

Precision in iterative modulation enhanced single-molecule localization microscopy

Dylan Kalisvaart

Master of Science Thesis

Precision in iterative modulation enhanced single-molecule localization microscopy

MASTER OF SCIENCE THESIS

For the degree of Master of Science in Systems and Control at Delft
University of Technology

Dylan Kalisvaart

September 23, 2020

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of
Technology

Front cover image was adapted from Howard Vindin (own work) and re-used under CC BY-SA 4.0. Obtained from <https://commons.wikimedia.org/w/index.php?curid=40722030>.



Copyright © Delft Center for Systems and Control (DCSC)
All rights reserved.



DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF
DELFT CENTER FOR SYSTEMS AND CONTROL (DCSC)

The undersigned hereby certify that they have read and recommend to the Faculty of
Mechanical, Maritime and Materials Engineering (3mE) for acceptance a thesis
entitled

PRECISION IN ITERATIVE MODULATION ENHANCED SINGLE-MOLECULE
LOCALIZATION MICROSCOPY

by

DYLAN KALISVAART

in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE SYSTEMS AND CONTROL

Dated: September 23, 2020

Supervisor(s):

dr. C.S. Smith

Reader(s):

prof.dr.ir. M.H.G. Verhaegen

dr. M. Kok

dr. G.J. Verbiest

Abstract

Super-resolution microscopy methods are able to image samples with improved resolution over the diffraction limit. Single-molecule localization microscopy (SMLM) methods break the diffraction limit by sparsely activating fluorescent emitters. The resulting sparsity in the emission signal can be exploited by estimation algorithms, which enables localization with improved precision. Modulation enhanced SMLM (meSMLM) methods further increase the localization precision by combining SMLM techniques with patterned illumination. By using prior information to refine emitter position estimates, iterative meSMLM (imeSMLM) methods such as iterative MINFLUX are able to locally improve the resolution.

For (me)SMLM methods, the Cramér-Rao lower bound (CRLB) is often used to assess the localization precision. The CRLB bounds the variance of arbitrary unbiased estimators from below. As the CRLB treats estimands as deterministic unknowns, a prior distribution on the estimands cannot be incorporated into the bound. Therefore, the effect of prior information on the localization precision of imeSMLM methods is not captured by the CRLB.

The Van Trees inequality (VTI) is a Bayesian variant of the CRLB. Because it treats estimands as random variables with a known prior distribution, it is able to account for the effect of prior information on the estimator precision. It is therefore able to accurately bound the localization precision of imeSMLM methods from below.

An imeSMLM method is considered, in which the positions of sinusoidal intensity patterns are controlled over the course of multiple iterations. Intensity minima of sinusoidal patterns are placed symmetrically around the current estimate of the emitter position, at a distance based on the localization precision of the previous iteration. This strategy balances the information content of signal photons with the need for robustness to estimation errors.

Using the VTI, we derive a fundamental limit on the localization precision of imeSMLM methods that make use of standing wave illumination patterns. This limit shows that in the absence of background, the information content of signal photons increases exponentially as a function of the iteration count. Using Monte Carlo simulations, the maximally achievable localization precision for different illumination pattern placement control strategies was evaluated. The VTI allows to assess the performance of pattern placement control strategies and is therefore a promising method for optimal control of imeSMLM methods.

Table of Contents

Preface	xi
Nomenclature	xv
1 Introduction	1
1-1 Fluorescence microscopy	1
1-2 Resolution	4
1-3 Super-resolution microscopy	6
1-4 Modulation-enhanced single-molecule localization microscopy	8
1-4-1 MINFLUX: Doughnut-shaped beam illumination	8
1-4-2 SIMFLUX: structured illumination	10
1-5 Thesis motivation	11
1-6 Outline	12
2 Parameter estimation	13
2-1 Likelihood functions	13
2-2 Cramér-Rao lower bound	14
2-3 Maximum likelihood estimation	15
2-4 Van Trees inequality	17
3 Precision in iterative modulation enhanced single-molecule localization microscopy	21
3-1 Introduction	22
3-2 Image formation model	23
3-3 Van Trees inequality for iterative localization precision	25
3-3-1 Van Trees inequality	25
3-3-2 Localization precision for imeSMLM	26
3-3-3 Choice of pattern positions	27
3-3-4 Numerical evaluation	30

3-4	Results and discussion	31
3-4-1	Theoretical limit	31
3-4-2	Simulations and parameter values	32
3-4-3	Effect of iterations and pattern positioning	33
3-4-4	Effect of initial guess and estimation errors	37
3-5	Conclusion	40
3-6	Outlook	40
3-7	Supplementary information	42
4	Conclusion	53
4-1	Summary of study	53
4-2	Outlook	55
A	Proofs and derivations	59
A-1	Proof of Theorem 2.1	59
A-2	Proof of Theorem 2.2	61
A-3	Proof of Theorem 2.3	65
A-4	Proof of Lemma A.1	68
	Glossary	77
	List of Acronyms	77

List of Figures

1-1	(a): Example of a Jabłoński diagram for a fluorophore. When a photon is absorbed by a fluorophore, an electron can move from the ground state GS_0 to a higher energy state ES_n (Transition 1). In a higher energy state ES_2 , the electron quickly dissipates energy through vibrational losses, such that it ends up at the lowest excited energy state (Transition 3). When the electron returns to the ground state, the dissipated energy can be dispersed as heat or to a quenching molecule (Transition 2) or it is emitted as a photon, of which the wavelength is determined by the Stokes shift (Transition 4). Source: Adapted from [1]. (b): As a result of energy losses, the emission wavelength is higher than the excitation wavelength. This effect is called the Stokes shift. Source: Adapted from [2].	2
1-2	Design of an epifluorescence microscope. White light from a light source passes an excitation filter, to discard all wavelengths that do not match the absorption spectrum of the used fluorophores. This light is reflected by the dichroic mirror and illuminates the sample. Fluorophores in the sample send out emission light, which is magnified by the lens array and which can pass the dichroic mirror as a result of the Stokes shift. This light is filtered once more by an emission filter, after which it is recorded by a detector. Source: [1].	3
1-3	Simulated Airy pattern, which occurs as a result of diffraction when focusing light from a point source into the image plane. Source: [3].	4
1-4	Example of how a point spread function (PSF) smooths an object, through the convolution of Equation (1-1).	5
1-5	Image processing pipeline to obtain images of samples and to process the raw images data into a super-resolution image using single-molecule localization microscopy (SMLM). In the data acquisition step, a sample (labeled with emitters) is imaged at multiple time instances. As a result of blinking or photobleaching, a small, random subset of emitters is active in each frame. During pre-processing, measured (scaled) intensity values are converted to photon counts, to be used in a fitting model. Next, during the detection step, regions of interest that contain emitters are cropped from each image frame. During localization, emitter positions are estimated for each region of interest. In the post-processing step, the outcomes of the localization step are assessed, checked for errors and corrected. Examples of this are quality control, drift correction and combining multiple localizations. Lastly, the results are visualized. Source: [4].	7

1-6	Figure from [5], illustrating how MINFLUX estimates an emitter position (located at the star) through triangulation with the STED doughnut shaped beam. If it is known that an emitter is contained in a ring (dashed circle), probing with the doughnut beam at four different locations (the blue, purple, red and yellow dots) result in different photon counts. If the emitter is close to the doughnut zero, it will emit a low amount of photons. If the emitter is far from the center of the doughnut beam, it will emit a high amount of photons. From these measurements, the molecule position can be estimated. Source: Adapted from [5].	9
1-7	In SIMFLUX, sinusoidal illumination patterns are used to modulate the emission signal. Standing wave illumination is used in two directions with three equidistant phase shifts between 0 and 2π . The emission signal produced by the emitter (red dot) is modulated because of this. If the Source: Adapted from [6].	10
1-8	In SIMFLUX, sinusoidal illumination patterns are used to modulate the emission signal. Standing wave illumination is used in two directions with three equidistant phase shifts between 0 and 2π . The emission signal produced by the emitter (red dot) is modulated because of this. Source: Adapted from [7].	11
3-1	Simulation of the localization precision of imeSMLM using the Van Trees inequality (VTI). In the first iteration, the Fisher information matrix $I_0(\theta)$ is computed. The resulting maximum likelihood estimate is asymptotically $\mathcal{N}(\theta, I_0^{-1}(\theta))$ distributed, and we choose this as a prior on the next iteration. In every following iteration, the VTI is evaluated using the prior, resulting in the data information matrix $J_{D,k}$ and the prior information matrix $J_{P,k}$ in iteration k . As the next prior, we choose $\mathcal{N}(\theta, (J_{D,k} + J_{P,k})^{-1})$. This iteration continues until the maximum amount of iterations M is reached.	27
3-2	Optimization procedure to select pattern positions that maximize the information content of signal photons. In the initialization step, no prior information is available and pattern positions $\phi_{x,0} = \phi_{y,0} = 0$ are chosen. The resulting CRLB is used to determine a prior for the next iteration. For all further iterations k , $\phi_{x,k}$ and $\phi_{y,k}$ are chosen to minimize the localization precision VTI for a fixed amount of photons, after which the realized VTI is used to determine a prior for the next iteration.	28
3-3	Simulated localization precision in x -direction as a function of the acquired amount of signal photons, when sinusoidal intensity patterns are positioned (a) using Equation (3-13) with $\alpha = 3$; or (b) using Equation (3-14) with $\beta = 3$. Simulated results are compared against the theoretical limit of Equation (3-19). The emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and 3 iterations were simulated. Intensity minima were placed symmetrically around the true emitter position, to eliminate the effect of estimation errors.	34
3-4	Simulated localization precision in x -direction as a function of the maximum amount of iterations, when sinusoidal intensity patterns are positioned using Equation (3-13) with $\alpha = 1, 3$ or 6 or using Equation (3-14) with $\beta = 2$ or 3. Simulated results are compared against the theoretical limit of Equation (3-19). Photons are equally distributed over iterations, such that 2000 photons are acquired when the maximum amount of iterations is reached. The emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and intensity minima were placed symmetrically around the true emitter position, to eliminate the effect of estimation errors. . .	36

3-5	Simulated localization precision in x -direction as a function of the global emitter x -phase $\omega\theta_x$, when sinusoidal intensity patterns are positioned (a) using Equation (3-13) with $\alpha = 3$; or (b) using Equation (3-14) with $\beta = 3$. Simulated results are compared against the theoretical limit of Equation (3-19). Photons are equally distributed over iterations, such that 2000 photons are acquired when the maximum amount of iterations is reached. The initial pattern position $\phi_{x,0} = 0$ was used and intensity minima were placed symmetrically around the true emitter position in for $k \geq 1$, to eliminate the effect of estimation errors.	38
3-6	Simulated localization precision in x -direction as a function of the acquired amount of signal photons, when sinusoidal intensity patterns are positioned (a) using Equation (3-13) with $\alpha = 1, 3$ or 6 ; or (b) using Equation (3-14) with $\beta = 2$ or 3 . Simulated results are compared against the theoretical limit of Equation (3-19). Intensity minima were placed symmetrically around the true emitter position plus an offset of $\sigma_{x,k-1}$ or $2\sigma_{x,k-1}$, to investigate the effect of estimation errors on the localization precision. The emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and 3 iterations were simulated.	39
3-7	Zoomed-in view of the back focal plane of the objective lens of a SIM-setup to establish notation and axis conventions. Boxed numbers indicate beam indices, corresponding to their diffraction orders.	44
4-1	Simulated 3D-SIM illumination pattern. (a): 3D view of the pattern. (b): Cross-section of pattern in (x, z) -plane, for $y = 0 \text{ nm}$. (c): Cross-section of pattern on x -axis, for $y = 0 \text{ nm}$, $z = 0 \text{ nm}$. (d): Cross-section of pattern on z -axis, for $x = 0 \text{ nm}$, $y = 0 \text{ nm}$	56

List of Tables

3-1	Model parameters used in the localization precision simulations for sinusoidal illumination with a Gaussian point spread function. Parameters were based on those used in [7].	33
-----	--	----

Preface

Originating from ancient Greek drama, the protagonist is the most important character of a story. Over the course of the narrative, the protagonist makes choices to propel the plot forwards. As a result of these choices and external factors, protagonists face significant adversities. These show the strengths and weaknesses of the character, testing him in the process. The protagonist faces, overcomes and learns from these problems, becoming a better person in the end.

In many ways, the story of the classical protagonist is a model for a thesis project. As the main character in my project, I made choices to progress my research. Consequently, difficult challenges arose. Many weeks were spent pondering over literature, carefully writing rigorous proofs and struggling with simulation code, all while in the midst of a global pandemic of which I could not have imagined the far-reaching consequences a year ago.

To the classical protagonist, no problem was unsolvable, and the same holds true for this project. Through perseverance, dedication and most importantly, the support of others, I was ultimately able to overcome these challenges, growing as a person and as an engineer because of it. The result of this process lies before you, and I hope you will enjoy reading it as much as I have developing it.

While only my name undersigns this preface, it should be noted that this thesis would not have existed without the continuous support of many others. I would like to express my sincerest gratitude to my supervisor Carlas, for his seemingly never-ending supply of ideas, his genuine care about me and my project and the life lessons he has taught me over the course of the past year. My thanks also goes out to Shih-Te and Jelmer for their support during my project, as well as to all the students I have worked with over the course of the past year, both inside and outside the group. Even though we have been physically separated for the majority of the time, your open and welcoming attitude has made even the coldest videocall a warm experience.

Last, but certainly not least, I want to deeply thank my amazing family and my fantastic friends. Your continuous and unconditional support has pulled me through this challenging time. I am very glad to have each of you in my life and I hope that I may enjoy your love and friendship over the course of a very long time.

Delft, University of Technology
September 23, 2020

Dylan Kalisvaart

“One can take the point of view that in the 20th century, it was the lenses which were decisive. And the lens makers were the “kings”. One had to go to them and ask them for the best lenses to get the best resolution. But how is it today? No, it is not the lens makers. This resolution game is not about lenses anymore.”

— *Stefan W. Hell, Nobel Lecture, December 8, 2014.*

Nomenclature

Symbol	Quantity	Unit
α	Objective half-angle	rad
A	Normalization constant for illumination pattern	-
c	Speed of light in vacuum	m/s
$C_{\hat{\theta}}$	Covariance matrix of $\hat{\theta}$	-
Δx	Pixel size in x -direction	m
Δy	Pixel size in y -direction	m
f_{θ}	Probability density function, dependent on θ	-
$h(\mathbf{r})^2$	Gaussian point spread function in position \mathbf{r}	-
$I(\boldsymbol{\theta})$	Fisher information matrix	-
J	Bayesian information matrix	-
J_D	Bayesian data information matrix	-
J_P	Bayesian prior information matrix	-
k	Wave number	1/m
L	Spatial pattern period	m
λ_{em}	Wavelength of emission light	m
$\lambda(\boldsymbol{\theta})$	Prior probability distribution on $\boldsymbol{\theta}$	-
$L(\boldsymbol{\theta} \mathbf{x})$	Likelihood function of $\boldsymbol{\theta}$, given measurements \mathbf{x}	-
$\ell(\boldsymbol{\theta} \mathbf{x})$	Log-likelihood function of $\boldsymbol{\theta}$, given measurements \mathbf{x}	-
m	Modulation contrast	-
M	Amount of iterations	-
μ	Poisson mean, in particular expected photon count	photons
n	Index of refraction	-
n_{MC}	Amount of Monte Carlo samples	-
n_x	Amount of pixels in x -direction	-
n_y	Amount of pixels in y -direction	-
NA	Numerical aperture	-

Symbol	Quantity	Unit
p_θ	Probability mass function, dependent on θ	-
P	Illumination pattern intensity	-
\mathbf{r}	Position vector	m
σ	Standard deviation of a random variable	-
σ_{CRLB}	Localization precision Cramér-Rao lower bound	m
σ_{VT}	Localization precision Van Trees inequality	m
σ_{PSF}	Standard deviation of Gaussian PSF	m
t	Time	s
θ	Parameter, used in estimation	-
$\boldsymbol{\theta}$	Parameter vector, used in estimation	-
$\hat{\boldsymbol{\theta}}$	Estimate of parameter vector $\boldsymbol{\theta}$	-
θ_x	Emitter x -position	m
θ_y	Emitter y -position	m
θ_z	Emitter z -position	m
θ_I	Expected signal photon count	photons
θ_b	Expected background photon count per area	photons/ μm^2
U	Monochromatic plane wave	-
x	x -coordinate	m
\mathbf{x}	Measurement vector in parameter estimation	-
x_s	Realization of random variable X_s	-
x_i	x -position of camera pixel i	m
X_s	Random variable with index s	-
y	y -coordinate	m
y_i	y -position of camera pixel i	m
z	z -coordinate	m
ϕ	Phase of a wave	rad
ϕ_0	Initial phase of a wave	rad
ϕ_k	Pattern phase in iteration k	rad
ω	Angular frequency of a wave	rad/s

Chapter 1

Introduction

Optical microscopes have greatly enhanced the knowledge about life on a microscopic scale. By overcoming a fundamental resolution limit, super-resolution fluorescence microscopy methods are able to image cells with nanoscopic precision. In this chapter, we introduce the topic of super-resolution microscopy by discussing important concepts and methodologies. This will ultimately build up to the motivation for this thesis.

In Section 1-1, fluorescence microscopy is introduced. Section 1-2 will discuss the diffraction limit, which is a fundamental limit on the resolution on optical microscopes. Next, in Section 1-3, super-resolution microscopy will be introduced as a way to circumvent the diffraction limit. Recently proposed modulation-enhanced single molecule localization microscopy (meSMLM) methods, which are able to further improve the localization precision with respect to classical super-resolution microscopy methods, are discussed in Section 1-4. Section 1-5 contains the thesis motivation and the outline of this text is given in Section 1-6.

1-1 Fluorescence microscopy

For decades, optical microscopes have played a key role in understanding life. After his development of the first single lens microscope in the seventeenth century, Antoni van Leeuwenhoek (a fellow Delftian) was able to observe bacteria, red blood cells and larvae [8]. The latter observation was solid evidence to dispute the theory of *generatio spontanea*, which was the classical idea that life could originate from dead or lifeless subjects. Van Leeuwenhoek and his primitive microscope provided key evidence to advance the knowledge about life.

Since then, light microscopes have undergone significant developments. For our purposes, the discovery of fluorescence microscopy is of particular interest. Before we discuss the design of an epifluorescence microscope, we briefly explain the principle of fluorescence. In short, fluorescence is a process where a fluorescent emitter absorbs light, after which it re-emits light again when a certain time has passed [1].

To give a detailed explanation of fluorescence, we have to introduce electron energy levels and transitions. This can be done through a so-called Jabłoński diagram [1], as shown in

Figure 1-1(a). A Jabłoński diagram displays the energy states of the electrons in an atom or molecule on the vertical axis, to indicate energy transitions.

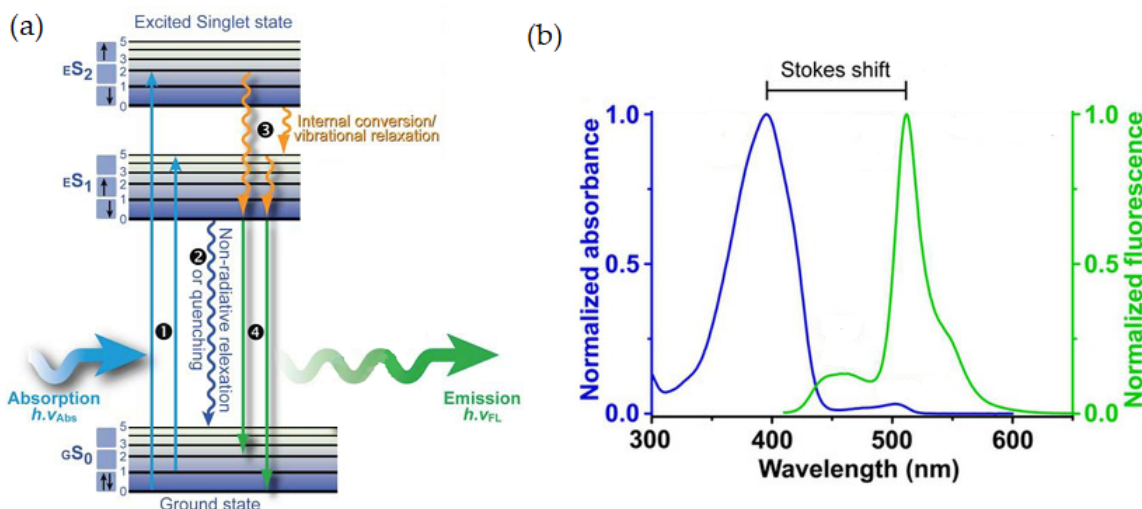


Figure 1-1: (a): Example of a Jabłoński diagram for a fluorophore. When a photon is absorbed by a fluorophore, an electron can move from the ground state GS_0 to a higher energy state ES_n (Transition 1). In a higher energy state ES_2 , the electron quickly dissipates energy through vibrational losses, such that it ends up at the lowest excited energy state (Transition 3). When the electron returns to the ground state, the dissipated energy can be dispersed as heat or to a quenching molecule (Transition 2) or it is emitted as a photon, of which the wavelength is determined by the Stokes shift (Transition 4). Source: Adapted from [1].

(b): As a result of energy losses, the emission wavelength is higher than the excitation wavelength. This effect is called the Stokes shift. Source: Adapted from [2].

In Figure 1-1(a), electrons initially reside in the ground state GS_0 . When light of a specific wavelength hits a fluorophore, a photon can be absorbed by the fluorophore and this moves an electron from the ground state to a higher energy state ES_n [1]. After an electron has moved to a higher energy state (ES_2 in Figure 1-1(a)), it quickly moves back to its lowest excited energy state (ES_1 in Figure 1-1(a)). This is accomplished through vibrational energy losses and the lost energy is dissipated as heat.

When going from the lowest excited state to the ground state, the electron once again needs to release energy [1]. This energy can be lost as heat (non-radiative relaxation) or it can be dispersed to a quenching molecule, both of which will not result in fluorescence. The last option to disperse energy is through releasing a photon, and this is called fluorescence.

Note that, as energy is lost during the transitions, the emitted photon will have a longer wavelength than the absorbed photon, under the assumption that a single photon is used in the initial excitation. The difference between the maxima of the emission and excitation spectra is called the Stokes shift [9], and it is illustrated in Figure 1-1(b).

The difference between the illumination and emission wavelengths can be exploited to image emitters using fluorescence. To explain this, we discuss the design of an epifluorescence microscope. An example of an epifluorescence microscope is shown in Figure 1-2.

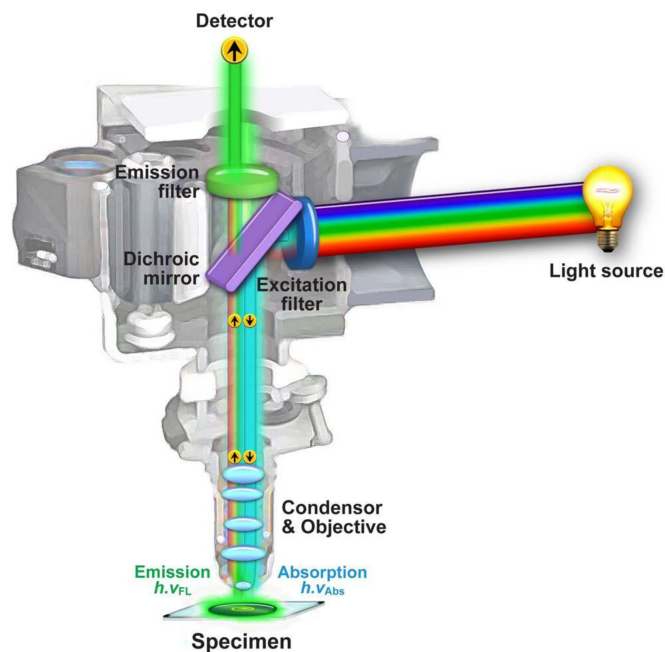


Figure 1-2: Design of an epifluorescence microscope. White light from a light source passes an excitation filter, to discard all wavelengths that do not match the absorption spectrum of the used fluorophores. This light is reflected by the dichroic mirror and illuminates the sample. Fluorophores in the sample send out emission light, which is magnified by the lens array and which can pass the dichroic mirror as a result of the Stokes shift. This light is filtered once more by an emission filter, after which it is recorded by a detector. Source: [1].

In Figure 1-2, light from a light source enters the microscope. Figure 1-2 depicts a white light source, but as an alternative one could use a single-mode laser as a light source [1]. This is advantageous, as it can produce (almost) monochromatic light of high enough intensity to be used in fluorescence microscopy.

As a first step, it is important to filter the excitation light, such that only light in the absorption spectrum of the fluorescent emitter remains. In this way, unwanted excitation is minimized, which may reduce the time it takes for a fluorophore to bleach. The filtering is done through an excitation filter, which is usually band-pass to achieve the aforementioned goal.

After passing through the excitation filter, light encounters a dichroic mirror (DM) [10]. This mirror is a special filter, which reflects certain wavelengths, while letting others pass. It should be chosen, such that the excitation wavelengths are reflected, while the emission wavelengths (which are different as a result of the Stokes shift) pass.

After being reflected by the dichroic mirror, excitation light passes through an array of lenses. These lenses focus the excitation light into the sample plane. The sample has been labeled with fluorophores. In particular, the fluorophores and sample have been treated, such that the fluorophores attach to particular molecules in the sample. While illuminated, the fluorophores emit photons, which are focused into the camera plane by the lens array.

By the choice of the dichroic mirror, the emission wavelengths can pass through, which means they have been separated from the excitation wavelengths and can be imaged on a camera. However, some unwanted wavelengths may still have passed the dichroic mirror. An emission

filter is placed in the emission path after the dichroic mirror, to filter out these unwanted wavelengths. After this, the light is focused on a camera, resulting in an image of the sample. The main advantages of (epi)fluorescence microscopy are its selectivity, its relatively high signal to background ratio (SBR) and its minimal invasivity [1]. As the sample needs to be labeled with fluorophores, they can be chosen or treated such that they attach to specific molecules in the sample. Furthermore, as the emission light is filtered, the microscope discriminates between excitation and emission light. Because of this, the background of the image is reduced (or the contrast in the image is enhanced), resulting in a high SBR. Lastly, fluorescence microscopy is particularly suited for biological purposes, as it is a minimally invasive technique.

A drawback of epifluorescence microscopy is the significant sample volume that is illuminated at a time. Most importantly, this reduces the contrast of the image, as light of fluorophores in the background cannot be distinguished from light of fluorophores in the image plane [10]. Consequently, the image resolution decreases. This issue is solved in other microscope types, such as TIRF microscopy [11] and confocal microscopy [12, 13].

1-2 Resolution

Due to the effects of diffraction, wavefronts of light coming from a point source cannot be focused into an infinitesimally small spot. The image of a point source is called the point spread function (PSF) of the imaging system [14]. The PSF is proportional to the two-dimensional Fourier transform of the complex amplitude of the wavefront in the pupil plane of the optical system. The intensity of a point spread function for an unaberrated optical system with a circular aperture is called the Airy pattern and it is shown in Figure 1-3.

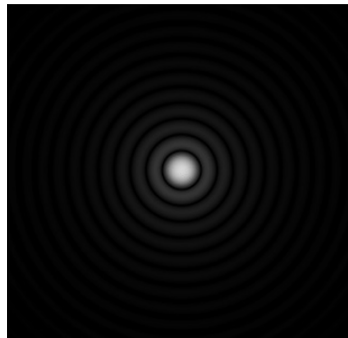


Figure 1-3: Simulated Airy pattern, which occurs as a result of diffraction when focusing light from a point source into the image plane. Source: [3].

The PSF carries information about the effect of the imaging system on the image. Note that a point source is very similar to a Dirac delta function: it is infinitesimally small and maximally bright. Because of this, the PSF is often referred to as the two-dimensional spatial impulse response of an imaging system [14].

As the imaging system does not add or subtract energy from the object, a common assumption is that the PSF $h(x, y)^2$ is normalized: $\iint_{\mathbb{R}^2} h(x, y)^2 dx dy = 1$ [14]. Furthermore, the PSF is not dependent on the position of the point source in the object plane.

As an object can be seen as collections of point sources, the PSF can be used to compute the image of any object by convolution, as shown in Equation (1-1).

$$g(x, y) = h(x, y)^2 \otimes f(x, y) \quad (1-1)$$

In Equation (1-1), $g(x, y)$ denotes the intensity of the image and $f(x, y)$ denotes the intensity of the object. The \otimes operator represents a two-dimensional convolution. From Equation (1-1), we can see that the image is a smoothed version of the object, with the smoothing being done by the PSF. This is also illustrated in Figure 1-4, which shows how a PSF smooths an object.

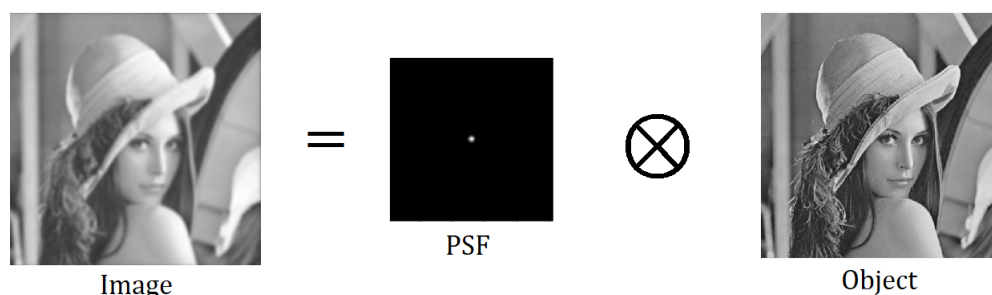


Figure 1-4: Example of how a point spread function (PSF) smooths an object, through the convolution of Equation (1-1).

Figure 1-4 shows that as a result of smoothing of the object by the PSF, features in an image may become indistinguishable. The degree to which features are distinguishable is quantified through the resolution of an optical system. Specifically, the resolution of an optical system specifies the distance two objects need to be apart, for their images to be distinguishable (or resolved) [14].

To quantify the best possible resolution that can be achieved by a light microscope, the Abbe criterion can be used. It is defined as the highest spatial frequency (i.e. the smallest periodicity) of an object that is able to pass through the objective [15]. The Abbe criterion, often also called the diffraction limit, is shown in Equation (1-2).

$$\sigma_{x,y} = \frac{\lambda_{em}}{2NA} \quad (1-2)$$

In Equation (1-2), $\sigma_{x,y}$ is the lateral diffraction limited resolution, λ_{em} is the wavelength of emission light and the numerical aperture NA is given by $NA = n \sin \alpha$. Here, n represents the refractive index of the medium in which the objective lens is working and α represents the maximal half-angle of light that can enter the objective.

Due to the diffraction limit, the resolution $\sigma_{x,y}$ of a light microscope takes values around [150, 400] nm. For example, if we take green light with a wavelength of 500 nm and a numerical aperture of 1.3, we find a resolution $\sigma_{x,y} = 192$ nm. While this certainly is a small number, it is not enough to differentiate between proteins, small molecules or viruses, all of which are (roughly speaking) in the range of 1 nm to 100 nm [16, 17].

1-3 Super-resolution microscopy

Super-resolution microscopy methods break the diffraction barrier of Equation (1-2) [16]. Originally, methods used either the illumination path [18, 19] or the emission path [17, 20, 21, 22] of the microscope. For our purposes, the latter category of methods is of interest. Methods that use the emission path (often called single-molecule localization microscopy (SMLM), or localization microscopy) sparsely activate fluorescent labels and exploit the resulting sparsity in the emission signal to estimate emitter positions with high precision [17, 23, 24].

The objective of SMLM is to isolate emitters in space and in time [16]. With isolation in space, we mean that molecules that are visible at the same time should be located outside of a range defined by the diffraction limit. Because of this, their PSF's appear separately in an image.

With isolation in time, we mean that only one molecule within a range defined by the diffraction limit may be visible at a time. Without temporal separation, the PSF's of neighboring emitters overlap in the same image. Temporal separation results in their PSF's appearing on separate images, thus making them individually distinguishable.

To achieve this spatio-temporal separation, a key role is played by fluorescence. SMLM methods sparsely activate fluorescent emitters, such that there is a high probability of (at least some) isolated PSF's appearing on every image frame. Using the recorded isolated PSF's, an estimation algorithm can be used to localize molecules with increased precision.

The localization precision of SMLM methods is proportional to $\frac{\sigma_{x,y}}{\sqrt{\theta_I}}$, where $\sigma_{x,y}$ is the diffraction limit of Equation (1-2) and where θ_I is the number of collected photons from one emitter. From this, it can be seen that SMLM is able to circumvent the diffraction limit, by recording multiple signal photons.

To obtain estimates of emitter positions and their localization precision, SMLM methods acquire and process data. Based on [4] we present the following image processing pipeline to obtain images of samples and to process the raw image data into a super-resolution image. This pipeline is illustrated in Figure 1-5.

1. Data acquisition
2. Pre-processing
3. Detection
4. Localization
5. Post-processing
6. Visualization

We briefly outline the steps of the image processing pipeline of Figure 1-5. In the data acquisition step, data from an SMLM experiment is acquired. During each measurement, a random, sparse subset of fluorescent emitters is activated. There is a high probability that (at least some) PSF's are spatially separated. When recording the next image, a new set

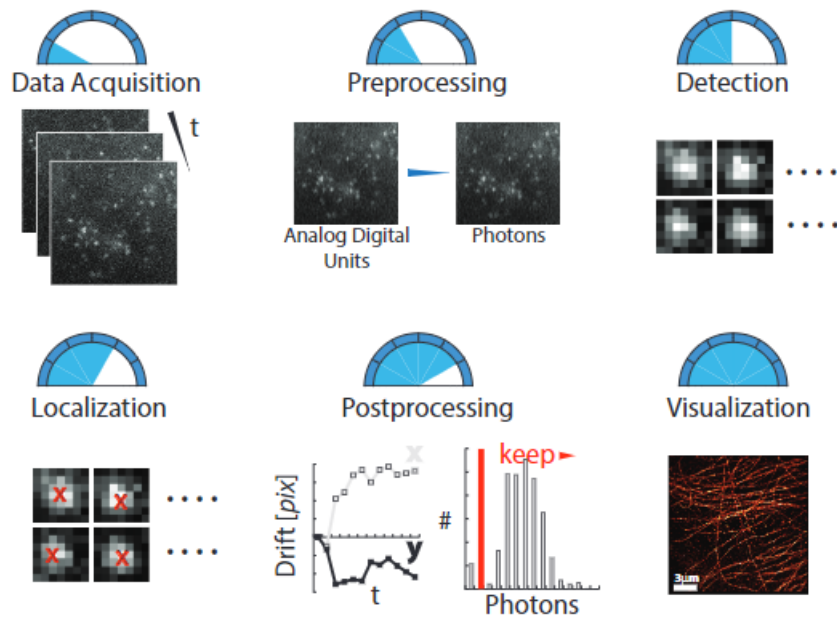


Figure 1-5: Image processing pipeline to obtain images of samples and to process the raw images data into a super-resolution image using single-molecule localization microscopy (SMLM). In the data acquisition step, a sample (labeled with emitters) is imaged at multiple time instances. As a result of blinking or photobleaching, a small, random subset of emitters is active in each frame. During pre-processing, measured (scaled) intensity values are converted to photon counts, to be used in a fitting model. Next, during the detection step, regions of interest that contain emitters are cropped from each image frame. During localization, emitter positions are estimated for each region of interest. In the post-processing step, the outcomes of the localization step are assessed, checked for errors and corrected. Examples of this are quality control, drift correction and combining multiple localizations. Lastly, the results are visualized. Source: [4].

of emitters will be activated, which means that temporal separation is also achieved. This ultimately results in a stack of images, each of which contains many (likely) separated PSF's [24].

Next, the images should be pre-processed. In [4], images (containing recorded intensity data) are corrected for offset and gain. In this way, photon counts instead of intensity data can be used in the image formation model.

In the detection step, regions of images that contain emitter signals should be selected [4]. There are many possible ways to select these regions of interest (RoIs), such as cropping out a region if a certain photon count threshold is exceeded or doing a generalized likelihood ratio test (GLRT) [25].

In the localization step, emitter positions are estimated for each region of interest. Simply picking the maximum intensity pixel is not sufficient, as the resolution is then limited by the pixel size [4]. A basic method that can estimate the emitter position with subpixel precision is the center of mass algorithm [26]. However, this does not take any information we have about the PSF or the imaging model into account. In [27], it is proposed to localized using maximum likelihood estimation (MLE). A detailed description of this method is given in Chapter 2.

Apart from estimating the emitter position, another important part of the localization step

is to determine the localization precision. After all, the goal is to localize emitters with improved precision over the diffraction limit. Finding out the estimation error is non-trivial in an experimental scenario, as the true estimator position is unknown. Hence, estimation error bounds are often used to estimate the localization precision [28]. We introduce two estimation error bounds, namely the Cramér-Rao lower bound and the Van Trees inequality, in Chapter 2.

In the post-processing step, the outcomes of the localization step are assessed, checked for errors and corrected. In [4], three parts of the post-processing step are discussed: quality control, drift correction and combining multiple localizations. In the quality control part, the outcomes of the detection step and the localization step are compared, to decrease the false positive and false negative rates in the detection step. In the drift correction part, images are corrected for moving emitters. The last part of post-processing is to combine multiple localizations belonging to the same emitter, to estimate the position of this emitter with improved precision.

In the final step, the results are visualized. Ultimately, one ends up with a set of localized molecules, an estimate of their localization precision and a visualization that combines these quantities [24].

1-4 Modulation-enhanced single-molecule localization microscopy

Recently, methods have been developed that make use of both the illumination path and the emission path of a microscope [5, 6, 7, 29, 30, 31]. Instead of using uniform illumination, like in the classical SMLM methods, these modulation enhanced single-molecule localization microscopy (meSMLM) methods use patterned illumination. In this way, the recorded emission signals are modulated, which increases the information content. meSMLM methods are therefore able to improve the resolution of optical microscopes even further.

In this section, we take a look at recently introduced modulation enhanced single-molecule localization microscopy (meSMLM) methods. In Subsection 1-4-1, we discuss MINFLUX [5, 6], which modulates the emitter signal with a doughnut-shaped illumination beam. In Subsection 1-4-2, we introduce SIMFLUX [7], which makes use of a sinusoidal illumination pattern.

1-4-1 MINFLUX: Doughnut-shaped beam illumination

In this subsection, we explain how MINFLUX is able to localize emitters. Furthermore, we discuss an iterative variant of MINFLUX, which locally improves the resolution around the emitter position.

In MINFLUX, a doughnut-shaped beam is combined with maximum likelihood estimation to decrease the amount of signal photons needed to achieve a given resolution [5]. A doughnut-shaped beam, as illustrated in Figure 1-6, is a circular illumination pattern with high intensity on the boundaries and low or zero intensity in the center.

To localize an emitter with MINFLUX, a triangulation-like procedure is used. This is illustrated in Figure 1-6. Suppose we know from an earlier scan of the sample that an emitter (star in Figure 1-6) is located in a ring (dashed circle in Figure 1-6). A doughnut-shaped

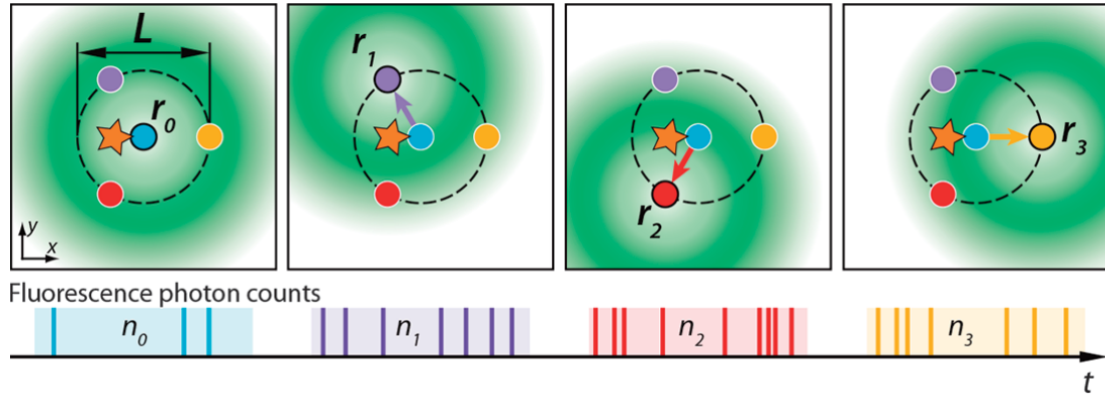


Figure 1-6: Figure from [5], illustrating how MINFLUX estimates an emitter position (located at the star) through triangulation with the STED doughnut shaped beam. If it is known that an emitter is contained in a ring (dashed circle), probing with the doughnut beam at four different locations (the blue, purple, red and yellow dots) result in different photon counts. If the emitter is close to the doughnut zero, it will emit a low amount of photons. If the emitter is far from the center of the doughnut beam, it will emit a high amount of photons. From these measurements, the molecule position can be estimated. Source: Adapted from [5].

beam is used with a diameter that is at least twice the diameter of this ring.

By positioning the illumination minimum of the doughnut-shaped beam in the center of the ring and at the boundary (blue, purple, red and yellow dots in Figure 1-6) and measuring the fluorescent response, we get information about the position of the emitter. If the emitter is close to the intensity minimum of the doughnut-shaped beam, it will emit a low amount of photons. If the emitter is far from the intensity minimum of the doughnut-shaped beam, it will emit a high amount of photons. The information content of signal photons is thus increased, as the emission signal carries additional information on the position of the emitter relative to the doughnut-shaped beam. As such, emitters can be localized with increased precision.

Being able to triangulate the emitter position with a small triangulation ring increases the localization accuracy [5]. Iterative MINFLUX [6] uses prior information from earlier measurements to redefine the triangulation ring during each iteration. This procedure is illustrated in Figure 1-7.

First, the sample is scanned until an emitter is found. In the initial step, the emitter position is estimated by probing around the emitter with a Gaussian laser beam in four positions. This gives an estimate of the emitter position that is accurate enough to start the MINFLUX triangulation. In each iteration, the radius of the triangulation ring is reduced by three times the estimated localization precision of the previous iteration. The MINFLUX triangulation is now performed on the smaller triangulation circle, using doughnut-shaped beams. In the next iteration, the radius of the triangulation circle is reduced again and the process repeats. In [6], it is argued that increasing the amount of iterations significantly improves the localization precision over MINFLUX. Furthermore, it is beneficial to do more iterations instead of collecting more photons per iteration, as photons become more informative when the amount of iterations increases.

In practice, the amount of iterations cannot be increased indefinitely. By increasing the

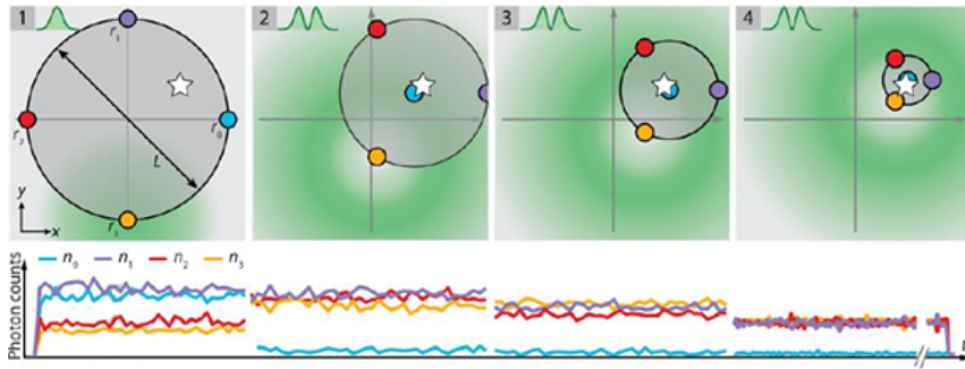


Figure 1-7: In SIMFLUX, sinusoidal illumination patterns are used to modulate the emission signal. Standing wave illumination is used in two directions with three equidistant phase shifts between 0 and 2π . The emission signal produced by the emitter (red dot) is modulated because of this. If the Source: Adapted from [6].

amount of iterations, we need to accept that the amount of signal photons per iteration is reduced as a result of e.g. bleaching of the fluorescent emitter. Because of this, the signal-to-background ratio for each iteration lowers and the resolution is perturbed by background (or, in the worst case, the emitter is no longer contained in the triangulation circle and the procedure does not converge to the emitter position anymore). Hence, the process is iterated until the signal-to-background ratio is below a predefined threshold.

1-4-2 SIMFLUX: structured illumination

In this subsection, we discuss how SIMFLUX uses structured illumination to increase the information contained in images.

The main drawback of both MINFLUX and iterative MINFLUX is that the processes are not suitable for imaging large samples. The triangulation regions are small (in [5], 50 nm, 100 nm and 150 nm were used) and the procedure is hard to parallelize in practice [7]. SIMFLUX is a wide-field technique and as such, it is able to image emitters over a large field of view.

In SIMFLUX [7], sinusoidal intensity patterns are used to modulate the emission signal. This increases the information content per signal photon over SMLM methods, enabling localization with improved resolution. This is illustrated in Figure 1-8.

Sinusoidal intensity patterns are generated in the sample plane by interference of two plane waves. Patterns are oriented either in the x - or y -direction. In each direction, the pattern is phase shifted three times with an equidistant spacing and emission signals are recorded per pattern orientation and position. Similar to what was seen for MINFLUX, the recorded emission signal is modulated by the sinusoidal illumination patterns: if an emitter was located near a minimum of the sinusoidal illumination pattern, the corresponding photon count is low, and vice versa. This additional information makes it possible to image emitters with increased precision.

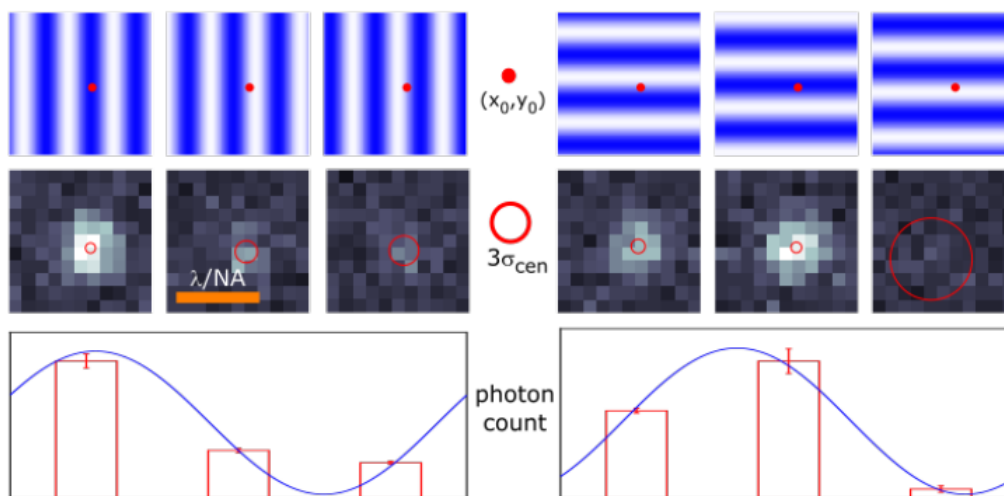


Figure 1-8: In SIMFLUX, sinusoidal illumination patterns are used to modulate the emission signal. Standing wave illumination is used in two directions with three equidistant phase shifts between 0 and 2π . The emission signal produced by the emitter (red dot) is modulated because of this. Source: Adapted from [7].

1-5 Thesis motivation

As discussed in the previous sections, localization precision plays an important role to assess the quality of super-resolution methods. To characterize the localization precision of (me)SMLM methods, the Cramér-Rao lower bound (CRLB) is often used [28]. Under mild assumptions on the likelihood function of the acquired data, the CRLB bounds the variance of unbiased estimators from below [32]. It is shown in [27] that in SMLM, the variance of the maximum likelihood estimator approximately converges to the CRLB for 100 or more signal photons. The CRLB can thus be used to estimate the localization precision of (me)SMLM methods.

However, the CRLB cannot be used to accurately estimate the localization precision of iterative modulation enhanced single-molecule localization microscopy (imeSMLM) methods, such as iterative MINFLUX. As imeSMLM methods use prior information to improve emitter position estimates over the course of multiple iterations, a suitable error bound should be able to incorporate prior information. The CRLB treats estimands as deterministic variables and as such, it is not able to incorporate prior information on the estimands into the localization precision. For this reason, the CRLB fails to estimate how prior information affects the localization precision.

In this text, we use the Van Trees inequality (VTI) as a Bayesian alternative to the CRLB if prior information on the estimands, such as the emitter position, is available. The VTI treats estimands as random variables with a known prior distribution, which makes it suited to bound the localization precision of imeSMLM methods from below. Specifically, we use the VTI to develop a fundamental limit on the localization precision of imeSMLM methods, where standing wave intensity patterns are used for the localization.

In addition, the VTI can be used as an objective function in optimal control of iterative meSMLM methods. In this text, we study an example where the positions of sinusoidal in-

tensity patterns (as used in SIMFLUX) are chosen to maximize the information content per signal photon. If the amount of signal photons that are recorded per iteration is fixed, this optimization problem can be solved by finding the pattern positions that minimize the VTI. We show that the information content of signal photons increases exponentially as a function of the iteration count in the absence of background. Furthermore, we find that intensity minima should be placed directly on the (estimate of the) emitter position to maximize the information content of photons. As this pattern placement may cause robustness issues due to the effect of estimation errors, we introduce two alternate pattern position control strategies that balance performance and robustness. Using Monte Carlo simulations, we are able to evaluate the performance of these pattern position control strategies using the VTI as a performance metric.

1-6 Outline

This text is structured as follows. Chapter 2 discusses the mathematical background on maximum likelihood estimation, the Cramér-Rao lower bound and the Van Trees inequality. Sections 2-1 introduces likelihood functions. In Sections 2-2 to 2-4, maximum likelihood estimation and estimation error bounds are discussed.

In Chapter 3, the main findings of this thesis are described in the form of a manuscript. In the manuscript, an image formation model for an imeSMLM method will be described and the Van Trees inequality will be used to estimate the localization precision of this method. Results include a fundamental limit on the localization precision of imeSMLM and simulations of the localization precision under the effects of background and image discretization. Three supplementary notes are included, which derive important results from the main text. In Chapter 4, the thesis is concluded with a summary of the main findings. Furthermore, we recommend ways to improve this research and we provide an outlook on future developments regarding control of imeSMLM methods. Lastly, in Appendix A, rigorous proofs and derivations for results in the main text are found.

Parameter estimation

In this chapter, the mathematical background for maximum likelihood estimation is outlined. In addition, lower bounds on the estimation error are formulated. In Section 2-1, the likelihood function is introduced. In Section 2-2, the Cramér-Rao lower bound is formulated. In Section 2-3, the maximum likelihood estimator and its properties are outlined. In Section 2-4, we introduce the Van Trees inequality as the Bayesian extension of the Cramér-Rao lower bound.

2-1 Likelihood functions

In this section, we introduce the likelihood function, which will be used in the coming sections to compute the Cramér-Rao lower bound, to introduce maximum likelihood estimators and to compute the Van Trees inequality, respectively. Suppose we have a data set $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, which can be seen as realizations of independent and identically distributed (i.i.d.) random variables $[X_1, X_2, \dots, X_n]^T$. Assume furthermore that the underlying distribution of these random variables is known, up to a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$. To indicate the dependence of the probability mass function (for discrete random variables) on $\boldsymbol{\theta}$, we write it as $p_{\boldsymbol{\theta}}(x)$. Similarly, we write $f_{\boldsymbol{\theta}}(x)$ for the probability density function for continuous random variables.

The likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$ for these measurements describes the probability of obtaining precisely the measurements $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ when realizing $[X_1, X_2, \dots, X_n]^T$, given a value of the parameter $\boldsymbol{\theta}$ [33]. The likelihood function is defined in Definition 2.1.

Definition 2.1. (Likelihood function) Suppose we have a data set $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, which can be seen as realizations of i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$, of which the distribution depends on the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$. For discrete random variables $[X_1, X_2, \dots, X_n]^T$, the likelihood $L(\boldsymbol{\theta}|\mathbf{x})$ is given by:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^n p_{\boldsymbol{\theta}}(x_i)$$

For continuous random variables $[X_1, X_2, \dots, X_n]^T$, the likelihood $L(\boldsymbol{\theta}|\mathbf{x})$ is given by:

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^n f_{\boldsymbol{\theta}}(x_i)$$

In maximum likelihood estimation (see Section 2-3), it is the objective to maximize the likelihood function by an optimal choice of $\boldsymbol{\theta}$. Maximizing the expression of Definition 2.1 by setting the first derivative to zero may be computationally intensive, as the product rule for differentiation needs to be applied repeatedly [34]. For discrete random variables, there is a second drawback. As probability mass functions take values between 0 and 1, the formulation of the likelihood function of Definition 2.1 goes to zero as the amount of measurements n increases, which is undesirable for numerical optimization routines.

Hence, the likelihood is often replaced by the log-likelihood $\ell(\boldsymbol{\theta}|\mathbf{x}) = \log(L(\boldsymbol{\theta}|\mathbf{x}))$ (where \log denotes the natural logarithm). As the logarithm is a monotonically increasing function, $L(\boldsymbol{\theta}|\mathbf{x})$ and $\ell(\boldsymbol{\theta}|\mathbf{x})$ attain their extreme values for identical $\boldsymbol{\theta}$. Furthermore, the use of a logarithm turns the product terms in Definition 2.1 into a sum of logarithms, which can be differentiated using the sum rule for differentiation. The log-likelihood functions for discrete and continuous random variables are shown in Equations (2-1) and (2-2), respectively.

$$\ell(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \log(p_{\boldsymbol{\theta}}(x_i)) \quad (2-1)$$

$$\ell(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \log(f_{\boldsymbol{\theta}}(x_i)) \quad (2-2)$$

2-2 Cramér-Rao lower bound

In this section, we introduce the Cramér-Rao lower bound (CRLB), which bounds the variance of any unbiased estimator from below. We do so for scalar and vector estimators, respectively. The CRLB is a lower bound on the variance of any unbiased estimator of an unknown, but deterministic, parameter [32]. Many similar bounds exist: noteworthy alternatives are the Van Trees inequality (see Section 2-4) or the Bhattacharyya bound [36], which have their individual advantages and disadvantages with respect to the CRLB.

The main advantage of the CRLB with respect to the aforementioned bounds is that it is relatively easy to obtain an analytical expression for the CRLB [32]. Furthermore, maximum likelihood estimators will converge to the CRLB when the amount of measurements increases [34]. Because of this, the CRLB is often used in conjunction with maximum likelihood estimators in estimation problems [28].

Theorem 2.1 presents the scalar variant of the CRLB, based on [32]. A proof of this theorem can be found in Appendix A-1.

Theorem 2.1. (Cramér-Rao lower bound for a scalar estimator). Let $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ represent a data set, which can be seen a realization of i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$ with a distribution depending on a parameter θ . Suppose that the likelihood function $L(\theta|\mathbf{x})$ is continuously differentiable with respect to $\theta \in \mathbb{R}$ and that its support $\{\mathbf{x} | L(\theta|\mathbf{x}) > 0\}$ does not depend on θ . Then, the variance of any unbiased estimator $\hat{\theta}$ of θ satisfies:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ell(\theta|\mathbf{x})}{\partial \theta^2}\right]} := \text{CRLB}(\theta)$$

where the derivative is evaluated at the true value of θ and where the expectation is taken with respect to the likelihood function $L(\theta|\mathbf{x})$. Here, $I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ell(\theta|\mathbf{x})}{\partial \theta^2}\right]$ is called the Fisher information.

Note that the CRLB as presented in Theorem 2.1 generally depends on the parameter θ [32]. As a result, it is difficult to find the underlying CRLB for a practical estimation problem, as the parameter θ is unknown (if not, there would be no reason to estimate it). This problem, among other problems, is addressed by the Van Trees bound in Section 2-4.

In super-resolution microscopy, it is a common practice to estimate multiple model parameters simultaneously [27]. Because of this, there is a need for a multivariate CRLB. The CRLB for multivariate estimation problems is shown in Theorem 2.2. A proof of this theorem is given in Appendix A-2.

Theorem 2.2. (Cramér-Rao lower bound for a vector estimator). Let $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ represent a data set, which can be seen a realization of i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$ with a distribution depending on a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^m$. Suppose that the likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$ is continuously differentiable with respect to $\boldsymbol{\theta}$ and that its support $\{\mathbf{x} | L(\boldsymbol{\theta}|\mathbf{x}) > 0\}$ does not depend on $\boldsymbol{\theta}$.

Define the entries of the Fisher information matrix $I(\boldsymbol{\theta})$ to be:

$$[I(\boldsymbol{\theta})]_{p,q} = -\mathbb{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_p \partial \theta_q}\right]$$

where the derivatives are evaluated at the true value of $\boldsymbol{\theta}$ and where the expectation is taken with respect to the likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$. Assume that $I(\boldsymbol{\theta})$ is invertible.

Then, the covariance matrix $C_{\hat{\boldsymbol{\theta}}}$ of any unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ satisfies $C_{\hat{\boldsymbol{\theta}}} - I^{-1}(\boldsymbol{\theta}) \succeq \mathbf{0}$ (that is, $C_{\hat{\boldsymbol{\theta}}} - I^{-1}(\boldsymbol{\theta})$ is positive semi-definite).

2-3 Maximum likelihood estimation

In this section, we introduce maximum likelihood estimation (MLE) as a practical way to estimate model parameters from data.

Recall from Section 2-1 that the likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$ indicates the probability of realizing precisely the data set $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ when realizing i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$, of which the probability distribution depends on some unknown parameter

vector $\boldsymbol{\theta} \in \mathbb{R}^m$. The intuition behind maximum likelihood estimation is to maximize the probability of realizing the obtained data by an optimal choice of the parameter $\boldsymbol{\theta}$.

To obtain the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$, we thus need to solve the maximization problem of Equation (2-3) [32].

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\mathbf{x}) \quad (2-3)$$

From both an analytical and a numerical point of view, maximizing the likelihood function itself may be a bad idea. Maximizing the likelihood function by setting the first derivative to zero may be computationally intensive, as the product rule for differentiation needs to be applied repeatedly [34]. For discrete random variables, there is a second drawback. As probability mass functions take values between 0 and 1, the formulation of the likelihood function for discrete random variables, as shown in Definition 2.1, goes to zero as the amount of measurements increases, which is undesirable for numerical optimization routines.

Note that these were precisely the reasons why the log-likelihood function $\ell(\boldsymbol{\theta}|\mathbf{x})$ was introduced before. Hence, an alternative approach is to maximize the log-likelihood instead. This gives rise to the optimization problem shown in Equation (2-4). As the logarithm is a monotonically increasing function, $L(\boldsymbol{\theta}|\mathbf{x})$ and $\ell(\boldsymbol{\theta}|\mathbf{x})$ attain their extreme values for identical $\boldsymbol{\theta}$.

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}|\mathbf{x}) \quad (2-4)$$

As such, it can be concluded that the maximum likelihood framework provides a systematic way to construct estimators. Furthermore, maximum likelihood estimators have some desirable properties [34], of which we will discuss four. First of all, the maximum likelihood estimator is asymptotically unbiased. Note that nowhere in the maximum likelihood framework, we impose that the estimator should be unbiased. As such, there is no guarantee that the maximum likelihood estimator is unbiased for finite amounts of measurements. However, as the amount of data points goes to infinity, the maximum likelihood estimator becomes unbiased [34].

A second desirable property is that the maximum likelihood estimator asymptotically attains minimal (co)variance, given by the Cramér-Rao lower bound, as long as the likelihood function satisfies the assumptions as stated in Theorem 2.1 or Theorem 2.2. This is a very strong property: the Cramér-Rao lower bound is a lower bound on the variance of arbitrary unbiased estimators, and as the maximum likelihood estimator attains it, it is an efficient estimator. These two properties of asymptotic unbiasedness and asymptotic minimum variance, are highly useful in super-resolution microscopy [27]. Namely, as shown in [27], the variance of the maximum likelihood estimator can be considered to have converged to the Cramér-Rao lower bound for relatively low amounts of photons (for classical SMLM methods, this happens around 100 photons [27]). Because of this, the Cramér-Rao lower bound can be used as a measure for the localization precision of a maximum likelihood estimate.

The next important property of the maximum likelihood estimator is that its distribution is known asymptotically [38]. Suppose that $\hat{\boldsymbol{\theta}}$ denotes the maximum likelihood estimator for $\boldsymbol{\theta}$ and that the Fisher information $I(\boldsymbol{\theta})$ for the corresponding log-likelihood function is known. Then, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is normally distributed if the amount of data points goes to infinity, with mean $\boldsymbol{\theta}$ and covariance $I^{-1}(\boldsymbol{\theta})$, or $\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta}))$.

The last property of maximum likelihood estimators that we will discuss is that they satisfy the so-called invariance principle for estimators [34]. Suppose that $\hat{\theta}$ denotes the maximum likelihood estimator for θ . Furthermore, let $g(x)$ be an arbitrary, invertible function. The invariance principle for estimators then states that $g(\hat{\theta})$ is the maximum likelihood estimator for $g(\theta)$.

2-4 Van Trees inequality

In this section, we introduce the Van Trees inequality, which can be seen as an extension of the traditional CRLB to account for prior information or biased estimators.

The Cramér-Rao lower bound, as introduced in Section 2-2, is often used to estimate the localization precision of (me)SMLM methods, as it is a tight lower bound on the variance of maximum likelihood estimators if more than 100 photons are collected [27]. However, it suffers from some significant drawbacks.

The main drawback of the CRLB is that it cannot incorporate prior information on the estimand θ . As the CRLB treats θ as a deterministic parameter, it cannot include a prior distribution on θ in its lower bound on the (co)variance. For imeSMLM, where information from previous measurements is used to improve the quality of the current measurement, the CRLB is not able to tightly bound the localization precision of the maximum likelihood estimate.

This problem is addressed by the Van Trees inequality (VTI) [38], which can be seen as an extension of the Cramér-Rao lower bound into a Bayesian statistics formulation. As the VTI treats the estimand θ as a random variable with a known prior distribution, it is able to incorporate prior information into a lower bound on the estimation error. Furthermore, it is able to bound the mean squared error of biased estimators from below.

Suppose a data set $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ was obtained, which can be seen as realizations of i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$, of which the probability distribution depends on some unknown parameter $\theta \in \mathbb{R}^m$. The corresponding likelihood function for this data set is $L(\theta|\mathbf{x})$ and the corresponding log-likelihood function for this data set is $\ell(\theta|\mathbf{x})$. Suppose furthermore that we have access to a prior distribution $\lambda(\theta)$ on the parameter vector θ . We define two expectation operators, as shown in Equations (2-5) and (2-6).

$$\mathbb{E}_L[\cdot] = \int_{\mathbb{R}^n} [\cdot] L(\theta|\mathbf{x}) d\mathbf{x} \quad (2-5)$$

$$\mathbb{E}_\lambda[\cdot] = \int_{\mathbb{R}^m} [\cdot] \lambda(\theta) d\theta \quad (2-6)$$

Note that Equation (2-5) describes the expectation operator, used in evaluating the CRLB. Equation (2-6) describes the expected value with respect to the prior distribution on θ . We can combine these expectation operators, as shown in Equation (2-7).

$$\mathbb{E}_{\lambda,L} = \mathbb{E}_\lambda[\mathbb{E}_L[\cdot]] = \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} [\cdot] L(\theta|\mathbf{x}) \lambda(\theta) d\mathbf{x} d\theta \quad (2-7)$$

Note that $L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^n f_{\boldsymbol{\theta}}(x_i) = p(\mathbf{x}|\boldsymbol{\theta})$ by the definition of the likelihood function, where $p(\mathbf{x}|\boldsymbol{\theta})$ denotes the joint probability density function of \mathbf{x} , conditioned on $\boldsymbol{\theta}$. By the definition of the conditional probability, $L(\boldsymbol{\theta}|\mathbf{x})\lambda(\boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})\lambda(\boldsymbol{\theta}) = p(\mathbf{x}, \boldsymbol{\theta})$, where $p(\mathbf{x}, \boldsymbol{\theta})$ denotes the joint probability density function of \mathbf{x} . Hence, Equation (2-7) is indeed an expectation operator.

Equations (2-5) to (2-7) are used in the Van Trees inequality. The Van Trees inequality, as formulated in [42], is stated in Theorem 2.3. A proof for this theorem can be found in Appendix A-3.

Theorem 2.3. (Van Trees inequality for a scalar estimator). *Let $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ represent a data set, which can be seen a realization of i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$ with a distribution depending on a parameter $\theta \in \mathbb{R}$. Suppose that the likelihood function $L(\theta|\mathbf{x})$ is continuously differentiable with respect to θ and that its support $\{\mathbf{x} | L(\theta|\mathbf{x}) > 0\}$ does not depend on θ .*

Let a prior probability density function $\lambda(\theta)$ on the parameter θ be known. Suppose that $\lambda(\theta)$ is absolutely continuous and that λ converges to zero at the endpoints of its domain¹.

Then, the mean squared error of any estimator $\hat{\theta}$ of θ satisfies:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}_{\lambda, L}[(\hat{\theta} - \theta)^2] \geq (J_D + J_P)^{-1}$$

where the data information J_D is given by

$$\begin{aligned} J_D &= \mathbb{E}_{\lambda}[I(\theta)] = \mathbb{E}_{\lambda} \left[\mathbb{E}_L \left[\left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right)^2 \right] \right] \\ &= \mathbb{E}_{\lambda} \left[\int_{\mathbb{R}^n} \left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right)^2 L(\theta|\mathbf{x}) d\mathbf{x} \right] \end{aligned}$$

and where the prior information J_P is given by

$$J_P = I(\lambda) = \mathbb{E}_{\lambda} \left[\left(\frac{\partial \log(\lambda(\theta))}{\partial \theta} \right)^2 \right] = \int_{\mathbb{R}} \left(\frac{\partial \log(\lambda(\theta))}{\partial \theta} \right)^2 \lambda(\theta) d\theta$$

Furthermore, $(J_D + J_P)$ can be rewritten as $J = \mathbb{E}_{\lambda, L} \left[\left(\frac{\partial \log(p(\mathbf{x}, \theta))}{\partial \theta} \right)^2 \right]$, where $p(\mathbf{x}, \theta)$ denotes the joint probability density function of \mathbf{x} and θ . The scalar J is called the Bayesian information.

As with the Cramér-Rao lower bound, the Van Trees inequality can be generalized for vector estimators. The proof follows similar steps as multivariate CRLB-proof and the scalar Van Trees inequality proof, as shown in Appendices A-2 and A-3. Hence, we state the multivariate Van Trees inequality in Theorem 2.4 and we refer to [38] for its proof.

¹A special case of this is that the domain of $\lambda(\theta)$ is compact (i.e. closed and bounded by the Heine-Borel theorem) and that $\lambda(\theta)$ is zero at the boundaries of its domain.

Theorem 2.4. (Van Trees inequality for a vector estimator). Let $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ represent a data set, which can be seen a realization of i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$ with a distribution depending on a parameter $\boldsymbol{\theta} \in \mathbb{R}^m$. Suppose that the likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$ is continuously differentiable with respect to $\boldsymbol{\theta}$ and that its support $\{\mathbf{x}|L(\boldsymbol{\theta}|\mathbf{x}) > 0\}$ does not depend on $\boldsymbol{\theta}$.

Let a prior probability density function $\lambda(\boldsymbol{\theta})$ on the parameter $\boldsymbol{\theta}$ be known. Suppose that $\lambda(\boldsymbol{\theta})$ is absolutely continuous and that λ converges to zero at the endpoints of its domain.

Then, the mean squared error matrix $\text{MSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E}_{\lambda,L}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$ of any estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ satisfies:

$$\left(\mathbb{E}_{\lambda,L}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] - (J_D + J_P)^{-1} \right) \succeq 0$$

That is, $\mathbb{E}_{\lambda,L}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] - (J_D + J_P)^{-1}$ is positive semi-definite, where the data information matrix J_D is given by

$$\begin{aligned} [J_D]_{p,q} &= [\mathbb{E}_{\lambda}[I(\boldsymbol{\theta})]]_{p,q} = \mathbb{E}_{\lambda} \left[\mathbb{E}_L \left[\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_p} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_q} \right] \right] \\ &= \mathbb{E}_{\lambda} \left[\int_{\mathbb{R}^n} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_p} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_q} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \right] \end{aligned}$$

and where the prior information matrix J_P is given by

$$[J_P]_{p,q} = [I(\lambda)]_{p,q} = \mathbb{E}_{\lambda} \left[\frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \theta_p} \frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \theta_q} \right] = \int_{\mathbb{R}^m} \frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \theta_p} \frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \theta_q} \lambda(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

Furthermore, $(J_D + J_P)$ can be rewritten as J with $J_{p,q} = \mathbb{E}_{\lambda,L} \left[\frac{\partial \log(p(\mathbf{x}, \boldsymbol{\theta}))}{\partial \theta_p} \frac{\partial \log(p(\mathbf{x}, \boldsymbol{\theta}))}{\partial \theta_q} \right]$, where $p(\mathbf{x}, \boldsymbol{\theta})$ denotes the joint probability density function of \mathbf{x} and $\boldsymbol{\theta}$. The matrix J is called the Bayesian information matrix.

From Theorems 2.3 and 2.4, it can be seen that the VTI does not depend on the parameter vector $\boldsymbol{\theta}$ anymore, as it is integrated away using the prior distribution. Hence, the Van Trees inequality successfully incorporates prior information into its bound on the mean squared error, as required.

One caveat in using the Van Trees inequality is that the expectation with respect to the prior distribution is generally difficult to compute. While this expectation can be computed analytically for special cases of the log-likelihood and the prior distribution, one generally resorts to numerical evaluation of the expectations [43, 47]. We will return to this in Chapter 3, where we present importance sampling Monte Carlo integration as a numerical method to estimate the value of integrals.

Chapter 3

Precision in iterative modulation enhanced single-molecule localization microscopy

In this chapter, the findings of my thesis are discussed in the form of a manuscript. Furthermore, three supplementary notes are appended in which important results are derived.

Precision in iterative modulation enhanced single-molecule localization microscopy

Abstract Modulation enhanced single-molecule localization microscopy (meSMLM) methods achieve increased resolution by combining SMLM techniques with patterned illumination. By using prior information to refine emitter position estimates, iterative meSMLM (imeSMLM) methods as iterative MINFLUX are able to locally improve resolution. The Cramér-Rao lower bound (CRLB) cannot be used to accurately estimate the localization precision of imeSMLM methods, because it treats estimands as deterministic unknowns. By treating estimands as random variables with a known prior distribution, the Van Trees inequality (VTI) can be used to bound the localization precision of imeSMLM methods from below. An imeSMLM method is considered, where the positions of standing wave illumination patterns are controlled over the course of multiple iterations. Using the VTI, we derive a fundamental limit on the localization precision of imeSMLM methods that make use of standing wave illumination patterns. This limit shows that in the absence of background, the information content of signal photons increases exponentially as a function of the iteration count. Using Monte Carlo simulations, the maximally achievable localization precision for different illumination pattern placement control strategies was evaluated. The VTI allows to assess the performance of pattern placement control strategies and is therefore a promising method for optimal control of imeSMLM methods.

3-1 Introduction

Superresolution fluorescence microscopy methods, such as stimulated emission depletion (STED) [18, 19], structured illumination microscopy (SIM) [44, 45], photoactivated localization microscopy (PALM) [17] and stochastic optical reconstruction microscopy (STORM) [23, 24], are able to circumvent the diffraction limit of light. STED and SIM use patterned illumination to increase the localization precision. On the other hand, single-molecule localization microscopy (SMLM) methods as PALM and STORM improve resolution over the diffraction limit by sparsely activating fluorescent emitters and exploiting sparsity in the emission light to estimate emitter positions with increased precision.

Recently, methods were proposed that make use both the illumination and the emission path of a microscope, to further increase the localization precision. These modulation enhanced SMLM (meSMLM) methods use patterned illumination to sparsely activate emitters in a sample, after which emitter positions are estimated from the sparsity in the emission light. Methods as SIMFLUX [7], SIMPLE [30] and repetitive optical selective exposure (ROSE) [31] use a standing wave intensity pattern for the illumination, while MINFLUX [5] uses a doughnut-shaped intensity pattern. Axial resolution was increased through modulated localization (ModLoc) [29, 46], which uses a pattern with structure in lateral and axial directions.

Localization precision can be increased locally around the emitter by iteratively adapting meSMLM methods, based on prior information on the emitter position that was generated

by previous measurements. These methods are called iterative meSMLM (imeSMLM) methods. In [6], an iterative variant of MINFLUX is discussed, where the position of an emitter is estimated through triangulation with doughnut-shaped intensity patterns. This estimate and its localization uncertainty are used as prior information to reposition and shrink the triangulation region, after which the emitter position is estimated again. This procedure locally improves precision in the neighborhood of the emitter. Furthermore, it is argued that increasing the iteration count is preferred over increasing the amount of signal photons per iteration, as the information content of signal photons increases over the course of iterations.

To characterize the localization precision of (me)SMLM methods, the Cramér-Rao lower bound (CRLB) is often used [28]. Under mild assumptions on the likelihood function of the acquired data, the CRLB states that for any unbiased estimator $\hat{\theta}$ of the parameter vector θ , it holds that $C_{\hat{\theta}} - I^{-1}(\theta)$ is positive semi-definite, where $C_{\hat{\theta}}$ denotes the estimator covariance and $I^{-1}(\theta)$ denotes the Fisher information [32]. In a scalar scenario, the CRLB thus bounds the estimator variance from below. As maximum likelihood estimators attain the CRLB asymptotically, the CRLB can be used to estimate the localization precision of SMLM methods. It is shown in [27] that in SMLM, the covariance of the maximum likelihood estimator approximately converges to the CRLB for 100 or more signal photons.

However, the CRLB cannot be used to accurately estimate the localization precision of imeSMLM methods. As imeSMLM methods use prior information to refine emitter position estimates over the course of multiple iterations, a suitable error bound should be able to incorporate prior information. The CRLB treats estimands as deterministic variables and as such, it is not able to incorporate prior information into the localization precision. For this reason, the CRLB fails to estimate how prior information affects the localization precision.

In this paper, we use the Van Trees inequality (VTI) as a Bayesian alternative to the CRLB if prior information on the estimands, such as the emitter position, is available. The VTI treats estimands as random variables with a known prior distribution, which makes it suited to bound the localization precision of imeSMLM methods from below. Specifically, we use the VTI to develop a fundamental limit on the localization precision of imeSMLM methods, where standing wave intensity patterns are used for the localization. Furthermore, we simulate the effects of illumination pattern position control on the localization precision. We show that in the absence of background, the information content of signal photons increases exponentially as a function of the iteration count. In addition, we are able to evaluate the performance of pattern position control strategies using the VTI as a performance metric.

The remainder of this paper is structured as follows. First, an imeSMLM method that uses standing wave illumination patterns is modeled in Section 3-2. Then in Section 3-3, the VTI is applied to this image formation model. Section 3-4 discusses the obtained results and the paper will be concluded in Section 3-5. Lastly, we elaborate on further applications of the VTI in Section 3-6.

3-2 Image formation model

In this section, we discuss the image formation model used to estimate the localization precision in this text. A detailed derivation of this model is included in Supplementary Note 1.

We consider an epifluorescence microscopy setup, where a single emitter is illuminated with patterned intensity profiles during M iterations and for which read-out noise is assumed to be negligible. As in [7], we consider standing wave intensity patterns with controllable spatial phase shifts $\phi_{x,k}, \phi_{y,k}$ in two orthogonal orientations. In each iteration $k \in \{0, 1, \dots, M-1\}$ of the localization procedure, one x -oriented pattern $P_{x,k}(x, y, \phi_{x,k})$ and one y -oriented pattern $P_{y,k}(x, y, \phi_{y,k})$ are used to illuminate the sample. These patterns can be described by:

$$\begin{cases} P_{x,k}(x, y, \phi_{x,k}) = \frac{1}{4M} [1 + m \cos(\omega x - \phi_{x,k})] \\ P_{y,k}(x, y, \phi_{y,k}) = \frac{1}{4M} [1 + m \cos(\omega y - \phi_{y,k})] \end{cases} \quad (3-1)$$

where $m \in [0, 1]$ denotes the modulation contrast of the pattern and ω (rad/m) denotes the spatial frequency of the pattern.

The point spread function (PSF) $h(x, y)^2$ is assumed to be Gaussian:

$$h(x, y)^2 = \frac{1}{2\pi\sigma_{PSF}^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_{PSF}^2}\right) \quad (3-2)$$

where σ_{PSF} (m) denotes the standard deviation of the Gaussian PSF.

In the absence of read-out noise, the photons collected on position (x, y) of the camera during iteration k can be modeled as realizations of a Poisson process with mean $\mu_k(x, y)$ [27], given by:

$$\mu_k(x, y) = \theta_I [P_{x,k}(\theta_x, \theta_y, \phi_{x,k}) + P_{y,k}(\theta_x, \theta_y, \phi_{y,k})] h(x - \theta_x, y - \theta_y)^2 + \theta_b B_k(x, y) \quad (3-3)$$

where coordinates (θ_x, θ_y) (m) denote the emitter position, θ_I (photons) and θ_b (photons/m²) respectively denote the expected signal photon count and the expected background photon count per area over all iterations in case an emitter is illuminated with maximum intensity light. These estimands are collected in the parameter vector $\boldsymbol{\theta} = [\theta_x, \theta_y, \theta_I, \theta_b]^T$. Furthermore, B_k (m²) depends on the PSF and the illumination pattern (see Supplementary Note 1) and accounts for the effect of imperfect illumination on recorded background. For the illumination pattern described in Equation (3-1), $B_k(x, y)$ is given by:

$$B_k(x, y) = \frac{1}{2M} + \frac{m}{4M} \exp\left(\frac{\omega^2 \sigma_{PSF}^2}{2}\right) \cos(\omega x - \phi_{x,k}) + \frac{m}{4M} \exp\left(\frac{\omega^2 \sigma_{PSF}^2}{2}\right) \cos(\omega y - \phi_{y,k}) \quad (3-4)$$

To model the effects of discretization of the image due to finite camera pixel size, Equation (3-3) has to be integrated over the pixel area. We consider a rectangular camera pixel array without dead space between pixels, with n_x pixels in x -direction and n_y pixels in y -direction. In addition, let all pixels be rectangular, with size Δx in x -direction and Δy in y -direction and with the center of pixel $i \in \{1, 2, \dots, n_x n_y\}$ located at (x_i, y_i) . Equation (3-3) then reads as follows:

$$\mu_{i,k} = \theta_I [P_{x,k}(\theta_x, \theta_y, \phi_{x,k}) + P_{y,k}(\theta_x, \theta_y, \phi_{y,k})] E_x(x_i, \theta_x) E_y(y_i, \theta_y) + \theta_b B_{i,k} \quad (3-5)$$

where $B_{i,k}$ (m^2) depends on the detector pixel area, the PSF and the illumination pattern (see Supplementary Note 1) and accounts for the effects of finite camera pixel area and imperfect illumination on recorded background. $B_{i,k}$ and $E_u(u_i, \theta_u)$ (with $u = x$ or $u = y$) are given by:

$$B_{i,k} = \frac{\Delta x \Delta y}{2M} + \frac{m \Delta x}{2M\omega} \exp\left(-\frac{\omega^2 \sigma^2}{2}\right) \cos(\omega y_i - \phi_{x,k}) \sin\left(\frac{\omega \Delta y}{2}\right) \\ + \frac{m \Delta y}{2M\omega} \exp\left(-\frac{\omega^2 \sigma^2}{2}\right) \cos(\omega x_i - \phi_{y,k}) \sin\left(\frac{\omega \Delta x}{2}\right) \quad (3-6)$$

$$E_u(u_i, \theta_u) = \frac{1}{2} \text{erf}\left(\frac{u_i - \theta_u + \frac{\Delta u}{2}}{\sigma_{PSF} \sqrt{2}}\right) - \frac{1}{2} \text{erf}\left(\frac{u_i - \theta_u - \frac{\Delta u}{2}}{\sigma_{PSF} \sqrt{2}}\right) \quad (3-7)$$

Using this image formation model, the CRLB for individual iterations can be computed. It is shown in [7, 27], that for a parameter vector θ , the entries of the Fisher information matrix $I_k(\theta)$ during each iteration k are given by:

$$[I_k(\theta)]_{p,q} = \sum_{i=1}^N \frac{1}{\mu_{i,k}} \frac{\partial \mu_{i,k}}{\partial \theta_p} \frac{\partial \mu_{i,k}}{\partial \theta_q} \quad (3-8)$$

3-3 Van Trees inequality for iterative localization precision

In this section, we discuss the Van Trees inequality (VTI) and its application to imeSMLM methods. In Subsection 3-3-1, we introduce the Van Trees inequality as a way to bound the mean squared error of estimators from below in case prior information is available. We show in Subsection 3-3-2 how the VTI can be applied to compute the maximally achievable localization precision of imeSMLM methods. In Subsection 3-3-3, we elaborate on how the pattern positions $\phi_{x,k}$ and $\phi_{y,k}$ should be chosen to maximize the information content per collected signal photon. Lastly, in Subsection 3-3-4, we explain how importance sampling Monte Carlo integration can be used to compute numerical estimates of the VTI.

3-3-1 Van Trees inequality

In imeSMLM, prior information on the emitter position plays an important role in maximizing the information content of signal photons. The CRLB cannot include prior information on estimand vector θ , as it treats estimands as deterministic unknowns (i.e. they do not have a probability distribution). As prior information cannot be incorporated into the CRLB, it will fail to accurately estimate the localization precision of these methods.

The VTI [38, 42] is a Bayesian variant of the CRLB. By treating the estimand vector θ as a random variable with a known prior distribution, it can incorporate prior information into

the localization precision. A second advantage is that the VTI can bound the mean squared error of biased estimators from below, while the CRLB only bounds the localization precision of unbiased estimators.

The VTI can be used to bound the mean squared error of any estimator $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$ from below if a data set $\mathbf{x} \in \mathbb{R}^n$, of which the distribution depends on $\boldsymbol{\theta}$, and a prior distribution $\lambda_{k-1}(\boldsymbol{\theta})$ on the parameter vector are available in each iteration k of the localization procedure [38, 42]. Under regularity conditions on the likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$ and the prior distribution $\lambda_{k-1}(\boldsymbol{\theta})$ (see Supplementary Note 2), the mean squared error matrix $\text{MSE}(\hat{\boldsymbol{\theta}}_k) = \mathbb{E}_{\lambda_{k-1}} \left[\int_{\mathbb{R}^n} (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^T L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \right]$ of any estimator $\hat{\boldsymbol{\theta}}_k$ of $\boldsymbol{\theta}$ during iteration k satisfies:

$$\left(\mathbb{E}_{\lambda_{k-1}} \left[\int_{\mathbb{R}^n} (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^T L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \right] - (J_{D,k} + J_{P,k})^{-1} \right) \succeq 0 \quad (3-9)$$

That is, $\mathbb{E}_{\lambda_{k-1}} \left[\int_{\mathbb{R}^n} (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^T L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \right] - (J_{D,k} + J_{P,k})^{-1}$ is positive semi-definite, where the data information matrix $J_{D,k}$ is given by:

$$[J_{D,k}]_{p,q} = \mathbb{E}_{\lambda_{k-1}} [[I_k(\boldsymbol{\theta})]_{p,q}] \quad (3-10)$$

where the Fisher information $I_k(\boldsymbol{\theta})$ is given by Equation (3-8). The prior information matrix $J_{P,k}$ is given by:

$$[J_{P,k}]_{p,q} = [I(\lambda_{k-1})]_{p,q} = \mathbb{E}_{\lambda_{k-1}} \left[\frac{\partial \log(\lambda_{k-1}(\boldsymbol{\theta}))}{\partial \theta_i} \frac{\partial \log(\lambda_{k-1}(\boldsymbol{\theta}))}{\partial \theta_j} \right] \quad (3-11)$$

3-3-2 Localization precision for imeSMLM

The VTI can be used to estimate the maximally achievable localization precision of imeSMLM methods in case a maximum likelihood estimator or a maximum a posteriori estimator is used for localization. To be able to formulate the VTI, a prior distribution on the parameter vector $\boldsymbol{\theta} = [\theta_x, \theta_y, \theta_I, \theta_b]^T$ needs to be chosen. It is shown in [38] that the maximum likelihood estimator asymptotically follows a multivariate normal distribution, with mean $\boldsymbol{\theta}$ and covariance $I^{-1}(\boldsymbol{\theta})$. Alternatively, one can say that for an increasing amount of signal photons, the maximum likelihood estimator becomes more unbiased and attains minimum covariance given by the CRLB.

Results from [27] show that in SMLM, the maximum likelihood estimator approximately attains these properties for 100 or more signal photons. For this reason, errors made in the localization will mostly be a result of estimator covariance, which means that the mean squared error and the covariance are approximately equal for high signal photon counts. Consequently, it can be expected that the mean squared error bound provided by the Van Trees inequality is dominated by the effect of covariance for high signal photon counts as well. This means that the VTI is an indication of the maximally achievable localization precision.

By combining maximum likelihood estimation or maximum a posteriori estimation with the VTI, the localization precision of imeSMLM methods can be simulated. This iterative procedure is illustrated in Figure 3-1.

	Data		Prior		Estimator and prior distribution	
Iteration 0: CRLB	$I_0(\theta)$	+	-	→	$\hat{\theta}_0 \sim \mathcal{N}(\theta, I_0^{-1}(\theta))$	
Iteration 1: VTI	$J_{D,1} = \mathbb{E}_{\lambda_0}[I_1(\theta)]$	+	$J_{P,1} = I(\lambda_0)$	→	$\hat{\theta}_1 \sim \mathcal{N}(\theta, (J_{D,1} + J_{P,1})^{-1})$	← $\lambda_0(\theta)$
Iteration 2: VTI	$J_{D,2} = \mathbb{E}_{\lambda_1}[I_2(\theta)]$	+	$J_{P,2} = I(\lambda_1)$	→	$\hat{\theta}_2 \sim \mathcal{N}(\theta, (J_{D,2} + J_{P,2})^{-1})$	← $\lambda_1(\theta)$
Iteration 3: VTI	$J_{D,3} = \mathbb{E}_{\lambda_2}[I_3(\theta)]$	+	$J_{P,3} = I(\lambda_2)$	→	$\hat{\theta}_3 \sim \mathcal{N}(\theta, (J_{D,3} + J_{P,3})^{-1})$	← $\lambda_2(\theta)$
⋮	⋮	+	⋮	→	⋮	← $\lambda_3(\theta)$

Figure 3-1: Simulation of the localization precision of imeSMLM using the Van Trees inequality (VTI). In the first iteration, the Fisher information matrix $I_0(\theta)$ is computed. The resulting maximum likelihood estimate is asymptotically $\mathcal{N}(\theta, I_0^{-1}(\theta))$ distributed, and we choose this as a prior on the next iteration. In every following iteration, the VTI is evaluated using the prior, resulting in the data information matrix $J_{D,k}$ and the prior information matrix $J_{P,k}$ in iteration k . As the next prior, we choose $\mathcal{N}(\theta, (J_{D,k} + J_{P,k})^{-1})$. This iteration continues until the maximum amount of iterations M is reached.

In the first iteration, the Cramér-Rao lower bound $I_0^{-1}(\theta)$ is evaluated. The corresponding maximum likelihood estimator $\hat{\theta}_0$ will approximately be Gaussian distributed, with mean θ and covariance $I_0^{-1}(\theta)$.

An important realization is that the approximate distribution of the maximum likelihood estimator can be used as a prior distribution for the next Van Trees evaluation. In iteration 1, we take the prior distribution $\lambda_0(\theta)$ to be the probability density function of a multivariate Gaussian distribution, with covariance $I_0^{-1}(\theta)$. As the asymptotic mean θ of the maximum likelihood estimator is unknown, we take the maximum likelihood estimate $\hat{\theta}_0$ as the mean of the prior distribution. We then evaluate $J_{D,1} = \mathbb{E}_{\lambda_0}[I_1(\theta)]$ and $J_{P,1} = I(\lambda_0)$. We can say that the new maximum likelihood estimate $\hat{\theta}_1$ is approximately Gaussian distributed, with mean θ and covariance $(J_{D,1} + J_{P,1})^{-1}$ and use this to formulate the new prior distribution $\lambda_2(\theta)$.

In each new iteration k , we take the prior distribution $\lambda_{k-1}(\theta)$ to be the probability density function of a multivariate Gaussian distribution, with mean $\hat{\theta}_{k-1}$ and covariance $(J_{D,k-1} + J_{P,k-1})^{-1}$. We then evaluate $J_{D,k} = \mathbb{E}_{\lambda_{k-1}}[I_k(\theta)]$ and $J_{P,k} = I(\lambda_{k-1})$. We can now say that the new maximum likelihood estimate $\hat{\theta}_k$ is approximately Gaussian distributed, with mean θ and covariance $(J_{D,k} + J_{P,k})^{-1}$ and we use this to formulate the new prior distribution $\lambda_k(\theta)$. This continues, until M iterations are completed.

3-3-3 Choice of pattern positions

Using prior information on the emitter position, the pattern positions $\phi_{x,k}$ and $\phi_{y,k}$ can be selected to maximize the information content of signal photons.

If we keep the amount of photons that are retrieved during a single iteration constant during all iterations, the patterns that minimize the localization precision VTI maximize the information content of signal photons. This procedure is illustrated in Figure 3-2.



	Optimization	Illumination	Precision	
Iteration 0	-		Parameter estimate: $\hat{\theta}_0$ CRLB: $I_0^{-1}(\theta)$	Prior 0
Iteration 1	For optimal expected precision, choose pattern positions $\phi_{x,1}$ and $\phi_{y,1}$		Parameter estimate: $\hat{\theta}_1$ VTI: $(J_{D,1} + J_{P,1})^{-1}$	Prior 1
\vdots	\vdots	\vdots	\vdots	

Figure 3-2: Optimization procedure to select pattern positions that maximize the information content of signal photons. In the initialization step, no prior information is available and pattern positions $\phi_{x,0} = \phi_{y,0} = 0$ are chosen. The resulting CRLB is used to determine a prior for the next iteration. For all further iterations k , $\phi_{x,k}$ and $\phi_{y,k}$ are chosen to minimize the localization precision VTI for a fixed amount of photons, after which the realized VTI is used to determine a prior for the next iteration.

In the initialization step, no prior information is available. Therefore, the pattern positions are chosen to be $\phi_{x,0} = \phi_{y,0} = 0$ by default. A parameter estimate $\hat{\theta}_0$ is obtained and the CRLB can now be computed, which results in a $\mathcal{N}(\hat{\theta}_0, I_0^{-1}(\theta))$ prior distribution for the next iteration.

To select patterns in a further iterations k , pattern positions $\phi_{x,k} = \phi_{y,k}$ should be chosen to minimize the mean squared error given by the VTI. The amount of signal photons is fixed per iteration and therefore a minimal mean squared error corresponds to a maximum amount of information contained in signal photons. Consequently, this procedure allows us to locally increase the localization precision around the emitter position.

The pattern positions that minimize the localization precision in both the x - and y -directions are given by:

$$\begin{cases} \phi_{x,k} = \omega \hat{\theta}_{x,k} - \pi \\ \phi_{y,k} = \omega \hat{\theta}_{y,k} - \pi \end{cases} \quad (3-12)$$

or 2π -multiples of these positions. That is, for the information content of signal photons to be maximal, the intensity minimum of the standing wave patterns should be placed on the current position estimate.

This result can easily be explained. Suppose that the pattern positions are chosen, such that each intensity minimum is placed exactly on the emitter position. For simplicity, let us also assume that the modulation contrast $m = 1$, such that the intensity minimum has zero intensity. As the emitter is illuminated with zero intensity light, it will not emit any photons. We therefore need to wait infinitely long to receive the fixed amount of signal photons.

For this to happen, two scenarios are possible. Either the emitter is located perfectly in the intensity minimum, or nothing is located in the intensity minimum. These scenarios are equivalent for the CRLB, as it does not incorporate prior knowledge on the emitter position into the localization precision. As no emitter signal is recorded during this iteration, the CRLB stays equal.

On the other hand, the VTI is able to differentiate between these scenarios. Using the prior information, the VTI expresses its confidence in that there is indeed an emitter located in this position, resulting in a decrease of the mean squared error. The prior information thus adds a significant amount of information content to signal photons in case an emitter is illuminated with (near-)zero intensity light.

If the objective is solely to maximize the information content of signal photons, the robustness of the iterative localization procedure could be destroyed. This is indeed the case for the pattern positions of Equation (3-12). If the current position estimate is imprecise, intensity minima will be placed away from the true emitter position and the newly obtained prior distribution will add little information.

To increase the robustness of the iterative localization method, the intensity minima of two patterns ϕ_k^+ and ϕ_k^- can be placed symmetrically around the current estimate of the emitter position. In each iteration, the distance between the intensity minima can be decreased, locally improving the localization precision around the current estimate of the emitter position.

In this text, we will explore two variants of this approach. First, we introduce a pattern position control strategy where illumination minima are placed symmetrically around the current estimate of the emitter position, in which the distance between the emitter position and an intensity minimum scales with the localization precision of the previous iteration. This control strategy is given by:

$$\begin{cases} \phi_{x,k}^\pm = \omega(\hat{\theta}_{x,k-1} \pm \alpha\sigma_{x,k-1}) - \pi \\ \phi_{y,k}^\pm = \omega(\hat{\theta}_{y,k-1} \pm \alpha\sigma_{y,k-1}) - \pi \end{cases} \quad (3-13)$$

In the pattern positions of Equation (3-13), the localization precision σ_{k-1} estimated in the previous iteration is used to place the intensity minima of current patterns in the proximity of the estimated emitter position. Therefore, robustness guarantees can be given under the assumption that the estimator distribution is approximately Gaussian. We will explore the performance using $\alpha = 1$, $\alpha = 3$ and $\alpha = 6$, which respectively entrap the true emitter position between the intensity minima with 68.3%, 99.7% and 99.9% probability during each iteration. As a drawback of this method, the localization precision σ_{k-1} must be computed before a new pattern can be placed, thus requiring online adaptations to the pattern positions. To circumvent the need for online adaptations, we propose a second pattern position control strategy where illumination minima are placed symmetrically around the current estimate of the emitter position and where the distance between the emitter position and an intensity minimum is given by a predetermined constant. This control strategy is given by:

$$\begin{cases} \phi_{x,k}^\pm = \omega\hat{\theta}_{x,k-1} \pm \frac{\pi}{\beta^k} - \pi \\ \phi_{y,k}^\pm = \omega\hat{\theta}_{y,k-1} \pm \frac{\pi}{\beta^k} - \pi \end{cases} \quad (3-14)$$

In the pattern positions of Equation (3-14), a predefined factor $\frac{\pi}{\beta^k}$ is used to position intensity minima around the emitter position estimate. As these pattern positions are known beforehand, they are suited for offline pattern control. However, robustness guarantees cannot be given for arbitrary model parameters. We will explore the performance using $\beta = 2$ and $\beta = 3$.

3-3-4 Numerical evaluation

In many practical scenarios, Equation (3-10) is difficult to compute analytically or it results in complicated expressions for the localization precision. Aside from certain special cases (of which two will be discussed in Subsection 3-4-1), one generally resorts to numerical methods to evaluate the VTI [43, 47].

Due to the fact that we use a Gaussian prior, an appropriately chosen numerical quadrature should be able to cover an infinite domain of integration, to ensure convergence to the true data information matrix. Furthermore, as we consider a multivariate integral over \mathbb{R}^4 , the curse of dimensionality may cause slow convergence. Therefore, numerical integration methods that efficiently distribute function evaluations are preferred.

As we can easily sample from a multivariate Gaussian distribution, importance sampling Monte Carlo integration (IMC) is a suitable method to evaluate Equation (3-10). In short, Monte Carlo integration methods sample from a known probability distribution. These samples are used to evaluate the integrand, after which the value of the integral is estimated by averaging the obtained integrand values [48]. If the domain of the sampling probability distribution equals or contains the domain of integration, asymptotic convergence to the integral value is guaranteed under the law of large numbers [49].

In conventional Monte Carlo integration, a uniform distribution is used as a sampling probability. For the reasons named earlier, this is not sufficient for our purposes. In IMC, the sampling probability distribution is chosen to maximize the similarity with the integrand. By choosing a sampling distribution with an infinite domain, integrals over infinite domains can be evaluated. Furthermore, by distributing sampling points such that important regions of the integrand are sampled more often, the variance of the IMC estimator can be reduced with respect to the conventional Monte Carlo estimator.

Equations (3-10) and (3-11) show that the Van Trees bound requires the following two expectations to be computed:

$$[J_{D,k}]_{p,q} = \mathbb{E}_{\lambda_{k-1}} [[I_k(\boldsymbol{\theta})]_{p,q}] = \int_{\mathbb{R}^m} [I_k(\boldsymbol{\theta})]_{p,q} \lambda_{k-1}(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3-15)$$

$$[J_{P,k}]_{p,q} = \mathbb{E}_{\lambda_{k-1}} \left[\frac{\partial \log(\lambda_{k-1}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_p} \frac{\partial \log(\lambda_{k-1}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_q} \right] = \int_{\mathbb{R}^m} \frac{\partial \log(\lambda_{k-1}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_p} \frac{\partial \log(\lambda_{k-1}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_q} \lambda_{k-1}(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3-16)$$

By randomly sampling data points $[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{n_{MC}}]^T$ from the i.i.d. random variables $[\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \dots, \boldsymbol{\Theta}_{n_{MC}}]^T$, each of which has the density function $\lambda_{k-1}(\boldsymbol{\theta})$, the IMC-estimates $\hat{J}_{D,k}$ and $\hat{J}_{P,k}$ of Equations (3-17) and (3-18) can be used to estimate $J_{D,k}$ and $J_{P,k}$, respectively.

$$[\hat{J}_{D,k}]_{p,q} = \frac{1}{n_{MC}} \sum_{s=1}^{n_{MC}} [I_k(\boldsymbol{\theta}_s)]_{p,q} \quad (3-17)$$

$$[\hat{J}_{P,k}]_{p,q} = \frac{1}{n_{MC}} \sum_{s=1}^{n_{MC}} \left(\left[\frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_i} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_s} \left[\frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_j} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_s} \right) \quad (3-18)$$

For the simulations of Section IV, $n_{MC} = 50000$ Monte Carlo samples were used. It was found empirically that over the entire range of acquired signal photon counts and iterations, the sample standard deviation of the evaluated integrand values lies below 2% of its sample mean, implying convergence of the Monte Carlo integration for the given amount of samples.

3-4 Results and discussion

In this section, we present the theoretical and numerical results of our study. In Subsection 3-4-1, analytical expressions for fundamental limits on the localization precision are described and analyzed. Subsection 3-4-2 gives an overview of the simulations and the used model parameters. In Subsection 3-4-3, the effect of the amount of iterations and pattern positioning on the localization precision is explored. Lastly, in Subsection 3-4-4, the effect of the initial pattern position and estimation errors on the localization precision are simulated.

3-4-1 Theoretical limit

Under some assumptions on the image formation model, analytical expressions for the localization precision can be derived using the VTI. We limit ourselves to one-dimensional localization. Furthermore, as discussed in Section 3-3-3, we consider two sinusoidal patterns ϕ_k^+ and ϕ_k^- of which the intensity minima have been placed symmetrically around the estimate of the emitter position.

Lastly, we assume the expected signal photon count θ_I is a known constant (i.e. we do not estimate it), we ignore background such that $\theta_b = 0$ and we assume that the modulation is perfect, such that $m = 1$. Under these three assumptions, the derived theoretical bound serves as a fundamental limit on the localization precision. That is, the localization precision of the imeSMLM method of Section II can only be worse than the fundamental limit, due to estimation uncertainty on θ_I , the effects of background and imperfect modulation. The effects of estimation uncertainty on θ_I and background on the localization precision will be explored through numerical simulations in Sections 3-4-3 and 3-4-4.

We will present results in two scenarios. In the first case, we disregard the effect of discretization due to the finite size of camera pixels. As a result, this expression represents the maximally achievable localization precision of an imeSMLM method where sinusoidal intensity patterns are used, given a Gaussian PSF and a sequence of pattern positions. The localization precision is given by (see Supplementary Note 3):

$$\sigma_{x,k} \geq \sigma_{x,k-1} \left(2 + \frac{\theta_I \omega^2 \sigma_{x,k-1}^2}{4M} \left(1 - (\cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^+) + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^-)) \exp\left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2}\right) \right) \right. \\ \left. + \frac{\theta_I \sigma_{x,k-1}^2}{4M \sigma_{PSF}^2} \left(1 + (\cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^+) + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^-)) \exp\left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2}\right) \right) \right)^{-\frac{1}{2}} \quad (3-19)$$

If the pattern positions of Equation (3-13) are substituted into Equation (3-19), we obtain:

$$\sigma_{x,k} \geq \sigma_{x,k-1} \left(2 + \frac{\theta_I \omega^2 \sigma_{x,k-1}^2}{4M} \left(1 + 2 \cos(\omega \alpha \sigma_{x,k-1}) \exp \left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2} \right) \right) + \frac{\theta_I \sigma_{x,k-1}^2}{4M \sigma_{PSF}^2} \left(1 - 2 \cos(\omega \alpha \sigma_{x,k-1}) \exp \left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2} \right) \right) \right)^{-\frac{1}{2}} \quad (3-20)$$

From Equation (3-20), several conclusions can be drawn. It can be seen that the contribution of the expected signal photon count θ_I to the localization precision grows exponentially as $\sigma_{x,k-1}$ decreases. This implies that the information content per signal photon grows as the amount of iterations increases. In the ideal scenario, it is thus favorable to increase the amount of iterations indefinitely.

However, increasing the amount of iterations results in a lower amount of signal photons per iteration, lowering the signal-to-background ratio in each iteration. The exponential scaling may therefore be destroyed by background. In practice, it thus makes sense to limit the amount of iterations. Furthermore, we assume here that the mechanical resolution of illumination positioning system is not limiting, such that every illumination pattern position between $-\pi$ and π can be reached. In practice, this becomes increasingly difficult for small $\sigma_{x,k-1}$, preventing the exponential limit to be reached for high amounts of iterations.

In the second scenario, the effects of image discretization by the camera are investigated. In addition to all previous assumptions, a camera with a single, infinitely large pixel is considered. Under this additional assumption, the localization precision is given by (see Supplementary Note 3):

$$\sigma_{x,k} \geq \sigma_{x,k-1} \left(2 + \frac{\theta_I \omega^2}{4M} \sigma_{x,k-1}^2 \left(1 - (\cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^+) + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^-)) \exp \left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2} \right) \right) \right)^{-\frac{1}{2}} \quad (3-21)$$

By comparing Equations (3-19) and (3-21), it becomes apparent that discretization only affects the influence of the PSF on the localization precision. In the worst-case discretization of Equation (3-21), the terms involving the standard deviation of the PSF have disappeared from the expression, causing a precision loss. This is mainly a problem for earlier iterations where $\sigma_{x,k-1}$ is large. However, as $\sigma_{x,k-1}$ decreases, the exponential scaling of the localization precision is retrieved.

3-4-2 Simulations and parameter values

To evaluate how the VTI increases the information content of signal photons with respect to the CRLB in an SMLM scenario, four simulations were done. These simulations explore the effects of the amount of iterations, the choice of pattern positioning, the effect of the initial pattern placement and the effect of estimation errors on the localization precision. As orthogonal patterns are used, the localization precision in x - and y -direction is equal. We will therefore only present the localization precision in the x -direction. In Figures 3-3 and

3-4, the effects of pattern positioning and the iteration count on the localization precision are simulated. The sensitivity of localization precision to the initial pattern position and estimation errors is explored in Figures 3-5 and 3-6.

An overview of the used model parameters is contained in Table 3-1. These model parameters were adopted from [7] are considered to be representative of an imeSMLM experiment where standing wave intensity patterns are used to illuminate the sample. The modulation contrast was set to 1 to isolate the effects of background and image discretization from the effect of imperfect modulation and to accelerate convergence of the IMC estimation. In all simulations, the acquired amount of signal photons was kept constant at 2000 photons and signal photons were distributed equally over the iterations.

Table 3-1: Model parameters used in the localization precision simulations for sinusoidal illumination with a Gaussian point spread function. Parameters were based on those used in [7].

Quantity	Symbol	Value
Wavelength of emission light	λ_{em}	680 nm
Emitter position (x -direction)	θ_x	0 nm
Emitter position (y -direction)	θ_y	0 nm
Expected background count	θ_b	8 photons/pixel
Pixel size (x -direction)	Δx	65 nm
Pixel size (y -direction)	Δy	65 nm
Amount of pixels in x -direction	n_x	11 pixels
Amount of pixels in y -direction	n_y	11 pixels
Numerical aperture	NA	1.49
Standard deviation of PSF in x - and y -directions	σ_{PSF}	114 nm
Pattern frequency	ω	$25.8 \cdot 10^6$ rad/m
Modulation contrast	m	1

3-4-3 Effect of iterations and pattern positioning

From the theoretical limit discussed in Subsection 3-4-1, it was found that illumination pattern placement control can exponentially increase the information content of signal photons, making an increase of the iteration count preferred over an increase of the amount of photons per iteration. The effects of pattern positioning and the iteration count on the localization precision will be simulated in this subsection.

In Figure 3-3, the information content of signal photons over the course of iterations is explored for different pattern position placement strategies. The pattern positions described by Equations (3-13), (3-14) with $\alpha = 3$ and $\beta = 3$ were used during 1, 3 and 5 iterations and the resulting localization precision was evaluated using the Van Trees inequality. These results are compared with the theoretical limit of Equation (3-19). The emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and 3 iterations were simulated. Intensity minima were placed symmetrically around the true emitter position, to eliminate the effect of estimation errors.

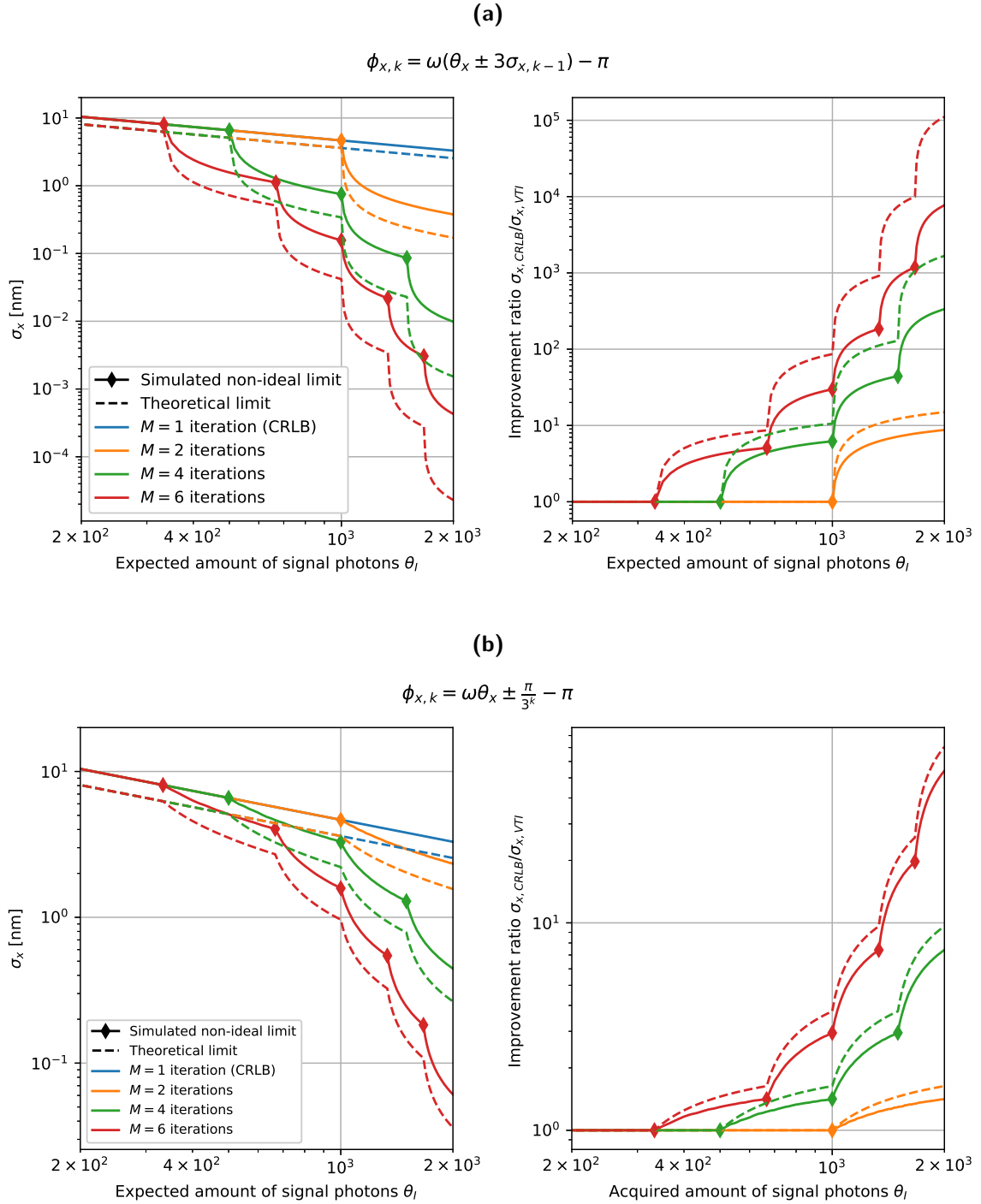


Figure 3-3: Simulated localization precision in x -direction as a function of the acquired amount of signal photons, when sinusoidal intensity patterns are positioned (a) using Equation (3-13) with $\alpha = 3$; or (b) using Equation (3-14) with $\beta = 3$. Simulated results are compared against the theoretical limit of Equation (3-19). The emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and 3 iterations were simulated. Intensity minima were placed symmetrically around the true emitter position, to eliminate the effect of estimation errors.

From Figure 3-3(a), it can be concluded that the information content of signal photons is increased as a result of the pattern placement of Equation (3-13). As discussed for the theoretical limit, increasing the iteration count further increases the information content per signal photon. In addition, the relative difference between the simulated non-ideal localization precision and the theoretical limit increases as the amount of iterations grows. This is explained by the fact that the influence of background and discretization on the localization precision accumulates as the amount of iterations increases, as lowered precision $\sigma_{x,k-1}$ in a previous iteration is carried over to the next iteration.

From Figure 3-3(b), it can be concluded that the information content of signal photons is increased as a result of the pattern placement of Equation (3-14). However, compared to the pattern placement of Figure 3-3(a), significantly less information is obtained by using this strategy for pattern placement. As the pattern placements are not adapted based on the current localization precision, the distance between the intensity minima decreases significantly slower during early iterations. In contrast to Figure 3-3(a), the relative difference between the simulated non-ideal localization precision and the theoretical limit stays approximately constant as the amount of iterations grows. As the pattern positions do not depend on the localization precision of earlier iterations, lowered precision in previous iterations does not significantly alter the localization precision achieved in the current iteration.

In Figure 3-4, the effect of increasing the iteration count on the localization precision was simulated for the pattern positions described by Equations (3-13), (3-14). Photons are equally distributed over iterations, such that 2000 photons are acquired when the maximum amount of iterations is reached. As before, these results are compared against the theoretical limit of Equation (3-19), the emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and intensity minima were placed symmetrically around the true emitter position.

From Figure 3-4, it can be seen that independent from the chosen pattern placement strategy, increasing the iteration count also increases the localization precision. Furthermore, it can be seen that the pattern placement strategy of Equation (3-13) outclasses the theoretical limit of the strategy of Equation (3-14), which is again a result of the fact that the distance between the intensity minima decreases significantly slower during the early iterations.

For higher amounts of iterations, it can be seen that the pattern placement of Equation (3-13) begins to stagnate as it deviates further from its theoretical limit, due to the accumulating effects of background and discretization. On the contrary, the pattern placement strategy of Equation (3-14) does not show an increasing deviation from its theoretical limit, as it is less affected by repeated loss of localization precision.

As a remark, it should be noted that the simulated localization precision values are unrealistically low. For example, localization precision values of 10^{-5} to 10^{-7} nm are not realistically achievable, given e.g. the size of emitters. As the emitter size was not included in the image formation model, these values do appear in simulations. It does show that in practice, either a low amount of iterations is needed or a low amount of signal photons needs to be acquired to achieve a desired localization precision.

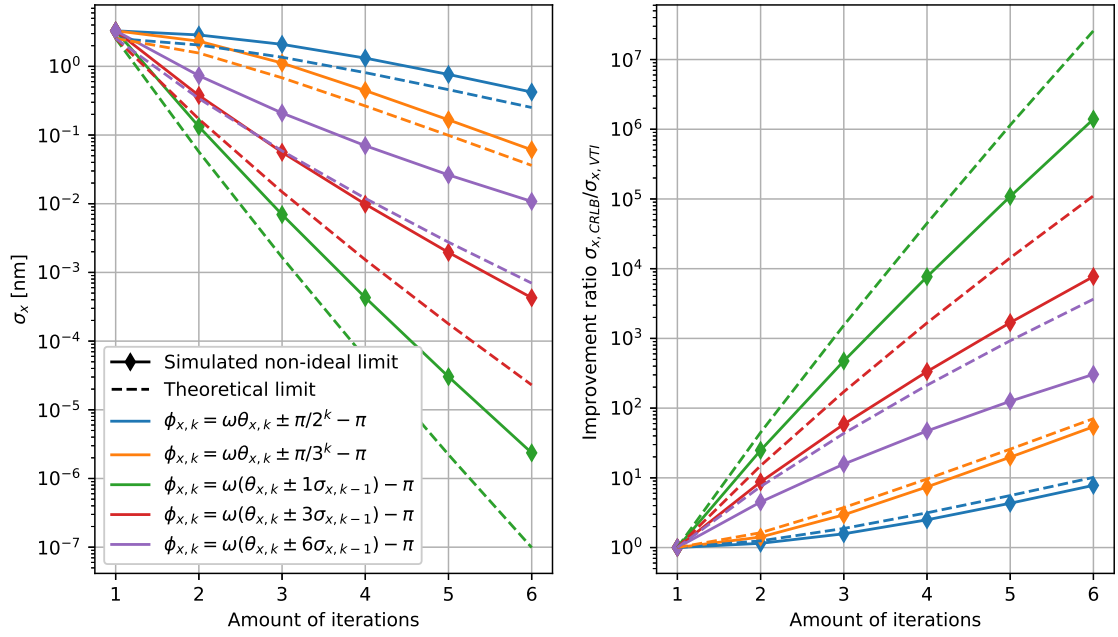


Figure 3-4: Simulated localization precision in x -direction as a function of the maximum amount of iterations, when sinusoidal intensity patterns are positioned using Equation (3-13) with $\alpha = 1, 3$ or 6 or using Equation (3-14) with $\beta = 2$ or 3 . Simulated results are compared against the theoretical limit of Equation (3-19). Photons are equally distributed over iterations, such that 2000 photons are acquired when the maximum amount of iterations is reached. The emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and intensity minima were placed symmetrically around the true emitter position, to eliminate the effect of estimation errors.

3-4-4 Effect of initial guess and estimation errors

In imeSMLM, the localization precision is also affected by the initial guess for the emitter position and errors made when estimating the emitter position. If we are able to place the intensity minimum of the initial pattern close to the true emitter position, then the prior distribution on the emitter position becomes more precise in early iterations. However, estimation errors can cause deterioration of the localization precision, as there now is an offset between the true emitter position and the center between the intensity minima. In this subsection, we investigate the sensitivity of the localization precision to these factors.

In Figure 3-5, the emitter position θ_x is varied over 50 equidistantly spaced values in $[-\lambda/2, \lambda/2]$ while the initial pattern position $\phi_{x,0} = 0$ is kept constant. In all further iterations, the intensity minima were again placed symmetrically around the true emitter position, to eliminate the effect of estimation errors. Photons are equally distributed over iterations, such that 2000 photons are acquired when the maximum amount of iterations is reached.

From Figure 3-5, it can be seen that the localization precision is strongly affected by the initial pattern position. If the intensity minimum of the initial pattern is placed closed to the emitter position, the information content of signal photons grows sharply. However, if the intensity maximum is placed over the emitter position, the initial localization precision is low. This effect grows larger in the absence of background, as indicated by the differences between the theoretical and simulated localization precision.

Figure 3-5(a) shows that the achievable localization precision with the pattern placement strategy of Equation (3-13) is permanently affected by the initial guess, as improved precision in an early iteration allows for closely spaced intensity minima during future iterations. This does not apply to the achievable localization precision with the pattern placement strategy of Equation (3-14), as shown in Figure 3-5(b) and it is thus able to asymptotically eliminate the effect of the initial pattern placement on the localization precision using this pattern position control strategy.

In Figure 3-6, the intensity minima for all pattern iterations were placed symmetrically around the true emitter position plus an offset of $\sigma_{x,k-1}$ or $2\sigma_{x,k-1}$. By doing so, the worst-case scenario of accumulating estimation errors can be simulated. As before, the emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and 3 iterations were simulated.

From Figure 3-6(a), it can be seen that estimation errors have a small, but increasing effect on the localization precision. In the first iteration, the difference between the fundamental precision limit and the precision that was disturbed by an estimation error is insignificant. As such, the intensity minima are placed closer together in the next iteration, approximately independent from the estimation error. As the region of increased localization precision has grown smaller, the effect of estimation errors increases in the next iterations. Based on the theoretical and simulated localization precision, we can conclude that the effect of estimation errors on the localization precision is not significantly influenced by background and image discretization. The same principle applies to the pattern placement strategy of Equation (3-14), as shown in Figure 3-6(b). Because the intensity minima converge slowly for this strategy, the effects of estimation errors on the localization precision are less significant for low amounts of iterations.

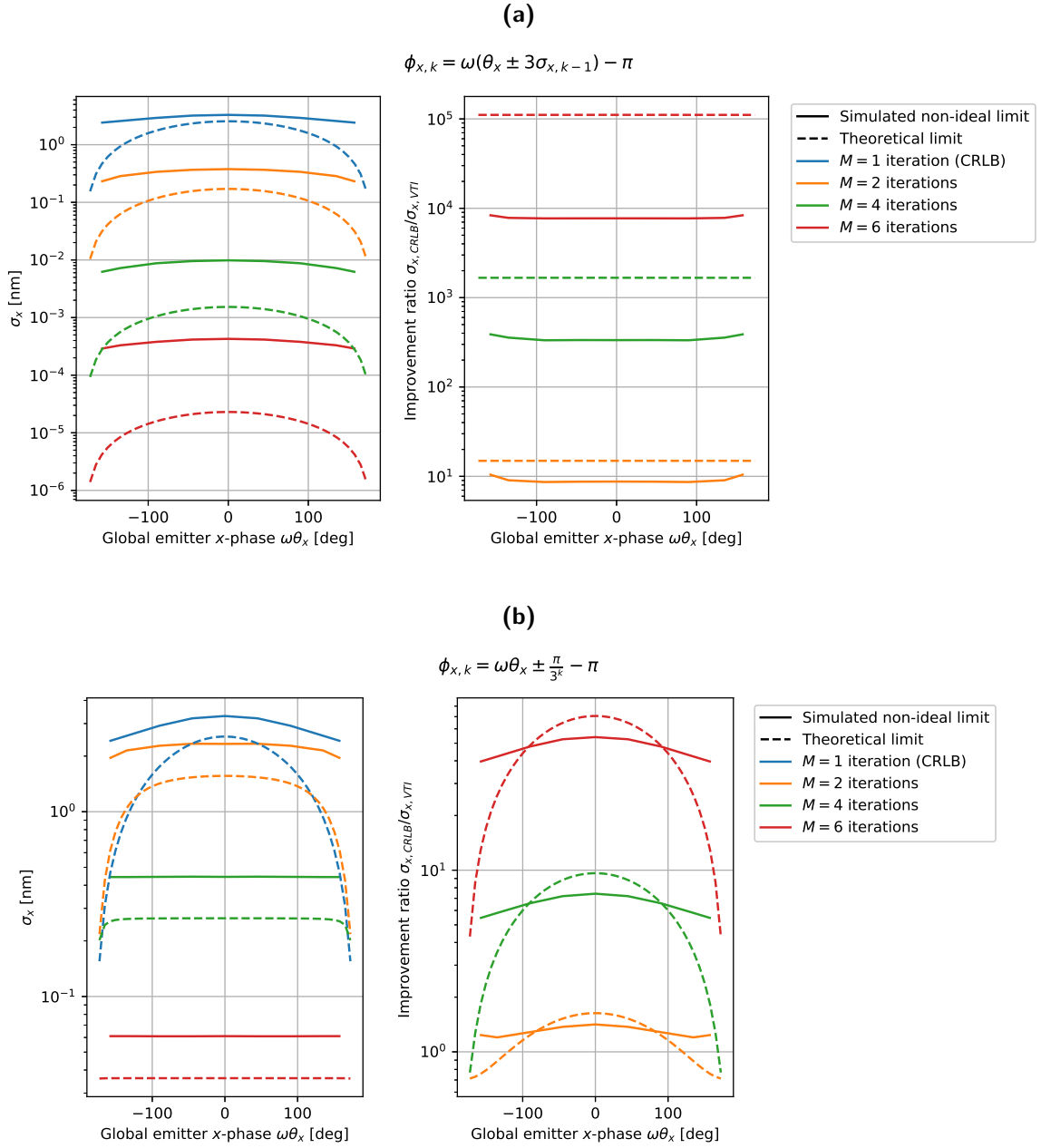


Figure 3-5: Simulated localization precision in x -direction as a function of the global emitter x -phase $\omega\theta_x$, when sinusoidal intensity patterns are positioned (a) using Equation (3-13) with $\alpha = 3$; or (b) using Equation (3-14) with $\beta = 3$. Simulated results are compared against the theoretical limit of Equation (3-19). Photons are equally distributed over iterations, such that 2000 photons are acquired when the maximum amount of iterations is reached. The initial pattern position $\phi_{x,0} = 0$ was used and intensity minima were placed symmetrically around the true emitter position in for $k \geq 1$, to eliminate the effect of estimation errors.

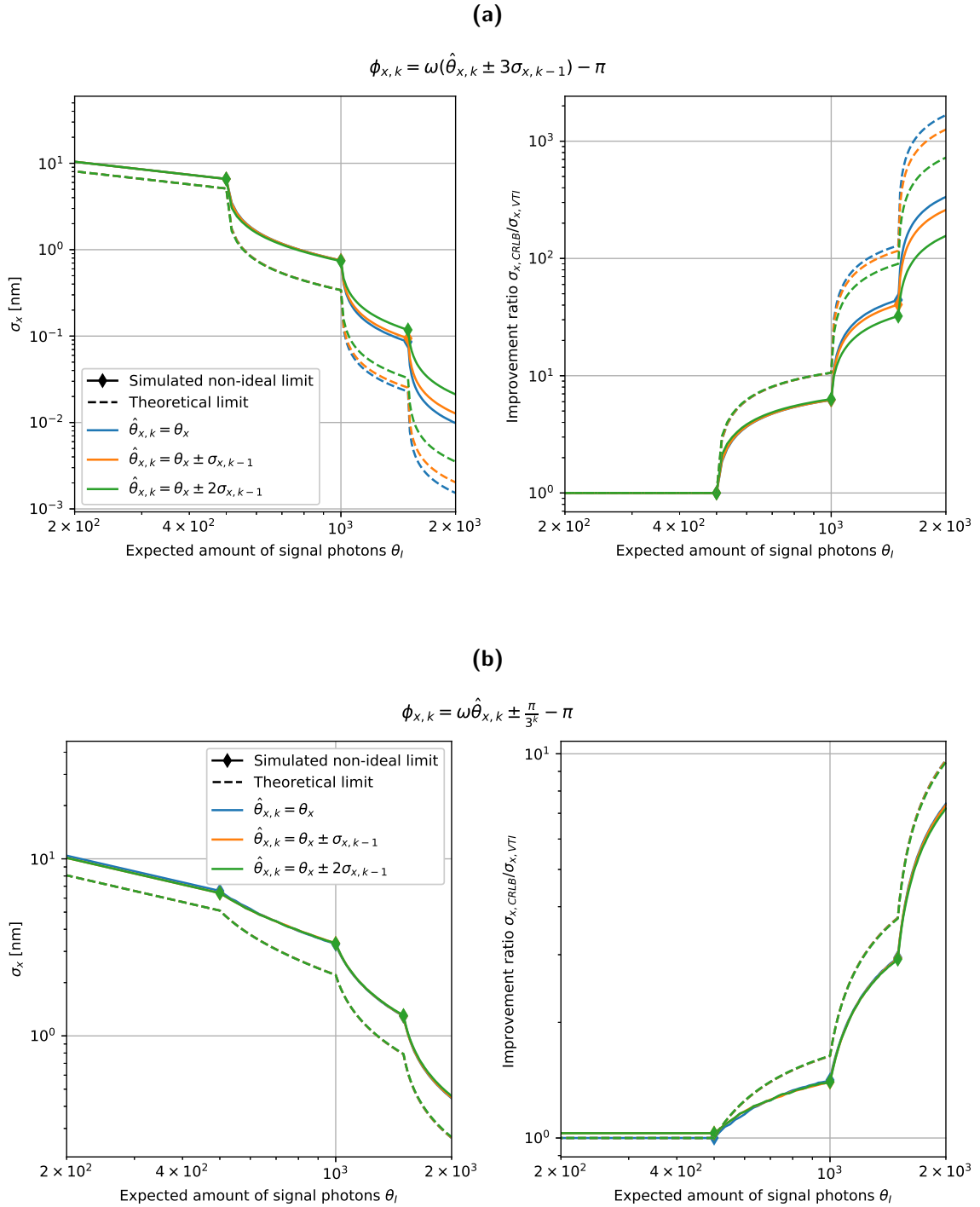


Figure 3-6: Simulated localization precision in x -direction as a function of the acquired amount of signal photons, when sinusoidal intensity patterns are positioned (a) using Equation (3-13) with $\alpha = 1, 3$ or 6 ; or (b) using Equation (3-14) with $\beta = 2$ or 3 . Simulated results are compared against the theoretical limit of Equation (3-19). Intensity minima were placed symmetrically around the true emitter position plus an offset of $\sigma_{x,k-1}$ or $2\sigma_{x,k-1}$, to investigate the effect of estimation errors on the localization precision. The emitter position $(\theta_x, \theta_y) = (0 \text{ nm}, 0 \text{ nm})$ was used and 3 iterations were simulated.

3-5 Conclusion

In imeSMLM methods, resolution is improved locally around an emitter position by using prior information that was derived from earlier iterations. The CRLB cannot be used as a precision measure for imeSMLM methods, as it treats estimands as deterministic unknowns. As such, prior information cannot be incorporated in the CRLB.

By treating estimands as random variables with a known prior distribution, the Van Trees inequality can be used to estimate the maximally achievable localization precision. The Van Trees inequality is useful for estimating the localization precision in imeSMLM, as it is able to account for the effect of prior information that is generated over the course of iterations. Furthermore, the VTI can be used to simulate the effects of illumination pattern position control on the localization precision and is therefore a promising method for optimal control of imeSMLM methods.

We derived a theoretical limit on the localization precision of imeSMLM methods that make use of standing wave illumination patterns, in the absence of estimation uncertainty on the signal photon count and background fluorescence and with perfect modulation. This fundamental limit cannot be surpassed, as the aforementioned effects can only worsen the localization precision of imeSMLM methods. Using this limit, we have shown that under the above assumptions, the information content of signal photons increases exponentially as a function of the iteration count.

In addition, the Van Trees inequality was used to assess the performance of illumination pattern position control strategies. While the information content per signal photon can be maximized by illuminating an emitter with an intensity minimum, robustness and performance criteria can be balanced by symmetrically surrounding an emitter with intensity minima of standing waves, after which the distance between these minima is reduced over the course of multiple iterations. By coupling this distance reduction to the localization precision of the previous iteration, a significant performance increase with respect to predefined scaling was simulated, although this makes the method susceptible to accumulating precision loss through the effects of background and image discretization by the camera.

3-6 Outlook

The Van Trees inequality can be used to estimate the localization precision for a wide range of imeSMLM methods. The Van Trees inequality can easily be adapted to deal with other illumination pattern types, such as the doughnut-shaped beam used in (iterative) MINFLUX [5, 6], or the 3D-SIM pattern [50], which can be used for three-dimensional localization.

Furthermore, the Van Trees inequality can be used to assess the performance of illumination pattern control strategies and is therefore a promising performance metric in optimal control of imeSMLM methods. While we limited ourselves to fixing the amount of photons per iteration, fixing the acquisition time per iteration may result in different optimal pattern positions, as the amount of received photons per iteration now depends on the pattern position. In addition, optimal pattern positions may change in case multiple emitters need to be localized by the same pattern placement control strategy. Lastly, we expect that optimal control for iterative three-dimensional localization methods is a challenging problem, as the objectives of improving lateral and axial resolution are not necessarily aligned.

Optimal control of iterative localization methods using the Van Trees inequality does not have to stay limited to illumination control only. In [51], the CRLB was used to engineer a PSF that minimizes the sum of the localization precisions in the x -, y - and z -directions. We expect that the Van Trees inequality can be used for a similar purpose, where an optimal PSF can be engineered for every iteration of an iterative localization method.

Supplementary Note 1

In this note, we give an extensive derivation of the image formation model of Section 3-2.

Model for photon collection An image $g_k(x, y)$ of an object function $f_k(x, y)$ in iteration k is formed through a convolution with the PSF $h(x, y)^2$ of the optical system, as shown in Equation (3-22).

$$g_k(x, y) = h(x, y)^2 \otimes f_k(x, y) \quad (3-22)$$

In Equation (3-22), \otimes denotes the two-dimensional convolution operator. We must assure that the total area under the PSF equals 1 to preserve conservation of energy. As a result, $h(x, y)^2$ describes the relative amount of PSF that falls on an infinitesimal area $dx dy$. This results in the normalization condition of Equation (3-23).

$$\iint_{\mathbb{R}^2} h(x, y)^2 dx dy = 1 \quad (3-23)$$

Next, we formulate the object function $f(x, y)$. Consider a point emitter, located at position (θ_x, θ_y) . Such an emitter can be modeled as $\delta(x - \theta_x, y - \theta_y)$, where δ denotes the Dirac delta function. Under uniform illumination, the expected amount of photons emitted by this emitter is θ_I and a background count θ_b is expected. Therefore, under uniform illumination, we find an object description $\theta_I \delta(x - \theta_x) + \theta_b$.

Let an illumination intensity pattern be denoted by $P(x, y)$, with $0 \leq P(x, y) \leq 1$ for all $(x, y) \in \mathbb{R}^2$. Note that this illumination will affect both the amount of photons emitted by the sample, as well as the background count. The resulting object function $f_k(x, y)$ is shown in Equation (3-24).

$$f_k(x, y) = P_k(x, y)(\theta_I \delta(x - \theta_x, y - \theta_y) + \theta_b) \quad (3-24)$$

To obtain an expression for the image function $g_k(x, y)$ from Equation (3-22), the point spread function $h(x, y)^2$ and the object function $f_k(x, y)$ are convolved. The resulting expression is given by Equation (3-25).

$$g_k(x, y) = \theta_I h(x - \theta_x, y - \theta_y)^2 P_k(\theta_x, \theta_y) + \theta_b \iint_{\mathbb{R}^2} h(\tau, \gamma)^2 P_k(x - \tau, y - \gamma) d\tau d\gamma \quad (3-25)$$

As a next step, we discretize the image function $g_k(x, y)$ using the camera pixel area A_i , where i denotes the pixel index. To discretize the image function, we integrate it over all $(x, y) \in A_i$ to obtain the expected amount of photons on the i 'th pixel during iteration k , $\mu_{i,k}$. We find:

$$\mu_{i,k} = \theta_I P_k(\theta_x, \theta_y) \iint_{(x,y) \in A_i} h(x - \theta_x, y - \theta_y)^2 dx dy + \theta_b \underbrace{\iint_{(x,y) \in A_i} \left(\iint_{\mathbb{R}^2} h(\tau, \gamma)^2 P_k(x - \tau, y - \gamma) d\tau d\gamma \right) dx dy}_{B_{i,k}} \quad (3-26)$$

Note that the parameters θ_x , θ_y , θ_I and θ_b appear in the expression for $\mu_{i,k}$. Therefore, the Poisson model under consideration can be used to estimate θ_x , θ_y , θ_I and θ_b from data, using maximum likelihood estimation. Furthermore, note that $B_{i,k}$ is a constant, which does not depend on the emitter position, but only on the pixel area, the PSF and the illumination pattern. We can thus give a compact expression for the Poisson mean $\mu_{i,k}$, as shown in Equation (3-27).

$$\mu_{i,k} = \theta_I P_k(\theta_x, \theta_y) \iint_{(x,y) \in A_i} h(x - \theta_x, y - \theta_y)^2 dx dy + \theta_b B_{i,k} \quad (3-27)$$

Let us assume a camera, for which all pixels have the same shape and size. Let all pixels be rectangular, with length Δx in x -direction and Δy in y -direction. Furthermore, let (x_i, y_i) denote the center coordinates of the i 'th pixel. Equation (3-27) then becomes Equation (3-28).

$$\mu_{i,k} = \theta_I P_k(\theta_x, \theta_y) \int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} \int_{y_i - \frac{\Delta y}{2}}^{y_i + \frac{\Delta y}{2}} h(x - \theta_x, y - \theta_y)^2 dx dy + \theta_b B_{i,k} \quad (3-28)$$

We assume that PSF $h(x, y)^2$ is Gaussian. In two dimensions, it is given by Equation (3-29). Note that for the Gaussian PSF model, the condition of Equation (3-23) is satisfied.

$$h(x, y)^2 = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}} \quad (3-29)$$

Furthermore, note that the exponential term in Equation (3-29) can be split in a product of two exponentials, of which one is dependent on x and of which the other is dependent on y . Using these insights, we can further simplify the expression for the Poissonian mean:

$$\mu_{i,k} = \theta_I P_k(\theta_x, \theta_y) \underbrace{\left(\int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \theta_x)^2}{2\sigma^2}} dx \right)}_{E_x(x_i, \theta_x)} \underbrace{\left(\int_{y_i - \frac{\Delta y}{2}}^{y_i + \frac{\Delta y}{2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y - \theta_y)^2}{2\sigma^2}} dy \right)}_{E_y(y_i, \theta_y)} + \theta_b B_i$$

We use the error function to write explicit expressions for $E_x(x_i, \theta_x)$ and $E_y(y_i, \theta_y)$:

$$E_x(x_i, \theta_x) = \frac{1}{2} \operatorname{erf} \left(\frac{x_i - \theta_x + \frac{\Delta x}{2}}{\sigma\sqrt{2}} \right) - \frac{1}{2} \operatorname{erf} \left(\frac{x_i - \theta_x - \frac{\Delta x}{2}}{\sigma\sqrt{2}} \right)$$

$$E_y(y_i, \theta_y) = \frac{1}{2} \operatorname{erf} \left(\frac{y_i - \theta_y + \frac{\Delta y}{2}}{\sigma\sqrt{2}} \right) - \frac{1}{2} \operatorname{erf} \left(\frac{y_i - \theta_y - \frac{\Delta y}{2}}{\sigma\sqrt{2}} \right)$$

Using this expression, we can simplify Equation (3-28) even further. We then obtain the Poissonian mean μ_i as shown in Equation (3-30).

$$\mu_{i,k} = \theta_I P_k(\theta_x, \theta_y) E_x(x_i, \theta_x) E_y(y_i, \theta_y) + \theta_b B_i \quad (3-30)$$

Model for illumination We consider standing wave illumination as discussed in [44]. To construct this pattern, two beams are focused such that they intersect in the focal plane. We assume the beams consist of monochromatic light and that the waves have the same polarization. The beams interfere in the focal plane and form the sinusoidal SIM illumination pattern. This pattern is sinusoidal in one lateral direction (e.g. the x -direction) and does not vary as a function of other lateral coordinate. Hence, we can analyze the pattern in one dimension first, after which it can be generalized for two dimensions. Before we model the illumination pattern, we establish some notation and axis conventions. These are shown in Figure 3-7.

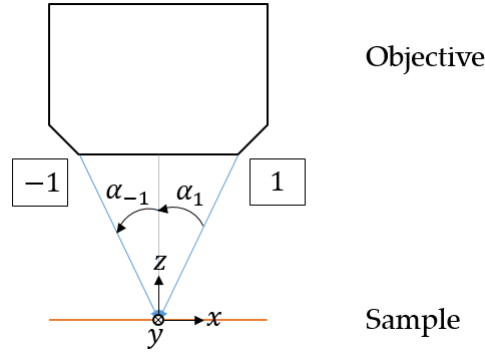


Figure 3-7: Zoomed-in view of the back focal plane of the objective lens of a SIM-setup to establish notation and axis conventions. Boxed numbers indicate beam indices, corresponding to their diffraction orders.

From [44], we infer that the individual beams after the objective lens may be approximated as plane waves. From [14], we use the fact that we can write monochromatic plane waves in complex exponential form as shown in Equation (3-31).

$$U(\mathbf{r}, t) = A e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi)} \quad (3-31)$$

In Equation (3-31), A describes the amplitude of the wave, \mathbf{k} is the wave vector with length $\|\mathbf{k}\|_2 = \frac{2\pi}{\lambda}$ (i.e. the wavenumber, with λ the wavelength), $\mathbf{r} = \begin{bmatrix} x \\ z \end{bmatrix}$ is a vector of spatial coordinates, $\omega = 2\pi f = \frac{2\pi c}{\lambda}$ is the angular frequency of the wave (with f the frequency of the wave and c the speed of light), t is the time and ϕ is the initial phase shift. From Figure 3-7, we can see that waves -1 and 1 are rotated with respect to the z -axis. This rotation can be captured by a linear transformation, included in the wave vector \mathbf{k} . This results in the wave vectors for waves -1 and 1 as shown in Equations (3-32) and (3-33), respectively.

$$\mathbf{k}_{-1} = \frac{2\pi}{\lambda} \begin{bmatrix} -\sin(\alpha_{-1}) \\ \cos(\alpha_{-1}) \end{bmatrix} \quad (3-32)$$

$$\mathbf{k}_1 = \frac{2\pi}{\lambda} \begin{bmatrix} \sin(\alpha_1) \\ \cos(\alpha_1) \end{bmatrix} \quad (3-33)$$

Using Equation (3-31), we end up with the expressions for waves -1 and 1 as shown in Equations (3-34) and (3-35), respectively.

$$U_{-1}(x, z, t) = A_{-1}e^{i(\mathbf{k}_{-1} \cdot \mathbf{r} - \omega t + \phi_{-1})} = A_{-1}e^{i\left(\frac{2\pi}{\lambda}(-\sin(\alpha_{-1})x + \cos(\alpha_{-1})z) - \omega t + \phi_{-1}\right)} \quad (3-34)$$

$$U_1(x, z, t) = A_1e^{i(\mathbf{k}_1 \cdot \mathbf{r} - \omega t + \phi_1)} = A_1e^{i\left(\frac{2\pi}{\lambda}(\sin(\alpha_1)x + \cos(\alpha_1)z) - \omega t + \phi_1\right)} \quad (3-35)$$

We now make an assumption to simplify the model. If the optical system used to construct this pattern is perfectly aligned and if all components are perfect (i.e. they contain no manufacturing defects), it must hold that $\alpha_{-1} = \alpha_1$. While perfect alignment and perfect components cannot be guaranteed, we assume that the differences are small enough such that the above assumption is valid.

The wave pattern due to interference can be described by summing up Equations (3-34) and (3-35). Under the above assumption, the result of summing up Equations (3-34) and (3-35) is found in Equation (3-36).

$$U(x, z, t) = A_{-1}e^{i\left(\frac{2\pi}{\lambda}(-\sin(\alpha_1)x + \cos(\alpha_1)z) - \omega t + \phi_{-1}\right)} + A_1e^{i\left(\frac{2\pi}{\lambda}(\sin(\alpha_1)x + \cos(\alpha_1)z) - \omega t + \phi_1\right)} \quad (3-36)$$

In the image formation model, we are interested in the intensity pattern $I(x, z)$. For scalar waves, it is defined in [14] as the infinite time average of the squared magnitude of $U(x, z, t)$. This definition is shown in Equation (3-37).

$$I(x, z) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} U(x, z, t)U^*(x, z, t)dt \quad (3-37)$$

In Equation (3-37), U^* denotes the complex conjugate of U .

From [44], we know that the pattern is time invariant when we assume monochromatic illumination. This also follows from the expression of $U(x, z, t)U^*(x, z, t)$, as shown in Equation (3-38).

$$U(x, z, t)U^*(x, z, t) = A_{-1}^2 + A_1^2 + 2A_{-1}A_1 \cos\left(\frac{4\pi}{\lambda} \sin(\alpha_1)x - \phi\right) \quad (3-38)$$

In Equation (3-38), $\phi = \phi_{-1} - \phi_1$. This parameter can be used to control the spatial pattern shift. As $U(x, z, t)U^*(x, z, t)$ is time invariant, Equation (3-37) simplifies to Equation (3-39).

$$I(x, z) = U(x, z, t)U^*(x, z, t) \quad (3-39)$$

Let us introduce the beam intensities $I_{-1} = A_{-1}^2$ and $I_1 = A_1^2$ for beams -1 and 1, respectively. Furthermore, for the intensity pattern to be consistent with the derived image formation model, we include a normalization constant K , which will be chosen such that the maximum total illumination received by an emitter equals at most 1. Using these expressions, we end up at the intensity pattern model of Equation (3-40).

$$P(x) = K \left[I_{-1} + I_1 + 2\sqrt{I_{-1}I_1} \cos \left(\frac{4\pi}{\lambda} \sin(\alpha_1)x - \phi \right) \right] \quad (3-40)$$

We will make some additional adjustments to the pattern description of Equation (3-40) for convenience. First of all, we substitute the pattern frequency $\omega = \frac{4\pi}{\lambda} \sin(\alpha_1)$. Note that ω is inversely proportional with the Abbe diffraction limit $\frac{\lambda}{2\text{NA}}$ in air. This is not unexpected, as the pattern itself is, at best, also diffraction limited. While we did not derive the result for an arbitrary immersion medium with index of refraction n , it can be shown that the inverse proportionality relation still holds in this case [7].

Next, we factor out $(I_{-1} + I_1)$. As the constant K has not been defined, we replace $\frac{K}{I_{-1}+I_1}$ by a new constant A for simplicity. Furthermore, we replace $\frac{2\sqrt{I_{-1}I_1}}{I_{-1}+I_1}$ by the modulation contrast m . Note that for perfect modulation, $I_{-1} = I_1$ and thus $m = 1$.

The expression resulting from these adjustments is found in Equation (3-41).

$$P(x) = A [1 + m \cos(\omega x - \phi)] \quad (3-41)$$

Next, we generalize this pattern to two dimensions. We consider an intensity pattern sequence, consisting of standing waves with different phase shifts in two orthogonal orientations. For each iteration $k \in \{0, 1, \dots, M-1\}$, with M the amount of pattern placement iterations, we define the standing wave illumination pattern $P_k = P_{x,k} + P_{y,k}$ as shown in Equation (3-42). Note that for each orientation individually, the standing wave pattern of Equation (3-41) is retrieved.

$$\begin{cases} P_{x,k}(x, y, \phi_{x,k}) = A [1 + m \cos(\omega x - \phi_{x,k})] \\ P_{y,k}(x, y, \phi_{y,k}) = A [1 + m \cos(\omega y - \phi_{y,k})] \end{cases} \quad (3-42)$$

We choose the normalization constant A to ensure that the maximum possible illumination received by any emitter is equal to 1. That is, $\sum_{k=0}^M P_{x,k}(x, y, \phi_{x,k}) + P_{y,k}(x, y, \phi_{y,k}) \leq 1$ for all $(x, y) \in \mathbb{R}^2$ and for all $\phi_{x,k}, \phi_{y,k} \in [-\pi, \pi]$, with equality for at least one $(x, y) \in \mathbb{R}^2$ and $\phi_{x,k}, \phi_{y,k} \in [-\pi, \pi]$. This can be achieved by setting $A = \frac{1}{4M}$.

Lastly, we derive the constant $B_{i,k}$ from Equation (3-26). As the Gaussian PSF is separable in x and y , we find:

$$\begin{aligned} B_{i,k} &= \iint_{\{x,y\} \in A_i} \left(\iint_{\mathbb{R}^2} h_x(\tau)^2 h_y(\gamma)^2 P_{x,k}(x - \tau, y - \gamma, \phi_{x,k}) d\tau d\gamma \right) dx dy \\ &\quad + \iint_{\{x,y\} \in A_i} \left(\iint_{\mathbb{R}^2} h_x(\tau)^2 h_y(\gamma)^2 P_{y,k}(x - \tau, y - \gamma, \phi_{y,k}) d\tau d\gamma \right) dx dy \\ &= \int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} \left(\int_{y_i - \frac{\Delta y}{2}}^{y_i + \frac{\Delta y}{2}} \int_{-\infty}^{\infty} h_y(\gamma)^2 d\gamma dy \right) \left(\int_{-\infty}^{\infty} h_x(\tau)^2 P_{x,k}(x - \tau, y, \phi_{x,k}) d\tau \right) dx \\ &\quad + \int_{y_i - \frac{\Delta y}{2}}^{y_i + \frac{\Delta y}{2}} \left(\int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} \int_{-\infty}^{\infty} h_x(\tau)^2 d\tau dx \right) \left(\int_{-\infty}^{\infty} h_y(\gamma)^2 P_{y,k}(x, y - \gamma, \phi_{y,k}) d\gamma \right) dy \end{aligned}$$

$$\begin{aligned}
&= \Delta y \int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} \left(\int_{-\infty}^{\infty} h_x(\tau)^2 P_{x,k}(x - \tau, y, \phi_{x,k}) d\tau \right) dx \\
&+ \Delta x \int_{y_i - \frac{\Delta y}{2}}^{y_i + \frac{\Delta y}{2}} \left(\int_{-\infty}^{\infty} h_y(\gamma)^2 P_{y,k}(x, y - \gamma, \phi_{y,k}) d\gamma \right) dy
\end{aligned}$$

Evaluating these improper integrals results in the following expression for $B_{i,k}$:

$$\begin{aligned}
B_{i,k} &= \frac{\Delta x \Delta y}{2M} + \frac{m \Delta x}{2M\omega} \exp\left(-\frac{\omega^2 \sigma^2}{2}\right) \cos(\omega y_i - \phi_{x,k}) \sin\left(\frac{\omega \Delta y}{2}\right) \\
&+ \frac{m \Delta y}{2M\omega} \exp\left(-\frac{\omega^2 \sigma^2}{2}\right) \cos(\omega x_i - \phi_{y,k}) \sin\left(\frac{\omega \Delta x}{2}\right)
\end{aligned} \tag{3-43}$$

Supplementary Note 2

In this note, we state the multivariate Van Trees inequality and its assumptions, based on its formulation in [38, 42]. A proof of this theorem can be found in [38].

Let $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ represent a data set, which can be seen a realization of i.i.d. random variables $[X_1, X_2, \dots, X_n]^T$ with a distribution depending on a parameter $\boldsymbol{\theta} \in \mathbb{R}^m$. Suppose that the likelihood function $L(\boldsymbol{\theta}|\mathbf{x})$ is continuously differentiable with respect to $\boldsymbol{\theta}$ and that its support $\{\mathbf{x}|L(\boldsymbol{\theta}|\mathbf{x}) > 0\}$ does not depend on $\boldsymbol{\theta}$.

Let a prior probability density function $\lambda(\boldsymbol{\theta})$ on the parameter $\boldsymbol{\theta}$ be known. Suppose that $\lambda(\boldsymbol{\theta})$ is absolutely continuous and that λ converges to zero at the endpoints of its domain. Then, the mean squared error matrix, $\text{MSE}(\hat{\boldsymbol{\theta}}) = \mathbb{E}_{\lambda,L}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$ of any estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ satisfies:

$$\left(\mathbb{E}_{\lambda,L}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] - (J_D + J_P)^{-1}\right) \succeq 0 \quad (3-44)$$

That is, $\mathbb{E}_{\lambda,L}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] - (J_D + J_P)^{-1}$ is positive semi-definite, where the data information matrix J_D is given by

$$\begin{aligned} [J_D]_{p,q} &= [\mathbb{E}_{\lambda}[I(\boldsymbol{\theta})]_{p,q}] = \mathbb{E}_{\lambda} \left[\mathbb{E}_L \left[\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_p} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_q} \right] \right] \\ &= \mathbb{E}_{\lambda} \left[\int_{\mathbb{R}^n} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_p} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_q} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \right] \end{aligned} \quad (3-45)$$

and where the prior information matrix J_P is given by

$$[J_P]_{p,q} = [I(\lambda)]_{p,q} = \mathbb{E}_{\lambda} \left[\frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_p} \frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_q} \right] = \int_{\mathbb{R}^m} \frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_p} \frac{\partial \log(\lambda(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_q} \lambda(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3-46)$$

Furthermore, $(J_D + J_P)$ can be rewritten as J with $J_{p,q} = \mathbb{E}_{\lambda,L} \left[\frac{\partial \log(p(\mathbf{x}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_p} \frac{\partial \log(p(\mathbf{x}, \boldsymbol{\theta}))}{\partial \boldsymbol{\theta}_q} \right]$, where $p(\mathbf{x}, \boldsymbol{\theta})$ denotes the joint probability density function of \mathbf{x} and $\boldsymbol{\theta}$. The matrix J is called the Bayesian information matrix.

Supplementary Note 3

In this note, theoretical limits on the localization precision are derived using the Van Trees inequality. We limit ourselves to one-dimensional localization. Furthermore, we assume the expected signal photon count θ_I is a known constant (i.e. we do not estimate it), we ignore background such that $\theta_b = 0$ and we assume that the modulation is perfect, such that $m = 1$. Lastly, we assume that the point spread function is a Gaussian with standard deviation σ_{PSF} . We will derive results in two scenarios. In the first case, we disregard the effect of discretization due to the finite size of camera pixels. In the second scenario, a camera with a single, infinitely large pixel is considered. In both situations, expressions are derived in case 1 or 2 pattern positions are used on the x -axis during a single iteration.

No pixelation In the one-dimensional case when one pattern position $\phi_{x,k}$ is used, the illumination pattern in iteration k is given by Equation (3-47).

$$P_{x,k}(x) = \frac{1}{2M} [1 + m \cos(\omega x - \phi_{x,k})] \quad (3-47)$$

Under the presented assumptions, we find the expected photon count (as a function of the x -coordinate, since the image is not discretized) of Equation (3-48), with its first partial derivative with respect to θ_x being given by Equation (3-49).

$$\mu_k(x) = \frac{\frac{\theta_I}{2M} [1 + \cos(\omega \theta_x - \phi_{x,k})]}{\sigma_{PSF} \sqrt{2\pi}} e^{-\frac{(x-\theta_x)^2}{2\sigma_{PSF}^2}} \quad (3-48)$$

$$\frac{\partial \mu_k}{\partial \theta_x}(x) = \frac{\frac{\theta_I}{2M}}{\sigma_{PSF} \sqrt{2\pi}} e^{-\frac{(x-\theta_x)^2}{2\sigma_{PSF}^2}} \left(\frac{x - \theta_x}{\sigma_{PSF}^2} + \frac{x - \theta_x}{\sigma_{PSF}^2} \cos(\omega \theta_x - \phi_{x,k}) - \omega \sin(\omega \theta_x - \phi_{x,k}) \right) \quad (3-49)$$

The Fisher information in iteration k is then given by Equation (3-50).

$$\begin{aligned} I_k(\theta_x) &= \int_{\mathbb{R}} \frac{1}{\mu_k(x)} \left(\frac{\partial \mu_k}{\partial \theta_x}(x) \right)^2 dx \\ &= \frac{\frac{\theta_I}{2M}}{1 + \cos(\omega \theta_x - \phi_{x,k})} \left(\omega^2 \sin^2(\omega \theta_x - \phi_{x,k}) + \frac{1}{\sigma_{PSF}^2} (1 + \cos(\omega \theta_x - \phi_{x,k}))^2 \right) \\ &= \frac{\theta_I}{2M} \omega^2 (1 - \cos(\omega \theta_x - \phi_{x,k})) + \frac{\theta_I}{2M \sigma_{PSF}^2} (1 + \cos(\omega \theta_x - \phi_{x,k})) \end{aligned} \quad (3-50)$$

We choose the prior information available in iteration $k \geq 1$ to be normally distributed with mean $\hat{\theta}_{x,k-1}$ and variance $\sigma_{x,k-1} = \frac{1}{\sqrt{I_0(\theta_x)}}$ (for $k = 1$) or $\sigma_{x,k-1} = \frac{1}{\sqrt{J_{k-1}}}$ (for $k \geq 2$). The resulting prior information, its natural logarithm and its partial derivative with respect to θ_x are given by Equations (3-51)-(3-53).

$$\lambda_{k-1}(\theta_x) = \frac{1}{\sigma_{x,k-1}\sqrt{2\pi}} \exp\left(-\frac{(\theta_x - \hat{\theta}_{x,k-1})^2}{2\sigma_{x,k-1}^2}\right) \quad (3-51)$$

$$\log(\lambda_{k-1}(\theta_x)) = \log\left(\frac{1}{\sigma_{x,k-1}\sqrt{2\pi}}\right) - \frac{(\theta_x - \hat{\theta}_{x,k-1})^2}{2\sigma_{x,k-1}^2} \quad (3-52)$$

$$\frac{\partial \log(\lambda_{k-1}(\theta_x))}{\partial \theta_x} = -\frac{2(\theta_x - \hat{\theta}_{x,k-1})}{2\sigma_{x,k-1}^2} \quad (3-53)$$

The data and prior information in iteration k are then given by Equations (3-54) and (3-55), respectively.

$$\begin{aligned} J_{D,k} &= \int_{\mathbb{R}} I_k(\theta_x) \lambda_{k-1}(\theta_x) d\theta_x \\ &= \frac{\theta_I \omega^2}{2M} \left(1 - \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp\left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2}\right)\right) \\ &\quad + \frac{\theta_I}{2M\sigma_{PSF}^2} \left(1 + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp\left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2}\right)\right) \end{aligned} \quad (3-54)$$

$$\begin{aligned} J_{P,k} &= \int_{\mathbb{R}} \left(\frac{\partial \log(\lambda_{k-1}(\theta_x))}{\partial \theta_x}\right)^2 \lambda_{k-1}(\theta_x) d\theta_x \\ &= \frac{1}{\sigma_{x,k-1}\sqrt{2\pi}} \int_{\mathbb{R}} \left(\frac{2(\theta_x - \hat{\theta}_{x,k-1})}{2\sigma_{x,k-1}^2}\right)^2 \exp\left(-\frac{(\theta_x - \hat{\theta}_{x,k-1})^2}{2\sigma_{x,k-1}^2}\right) d\theta_x \\ &= \frac{1}{\sigma_{x,k-1}^2} \end{aligned} \quad (3-55)$$

Then the Bayesian information is given by Equation (3-56).

$$\begin{aligned} J &= J_D + J_P \\ &= \frac{1 + \frac{\theta_I \omega^2}{2M} \sigma_{x,k-1}^2 \left(1 - \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp\left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2}\right)\right)}{\sigma_{x,k-1}^2} \\ &\quad + \frac{\frac{\theta_I \sigma_{x,k-1}^2 A}{\sigma_{PSF}^2} \left(1 + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp\left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2}\right)\right)}{\sigma_{x,k-1}^2} \end{aligned} \quad (3-56)$$

By the Van Trees inequality, we find for the localization precision:

$$\sigma_{x,k} \geq \sigma_{x,k-1} \left(1 + \frac{\theta_I \omega^2 \sigma_{x,k-1}^2}{2M} \left(1 - \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp \left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2} \right) \right) \right. \\ \left. + \frac{\theta_I \sigma_{x,k-1}^2}{2M \sigma_{PSF}^2} \left(1 + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp \left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2} \right) \right) \right)^{-\frac{1}{2}} \quad (3-57)$$

Suppose now that two patterns are used of which the intensity minima are symmetrically placed around the current estimate of the emitter position (such as in Equations (3-13), (3-14)). Let us call the pattern placed to the right of the estimate of the emitter position ϕ_k^+ and the pattern placed to the left of the estimate of the emitter position ϕ_k^- . Their combined illumination pattern can then be described as:

$$P_k(x) = \frac{1}{4M} [2 + m \cos(\omega x - \phi_{x,k}^+) + m \cos(\omega x - \phi_{x,k}^-)] \quad (3-58)$$

The derivation is entirely analogous to the derivation for one pattern. We therefore immediately present the result:

$$\sigma_{x,k} \geq \sigma_{x,k-1} \left(2 + \frac{\theta_I \omega^2 \sigma_{x,k-1}^2}{4M} \left(1 - (\cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^+) + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^-)) \exp \left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2} \right) \right) \right. \\ \left. + \frac{\theta_I \sigma_{x,k-1}^2}{4M \sigma_{PSF}^2} \left(1 + (\cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^+) + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^-)) \exp \left(-\frac{\sigma_{x,k-1}^2 \omega^2}{2} \right) \right) \right)^{-\frac{1}{2}} \quad (3-59)$$

Single camera pixel We now assume a single, infinitely sized pixel is used for the recording. Consequently, the expected photon count needs to be integrated with respect to x over the entire real axis, losing information about the PSF in the process. The updated expected photon count is given by Equation (3-60), with its first partial derivative with respect to θ_x being given by Equation (3-61).

$$\mu_k = \frac{\theta_I}{2M} [1 + \cos(\omega \theta_x - \phi_{x,k})] \quad (3-60)$$

$$\frac{\partial \mu_k}{\partial \theta_x} = -\frac{\omega \theta_I}{2M} \sin(\omega \theta_x - \phi_{x,k}) \quad (3-61)$$

The Fisher information in iteration k is then given by Equation (3-62).

$$I_k(\theta_x) = \frac{1}{\mu_k} \left(\frac{\partial \mu_k}{\partial \theta_x} \right)^2 = \frac{\left(-\frac{\omega \theta_I}{2M} \sin(\omega \theta_x - \phi_{x,k}) \right)^2}{\frac{\theta_I}{2M} [1 + \cos(\omega \theta_x - \phi_{x,k})]} \\ = \frac{\omega^2 \theta_I}{2M} (1 - \cos(\omega \theta_x - \phi_{x,k})) \quad (3-62)$$

Using the prior information of Equation (3-51), the data and prior information in iteration k are given by Equations (3-63) and (3-64), respectively.

$$J_{D,k} = \int_{\mathbb{R}} I_k(\theta_x) \lambda_{k-1}(\theta_x) d\theta_x = \frac{\omega^2 \theta_I}{2M} \left(1 - \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp\left(\frac{-\sigma_{x,k-1}^2 \omega^2}{2}\right) \right) \quad (3-63)$$

$$J_{P,k} = \int_{\mathbb{R}} \left(\frac{\partial \log(\lambda_{k-1}(\theta_x))}{\partial \theta_x} \right)^2 \lambda_{k-1}(\theta_x) d\theta_x = \frac{1}{\sigma_{x,k-1}^2} \quad (3-64)$$

Then the Bayesian information is given by Equation (3-65).

$$J = J_D + J_P = \frac{1 + \frac{\theta_I \omega^2 \sigma_{x,k-1}^2}{2M} \left(1 - \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp\left(\frac{-\sigma_{x,k-1}^2 \omega^2}{2}\right) \right)}{\sigma_{x,k-1}^2} \quad (3-65)$$

By the Van Trees inequality, we find:

$$\sigma_k \geq \frac{\sigma_{x,k-1}}{\sqrt{1 + \frac{\theta_I \omega^2 \sigma_{x,k-1}^2}{2M} \left(1 - \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}) \exp\left(\frac{-\sigma_{x,k-1}^2 \omega^2}{2}\right) \right)}} \quad (3-66)$$

Suppose now that two patterns are used of which the intensity minima are symmetrically placed around (the estimate of) the emitter position, as described by Equation (3-58). An analogous derivation then results in the following expression for the localization precision:

$$\sigma_{x,k} \geq \sigma_{x,k-1} \left(2 + \frac{\theta_I \omega^2 \sigma_{x,k-1}^2}{4M} \left(1 - (\cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^+) + \cos(\omega \hat{\theta}_{x,k-1} - \phi_{x,k}^-)) \exp\left(\frac{-\sigma_{x,k-1}^2 \omega^2}{2}\right) \right) \right)^{-\frac{1}{2}} \quad (3-67)$$

Chapter 4

Conclusion

In this chapter, we conclude the thesis. First, in Section 4-1, we will summarize the main findings of the study. In Section 4-2, we recommend further improvements on topics discussed in this text and we provide an outlook on the future of control in iterative localization microscopy.

4-1 Summary of study

Optical microscopy is a key tool in observing and understanding life on a microscopic scale. Fluorescence microscopy is particularly suited for biological applications, as it is a minimally invasive technique and as it can differentiate between structures by its high selectivity. However, due to the wave nature of light, diffraction in the optical system places a physical limit on the resolution of an optical microscope. The Abbe diffraction limit states that, for two points to be individually distinguishable, they should be a distance $\sigma_{x,y} = \frac{\lambda_{\text{em}}}{2\text{NA}}$ apart, where λ_{em} denotes the wavelength of emission light and where the numerical aperture NA depends on the microscope objective and the immersion medium. While diffraction-limited imaging may result in resolutions around 200 nm, it is not enough to differentiate between proteins, molecules or viruses, all of which are in the range of 1 nm to 100 nm [17].

Fortunately, methods have been discovered to circumvent the diffraction barrier. These methods are called super-resolution microscopy methods. Super-resolution microscopy methods break the diffraction barrier by increasing the information contained in emission signals [16]. In single-molecule localization microscopy (SMLM), fluorescent emitters in the sample are sparsely activated. By only allowing a random, sparse subset of fluorescent labels to emit at a time, there is a high probability that only a single emitter is active within a region defined by the diffraction limit. This creates a sparsity in the emission signal, which separates emission signals of different emitters in time and space. As a result of this sparsity, estimation algorithms can estimate emitter positions with increased precision over the diffraction limit. Recently, methods have been developed that combine SMLM techniques with patterned illumination. These modulation enhanced SMLM (meSMLM) methods are able to improve the

resolution of optical microscopes even further. Patterned illumination modulates the emission signal, increasing the information content of signal photons. Estimation algorithms can be used to extract this additional information, resulting in improved localization precision.

By iteratively controlling the positioning of illumination patterns, resolution can be improved locally around an emitter. This is done in iterative meSMLM (imeSMLM) methods, such as iterative MINFLUX [6]. Estimates of the emitter position and the localization precision in previous iterations are used as prior information in the current iteration. Based on the prior information, new pattern positions are selected to achieve improved localization precision with respect to earlier iterations.

Localization precision plays an important role to assess the quality of super-resolution methods. For (me)SMLM methods, the Cramér-Rao lower bound (CRLB) is often used to estimate the localization precision. The CRLB bounds the (co)variance of arbitrary unbiased estimators from below. The (co)variance of maximum likelihood estimators converges asymptotically to the CRLB and [27] shows that it is approximately attained for 100 or more signal photons in SMLM.

For imeSMLM methods, the CRLB cannot be used to accurately estimate the localization precision. As the CRLB treats estimands as deterministic unknowns (i.e. without a distribution), a prior distribution on the estimands cannot be incorporated into the bound. As such, the effect of prior information on the localization precision in imeSMLM methods is not captured by the CRLB.

The Van Trees inequality (VTI) is a Bayesian variant of the CRLB. As it treats estimands as random variables with a known prior distribution, it is able to incorporate prior information on e.g. the emitter position into its estimation of the localization precision of imeSMLM methods. It is therefore able to accurately bound the localization precision of imeSMLM methods from below. In addition, as the VTI allows us to predict the localization precision of imeSMLM methods, it opens up the possibility of optimal control for iterative localization methods.

We have modeled an imeSMLM method, that controls the position of sinusoidal intensity patterns to locally increase the localization precision around the emitter. Intensity minima of sinusoidal patterns were placed symmetrically around the current estimate of the emitter position, at a distance based on either the localization precision of the previous iteration or a predetermined constant. These strategies balance the information content of signal photons with the need for robustness to estimation errors. Using the VTI, the localization precision of this method was estimated.

We derived a fundamental limit on the localization precision of imeSMLM methods that make use of standing wave illumination patterns, in the absence of estimation uncertainty on the signal photon count and background fluorescence and with perfect modulation. This fundamental limit cannot be surpassed, as the aforementioned effects can only worsen the localization precision of imeSMLM methods. Using this limit, we have shown that under the above assumptions, the information content of signal photons increases exponentially as a function of the iteration count.

In addition, the Van Trees inequality was used to assess the performance of illumination pattern position control strategies. It was found that if the distance between intensity minima and the emitter position scales with the localization precision of the previous iteration, the

information content of signal photons is increased the most. However, this makes the method susceptible to accumulating precision loss through the effects of background and image discretization by the camera.

4-2 Outlook

In this section, we recommend further improvements on topics discussed in this text and we provide an outlook on the future of control in iterative localization microscopy. First, we discuss how the numerical evaluation of the VTI can be improved to allow for fast optimal control. We then elaborate on how the 3D-SIM illumination pattern can be used for imeSMLM in three dimensions. Lastly, we discuss iterative learning control and reinforcement learning as promising control strategies for pattern control in imeSMLM methods.

As discussed in Chapter 3, evaluation of the VTI is done through numerical evaluation in many practical scenarios. For the VTI to be usable in as a performance metric in (online) control, it is essential that it can be evaluated quickly and reliably.

The performance of the IMC estimator is not satisfactory on these criteria. For low signal photon counts, high background counts and imperfect modulation, the covariance of the IMC estimator grows sharply. In these cases, the covariance of the prior distribution $\lambda_{k-1}(\boldsymbol{\theta})$ (which is used as a sampling distribution in IMC) grows, which widens the distribution and increases the amount of samples needed for a precise evaluation of the VTI. This is undesirable, as the covariance of the IMC estimator scales with the sample covariance. Therefore, to reduce the covariance of the VTI by 10 times, approximately 100 times as many samples are needed. This slows evaluation of the VTI down significantly.

Furthermore, the covariance of the IMC estimator decreases the reliability of control methods based on the VTI. The covariance of the IMC estimator perturbs function evaluations of the VTI by a random error. This is problematic for optimization routines that rely on function evaluations, as evaluated values of the VTI behave like random variables. This uncertainty on the function value decreases convergence speed and it may even prevent convergence.

To enable (optimal) control that uses the VTI as a performance metric, the speed and reliability of its evaluation should thus be increased. We advise to use integration methods for this, that efficiently distribute samples over an infinite domain of integration. An example of such a method is Bayesian quadrature [52]. In contrast to IMC, Bayesian quadrature treats the numerical estimation problem as a statistical inference problem. A prior distribution (usually a Gaussian process) is placed on the integrand and it is used to select function evaluations that minimize the (co)variance of the posterior distribution of the integral. Doing this maximizes the information gained by function evaluations, as the similarity of neighboring function values is exploited.

Next, we elaborate further on the potential of 3D-SIM pattern control. In this text, sinusoidal intensity patterns were used in an imeSMLM method to locally improve localization precision in two dimensions. A similar methodology can be used to locally improve localization precision in three dimensions, by making use of the 3D-SIM pattern [50]. The 3D-SIM pattern has structure in lateral and axial directions, as shown in the simulation of Figure 4-1.

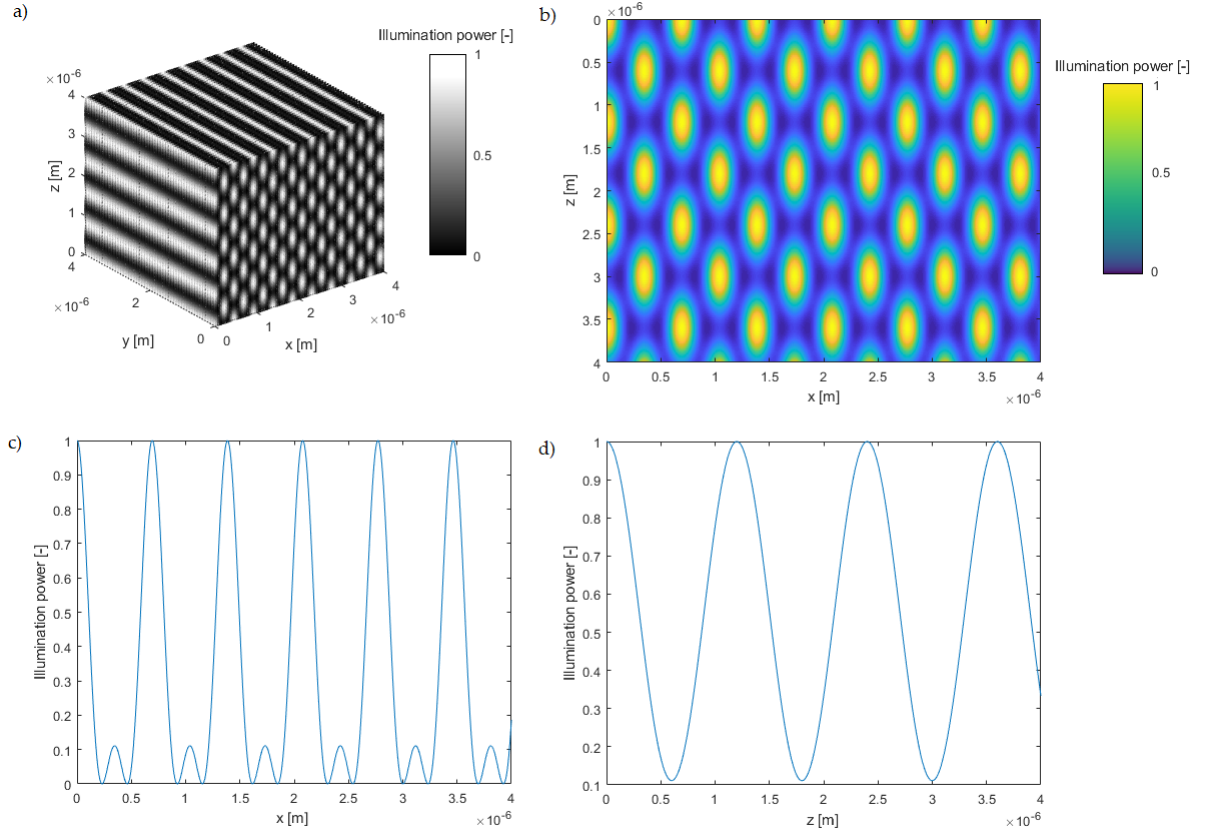


Figure 4-1: Simulated 3D-SIM illumination pattern. (a): 3D view of the pattern. (b): Cross-section of pattern in (x, z) -plane, for $y = 0$ nm. (c): Cross-section of pattern on x -axis, for $y = 0$ nm, $z = 0$ nm. (d): Cross-section of pattern on z -axis, for $x = 0$ nm, $y = 0$ nm.

From Figure 4-1, it can be seen that the 3D-SIM intensity is sinusoidal in the z -direction and that it consists of two overlaid sinusoids of different frequencies in the x - and y -directions. As such, a similar approach to that of Chapter 3 can be followed, where intensity minima of the 3D-SIM pattern are placed symmetrically around the emitter position in the x -, y - and z -directions, effectively boxing in the emitter between intensity minima.

For optimal control with the 3D-SIM pattern, a performance metric should weigh lateral and axial precision. In [53], the σ_{3D} metric is used for this. σ_{3D} is given by:

$$\sigma_{3D} = \frac{4\pi}{3} \sqrt{\det(I_{xyz}^{-1}(\boldsymbol{\theta}))} \quad (4-1)$$

where $I_{xyz}(\boldsymbol{\theta})$ denotes the Fisher information related to the emitter position $(\theta_x, \theta_y, \theta_z)$. A similar criterion could be used for the VTI, where only the components of the Bayesian information matrix related to the emitter position $(\theta_x, \theta_y, \theta_z)$ are included in the measure. The σ_{3D} metric weighs the individual localization precisions in x -, y - and z -directions based on their covariance. Using σ_{3D} could thus make for a challenging multivariable pattern placement control problem, where the total signal photon budget needs to be distributed among axial and lateral localization precision.

We have seen that the VTI can be used to assess the performance of imeSMLM methods, based on their localization precision for a single emitter. In a simulation, minimization of the VTI is therefore a satisfactory way to determine optimal pattern positions. In practice, model inaccuracies and disturbances likely require control to acquire improved resolution.

Three control objectives are of particular relevance for control of imeSMLM methods. First of all, the localization precision should be as high as possible for the chosen pattern positions. We have also discussed the importance of control strategies that are robust to estimation errors. In practice, multiple emitters are imaged at a time. Hence, there is practical importance to maximize robustness and performance over a range of emitter positions.

Based on these control objectives, we expect intelligent control methods to be particularly relevant. We will briefly discuss the potential of iterative learning control (ILC) and reinforcement learning (RL).

ILC is an open-loop control strategy suited for repetitive processes, that improves the performance of a system by learning from earlier iterations [54]. By using the error signal of the previous iteration, the previous control action is slightly altered to obtain a new control action, which should result in improved system performance. By repeating this process, optimal control actions can effectively be learned.

In the proposed imeSMLM method, pattern positions are repeatedly selected to minimize the VTI. As such, iterative learning control may be used to iteratively improve pattern positions. The choice of the error signal is important here, as it should be a measure for the distance between the true emitter position and its estimate. A possible choice for the error signal could be the difference between the likelihood functions of the current and previous measurements. If the current pattern position is better than the previous one, the likelihood of the current measurements is larger than that of the previous measurements, and vice versa. Furthermore, the difference between the likelihood functions of the current and previous measurements gets smaller as the patterns converge to the optimal pattern, which will also cause convergence of the ILC approach.

Another promising intelligent control method is reinforcement learning (RL) [55]. In RL, an agent (the controller) interacts with its environment (the process). This changes the state of the environment, for better or for worse. A reward policy is used to assess the new state of the environment, after which the agent is adapted based on this reward policy.

RL is a suitable candidate for pattern position control in imeSMLM. The controller determines pattern positions to improve the localization precision. The likelihood of the current measurements can be used as a reward policy, to assess the effectiveness of a pattern placement. This can be used to update the controller, such that it learns optimal pattern positions over time. Alternatively, the log-likelihood of the current measurements can be used as a reward policy, to accelerate the convergence speed of RL at the cost of robustness.

Appendix A

Proofs and derivations

This appendix contains proofs and derivations which were referred to in the main text.

A-1 Proof of Theorem 2.1

Let $\theta \in \mathbb{R}$ arbitrary. Before we prove the theorem, we will first show that the regularity condition $\mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right] = 0$ follows from the assumptions on $L(\theta|\mathbf{x})$.

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right] &= \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} L(\theta|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \frac{1}{L(\theta|\mathbf{x})} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta} L(\theta|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta} d\mathbf{x} \end{aligned}$$

By assumption, we know that $L(\theta|\mathbf{x})$ is continuously differentiable with respect to θ and that its support $\{\mathbf{x} | L(\theta|\mathbf{x}) > 0\}$ does not depend on θ . It must therefore hold that $L(\theta|\mathbf{x})$ either has a bounded support in \mathbf{x} without dependency on θ , or that it has infinite support, is continuously differentiable with respect to θ and that the integral converges uniformly for all θ . In these cases, we may use Leibniz's integral rule to interchange the differentiation and integration operators [56]. We then obtain:

$$\mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right] = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} L(\theta|\mathbf{x}) d\mathbf{x} = \frac{\partial}{\partial \theta} (1) = 0$$

We now start the proof of the theorem. We start by a general proof of the CRLB for a scalar parameter $\alpha = g(\theta)$, where the probability density function is parametrized by θ . Later, we will apply this to the specific case $g(\theta) = \theta$, which will prove the theorem.

Consider any unbiased estimator $\hat{\alpha}$ of α . By unbiasedness, it must follow that:

$$\mathbb{E}[\hat{\alpha}] = \int_{\mathbb{R}^n} \hat{\alpha} L(\theta|\mathbf{x}) d\mathbf{x} = \alpha = g(\theta)$$

We differentiate both sides of the equation with respect to θ . Note that we can again apply Leibniz's integral rule to interchange differentiation and integration, by the assumptions on $L(\theta|\mathbf{x})$. We obtain:

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} \hat{\alpha} L(\theta|\mathbf{x}) d\mathbf{x} &= \frac{\partial}{\partial \theta} g(\theta) \\ \int_{\mathbb{R}^n} \hat{\alpha} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta} d\mathbf{x} &= \frac{\partial g(\theta)}{\partial \theta} \end{aligned}$$

We can now subtract a term $\alpha \mathbb{E} \left[\frac{\partial L(\theta|\mathbf{x})}{\partial \theta} \right] = \int_{\mathbb{R}^n} \alpha \frac{\partial L(\theta|\mathbf{x})}{\partial \theta} L(\theta|\mathbf{x}) d\mathbf{x}$. By the derived regularity condition, this term must be equal to zero, which means we can freely subtract it from the left-hand side of the equation. We then find:

$$\int_{\mathbb{R}^n} (\hat{\alpha} - \alpha) \frac{\partial L(\theta|\mathbf{x})}{\partial \theta} d\mathbf{x} = \frac{\partial g(\theta)}{\partial \theta}$$

Note as before that $\frac{\partial L(\theta|\mathbf{x})}{\partial \theta} = \frac{1}{L(\theta|\mathbf{x})} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta} L(\theta|\mathbf{x}) = \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} L(\theta|\mathbf{x})$. We substitute this into the expression on the left-hand side, to get:

$$\begin{aligned} \int_{\mathbb{R}^n} (\hat{\alpha} - \alpha) \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} L(\theta|\mathbf{x}) d\mathbf{x} &= \frac{\partial g(\theta)}{\partial \theta} \\ \mathbb{E} \left[(\hat{\alpha} - \alpha) \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right] &= \frac{\partial g(\theta)}{\partial \theta} \end{aligned}$$

Next, we make use of the Cauchy-Schwarz inequality. We use the following formulation, as presented in [56]:

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2]$$

Application of the Cauchy-Schwarz inequality on the left-hand side of the equation, with $X = (\hat{\alpha} - \alpha)$ and $Y = \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta}$ then gives:

$$\left(\frac{\partial g(\theta)}{\partial \theta} \right)^2 = \mathbb{E} \left[(\hat{\alpha} - \alpha) \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right]^2 \leq \mathbb{E} [(\hat{\alpha} - \alpha)^2] \mathbb{E} \left[\left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right)^2 \right]$$

Note that $\text{Var}(\hat{\alpha}) = \mathbb{E} [(\hat{\alpha} - \alpha)^2]$ by definition. Substitution of this fact and some algebraic reordering then gives:

$$\left(\frac{\partial g(\theta)}{\partial \theta} \right)^2 \leq \text{Var}(\hat{\alpha}) \cdot \mathbb{E} \left[\left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right)^2 \right]$$

$$\text{Var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{\mathbb{E}\left[\left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta}\right)^2\right]}$$

Lastly, under the assumptions on $L(\theta|\mathbf{x})$ that allow application of Leibniz's integral rule, we can apply Lemma A.1 (which is shown and proven in Appendix A-4) to conclude that $\mathbb{E}\left[\left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta}\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2 \ell(\theta|\mathbf{x})}{\partial \theta^2}\right]$. This gives the following formulation of the variance bound for the general case $\alpha = g(\theta)$:

$$\text{Var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-\mathbb{E}\left[\frac{\partial^2 \ell(\theta|\mathbf{x})}{\partial \theta^2}\right]}$$

For the special case under consideration, choose $g(\theta) = \theta$. Then $\frac{\partial g(\theta)}{\partial \theta} = 1$, which returns the Cramér-Rao lower bound:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left[\frac{\partial^2 \ell(\theta|\mathbf{x})}{\partial \theta^2}\right]} := \text{CRLB}(\theta)$$

as required. \square

A-2 Proof of Theorem 2.2

Let $\theta \in \mathbb{R}^m$ arbitrary. Before we prove the theorem, we will first show that the regularity condition $\mathbb{E}\left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i}\right] = 0$ follows from the assumptions on $L(\theta|\mathbf{x})$ for all $i \in \{1, 2, \dots, m\}$.

$$\begin{aligned} \mathbb{E}\left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i}\right] &= \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i} L(\theta|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \frac{1}{L(\theta|\mathbf{x})} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_i} L(\theta|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_i} d\mathbf{x} \end{aligned}$$

By assumption, we know that $L(\theta|\mathbf{x})$ is continuously differentiable with respect to θ and that its support $\{\mathbf{x} | L(\theta|\mathbf{x}) > 0\}$ does not depend on θ . It must therefore hold that $L(\theta|\mathbf{x})$ either has a bounded support in \mathbf{x} without dependency on θ , or that it has infinite support, is continuously differentiable with respect to θ and that the integral converges uniformly for all θ . In these cases, we may use Leibniz's integral rule to interchange the differentiation and integration operators [56]. We then obtain:

$$\mathbb{E}\left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i}\right] = \frac{\partial}{\partial \theta_i} \int_{\mathbb{R}^n} L(\theta|\mathbf{x}) d\mathbf{x} = \frac{\partial}{\partial \theta_i}(1) = 0$$

We now start the proof of the theorem. We start by a general proof of the CRLB for a parameter vector $\alpha = g(\theta)$, where g acts on all entries of θ separately and where the

probability density function is parametrized by $\boldsymbol{\theta}$. Later, we will apply this to the specific case $g(\boldsymbol{\theta}) = \boldsymbol{\theta}$, which will prove the theorem.

Consider any unbiased estimator $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. By unbiasedness, it must follow that for all $i \in \{1, 2, \dots, m\}$:

$$\mathbb{E}[\hat{\boldsymbol{\alpha}}_i] = \int_{\mathbb{R}^n} \hat{\boldsymbol{\alpha}}_i L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} = \boldsymbol{\alpha}_i = g(\boldsymbol{\theta}_i)$$

We differentiate both sides of the equation with respect to $\boldsymbol{\theta}_i$. Note that we can apply Leibniz's integral rule to interchange differentiation and integration, by the assumptions on $L(\boldsymbol{\theta}|\mathbf{x})$. We obtain:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}_i} \int_{\mathbb{R}^n} \hat{\boldsymbol{\alpha}}_i L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} &= \frac{\partial}{\partial \boldsymbol{\theta}_i} g(\boldsymbol{\theta}_i) \\ \int_{\mathbb{R}^n} \hat{\boldsymbol{\alpha}}_i \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} d\mathbf{x} &= \frac{\partial g(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i} \end{aligned}$$

We can now subtract a term $\boldsymbol{\alpha}_i \mathbb{E} \left[\frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} \right] = \int_{\mathbb{R}^n} \boldsymbol{\alpha}_i \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x}$. By the derived regularity condition, this term must be equal to zero, which means we can freely subtract it from the left-hand side of the equation. We then find:

$$\int_{\mathbb{R}^n} (\hat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i) \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} d\mathbf{x} = \frac{\partial g(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i}$$

Note as before that $\frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} = \frac{1}{L(\boldsymbol{\theta}|\mathbf{x})} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} L(\boldsymbol{\theta}|\mathbf{x}) = \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} L(\boldsymbol{\theta}|\mathbf{x})$. We substitute this into the expression on the left-hand side, to get:

$$\int_{\mathbb{R}^n} (\hat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i) \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} = \frac{\partial g(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i}$$

Consider now the case $i \neq j$, where both $i, j \in \{1, 2, \dots, m\}$. Let us consider the following integral:

$$\begin{aligned} \int_{\mathbb{R}^n} (\hat{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i) \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} &= \int_{\mathbb{R}^n} \hat{\boldsymbol{\alpha}}_i \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} - \int_{\mathbb{R}^n} \boldsymbol{\alpha}_i \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \hat{\boldsymbol{\alpha}}_i \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} d\mathbf{x} - \boldsymbol{\alpha}_i \int_{\mathbb{R}^n} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \hat{\boldsymbol{\alpha}}_i \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} d\mathbf{x} - \boldsymbol{\alpha}_i \mathbb{E} \left[\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} \right] \end{aligned}$$

Note that by Leibniz's integral rule, we may interchange the differentiation and integration operators in the first term. Furthermore, by the derived regularity condition, the second term is equal to zero. The integral thus reduces to:

$$\begin{aligned} \int_{\mathbb{R}^n} (\hat{\alpha}_i - \alpha_i) \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} &= \frac{\partial}{\partial \theta_j} \int_{\mathbb{R}^n} \hat{\alpha}_i L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\ &= \frac{\partial}{\partial \theta_j} \mathbb{E}[\hat{\alpha}_i] = \frac{\partial g(\theta_i)}{\partial \theta_j} \end{aligned}$$

We have now derived the result $\int_{\mathbb{R}^n} (\hat{\alpha}_i - \alpha_i) \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} = \frac{\partial g(\theta_i)}{\partial \theta_j}$ for $i = j$ and $i \neq j$. For convenience, let us combine these results in vector-matrix form, as follows:

$$\int_{\mathbb{R}^n} (\hat{\alpha} - \alpha) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} = \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Let us now pre-multiply the above equation by \mathbf{a}^T and post-multiply by \mathbf{b} , where \mathbf{a} and \mathbf{b} are arbitrary real vectors of appropriate dimensions, which do not depend on \mathbf{x} . We then have:

$$\begin{aligned} \int_{\mathbb{R}^n} \mathbf{a}^T (\hat{\alpha} - \alpha) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{b} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} &= \mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \\ \mathbb{E} \left[\mathbf{a}^T (\hat{\alpha} - \alpha) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{b} \right] &= \mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \end{aligned}$$

From [56], we infer that the expectation operator is an inner product on \mathbb{R}^n and hence, we may use the Cauchy-Schwarz inequality. In vector form, it is given by the following expression [57]:

$$\left(\mathbb{E} [\mathbf{X}^T \mathbf{Y}] \right)^2 \leq \mathbb{E} [\mathbf{X}^T \mathbf{X}] \mathbb{E} [\mathbf{Y}^T \mathbf{Y}]$$

Here, \mathbf{X} and \mathbf{Y} are arbitrary real vectors of appropriate dimensions.

Application of the Cauchy-Schwarz inequality on the left-hand side of the equation, with $\mathbf{X} = (\hat{\alpha} - \alpha)^T \mathbf{a}$ and $\mathbf{Y} = \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{b}$ then gives (where $C_{\hat{\alpha}}$ denotes the covariance matrix of $\hat{\alpha}$):

$$\begin{aligned} \left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \right)^2 &= \left(\mathbb{E} \left[\mathbf{a}^T (\hat{\alpha} - \alpha) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{b} \right] \right)^2 \\ \left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \right)^2 &\leq \mathbb{E} \left[\mathbf{a}^T (\hat{\alpha} - \alpha) (\hat{\alpha} - \alpha)^T \mathbf{a} \right] \mathbb{E} \left[\mathbf{b}^T \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{b} \right] \\ \left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \right)^2 &\leq \mathbf{a}^T C_{\hat{\alpha}} \mathbf{a} \mathbf{b}^T \mathbb{E} \left[\left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \right] \mathbf{b} \end{aligned}$$

Note that for each pair $i, j \in \{1, 2, \dots, m\}$ we have by Lemma A.1 (which is proven in Appendix A-4):

$$\mathbb{E} \left[\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_j} \right] = -\mathbb{E} \left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \right] = I_{ij}(\boldsymbol{\theta})$$

where $I_{ij}(\boldsymbol{\theta})$ denotes entry (i, j) of the Fisher information matrix. Substituting this into the inequality gives:

$$\left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \right)^2 \leq \mathbf{a}^T C_{\hat{\alpha}} \mathbf{a} \mathbf{b}^T I(\boldsymbol{\theta}) \mathbf{b}$$

Consider now the off-diagonal entries of $I(\boldsymbol{\theta})$. If $i \neq j$, then $I_{ij}(\boldsymbol{\theta}) = I_{ji}(\boldsymbol{\theta})$, as the partial derivatives are interchangeable (by Fubini's theorem [56]). Therefore, $I(\boldsymbol{\theta})$ and its inverse are symmetric, which means $I^{-T}(\boldsymbol{\theta}) = I^{-1}(\boldsymbol{\theta})$. Also, for some vector \mathbf{u} of appropriate dimensions, we have:

$$\begin{aligned} \mathbf{u}^T I(\boldsymbol{\theta}) \mathbf{u} &= \mathbf{u}^T \mathbb{E} \left[\left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \right] \mathbf{u} \\ &= \mathbb{E} \left[\mathbf{u}^T \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{u} \right] \\ &= \mathbb{E} \left[\left\| \left(\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{u} \right\|_2^2 \right] \geq 0 \end{aligned}$$

Hence, $I(\boldsymbol{\theta})$ is positive semi-definite. Therefore, all its eigenvalues are non-negative. Furthermore, as $I(\boldsymbol{\theta})$ is assumed to be invertible, it must have no eigenvalues equal to zero. This means $I(\boldsymbol{\theta})$ has only strictly positive eigenvalues, which means it is positive definite.

Recall that \mathbf{a} and \mathbf{b} were chosen to be arbitrary real vectors of appropriate dimensions, which do not depend on \mathbf{x} . We can now retrieve the Cramér-Rao lower bound from this expression by a smart choice of \mathbf{b} . Choose $\mathbf{b} = I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{a}$. This gives us:

$$\begin{aligned} \left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{a} \right)^2 &\leq \mathbf{a}^T C_{\hat{\alpha}} \mathbf{a} \mathbf{a}^T \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) I^{-T}(\boldsymbol{\theta}) I(\boldsymbol{\theta}) I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{a} \\ \left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{a} \right)^2 &\leq \mathbf{a}^T C_{\hat{\alpha}} \mathbf{a} \mathbf{a}^T \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) I^{-T}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{a} \end{aligned}$$

As $I(\boldsymbol{\theta})$ is symmetric, so is its inverse, which means $I^{-T}(\boldsymbol{\theta}) = I^{-1}(\boldsymbol{\theta})$. This gives:

$$\left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{a} \right)^2 \leq \mathbf{a}^T C_{\hat{\alpha}} \mathbf{a} \left(\mathbf{a}^T \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \mathbf{a} \right)$$

Furthermore, as $I(\boldsymbol{\theta})$ is positive definite, $I^{-1}(\boldsymbol{\theta})$ is also positive definite. This means the term $\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T$ is at least positive semi-definite (as $\left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)$ could be the zero

vector). Lastly, it must then also hold that $\left(\mathbf{a}^T \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \mathbf{a}\right)$ is at least positive semi-definite and since it is scalar, it must be non-negative. We therefore have:

$$\mathbf{a}^T \left(C_{\hat{\alpha}} - \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \right) \mathbf{a} \geq 0$$

As \mathbf{a} was arbitrary, we can conclude that $\left(C_{\hat{\alpha}} - \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} I^{-1}(\boldsymbol{\theta}) \left(\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \right) \succeq \mathbf{0}$.

For the special case under consideration, choose $g(\boldsymbol{\theta}) = \boldsymbol{\theta}$. Then $\frac{\partial g(\boldsymbol{\theta})}{\partial \theta_i} = 1$ for all i , which returns the Cramér-Rao lower bound:

$$\left(C_{\hat{\alpha}} - I^{-1}(\boldsymbol{\theta}) \right) \succeq \mathbf{0}$$

as required. □

A-3 Proof of Theorem 2.3

First, recall from the proof in Appendix A-1 that by the assumptions on the likelihood function, the regularity condition $\mathbb{E}_L \left[\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}} \right] = \mathbf{0}$ is satisfied. Furthermore, by the assumptions on the prior distribution, we have:

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{\theta} \\ &= [L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})]_{-\infty}^{\infty} = 0 \end{aligned}$$

Suppose we have an estimator $\hat{\boldsymbol{\theta}}(\mathbf{x})$. This estimator does not depend on $\boldsymbol{\theta}$ by definition and hence we have:

$$\int_{-\infty}^{\infty} \hat{\boldsymbol{\theta}}(\mathbf{x}) \frac{\partial L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(\mathbf{x}) \int_{-\infty}^{\infty} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{\theta} = \mathbf{0}$$

We will expand the term $\int_{-\infty}^{\infty} \boldsymbol{\theta} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{\theta}$ using integration by parts. Under the assumptions on $\lambda(\boldsymbol{\theta})$, we have:

$$\begin{aligned} \int_{-\infty}^{\infty} \boldsymbol{\theta} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{\theta} &= [\boldsymbol{\theta} L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= - \int_{-\infty}^{\infty} L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned}$$

We combine the previous two equations to get:

$$\int_{-\infty}^{\infty} (\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}) \frac{\partial L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\boldsymbol{\theta} = \int_{-\infty}^{\infty} L(\boldsymbol{\theta}|\mathbf{x}) \lambda(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

We integrate both sides of the previous equation over \mathbf{x} with the domain of integration being \mathbb{R}^n . We then obtain:

$$\begin{aligned} \int_{\mathbb{R}^n} \int_{-\infty}^{\infty} (\hat{\theta}(\mathbf{x}) - \theta) \frac{\partial L(\theta|\mathbf{x})\lambda(\theta)}{\partial \theta} d\theta d\mathbf{x} &= \int_{\mathbb{R}^n} \int_{-\infty}^{\infty} L(\theta|\mathbf{x})\lambda(\theta) d\theta d\mathbf{x} \\ \int_{\mathbb{R}^n} \int_{-\infty}^{\infty} (\hat{\theta}(\mathbf{x}) - \theta) \frac{\partial L(\theta|\mathbf{x})\lambda(\theta)}{\partial \theta} d\theta d\mathbf{x} &= \int_{\mathbb{R}^n} \int_{-\infty}^{\infty} p(\mathbf{x}, \theta) d\theta d\mathbf{x} \\ \int_{\mathbb{R}^n} \int_{-\infty}^{\infty} (\hat{\theta}(\mathbf{x}) - \theta) \frac{\partial L(\theta|\mathbf{x})\lambda(\theta)}{\partial \theta} d\theta d\mathbf{x} &= 1 \end{aligned}$$

where $p(\mathbf{x}, \theta)$ denotes the joint probability density function of \mathbf{x} and θ . We continue to work on the previous equation, by dividing and multiplying by $L(\theta|\mathbf{x})\lambda(\theta)$:

$$\begin{aligned} \int_{\mathbb{R}^n} \int_{-\infty}^{\infty} (\hat{\theta}(\mathbf{x}) - \theta) \frac{1}{L(\theta|\mathbf{x})\lambda(\theta)} \frac{\partial L(\theta|\mathbf{x})\lambda(\theta)}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\theta d\mathbf{x} &= 1 \\ \int_{\mathbb{R}^n} \int_{-\infty}^{\infty} (\hat{\theta}(\mathbf{x}) - \theta) \frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\theta d\mathbf{x} &= 1 \end{aligned}$$

Note that we can freely swap the order of integration, as no function depends on both \mathbf{x} and θ and as no bound contains a dependency on either \mathbf{x} or θ . Therefore, by the Fubini-Tonelli theorem [56], we have:

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} (\hat{\theta}(\mathbf{x}) - \theta) \frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta &= 1 \\ \mathbb{E}_{\lambda, L} \left[(\hat{\theta}(\mathbf{x}) - \theta) \frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} \right] &= 1 \end{aligned}$$

We now apply the Cauchy-Schwarz inequality, as discussed in Appendix A-1. That is, $(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$. We then have:

$$\mathbb{E}_{\lambda, L} \left[(\hat{\theta}(\mathbf{x}) - \theta)^2 \right] \mathbb{E}_{\lambda, L} \left[\left(\frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} \right)^2 \right] \geq 1$$

Note that by definition, $\text{MSE}(\hat{\theta}) = \mathbb{E}_{\lambda, L} \left[(\hat{\theta}(\mathbf{x}) - \theta)^2 \right]$. By rewriting the above inequality, we obtain:

$$\text{MSE}(\hat{\theta}) \geq \frac{1}{\mathbb{E}_{\lambda, L} \left[\left(\frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} \right)^2 \right]}$$

Note that, by the definition of the conditional probability, $p(\mathbf{x}, \theta) = L(\theta|\mathbf{x})\lambda(\theta)$. Therefore, the Bayesian information matrix J is given by

$$J = \mathbb{E}_{\lambda, L} \left[\left(\frac{\partial \log(p(\mathbf{x}, \theta))}{\partial \theta} \right)^2 \right] = \mathbb{E}_{\lambda, L} \left[\left(\frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} \right)^2 \right]$$

which proves the first part of Theorem 2.3.

We now show that this Bayesian information matrix can be split up in a data matrix J_D and a prior matrix J_P . We start at the expression for J :

$$\begin{aligned}
\mathbb{E}_{\lambda, L} \left[\left(\frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} \right)^2 \right] &= \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \left(\frac{\partial \ln(L(\theta|\mathbf{x})\lambda(\theta))}{\partial \theta} \right)^2 L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&= \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \left(\frac{\partial \ln(L(\theta|\mathbf{x}))}{\partial \theta} + \frac{\partial \ln(\lambda(\theta))}{\partial \theta} \right)^2 L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&= \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \left(\frac{\partial \ln(L(\theta|\mathbf{x}))}{\partial \theta} \right)^2 L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&\quad + 2 \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\partial \ln(L(\theta|\mathbf{x}))}{\partial \theta} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&\quad + \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \left(\frac{\partial \ln(\lambda(\theta))}{\partial \theta} \right)^2 L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&= \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right)^2 L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&\quad + 2 \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\partial \ln(L(\theta|\mathbf{x}))}{\partial \theta} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&\quad + \int_{-\infty}^{\infty} \left(\frac{\partial \ln(\lambda(\theta))}{\partial \theta} \right)^2 \lambda(\theta) \int_{\mathbb{R}^n} L(\theta|\mathbf{x}) d\mathbf{x} d\theta \\
&= \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \left(\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right)^2 L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&\quad + 2 \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta \\
&\quad + \int_{-\infty}^{\infty} \left(\frac{\partial \ln(\lambda(\theta))}{\partial \theta} \right)^2 \lambda(\theta) d\theta \\
&= J_D + 2 \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta + J_P
\end{aligned}$$

We now only have to show that $\int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x})\lambda(\theta) d\mathbf{x} d\theta = 0$ to complete the proof. We do so as follows:

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} L(\theta|\mathbf{x}) \lambda(\theta) d\mathbf{x} d\theta \\
&= \int_{-\infty}^{\infty} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} \lambda(\theta) \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} L(\theta|\mathbf{x}) d\mathbf{x} d\theta \\
&= \int_{-\infty}^{\infty} \frac{\partial \ln(\lambda(\theta))}{\partial \theta} \lambda(\theta) \mathbb{E}_L \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right] d\theta = 0
\end{aligned}$$

where the last result follows from the regularity condition $\mathbb{E}_L \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta} \right] = 0$. This completes the proof. \square

A-4 Proof of Lemma A.1

In this section, we prove the following auxiliary lemma:

Lemma A.1. *Suppose that a likelihood function $L(\theta|\mathbf{x})$ is continuously differentiable with respect to θ and that its support $\{\mathbf{x} | L(\theta|\mathbf{x}) > 0\}$ does not depend on θ . For the entries of the Fisher information matrix, we then have:*

$$I_{ij}(\theta) = -\mathbb{E} \left[\frac{\partial^2 \ell(\theta|\mathbf{x})}{\partial \theta_i \partial \theta_j} \right] = \mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_j} \right]$$

Before we prove the lemma, we will first show that the regularity condition $\mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i} \right] = 0$ follows from the assumptions on $L(\theta|\mathbf{x})$ for all i .

$$\begin{aligned}
\mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i} \right] &= \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i} L(\theta|\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{1}{L(\theta|\mathbf{x})} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_i} L(\theta|\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_i} d\mathbf{x}
\end{aligned}$$

By assumption, we know that $L(\theta|\mathbf{x})$ is continuously differentiable with respect to θ and that its support $\{\mathbf{x} | L(\theta|\mathbf{x}) > 0\}$ does not depend on θ . It must therefore hold that $L(\theta|\mathbf{x})$ either has a bounded support in \mathbf{x} without dependency on θ , or that it has infinite support, is continuously differentiable with respect to θ and that the integral converges uniformly for all θ . In these cases, we may use Leibniz's integral rule to interchange the differentiation and integration operators [56]. We then obtain:

$$\mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_i} \right] = \frac{\partial}{\partial \theta_i} \int_{\mathbb{R}^n} L(\theta|\mathbf{x}) d\mathbf{x} = \frac{\partial}{\partial \theta_i} (1) = 0$$

Let i, j arbitrary. We start the proof with the definition of the Fisher information as given in Theorem 2.2:

$$\begin{aligned}
I_{ij}(\boldsymbol{\theta}) &= -\mathbb{E} \left[\frac{\partial^2 \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \right] \\
&= -\mathbb{E} \left[\frac{\partial^2 \log(L(\boldsymbol{\theta}|\mathbf{x}))}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \right]
\end{aligned}$$

Note that for the case $i = j$, we have

$$\begin{aligned}
\frac{\partial^2 \log(L(\boldsymbol{\theta}|\mathbf{x}))}{\partial \boldsymbol{\theta}_i^2} &= \frac{\partial}{\partial \boldsymbol{\theta}_i} \left[\frac{1}{L(\boldsymbol{\theta}|\mathbf{x})} \cdot \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} \right] \\
&= \frac{\frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} - \left(\frac{\frac{\partial}{\partial \boldsymbol{\theta}_i} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right)^2 \\
&= \frac{\frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} - \left(\frac{\partial}{\partial \boldsymbol{\theta}_i} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right)^2
\end{aligned}$$

And therefore

$$\begin{aligned}
I_{ii}(\boldsymbol{\theta}) &= -\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} - \left(\frac{\partial}{\partial \boldsymbol{\theta}_i} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right)^2 \right] \\
&= -\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] + \mathbb{E} \left[\left(\frac{\partial}{\partial \boldsymbol{\theta}_i} \ell(\boldsymbol{\theta}|\mathbf{x}) \right)^2 \right]
\end{aligned}$$

For the statement to be true, we now need to show $\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] = 0$. We do so as follows:

$$\begin{aligned}
\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] &= \int_{\mathbb{R}^n} \frac{\frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{\partial^2}{\partial \boldsymbol{\theta}_i^2} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{\partial}{\partial \boldsymbol{\theta}_i} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{\partial}{\partial \boldsymbol{\theta}_i} \frac{1}{L(\boldsymbol{\theta}|\mathbf{x})} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{\partial}{\partial \boldsymbol{\theta}_i} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}_i} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x}
\end{aligned}$$

By again applying Leibniz's integral rule, we interchange differentiation and integration and we have:

$$\begin{aligned}\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] &= \frac{\partial}{\partial \theta_i} \int_{\mathbb{R}^n} \frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_i} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\ &= \frac{\partial}{\partial \theta_i} \mathbb{E} \left[\frac{\partial \ell(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_i} \right] = 0\end{aligned}$$

where the last equality results from the regularity condition. As $\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i^2} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] = 0$, we can write for $I_{ii}(\boldsymbol{\theta})$:

$$I_{ii}(\boldsymbol{\theta}) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_i} \ell(\boldsymbol{\theta}|\mathbf{x}) \right)^2 \right]$$

which completes the proof for the case $i = j$.

Now consider the case $i \neq j$. For this case, we have

$$\begin{aligned}\frac{\partial^2 \log(L(\boldsymbol{\theta}|\mathbf{x}))}{\partial \theta_i \partial \theta_j} &= \frac{\partial}{\partial \theta_i} \left[\frac{1}{L(\boldsymbol{\theta}|\mathbf{x})} \cdot \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_j} \right] \\ &= \frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} - \frac{\frac{\partial}{\partial \theta_i} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \frac{\frac{\partial}{\partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \\ &= \frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} - \left(\frac{\partial}{\partial \theta_i} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right) \left(\frac{\partial}{\partial \theta_j} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right)\end{aligned}$$

And therefore

$$\begin{aligned}I_{ij}(\boldsymbol{\theta}) &= -\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} - \left(\frac{\partial}{\partial \theta_i} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right) \left(\frac{\partial}{\partial \theta_j} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right) \right] \\ &= -\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] + \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_i} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right) \left(\frac{\partial}{\partial \theta_j} \log(L(\boldsymbol{\theta}|\mathbf{x})) \right) \right]\end{aligned}$$

For the statement to be true, we now need to show $\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] = 0$. We do so as follows:

$$\begin{aligned}\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} \right] &= \int_{\mathbb{R}^n} \frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x})}{L(\boldsymbol{\theta}|\mathbf{x})} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\boldsymbol{\theta}|\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta_i} \frac{\partial L(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_j} d\mathbf{x}\end{aligned}$$

$$\begin{aligned}
&= \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta_i} \frac{1}{L(\theta|\mathbf{x})} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_j} L(\theta|\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta_i} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_j} L(\theta|\mathbf{x}) d\mathbf{x}
\end{aligned}$$

By again applying Leibniz's integral rule, we interchange differentiation and integration and we have:

$$\begin{aligned}
\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\theta|\mathbf{x})}{L(\theta|\mathbf{x})} \right] &= \frac{\partial}{\partial \theta_i} \int_{\mathbb{R}^n} \frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_j} L(\theta|\mathbf{x}) d\mathbf{x} \\
&= \frac{\partial}{\partial \theta_i} \mathbb{E} \left[\frac{\partial \ell(\theta|\mathbf{x})}{\partial \theta_j} \right] = 0
\end{aligned}$$

where the last equality results from the regularity condition. As $\mathbb{E} \left[\frac{\frac{\partial^2}{\partial \theta_i \partial \theta_j} L(\theta|\mathbf{x})}{L(\theta|\mathbf{x})} \right] = 0$, we can write for $I_{ij}(\theta)$:

$$I_{ij}(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_i} \ell(\theta|\mathbf{x}) \right) \left(\frac{\partial}{\partial \theta_j} \ell(\theta|\mathbf{x}) \right) \right]$$

which completes the proof for the case $i = j$. As the statement holds for both $i = j$ and $i \neq j$, this completes the proof. \square

Bibliography

- [1] H. C. Ishikawa-Ankerhold, R. Ankerhold, and G. P. C. Drummen, “Advanced fluorescence microscopy techniques—FRAP, FLIP, FLAP, FRET and FLIM,” *Molecules*, vol. 17, pp. 4047–4132, Apr. 2012.
- [2] L. Tang, L. Zhu, M. Taylor, Y. Wang, S. Remington, and C. Fang, “Excited state structural evolution of a GFP single-site mutant tracked by tunable femtosecond-stimulated raman spectroscopy,” *Molecules*, vol. 23, p. 2226, Sept. 2018.
- [3] J. Sasián, *Image Evaluation*. Cambridge University Press, 2019.
- [4] C. Smith, *Optimum Single Molecule Localization Microscopy*. PhD thesis, 2016.
- [5] F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, F. D. Stefani, J. Elf, and S. W. Hell, “Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes,” *Science*, vol. 355, no. 6325, pp. 606–612, 2017.
- [6] K. C. Gwosch, J. K. Pape, F. Balzarotti, P. Hoess, J. Ellenberg, J. Ries, and S. W. Hell, “Minflux nanoscopy delivers multicolor nanometer 3d-resolution in (living) cells,” *bioRxiv*, 2019.
- [7] J. Cnossen, T. Hinsdale, R. Thorsen, F. Schueder, R. Jungmann, C. S. Smith, B. Rieger, and S. Stallinga, “Localization microscopy at doubled precision with patterned illumination,” *bioRxiv*, 2019.
- [8] A. van Leeuwenhoek and S. Hoole, *The select works of Antony van Leeuwenhoek, containing his microscopical discoveries in many of the works of nature*. Whittingham and Arliss, 1816.
- [9] G. G. Stokes, “On the change of refrangibility of light,” *Philosophical transactions of the Royal Society of London*, no. 142, pp. 463–562, 1852.
- [10] J. W. Lichtman and J.-A. Conchello, “Fluorescence microscopy,” *Nature Methods*, vol. 2, pp. 910–919, Nov. 2005.

- [11] M. F. Paige, E. J. Bjerneld, and W. E. Moerner, “A comparison of through-the-objective total internal reflection microscopy and epifluorescence microscopy for single-molecule fluorescence imaging,” *Single Molecules*, vol. 2, pp. 191–201, Oct. 2001.
- [12] M. Minsky, “Memoir on inventing the confocal scanning microscope,” *Scanning*, vol. 10, no. 4, pp. 128–138, 1988.
- [13] M. Minsky, “Microscopy apparatus.” United States Patent US3013467A, December 19, 1961.
- [14] J. W. Goodman, *Introduction to Fourier Optics*. W.H. Freeman, 2017.
- [15] M. Born and E. Wolf, *Principles of Optics*. Cambridge University Press, 1999.
- [16] S. W. Hell, “Nanoscopy with focused light,” *Annalen der Physik*, vol. 527, pp. 423–445, June 2015.
- [17] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, “Imaging intracellular fluorescent proteins at nanometer resolution,” *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.
- [18] S. W. Hell and J. Wichmann, “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy,” *Optics Letters*, vol. 19, p. 780, June 1994.
- [19] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, “Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission,” *Proceedings of the National Academy of Sciences*, vol. 97, pp. 8206–8210, July 2000.
- [20] E. Betzig, “Proposed method for molecular optical imaging,” *Opt. Lett.*, vol. 20, pp. 237–239, Feb 1995.
- [21] W. E. Moerner and L. Kador, “Optical detection and spectroscopy of single molecules in a solid,” *Physical Review Letters*, vol. 62, pp. 2535–2538, May 1989.
- [22] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. E. Moerner, “On/off blinking and switching behaviour of single molecules of green fluorescent protein,” *Nature*, vol. 388, pp. 355–358, July 1997.
- [23] M. J. Rust, M. Bates, and X. Zhuang, “Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm),” *Nature methods*, vol. 3, no. 10, pp. 793–796, 2006.
- [24] B. Huang, M. Bates, and X. Zhuang, “Super-resolution fluorescence microscopy,” *Annual Review of Biochemistry*, vol. 78, pp. 993–1016, June 2009.
- [25] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Prentice Hall, 1998.
- [26] 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, p. 14064, Aug. 2008.

-
- [27] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, “Fast, single-molecule localization that achieves theoretically minimum uncertainty,” *Nature Methods*, vol. 7, no. 5, pp. 373–375, 2010.
 - [28] R. J. Ober, S. Ram, and E. S. Ward, “Localization accuracy in single-molecule microscopy,” *Biophysical Journal*, vol. 86, pp. 1185–1200, Feb 2004.
 - [29] P. Jouchet, C. Cabriel, N. Bourg, M. Bardou, C. Poüs, E. Fort, and S. Lévêque-Fort, “Nanometric axial localization of single fluorescent molecules with modulated excitation,” Dec. 2019.
 - [30] L. Reymond, J. Ziegler, C. Knapp, F.-C. Wang, T. Huser, V. Ruprecht, and S. Wieser, “Simple: Structured illumination based point localization estimator with enhanced precision,” *Opt. Express*, vol. 27, pp. 24578–24590, Aug 2019.
 - [31] L. Gu, Y. Li, S. Zhang, Y. Xue, W. Li, D. Li, T. Xu, and W. Ji, “Molecular resolution imaging by repetitive optical selective exposure,” *Nature Methods*, 2019.
 - [32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.
 - [33] R. A. Fisher and F. Yates, *Statistical Methods for Research Workers*. Oliver and Boyd, 1954.
 - [34] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, and L. E. Meester, *A Modern Introduction to Probability and Statistics*. Springer, 2005.
 - [35] L. P. Seidman, “Performance limitations and error calculations for parameter estimation,” *Proceedings of the IEEE*, vol. 58, pp. 644–652, May 1970.
 - [36] S. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*. Macmillan, 1979.
 - [37] D. C. Lay, S. R. Lay, and J. J. McDonald, *Linear Algebra and its Applications*. Pearson Education Limited, 2015.
 - [38] H. L. Van Trees, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.
 - [39] W. Melching, “Data-driven psf modeling using b-splines,” Master’s thesis, Delft University of Technology, 2019.
 - [40] R. T. Cox, “Probability, frequency and reasonable expectation,” *American Journal of Physics*, vol. 14, pp. 1–13, Jan. 1946.
 - [41] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. Chapman and Hall/CRC, 2013.
 - [42] R. D. Gill and B. Y. Levit, “Applications of the van trees inequality: A bayesian cramér-rao bound,” *Bernoulli*, vol. 1, p. 59, Mar. 1995.
 - [43] J. Dauwels, “Computing bayesian cramer-rao bounds,” in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.*, IEEE, 2005.

- [44] M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. SHORT COMMUNICATION," *Journal of Microscopy*, vol. 198, pp. 82–87, May 2000.
- [45] F. Ströhl and C. F. Kaminski, "Frontiers in structured illumination microscopy," *Optica*, vol. 3, p. 667, June 2016.
- [46] P. Jouchet, C. Cabriel, N. Bourg, M. Bardou, C. Poüs, E. Fort, and S. Lévêque-Fort, "In depth 3d single molecule localization microscopy with time modulated excitation," *Biophysical Journal*, vol. 118, p. 149a, Feb. 2020.
- [47] S. Bay, C. Herzet, J.-M. Brossier, J.-P. Barbot, and B. Geller, "Analytic and asymptotic analysis of bayesian cramér–rao bound for dynamical phase offset estimation," *IEEE Transactions on Signal Processing*, vol. 56, pp. 61–70, Jan. 2008.
- [48] C. Robert, *Monte Carlo Statistical Methods*. New York, NY: Springer New York, 2004.
- [49] J. M. Hammersley, *Monte Carlo Methods*. Dordrecht: Springer Netherlands, 1964.
- [50] M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination," *Biophysical journal*, vol. 94, pp. 4957–4970, Jun 2008. 18326650[pmid].
- [51] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. E. Moerner, "Optimal point spread function design for 3d imaging," *Phys. Rev. Lett.*, vol. 113, p. 133902, Sep 2014.
- [52] A. O-Hagan, "Bayes-hermite quadrature," *Journal of Statistical Planning and Inference*, vol. 29, pp. 245–260, Nov. 1991.
- [53] S. Liu and K. A. Lidke, "A multiemitter localization comparison of 3d superresolution imaging modalities," *ChemPhysChem*, vol. 15, no. 4, pp. 696–704, 2014.
- [54] K. Moore, *Iterative Learning Control for Deterministic Systems*. London: Springer London, 1993.
- [55] R. S. Sutton and A. G. Barto, *Reinforcement learning : an introduction*. Cambridge, Massachusetts London, England: The MIT Press, 2018.
- [56] J. Jacod and P. Protter, *Probability Essentials*. Springer, 2004.
- [57] G. H. Golub and C. F. Van Loan, *Matrix Computations*. John-Hopkins University Press, 2013.

Glossary

List of Acronyms

CRLB	Cramér-Rao lower bound
DM	dichroic mirror
GLRT	generalized likelihood ratio test
i.i.d.	independent and identically distributed
imeSMLM	iterative modulation enhanced single-molecule localization microscopy
ILC	iterative learning control
meSMLM	modulation enhanced single-molecule localization microscopy
MLE	maximum likelihood estimation
PSF	point spread function
RL	reinforcement learning
RoIs	regions of interest
SBR	signal to background ratio
SMLM	single-molecule localization microscopy
VTI	Van Trees inequality

