l, \tau_{l+1})X(\tau_l) + W^\theta(\tau_l, \tau_{l+1}), \quad \tau_0 := t_0 \le \tau_1 < \cdots < \tau_l, \quad l = 1, 2, \cdots,$$

where the 2D random variable $X(\tau_l)$ denotes the location of the object at time τ_l, $\phi^\theta(\tau_l, \tau_{l+1}) \in \mathbb{R}^{2\times2}$ is the state transition matrix, and $\{W^\theta(\tau_l, \tau_{l+1}), l = 1, 2, \cdots\}$ is a sequence of independent zero mean Gaussian random variables with covariance $Q^\theta(\tau_l, \tau_{l+1}) \in \mathbb{R}^{2\times2}, Q^\theta(\tau_l, \tau_{l+1}) > 0$. Also, assume that the initial location of the object is given by $X(\tau_0) = x_0 \in \mathbb{R}^2$. Then,

$$\frac{\partial p^\theta_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L)}{\partial \theta_i}$$

$$= p^\theta_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L) \frac{\partial \log p^\theta_{X(\tau_1),\cdots,X(\tau_L)}(x_1,\cdots,x_L)}{\partial \theta_i}$$

$$= p^\theta_{X(\tau_1)}(x_1) \prod_{l=2}^{L} p^\theta_{X(\tau_l)|X(\tau_{l-1})}(x_l|x_{l-1}) \left(\frac{\partial \log p^\theta_{X(\tau_1)}(x_1)}{\partial \theta_i} + \sum_{l=2}^{L} \frac{\partial \log p^\theta_{X(\tau_l)|X(\tau_{l-1})}(x_l|x_{l-1})}{\partial \theta_i}\right), \quad (5.16)$$

where

$$\frac{\partial p^\theta_{X(\tau_1)}(x_1)}{\partial \theta_i} = -\frac{1}{2} trace\left[\left((Q^\theta(\tau_0,\tau_1))^{-1} \frac{\partial Q^\theta(\tau_0,\tau_1)}{\partial \theta_i}\right)\left(I - (Q^\theta(\tau_0,\tau_1))^{-1} e^\theta_{\tau_0,\tau_1}(x_0,x_1)\left(e^\theta_{\tau_0,\tau_1}(x_0,x_1)\right)^T\right)\right]$$

$$- \frac{\partial\left(e^\theta_{\tau_0,\tau_1}(x_0,x_1)\right)^T}{\partial \theta_i}(Q^\theta(\tau_0,\tau_1))^{-1} e^\theta_{\tau_0,\tau_1}(x_0,x_1), \quad (5.17)$$

and

$$\frac{\partial \log p^\theta_{X(\tau_l)|X(\tau_{l-1})}(x_l|x_{l-1})}{\partial \theta_i}$$

$$= -\frac{1}{2} trace\left[\left((Q^\theta(\tau_{l-1},\tau_l))^{-1} \frac{\partial Q^\theta(\tau_{l-1},\tau_l)}{\partial \theta_i}\right)\left(I - (Q^\theta(\tau_{l-1},\tau_l))^{-1} e^\theta_{\tau_{l-1},\tau_l}(x_{l-1},x_l)\left(e^\theta_{\tau_{l-1},\tau_l}(x_{l-1},x_l)\right)^T\right)\right]$$

$$- \frac{\partial\left(e^\theta_{\tau_{l-1},\tau_l}(x_{l-1},x_l)\right)^T}{\partial \theta_i}(Q^\theta(\tau_{l-1},\tau_l))^{-1} e^\theta_{\tau_{l-1},\tau_l}(x_{l-1},x_l), \quad l = 2,\cdots,L, \quad (5.18)$$

where $e^\theta_{\tau_{l-1},\tau_l}(x_{l-1},x_l) := x_l - \phi^\theta(\tau_{l-1},\tau_l)x_{l-1}, l = 1,\cdots,L$, and I denotes the identity matrix with the corresponding size.

Proof. See the proofs of Theorem 3 and Corollary 4. $\qquad\square$

We next use the results obtained in the previous section to calculate the Fisher information matrix for the practical data model. We first, in the following definition, introduce a notation for the Fisher information matrix of the practical data model.

Definition 6. *Let the parameter space Θ describe an open subset of $\mathbb{R}^n$ containing the true parameters. We introduce the following notation for the Fisher information matrix of the practical*

data model, for a row parameter vector $\theta \in \Theta$,

$$
\begin{aligned}
I^p(\theta) :=\ & E_{Pr^\theta[S_1=z_1,\cdots,S_K=z_K]}\left\{\left(\frac{\partial \log Pr^\theta[S_1=z_1,\cdots,S_K=z_K]}{\partial \theta}\right)^T \left(\frac{\partial \log Pr^\theta[S_1=z_1,\cdots,S_K=z_K]}{\partial \theta}\right)\right\} \\
=\ & \sum_{z_1=0}^\infty \cdots \sum_{z_K=0}^\infty Pr^\theta[S_1=z_1,\cdots,S_K=z_K]\left(\frac{\partial \log Pr^\theta[S_1=z_1,\cdots,S_K=z_K]}{\partial \theta}\right)^T \\
& \times \left(\frac{\partial \log Pr^\theta[S_1=z_1,\cdots,S_K=z_K]}{\partial \theta}\right).
\end{aligned}
$$

(5.19)

In the following theorem, we calculate the Fisher information matrix of the practical data model introduced in the above definition.

Theorem 7. *1. For a row parameter vector $\theta \in \Theta$, the Fisher information matrix $I^{fp}(\theta)$ of the full practical data model can be calculated as*

$$
\begin{aligned}
I^{fp}(\theta) =\ & \sum_{z_1=0}^\infty \cdots \sum_{z_K=0}^\infty \frac{L!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^L} \\
& \times \frac{\sum_{v:=(v_1,\cdots,v_L)\in A_K^L(z_{1:K})}\sum_{v':=(v_1',\cdots,v_L')\in A_K^L(z_{1:K})}\int_{C_{v_1}}\cdots\int_{C_{v_L}}\int_{C_{v_1'}}\cdots\int_{C_{v_L'}}I_\theta^f\left(r_{1:L},r_{1:L}'\right)dr_L'\cdots dr_1'dr_L\cdots dr_1}{\sum_{v'':=(v_1'',\cdots,v_L'')\in A_K^L(z_{1:K})}\int_{C_{v_1''}}\cdots\int_{C_{v_L''}}p_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}^\theta\left(r_1'',\cdots,r_L''|L\right)dr_L''\cdots dr_1''},
\end{aligned}
$$

(5.20)

where $z_{1:K} := (z_1,\cdots,z_K)$, $L = \sum_{k=1}^K z_k$, $z_1,\cdots,z_K = 0,1,\cdots$, and $p_{U(X(T_1)),\cdots,U(X(T_L))|N(t)}^\theta$, I_θ^f are given by Eq. (5.12) of Theorem 6.

2. The Fisher information matrix $I^p(\theta)$ of the practical data model can be calculated as

$$
\begin{aligned}
I^p(\theta) =\ & \sum_{z_1=0}^\infty \cdots \sum_{z_K=0}^\infty \sum_{z=0}^\infty \sum_{z'=0}^\infty \frac{(L+z)!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^{L+z}}\frac{(L+z')!}{\left(\int_{t_0}^t \Lambda(\psi)d\psi\right)^{L+z'}} \\
& \times \frac{\sum_{v:=(v_1,\cdots,v_{L+z})\in A_{K+1}^{L+z}(z_{1:K},z)}\sum_{v':=(v_1',\cdots,v_{L+z'}')\in A_{K+1}^{L+z'}(z_{1:K},z')}\int_{C_{v_1}}\cdots\int_{C_{v_{L+z}}}\int_{C_{v_1'}}\cdots\int_{C_{v_{L+z'}'}}I_\theta^f\left(r_{1:L+z},r_{1:L+z'}'\right)}{\sum_{z''=0}^\infty \sum_{v'':=(v_{1:L+z''}'')\in A_{K+1}^{L+z''}(z_{1:K},z'')}\int_{C_{v_1''}}\cdots\int_{C_{v_{L+z''}''}}p_{U(X(T_1)),\cdots,U(X(T_{L+z''}))|N(t)}^\theta\left(r_{1:L+z''}''|L+z''\right)dr_{L+z''}''\cdots dr_1''} \\
& \times dr_{L+z'}'\cdots dr_1'dr_{L+z}\cdots dr_1,
\end{aligned}
$$

(5.21)

where $r_{1:L+z''}'' := \left(r_1'',\cdots,r_{L+z''}''\right)$, $r_1'',\cdots,r_{L+z''}'' \in \mathbb{R}^2$, $z'' = 0,1,\cdots$.

Proof. It results by substituting Eqs. (5.3) and (5.4) of Theorem 5 into Eq. (5.19) of Definition 6, and using the same procedure as the proof of Theorem 6. $\qquad\square$

6. SUMMARY AND CONCLUSIONS

Single molecule microscopy is a powerful tool to study the dynamics of individual molecules within living cells, which previously were masked through bulk microscopy due to the averaging over a group of closely spaced molecules. In this thesis, we have developed parameter estimation methods to analyze single molecule microscopy data. In particular, we have focused on two important applications of single molecule microscopy: single molecule super-resolution microscopy and single molecule tracking.

In order to solve the localization problem for single molecule super-resolution images, we have developed a non-fitting state space algorithm. In single molecule fluorescence images, the locations of peaks of intensity correspond to the locations of molecules. Our proposed algorithm models such an image by the frequency response of a multi-order system obtained using a state space realization algorithm based on the singular value decomposition of a Hankel matrix. The locations of the poles of such a system determine the peak locations in the frequency domain, and therefore correspond to the locations of molecules. To assess the performance of the proposed localization algorithm, we applied the algorithm to both simulated and experimental data comprising images of closely spaced molecules. In the case of simulated data, we evaluated the detection rate of the algorithm for molecules with different mean photon counts and different distances between the molecules. We also analyzed the bias of the algorithm. The bias was evaluated as the average of the deviations of the estimated molecule locations from the ground truth. In the case where there was only one molecule per image, our results suggested that there was no systematic bias associated with the algorithm. In the case of data sets consisting of repeat images of multiple molecules, however, the results showed that bias existed which we found to be dependent on the distances between the molecules relative to the image size. Also, the accuracy of the algorithm, determined by how far the estimates were spread out from the ground truth, was assessed by looking at the square root of the average of the squared deviations from the ground truth. In the case that we had repeat images of the same molecules, we looked at the squared deviations of the estimates from

their average, i.e., the accuracy was given by the standard deviation of the estimates. Importantly, for data sets comprising repeat images of one molecule, the standard deviation of the estimates was compared with the limit of the localization accuracy, a theoretical accuracy benchmark given by the square root of the CRLB. The results showed that the accuracy of the algorithm was reasonable, but the difference between the accuracy and the limit of accuracy was nevertheless around twice the limit of accuracy. We showed, however, that by using the obtained location estimates as the initial conditions for a maximum likelihood estimator, we can decrease the standard deviation of the estimates and approach the limit of accuracy, as is usually possible with the maximum likelihood estimator for standard single molecule estimation problems. We further demonstrated with experimental data that the algorithm could recover the locations of the significant peaks in the original image that correspond to the locations of individual Alexa Fluor 647 dye molecules.

To analyze single molecule tracking data, we have proposed a maximum likelihood estimation method to estimate the parameters of the molecule trajectories in cellular environments. In our method, we have considered a dynamical system where the motion of an object of interest, e.g., single molecule, is modeled generally by stochastic differential equations, and the measurements are the time points and locations of the detected photons emitted by the moving fluorescently labeled object using an ideal pixelated detector, referred to as the fundamental data model. More importantly, we have calculated the CRLB-based limit of accuracy, given by the inverse of the Fisher information matrix, for the estimation of the unknown parameters. To examine the performance of the proposed estimation method, we applied it to simulated data sets comprising linear two-dimensional trajectories of a molecule with Gaussian, Airy and classical model of Born and Wolf measurements. The results showed that there was no systematic bias associated with the method. In addition, we showed that the means of the distributions of the prediction of the molecule locations were able to follow the true locations of the molecule for all different types of measurements. In particular, for data sets comprising repeat trajectories of a molecule with Gaussian measurements, it was shown that the standard deviations of the diffusion and drift estimates were close to the square roots of their corresponding CRLBs. We also showed that, in the case

that we had one detected photon, the Fisher information matrices obtained for an Airy and its corresponding approximating Gaussian profile were different from each other, and therefore, the use of the Gaussian approximation can be problematic in some applications. We showed that equally sampled time points, which have been commonly used in most dynamical systems, and Poisson distributed time points can lead to significantly different Fisher information matrices. We further showed that even the results obtained for different realizations of a Poisson process could vary notably.

In the fundamental data model, we assumed that the time points and locations of the photons, which are detected by an ideal unpixelated detector, are available. However, in practice, pixelated detectors, such as charge-coupled device (CCD) and electron multiplying CCD (EMCCD) cameras, are commonly used for acquiring the image of the object. In this case, referred to as the practical data model, the measurements, i.e., the fluorescence microscopy images, include the numbers of the photons detected in each pixel. Therefore, the temporal information of the detection of the photons is not available anymore. Moreover, instead of having the exact locations of detection of the photons, we only know the pixel areas in which the photons impact the detector. These challenges make the analysis of single molecule trajectories from pixelated images a complex problem. We addressed these challenges by extending our framework to the practical data model. In cases that we had a small number of photons, the results showed that we were able to estimate the parameters of the molecule trajectory from simulated fluorescence microscopy images using our proposed method.

REFERENCES

[1] S. Ram, *Resolution and localization in single molecule microscopy*. Ph. D. thesis, University of Texas at Arlington/University of Texas Southwestern Medical Center at Dallas, 2007.

[2] W. E. Moerner and L. Kador, "Optical detection and spectroscopy of single molecules in a solid," *Physical review letters*, vol. 62, no. 21, pp. 2535–2538, 1989.

[3] G. Seisenberger, M. U. Ried, T. Endress, H. Büning, M. Hallek, and C. Bräuchle, "Real-time single-molecule imaging of the infection pathway of an adeno-associated virus," *Science*, vol. 294, no. 5548, pp. 1929–1932, 2001.

[4] W. E. Moerner, "New directions in single-molecule imaging and analysis," *Proceedings of the National Academy of Sciences*, vol. 104, no. 31, pp. 12596–12602, 2007.

[5] N. G. Walter, C. Y. Huang, A. J. Manzo, and M. A. Sobhy, "Do-it-yourself guide: how to use the modern single-molecule toolkit," *Nature methods*, vol. 5, no. 6, pp. 475–489, 2008.

[6] R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophysical Journal*, vol. 8, no. 2, pp. 1185–1200, 2004.

[7] S. Ram, E. S. Ward, and R. J. Ober, "A stochastic analysis of performance limits for optical microscopes," *Multidimensional Systems and Signal Processing*, vol. 17, no. 1, pp. 27–57, 2006.

[8] S. Ram, D. Kim, R. J. Ober, and E. S. Ward, "3D single molecule tracking with multifocal plane microscopy reveals rapid intercellular transferrin transport at epithelial cell barriers," *Biophysical Journal*, vol. 103, no. 7, pp. 1594–603, 2012.

[9] P. Prabhat, S. Ram, E. S. Ward, and R. J. Ober, "Simultaneous imaging of different focal planes in fluorescence microscopy for the study of cellular dynamics in three dimensions," *IEEE Transactions on NanoBioscience*, vol. 3, no. 4, pp. 237–242, 2004.

[10] A. Small and S. Stahlheber, "Fluorophore localization algorithms for super-resolution microscopy," *Nature Methods*, vol. 11, pp. 267–279, 2014.

[11] F. Huang, S. L. Schwartz, J. M. Byars, and K. A. Lidke, "Simultaneous multiple-emitter fitting for single molecule super-resolution imaging," *Biomedical Optics Express*, vol. 2, no. 5, pp. 1377–1393, 2011.

[12] S. J. Holden, S. Uphoff, and A. N. Kapanidis, "DAOSTORM: an algorithm for high-density super-resolution microscopy," *Nature Methods*, vol. 8, pp. 279–280, 2011.

[13] T. A. Laurence and B. A. Chromy, "Efficient maximum likelihood estimator fitting of histograms," *Nature Methods*, vol. 7, p. 338âĂŞ339, 2010.

[14] R. Henriques, M. Lelek, E. F. Fornasiero, F. Valtorta, C. Zimmer, and M. M. Mhlanga, "QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ," *Nature Methods*, vol. 7, pp. 339–340, 2010.

[15] A. J. Berglund, M. D. McMahon, J. J. McClelland, and J. A. Liddle, "Fast, bias-free algorithm for tracking single particles with variable size and shape," *Optics Express*, vol. 16, no. 18, pp. 14064–14075, 2008.

[16] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[17] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," *Nature Methods*, vol. 9, pp. 721–723, 2012.

[18] T. Quan, H. Zhu, X. Liu, Y. Liu, J. Ding, S. Zeng, and Z.-L. Huang, "High-density localization of active molecules using Structured Sparse Model and Bayesian Information Criterion," *Optics Express*, vol. 19, no. 18, pp. 16963–16974, 2011.

[19] J. Min, C. Vonesch, H. Kirshner, L. Carlini, N. Olivier, S. Holden, S. Manley, J. C. Ye, and M. Unser, "FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data," *Scientific Reports*, vol. 4, p. 4577, 2014.

[20] J. Huang, K. Gumpper, Y. Chi, M. Sun, and J. Ma, "Fast two-dimensional super-resolution image reconstruction algorithm for ultra-high emitter density," *Optics Letters*, vol. 40, no. 13, pp. 2989–2992, 2015.

[21] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[22] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2267–2280, 1992.

[23] M. J. Saxton and K. Jacobson, "Single-particle tracking: applications to membrane dynamics," *Annual Review of Biophysics and Biomolecular Structure*, vol. 26, no. 1, pp. 373–399, 1997.

[24] S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, "High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells," *Biophysical journal*, vol. 95, no. 12, pp. 6025–6043, 2008.

[25] A. Kusumi, T. A. Tsunoyama, K. M. Hirosawa, R. S. Kasai, and T. K. Fujiwara, "Tracking single molecules at work in living cells," *Nature chemical biology*, vol. 10, no. 7, pp. 524–532, 2014.

[26] C. Manzo and M. F. Garcia-Parajo, "A review of progress in single particle tracking: from methods to biophysical insights," *Reports on Progress in Physics*, vol. 78, no. 12, p. 124601, 2015.

[27] W. C. Olson and R. J. Israel, "Antibody-drug conjugates targeting prostate-specific membrane antigen," *Frontiers in Bioscience*, vol. 19, pp. 12–33, 2014.

[28] Y. Gao, Y. Li, Y. Li, L. Yuan, Y. Zhou, J. Li, L. Zhao, C. Zhang, X. Li, and Y. Liu, "PSMA-mediated endosome escape-accelerating polymeric micelles for targeted therapy of prostate cancer and the real time tracing of their intracellular trafficking," *Nanoscale*, vol. 7, no. 2, pp. 597–612, 2015.

[29] Z. Schuss, *Theory and applications of stochastic processes: an analytical approach*. Springer New York, 2009.

[30] V. Briane, C. Kervrann, and M. Vimond, "Statistical analysis of particle trajectories in living cells," *Physical Review E*, vol. 97, no. 6, p. 062121, 2018.

[31] H. Risken and T. Frank, *The Fokker-Planck equation: methods of solution and applications*. Springer-Verlag Berlin Heidelberg, 1996.

[32] C. P. Calderon, M. A. Thompson, J. M. Casolari, R. C. Paffenroth, and W. E. Moerner, "Quantifying transient 3D dynamical phenomena of single mRNA particles in live yeast cell measurements," *The Journal of Physical Chemistry B*, vol. 117, no. 49, pp. 15701–15713, 2013.

[33] 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*, vol. 10, no. 9, p. e0137633, 2015.

[34] C. P. Calderon, "Motion blur filtering: A statistical approach for extracting confinement forces and diffusivity from a single blurred trajectory," *Physical Review E*, vol. 93, no. 5, p. 053303, 2016.

[35] J. Chao, E. S. Ward, and R. J. Ober, "Fisher information theory for parameter estimation in single molecule microscopy: tutorial," *Journal of the Optical Society of America A*, vol. 33, no. 7, pp. B36–B57, 2016.

[36] M. Born and E. Wolf, *Principles of Optics*. Cambridge University Press, 1999.

[37] A. V. Abraham, S. Ram, J. Chao, E. S. Ward, and R. J. Ober, "Quantitative study of single molecule location estimation techniques," *Optics Express*, vol. 17, no. 26, pp. 23352–23373, 2009.

[38] H. Qian, M. P. Sheetz, and E. L. Elson, "Single particle tracking. Analysis of diffusion and flow in two-dimensional systems," *Biophysical journal*, vol. 60, no. 4, pp. 910–921, 1991.

[39] A. J. Berglund, "Statistics of camera-based single-particle tracking," *Physical Review E*, vol. 82, no. 1, p. 011917, 2010.

[40] X. Michalet and A. J. Berglund, "Optimal diffusion coefficient estimation in single-particle tracking," *Physical Review E*, vol. 85, no. 6, p. 061916, 2012.

[41] X. Michalet, "Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium," *Physical Review E*, vol. 82, no. 4, p. 041914, 2010.

[42] 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*, vol. 93, no. 4, p. 042401, 2016.

[43] T. T. Ashley and S. B. Andersson, "Method for simultaneous localization and parameter estimation in particle tracking experiments," *Physical Review E*, vol. 92, no. 5, p. 052707, 2015.

[44] P. C. Nelson, C. Zurla, D. Brogioli, J. F. Beausang, L. Finzi, and D. Dunlap, "Tethered particle motion as a diagnostic of DNA tether length," *The Journal of Physical Chemistry B*, vol. 110, no. 34, pp. 17260–17267, 2006.

[45] T. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, 2011.

[46] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin, "On particle methods for parameter estimation in state-space models," *Statistical Science*, vol. 30, no. 3, pp. 328–351, 2015.

[47] L. Jiang, S. S. Singh, and S. Yildirim, "Bayesian tracking and parameter learning for nonlinear multiple target tracking models," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5733–5745, 2015.

[48] R. J. Ober, X. Lai, Z. Lin, and E. S. Ward, "A state space approach to noise reduction of 3D fluorescent microscopy images," *Proceeding IEEE International Conference on Image Processing (ICIP '04)*, vol. 2, pp. 1153–1156, 2006.

[49] X. Lai, E. S. Ward, Z. Lin, and R. J. Ober, "Three-dimensional state space realization algorithm: noise suppression of fluorescence microscopy images and point spread functions," *Proc. SPIE*, vol. 5701, 2005.

[50] R. J. Ober, X. Lai, Z. Lin, and E. S. Ward, "State space realization of a three-dimensional image set with application to noise reduction of fluorescent microscopy images of cells," *Multidimensional Systems and Signal Processing*, vol. 16, no. 1, pp. 7–47, 2005.

[51] T. McKelvey, H. Akcay, and L. Ljung, "Subspace-based multivariable system identification from frequency response data," *IEEE Transactions on Automatic Control*, vol. 41, no. 7, pp. 960–979, 1996.

[52] J. M. Maciejowski, "Guaranteed stability with subspace methods," *Systems & Control Letters*, vol. 26, no. 2, pp. 153–156, 1995.

[53] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser, "Quantitative evaluation of software packages for single-molecule localization microscopy," *Nature Methods*, vol. 12, pp. 717–724, 2015.

[54] M. R. Vahid, J. Chao, D. Kim, E. S. Ward, and R. J. Ober, "State space approach to single molecule localization in fluorescence microscopy," *Biomedical Optics Express*, vol. 8, no. 3, pp. 1332–1355, 2017.

[55] D. L. Snyder and M. I. Miller, *Random point processes in time and space*. Springer-Verlag New York, 1991.

[56] R. L. Streit, *Poisson Point Processes*. Springer US, 2010.

[57] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*. Springer-Verlag Berlin, 1992.

[58] A. M. Basharov, "Derivation of kinetic equations from non-Wiener stochastic differential equations," *Journal of Physics: Conference Series*, vol. 478, no. 1, p. 012011, 2013.

[59] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[60] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1246–1250, 1998.

[61] 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*, vol. 59, no. 3, pp. 895–911, 2011.

[62] M. R. Vahid, J. Chao, E. S. Ward, and R. J. Ober, "A state space based approach to localizing single molecules from multi-emitter images," *Proc. SPIE*, p. 10070: 100700J.

[63] A. C. Harvey, *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, 1990.

[64] A. Tahmasbi, E. S. Ward, and R. J. Ober, "Determination of localization accuracy based on experimentally acquired image sets: applications to single molecule microscopy," *Optics Express*, vol. 23, no. 6, pp. 7630–7652, 2015.

[65] S. A. Douglass, *Introduction to mathematical analysis*. Addison-Wesley Publishing Company, 1996.

[66] R. L. M. Peeters and B. Hanzon, "Symbolic computation of Fisher information matrices for parametrized state-space systemss," *Automatica*, vol. 35, no. 6, pp. 1059–1071, 1999.

APPENDIX A

FREQUENCY RESPONSE OF A MULTI-ORDER SYSTEM AS THE PSF MODEL

In this section, we present the results of the proposed algorithm applied to images simulated using the frequency response of a multi-order system. In this case, instead of Eq. (2.36), $\mu_{\hat{\theta}^h}$ is given as

$$\mu_{\hat{\theta}^h}(k_1, k_2) := \frac{1}{C} \left| \sum_{n=1}^{h} \frac{N_{p,n}}{(e^{i2\pi k_1/N_1} - \bar{a}_n^1)(e^{i2\pi k_2/N_2} - \bar{a}_n^2)} \right|, \quad k_i = 1, ..., N_i, \quad i = 1, 2, \quad \text{(A.1)}$$

where $C := \sum_{k_1=1}^{N_1} \sum_{k_2=1}^{N_2} \left| \sum_{n=1}^{h} \frac{1}{(e^{i2\pi k_1/N_1} - \bar{a}_n^1)(e^{i2\pi k_2/N_2} - \bar{a}_n^2)} \right|$ is the normalization factor. Here, to analyze the performance of the algorithm, we simulated data sets containing repeat images of one molecule using the frequency response of a first-order system, i.e., using Eq. (A.1) with h=1. The data sets differ by the mean photon count $N_{p,1}$ for the molecule. For each mean photon count, the data set consists of 1000 repeat images of size 20×20 pixels. In Figs. A.1(a) and A.1(b), an example of an image with a mean photon count of $N_{p,1} = 1000$ is shown. To assess the bias of the algorithm, we calculated the differences between the averages of the x- and y-estimates and the corresponding true x- and y-coordinates. Similar to the case of images simulated with the Airy profile [Fig. 2.3], the evenness of the spread of the estimated bias about 0 nm for both coordinates [Fig. A.1(c)] demonstrates that there is no systematic bias associated with our proposed algorithm when there is only one molecule per image.

Also, we calculated the standard deviation of the estimates for nine of the data sets and compared the results with the limit of the localization accuracy calculated using the approach for experimental PSFs presented in [64]. It can be seen in Fig. A.2 that when the image of the molecule is simulated as the frequency response of a first-order system, the accuracy of the algorithm comes close to the limit of the localization accuracy.
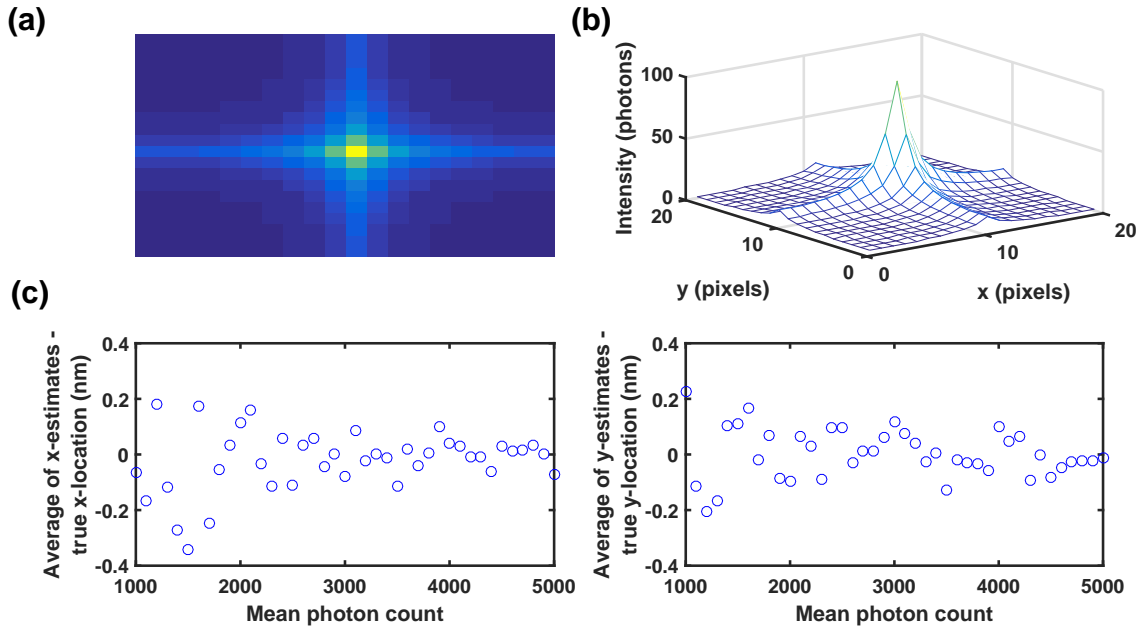
(a)
(b)
(c)

Figure A.1: Analysis of the bias of location estimates obtained from repeat images containing exactly one molecule, simulated using the frequency response of a first-order system. (a) Image of a point source simulated using the frequency response of a first-order system, i.e., using Eq. (A.1) with $h = 1$, and a mean photon count of $N_{p,1} = 1000$. (b) Mesh view of the image shown in (a). (c) Difference between the average of the x-estimates and the true x-value, and the difference between the average of the y-estimates and the true y-value for data sets that differ by the mean photon count per image assumed for the molecule. For each mean photon count, the data set consists of 1000 repeat images of size 20×20 pixels, simulated using the frequency response of a first-order system.

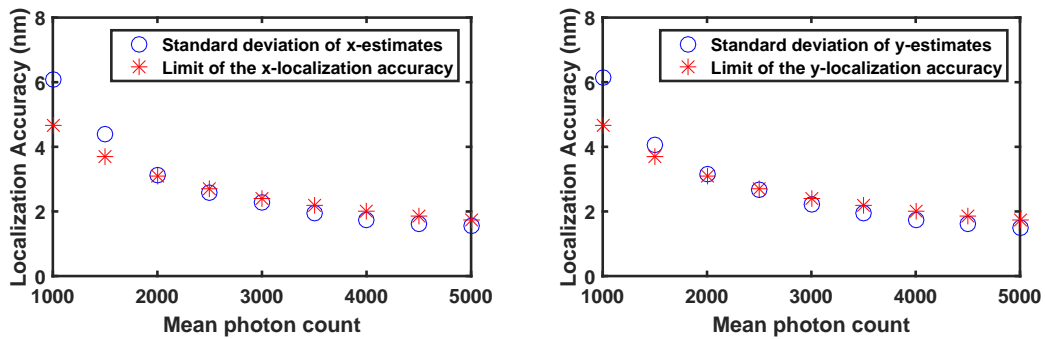Figure A.2: Analysis of the standard deviation of location estimates obtained from repeat images containing exactly one molecule, simulated using the frequency response of a first-order system. Shown in the left and right plots are the standard deviations of the x- and y-estimates and the limits of the x- and y-localization accuracy, respectively, for nine of the data sets from Fig. A.1.

127

APPENDIX B

COMPUTATION OF MAXIMUM LIKELIHOOD ESTIMATES AND

FISHER INFORMATION MATRIX

B.1 Example of maximum likelihood estimation

Let the 2D motion of an object be given by the following continuous-time stochastic differential equation

$$dX(\tau) = FI_{2\times2}X(\tau)d\tau + \sqrt{2D}I_{2\times2}dB(\tau), \qquad \tau \geq t_0, \tag{B.1}$$

where $I_{2\times2}$ denotes the 2×2 identity matrix, $F \in \mathbb{R}$ and $D > 0$ denote the first order drift and diffusion coefficients, respectively, and $\{B(\tau) \in \mathbb{R}^2, \tau \geq t_0\}$ is a 2-vector Brownian motion process with $E\{dB(\tau)dB(\tau)^T\} = I_{2\times2}$. Also, let $X(t_0)$ be Gaussian distributed with mean $x_0 \in \mathbb{R}^2$ and diagonal covariance matrix $P_0 := \rho_0 I_{2\times2}, \rho_0 > 0$, which is assumed to be independent of $B(\tau)$. Assume that the photon detection rate Λ, the magnification matrix $M = mI_{2\times2}, m > 0$, and the covariance matrix $\Sigma_g = vI_{2\times2}, v > 0$, of the measurement noise are independent of the parameter vector $\theta \in \Theta$. Also, for the corresponding discrete system at time points $\tau_0 := t_0 \leq \tau_1 < \tau_2 < \cdots < \tau_K < \cdots$, let the transition matrix be given by $\phi(\tau_{l-1}, \tau_l) := \phi^s(\tau_{l-1}, \tau_l)I_{2\times2}, \phi^s(\tau_{l-1}, \tau_l) \in \mathbb{R}$, and the process noise covariance matrix be given by $Q_g(\tau_{l-1}, \tau_l) := q^s(\tau_{l-1}, \tau_l)I_{2\times2}, q^s(\tau_{l-1}, \tau_l) > 0$. Then, the covariances of the states, which can be calculated through the Kalman filter formulae recursively, are also scalar matrices, i.e., can be defined as $P_{\theta,l}^{l-1} := \rho_{\theta,l}^{l-1}I_{2\times2}, \rho_{\theta,l}^{l-1} > 0, l = 1, 2, \cdots$. Also, let $\hat{x}_{\theta,l}^{l-1}$ denote the means of the states.

Then, the maximum likelihood estimate $\hat{\theta}_{mle}$ of $\theta = (\theta_1, \cdots, \theta_n)$ is the solution of the following equation, according to Eq. (4.38), for the acquired data denoted by $d_K \in \mathcal{C}^K \times \mathbb{R}_{[\infty]}^K, K =$

$1, 2, \cdots,$

$$\frac{\partial \log \mathcal{L}\left(\theta | d_K\right)}{\partial \theta_i}$$

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

$$- \frac{m d\hat{x}_{\theta,l}^{l-1,i} \left(r_l - m\hat{x}_{\theta,l}^{l-1} \right)}{m^2 \rho_{\theta,l}^{l-1} + v}$$

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

where $i = 1, \cdots, n$, $\|.\|$ denotes the Euclidean norm, and for $l = 1, 2, \cdots, \hat{X}_{\theta,l}^{(i)} := \begin{bmatrix} \hat{x}_{\theta,l}^{l-1} & d\hat{x}_{\theta,l}^{l-1,i} \end{bmatrix}^T$, $d\hat{x}_{\theta,l}^{l-1,i} := \frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_i}$, $P_{\theta,l}^{(i)} := \begin{bmatrix} \rho_{\theta,l}^{l-1} & d\rho_{\theta,l}^{l-1,i} \end{bmatrix}^T$, and $d\rho_{\theta,l}^{l-1,i} := \frac{\partial \rho_{\theta,l}^{l-1}}{\partial \theta_i}$ can be calculated through the following recursive formulas, by combining the Kalman filtering equations (Eqs. (3.28) and (3.29)) and their derivatives, and using Lemma 5 (see Section B.7),

$$\hat{X}_{\theta,l+1}^{(i)} = A_{\theta,l+1}^{(i)} \hat{X}_{\theta,l}^{(i)} + B_{\theta,l+1}^{(i)} \left(r_l - m\hat{x}_{\theta,l}^{l-1} \right),$$

$$P_{\theta,l+1}^{(i)} = C_{\theta,l+1}^{(i)} P_{\theta,l}^{(i)} + \begin{bmatrix} q_\theta^s \left(\tau_l, \tau_{l+1} \right) \\ \frac{\partial q_\theta^s (\tau_l, \tau_{l+1})}{\partial \theta_i} \end{bmatrix}, \qquad \text{(B.3)}$$

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

$$A_{\theta,l+1}^{(i)} := \begin{bmatrix} \phi_\theta(\tau_l, \tau_{l+1}) & 0_{2 \times 2} \\ \frac{\partial \phi_\theta(\tau_l, \tau_{l+1})}{\partial \theta_i} & \phi_\theta(\tau_l, \tau_{l+1}) \left(I_{2 \times 2} - K_{\theta,l} M \right) \end{bmatrix},$$

$$B_{\theta,l+1}^{(i)} := \begin{bmatrix} \phi_\theta(\tau_l, \tau_{l+1}) K_{\theta,l} \\ \phi_\theta(\tau_l, \tau_{l+1}) \frac{\partial K_{\theta,l}}{\partial \theta_i} + \frac{\partial \phi_\theta(\tau_l, \tau_{l+1})}{\partial \theta_i} K_{\theta,l} \end{bmatrix},$$

$$C_{\theta,l+1}^{(i)} := \begin{bmatrix} \left(1 - m k_{\theta,l}^s \right) \left(\phi_\theta^s(\tau_l, \tau_{l+1}) \right)^2 & 0 \\ \left(1 - m k_{\theta,l}^s \right) \frac{\partial (\phi_\theta^s(\tau_l, \tau_{l+1}))^2}{\partial \theta_i} - m \frac{\partial k_{\theta,l}^s}{\partial \theta_i} \left(\phi_\theta^s(\tau_l, \tau_{l+1}) \right)^2 & \left(1 - m k_{\theta,l}^s \right) \left(\phi_\theta^s(\tau_l, \tau_{l+1}) \right)^2 \end{bmatrix}, \qquad \text{(B.4)}$$

where the Kalman gain and its derivative are given by

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

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

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

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

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

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

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

and

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

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

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

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

130

and

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

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

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

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

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

and

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

B.2 Sequential Monte Carlo method

Here, for the acquired data denoted by $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[\infty]}$ (or $d_l \in \mathcal{C}^l \times \mathbb{R}^l_{[t]}$), $l = 1, 2, \cdots$, we approximate the distribution $p_{pr_{l+1}}\left(x_{l+1}|\tau_{l+1}, d_l\right)$ through the sequential Monte Carlo method provided in [45]. Note that

$$
\begin{aligned}
p_{pr_{l+1}}\left(x_{l+1}|\tau_{l+1}, d_l\right) &= p_{X(T_{l+1})|T_{l+1},\mathcal{D}_l}\left(x_{l+1}|\tau_{l+1}, d_l\right) \\
&= \int_{\mathbb{R}^3} p_{X(T_{l+1}),X(T_l)|T_{l+1},\mathcal{D}_l}\left(x_{l+1}, x|\tau_{l+1}, d_l\right)dx \\
&= \int_{\mathbb{R}^3} p_{X(T_{l+1})|X(T_l),T_{l+1},\mathcal{D}_l}\left(x_{l+1}|x, \tau_{l+1}, d_l\right)p_{X(T_l)|T_{l+1},\mathcal{D}_l}\left(x|\tau_{l+1}, d_l\right)dx \\
&= \int_{\mathbb{R}^3} p_{X(T_{l+1})|X(T_l),T_{l+1},\mathcal{D}_l}\left(x_{l+1}|x, \tau_{l+1}, d_l\right)p_{fi_l}\left(x|d_l\right)dx, \tag{B.9}
\end{aligned}
$$

where $p_{fi_l}(x|d_l) := p_{X(T_l)|\mathcal{D}_l}(x|d_l), x \in \mathbb{R}^3$, and for the linear stochastic system with state $X(\tau) \in \mathbb{R}^3, \tau \geq t_0$, zero-mean Gaussian process noise with covariance matrix $Q_g(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3\times3}, Q_g(\tau_l, \tau_{l+1}) > 0$, and state-transition matrix $\phi(\tau_l, \tau_{l+1}) \in \mathbb{R}^{3\times3}$, we have, for $x_l \in \mathbb{R}^3$,

$$
\begin{aligned}
& p_{X(T_{l+1})|X(T_l),T_{l+1},\mathcal{D}_l}\left(x_{l+1}|x_l, \tau_{l+1}, d_l\right) \\
&= p_{X(T_{l+1})|X(T_l)}\left(x_{l+1}|x_l\right) \\
&= \frac{1}{2\pi\left[\det\left(Q_g(\tau_l, \tau_{l+1})\right)\right]^{1/2}} \\
&\quad \times \exp\left(-\frac{1}{2}(x_{l+1} - \phi(\tau_l, \tau_{l+1})x_l)^T Q_g^{-1}(\tau_l, \tau_{l+1})(x_{l+1} - \phi(\tau_l, \tau_{l+1})x_l)\right).
\end{aligned}
\tag{B.10}
$$

The distribution p_{fi_l} of the filtered object location can be approximated as [45]

$$
p_{fi_l}\left(x_l|d_l\right) \approx \sum_{i=1}^{N} w_l^i(r_l)\delta\left(x_l - \dot{x}_l^i\right),
\tag{B.11}
$$

where δ is the Dirac delta function, and the samples $\dot{x}_l^i$ and their corresponding weights $w_l^i(r_l), i = 1, \cdots, N$, are given through the following sequential Monte Carlo algorithm. Finally, by substituting Eqs. (B.10) and (B.11) into Eq. (B.9), the distribution $p_{pr_{l+1}}$ can be approximated as

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

Sequential Monte Carlo (particle filter) algorithm: [45]

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

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

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

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

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

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

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

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

B.3 Estimation results for Airy measurements

Here, we analyze the error of the diffusion and drift coefficient estimates for simulated data sets with the Airy measurement profile, with the same standard deviation as the Born and Wolf and Gaussian data presented in Figs. 3.3-3.6, and obtain similar results (see Figs. B.1 and B.2). We also show the differences between the means of the distributions of the prediction of the molecule locations and the true locations of the molecule in Fig. B.4 (see Section B.11).

B.4 Example for Fisher information calculation

For the data model described in the example provided in Section B.1, the Fisher information matrix is given by Eqs. (4.31) and (4.32), where $S_{\theta,l}, l = 1, 2, \cdots$, is given recursively by, for $\theta = (\theta_1, \cdots, \theta_n) \in \Theta$ and $i, j = 1, \cdots, n$,

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

and

$$S_{\theta,1}^{(ji)} = \begin{bmatrix} \phi_\theta(\tau_0, \tau_1)x_{\theta,0} \\ \frac{\partial\left(\phi_\theta(\tau_0,\tau_1)x_{\theta,0}\right)}{\partial\theta_j} \end{bmatrix} \begin{bmatrix} \left(\phi_\theta(\tau_0, \tau_1)x_{\theta,0}\right)^T & \left(\frac{\partial\left(\phi_\theta(\tau_0,\tau_1)x_{\theta,0}\right)}{\partial\theta_i}\right)^T \end{bmatrix},$$
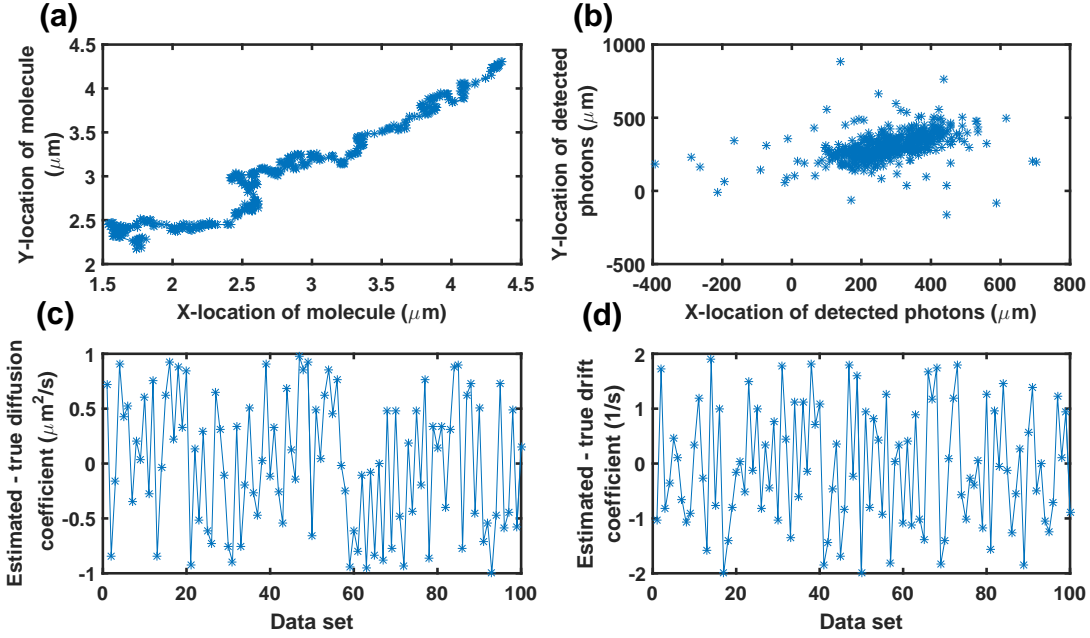
Figure B.1: Analysis of the error of diffusion coefficient and drift coefficient estimates produced by the maximum likelihood estimation method for the Airy measurement model. (a) The two-dimensional single molecule trajectory simulated in Fig. 3.3(a). (b) Detected locations of the photons emitted from the molecule trajectory of part (a) in the image space which are simulated using Eq. (3.23) with the Airy profile (Eq. (3.11)) and $\alpha := \frac{2\pi n_a}{\lambda} = 2.59$. (c) Differences between the diffusion coefficient estimates and the true diffusion coefficient value for 100 data sets, each containing a trajectory of a molecule simulated using Eqs. (B.1) and (3.23) with the Airy profile, and the parameters given in parts (a) and (b). (d) Differences between the first order drift coefficient estimates and the true first order drift coefficient value for the data sets of part (c).
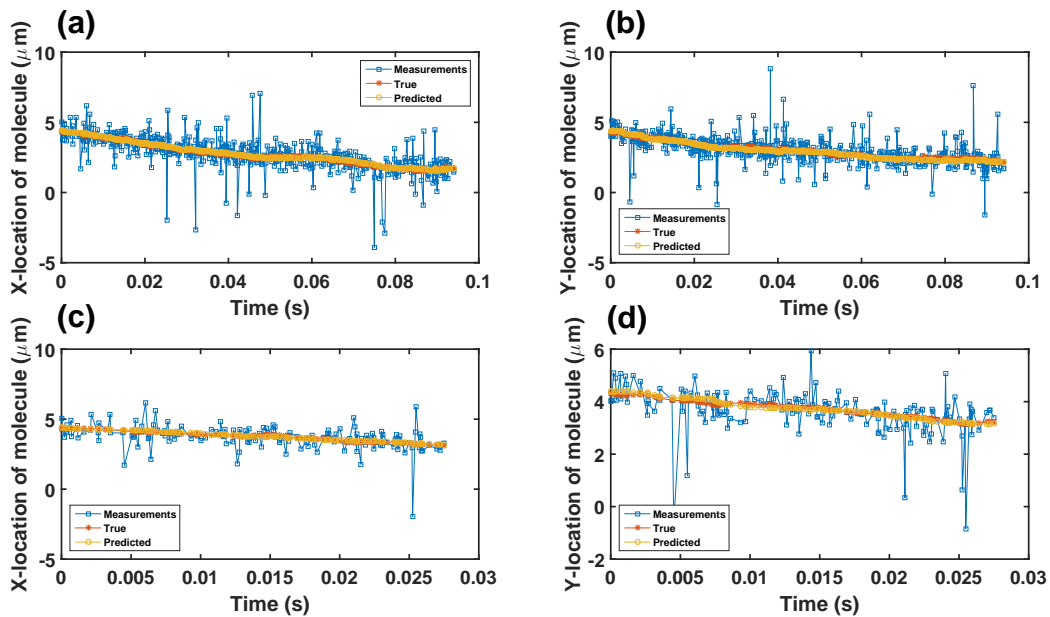
Figure B.2: Predicted locations of the molecule for the Airy measurement model. (a) and (b) Means of the distributions of the prediction of the molecule x- and y-locations, where the first order drift coefficient is unknown, and the true x- and y-locations of the molecule for the same data set as in Figs. B.1(a) and B.1(b). The measurements transformed from the image space to the object space are also shown. (c) and (d) Means of the distributions of the prediction of the molecule x- and y-locations and the true x- and y-locations of the molecule over the time interval $[0, 27.5]$ ms.

with coefficient matrices given through Eqs. (B.3)-(B.5). In Section B.1, we calculated these coefficient matrices for drift and diffusion coefficients estimation problem in different scenarios.

B.5 Computation of general Fisher information matrix

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

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

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

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

$$
\begin{aligned}
dp_{X(\tau_1)}(x_1) &= \int_{\mathbb{R}^2} dp_{X(\tau_1)|X(t_0)}(x_1|x)\, p_{X(t_0)}(x)\, dx \\
&\approx h^2 \sum_{i=0}^{n} \sum_{j=0}^{n} dp_{X(\tau_1)|X(t_0)}(x_1|(x_i, y_j))\, p_{X(t_0)}(x_i, y_j), \quad x_1 \in \mathbb{R}^2. \quad \text{(B.12)}
\end{aligned}
$$

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

$$p_{U_1|T_1}(r|\tau_1) = \int_{\mathbb{R}^2} p_{X(\tau_1)}(x)\, p_{U_1|X(\tau_1)}(r|x)\, dx$$

$$= \int_{\mathbb{R}^2} p_{X(\tau_1)}(x)\, f_x(r)\, dx$$

$$= \frac{1}{|\det(M)|} \int_{\mathbb{R}^2} p_{X(\tau_1)}(x)\, q\left(M^{-1}r - x\right) dx$$

$$\approx \frac{h^2}{|\det(M)|} \sum_{i=0}^{n} \sum_{j=0}^{n} p_{X(\tau_1)}(x_i, y_j)\, q\left(M^{-1}r - (x_i, y_j)\right), \quad r \in \mathcal{C}.$$

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

$$dp_{U_1|T_1}(r|\tau_1) = \int_{\mathbb{R}^2} dp_{X(\tau_1)}(x)\, f_x(r)\, dx$$

$$= \frac{1}{|\det(M)|} \int_{\mathbb{R}^2} dp_{X(\tau_1)}(x)\, q\left(M^{-1}r - x\right) dx$$

$$\approx \frac{h^2}{|\det(M)|} \sum_{i=0}^{n} \sum_{j=0}^{n} dp_{X(\tau_1)}(x_i, y_j)\, q\left(M^{-1}r - (x_i, y_j)\right), \quad r \in \mathcal{C}.$$

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

$$I(D) = \int_{\mathcal{C}} \frac{1}{p_{U_1|T_1}(r|\tau_1)} dp^2_{U_1|T_1}(r|\tau_1)\, dr$$

$$\approx h_r^2 \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{1}{p_{U_1|T_1}\left((r_{x_i}, r_{y_j})|\tau_1\right)} dp^2_{U_1|T_1}\left((r_{x_i}, r_{y_j})|\tau_1\right).$$

B.6 Joint probability distribution of arrival time points for a Poisson process

Lemma 4. *For $t_0 \in \mathbb{R}$, let $\{N(\tau), \tau \geq t_0\}$ be a Poisson process with intensity function $\Lambda(\tau), \tau \geq t_0$. Let $\mathcal{T}_l := (T_1, \cdots, T_l)^T, l = 1, \cdots, N(\tau), \tau \geq t_0$, where the 1D random variable T_l describes*

the lth arrival time points of $\{N(\tau), \tau \geq t_0\}$.

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

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

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

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

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

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

Proof. See Section 2 of [55]. □

B.7 Derivative of state estimates

Lemma 5. *Let Θ denote a parameter space that is an open subset of $\mathbb{R}^n$, and let $\tau_1 \in \mathbb{R}$. For $\theta = (\theta_1, \cdots, \theta_n) \in \Theta, r_l \in \mathcal{C}, l = 1, 2, \cdots$, and $\tau_1 < \tau_2 < \cdots$, let*

$$\hat{x}_{\theta,l+1}^l = \tilde{\phi}_\theta(\tau_l, \tau_{l+1}) \left(\hat{x}_{\theta,l}^{l-1} + K_{\theta,l}(r_l - C\hat{x}_{\theta,l}^{l-1}) \right), \quad \hat{x}_{\theta,l+1}^l \in \mathbb{R}^k, \tag{B.13}$$

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

$$X_{\theta,l+1}^{(i)} = A_{\theta,l+1}^{(i)} X_{\theta,l}^{(i)} + B_{\theta,l+1}^{(i)} e_{\theta,l}, \quad l = 1, 2, \cdots,$$

138

where

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

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

Proof. By differentiating Eq. (B.13) (Kalman state estimate update formula) with respect to θ_i, $i = 1, \cdots, n$, we have, for $l = 1, 2, \cdots,$

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

$$+ \left(\tilde{\phi}_\theta(\tau_l, \tau_{l+1})\frac{\partial K_{\theta,l}}{\partial \theta_i} + \frac{\partial \tilde{\phi}_\theta(\tau_l, \tau_{l+1})}{\partial \theta_i}K_{\theta,l} \right) e_{\theta,l}, \tag{B.14}$$

Then, by combining Eqs. (B.13) and (B.14), for $X_{\theta,l}^{(i)} = \begin{bmatrix} \hat{x}_{\theta,l}^{l-1} \\ \frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_i} \end{bmatrix}$, we have the following recursive

formulation

$$X_{\theta,l+1}^{(i)} = A_{\theta,l+1}^{(i)}X_{\theta,l}^{(i)} + B_{\theta,l+1}^{(i)}e_{\theta,l},$$

where

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

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

$\square$

B.8 Chain rule

Theorem 8. *Let S be an open set in $\mathbb{R}^K$ and let c be a point of S. Let $d = (d_1, \cdots, d_M)$ be a function mapping S into an open set H in $\mathbb{R}^M$, i.e., $d: S \mapsto H$, that is differentiable at c. Let*

$h = (h_1, \cdots, h_N)$ *be a function mapping H into an open set Q in $\mathbb{R}^N$, i.e., $h\colon H \mapsto Q$, that is differentiable at $d(c)$. Let q be a real-valued function defined on Q that is differentiable at $h(d(c))$. Then,*

$$(D_k(q \circ h \circ d))(c) = \sum_{i=1}^{N} \sum_{j=1}^{M} (D_i q)(h(d(c)))(D_j h_i)(d(c))(D_k d_j)(c), \quad k = 1, \cdots, K.$$

Proof. See the proof of Corollary 8.4.3 of [65]. □

B.9 Integral transformation theorem

Theorem 9. *Let $g = (g_1, g_2, \cdots, g_n)\colon B \subseteq \mathbb{R}^n \mapsto \mathbb{R}^n$ be an injective and continuously differentiable function. Let $w\colon \mathbb{R}^n \mapsto \mathbb{R}$ be an integral function and $A \subseteq \mathbb{R}^n$, then the* integral transformation theorem *is given by*

$$\int_{g(A)} w(y_1, y_2, \cdots, y_n) dy_1 dy_2 \cdots dy_n = \int_A w(g(x_1, x_2, \cdots, x_n))$$
$$\times |\det(J(g)(x_1, x_2, \cdots, x_n))| \, dx_1 dx_2 \cdots dx_n,$$

where the Jacobian matrix *is given by*

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

Proof. See Section 10.3 of [65]. □

B.10 Innovation representation of the state space model

Lemma 6. *Let $\mathcal{G}_{[t]}^g \left(\left(\tilde{X}, g, \tilde{W}_g, Z_g \right), \left(\mathcal{U}_{[t]}, \mathcal{T}_{[t]} \right), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$ (or $\mathcal{G}_L^g \left(\left(\tilde{X}, g, \tilde{W}_g, Z_g \right), \left(\mathcal{U}_L, \mathcal{T}_L \right), \tilde{\Phi}, M', \mathcal{C}, \Theta \right)$) be an image detection process with expanded state space $\tilde{X}$ and Gaussian process and measurement noise models for a time interval $[t_0, t]$ (or for a fixed number L of photons).*

Assume that

$$X(\tau) = g(\tilde{X}(\tau)) := G\tilde{X}(\tau), \quad \tau \geq t_0,$$

where $G \in \mathbb{R}^{3 \times k}$, and let $C := M'G$. Assume that C and Z_g are independent of the parameter vector $\theta \in \Theta$. For $\theta = (\theta_1, \cdots, \theta_n)$ and $\hat{x}_{\theta,l}^{l-1} := E\left[\tilde{X}_\theta(\tau_l)|r_{l-1}, \cdots, r_1\right]$, let $X_{\theta,l}^{(i)} := \begin{bmatrix} \hat{x}_{\theta,l}^{l-1} \\ \frac{\partial \hat{x}_{\theta,l}^{l-1}}{\partial \theta_i} \end{bmatrix}, i = 1, \cdots, n$, be the extended state vector and $e_{\theta,l} := r_l - C\hat{x}_{\theta,l}^{l-1}$ be the prediction error. Then, $E\left[e_{\theta,l}X_{\theta,l}^{(i)}\right] = 0, i = 1, \cdots, n$.

Proof. Since the measurement noise Z_g, the process noise $\tilde{W}_g$, and the initial condition of the state vector $\tilde{X}$ are independent, the prediction error $e_{\theta,l}$ and the extended state vector $X_{\theta,l}^{(i)}, i = 1, \cdots, n$, are independent (see the proof of Theorem 5 of [66]), and we have

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

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

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

$\square$

B.11 Analysis of the error of the predicted locations of the molecule

In this section, the errors between the means of the distributions of the prediction of the molecule locations and the true locations of the molecule for Born and Wolf, Airy and Gaussian measurements are shown in Figs. B.3, B.4 and B.5, respectively.
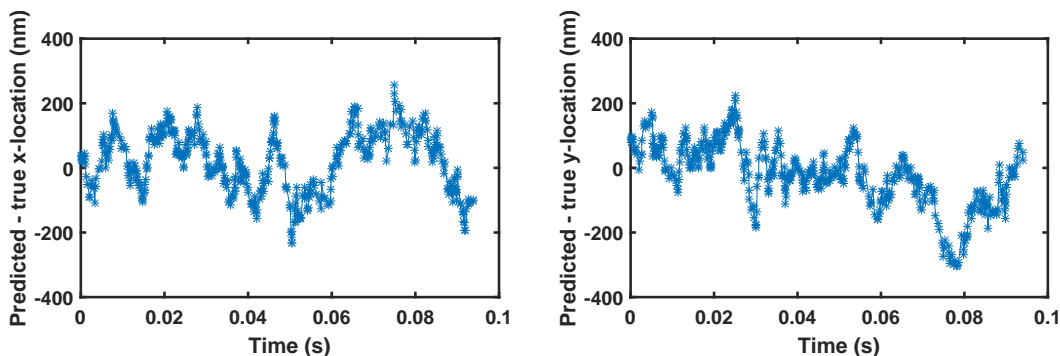
Figure B.3: Analysis of the error of the predicted locations of the molecule for the Born and Wolf measurement model. Shown in the left and right plots are the differences between the means of the distributions of the prediction of the molecule x-locations, where the first order drift coefficient is unknown, and the true x-values, and the means of the distributions of the prediction of the molecule y-locations and the true y-values, respectively, for the data sets of Fig. 3.4.
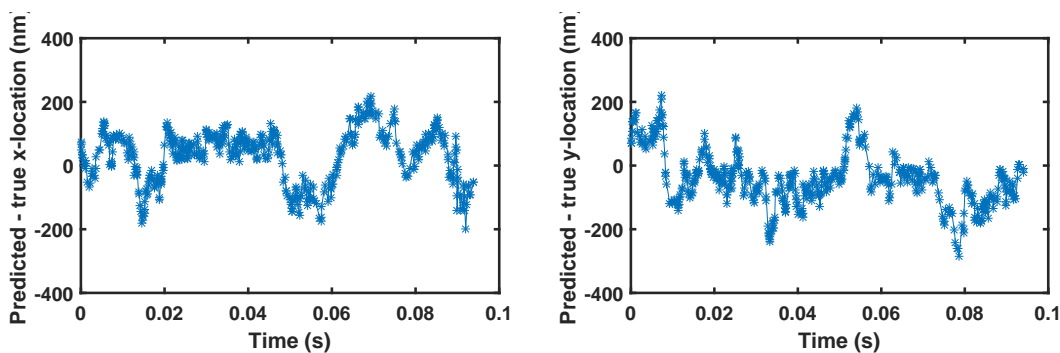

Figure B.4: Analysis of the error of the predicted locations of the molecule for the Airy measurement model. Shown in the left and right plots are the differences between the means of the distributions of the prediction of the molecule x-locations, where the first order drift coefficient is unknown, and the true x-values, and the means of the distributions of the prediction of the molecule y-locations and the true y-values, respectively, for the data sets of Fig. B.2.
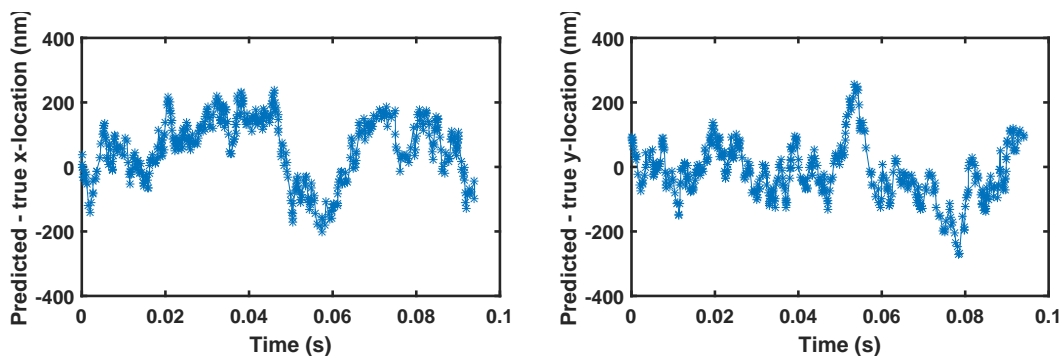
Figure B.5: Analysis of the error of the predicted locations of the molecule for the Gaussian measurement noise case. Shown in the left and right plots are the differences between the means of the distributions of the prediction of the molecule x-locations, where the diffusion coefficient is unknown, and the true x-values, and the means of the distributions of the prediction of the molecule y-locations and the true y-values, respectively, for the data sets of Fig. 3.6.