THE UNIVERSITY *of* EDINBURGH

Super-resolution methods for fluorescence microscopy

Ondřej Mandula

Doctor of Philosophy

Institute for Adaptive and Neural Computation

School of Informatics

University of Edinburgh

2012

Abstract

Fluorescence microscopy is an important tool for biological research. However, the resolution of a standard fluorescence microscope is limited by diffraction, which makes it difficult to observe small details of a specimen's structure. We have developed two fluorescence microscopy methods that achieve resolution beyond the classical diffraction limit.

The first method represents an extension of localisation microscopy. We used non-negative matrix factorisation (NMF) to model a noisy dataset of highly overlapping fluorophores with intermittent intensities. We can recover images of individual sources from the optimised model, despite their high mutual overlap in the original dataset. This allows us to consider blinking quantum dots as bright and stable fluorophores for localisation microscopy. Moreover, NMF allows recovery of sources each having a unique shape. Such a situation can arise, for example, when the sources are located in different focal planes, and NMF can potentially be used for three dimensional super-resolution imaging. We discuss the practical aspects of applying NMF to real datasets, and show super-resolution images of biological samples labelled with quantum dots. It should be noted that this technique can be performed on any wide-field epifluorescence microscope equipped with a camera, which makes this super-resolution method very accessible to a wide scientific community.

The second optical microscopy method we discuss in this thesis is a member of the growing family of structured illumination techniques. Our main goal is to apply structured illumination to thick fluorescent samples generating a large out-of-focus background. The out-of-focus fluorescence background degrades the illumination pattern, and the reconstructed images suffer from the influence of noise. We present a combination of structured illumination microscopy and line scanning. This technique reduces the out-of-focus fluorescence background, which improves the quality of the illumination pattern and therefore facilitates reconstruction. We present super-resolution, optically sectioned images of a thick fluorescent sample, revealing details of the specimen's inner structure.

In addition, in this thesis we also discuss a theoretical resolution limit for noisy and pixelated data. We correct a previously published expression for the so-called fundamental resolution measure (FREM) and derive FREM for two fluorophores with intermittent intensity. We show that the intensity intermittency of the sources (observed for quantum dots, for example) can increase the "resolution" defined in terms of FREM.

Acknowledgements

I would like to thank to my supervisors Chris Williams and Rainer Heintzmann for their guidance. I would like to give special thanks to Ivana Šumanovac for preparation of 3T3 fibroblast cells and Otto Baumann and Eva Simbürger for providing the Calliphora sample. My thanks go to Ingo Kleppe and Gerhard Krampert for providing us with ZEISS ELYRA-S system with a line-scanning module. I would also like to acknowledge Aurélie Jost, Helen Ramsden, Martin Kielhorn, Jakub Nedbal, Kai Wicker and Hugh Pastoll for help with revision of this manuscript.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Ondřej Mandula*)

Table of Contents

Chapter 1

Introduction

1.1 Optical microscope

A microscope is an instrument allowing us to see objects, which are too small for a naked eye. An optical microscope (often referred to as a "light microscope") uses light in the visible spectral range (wavelength $\approx 400 - 700\,\text{nm}$), which makes it particularly suitable for biological exploration. Visible light is minimally invasive for sensitive biological samples and allows observation of living specimens. Visible light is also minimally absorbed by water, which prevents heating of the sample.

The most common optical microscope is a "far-field" microscope, where the light has to propagate over a distance significantly longer than its wavelength. The specimen is observed with transmitted, reflected or fluorescent light. Fluorescence microscopy is discussed further in Sect. 1.5. The focus of this thesis is on far-field fluorescence optical microscopy.

1.2 Brief historical overview

Optical microscopy has been around for over 400 years. Since the very early versions of Zacharias Janssen's or Galileo's compound microscopes from the beginning of 17^{th} century, optical microscopy has undergone a long and steady process of development.

Despite the speculation as to who was the actual inventor of the optical microscope, it was Anton van Leeuwenhoek who largely popularised the use of the microscope as an instrument for observing the minute details of the specimen. Leeuwenhoek also introduced his simple instrument into biological research during the 17th century.

An important milestone was the pioneering work of Ernst Abbe [1] in the second half of the 19th century. Abbe set the theoretical resolution limit for the optical microscope and mastered the design of objective lenses highly corrected for optical aberrations.

With advances in technology in the 20th century, the manufactures have produced lenses reaching the theoretical limits of the optical microscope performance. The 1953 Nobel prize in physics was awarded to Frits Zernike for discovery of the phase contrast [2]. This method allows observation of transparent specimens, and had major impact on biological research such as in vivo study of cell cycle.

The emergence of new microscopy methods surpassing the classical resolution limit (super-resolution microscopy) at the end of the 20th century and at the beginning of the 21st century has given another boost to optical microscopy research. The resolution of the super-resolution optical microscopes has reached the order of ten nanometres and some researchers have proposed the term "optical nanoscopy" to be used [3, 4, 5]. However, super resolution micro/nano-scopy remains a challenging task, especially when applied to living biological specimens. While most of the super-resolution techniques require a highly specialised and expensive hardware, some of the techniques, such as localisation microscopy (discussed in Chapter 2) can be performed with a conventional fluorescent microscope.

The number of scientific publications in recent years shows that even after four centuries of development the optical microscopy remains a vibrant and exciting scientific domain.

1.3 Point spread function

An important characteristic of a microscope is the so-called "point spread function" (PSF). The PSF represents an image of a point source. The image $i(x)$ of a specimen produced by an optical microscope can be described as a convolution between the object (specimen) $o(x)$ and the point spread function $q(x)$:

$$i(x) = \int q(x-x')o(x')dx'. \tag{1.1}$$

The PSF therefore defines how much the image of the specimen is "blurred" during the imaging process. The integration in Eq. (1.1) is over the whole space of acquired data (typically 2D or 3D). Note, that Eq. (1.1) applies to the situation with spatially invariant PSF. It also assumes that PSF is fully determined by the optical system. The influence of the specimen on the shape of the PSF is neglected. In a real experiment, PSF can be locally distorted by the aberrations introduced by the specimen or by the imperfections in the optical setup.

Neglecting the effect of polarisation (scalar theory), the two-dimensional PSF of an optical microscope, known as the "Airy pattern", is described by [6]

$$q(\rho) = \frac{1}{Z}\left(\frac{J_1(\alpha\rho)}{\alpha\rho}\right), \tag{1.2}$$

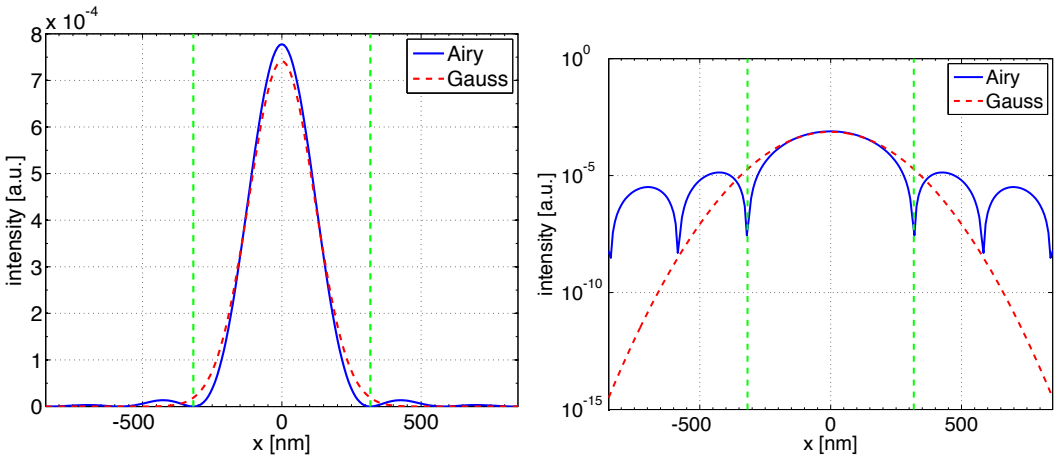
where $Z = \int q(\rho)d\rho$ is the normalising constant, J_1 is the Bessel function of the first kind of order one and ρ is the distance from the centre of the image. The parameter α depends on the emission wavelength λ_{em} and the numerical aperture of the objective NA:

$$\alpha = 2\pi\frac{\mathrm{NA}}{\lambda_{em}}. \tag{1.3}$$

The NA is defined as

$$\mathrm{NA} = n\sin(\theta), \tag{1.4}$$

where n is the refractive index of the immersion medium and 2θ is the angle of the light cone entering the objective. Note that the refractive index is a function of wavelength



(a) Linear **(b)** Logarithmic

Figure 1.1: Intensity profile of a PSF for 1.2 NA objective using $\lambda_{em} = 625$ nm emission light. The blue line represents profile of the Airy pattern, red dashed line shows the Gaussian approximation. Green vertical lines mark the first minima of the Airy pattern at $\delta = 318$ nm. δ corresponds to the radius of the Airy disk. (a) Linear, (b) logarithmic plot of the intensity highlighting the secondary maxima in the Airy pattern and the differences of the Gaussian approximation at the periphery of the function.

$n \sim n(\lambda)$. Dispersion in the sample and in the optics can lead to the PSF corrupted with spherical aberration when a range of wavelengths is used [7].

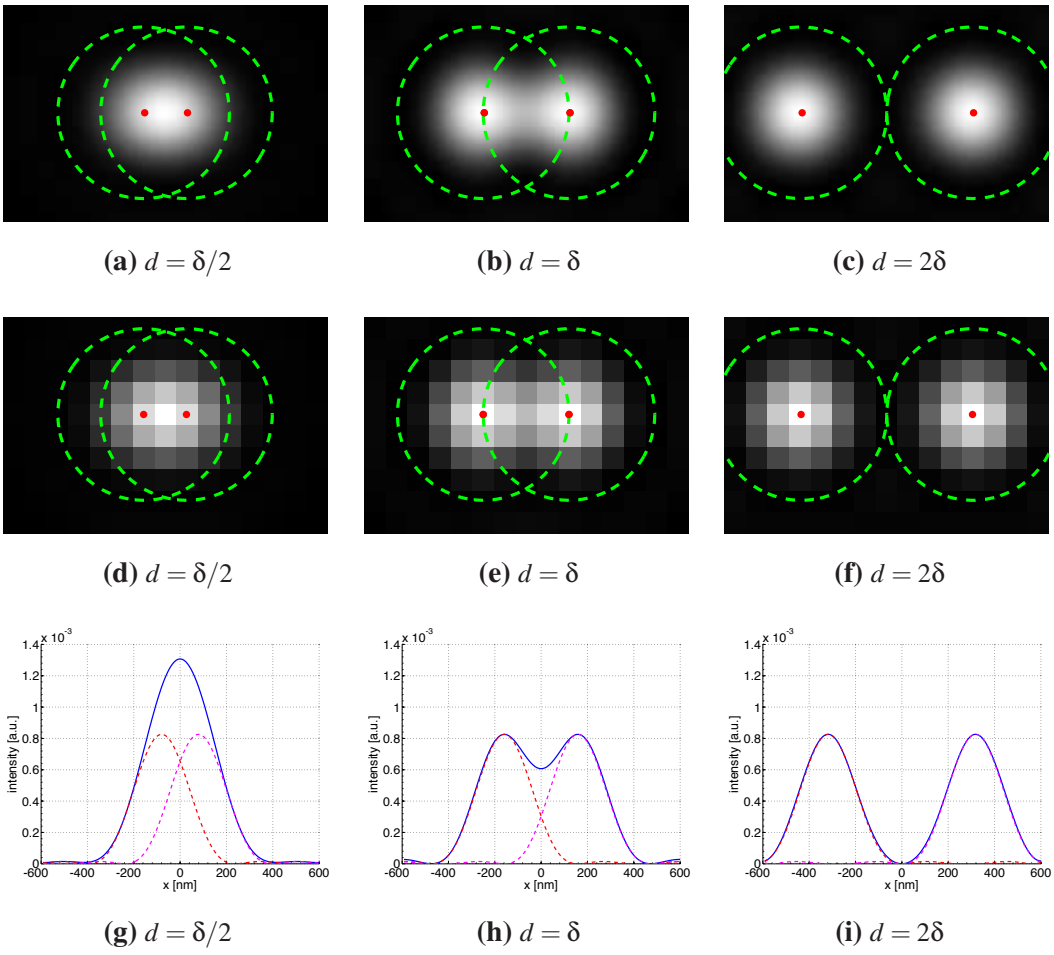
An intensity profile of an unaberrated PSF is shown in Fig. 1.1. The Airy pattern is also compared to the popular approximation of the PSF with a Gaussian function [8] defined by a standard deviation

$$\sigma = \frac{\sqrt{2}}{2\pi}\frac{\lambda_{em}}{\mathrm{NA}}. \tag{1.5}$$

1.4 Resolution limit

The ability of an optical microscope to show spatial details in the specimen structure is fundamentally limited by diffraction [6]. The radius δ of the Airy disk (see Fig. 1.1) is often considered as the "classical resolution limit". It is given by [6]

$$\delta = 0.61\frac{\lambda_{em}}{\mathrm{NA}}, \tag{1.6}$$



(a) $d = \delta/2$ **(b)** $d = \delta$ **(c)** $d = 2\delta$

(d) $d = \delta/2$ **(e)** $d = \delta$ **(f)** $d = 2\delta$

(g) $d = \delta/2$ **(h)** $d = \delta$ **(i)** $d = 2\delta$

Figure 1.2: Two PSFs (1.2 NA objective, $\lambda = 625$ nm) separated by distance d. Location of the sources is indicated with red dots. Airy disk is indicated with green dashed circles. The radius of the Airy disk corresponds to $\delta = 318$ nm. (a) Continuous representation of the PSFs. (b) Pixelated version with pixel-size 80 nm. (c) Blue lines show the intensity profiles along the line intersecting the sources' locations. The profiles of the individual PSFs are shown as red and magenta dashed lines.

where λ_{em} is the wavelength of the emitted light and NA is the numerical aperture of the objective Eq. (1.4). The resolution limit Eq. (1.6) comes from the *empirical* observation: two sources separated by a distance greater than δ can be "resolved" as two individual objects (see Fig. 1.2c). If the separation is smaller than δ, the point spread functions overlap significantly and the sources become "unresolved" (see Fig. 1.2a). The resolution limit defined by Eq. (1.6), sometimes called the "Raleigh resolution

limit", is often taken as the benchmark for different resolution techniques. Note that the "Abbe resolution limit" $\delta = 0.5\lambda_{em}/\mathrm{NA}$ related to the passband of spatial frequencies is also used.

The resolution limit Eq. (1.6) relates to the noise-free situation with continuous representation of the PSF (the top line of Fig. 1.2). However, the pixelation and noise associated with the photon-detection process can deteriorate the resolution significantly. The influence of noise, pixelation or brightness intermittency to the resolution is discussed in Chapter 3.

1.5 Fluorescence microscopy

The introduction of fluorescence microscopy in the 20$^{\text{th}}$ century has revolutionised the use of optical microscopy in biological and medical science. Fluorescence is generated by fluorescent molecules or nanostructures (commonly called fluorophores) during the relaxation of their electronic structure to the ground state. The fluorophores are driven into higher energetic state with an excitation light of a specific wavelength. The fluorescent light is typically shifted towards the longer wavelengths (Stokes shift), which allows an efficient filtering of the excitation light from the fluorescence signal. Fluorescence microscopy detects the fluorescence from the sample itself (auto-fluorescence) or more commonly from the fluorophores attached to the specimen [9].

Fluorescence labels can be attached to the structures of interest with high specificity and provide a strong intensity contrast in the microscopic image of the specimen. Fluorescent proteins (FPs) allow for direct expression of the fluorescent marker by the organism itself [10]. FPs have further redefined the use of fluorescence microscopy in cell biology as a nearly non-invasive and highly specific technique.

The first fluorescent protein used in microscopy (green fluorescent protein - GFP) has been isolated from the jellyfish *Aequorea victoria* [11]. Nowadays, modified versions of GFP and a rich variety of fluorescent dyes cover the emission across the whole visi-

ble spectral range [10]. Photo-activable and photo-switchable fluorescent proteins and dyes have also been discovered, and are used in specific applications and microscopy methods, see e.g. [12, 13, 14]. The 2008 Nobel Prize in Chemistry was awarded to Martin Chalfie, Osamu Shimomura, and Roger Y. Tsien for their discovery and development of the GFP.

1.6 Quantum Dots

Quantum dots (QDs), recently used in biological research, are promising fluorescent labels. QDs are inorganic crystals composed of 100-100,000 atoms of substances such as cadmium selenide (CdSe), with diameter $\sim 2-10\,\text{nm}$ [15]. QDs are often coated with a zinc sulphide (ZnS) shell conjugated with an antibody molecule. The diameter of the coated QDs is $\sim 10-30\,\text{nm}$. The emission wavelength can be tuned by the diameter of the QD core. QDs tend to have a broad excitation spectrum and a narrow emission spectrum when compared to the standard fluorescent dyes/proteins. This facilitates the multicolour imaging with QDs. QDs are also exceptionally photo-stable. It has been reported that QDs illuminated with continuous $\sim 50\,\text{mW}$ laser do not bleach even after 14 h, whereas the standard fluorescent dye "fluorescein" completely bleaches in less than 20 mins [16]. Moreover, QDs are an order of magnitude brighter than the traditional fluorophores [17, 18].

Commercially available QDs have a polymer coating with a covalently attached linker, or are conjugated with an antibody molecule. This allows them to be specifically attached to the structure of interest in the specimen the same way as the standard fluorescent labels. However, coated QDs are relatively big ($\sim 10-30\,\text{nm}$) compared to the fluorescent dyes (fluorescein size is $\sim 1\,\text{nm}$) or fluorescent proteins (GFP size is $\sim 5\,\text{nm}$). Therefore QDs cannot diffuse through the cell membrane, which complicates the labelling of the structures in the interior of the cell.

An interesting property of the QDs is that they exhibit "fluorescence blinking" (fluo-

rescence intermittency) under continuous excitation. QDs switch between the ON episodes of a rapid absorption-fluorescence cycling and the OFF episodes, where no light is emitted despite the continuous excitation. Both ON-time (τ_{ON}) and OFF-time (τ_{OFF}) probability densities follow an inverse power law $P(\tau_{ON/OFF}) \propto 1/\tau^{m}_{ON/OFF}$ [19, 20]. A comparison of QDs and the standard fluorescent dyes is described in the Resch-Genger et al. review article [17].

1.7 Super-resolution

The resolution limit Eq. (1.6) for a far-field optical microscope has been challenged in the past two decades. Several research teams have reported sub-diffraction resolution in the fluorescent samples. In fact, the resolution limit is no longer dictated by diffraction but by the signal-to-noise ratio.

There is a rich variety of super-resolution strategies in fluorescence microscopy. However, all these methods are based on driving the nearby fluorophores into emitting (ON) and non-emitting (OFF) states. This allows distinguishing the individual emitters separated by a sub-resolution distance.

"Selective" activation can be achieved either by spatially structured excitation (structured illumination) or by stochastic activation of the individual fluorophores. In the stochastic activation approach, a small random subset of fluorophores is activated for each acquisition frame. For the conventional methods, the activated subset must be sufficiently small to ensure that the majority of the activated emitters are separated by distances larger than the diffraction limit. In this case each acquired frame consists of several well-separated (non-overlapping) PSFs. The individual fluorophores can be localised by, for example, fitting each PSF with a Gaussian function. Given enough detected photons, the localisation precision can be significantly higher than the resolution limit. The activation-acquisition cycle is typically repeated for several thousands acquisition frames. Super-resolution fluorescent images are produced by visualisation

all the estimated fluorophores locations.

Methods based on this simple strategy are called by a general term "localisation microscopy" (LM). Various names for LM have been proposed: "Photo-Activation Localisation Microscopy" (PALM) [21], "fluorescence PALM" (fPALM) [22] or "Stochastic Optical Reconstruction Microscopy" (STORM) [23].

"Super-resolution Optical Fluctuation Imaging" (SOFI) [24] is a LM related technique. SOFI is based on higher-order statistical analysis of temporal intensity fluctuations caused by blinking behaviour of the fluorophores.

The structured illumination based methods modulate the fluorescence behaviour of the molecules within the diffraction-limited area. The nearby molecules are driven to either ON or OFF states, which facilitates their discrimination. These methods include "Saturated Structured Illumination microscopy" (SSIM) [25, 26] and "STimulated Emission Depletion microscopy" (STED) [27].

1.8 Overview of the thesis

In Chapter 2 we discuss a new approach to localisation microscopy. Application of the machine learning technique non-negative matrix factorisation (NMF) enables us to computationally separate images of individual blinking fluorophores despite their high mutual overlap in the recorded data. We show that with this approach we can use quantum dots (QDs) as extremely bright and stable fluorescent labels for super-resolution microscopy.

Chapter 3 discusses the resolution criterion for noisy and pixelated data in terms of so-called fundamental resolution measure (FREM). We show that intermittency of the sources' brightness (blinking) can be beneficial and provide higher resolution when compared to sources with static intensity.

In Chapter 4 we present structured illumination microscopy (SIM) combined with line scanning (LS). Our goal is to introduce the SIM into more realistic and biologically

relevant settings. Line scanning reduces the out-of-focus fluorescence background, which improves the quality of the illumination pattern. The method enables resolution improvement in relatively thick and densely labelled fluorescent samples. The reconstructed images reveal high details of the specimen's inner structure, and suffer less from the artefacts when compared to the conventional SIM methods.

Chapter 5 gives a short overview of possible extensions of the current work and contains the final summary of the thesis.

Chapter 2

Non-Negative Matrix Factorisation for Localisation Microscopy

We propose non-negative matrix factorisation (NMF) as a natural model for localisation microscopy of samples labelled with quantum dots (QDs) or other intermittent fluorophores. NMF can separate the individual highly overlapping sources with individual different shapes. We use the Lee-Seung NMF algorithm [28], which accounts for Poisson noise in the recorded images. This allows us to recover the individual intermittent sources from the noisy recordings. The separated sources can be localised with uncertainty smaller than the diffraction limit and provide information about sub-resolution details of the sample structure. We show that for flat (2D) samples the localised sources can be further processed and used to reconstruct the super-resolution image of the specimen.

The chapter is divided into the following sections: Section 2.1 introduces the localisation microscopy (LM) technique and discusses the advantages and challenges of using quantum dots as fluorescent labels. It also contains a short review of recent methods dealing with LM data containing overlapping sources.

Section 2.2 introduces non-negative matrix factorisation (NMF). We discuss several algorithms applying particular constraints during the optimisation. We also show a

generative probabilistic model for NMF.

Section 2.3 shows NMF as a natural model for intermittent overlapping QDs and Sect. 2.4 discusses the alternative methods used for treating QD data. A link to a standard deconvolution technique is also mentioned.

We used synthetic data for analysing the performance of the algorithm in different experimental settings. The main simulations used in this chapter are described in Sect. 2.5. Section 2.6 explains the evaluation techniques for the comparison of the results. We used averaged precision (AP) as a quantitative measure of the performance of different algorithms using simulated data. AP summarises both precision and recall curves.

Application of the NMF algorithm to real microscopic data is explored in Sect. 2.7. This includes the comparison of the different techniques for estimation of the number of sources in the dataset. We also present results on how specific problems with the use of NMF on real data were overcome. We also propose different methods for visualisation of the results.

Finally, the NMF reconstructed images of synthetic and real data are presented in Sect. 2.8. In this section we also compare NMF with two other techniques (CSSTORM, 3B analysis) dealing with a similar problem.

2.1 Localisation microscopy

Localisation microscopy (LM), is a conceptually simple and accessible technique for super-resolution imaging of fluorescent samples. LM takes as input a stack of images ($10^2 - 10^4$ frames) containing an unknown number of fluorescent sources (fluorophores) with time-varying intensity and identifies the locations of these sources. If the sources are attached to structures of interest (e.g. in biological samples), then this provides useful information about the target structures. Provided enough photons are collected, the localisation of an individual source can be an order of magnitude be-

low the classical diffraction limit Eq. (1.6) [29]. The super-resolution image is usually produced by visualisation of all estimated positions of the individual fluorophores ("pointillism" [30]).

LM techniques are based on fluorescent sources with a transition between bright (ON) and dark (OFF) intensity states. Fluorescent proteins or organic dyes are used as fluorophores in the standard techniques (fPALM [22], STORM [23]). In this case the density of the ON sources in each captured frame is controlled by photo-switching and must be optimised experimentally. High density of the ON fluorophores results in overlapping sources and complicates localisation (overlapping sources are usually discarded), whereas low density leads to a long total acquisition time [31]. Several thousands frames are typically required for an image reconstruction. The total acquisition time can exceed ten minutes (17 mins for 10^4 frames with 100 ms acquisition time per frame).

2.1.1 Quantum dots for localisation microscopy

There has been interest in using quantum dots (QDs) as sources for localisation microscopy in recent years [30, 32, 33, 34]. QDs are an order of magnitude brighter and more photo-stable compared to the organic dyes or fluorescent proteins used in conventional LM [17]. Under continuous excitation QDs exhibit stochastic blinking between ON and OFF states [19, 20]. Excellent photo-stability, low cyto-toxicity and distinctive spectral properties make QDs very attractive for biological research. However, the stochastic blinking of QDs is impractical for standard LM techniques because the rate of switching, and hence the density of ON sources, is difficult to control. Thus QD-labelled data typically consist of highly overlapping sources, which cannot be localised with standard techniques.

2.1.2 Overlapping sources

Several techniques dealing with overlapping sources have been proposed recently [30, 33, 35, 36, 37, 38]. Most of these methods model the LM data using a known image of a single source, the so called point spread function (PSF), see Sect. 1.3. Often a single point-spread function (PSF) is assumed to be shared by all sources in the dataset [33, 35, 36, 37, 38].

There are two main groups of the algorithms addressing the overlapping sources in the LM data. The first group operates separately on each frame of the LM dataset: a method proposed by Huang et al. [35] tries to fit multiple PSFs into each frame of the dataset. The DAOSTORM algorithm [36] applies iterative fitting and subtracting procedure in each frame. CSSTORM [38] makes use of compressed sampling to recover the sparse vector representing the distribution of the fluorophores' locations. CSSTORM is supposed to deal with higher densities than DAOSTORM (see supplementary materials to [38]).

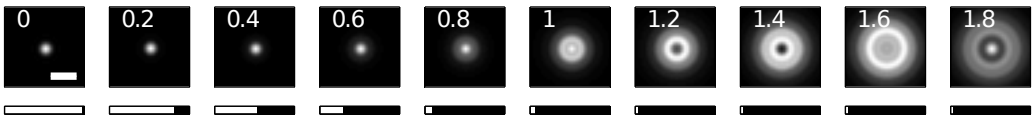
These methods ignore the fact that some sources can stay ON for several successive frames or can even reappear in different frames due to blinking because each frame of the dataset is treated independently. Therefore they can generally deal with only moderately overlapping sources with densities $< 10\,\text{sources}/\mu\text{m}^2$ [35, 36, 38].

The second group of the algorithms models LM dataset as a collection of blinking sources. They can improve the localisation for higher densities of the overlapping sources by taking the reappearance of fluorophores into account. However, these algorithms are, in general, computationally more expensive.

Modelling the whole dataset from a known PSF with maximum posterior (MAP) fitting has been proposed by Harrington et al. [33]. Separation of several (up to five) simulated emitters contained in a disk of 100 nm radius has been shown. However, the technique becomes computationally very challenging for higher numbers of sources. Bayesian analysis of intermittent sources (Bayesian Blinking and Bleaching (3B) analysis) has been suggested by Cox et al. [37]. The blinking behaviour of the fluoro-

phores is modelled as a hidden Markov model with three distinct states: emitting, not emitting and bleached. Each source is described by its position, size of the PSF, and intensity. MAP estimates of the positions obtained from different sampling of the state sequences are used as estimated locations of the fluorophores. While the 3B analysis adjusts the width of the PSF (Gaussian approximation of the PSF [8]), it cannot deal with individually different shapes of the sources. This situation can arise, for example, in three-dimensional samples, where the overlapping sources can be located in different focal planes (see Fig. 2.1). Moreover, 3B assumes a mono-exponential decay of

(a) Intensity scaled to the in-focus PSF



(b) Intensity scaled for each frame

Figure 2.1: Simulated PSF in different depths of focus. The number in each figure indicates the distance in μm from the in-focus plane. (a) Intensity scaled to the in-focus PSF. (b) Intensity scaled in each frame. The maximum intensity relative to the in-focus PSF is indicated in the bars below and corresponds to about 10% at 1 μm and 3% at 1.5 μm. Scale bar 1 μm.

the fluorescence for the individual sources. QDs have a complex blinking behaviour with power-law distribution of the ON and OFF times [39]. This can possibly complicate the 3B analysis of the QD data. Independent component analysis (ICA) has been proposed for analysis of overlapping intermittent sources in [30, 32]. However, as we demonstrate in Sect. 2.4.3, ICA is not a suitable model for noisy QD data.

Yet another approach to the LM data with overlapping sources has been proposed in a method called SOFI (Super-resolution Optical Fluctuation Imaging) [40]. Instead of separating the individual emitters, SOFI analyses higher order statistics of the intensity fluctuation. The intensity values in the SOFI image, however, reflect the fluctuation

behaviour, rather than the strength of the emitters. The non-linear relation between the sources' strength and the intensity in the reconstructed image leads to structural artefacts such as apparent discontinuities, cavities and holes in otherwise continuous structures. Sources, which do not blink will not appear in the SOFI image at all. Recently, this issue has been, to a certain extent, addressed by balanced SOFI (bSOFI) in [41].

In this chapter we propose non-negative matrix factorisation (NMF) as a model for overlapping sources. NMF models the whole dataset, taking into account the reappearance of the sources during the acquisition (fluorescence blinking). The intensity in the reconstructed image can be related to the strength of the individual emitters. NMF can deal with highly overlapping sources with unknown different shapes and variety of blinking behaviour. Moreover our algorithm accounts for Poisson noise in the recorded data.

2.2 Non-negative matrix factorisation

Non-negative matrix factorisation (NMF) solves the approximative factorisation of an $N \times T$ data matrix $\boldsymbol{D}$ with non-negative entries:

$$\boldsymbol{D} \approx \boldsymbol{W}\boldsymbol{H}, \tag{2.1}$$

where $\boldsymbol{W}$ and $\boldsymbol{H}$ are $N \times K$ and $K \times T$ matrices, respectively. More explicitly

$$\begin{pmatrix} d_{11} & d_{12} & \cdots & \cdots & d_{1T} \\ d_{21} & d_{22} & \cdots & \cdots & d_{2T} \\ \vdots & & & & \vdots \\ d_{N1} & \cdots & \cdots & \cdots & d_{NT} \end{pmatrix} \approx \begin{pmatrix} w_{11} & \cdots & w_{1K} \\ w_{21} & \cdots & w_{2K} \\ \vdots & & \vdots \\ w_{N1} & \cdots & w_{NK} \end{pmatrix} \cdot \begin{pmatrix} h_{11} & h_{12} & \cdots & \cdots & h_{1T} \\ \vdots & & & & \vdots \\ h_{K1} & h_{K2} & \cdots & \cdots & h_{KT} \end{pmatrix}. \tag{2.2}$$

Usually $K < N, T$. The factorisation is constrained to $\boldsymbol{W}$ and $\boldsymbol{H}$ with non-negative entries.

Initial factorisation algorithms (so called positive matrix factorisation) [42] were published in 1994. However, it was in 1999 when NMF attracted attention of researchers after publication of the *Nature* article by Daniel Lee and Sebastian Seung [43]. NMF was presented as an efficient and powerful method for approximation of non-negative data (in their case a database of facial images) by linear combination of non-negative localised basis vectors (images of the nose, mouth, ears, eyes, etc.) An individual face from the dataset can be recovered as a non-subtractive composition of individual basis vectors.

Lee and Seung also proposed simple multiplicative updates [28] for the elements of $\boldsymbol{W}$ and $\boldsymbol{H}$

$$\begin{aligned} w_{xk} &= \frac{w_{xk}}{\sum_{t=1}^{T} h_{kt}} \left[(\boldsymbol{D} \oslash \boldsymbol{W}\boldsymbol{H}) \boldsymbol{H}^{\top} \right]_{xk} \\ h_{kt} &= \frac{h_{kt}}{\sum_{x=1}^{N} w_{xk}} \left[\boldsymbol{W}^{\top} (\boldsymbol{D} \oslash \boldsymbol{W}\boldsymbol{H}) \right]_{kt} . \end{aligned} \tag{2.3}$$

minimising the (generalised) Kullback – Leibler (KL) divergence (see Eq. (2.7)) between data matrix $\boldsymbol{D}$ and its factorised approximation $\boldsymbol{W}\boldsymbol{H}$ (see Appendix A for details). The symbol "$\oslash$" denotes the element-wise division of matrices.

Equation (2.3) suggests that the complexity of the updates is $O(NKT)$, or more precisely $O(N(2KT+T+K))$. Note that updates Eq. (2.3) automatically ensure that $\boldsymbol{W}$ and $\boldsymbol{H}$ remains non-negative if initialised so. Also once they become zero they remain zero for the rest of iterations. Sufficient conditions for uniqueness of solutions to the NMF problem have been studied in [44].

Various alternative minimisation strategies have been explored in an effort to speed up the convergence properties of the Lee & Seung updates. A comprehensive discussion on the variety of these algorithms can be found in [45].

2.2.1 Additional constraints to the NMF model

Additional constraints can be imposed on $\boldsymbol{W}$ and $\boldsymbol{H}$ matrices. Imposing a defined "sparsity" on either columns of $\boldsymbol{W}$ or rows of $\boldsymbol{H}$ has been proposed in [46] and is dis-

cussed in Sect. 2.4.2. Enforcing the temporal smoothness of $\boldsymbol{H}$ in the analysis of EEG recordings has been published in [47]. Multiplicative updates for various constraints have been suggested in [47, 48] (see Appendix A).

2.2.2 Gamma - Poisson model

A generative model for NMF Eq. (2.4) is represented by the gamma-Poisson (GaP) model. This model has been proposed by John Canny [49] as a probabilistic model for documents. The entries h_{kt} of the intensity matrix $\boldsymbol{H}$ in Eq. (2.5) are regarded as latent variables generated from a Gamma distribution with parameters α_k, β_k and the data are modelled as a Poisson variable with mean $\boldsymbol{WH}$. Variables $\theta = \{\boldsymbol{w}_k, \alpha_k, \beta_k\}; k = 1..K$ are then parameters of the GaP model.

2.3 NMF as a natural model for QD data

$$\begin{bmatrix} i_{11} & i_{12} & \cdots & i_{1n_x} \\ i_{21} & i_{22} & \cdots & i_{2n_x} \\ \vdots & & & \vdots \\ i_{n_y,1} & \cdots & \cdots & i_{n_y,n_x} \end{bmatrix}_t \longrightarrow \begin{bmatrix} d_{1t} \\ d_{2t} \\ \vdots \\ \vdots \\ d_{Nt} \end{bmatrix}$$

Figure 2.2: Reshaping of one $n_x \times n_y$ frame I_t of the movie $\{I_t\}_{t=1}^{T}$ into a $N \times 1$ ($N = n_x n_y$) column $\boldsymbol{d}_t$ of the matrix $\boldsymbol{D}$ by concatenating the columns of the image.

Non-negative matrix factorisation (NMF) [43, 28] is a natural model for QD data. NMF decomposes a movie of blinking QDs into spatial and temporal parts, i.e., time independent emission profiles of the individual sources (PSFs) and fluctuating intensities of each source, respectively. NMF imposes non-negativity constraints on both the spatial and the temporal components, which are natural constraints for the source

profiles and intensities of blinking QDs.

Consider a $N \times T$ data matrix $\boldsymbol{D}$, where N is the number of pixels in each frame, and T is the number of time frames. The columns of $\boldsymbol{D}$ are the individual frames of the movie reshaped into $N \times 1$ vector by concatenating the columns of the image (see Fig. 2.2). All entries in $\boldsymbol{D}$ are non-negative, i.e., $d_{xt} \geq 0$. Under the NMF model Eq. (2.2), $N \times T$ matrix $\boldsymbol{D}$ is factorised into a $N \times K$ spatial component matrix $\boldsymbol{W}$ (images of the K individual sources) and the $K \times T$ temporal component matrix $\boldsymbol{H}$ (the intensities of the sources). In fact, we relax the demand for exact factorisation by factorisation of the noisy dataset expectation value:

$$\mathbb{E}[\boldsymbol{D}] = \boldsymbol{W}\boldsymbol{H};\ w_{xk}, h_{kt} \geq 0 \tag{2.4}$$

or in element-wise form

$$\mathbb{E}[d_{xt}] = \sum_{k=1}^{K} w_{xk} h_{kt};\ w_{xk}, h_{kt} \geq 0 \tag{2.5}$$

The predominant noise model in microscopy imaging is Poisson noise [9]. Therefore the log-likelihood function can be expressed as

$$\log p(\boldsymbol{D}|\boldsymbol{W},\boldsymbol{H}) = \sum_{xt} \left(d_{xt} \log \sum_{k=1}^{K} w_{xk} h_{kt} - \sum_{k=1}^{K} w_{xk} h_{kt} \right) + C_1, \tag{2.6}$$

where C_1 is independent of $\boldsymbol{W}$and $\boldsymbol{H}$.

The Lee and Seung NMF updates Eq. (2.3) minimise the KL divergence between the data and the NMF model

$$\mathrm{KL}(\boldsymbol{D} \parallel \boldsymbol{W}\boldsymbol{H}) = -\sum_{xt} \left(d_{xt} \log \sum_{k=1}^{K} w_{xk} h_{kt} - \sum_{k=1}^{K} w_{xk} h_{kt} \right) + C_2 \tag{2.7}$$

where C_2 is independent of $\boldsymbol{W}$and $\boldsymbol{H}$. Comparison with the log-likelihood Eq. (2.6) shows that the minimum of the divergence with positivity constrains on $\boldsymbol{W}$ and $\boldsymbol{H}$ is equivalent to the maximum of the log-likelihood. Note, that there is an alternative version of the algorithm minimising the least squares $|\boldsymbol{D} - \boldsymbol{W}\boldsymbol{H}|^2 = \sum_{x,t} \left[d_{xt} - \sum_{k=1}^{K} w_{xk} h_{kt} \right]^2$ proposed by Lee and Seung [28]. This objective function corresponds to the minimising the likelihood of the model with respect to data corrupted with Gaussian noise. The

original Paatero and Tapper "positive matrix factorisation" [42] used the least square approach.

There is a scaling indeterminacy between $\boldsymbol{W}$ and $\boldsymbol{H}$ in the NMF model. We fix this by setting the L_1 norm of each column of $\boldsymbol{W}$ to 1. The background fluorescence in the images is modelled as a "flat" component $\boldsymbol{w}_K = \mathbf{1}/N$ with corresponding intensity $\boldsymbol{h}_K$. The spatial part of the background component $\boldsymbol{w}_K$ is not updated during the optimisation, while the temporal part $\boldsymbol{h}_K$ is updated to account for changes in background levels during the data acquisition (due to bleaching or fluctuation of the excitation light, for example).

The NMF model is fitted to data iteratively using multiplicative updates Eq. (2.3) sequentially:

$$\boldsymbol{W}^{n+1} \longleftarrow (\boldsymbol{W}^n, \boldsymbol{H}^n, \boldsymbol{D}) \tag{2.8}$$

$$\boldsymbol{H}^{n+1} \longleftarrow \left(\boldsymbol{W}^{n+1}, \boldsymbol{H}^n, \boldsymbol{D}\right), \tag{2.9}$$

where n denotes the iteration of the update.

Note that the divergence Eq. (2.7) is convex with respect to $\boldsymbol{H}$ and $\boldsymbol{W}$ individually, but not in both variables together [28], leading to local optima.

2.4 Related work

This section points to published work using NMF as a versatile tool for analysis of biological data in Sect. 2.4.1. It also points to work relevant to the NMF application to QD data. An algorithm for NMF with sparsity constraints is reviewed and demonstrated on simulated data in Sect. 2.4.2. Section 2.4.3 discusses the proposed independent component analysis (ICA) as a model for QD data and Sect. 2.4.4 shows a link between NMF and the Richardson – Lucy deconvolution.

2.4.1 NMF in biological research

NMF has been recently used for analysis of various biological data. A review of NMF as a data analytical and interpretive tool in computational biology can be found in [50]. The most common application of NMF in computational biology has been in the area of molecular pattern discovery, especially for gene and protein expression microarray studies. NMF is a well-suited method for this problem as there is no a priori information about the expected expression pattern for a given set of genes. The $N \times T$ data matrix $\boldsymbol{D}$ in the NMF model Eq. (2.1) is created from the observation of the expression levels of N genes (typically thousands) in T different samples (typically less than 100). The columns of $N \times K$ matrix $\boldsymbol{W}$ are the linear combination of individual genes and are often called "metagenes". The columns of the $K \times T$ matrix $\boldsymbol{H}$ represent the metagene expression pattern in the corresponding sample.

NMF has proved to be a successful method in the recovery of biologically meaningful classes in gene-expression data. For example, Kim and Tidor [51] applied NMF as a tool to cluster genes and predict functional cellular relationships in yeast, while Heger and Holm [52] used it for the recognition of sequence patterns among related proteins. Brunet et al. [53] applied NMF to cancer microarray data for the recovery of tumour subtypes. They also developed a method for selecting the number of metagenes K: each sample (total T samples) is placed into a cluster corresponding to the most highly expressed metagene in the sample. E.g. sample j is placed in cluster i if the h_{ij} is the largest entry in column j of the matrix $\boldsymbol{H}$. They define a $T \times T$ connectivity matrix $\boldsymbol{C}$ with entry $c_{ij} = 1$ if samples i and j belong to the same cluster and $c_{ij} = 0$ if they belong to different clusters. The consensus matrix $\bar{\boldsymbol{C}}$ is computed as an average over $\boldsymbol{C}$ resulting from many (~ 100) runs of NMF. The entries of $\bar{\boldsymbol{C}}$ represent the probability that samples i and j will cluster together. If the clustering is stable then the values of $\bar{\boldsymbol{C}}$ will be close to either 0 or 1. They proposed a measure of stability based on a cophonetic correlation coefficient ρ_K which indicates the dispersion of the consensus matrix for a given number of components K. In a perfect consensus matrix, where all

entries are either 0 or 1 the cophonetic correlation coefficient $\rho_K = 1$. If the entries of $\bar{\boldsymbol{C}}$ are scattered between 0 and 1 the value is $\rho_K < 1$. They compute ρ_K for a range of Ks and select the one where ρ_K begins to fall. However, inspecting the provided figures in [53] we can observe that ρ_K drops rather gradually and unambiguous estimation of K is difficult. Also note that for each value of K the NMF has to be evaluated ~ 100 times to estimate the consensus matrix $\bar{\boldsymbol{C}}$. This is feasible only for small K (< 5), which is typically case in the gene expression data. However, this method of K estimation is prohibitive for our application of NMF to QD data, where the typical Ks are in the range of $10 - 100$.

Rabinovich et al. [54] has proposed NMF for accurate automated spectral decomposition of histologically stained tissue sections. The images (N pixels) stained with *known* number of colours K are recorded in T spectral channels creating the $N \times T$ data matrix $\boldsymbol{D}$. NMF decomposes this matrix into $\boldsymbol{W}$, with entries w_{ij} indicating the contribution of the jth the dye into the ith pixel and the rows of the matrix $\boldsymbol{H}$ represent the spectra of each dye. They have found similar performance of NMF and independent component analysis (ICA) by comparing the results to the ground truth data. Note, that the selection of K is dictated by the number of colours used for histological staining and therefore no algorithm for K estimation is needed. Similarly the NMF has been used for fluorescence microcopy data by Neher et al. [55]. In this case, the fluorescence images were recorded under excitation with different wavelength of the excitation light. The spatial distribution and excitation spectral properties of different fluorescent labels (K is known) were estimated from the factorised matrices $\boldsymbol{W}$ and $\boldsymbol{H}$, respectively.

NMF has also been used in variety of disciplines directly related to optical microscopy. Pande et al. [56] use NMF for analysis of multispectral time-domain fluorescence life time imaging (FLIM) data. The spatio-temporal/spectral data were obtained by measuring the fluorescent decay in different positions (raster scan) of the sample. The fluorescent decay in three spectral channels has been recorded in each position. The T columns of the $N \times T$ data matrix $\boldsymbol{D}$ in Eq. (2.1) has been created by concatenating

the decays from the three channels into $N \times 1$ vector for each pixel (T pixels in total). Note that the spatial and temporal/spectral domain are reversed when compared to our application to QD data in Sect. 2.3. The $N \times K$ NMF estimated components ($\boldsymbol{W}$ in Eq. (2.1)) were interpreted as spectral signatures of pure components. The $K \times T$ components $\boldsymbol{H}$ (see Eq. (2.1)) were interpreted as the quantity of the pure components in each pixel and therefore provide information about the composition in each position. The three recovered pure components were identified with the spectral signature ($\boldsymbol{W}$) as collagen, lipids and elastin. The spatial matrix $\boldsymbol{H}$ then provides information about the content of each constituent in different positions of the specimen. The recovery of $K = 3$ components from the measured data was based on the biochemical understanding of the specimen (histological section of a human coronary artery tissue). The work does not address the estimation of K in more general cases.

2.4.2 Hoyer's sparse NMF

The in-focus PSF (see leftmost plots in Fig. 2.1) is a fairly compact structure with only few pixels of significant values. Constraints on sparsity of the estimated $\boldsymbol{w}_k$s (individual PSFs) would likely facilitate the estimation of the credible sources and might lead to a faster convergence to a better local minimum.

NMF with explicit sparsity constraints has been developed by Hoyer [46]. The "sparsity" of a vector $\boldsymbol{x}$ was defined as

$$s(\boldsymbol{x}) = \frac{\sqrt{n} - L_1/L_2}{\sqrt{n} - 1}, \tag{2.10}$$

where $L_1 = \sum_i |x_i|$, $L_2 = \sqrt{\sum_i x_i^2}$ and n is the dimensionality of the vector $\boldsymbol{x}$.

Specific fixed constraints on the sparsity of the columns of $\boldsymbol{W}$ can be imposed during the optimisation. After each iteration, the columns $\boldsymbol{w}_k$s of the estimated matrix $\boldsymbol{W}$ are projected to be non-negative, have unchanged L_2 norm, but L_1 norm set to achieve the desired sparseness Eq. (2.10).

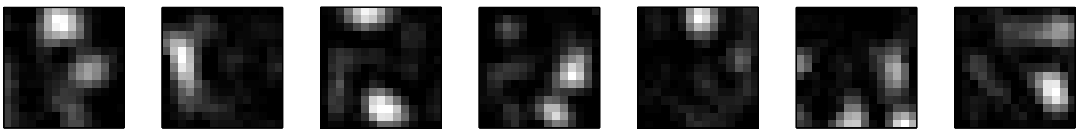
Note that the assumption that all columns have identical "sparseness" might be restric-

tive when out-of-focus PSFs are present. For example, the in-focus PSF in Fig. 2.1 has Hoyer's sparsity $s = 0.83$ while the PSF from $1\,\mu$m out-of-focus plane has $s = 0.4$ and the PSF from $1.8\,\mu$m out-of-focus plane has $s = 0.1$.

Hoyer's sparse NMF algorithm minimises $\|\boldsymbol{D} - \boldsymbol{W}\boldsymbol{H}\|^2$ rather than the KL divergence Eq. (2.7). This cost function corresponds to the Gaussian rather than Poisson noise assumption, which can be significant especially for low-intensity images (fast acquisition time, for example).

(a) 10 sources/μm^2

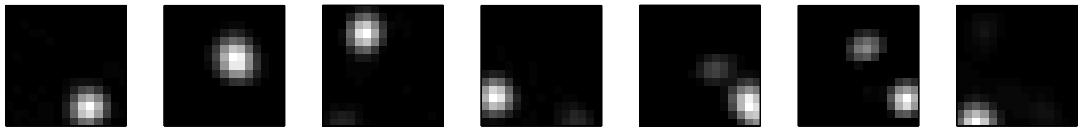
(b) 30 sources/μm^2



(c) 50 sources/μm^2

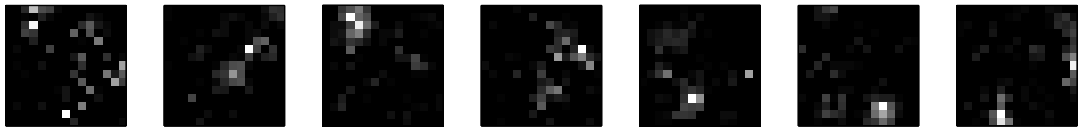
Figure 2.3: $\boldsymbol{W}$ estimated with Hoyer's algorithm with no sparsity constraints. This corresponds to conventional NMF. Evaluation of the simulated data of randomly scattered sources with different densities. Shown first 7 estimated components.

We used simulated data of randomly scattered sources with densities $10 - 50\,\mu\text{m}^{-2}$ to explore the ability of the Hoyer's algorithm to recover credible sources. The blinking intensity was uniformly distributed on the interval $[0, 5000]$ photons. The background was set to 100 photons/pixel and data were corrupted with Poisson noise. Prior to the evaluation with Hoyer's algorithm, the true background value was subtracted from the data, clipping negative pixels to zero. The number of components K was set to the true value used for simulation $K = K_{true}$. The algorithm was run for 1000 iterations.



(a) 10 sources/μm^2

(b) 30 sources/μm^2



(c) 50 sources/μm^2

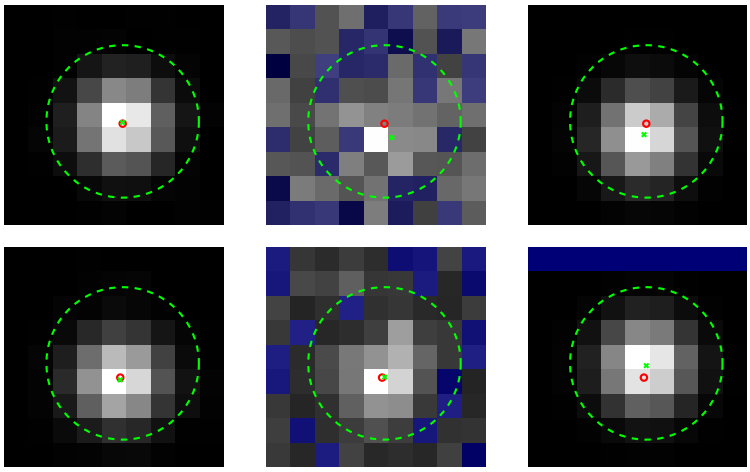
Figure 2.4: Sparsity constraints $s = 0.7$ on $\boldsymbol{W}$ estimated with Hoyer's algorithm from simulated data of randomly scattered sources with different densities. Shown first 7 estimated components.

Running the algorithm for longer (2000, 5000) iterations did not improve the estimated results.

Figure 2.3 shows the estimated $\boldsymbol{W}$ with Hoyer's algorithm without sparsity constraints. This corresponds to conventional NMF. Note that most of the $\boldsymbol{w}_k$ for higher densities contain multiple sources Fig. 2.3b,c. Imposing the sparsity constraints $s = 0.7$ on the columns of $\boldsymbol{W}$, estimated from the true PSF, gives better estimated sources for densities$< 30\,\mu$m^{-2} Fig. 2.4a, however, for dense data the method fails to recover the individual sources and gives unsatisfactory results, see Fig. 2.4c.

2.4.3 Independent component analysis

The independent component analysis (ICA) algorithm [57] has been used for separating the overlapping QDs [30, 32]. ICA allows each source to have a different individual PSF. However, the ICA model allows negative entries in the individual PSFs and does not account for noise in the measured data, which can make recovery of the individual



(a) NMF (noise) **(b)** ICA (noise) **(c)** ICA (noise free)

Figure 2.5: Comparison of the components separated with (a) NMF and (b) ICA for simulated noisy data of two blinking QDs separated by $0.5\,\mathrm{pixel}$ (which corresponds to $50\,\mathrm{nm}$ or $\lambda/12$). (c) ICA for noise-free data. Blue pixels contain negative values. The true and the estimated positions are shown as red circles and green crosses, respectively. The airy disk is shown as a green circle (radius $333\,\mathrm{nm}$).

sources difficult in realistic noise levels (see Fig. 2.5).

Figure 2.5b shows results from 10^3 simulated frames containing two sources with blinking intensity uniformly distributed on the interval $[0, 1500]$ photons and with background 100 photons/pixel. The true background level was subtracted (clipping any negative values to zero) prior to the ICA evaluation. We used `'tanh'` as a nonlinearity option in the fixed-point algorithm [57], and the number of sources was set to $K = K_{true} = 2$.

2.4.4 Richardson – Lucy deconvolution

There is a link between NMF and the classical Richardson – Lucy deconvolution algorithm. An observed "blurred" (diffraction limited) image $\boldsymbol{i}$ ($N \times 1$ vector) can be expressed as a (discretised) convolution

$$i_x = \sum_{j=1}^{N} o_j w_{x-j}, \tag{2.11}$$

where $\boldsymbol{o}$ ($N \times 1$) is the original (unblurred) object which represents locations and intensities of fluorescent sources. $\boldsymbol{w}$ ($N \times 1$) is an image of point spread function (PSF) centred in the middle of the image. Richardson [58] and Lucy [59] published an iterative deconvolution technique for astronomical images with known PSF. They used Bayes' theorem as a "hint" for an iterative update of $\boldsymbol{o}$. This update is usually referred to as Richardson – Lucy (RL) deconvolution algorithm and is identical to the Lee – Seung NMF update with generalised KL divergence objective function [28].

Holmes [60] derived the RL updates based on maximum likelihood estimation of the model with Poisson noise using the expectation-maximisation algorithm. He also proposed an update for $\boldsymbol{w}$ so that the method can be used as a blind deconvolution algorithm (PSF is not known). This is sometimes referred to as a "blind RL algorithm".

The updates for $\boldsymbol{o}$ and $\boldsymbol{w}$ are technically identical to the Lee and Seung NMF updates (KL divergence as an objective function). However, (blind) RL deconvolution solves a different problem than NMF. RL deconvolution estimates one PSF ($\boldsymbol{w}$), which is shared

by all sources. The deconvolution is performed for each frame separately, independent on the rest of the dataset. NMF models the whole dataset as a collection of individual (and in general different) PSFs (columns of $\boldsymbol{W}$) each changing intensity over time (rows of $\boldsymbol{H}$). While one source which appears in n different frames is treated as n different individual sources by RL, NMF can identify it as a single source.

Modified updates imposing radial symmetry constraints on the PSF were also proposed. There exist several modified updates derived using EM algorithm which impose some constraints on $\boldsymbol{o}$ or $\boldsymbol{w}$. Joshi [61] gives updates, where Good's roughness measure ($\int \frac{|\nabla f(x)|^2}{f(x)} dx$) on the original image $\boldsymbol{o}$ is used as a regularisation term. This biases the solution towards the smooth images and avoids speckle artefacts in the reconstructions, that are sometimes experienced in deconvolution methods.

Fish et al. [62] use blind RL algorithm (updates on both $\boldsymbol{o}$ and $\boldsymbol{w}$) but after some number of iterations they fit an approximation of the PSF to the estimated $\boldsymbol{w}$ and use this fit as a new $\boldsymbol{w}$. They claim that in noisy images this "semi-blind" deconvolution can perform better than the one with known PSF. The comparison of the regularised RL versions and some other deconvolution techniques has been shown in [63, 64]. RL usually performs well for noisy images.

2.5 Simulations

In this section we describe how we generated the simulated datasets. Simulated data were used for testing the performance of the algorithm in different experimental settings.

The parameters of the simulations were chosen to correspond to real experimental data with quantum dots (QD625, *Invitrogen*). Table 2.1 summarises the main simulation parameters. Radius of an Airy disk (classical resolution limit) for parameters from Tab. 2.1 is $\delta = 293\,\text{nm}$. FWHM of a Gaussian approximation of the in-focus PSF is 260 nm ($\sigma = 111\,\text{nm}$) [8]. The pixels size in the image plane (80 nm) is below Nyquist

limit (120 nm) estimated from the Abbe resolution limit in Sect. 1.4.

Parameter	Note	Value
λ_{em}	wavelength of the emission light	625 nm
NA	numerical aperture	1.3
RI	refractive index of the immersion oil	1.5
δ	radius of the Airy disk	293 nm
pixel-size	size of a pixel in image plane	80 nm
T	number of frames	$50 - 1000$
$\text{mean}(n_{phot})$	mean intensity	2500 photons / source / frame
$\max(n_{phot})$	max intensity	5000 photons / source / frame
b	background	100 photons / pixel / frame
noise	-	Poisson

Table 2.1: Main parameters used for data simulations.

The blinking behaviour of the QDs was simulated as either:

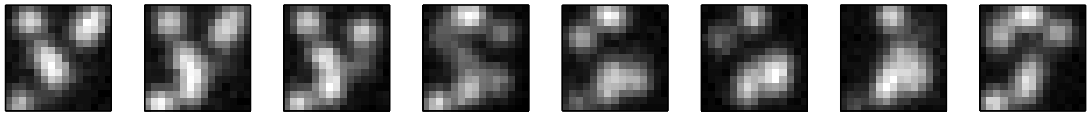
1. The uniform random distribution between 0 and $\max(n_{phot})$.

2. The telegraph process with the switching rate γ. The telegraph process is a Markov process, which takes only two values (in our case ON and OFF fluorescence states). The difference between the sampling rate of the experiment and the blinking of the fluorophores is considered by simulating the blinking behaviour on the oversampled time axis followed by averaging over several bins (see Sect. 2.8.1 for details).

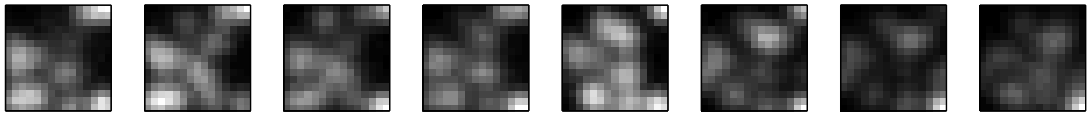
2.5.1 Randomly scattered sources

The ability of the algorithm to separate individual overlapping sources was tested on simulated data of randomly scattered fluorophores. The true positions of the sources were generated as a realisations of a homogeneous Poisson process [65]. The density

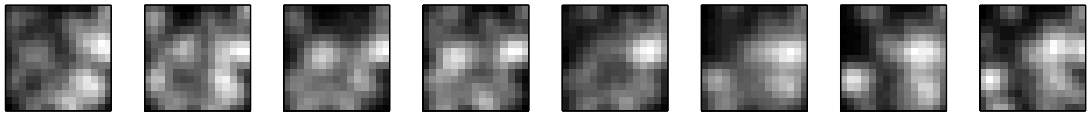
of the sources was in a range $\rho = 10 - 50\,\text{sources}/\mu\text{m}^2$. This density range corresponds to $\sim 3 - 14$ sources in an Airy disk, respectively, for parameters from Tab. 2.1. Several frames of the simulated dataset for three different densities are shown in Fig. 2.6. The mean projection of the frames, which corresponds to a wide-field image, is shown in Fig. 2.7.



(a) density $10\,\mu\text{m}^{-2}$ (14 sources)

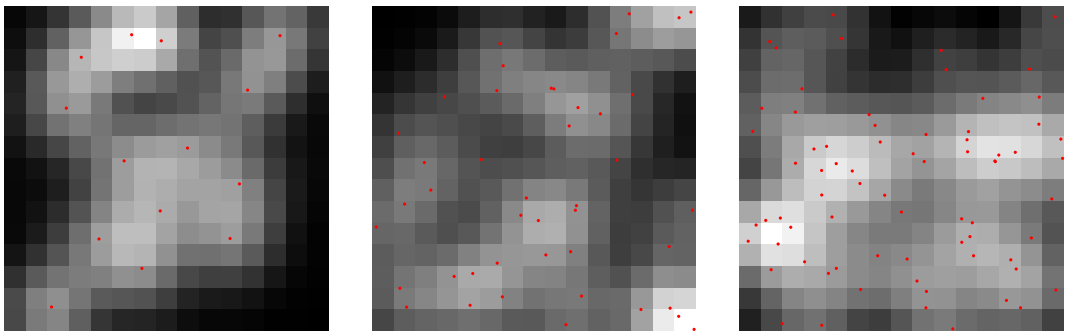


(b) density $30\,\mu\text{m}^{-2}$ (43 sources)



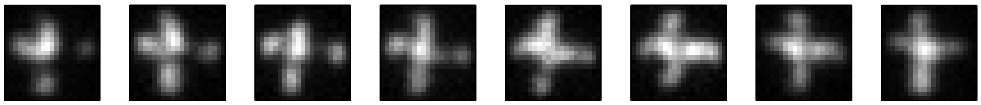
(c) density $50\,\mu\text{m}^{-2}$ (72 sources)

Figure 2.6: First eight frames of simulated randomly scattered sources with density $10 - 50\,\mu\text{m}^{-2}$. The area of the frame is $1.2 \times 1.2\,\mu\text{m}$ (15×15 pixels).

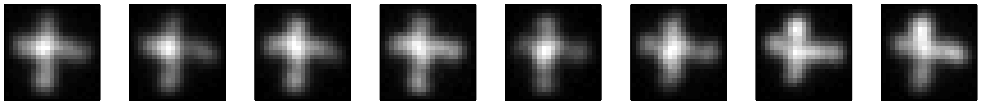


(a) density $10\,\mu\mathrm{m}^{-2}$ **(b)** density $30\,\mu\mathrm{m}^{-2}$ **(c)** density $50\,\mu\mathrm{m}^{-2}$

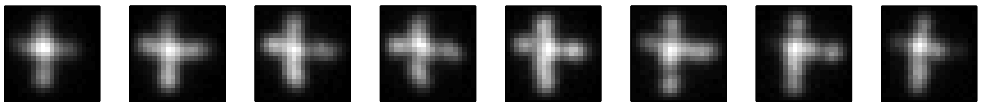
Figure 2.7: Mean projection of the simulated frames Fig. 2.6 shows the high overlap of the individual emitters. The sources' positions are marked with red dots. The area of the frame is $1.2 \times 1.2\,\mu\mathrm{m}$ (15×15 pixels).



(a) linear density $\mu = 7.5\,\mu\text{m}^{-1}$, distance between lines $d = 150\,\text{nm}$



(b) linear density $15\,\mu\text{m}^{-1}$, distance between lines $d = 150\,\text{nm}$

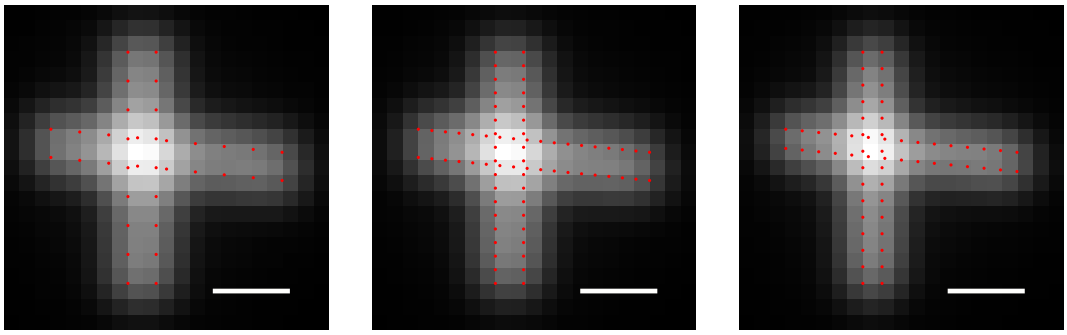


(c) linear density $12.5\,\mu\text{m}^{-1}$, distance between lines $d = 100\,\text{nm}$

Figure 2.8: First eight frames of the simulated dataset. The area of the frame is $1.7 \times 1.7\,\mu\text{m}$ (21×21 pixels). Note that sub-resolution features such as double line structure marked with red dots in Fig. 2.9 cannot be observed.

2.5.2 Artificial structure

A dataset with sources arranged in a shape of a hash symbol (#) was used for testing the algorithm to recover structural details in the sample. The vertical parallel lines were aligned with the pixels grid, the horizontal lines were slightly tilted to investigate the possible effect caused by the geometrical configuration of the sources with respect to the pixel grid.



(a) $\mu = 7.5\,\mu\text{m}^{-1}$, $d = 150\,\text{nm}$ **(b)** $\mu = 15\,\mu\text{m}^{-1}$, $d = 150\,\text{nm}$ **(c)** $\mu = 12.5\,\mu\text{m}^{-1}$, $d = 100\,\text{nm}$

Figure 2.9: Sum projection of the simulated frames Fig. 2.8. The sources' positions are marked with red dots. Scale bar $400\,\text{nm}$.

The distance d between the parallel lines and the linear density of the sources μ were two main parameters of the structure. The brightness and the background values are shown in Tab. 2.1. Figure 2.8 shows several frames of the simulated dataset for different linear densities μ and distances between parallel lines d. The distance $d = 150\,\text{nm}$ corresponds to the half of the Airy disk radius. The mean projections of the frames are shown in Fig. 2.9.

2.6 Evaluation of the results

The performance of the algorithm applied on a simulated dataset can be quantitatively measured, because the true locations of the sources are known. We used several measures to compare the performance on simulated datasets consisting of randomly scattered in-focus PSFs (see Fig. 2.6).

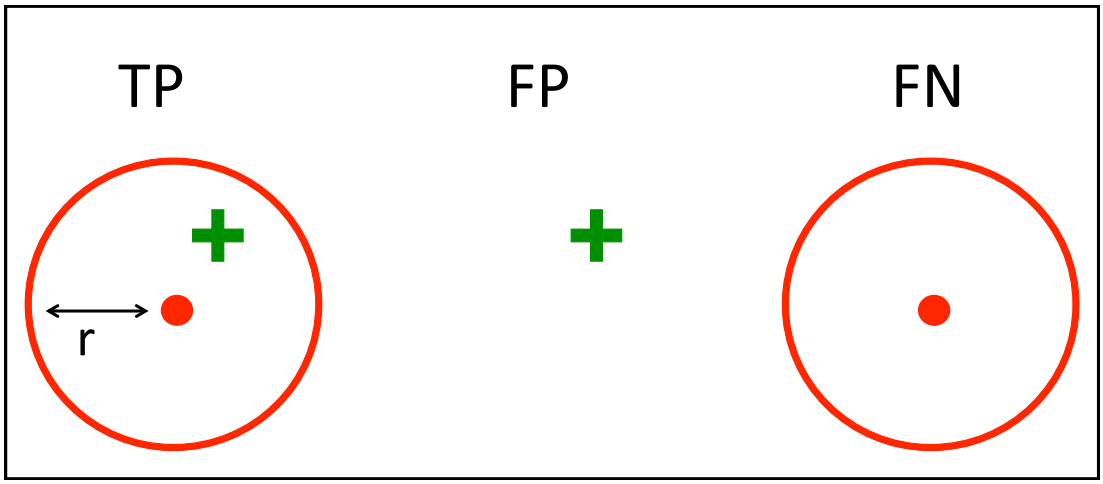


Figure 2.10: True positives (TP), false positives (FP) and false negatives (FN) illustration. A red dot represents the true location with a circle of radius r, a green cross denotes an estimated position.

The individual estimated sources $\boldsymbol{w}_k$ were localised by ML fitting of a Gaussian approximation of the PSF [8]. We used a greedy algorithm to assign the estimated locations (E) to their nearest true positions (T). Only one estimated position was assigned to each true position. If the distance between the estimated and the true position was

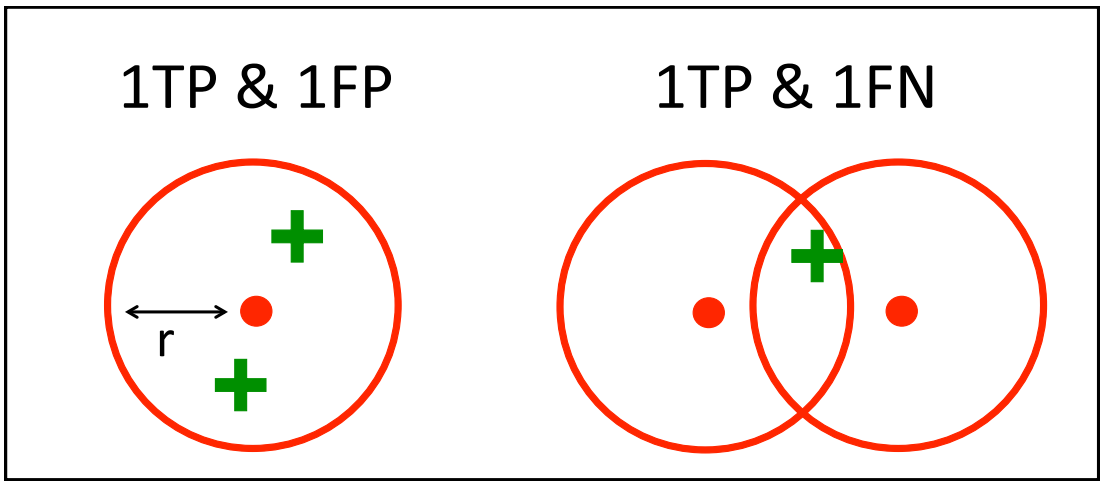


Figure 2.11: There is only one estimated position assigned to each true position. Two estimated sources in the proximity of one true source are counted as 1TP and 1FP (left). One estimated source in proximity of two true sources gives 1TP and 1FN (right)

smaller than a threshold r, then the source was consider as a true positive (TP). Each true position with no estimated source within a disk of radius r was counted as false negative (FN), whereas an estimated position further than r from any true position was considered as false positive (FP), see Fig. 2.10. M estimated sources in the proximity of one true source are counted as 1TP and $(M-1)$FP (Fig. 2.11, left). One estimated source in proximity of M true sources gives 1TP and $(M-1)$FN (Fig. 2.11, right).

We set the threshold $r = \sigma/2$, where $\sigma = \frac{\sqrt{2}}{2\pi}\frac{\lambda_{em}}{NA}$ is the standard deviation of the in-focus PSF Gaussian approximation [8]. For the parameters used in our simulations (see Tab. 2.1) the threshold corresponds to $r = 56\,\text{nm}$ ($0.7\,\text{pixels}$).

The number of all TP divided by the area of the image gives the estimation of the sources' density. This simple measure shows the ability of the method to separate the individual sources despite their overlap in the original data. The estimation of TP is independent on the brightness of each source. The localisation precision, however, scales as $1/\sqrt{N}$ [29]. We can therefore consider the position of bright sources to be estimated with better localisation precision and therefore higher confidence. We can compare the ability to recover the individual sources at different "confidence levels" by considering only the sources localised with certain precision. The average precision

(AP) [66, 67] summarises both localisation precision and ability to recover the individual sources. The estimated positions e_k are ranked according to the square root of the source's mean brightness

$$b_k = \sqrt{\bar{N}_k} \tag{2.12}$$

because the Cramér – Rao lower bound on localisation precision scales as $1/\sqrt{N}$, where N is a number of emitted photons (see Chapter 3 for details). For the results of the NMF evaluation, N is retrieved from the matrix $\boldsymbol{H}$ as a mean along the rows

$$\bar{N}_k = \underset{t}{\text{mean}}(h_{kt}). \tag{2.13}$$

The interval $[l_{min}, l_{max}]$ between the dimmest l_{min} and the brightest l_{max} source intensity is divided into a number of intervals (confidence levels) l_i defined by the steps in the sorted intensities of all sources. For each confidence level l_i only the sources with b_k above l_i are considered. True positives (TP_i), false negatives (FN_i) and false positives (FP_i) are computed for each confidence level l_i.

Precision P and recall R are computed from $\text{TP}(l_i)$, $\text{FP}(l_i)$ and $\text{FN}(l_i)$ for each confidence level l_i:

$$P(l_i) = \frac{\text{TP}(l_i)}{\text{TP}(l_i) + \text{FP}(l_i)}, \tag{2.14}$$

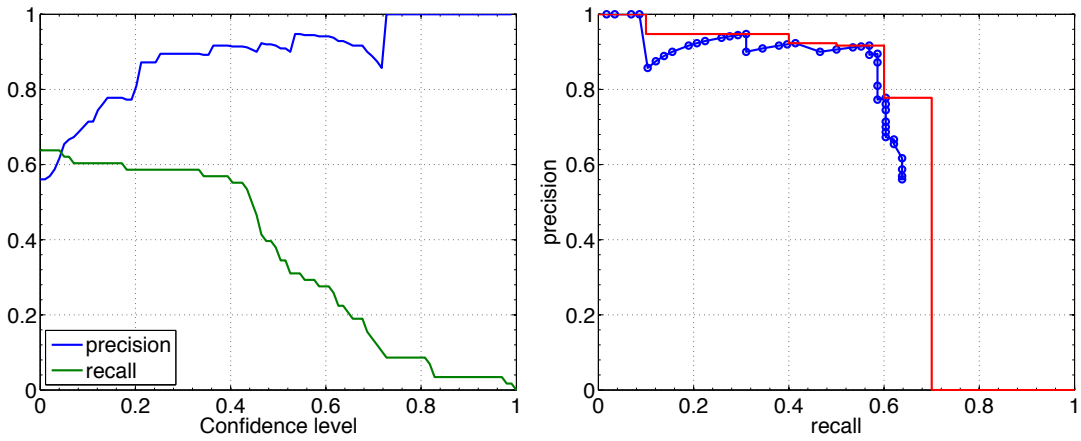
$$R(l_i) = \frac{\text{TP}(l_i)}{\text{TP}(l_i) + \text{FN}(l_i)}. \tag{2.15}$$

An example of precision $P(l_i)$ and recall $R(l_i)$ curves for different confidence levels is shown in Fig. 2.12a.

Following [67], the precision/recall (PR) curve $P(R)$ (Fig. 2.12a) is interpolated for 11 equally spaced recall levels $\tilde{R}_i \in [0 : 0.1 : 1]$ by taking the maximum precision for which the corresponding recall exceeds $\tilde{R}_i$ (Fig. 2.12b):

$$P_{interp}(\tilde{R}) = \max_{R;R \geq \tilde{R}} P(R). \tag{2.16}$$

The precision/recall (PR) curve is interpolated in order to reduce the impact of "wiggles" in the PR curve (see Fig. 2.12b). Note that to obtain a high AP, the method must



(a) (b)

Figure 2.12: (a) Example of the precision $P(l_i)$ (blue) and recall $R(l_i)$ (green) curve. (b) The precision/recall curve $P(R)$ (blue) with interpolated precision $P_{interp}(\tilde{R})$ (red).

have precision at all levels of recall. This penalises the methods that can accurately estimate only few very bright sources.

Average precision (AP) is then defined as the mean of interpolated precision:

$$AP = \frac{1}{11} \sum_{\tilde{R}} P_{interp}(\tilde{R}). \tag{2.17}$$

2.7 NMF for realistic microscopy datasets

NMF becomes challenging when applied to a dataset with large number ($\sim 10^3$) of images, each containing more than 10^4 pixels and more than 10^2 QDs. Beside long computational time (the computational complexity of the NMF updates scales linearly with the number of pixels - see Sect. 2.2), the local minima in NMF fitting complicate the optimisation [68].

We address this partly by dividing the data into overlapping patches, so that NMF is applied to each patch individually (see Sect. 2.7.1). In the end, the results from the patches are "stitched" back together.

Several methods for estimation of the number of components in the NMF model is discussed in Sect. 2.7.2. We have also developed methods to reduce local optima problems in the fitting procedure (iNMF algorithm discussed Sect. 2.7.3). In Sect. 2.7.4 we show a simple approach to asses the quality of the individual estimated components. The results of the NMF can be used in two different ways. The separated individual sources $\boldsymbol{w}_k$ can be localised and the estimated positions can be used either directly or to create a sub-resolution image very much like in the conventional LM techniques (see Sect. 2.7.5). A different approach avoids the localisation step and creates the super-resolution image directly from the estimated $\boldsymbol{w}_k$s by combining their "squeezed" versions (see Sect. 2.7.6).

The whole pipeline for NMF evaluation of a realistic dataset is described in this section. The individual steps of the procedure are illustrated on simulated data.

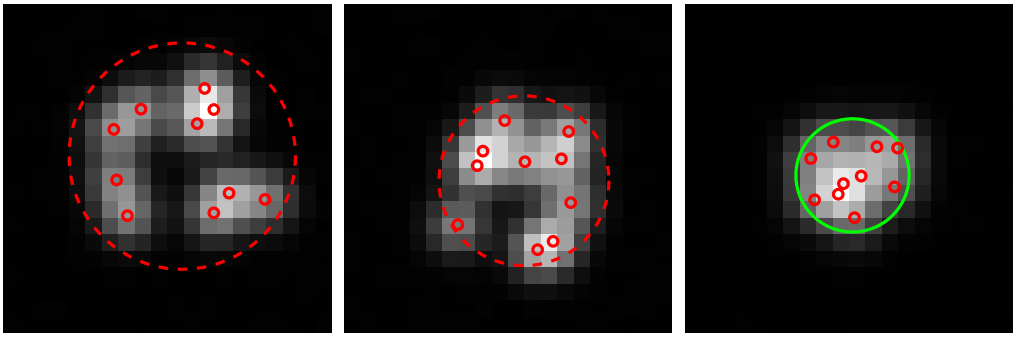
2.7.1 Pre-processing

Raw data are calibrated such that the image intensity corresponds to the photon counts. Each image is divided into patches of $n_x \times n_y$ pixels with o pixels overlap (see Fig. 2.39). We usually use $n_x = n_y = 25$ and $o = 5$. The overlap has been chosen as the estimated extent of a single in-focus point spread function. Each time frame of the patch p is reshaped into an $N \times 1$ vector ($N = n_x n_y$) by concatenation of the columns (see Fig. 2.2). All T frames then create a $N \times T$ data matrix $\boldsymbol{D}_p$.

To detect patches with low signal, the maximum intensity pixel in the time average of each patch $m_p = \max_i \langle \boldsymbol{D}_p(i,t) \rangle_t$ is compared to the maximum intensity pixel of the average of the whole dataset $m = \max_i \langle \boldsymbol{D}(i,t) \rangle_t$. The patches with $m_p/m < t_m$ contain weak signal and are not considered for further evaluation. For our evaluation we usually set $t_m = 0.25$.

2.7.2 Estimation of number of sources K

The NMF model requires prior knowledge about the number of sources K to be separated. The low rank factorisation of Eq. (2.1) will not hold exactly due to the presence of noise. This makes estimation of K is a difficult task for noisy datasets with parameters set to realistic experimental conditions. In preliminary work we explored this on simulated data with the NMF model fitted for a range of K values. We used following parameters for the simulation: $\lambda_{em} = 655\,\text{nm}$, $NA = 1.2$, maximum intensity 1500 photons/source/frame, background 100 photons/frame/pixel. A frame of two sources with corresponding noise levels are shown in Fig. 3.4. We considered Poisson noise Eq. (B.1) as a noise model in simulated data. Due to the property $\text{var}(X) = \text{mean}(X)$ of Poisson distributed variable X[69] the intensity values in the background pixels fluctuate around the mean value of 100 photons with standard deviation $\sigma = \sqrt{100} = 10$ photons while the pixels at the maximum intensity $1500 + 100$ photons fluctuate with $\sigma = \sqrt{1600} = 40$ photons.



(a) 2δ, wide-field **(b)** 1.5δ, wide-field **(c)** δ, wide-field

Figure 2.13: Sum of the simulated frames. Red marks indicate the locations of the sources. Green circle shows the Airy disk (with radius δ). Ten sources are randomly distributed on a disk with radius 2δ (left column), 1.5δ (middle column) and δ (right column). The border of the disk in (a) and (b) is marked with red dashed circle.

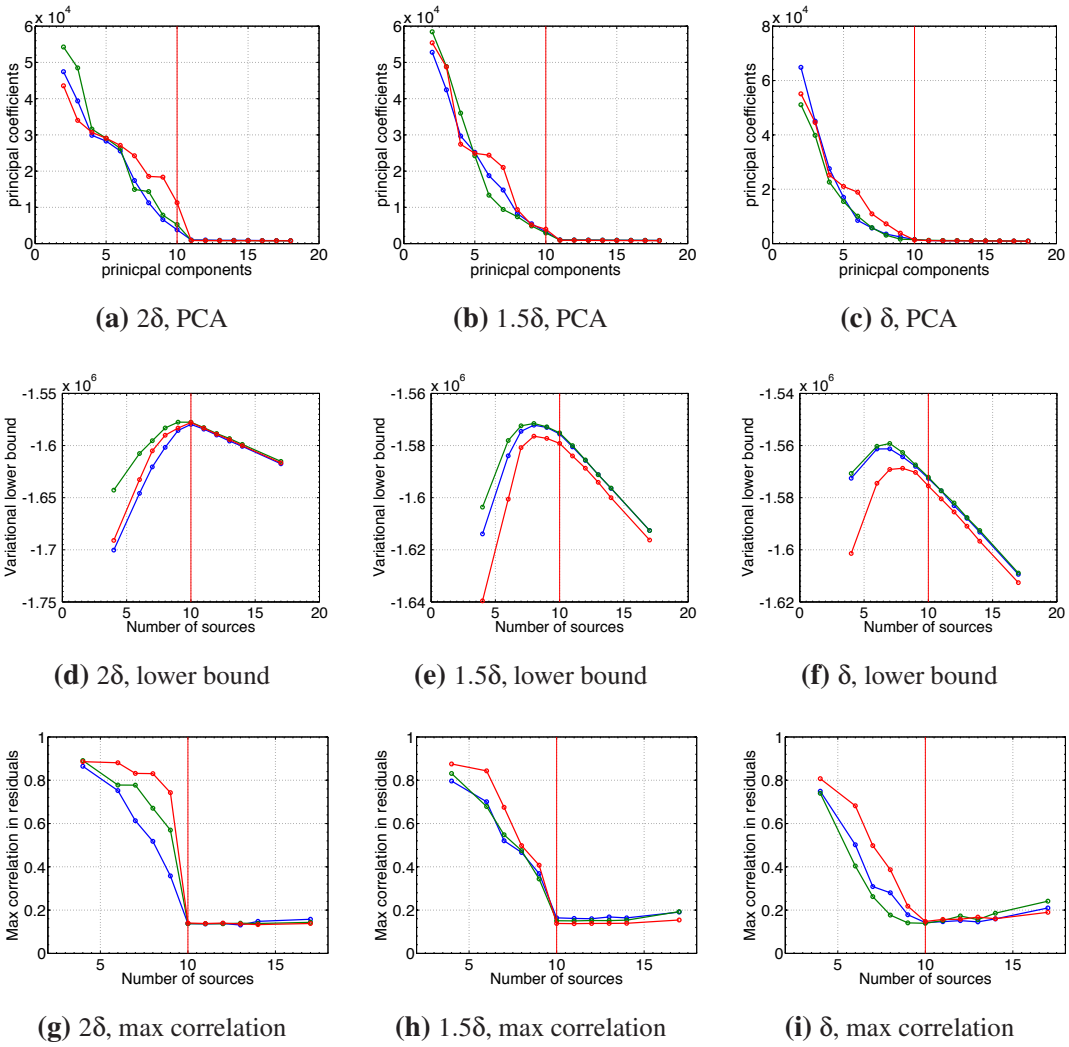
Figure 2.13 illustrates three different simulated datasets with 10 sources randomly scattered within an area of radius 2δ, 1.5δ and δ, where δ was equal to the diameter of an

Airy disk (diffraction limit), shown as a green circle in Fig. 2.13c ($\delta = 0.61\lambda_{em}/\mathrm{NA}$). This corresponds to the sources densities of 2.4, 4.4 and 10 sources per Airy disk or 7, 13 and 29 sources per μm^2, respectively. From my experience all the methods delivered very precise estimate of K for 2δ situation. Therefore I chose this as the upper limit. For 1δ all the methods more or less failed and therefore I set that value as a lower limit. I choose 1.5δ as an intermediate value between 1δ and 2δ. The mean of the simulated frames, which corresponds to a wide-field image, is shown as a grey-value image. Red marks indicate the true positions of the sources. Ten datasets with different geometrical configurations of randomly scattered sources were simulated for each source density.

The likelihood of the model Eq. (2.6) is increasing with higher K, because the noisy data can always be fitted better with a model containing higher number of components. The Bayesian Information Criterion (BIC) [69] is a simple model comparison method, adding a penalty term to the likelihood penalising for the NK parameters contained in $\boldsymbol{W}$. The models with larger K are therefore more heavily penalised. BIC, however, did not provide satisfactory results. We therefore tried to estimate the number of sources K using:

1. *Principal Component Analysis (PCA)* A crude estimation of K can be obtained from the position of the "kink" in the plot of sorted principal values Fig. 2.14a-c. However, the "kink" is not obvious in the presence of noise or for data with high density of blinking sources, see Fig. 2.14c.

2. *A variational lower bound (LB)* A variational approximation of the GaP model Sect. 2.2.2 provides lower bound $\mathcal{L}$ on the likelihood $p(\boldsymbol{D}|K,\theta)$ by approximately integrating out the latent variables $\boldsymbol{h}_k$ [70]. To obtain the marginal likelihood $p(\boldsymbol{D}|K)$ it would be necessary to also integrate out θ, but this is computationally challenging. We show in Fig. 2.14d-f that in fact the lower bound already underestimates the value of K, so that $p(\boldsymbol{D}|K)$ would likely peak at even lower values of K and thus systematically underestimate the number of sources.



(a) 2δ, PCA **(b)** 1.5δ, PCA **(c)** δ, PCA

(d) 2δ, lower bound **(e)** 1.5δ, lower bound **(f)** δ, lower bound

(g) 2δ, max correlation **(h)** 1.5δ, max correlation **(i)** δ, max correlation

Figure 2.14: K estimation for 10 sources contained within a disk with radius 2δ (*left column*), 1.5δ (*middle column*) and δ (*right column*). Lines for three datasets with different configuration of the sources are shown in different colours. K_{true} is marked with red vertical line.

3. *Analysis of correlations in residuals (ACR)* An alternative approach for estimating K is to analyse the residuals (data minus model). The entries of the $N \times T$ residual matrix $\boldsymbol{S}$:

$$s_{nt} = \frac{d_{nt} - \sum_{k=1}^{K} w_{nk} h_{kt}}{\sqrt{\sum_{k=1}^{K} w_{nk} h_{kt}}}. \tag{2.18}$$

The factor $1/\sqrt{\sum_{k=1}^{K} w_{nk} h_{kt}}$ is applied in order to standardise the residuals (zero mean and unit variance) of Poisson distributed data. We can then compute the

$N \times N$ correlation matrix

$$\boldsymbol{C}_S = \boldsymbol{S}\boldsymbol{S}^T, \tag{2.19}$$

and the $N \times N$ matrix of the correlation coefficients $\boldsymbol{R}_S$ with entries

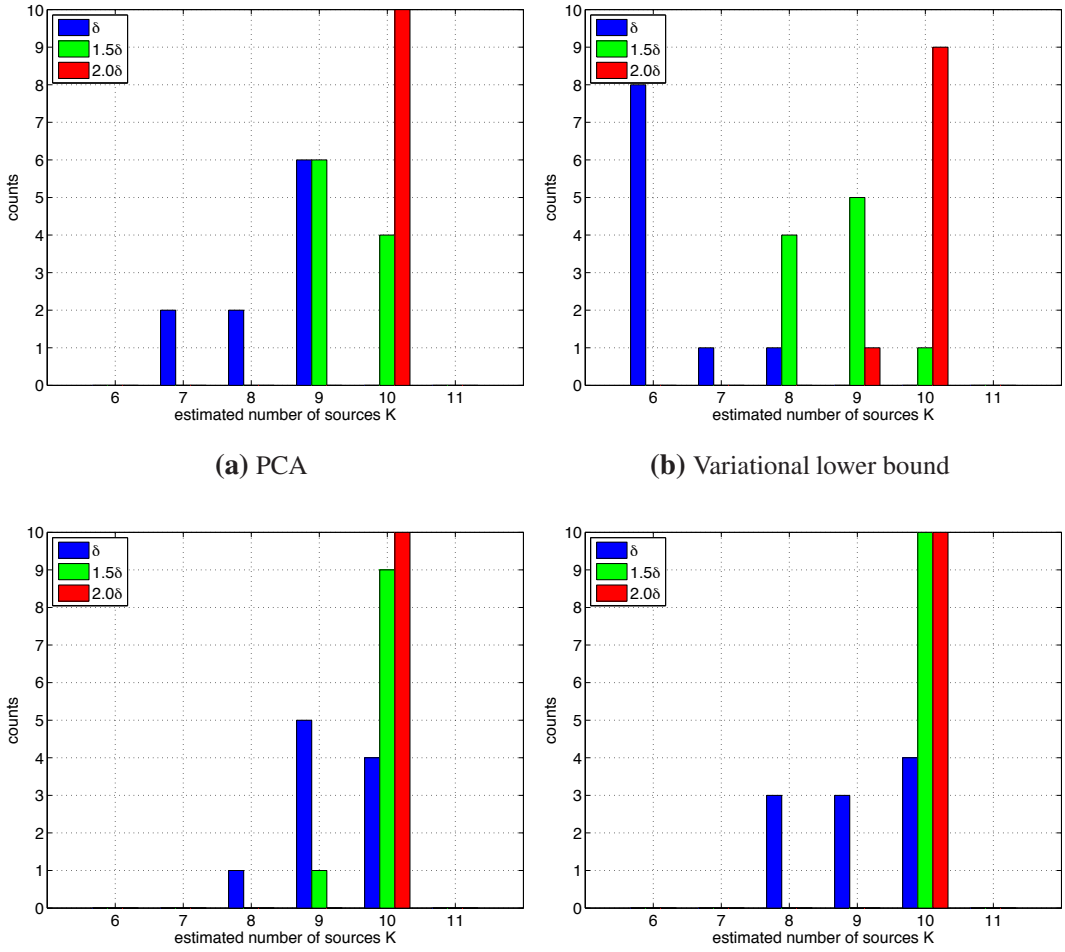
$$r_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}c_{jj}}}. \tag{2.20}$$

Underestimation of the number of sources ($K < K_{true}$) leads to correlations between some pixels as the model tries to explain multiple sources with one component. For $K \geq K_{true}$ the correlations are expected to drop to a base level and the residuals become uncorrelated. We can pick the value of K for which the maximum of the residual correlations decreases to a certain level and where further increase of K does not give any further improvement Fig. 2.14g-i.

A reliable estimation of K is a difficult task for higher source densities. Figure 2.15 shows the histograms of the estimated Ks for ten different geometrical configurations of the sources with a given density. From the three methods presented in this section (Fig. 2.15a-c), the analysis of the correlations in residuals (ACR) shows the best performance. However, while ACR shows good performance for simulated datasets, it can be challenging for analysis of real data if there are correlations that are not captured by the model. Also note that ACR requires evaluation of NMF for a suitable range of Ks, which makes the method cumbersome.

Both LB and ACR require evaluation of the model for a reasonable range of possible Ks. The range can be estimated from PCA (Fig. 2.15c), because the principal coefficients can be computed directly from the data matrix $\boldsymbol{D}$.

In the following section we will be discussing an iterative procedure of the NMF algorithm (iNMF) which can deal with moderate overestimation of K (estimated from PCA). The correct number of sources can be estimated additionally by analysing the optimised matrix $\boldsymbol{W}$ and selecting the "credible" sources $\boldsymbol{w}_k$. Therefore evaluation for only one overestimated value of K, rather than a range of Ks, is required. The



(a) PCA

(b) Variational lower bound

(c) Maximum correlations in residuals

(d) NMF with iterative restarts (iNMF)

Figure 2.15: Histograms of the K estimations ($K_{true} = 10$) with (a) PCA, (b) variational lower bound, (c) analysis of correlations in residuals and (d) iterative NMF (discussed further in the text in Sect. 2.7.3). Histograms are from the evaluation of simulated data of randomly scattered emitters: ten sources within a disk of δ (blue), 1.5δ (green) and 2δ (red). Ten different geometrical configurations were simulated for each density.

histogram of the Ks estimated with the iterative algorithm is shown in Fig. 2.15d for comparison. iNMF gives the most accurate estimates.

2.7.3 Tackling local optima in NMF fitting with iterative restarts

Although the Lee and Seung algorithm is convex with respect to $\boldsymbol{W}$ and $\boldsymbol{H}$ separately, it is non-convex in both simultaneously [28]. Multiple restarts can be used to address the problem of local optima, but we have not found good solutions with this approach. NMF tends to get persistently stuck in the local optima of the optimisation landscape. Instead, we exploit some prior knowledge about the problem, namely that the PSFs are likely to have a fairly compact structure (see Fig. 2.1). As the estimated sources $\boldsymbol{w}_k$ are normalised to have the L_1 norm equal to one (i.e., $\sum_j w_{jk} = 1$, see Sect. 2.3), we use the inverse L_2 norm to rank the columns $\boldsymbol{w}_k$'s of the matrix $\boldsymbol{W}$. Note that Hoyer's sparsity Eq. (2.10) is an L_1/L_2 measure normalised to the $[0..1]$ interval [68].

This leads to an iterative NMF algorithm (we denote it as iNMF), where on iteration $(j+1)$ the first j sorted sources $\{\boldsymbol{w}\}_1^j$ in decreasing manner (and corresponding $\{\boldsymbol{h}\}_1^j$) are used as initial values for the first j columns of $\boldsymbol{W}$ (and the corresponding rows of $\boldsymbol{H}$). The remaining components are re-initialised from a uniform random distribution. Initial values of $\boldsymbol{W}$ and $\boldsymbol{H}$ for the $(j+1)$th iteration are therefore composed of the j "sparsest" components of the previous iteration and $(K-j)$ randomly initialised components. The procedure runs until $j = K$. The iNMF algorithm is summarised in Algorithm 1.

We used a crude over-estimation of K with PCA because it can be computed directly from data $\boldsymbol{D}$ prior to the evaluation:

1. We compute the sorted principal coefficients λ_j of $\boldsymbol{D}$ ($\lambda_1 > \lambda_2 > \ldots$).
2. K is set to the number of components which satisfy $\lambda_j/\lambda_1 > t_{PCA}$, where t_{PCA} is a threshold.

User should be able to test the source estimation procedure on a patch where the num-

ber of sources can be guessed (e.g. an area with sparse sources) to get a notion about the threshold. The threshold t_{PCA} should be set such that it slightly overestimates the true number of sources.

Algorithm 1 Iterative restarts of the NMF (iNMF algorithm).

1. Set $\boldsymbol{W}_{init}$ and $\boldsymbol{H}_{init}$ as random positive matrices.
2. Iterate for $j = 1 : K$, where K is the (over) estimated number of sources.
   (a) Run NMF with $\boldsymbol{W}_{init}$ and $\boldsymbol{H}_{init}$ as initial values.
   (b) Sort columns of $\boldsymbol{W}$ according to decreasing L_2 norm and permute rows of $\boldsymbol{H}$ correspondingly.
   (c) Replace first j columns of $\boldsymbol{W}_{init}$ with first j columns of sorted $\boldsymbol{W}$.
   (d) Replace last $j+1 : K$ columns of $\boldsymbol{W}_{init}$ with positive random vectors.
   (e) Replace first j rows of $\boldsymbol{H}_{init}$ with first j rows of sorted $\boldsymbol{H}$.
   (f) Replace last $j+1 : K$ rows of $\boldsymbol{H}_{init}$ with positive random vectors.

The motivation for the iterative procedure Algorithm 1 is to progressively exploit the credible (and therefore sparse) components from the data while keeping full flexibility of NMF. It should be noted that in contrast to Hoyer's sparse NMF (Sect. 2.4.2), where the "sparsity" on the $\boldsymbol{w}_k$ is imposed as a "hard" constraint, iNMF leads to a "soft" enhancement of $\boldsymbol{w}_k$'s sparsity. The sparse components are preferably reused in the following iterative restarts but are still allowed to change during the further iterations. This leads to further refinement of the already estimated sources.

iNMF reduces the redundant sources introduced by overestimation of K. For example, a single emitter can be represented by multiple components $\boldsymbol{w}_k$s. The intensity profiles $\boldsymbol{h}_k$s of these redundant components will be correlated and their addition will yield the intensity of the source in each frame. However, iNMF will keep only one component $\boldsymbol{w}_m$ from all redundant sources. The source $\boldsymbol{w}_m$ has already approximately correct

shape and during the next iterations the algorithm it will be mostly changing the profile $\boldsymbol{h}_m$. The several redundant sources will be therefore replaced by a single sources $\boldsymbol{w}_m$. This mechanism can be observed in the step from Fig. 2.16b to c, for example.

iNMF is illustrated on simulated data of a slanted line with eight attached PSFs in Fig. 2.16. The parameters of the simulations are discussed in Sect. 2.8. An illustration of typical frames of the dataset is shown in Fig. 2.35a and the true sources are displayed in Fig. 2.35b.

For this data we set $K = 15$, with the last component reserved for background. The results of the first run (random initialisation) are shown in Fig. 2.16a. The individual PSFs (see Fig. 2.35b) are spread across all $\boldsymbol{w}_k$s, and many of them contain a mixture of multiple PSFs. This is a typical solution corresponding to a local minimum of the objective function Eq. (2.7). As the iterative procedure progresses, realistic sources are gradually recovered, see Fig. 2.16b. After eight iterations, the first eight $\boldsymbol{w}_k$'s show credible PSFs, while the rest represent only noise, see Fig. 2.16c. Further iterations do not have a significant effect on the already estimated PSFs, see Fig. 2.16d.

The iNMF procedure leads to better local minima of the NMF optimisation problem Fig. 2.17. The L_2 norm sorting of the recovered $\boldsymbol{w}_k$s after every iteration (step 2a in Algorithm 1) ensures that the sparsest components will be reused in subsequent evaluation. Gradually increasing number of sparse $\boldsymbol{w}_k$s with small L_2 norm is reused in subsequent restarts (step 2c in Algorithm 1), whereas the $\boldsymbol{w}_k$s with large L_2 norm replaced by a random vector after each run (step 2d in Algorithm 1). This "soft" sparsity enhancement allows for higher flexibility of the evaluated $\boldsymbol{w}_k$s. It also allows recovery of the sources with different individual sparsities such as the sources from different focal depths shown in Fig. 2.16d. This "flexible" sparsity enhancement is one of the iNMF advantages when compared to the "hard" sparsity constraints used in the Hoyer's algorithm (Sect. 2.4.2).

The "good" sources, representing the individual PSFs, can be identified after the termination of iNMF by analysing the resulting $\boldsymbol{W}$. In Fig. 2.16d only first eight $\boldsymbol{w}_k$s look

(a) run 1

(b) run 4

(c) run 8

(d) run 14

Figure 2.16: Illustration of the iterative restart procedure for simulated of $K_{true} = 8$ sources (see Fig. 2.35) evaluated for $K = 15$. Estimated sources after (a) 1, (b) 4, (c) 8 and (c) 14 runs of the algorithm. While the estimated sources after the first run are spread over all the available components (a), the number meaningful sources progressively reduces during the runs of the iNMF procedure (c,d). Note that some components estimated after first run represent multiple emitters (a). After fourteen runs (d), first eight components correspond to true sources. The rest of the components are used for representation of noise (see Fig. 2.35 for more details). Bars below the figures show the maximum of the intensity image $\boldsymbol{w}_k$.

like the "credible" PSFs, while rest of the $\boldsymbol{w}_k$s represent noise (except for the last one, which models the homogeneous background offset).

Figure 2.19 illustrates the "robustness" of iNMF with respect to the initial number of estimated sources K. Resulting $\boldsymbol{w}_k$s of the dataset Fig. 2.35a evaluation for initial number of sources set to $K = 15$, 30 and 45 are shown in Fig. 2.19a,b and c, respectively. In all cases, the eight different PSFs shown in Fig. 2.35b were recovered, while the remaining $K - 8$ estimated $\boldsymbol{w}_k$s are representing noise (last component models the homogeneous background offset).

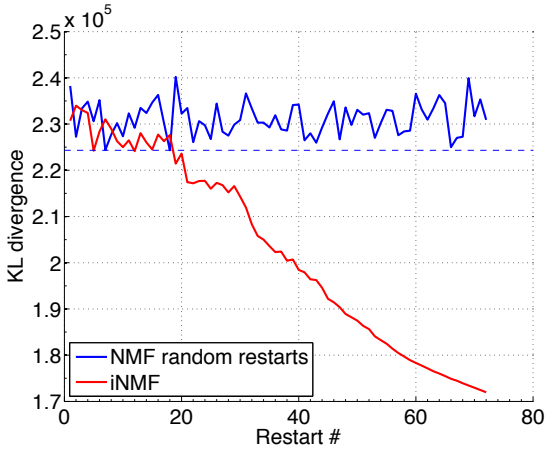
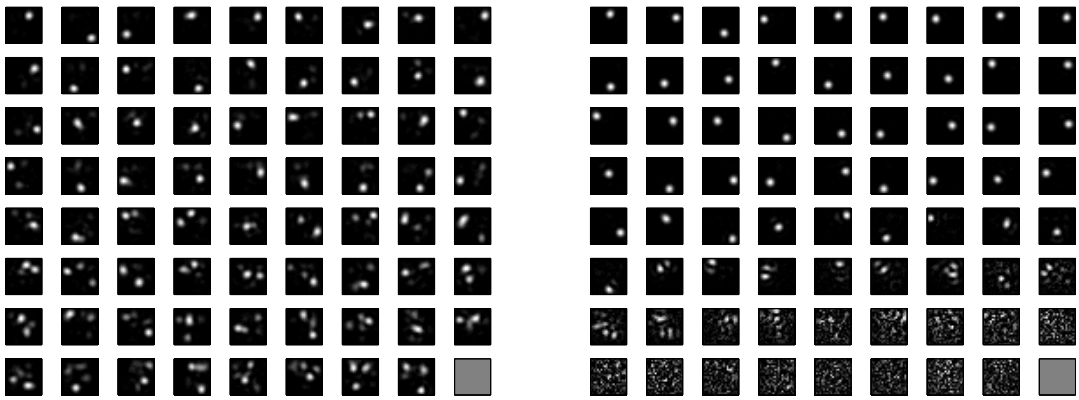


Figure 2.17: KL divergence of the NMF model Eq. (2.1) for NMF with 72 random restarts (blue line) and iNMF algorithm (red line). This is an evaluation of randomly scattered sources with density 50 sources/μm^2. The number of sources was set to the true value $K = K_{true} = 72$. The minimum of the KL divergence for NMF with random restarts is plotted as the blue dashed line. The estimated sources corresponding to the minimum of the blue and red curves are shown in Fig. 2.18a and b, respectively.

To make a fair comparison with standard NMF, we made 15 conventional NMF evaluations (for $K = 15$) of the dataset with matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ initialised with random values every time. Figure 2.20 shows the result of the evaluation with the highest likelihood Eq. (2.6) (lowest cost function Eq. (2.7)). The "credible" PSFs are distributed across all the available $\boldsymbol{w}_k$s and several $\boldsymbol{w}_k$s ($\boldsymbol{w_1}$, $\boldsymbol{w_3}$ and $\boldsymbol{w_5}$, for example) contain combination of multiple PSFs. Comparison with Fig. 2.19a demonstrates the superiority of the iNMF results.

2.7.4 Classification of the estimated sources

The estimated sources can greatly vary in quality. While some $\boldsymbol{w}_k$'s are credible representation of the PSF, there are often $\boldsymbol{w}_k$s which contain multiple PSFs or correspond to background noise. These redundant components are present due to the overestimation of K (Sect. 2.7.2). Sources located close to the patch border, and therefore partially missing, should also be identified. These sources will likely appear in the adjacent



(a) NMF with random restarts **(b)** iNMF

Figure 2.18: Estimated sources for the randomly scattered sources with density 50 sources/μm^2. The number of sources was set to the true value $K = K_{true} = 72$. The L_2 norm sorted results corresponding to the minimum of the KL divergence from the 72 NMF evaluations with random restarts (see blue dashed curve in Fig. 2.17) is shown in (a). Note that all components were used for approximation of the sources. Many components contain multiple sources. iNMF evaluation is shown in (b). Most of the components contain only a single source. Components in the last two lines were used for approximation of noise.

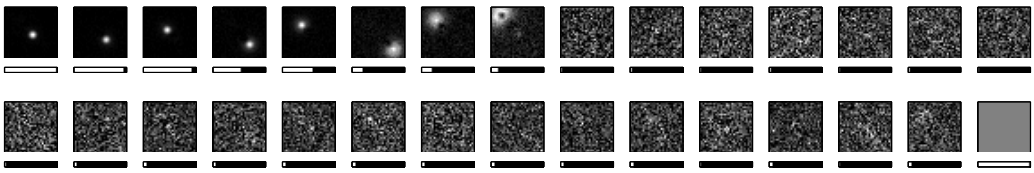
patch entirely, because the overlap of the patches is set to approximately the extent of the (in-focus) PSF (Sect. 2.7.1).

If all the sources are expected to be in-focus and therefore have a fairly compact PSF with one global maximum (left side of Fig. 2.1), we can use a simple procedure for identification of reasonable $\boldsymbol{w}_k$s:

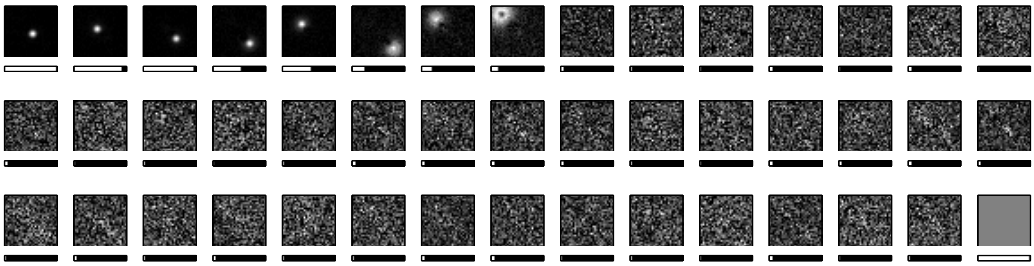
1. Each estimated source $\boldsymbol{w}_k$ is convolved with an in-focus point spread function (PSF) (generated from the parameters of the experimental setup). This is to smooth the noise in the results and to enhance the structures at the scale of PSF.

2. The number of local maxima with intensity larger than 50% of the global maximum are counted. The threshold 50% is arbitrary and reflects our empirical experience that the secondary peaks with intensity less than half of the brightest peak are not very visible in the scaled image of $\boldsymbol{w}_k$s.

(a) $K = 15$



(b) $K = 30$



(c) $K = 45$

Figure 2.19: iNMF evaluation of the simulated dataset Fig. 2.35a for different numbers of over-estimated sources K. Bars below the figures show the maximum of the intensity image $\boldsymbol{w}_k$s multiplied with the mean intensity estimated from the corresponding $\boldsymbol{h}$s.

Figure 2.20: Multiple random restarts. Bars below the figures show the maximum of the intensity image $\boldsymbol{w}_k$s multiplied with the mean intensity estimated from the corresponding $\boldsymbol{h}$s. The number of components was set to $K = 15$.

Only the sources with one major local maximum in the images of w_ks convolved with the PSF are considered for further evaluation. The distance of the maximum from the edge can indicate a partially missing source.

The process is illustrated in Fig. 2.21 on $\boldsymbol{w}_k$ estimated from the simulated dataset of 72 randomly scattered sources with density $50\,\mu\text{m}^{-2}$ (Fig. 2.6c). The $\boldsymbol{w}_k$s considered as "credible" are indicated with blue or green frame. The blue frame shows the sources

with maximum closer than two pixels from the border. The red frame shows the $\boldsymbol{w}_k$s with two local maxima of similar strength (at least 50% of the strength of the stronger maximum).

This approach would, however, fail if used on data with out-of-focus PSF, because the images of the out-of-focus PSF do not have one compact global maximum (Fig. 2.1 right). To accommodate for the individually different shapes of the PSFs we have to use a different approach.

One possibility is to compute a set of "features" on each estimated $\boldsymbol{w}_k$ (and possibly on the corresponding $\boldsymbol{h}_k$) and use a linear classifier to identify the class of each estimated source. The possible "features" can include, for example, the L_2 norm, the number of

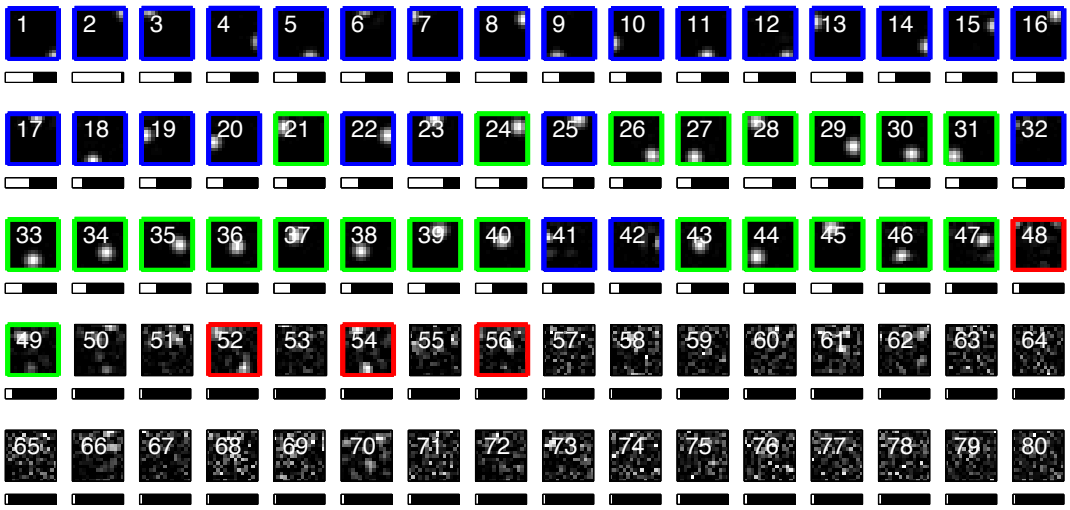


Figure 2.21: Selection of the credible $\boldsymbol{w}_k$s (here ordered by L_2 norm). The green and blue boxes indicate the estimated "credible" sources (with only one major global maximum). The sources with blue frame have the maximum closer than two pixels to the border and can be therefore considered as partly missing sources. The sources with red frame have two local maxima of comparable strength. Bars under the figures show the normalised maximum value of the estimated $\boldsymbol{w}_k$ multiplied with mean brightness of the source estimated from the intensity matrix $\boldsymbol{H}$. The index of the component k is printed in each frame. The true number of sources in this simulated dataset was $K_{true} = 72$.

.

clusters in the thresholded image, the maximum of the cross-correlation with the PSF, a measure of smoothness of the estimated result, the distance of the global maximum from the edge, and many others. Each $\boldsymbol{w}_k$ then represents a point in a high-dimensional feature space. The linear classifier assumes that the individual classes can be separated with linear manifolds in the feature space.

However, the linear classifier has to be trained on a set of labelled data. The training therefore requires a manual labelling of at least several hundreds of $\boldsymbol{w}_k$s (manual assignment of a class to each $\boldsymbol{w}_k$). In an ideal world, one training set would be sufficient for different datasets. The classifier, once trained, would be applicable for results from different datasets taken in a range of experimental conditions. However, our experience is that the classification performance varies significantly with the change of the experimental parameters (size of the patch, pixel-size, background levels). The performance, of course, depends on the quality of the features. The development of some "universal" features might be a topic of future work.

For simulated data we defined a "credibility index" as a maximum of the dot product between the estimated source and all the true sources normalised to the L_2 norm of the true and estimated sources:

$$c_k = \max_l \left[\frac{\boldsymbol{w}_k \cdot \boldsymbol{w}_l^{true}}{\sqrt{\boldsymbol{w}_l^{true} \cdot \boldsymbol{w}_l^{true}} \sqrt{\boldsymbol{w}_k \cdot \boldsymbol{w}_k}} \right] \tag{2.21}$$

For an ideal estimated source ($\boldsymbol{w}_k = \boldsymbol{w}_l^{true}$) the credibility index is equal to one. In Fig. 2.22 we show the correlation between the L_2 norm and the "credibility index" for sources from Fig. 2.21. The sources with high "credibility index" have high L_2 norm and L_2 norm can be used as a simple measure of "credibility". For example in Fig. 2.22 the "good" sources are above the threshold $L_2 > 0.02$. The scattered points in Fig. 2.22a are colour-coded in the same manner as in Fig. 2.21. We can observe that the "good" sources (blue and green points) cluster in upper right area of the graph. The sources with two local maxima of comparable strength (red) cluster together with noise contributions (black). Figure 2.22b shows the same scatter plot but points are coloured according to the index k of the components $\boldsymbol{w}_k$ shown in Fig. 2.21. The sources from

Fig. 2.21 show the "credibility index" close to one up to approximately $k = 45$.

2.7.5 Localisation and stitching

The individual estimated sources classified as credible representations of the PSFs can be localised. Conventional LM techniques often apply the maximum likelihood fitting of an in-focus PSF (usually the Gaussian approximation) to the estimated images [22]. The localisation precision is typically estimated from the number of photons emitted by the sources in the frames where the sources were localised. In constrast, the iNMF estimated intensity matrix $\boldsymbol{H}$ gives us access to the entire intensity profile of the source. We can therefore estimate the number of all photons emitted by the source during the measurement, maximum intensity of each source or a variance of the blinking over time.

The sources close to the edge can be problematic to localise. If the source represents an

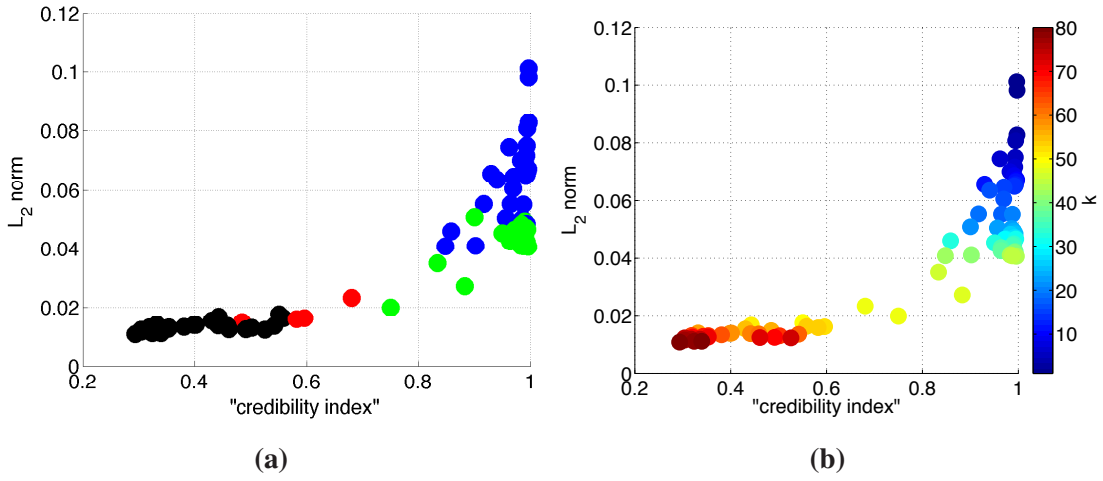
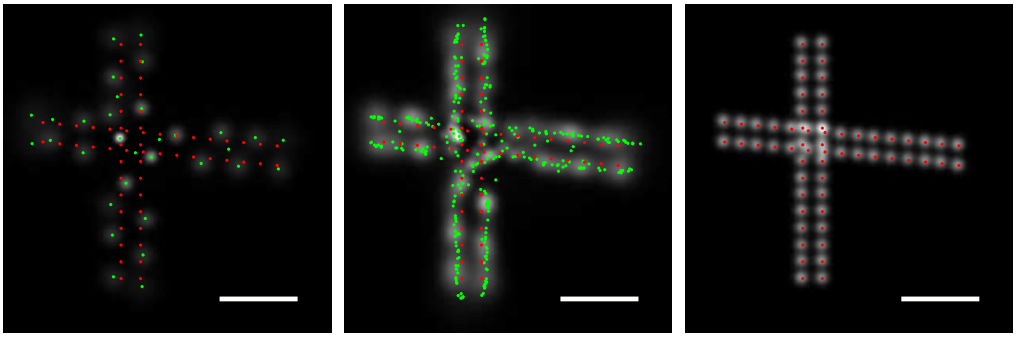


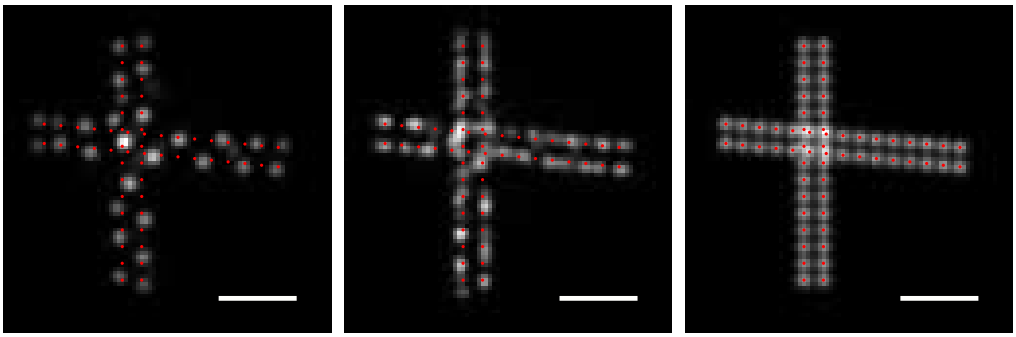
Figure 2.22: L_2 norm as a function of the "credibility index" Eq. (2.21) for sources from Fig. 2.21. (a) Colours of the points correspond to the colours of the frames in Fig. 2.21: the green and blue indicate the estimated "credible" sources (with only one major global maximum). The blue points correspond to sources considered as partly missing. The red points correspond to sources with two local maxima of comparable strength. The black points corresponds to "noise" contribution. (b) Colours of the points codes for the index k of the components in in Fig. 2.21.

in-focus PSF, then it should appear entirely in the adjacent patch and can be localised there. Therefore when dealing with images with mostly in-focus PSFs we can simply discard the sources classified as "partly missing" (Sect. 2.7.4). More problematic are the out-of-focus PSFs with extent larger than the overlap area. These sources have to be first stitched together before further processing.

2.7.6 Visualisation of the results



(a) Gaussian, $\sigma^2 \propto 1/\text{intensity}$ **(b)** Gaussian, $\sigma^2 \propto 1/\text{intensity}$ **(c)** True - Gaussian



(d) Powers, $p = 30$ **(e)** Powers, $p = 30$ **(f)** True - Powers

Figure 2.23: Visualisation of the results. (a,d) shows the results of one iNMF evaluation. (b,e) shows the sum of ten iNMF evaluations of the same dataset. (a,b) The conventional visualisation by placing Gaussians located at the positions of the estimated sources (green dots). (d,e) Powers of $\boldsymbol{w}_k$s. The true sources' locations are indicated with red dots. The corresponding images created from the true locations are shown in (c,f). Scale bar $400\,\text{nm}$.

The conventional way for visualisation of the LM results (STORM, PALM) is to sum

Gaussian functions placed in the estimated locations. The variance σ^2 of each Gaussian reflects the "uncertainty" of the estimated position. This is usually set to be proportional to the inverse of number of photons N emitted by the source. The motivation behind this is the Cramér – Rao (CR) lower bound on the localisation accuracy (see Chapter 3 for details)

$$\sigma_{CR}^2 \approx \sigma_{Airy}^2/N, \tag{2.22}$$

where the σ_{Airy}^2 is the variance of the PSF Gaussian approximation. As the σ_{CR} is typically considerably smaller than the resolution limit, the rendered image can provide super-resolution information about the specimen's structure.

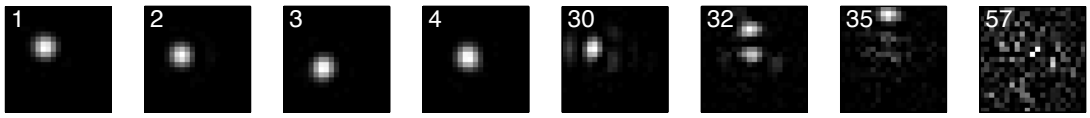
In terms of the iNMF procedure, this method replaces the credible estimated $\boldsymbol{w}_k$s with ideal, sub-resolution PSFs centred at the estimated source's location. The intensity values for each source can be estimated from the intensity time profiles of each source (rows of $\boldsymbol{H}$).

The conventional visualisation of the iNMF evaluation of the synthetic dataset (illustrated in Fig. 2.8) is shown in Fig. 2.23a,b. The data represents an artificial structure (a hash symbol with $\mu = 12.5\,\mu\text{m}^{-1}$, $d = 100\,\text{nm}$, see Sect. 2.5.2). The standard deviation of each Gaussian was set to $\sigma = 20\sigma_{CR}$. Fig. 2.23adisplays the result of one evaluation, while Fig. 2.23bshows the sum of ten evaluations (discussed below) of the same dataset.

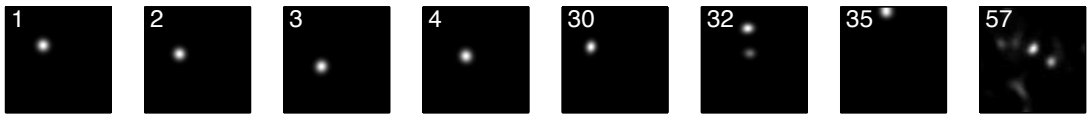
Another way to visualise the result is to use the estimated sources $\boldsymbol{w}_k$s directly without replacing them with "ideal" PSFs. By taking the pixel-wise power $p > 1$ of the estimated sources $\boldsymbol{w}^p$ we achieve "shrinking" of the individual $\boldsymbol{w}_k$ while keeping some characteristics of each source's shape (elongation along a certain direction, for example).

Up-sampling of $\boldsymbol{w}_k$s is needed before taking the higher powers p. We used bicubic interpolation of the $\boldsymbol{w}_k$'s image. Figure 2.24 shows the original estimated $\boldsymbol{w}_k$ and the corresponding up-sampled (by a factor of $r = 4$) version taken to the power $p = 5$: $\boldsymbol{w}_k^{\boldsymbol{p}}$.

This approach allows taking into account even the $\boldsymbol{w}_k$s containing multiple sources



(a) Estimated $\boldsymbol{w}_k$s.



(b) $\boldsymbol{w}_k$s up-sampled by a factor $r = 4$ and taken to the power $p = 5$

Figure 2.24: Illustration of the $\boldsymbol{w}_k$s "squeezing". Eight (out of 60) selected $\boldsymbol{w}_k$s shown. The number in the top left corner is the index k in the L_2 norm sorted $\boldsymbol{w}_k$s. (a) shows the iNMF estimated $\boldsymbol{w}_k$, (b) is the "squeezed" version $\boldsymbol{w}_k^p$ by taking the up-sampled ($r = 4$) results (a) to the power $p = 5$.

($\boldsymbol{w}_{32}$ in Fig. 2.24, for example).

If we normalise the L_1 norm of $\boldsymbol{w}^p$ to one ($\sum_x \boldsymbol{w}^p(x) = 1$), we can reconstruct a "super-resolution" image by summing all $\boldsymbol{w}_k^p$, weighted by the corresponding mean intensity mean($\boldsymbol{h_k}$).

As we show in Fig. 2.23d, the visualisation of a single iNMF evaluation can lead to a rather discontinuous image of the underlying structure. This is often the case for structures with high density of sources because only a subset of the sources is recovered. An average of multiple iNMF runs with different random initialisation is required to give smoother representation of the structure, see Fig. 2.23d. The number of iNMF runs has to be set by user and will depend on desired "smoothness" of the reconstructed images. The denser labelling requires more iNMF runs (see Sect. 2.9 for further discussion). Folowing the discussion in Sect. 2.7.4, we can also compute the "credibility index" Eq. (2.21) for the reconstructed images rather then for the individual sources by replacing $\boldsymbol{w}_k$ and $\boldsymbol{w}^{true}$ by the reconstructed image $\boldsymbol{I}$ and the "true" image $\boldsymbol{I}^{true}$, respectively:

$$c = \frac{\boldsymbol{I} \cdot \boldsymbol{I}^{true}}{\sqrt{\boldsymbol{I}^{true} \cdot \boldsymbol{I}^{true}}\sqrt{\boldsymbol{I} \cdot \boldsymbol{I}}}. \tag{2.23}$$

$\boldsymbol{I}$ are the vectors created by concatenating the columns of the image into a vector.

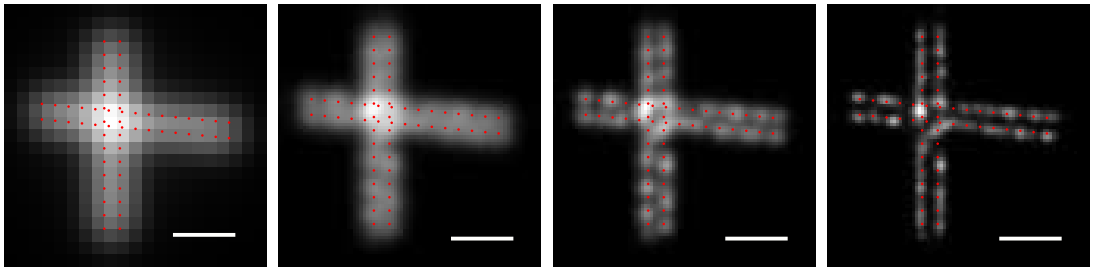
The true images computed from the true positions and the true intensities are shown in Fig. 2.23c and Fig. 2.23f. The computed indices for results from Fig. 2.23 are shown in

	one iNMF evaluation	mean of ten iNMF evaluations
Gaussian (Fig. 2.23a-c)	0.56	0.73
Powers (Fig. 2.23d-f)	0.65	0.87

Table 2.2: "Credibility index" Eq. (2.23) computed for the results shown in Fig. 2.23.

Tab. 2.2. The mean of ten iNMF evaluation gives $1.12\times$ and $1.34\times$ higher values of the index compared to the single evaluation for Gaussian and power - based visualisation (Fig. 2.23).

The power parameter p defines the final "resolution" in the reconstructed image. Larger values of p give "higher resolution" but the reconstructed structures are more "discontinuous". Figure 2.25 shows the sum projection of ten iNMF runs for different values of p. This image can be compared with the conventional visualisation in Fig. 2.23c,d.



(a) Wide-field **(b)** $p = 5$ **(c)** $p = 10$ **(d)** $p = 30$

Figure 2.25: (a) Sum projection of the simulated dataset. (b-d) Visualisation of the sum projections of ten different iNMF evaluations using $\boldsymbol{w}^p$ for different values of p. Images of $\boldsymbol{w}_k$s were up-sampled by a factor of $r = 4$. Scale bar $400\,\mathrm{nm}$.

2.8 Results

We used simulated data for exploring the behaviour of iNMF in different experimental regimes. In Sect. 2.8.1 and 2.8.2 we used the average precision and the estimated density as a quantitative quality assessments of the algorithm performance on simulated data.

Section 2.8.3 shows a qualitative comparison of iNMF results with two other techniques (CSSTORM and 3B analysis) dealing with overlapping sources. Average precision has been used as a quantitative measure of the performance on simulated datasets consisting of randomly scattered sources with different densities. We also used the estimated sources' density as a simple comparison criterion.

The comparison of the three techniques on a simulated dataset of an artificial sub-resolution structure is shown in Sect. 2.8.4. The results of the RL deconvolution and second order SOFI are also shown for further comparison.

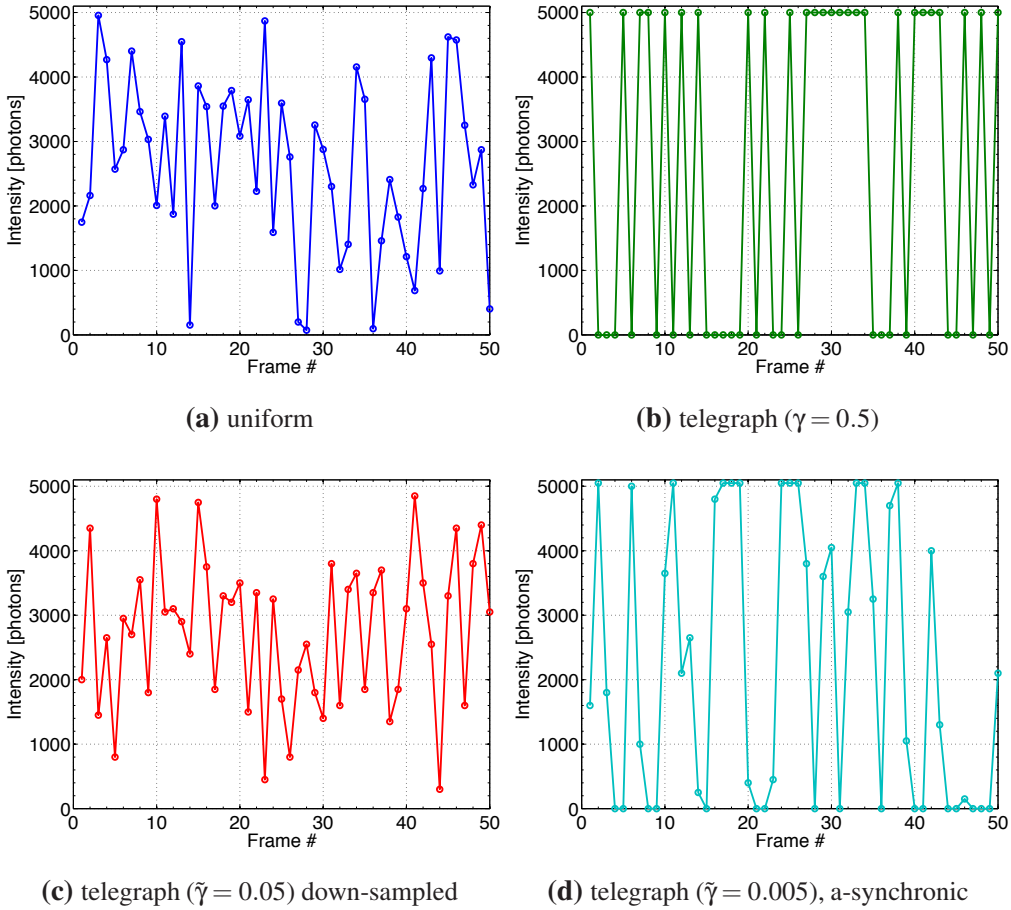
The ability of iNMF to recover different individual overlapping PSFs is illustrated on simulated data and on randomly scattered out-of-focus QDs in Sect. 2.8.7.

Section 2.8.8 shows the iNMF reconstruction of a real biological sample labelled with QDs, revealing sub-diffraction details of tubulin structures.

2.8.1 Effect of the blinking behaviour

The mechanism of the QD blinking is a complex and still not fully understood process [20]. Both ON and OFF time probability densities follow an inverse power law rather than an exponential decay observed in conventional fluorophores [19]. The blinking of different QDs can therefore vary greatly with a large range of ON and OFF time periods. NMF does not make any assumption about the intensity time profiles (rows of $\boldsymbol{H}$) and can recover a variety of blinking patterns. As we discuss later, the actual time ordering of the acquired frames is irrelevant for the iNMF evaluation.

The effect of the different blinking patterns on the performance of the iNMF algorithm



(a) uniform

(b) telegraph ($\gamma = 0.5$)

(c) telegraph ($\tilde{\gamma} = 0.05$) down-sampled

(d) telegraph ($\tilde{\gamma} = 0.005$), a-synchronic

Figure 2.26: Examples of blinking behaviour of one source (50 points out of 1000).

was tested on simulated data. We used randomly scattered sources with different densities in a range of $10-50\,\mu\text{m}^{-2}$ ($K_{true} = 14-72$). The simulated datasets are illustrated in Fig. 2.6 and the parameters of the simulations are in Tab. 2.1.

Four different blinking behaviours with fixed number of emitted photons (equal mean value) were considered:

(a) Uniform random distribution of intensities between 0 and $\max(n_{phot})$, illustrated in Fig. 2.26a.

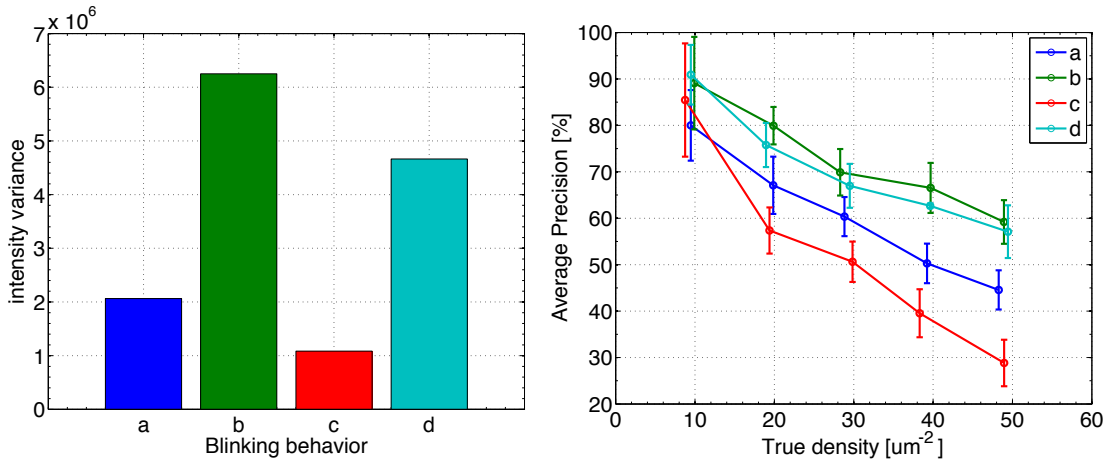
(b) A telegraph process with transition probability $\gamma = 0.5$, where the intensity is switching between 0 and $\max(n_{phot})$, shown in Fig. 2.26b.

(c) A down-sampled telegraph process. The time axis was oversampled q times and

a telegraph process with a rate $\tilde{\gamma}$ was generated. Finally, the blinking was undersampled q times. Figure 2.26c illustrates the result for $\tilde{\gamma} = 0.05$ and $q = 100$, which corresponds to $10\times$ faster blinking than the sampling frequency of the measurement. This leads to the averaging (smoothing) of the intermittent behaviour.

(d) Similar to (c), but for $\tilde{\gamma} = 0.005$. This represents more realistic intermittent behaviour than the "binary" telegraph process described in (b), keeping the same blinking rate. The switching between two states is no longer synchronised with the sampling, which gives rise to intermediate intensity values, Fig. 2.26d.

The variance of the four different intensity profiles is shown in Fig. 2.27a.



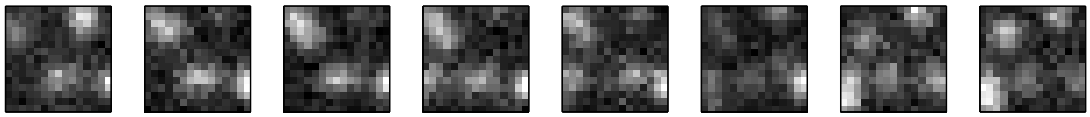
(a) Variance of blinking **(b)** Average precision

Figure 2.27: (a) Variance of the intensity time profiles for four different blinking behaviours shown in Fig. 2.26 and (b) corresponding average precision of the estimated results.

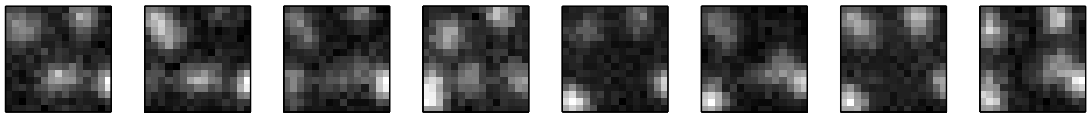
All datasets were evaluated with iNMF. The number of sources K was set to $K = K_{true} + 10$, where K_{true} is the true number of emitters used for simulation.

The average precision Eq. (2.17) for four blinking behaviours shown in Fig. 2.26 is plotted in Fig. 2.27b. The mean and the standard deviation of the results from five different geometric configurations of randomly scattered sources are shown.

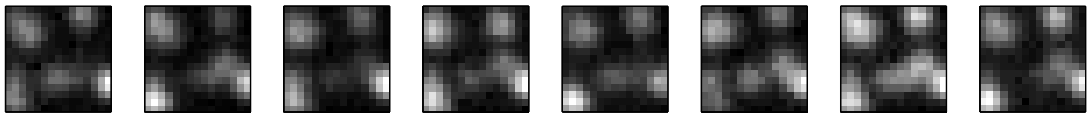
Figure 2.27 suggests that the AP is proportional to the variance of the blinking rather than to the time series structure of the blinking. In fact, NMF updates Eq. (2.3) are



(a) original data



(b) down-sampled 2×



(c) down-sampled 10×

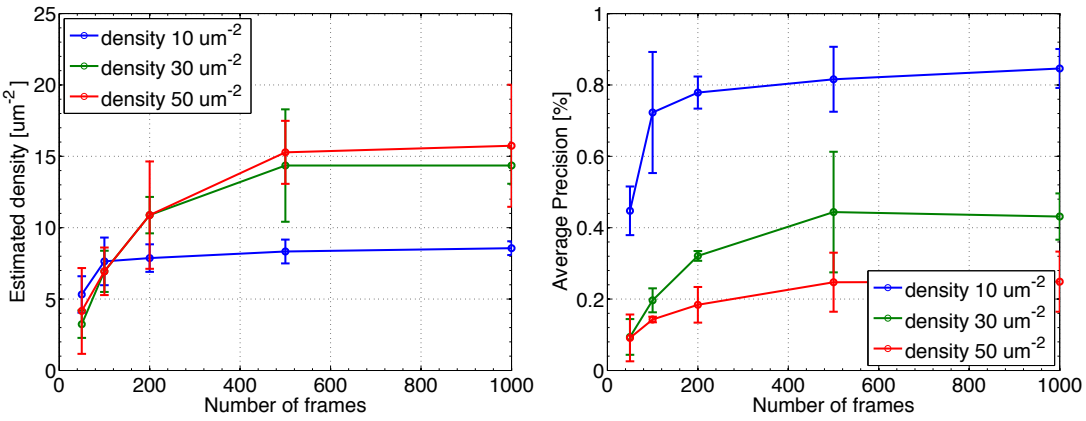
Figure 2.28: First eight frames of (a) original data and (b-c) down-sampled simulated data.

insensitive to permutations of time frames. The iNMF algorithm therefore does not take the time-series structure of the data into account. This is a drawback of the NMF model, because the correlations between the adjacent time frames provide valuable information. Note that the 3B algorithm [37] exploits this information by modelling the blinking behaviour of the fluorophores with a Markov chain.

2.8.2 Effect of the number of frames

What is the optimal way of acquiring data when we have a limited total acquisition time? Is it better to use longer acquisition time per frame to acquire smaller dataset with better signal-to-noise ratio in each frame? Or rather to record large number of noisy frames with as fast acquisition as possible?

To address these questions we tested the iNMF algorithm on simulated data of randomly scattered sources (Sect. 2.5.1). The parameters of the simulation were taken from Tab. 2.1 but the maximum of the sources' intensity $\max(n_{phot})$ was set to 300. This represents weak sources recorded with fast acquisition time. The telegraph process shown in Fig. 2.26c was used for simulated intensity profiles.



(a) Estimated density **(b)** Average precision

Figure 2.29: (a) Estimated density and (b) average precision as a function of number of down-sampled frames. The mean and the standard deviation of three different geometrical configurations of randomly scattered sources is shown.

Several frames of the dataset are shown in Fig. 2.28a. From this dataset consisting of 1000 frames we generated four more datasets by down-sampling the data $q = 2, 5, 10$ and 20 times by summing the q subsequent frames. If we neglect the read-out noise of the camera, this corresponds to data taken with q times longer acquisition time per frame. The down-sampled data consist of $1000/q$ frames. Several frames of the down-sampled data for $q = 2$ and 10 are shown in Fig. 2.28b and c, respectively.

Three different densities of the sources were considered. Each dataset was simulated three times with different geometrical configurations of the sources. Mean values with standard deviations of the average precision and the estimated density are plotted in Fig. 2.29.

The down-sampling of the dataset decreases the performance of the iNMF algorithm. The signal-to-noise ratio increases in each frame, due to down-sampling, however, the variance of the blinking decreases because the ON and OFF states average out. The increase of the signal-to-noise ratio does not compensate for the deterioration of the AP due to decreased blinking variance (Sect. 2.8.1).

Note, that only the blue curve in Fig. 2.29a corresponding to 10 μm^{-2} reaches the true

density. Both green and red curves underestimate the true density by a factor of two and three, respectively, even for data with fastest sampling.

2.8.3 Comparison with other methods - randomly scattered sources

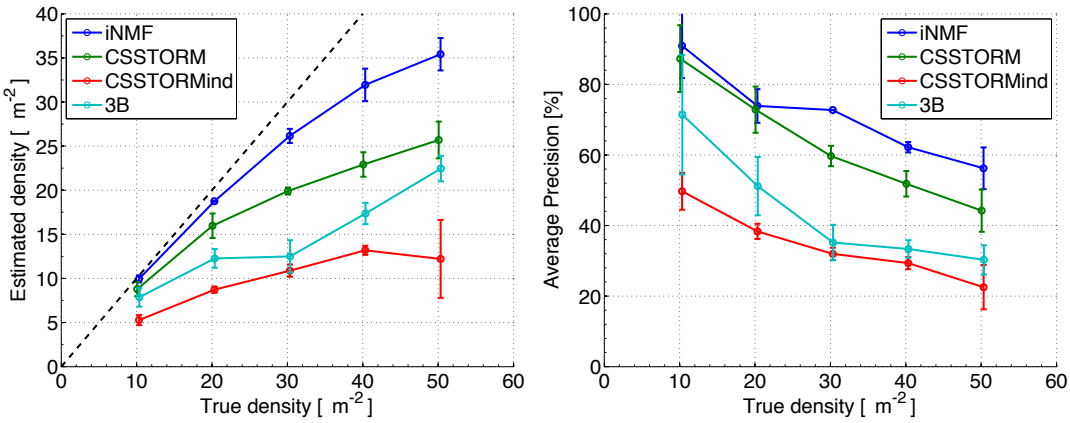
We used simulated data of randomly scattered overlapping sources with densities 10 to 50 sources/μm^2 (Sect. 2.5.1) to quantitatively compare performance of iNMF with CSSTORM [38], and the 3B analysis [37] (discussed in Sect. 2.1.2). The code for both 3B and CSSTORM is freely available.

A margin of three pixels was left empty in each simulated frame to ensure that there are no partially missing PSFs. The sum projections of the frames for densities $10\,\mu$m^2 and $40\,\mu$m^2 are shown in Fig. 2.31a and 2.32a, respectively. Several individual data frames are shown in Fig. 2.6, illustrating highly overlapping sources (dataset displayed in Fig. 2.6 does not contain the 3 pixel empty margin). We used three different geometrical configurations of the randomly scattered sources. The blinking behaviour was simulated as a telegraph process with asynchronous recording as described in Sect. 2.8.1 (d) and illustrated in Fig. 2.26d.

The true background value of 100 photons per pixel per frame was subtracted (clipping any negative values to zero) before CSSTORM and 3B evaluation. The true PSF was provided to both CSSTORM and 3B algorithms.

The estimated density and the AP values obtained from the results of the iNMF algorithm are shown as blue lines in Fig. 2.30. The mean and the standard deviation from three different configurations of the randomly scattered sources are shown. The visualisation of the results (Sect. 2.7.6, $p = 4$, $q = 4$) for datasets with four different densities is shown as a grey-scale image in Fig. 2.31b-2.32b. The green crosses show the maximum likelihood fit of a Gaussian function to the iNMF estimated $\boldsymbol{w}_k$. Only "credible" $\boldsymbol{w}_k$s were used. The selection of the "credible" sources is described in Sect. 2.7.4.

CSSTORM processes each input frame individually, independent of the rest of the dataset. This method tries to recover a sparse distribution of the active (ON) fluoro-



(a) density **(b)** AP

Figure 2.30: Comparison of the iNMF, CSSTORM and 3B evaluation of the randomly scattered PSFs. The blue line corresponds to iNMF, green line to the projected CSSTORM image, red curve to the values estimated from the individual frames of the CSSTORM and cyan line to the 3B evaluation. (a) Estimated density, (b) average precision. The mean and the standard deviation from three different configurations of the randomly scattered sources are shown. A small random offset (± 0.5) to the horizontal values was added to each dataset for the sake of clarity.

phores in each frame considering a known PSF (shared with all sources). The output for each frame is an image showing the possible positions of the sources on a sub-pixel grid (8 times oversampled). Following [38], we estimated the position of each source as a centre of mass of the small clusters formed on a sub-pixel grid. The AP (the mean of AP from individual frames) and the estimated density (the mean of estimated densities from the individual frames) are denoted as CSSTORMind and are shown as red curves in Fig. 2.30.

We also processed the sum of all CSSTORM output frames, which summarises all the estimated sources. The summed image was filtered with Gaussian kernel ($\sigma = 1$ pixel, which corresponds to $\sigma = 10$ nm). The result for different densities of the sources is shown in Fig. 2.31c-2.32c. The local maxima stronger than 5% of the global maximum were identified (green crosses in Fig. 2.31c-2.32c). We chose the threshold of 5%

because for this value the number of local maxima roughly corresponds to the true number of sources K_{true}. The positions of the local maxima were used as the estimated sources' positions for computation of the AP. The true positives were considered for density estimation (Sect. 2.6). The results are denoted as CSSTORM and are shown as green curves in Fig. 2.30.

As the last comparison technique, we used the 3B analysis for the simulated datasets. The prior parameters for the size of the PSF were adjusted to the true values. Also the true number of sources K_{true} was used as an initial number of spots in the model. The 3B algorithm was run for at least 30 iterations. Following [37], the output coordinates of the 3B analysis were placed on a 100× oversampled grid (0.8 nm pixel-size) and convolved with a Gaussian ($\sigma = 10$ pixels, which corresponds to $\sigma = 8$ nm). The resulting image is shown as a grey-scale image in Fig. 2.31d-2.32d. Similar to analysis of the projected CSSTORM, we identified local maxima in the images (green crosses in Fig. 2.31d-2.32d). Only the maxima above a certain threshold were considered for evaluation. The threshold was set individually for each image, such that the number of local maxima roughly corresponds to the number of sources considered by the 3B analysis after the last iteration. The estimated density and the AP are denoted as 3B and are shown in Fig. 2.30 as cyan lines.

Figure 2.30 suggests that CSSTORM cannot recover enough sources in individual frames (CSSTORMind, red lines). The density is severely underestimated, which leads to many false negatives (FN) and therefore low recall values Eq. (2.14), which penalises AP. However, CSSTORM recovers some subset of the sources in each frame and the sum projection show dramatically improved AP and density estimation (green lines in Fig. 2.30).

The AP of the CSSTORM results is comparable with the AP of the iNMF algorithm Fig. 2.30b. iNMF performs slightly better at the higher densities of the sources (56% as opposed to 44% at density $50\,\mu\text{m}^{-2}$). However, visual inspection of the results shown in Fig. 2.31b,c reveals that iNMF can discriminate even very close sources,

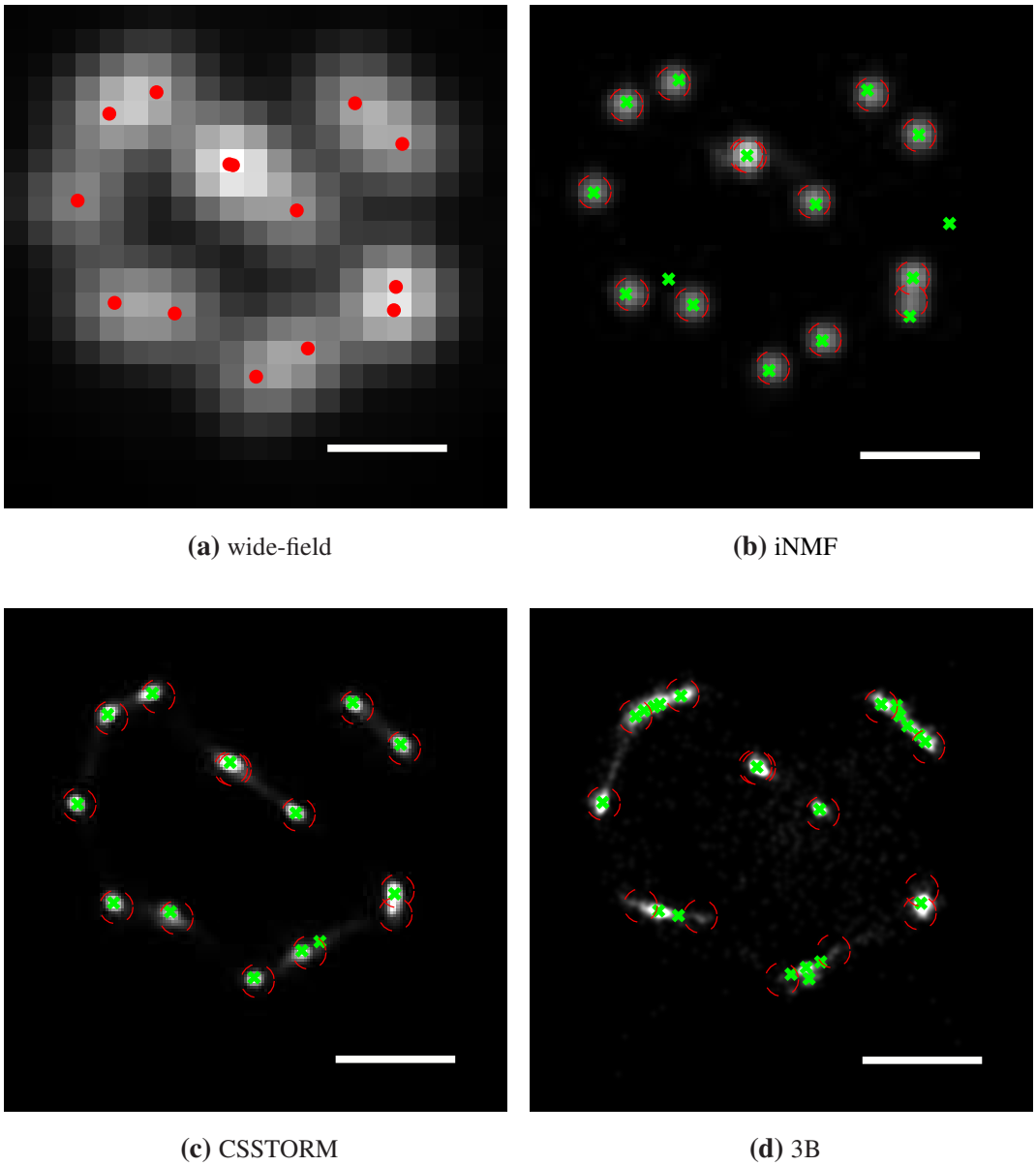
while CSSTORM approximates these sources by one intensity maximum in the middle (bottom right corner in Fig. 2.31b,c, for example). This is even more pronounced in the regions with higher densities. For example, in the bottom part of Fig. 2.32b,c the sources organised in approximately parallel lines are represented by one intensity "crest" in the middle in the CSSTORM image Fig. 2.31c, whereas iNMF managed to pick almost all the individual sources Fig. 2.31b.

For this simulated data 3B performs significantly worse than both iNMF and CSSTORM. The visualisation of the 3B results, shown in Fig. 2.31d, underlines the poor performance of 3B for this simulated dataset.

It should be noted that the estimated density and AP from the sum projection for CSSTORM and 3B results are dependent on the threshold for considering local maxima (see above). There a number of possibilities how to choose the threshold value, but we tried to relate the number of local maxima to quantities that are possible to interpret in terms of each algorithm. For CSSTORM we chose the threshold to obtain the number of local maxima approximately equivalent to K_{true}. For 3B we matched the number of local maxima to the numbers of sources considered by 3B algorithm after last iteration. In our opinion, these threshold settings can be used for a fair comparison with iNMF.

2.8.4 Comparison with other methods - artificial structure

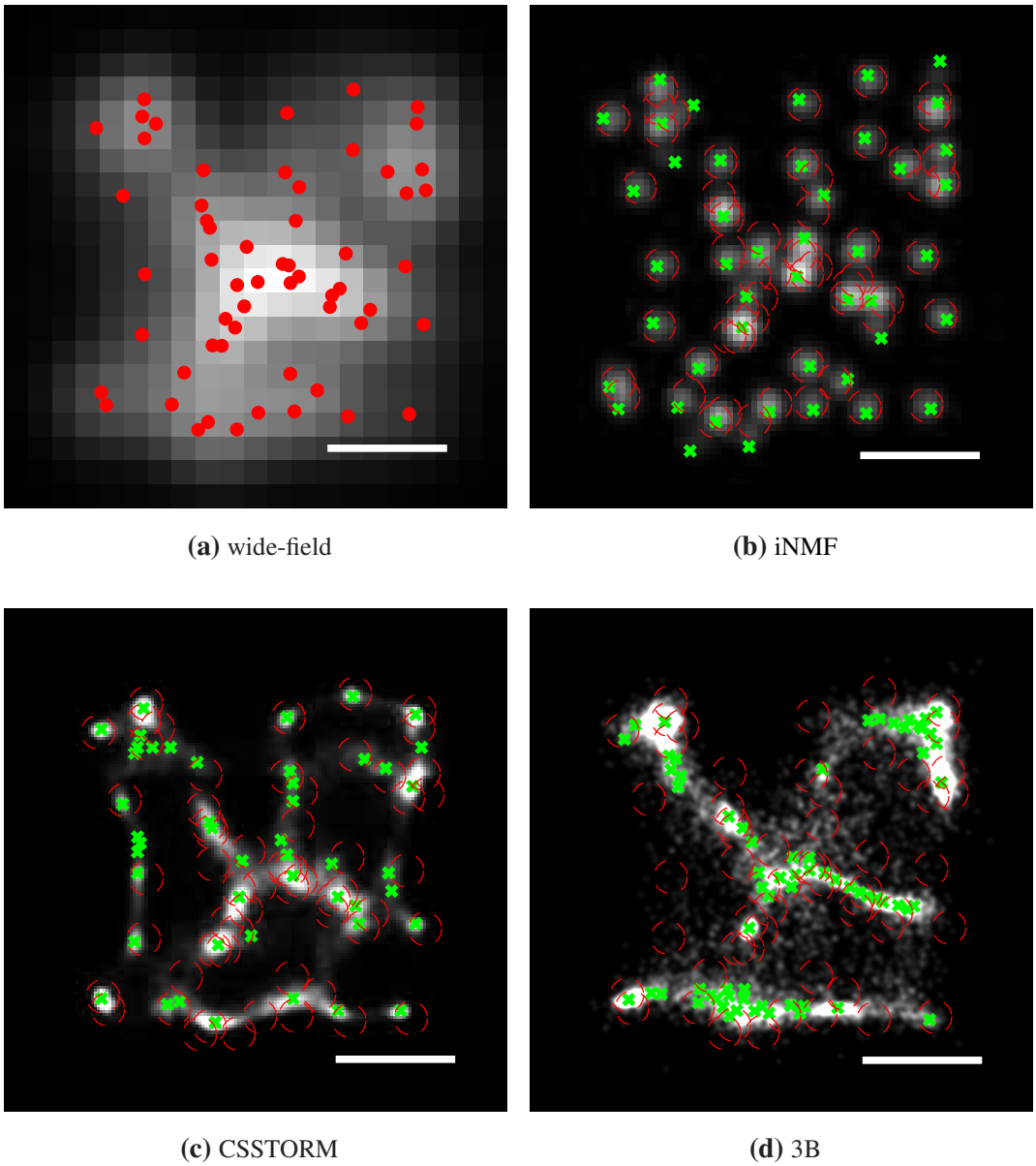
The experiments shown above are useful for a quantitative comparison of the different methods. The criterion for the quality assessment of the results was the ability to precisely estimate locations of the individual emitters. Randomly scattered sources are, however, of little practical interest. The main motivation of super-resolution microscopy is to recover sub-diffraction details of a sample structure. Therefore we used simulated data with sources attached to an artificial structure to further compare the performance of the three methods. The simulated data are illustrated in Sect. 2.5.2 with main parameters shown in Tab. 2.1. The distance between the parallel lines was set to $d = 150\,\text{nm}$ ($1.8\,\text{pixels}$), which corresponds to half the Airy disk's radius (see



(a) wide-field **(b)** iNMF

(c) CSSTORM **(d)** 3B

Figure 2.31: Comparison of the results for simulated of randomly scattered sources with density $10\,\mu\mathrm{m}^{-2}$ (14 sources in total). Sum projection of the dataset with true sources' positions marked with red dots is shown in (a). Red circles show the true locations of the sources. The radius of the circles $r = 0.7\,\mathrm{pixels}$ (56 nm) indicates the true-positive threshold distance. For further information see Sect. 2.6. Scale bar 400 nm.

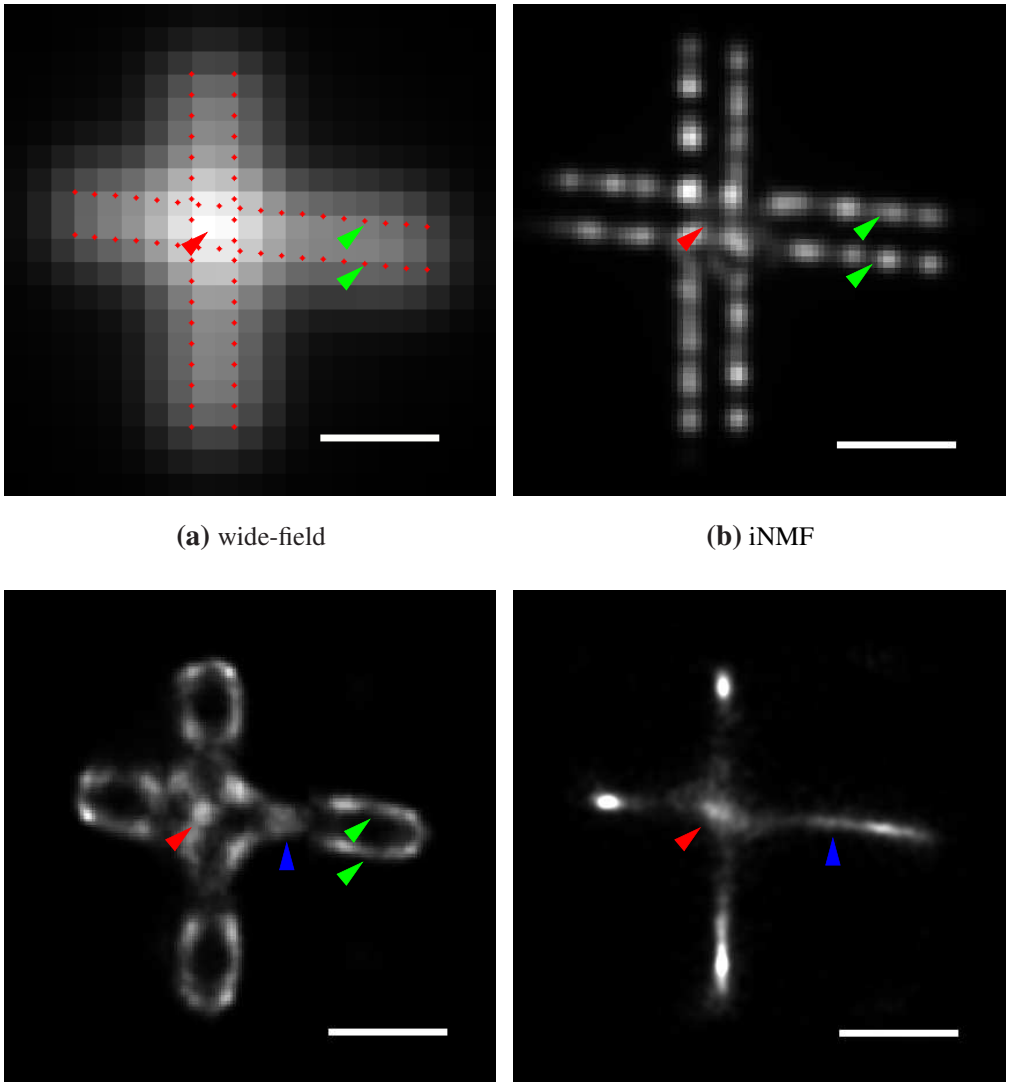


(a) wide-field **(b)** iNMF

(c) CSSTORM **(d)** 3B

Figure 2.32: Comparison of the results for simulated of randomly scattered sources with density $40\,\mu\mathrm{m}^{-2}$ (58 sources in total). Sum projection of the dataset with true sources' positions marked with red dots is shown in (a). Red circles show the true locations of the sources. The radius of the circles $r = 0.7\,\mathrm{pixels}$ (56 nm) indicates the true-positive threshold distance. For further information see Sect. 2.6. Scale bar 400 nm.

Tab. 2.1). The linear density was set to $\mu = 15\,\mu\text{m}^{-1}$ which corresponds to approximately 67 nm spacing between adjacent sources. We used the same blinking behaviour (Fig. 2.26d) as for the randomly scattered sources describe above. The sum projection of the data frames, equivalent to the wide-field image is shown in Fig. 2.33a.



(a) wide-field **(b)** iNMF

(c) CSSTORM **(d)** 3B

Figure 2.33: Evaluation of the artificial structure data with three different methods. The parallel lines are separated by $d = 150\,\text{nm}$ $(1.8\,\text{pixels})$. The sources (indicated as red dots in (a)) are distributed along the lines with a linear density $15\,\mu\text{m}^{-1}$. Arrows in a wide-field image (a) point at sub-resolution features of the specimen (further discussed in the text). Scale bar 400 nm.

To achieve a smoother representation of the underlying structure, we used the sum of

ten iNMF evaluations (see discussion in Sect. 2.7.6) rather than only one iNMF run used for the data of randomly scattered sources in the section above. The result is visualised in Fig. 2.33b. The same dataset was evaluated with 3B and CSSTORM, shown in Fig. 2.33c and Fig. 2.33d, respectively. Only one run of 3B (24 iterations) and CSSTORM has been used.

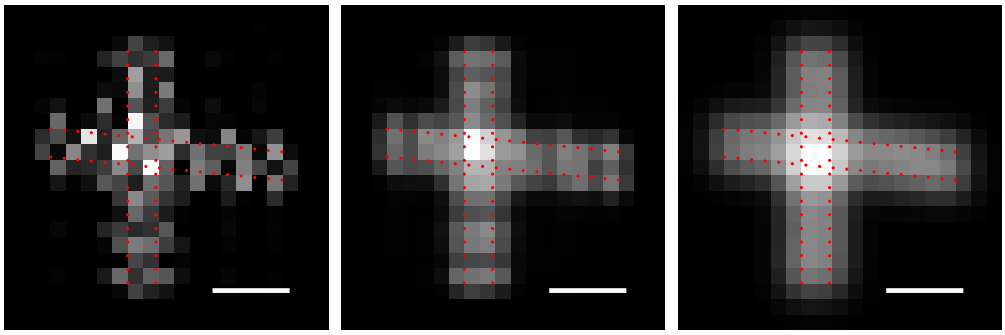
3B completely fails to recover the double parallel lines, replacing them with one intensity crest in between of the lines (blue arrow in Fig. 2.33d). CSSTORM shows the double line structure of the hash symbol at the periphery of the specimen (green arrows in Fig. 2.33c), however the double lines joins into a single line close to the centre of the cross (blue arrow in Fig. 2.33c). The hole (150×150 nm) in the middle of the specimen is completely unresolved and is replaced by intensity maximum (red arrow in Fig. 2.33c).

iNMF shows the double line structure all the way along the artificial specimen (green arrows in Fig. 2.33b) and the hole in the middle of the structure is clearly visible (red arrow in Fig. 2.33b). Figure 2.25d demonstrates that the double lines and the hole in the middle can be observed even for a structure with lines as close as $d = 100$ nm (1.25 pixels).

The results of one evaluation of iNMF did not provide satisfactory representation of the structure. As shown in an image constructed from the powers of w (Fig. 2.23c), only several individual PSFs are recovered from the highly overlapping sources (the adjacent sources are closer than $\sim \lambda_{em}/10$). The visualised image consists of disconnected individual blobs. Several runs of iNMF for the same dataset are needed to gradually fill the disconnected structure (see Fig. 2.23d and 2.33b).

The dataset was also evaluated with Richardson – Lucy (RL) deconvolution (`deconvlucy` function in MATLAB) with provided known (true) PSF and run for 1000 iterations. The true background offset of 100 photons was subtracted (clipping the negative values to zeros) before the evaluation.

RL deconvolution of the dataset sum projection (wide-field image) is shown in Fig. 2.34a.



(a) RL **(b)** RL ind **(c)** SOFI 2nd order

Figure 2.34: (a) RL deconvolution of the artificial structure sum projection. (b) Sum of the RL deconvolutions of the individual frames. (c) Second order SOFI image. Red dots show the locations of the sources. Scale bar 400 nm.

We also applied RL deconvolution to each frame of the dataset. The sum projection of deconvolved frames is shown in Fig. 2.34b.

Figure 2.34c shows the application of the second order SOFI image (see [40] and discussion in Sect. 2.1.2). The second order SOFI corresponds to the variance of the pixels intensity along the frames.

Neither RL deconvolved images nor the second order SOFI was capable of discriminating the sub-resolution features of the artificial structure (compare with Fig. 2.33).

2.8.5 Comparison with other methods - computational time

For our computer (Intel(R) Core(TM)2 Duo @ 2GHz processor with 3GB of RAM), the computational time for the simulated dataset ($21 \times 21 \times 1000$ frames) was:

iNMF ~ 20 mins for one complete run with $K = 50$ sources and K restarts.

CSSTORM ~ 260 mins.

3B analysis > 12 hours for 30 iterations.

Note that the iNMF images shown in Fig. 2.33b are results of 10 iNMF evaluation. The computation time is therefore comparable (~ 200 mins) to the CSSTORM method.

2.8.6 Comparison with other methods - parameters setting

Each of the method requires a number of parameters to be explicitly set by user prior to the evaluation. The explicit parameters are summarised in the following table (n.a. stands for not applicable):

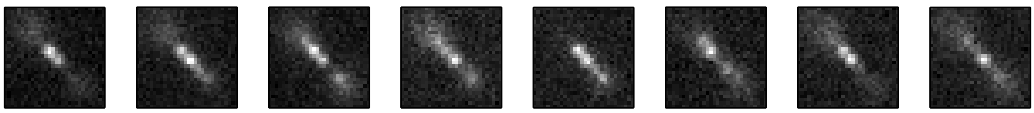
parameter	**iNMF**	**CSSTORM**	**3B analysis**
PSF description	NO	YES	YES
# of iterations in one run	YES	NO	YES
# of runs	YES	n.a.	n.a.
# of sources estimation	YES	NO	YES
patch size and overlap	YES	YES	YES

For visualisation purposes all methods require setting of the oversampling rate of the resulting images of the results in addition. There is also a parameter for a slight "blurring" of the results: the variance of the Gaussian kernel for 3B and CSSTORM (see Sect. 2.8.3) and the "power" parameter p for iNMF (see Fig. 2.25).

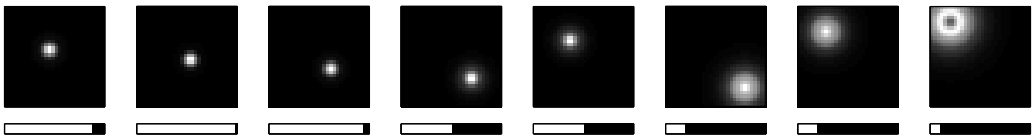
Note that these are only the parameters explicitly set by user. There are more parameters within each algorithm that are pre-set to their "optimal" values.

2.8.7 Out-of-focus PSFs

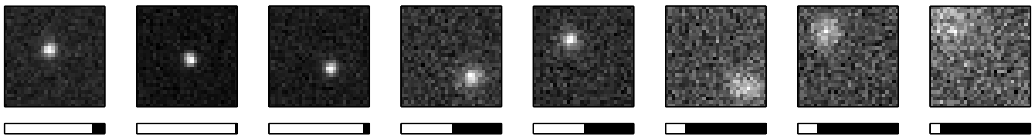
iNMF has a unique capability of recovering sources with different individual PSFs, because there is no assumption about the shape of the estimated components $\boldsymbol{w}_k$ in the NMF updates Eq. (2.3). We demonstrate this interesting feature on simulated data of eight blinking QDs attached to a bar slanting in depth, see Fig. 2.35a. The individual simulated sources were separated by 370 nm (1.15× radius of the Airy disk δ, 2.6 μm total length) in the projected plane and the axial difference between the tips of the bar was 1.6 μm. Other parameters of the simulation are shown in Tab. 2.1 with edge size of a pixel in the image plane 100 nm, $T = 500$, mean number of photons per source



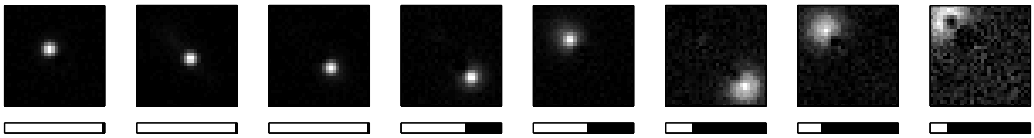
(a) Data



(b) True sources

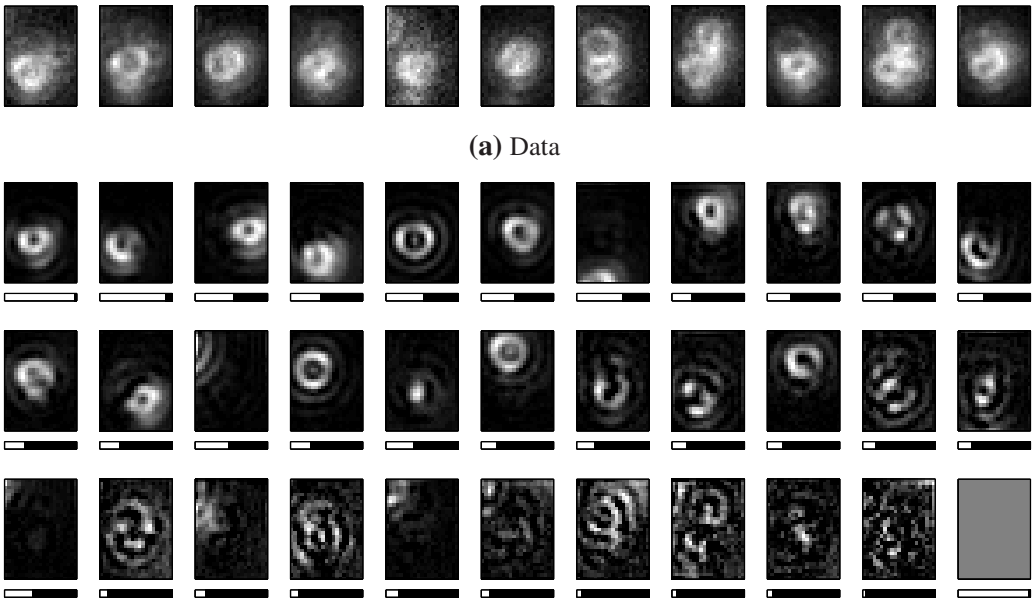


(c) True sources corrupted with noise



(d) Estimated sources

Figure 2.35: Simulated data of eight sources. (a) Eight frames (out of 500) of the simulated data set. (b) The true sources. (c) Noisy version of the true sources with their maximum intensity. (d) The first 8 estimated sources (see Fig. 2.16d for all $\boldsymbol{w}_k$'s.) Bars under the figures show the maximum of the intensity image of $\boldsymbol{w}_k$.



(a) Data

(b) Estimated sources

Figure 2.36: Real data of randomly scattered QDs. (a) Eleven randomly selected frames (out of $1,000$) of the overlapping PSFs produced by blinking QDs. (b) Estimated sources $\boldsymbol{w}_k$ sorted according to their estimated mean brightness. Bars below each figure show the maximum of the $\boldsymbol{w}_k$ multiplied by the mean brightness of the source estimated from $\boldsymbol{H}$.

per frame 1500, background photons per pixel per frame 70 and uniform distribution of blinking (Fig. 2.26a).

The true sources (individual PSFs) are shown in Fig. 2.35b and their noisy versions (obtained from the frame with the maximum intensity of each source) are shown in Fig. 2.35c. The iNMF result is shown in Fig. 2.35d (Several steps of the procedure are illustrated in Fig. 2.16). The correspondence of the estimated sources $\boldsymbol{w}_k$ (first eight out of 16 sources form Fig. 2.16) to the true sources shown in Fig. 2.35 demonstrates the ability of iNMF to recover sources with individually different shapes from noisy data with highly overlapping emitters (see Fig. 2.35).

To demonstrate the recovery of individual different PSF in realistic experimental settings, we applied iNMF on a movie of real out-of-focus blinking QDs. We analysed 1000 frames acquired with $50\,\mathrm{ms/frame}$ acquisition time (the total acquisition time

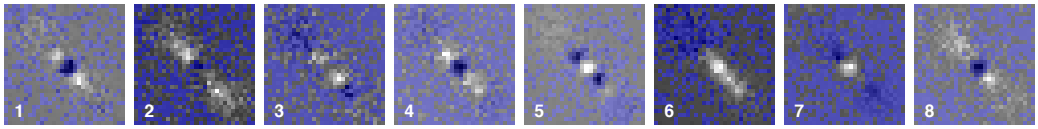


Figure 2.37: ICA evaluation of the simulated slanted bar (Fig. 2.35). Blue pixels indicate negative values.

was $\sim$ one minute). Several frames of the dataset are shown in Fig. 2.36a. The overestimated number of sources $K = 33$ was estimated from the principal values of the data as described in Sect. 2.7.2.

The images of evaluated $\boldsymbol{w}_k$s are shown in Fig. 2.36a. Credible out-of-focus PSFs from different focal depths (cf. Fig. 2.1) have been recovered (the first two rows in Fig. 2.36b). The $\boldsymbol{w}_k$s in the last row of Fig. 2.36b are mostly noise contribution. The mean brightness of these sources (estimated form $\boldsymbol{H}$) is less than 10% of the brightest $\boldsymbol{w}_k$ (see bars under individual images in Fig. 2.36b).

It should be noted that the recovery of different individual PSFs is beyond ability of either 3B or CSSTORM. Both methods require known PSF, which is shared by all emitters. 3B can adjust for the size of the PSF, however, the shape (Gaussian) remains identical for all sources.

In theory, independent component analysis (ICA), discussed in Sect. 2.4.3, allows recovery of different individual PSFs. However, as we demonstrated in Sect. 2.4.3, ICA's performance is poor when applied to noisy data. The results of the ICA evaluation (FastICA algorithm [57]) of data from Fig. 2.35 and Fig. 2.36 are shown in Fig. 2.37 and Fig. 2.38, respectively. The background was subtracted (clipping any negative values to zero) prior to the ICA evaluation. The number of sources was set to $K = K_{true} = 8$ in Fig. 2.37 and $K = 20$ in Fig. 2.38 (we set $K = 20$ because this corresponds to the number of "credible" PSF recovered with iNMF in Fig. 2.36b). We used `'tanh'` as the nonlinearity option in the fixed-point algorithm.

For the simulated data of slanted bar Fig. 2.35a, ICA completely fails to discriminate the out-of-focus overlapping sources Fig. 2.37. Only the in-focus PSFs are more or

less recovered (components 3 and 7 in Fig. 2.37). Other components represent a combination of several PSFs together. All the components contain negative values (blue pixels in Fig. 2.37).

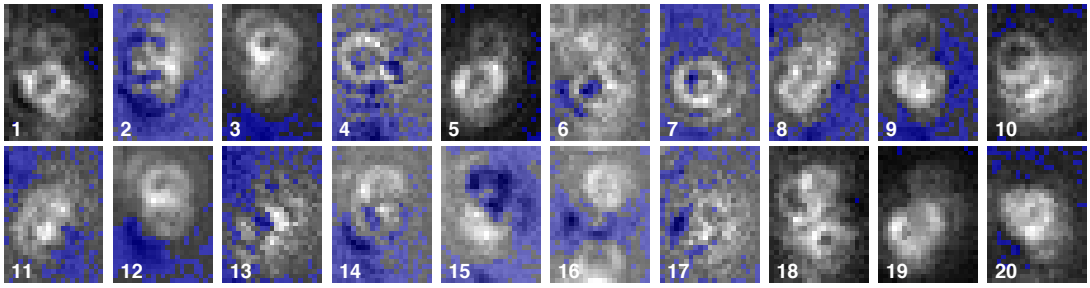


Figure 2.38: ICA evaluation of the randomly scattered out-of-focus QDs (Fig. 2.36). Blue pixels indicate the negative values.

For the real data of randomly scattered QDs Fig. 2.36a, some of the ICA estimated sources resemble the out-of-focus PSFs (for example, components 4 and 7 in Fig. 2.38), however most of the sources contain large regions of negative values (blue pixels in Fig. 2.38) and the overall quality is inferior to the iNMF results Fig. 2.36b. Most of the estimated components clearly combine several overlapping PSFs together (components 1, 9 and 12, for example).

2.8.8 Real data: QD stained tubulin fibres

We applied the pipeline described in Sect. 2.7 to a stack of $T = 10^3$ frames (128×128 pixels) of α-tubulin fibres of a 3T3 fibroblast cell imuno-labelled with QDs (QD625, *Invitrogen*). The experimental parameters are shown in Tab. 2.3.

The time average of the dataset, which corresponds to the wide-field image, is shown as a grey-valued image in Fig. 2.41a. The quantum dots are attached to the tubulin creating fine linear structures with sub-diffraction details.

The dataset was divided into 25×25 patches (Fig. 2.39), and only patches with sufficiently strong signal (thick boxes in Fig. 2.39) were considered for further evaluation. The number of sources within each patch is over-estimated via principal components

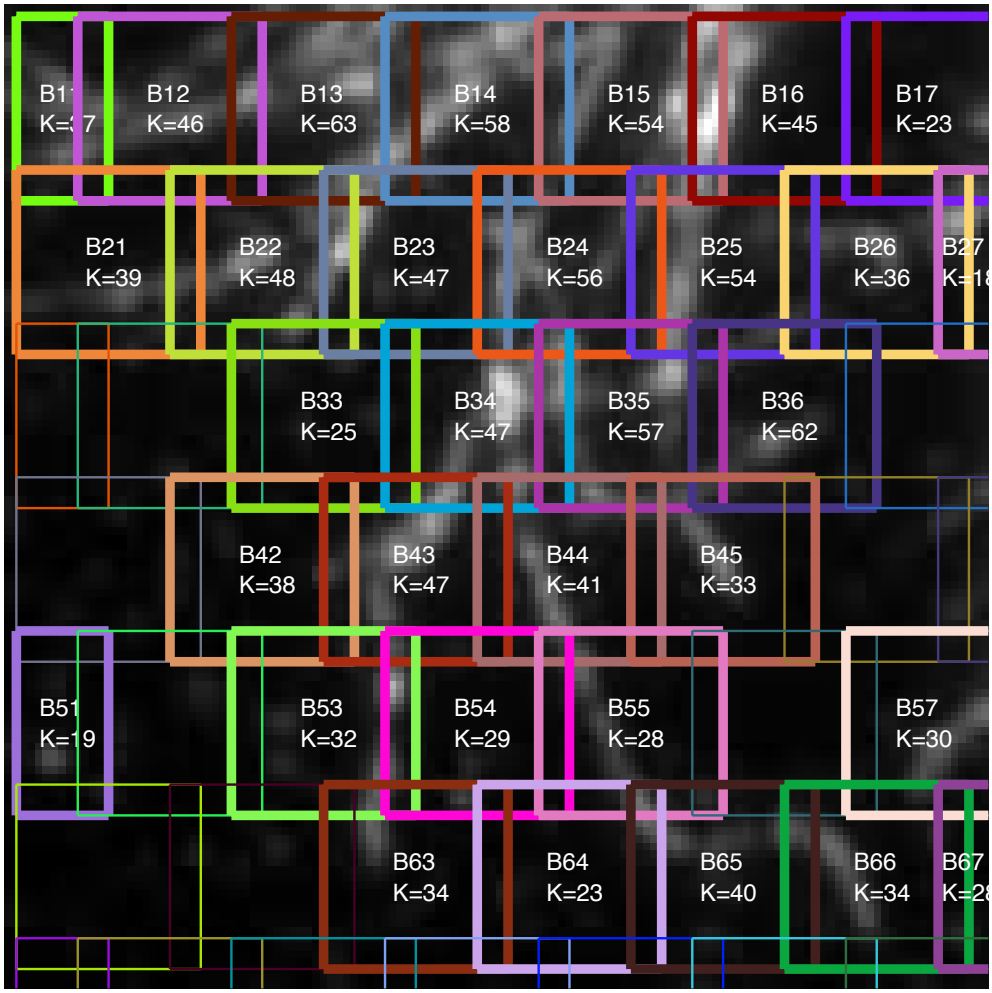
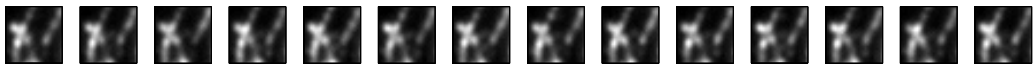


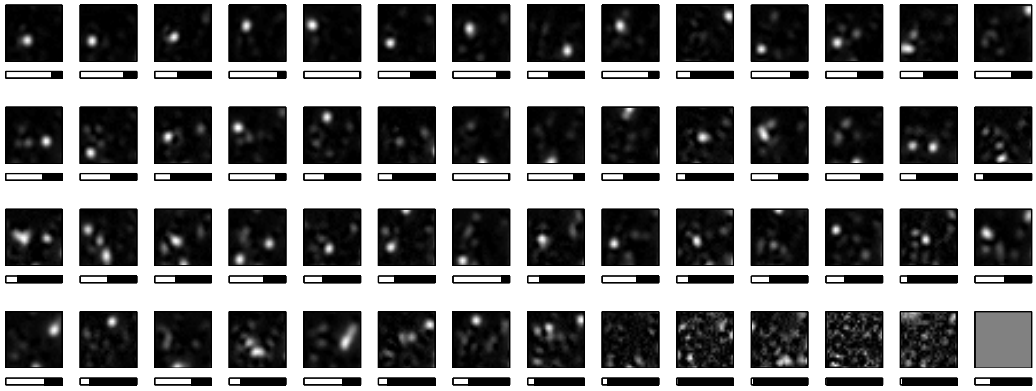
Figure 2.39: Division of the dataset into smaller patches (25×25 pixels, which corresponds to $\approx 2 \times 2\,\mu\mathrm{m}$). Time average of all frames is shown as a grey-valued image. Boxes with thick lines were used for iNMF evaluation (boxes with thin lines were considered to be empty). The index of the patches and the (over) estimated numbers of sources (K) are shown in each box.

analysis (see Sect. 2.7.2). Each patch was evaluated with the iNMF algorithm.

Several time frames of the patch B24 from Fig. 2.39 are shown in Fig. 2.40a. Figure 2.40b displays the iNMF-estimated $\boldsymbol{w}_k$s (for $K = 56$).



(a) Data

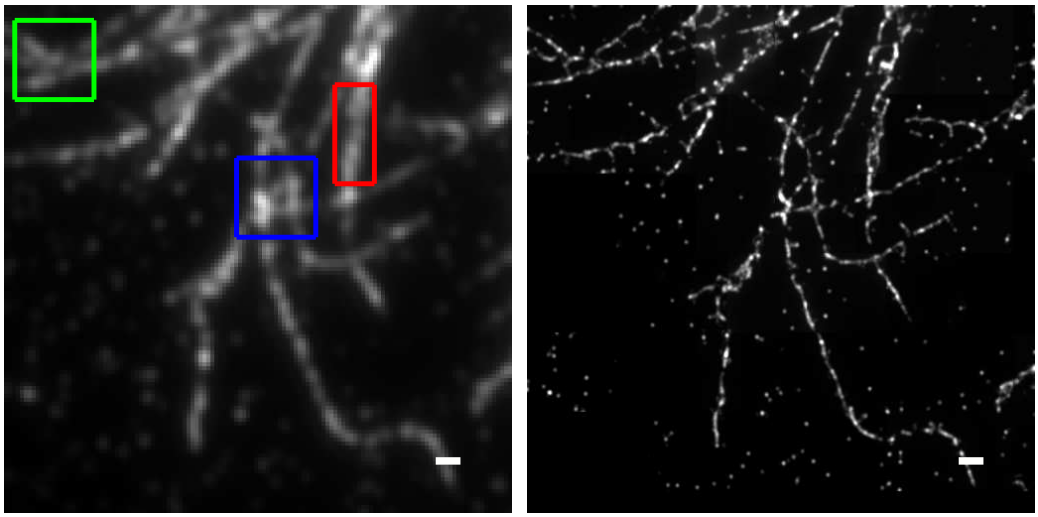


(b) Estimated sources

Figure 2.40: Real data - patch B24 from Fig. 2.39. (a) 14 randomly selected frames (out of 10^3) of the tubulin structure stained with QDs. (b) Estimated sources $\boldsymbol{w}_k$ sorted according to their L_2 norm (shown all $K = 56$ sources). Bars below each figure show the maximum of the $\boldsymbol{w}_k$ intensity image multiplied with the mean intensity of the source estimated from the matrix $\boldsymbol{H}$.

Each patch was evaluated five times with iNMF. The results were visualised by "squeezing" $\boldsymbol{w}_k$s, as described in Sect. 2.7.6, using the over-sampling by a factor $r = 4$ and power $p = 30$. The resulting images for five different iNMF evaluations were summed together to create a sub-resolution image. The final image of the whole dataset was created by tiling results for the individual patches. The border pixels of neighbouring patches were removed to avoid overlaps of the results.

Figure 2.41 compares the wide-field (WF) image with iNMF-evaluated results. The close-up of the highlighted regions in Fig. 2.41a for WF and iNMF is displayed in Fig. 2.42 (using false colours to enhance contrast of the dim features in the iNMF results). Sub-resolution details of the tubulin structure such as fibre crossing (left part



(a) wide-field **(b)** NMF

Figure 2.41: Tubulin labelled with QDs. Comparison of (a) wide-field and (b) iNMF evaluation. Details in the highlighted regions are shown in Fig. 2.42. Scale bar 500 nm.

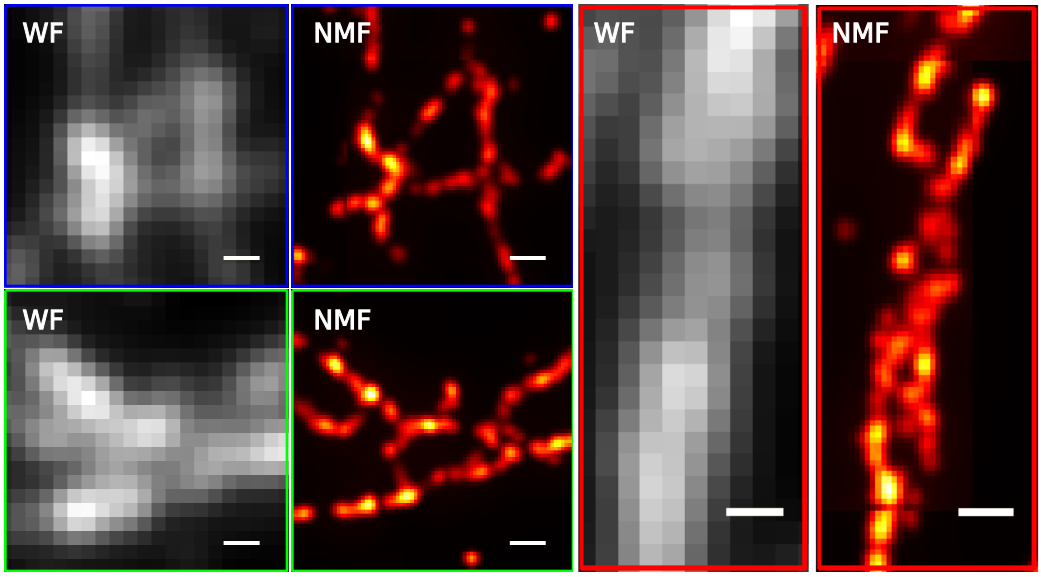
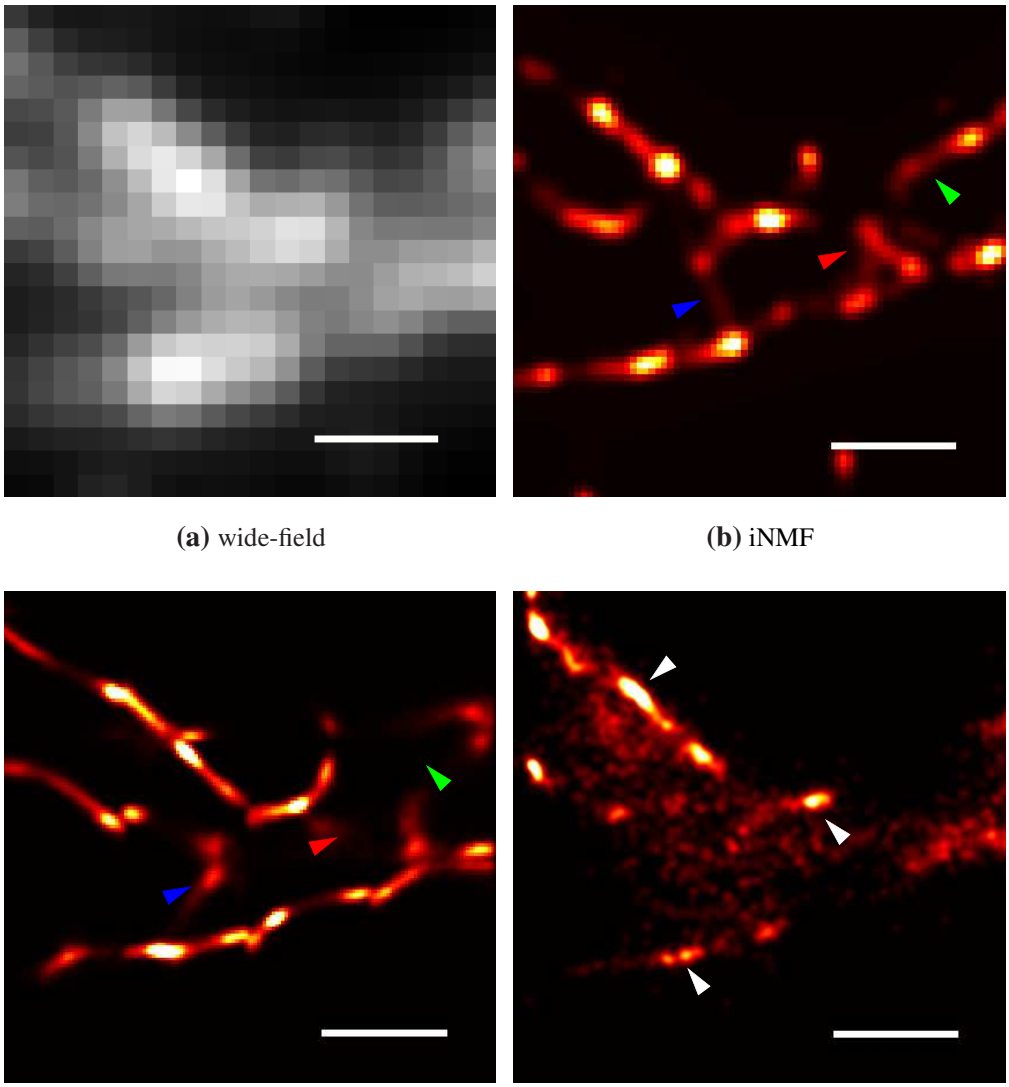


Figure 2.42: Wide field (WF) and the iNMF results for regions marked in Fig. 2.41 with coloured boxes. iNMF results are shown in false colours to enhance the contrast of dim features. Scale bars 200 nm.

of Fig. 2.42) or twisting of fibres (right part of Fig. 2.42) are revealed in visualised iNMF results.



(a) wide-field **(b)** iNMF

(c) CSSTORM **(d)** 3B

Figure 2.43: Comparison of the evaluation of a patch shown in green box in Fig. 2.41. Arrows in (b) and (c) point at differences in structure recovered with CSSTORM and iNMF, respectively. Arrows in (d) point at local maxima of the 3B results, corresponding to the bright pixels in the wide-field image show in (a). Scale bar 400 nm.

We compared performance of iNMF, CSSTORM and the 3B analysis on one patch of real data. Visualisation of results from 20×20 patch covering the green box in Fig. 2.41 are shown in Fig. 2.43. The structure revealed with iNMF and CSSTORM

(completely unresolved in wide-field image) differs in several places. The arrows in Fig. 2.43b and Fig. 2.43c point at some differences in iNMF and CSTORM recovered sample structure, respectively. However, unlike in the simulated artificial structure (Fig. 2.33), we do not have ground truth for this dataset. Quantitative comparison of the results is therefore difficult. The 3B analysis (20 iterations of the algorithm) delivered very poor results Fig. 2.43d and did not provide any further information about the sub-diffraction structure. Coloured arrows point at local maxima, which correspond to the bright pixels in wide-field image Fig. 2.43a. These are the only significant features of the 3B-recovered image.

Parameter	**Note**	**Value**
λ_{ex}	excitation light	405 nm
λ_{em}	emission light	625 nm
t_{exp}	exposure time	50 ms
NA	numerical aperture	1.4
RI	refractive index	1.52
pixel-size	size of a pixel in image plane	79 nm
QD	quantum dots	QD625
T	number of frames	10^3

Table 2.3: Parameters of the experiment.

2.9 Discussion

iNMF can recover individual highly overlapping sources of arbitrary shape. In the preliminary work [71], we demonstrated on simulated data that iNMF can separate *two* sources as close as $\sim \delta/7$ ($\sim \lambda_{em}/15 = 40\,\text{nm}$ for $\lambda_{em} = 625\,\text{nm}$, 1000 frames, mean intensity 500 photons/source/frame and background 100 photons/pixel/frame). The emitters were simulated with uniformly distributed blinking behaviour (see Fig. 2.26a),

where both sources are present in almost all frames. However, as we discuss below, the situation becomes increasingly difficult with larger number of sources in the sub-resolution area.

Figure 2.32b suggests that it is possible to identify most of the sources for densities up to $\rho = 40\,\mu\text{m}^{-2}$ (46 identified sources out of 58). Further increase of ρ leads to the significant underestimation of the density (blue lines in Fig. 2.30a). The average distance $\bar{r}$ between two nearest neighbours for randomly scattered sources with density ρ is given by [72]

$$\bar{r} = \frac{1}{2\sqrt{\rho}}. \tag{2.24}$$

The mean nearest neighbour distance corresponding to the density ρ of randomly scattered sources used for our simulations (see Sect. 2.8.3) is summarised in the following table:

ρ [μm^{-2}]	10	20	30	40	50
$\bar{r}$ [nm]	158	112	91	79	71

Table 2.4: 2D density ρ of the randomly scattered sources and the corresponding mean nearest neighbour distance $\bar{r}$.

The density $40\,\mu\text{m}^{-2}$ corresponds to ~ 10 sources within an Airy disk or in other terms to the mean distance between the nearest neighbours of 79 nm (see Tab. 2.4). However, the probability of the source to appear ON in each frame is 0.5 (see Fig. 2.26d), which on average reduces the density of the sources in each recorded frame by a factor of two. For data with simulated density of $40\,\mu\text{m}^{-2}$ the mean distance between the nearest neighbouring ON sources in each frame is therefore 112 nm, (see Tab. 2.4).

Figure 2.23c shows that iNMF cannot separate all the individual sources uniformly distributed on a line with $\sim \delta/3$ spacing (80 nm, linear density $12.5\,\mu\text{m}^{-1}$ or average spacing of ON sources 160 nm). One iNMF evaluation recovered only a subset of all sources. In some cases, one iNMF estimated source represents, in fact, several close emitters. Therefore some of the estimated locations (green dots in Fig. 2.23b) fall in

between two neighbouring emitters (red dots in Fig. 2.23b). Multiple runs of iNMF with different starting values can recover varying subsets of sources. Therefore the sum of these evaluations can provide more complete information about the structure, see Fig. 2.23b, 2.33b and 2.41.

QDs are characterised by broad absorption profiles and a narrow and a spectrally tuneable emission spectrum. A range of colours (determined by the size of QD's core) is readily available on the market. It is therefore possible to label the specimen with a mixture of QDs with a variety of colours and record the intermittent sources in several different spectral channels. This would lead to a reduction in the density of the QDs in each colour channel and facilitate the separation of the individual sources. On the other hand, such technique would have to deal with chromatic aberrations.

iNMF does not have any constraints on the shape of the estimated sources. We presented this as an advantage in Sect. 2.8.7, because such flexibility makes the recovery of different PSFs possible (Fig. 2.35 and 2.36b). However, there is a lot of information about PSF (compact and sparse object, circular symmetry), which can constrain the space of the "credible" $\boldsymbol{w}_k$s, and therefore make the recovery of the sources easier. The sparse NMF algorithm, discussed in Sect. 2.4.2, is an example of reducing the excessive degrees of freedom. As we demonstrated in Fig. 2.4, Hoyer's sparse NMF did not provide satisfactory results, though. Note that iNMF uses only a "soft" enhancement of the $\boldsymbol{w}_k$s sparsity (see discussion in Sect. 2.7.3).

Additional information about $\boldsymbol{W}$ or $\boldsymbol{H}$ (for example, upper bound on number of emitted photons) can be used in two different approaches. In the first approach, we specify the constraints on the $\boldsymbol{W}$ and (or) $\boldsymbol{H}$ and formulate NMF as an optimisation problem (minimising Eq. (2.7)) subject to these constraints. Hoyer's sparse NMF Sect. 2.4.2 is in this category.

In the other (rather heuristic) approach, we can use the standard unconstrained NMF and employ the additional information as a quality criterion for the estimated $\boldsymbol{W}$ and $\boldsymbol{H}$. Results, which do not satisfy this criterion can be recomputed. We can, for example,

randomise the "unsatisfactory" sources or split $\boldsymbol{w}_k$s with multiple local maxima into two or more individual components. iNMF belongs to the second category.

The disadvantage of the second approach is that it does not reduce the parameter space of the optimisation. Instead of reducing the number of local minima, it tries to search for the "better" ones. On the other hand, all the flexibility of NMF is maintained. It also reduces the number of parameters, which need to be set by the user.

It should be noted that NMF is applicable to any intermittent fluorescent dyes, such as blinking fluorescent dyes used in dSTORM technique [73]. Data with high densities of activated (and therefore overlapping) sources can be processed with iNMF. Such data require shorter total acquisition time compared to the conventional LM methods, where the individual emitters are separated physically by keeping the number of activated sources sufficiently small. However, iNMF is impractical for data with high bleaching rates. For example, in the standard fPALM techniques [22], each emitter is activated for one data frame (or few adjacent frames) and then irreversibly destroyed by photo-bleaching. Even though iNMF is applicable to such data, it does not make use of the method's major strength, which is the identification of the sources reappearing throughout the dataset.

Poor performance of the 3B analysis in Sect. 2.8.3 was surprising. Many of well separated sources in Fig. 2.32d were completely missed in the evaluation of simulated data. The method showed weak results even when applied to an artificial structure Fig. 2.33d or a real dataset of QD labelled tubulin Fig. 2.43d. Despite the ability to resolve structures on 50 nm scale claimed in [37], the double lines separated by 150 nm were completely unresolved (see Fig. 2.33). These data proved to be too difficult for the 3B method. The 3B method failed despite the fact that the simulated blinking of the sources was generated with a Markov process (without bleaching), one of the assumptions of the 3B model. The prior parameters (PSF, number of sources, blinking rates) of the 3B algorithm were set close to the true values. We spent a considerable amount of time to test different parameters' prior values, but could not improve the 3B

performance.

On the other hand, the performance of the CSSTORM was surprisingly high. Average precision and estimated density curves Fig. 2.30 were only marginally inferior to the iNMF results, even though we tested the method on sources densities four times higher than in the original publication [38] ($50\,\mu\text{m}^{-2}$ as opposed to $12\,\mu\text{m}^{-2}$). However, iNMF provided higher details in the recovered artificial structure Fig. 2.33. Comparison of the CSSTORM and iNMF methods on real data of QD labelled tubulin in Fig. 2.43 is difficult due the fact that the true underlying structure is unknown.

The computational time for one run of iNMF is approximately $10\times$ faster than for CSSTORM (Sect. 2.8.3). However, for the visualisation shown in Fig. 2.43 we used the mean of five evaluations of Algorithm 1 (see Sect. 2.8.4). Therefore the computational time (about 2 hours for 10^3 frames with 21×21 pixels) is comparable for both methods.

The ability to recover the individually different overlapping sources makes iNMF unique when compared with other methods. The only alternative method, ICA, proved to be unsuitable model for noisy data Fig. 2.5 and 2.37. The recovery of different shapes of the PSF can be used for the determination of the axial position of the emitters Fig. 2.1. For example, Speidel et al. [74] determines the axial position of the emitter from the diameter of the outermost ring. However, separation of the overlapping out-of-focus and in-focus PSFs might be problematic due to the large difference in brightness. Photons in the out-of-focus PSF are distributed over a much larger area making the PSF considerably dimmer. The maximum brightness of the 1 μm out-of-focus PSF is only 10% of the in-focus PSF (see Fig. 2.1). iNMF separated components $\boldsymbol{w}_6$, $\boldsymbol{w}_7$ and $\boldsymbol{w}_8$ in Fig. 2.35d are partially missing in the region of overlap with the in-focus source. This bright source took over a part of the weaker one. This effect is apparent in all evaluations shown in Fig. 2.19.

The more promising strategy for determining the axial position of the emitters might be using a tailored PSF, such as the double helix PSF [75] or the PSF with introduced astigmatism [76]. Such PSFs specifically change their shapes with the axial loca-

tion. Moreover, the PSF, however distorted, remains fairly compact over an interval of several micrometres. The difference in the brightness of the in-focus and out-of-focus PSFs is less pronounced (the in-focus PSF is less bright than the one in the system without aberrations), facilitating the separation of individual overlapping emitters. Testing iNMF on data with tailored PSF [76, 75] might be a subject for future work.

Note that iNMF assumes the sources are spatially fixed during the acquisition ($\sim$ 1 min). Significant movement of the specimen during the data acquisition would produce artefacts. iNMF can therefore be used only to live cells with movement on the scale of minutes ("slowly" moving cells). The total acquisition time can be shortened by reducing the number of frames. However, this can make recovery of individual sources from high density data difficult (Fig. 2.29). As pointed out by Shroff et al. [77] there is a trade-off between the spatial and the temporal resolution in LM.

2.10 Conclusion

In this chapter we demonstrated the non-negative matrix factorisation (NMF) as a natural model for microscopic samples labelled with quantum dots. We described a practical pipeline for the evaluation and the visualisation of realistic datasets. The individual steps of the pipeline were illustrated on simulated data.

We introduced a procedure of NMF with iterative restarts (iNMF), which leads to better local minima in the optimisation procedure and shows robustness in terms of the estimated number of sources.

We introduced average precision (AP) as a quantitative measure of the algorithm performance and used it for exploring the behaviour of iNMF in different experimental settings. We also used the AP for quantitative comparison of iNMF with CSSTORM and the 3B analysis demonstrating the superior performance of iNMF on simulated data with highly overlapping sources.

The unique ability of iNMF to recover individually different sources from data with

highly overlapping emitters was demonstrated on simulated three-dimensional object and on real data consisting of randomly scattered out-of-focus QDs.

Finally, we used iNMF for evaluation of a larger area of a biological sample with QD labelled tubulin structures. We demonstrated the ability of iNMF to show sub-resolution features in the specimen.

In conclusion, the non-negative matrix factorisation enlarges the family of localisation microscopy techniques and enables using quantum dots as fluorescent labels for localisation microscopy. It is a promising technique with the potential to deliver super-resolution images of three-dimensional samples.

Chapter 3

Theoretical Limits for LM

In this chapter, we discuss the resolution limit and its application to localisation microscopy (LM) from the theoretical point of view.

In Sect. 3.1 we compare the classical resolution limit with fundamental resolution measure (FREM), introduced by Ram et al. [78]. FREM accommodates the resolution criterion for situation of pixelated data corrupted with noise. Section 3.2 introduces the Cramér – Rao (CR) lower bound as a theoretical framework for description of the estimator covariance matrix. In Sect. 3.3 we use the CR lower bound to show the derivation of the Ram's original FREM formula. We also discuss the limitations and problems of the original FREM. In Sect. 3.4 we derive an alternative version of FREM and demonstrate that this version fixes strange and inconsistent behaviour of the original FREM. Throughout the chapter we use our version of FREM. We call the Ram's version the "original FREM" and denote it as $\text{FREM}^{\text{orig}}$. Our alternative version is denoted as $\text{FREM}^{\text{static}}$. In Sect. 3.5 we derive $\text{FREM}^{\text{blink}}$ for two emitters with intermittent intensity. This expression is relevant to localisation microscopy with blinking fluorophores, such as QDs (discussed in Chapter 2). The parameters used for data simulation are discussed in Sect. 3.6. Section 3.7 compares FREM for blinking and static sources in different experimental conditions and identifies the regions and experimental parameters setting, where the intermittent behaviour of the intensity allows

considerable higher resolution. The discussion of the results is in Sect. 3.8.

Details of the derivations can be found in Appendix B.

3.1 Fundamental resolution measure (FREM)

The classical resolution limit Eq. (1.6) discussed in Sect. 1.4 relates to an empirical observation and does not take into account the statistical nature of the photon detection process. The classical resolution limit applies to a noise-free situation and neglects the pixelation of data.

Ram et al. [78, 79] revised the resolution limit and defined a new measure, which considers the statistical process of photon detection on a pixelated grid of a camera. The so-called *Fundamental resolution measure* (FREM) refers to the achievable precision of the estimator on distance between two sources. FREM reflects the fact, that the "resolution limit" is different for sources with different noise levels. If we want to "resolve" two sources, the necessary separation must be larger for noisy data (weak emitters with high background) than for data with high signal-to-noise ratio (bright sources with low background values).

Ram et al. defined FREM as the Cramér – Rao (CR) lower bound on the standard deviation of the source separation estimator. FREM therefore does not provide "resolution criterion" such as Eq. (1.6), but gives us a notion about variability we can expect if we try to measure the distance between two emitters. We can set the "resolution" limit arbitrarily according to the measurement precision we are willing to accept. A natural choice for the "acceptable precision" is the distance between the sources. I.e. the standard deviation of the source separation measurement is equal to the separation itself. We use this "natural resolution criterion" throughout this chapter.

It is important to note that FREM defined as the CR lower bound does not consider any specific algorithm for the estimation of the source separation. FREM is derived from the generative model of the dataset. The standard deviation lower bound can be

achieved only with the "optimal" algorithm.

3.2 Cramér – Rao lower bound

Cramér – Rao lower bound is a theoretical framework for description of the estimator covariance matrix. If $\mathcal{L}(\theta) = \log p(x|\theta)$ is a log-likelihood function for data X, then a covariance matrix $\boldsymbol{Q}$ of an unbiased estimator of $\hat{\theta}$ is bounded by [80, 81]

$$\boldsymbol{Q} \geq \boldsymbol{I}^{-1}(\theta), \tag{3.1}$$

where the Fisher information matrix $\boldsymbol{I}(\theta)$ can be expressed in two equivalent formulas

$$I_{ij}(\theta) = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j}\right] = \mathbb{E}\left[\frac{\partial \mathcal{L}}{\partial\theta_i}\frac{\partial \mathcal{L}}{\partial\theta_j}\right]. \tag{3.2}$$

The inequality Eq. (3.1) is in the sense that $\boldsymbol{Q} - \boldsymbol{I}^{-1}(\theta)$ is a non-negative definite matrix.

3.3 Original FREM formula (FREM$^{\text{orig}}$)

Ram et al. [78] considered two sources separated by a distance d and derived the Fisher information

$$I(d) = \frac{1}{4}\sum_{k=1}^{N} \frac{\left[\Lambda_1 q_k'(-\frac{d}{2}) - \Lambda_2 q_k'(\frac{d}{2})\right]^2}{\Lambda_1 q_k(-\frac{d}{2}) + \Lambda_2 q_k(\frac{d}{2}) + b}, \tag{3.3}$$

where Λ_i is the intensity of the ith source, b is the background level in each pixel, $q_k(z) = \int_{\Gamma_k} q(x-z)dx$ is the pixelated version of a point spread function translated by z with Γ_k being the region of the kth pixel. The corresponding pixelated derivative is $q_k'(z) = \int_{\Gamma_k} \frac{\partial q(x-z)}{\partial x} dx$.

The inverse of the Fisher information bounds the variance of the estimator on d

$$\text{var}(d) \geq I^{-1}(d). \tag{3.4}$$

FREM$^{\text{orig}}$ as a lower bound on the standard deviation is therefore

$$\text{FREM}^{\text{orig}} = \sqrt{I^{-1}(d)}. \tag{3.5}$$

A short summary of the derivation is shown in Appendix B.

Closer inspection of FREM$^{\text{orig}}$ derived from the Fisher information given by Eq. (3.3) reveals problematic behaviour of FREM$^{\text{orig}}$ in the limits $d \to 0$ (see discussion in Appendix B). The limit of very close emitters $d \to 0$ gives, as we would expect, zero Fisher information $I(d) \to 0$, and therefore FREM$^{\text{orig}} \to \infty$. However, this is only for situation, when the sources have equal intensities $\Lambda_1 = \Lambda_2$. For emitters of unequal strength $\Lambda_1 \neq \Lambda_2$ the variance remains finite even for sources infinitely close.

Another problem with this expression is that the sources are assumed to be located at $\pm d/2$, which implies the knowledge of the origin. It is therefore not surprising that the formula Eq. (3.3) gives non-zero Fisher information $I(d) \neq 0$ (i.e. finite FREM$^{\text{orig}}$) even when one source is missing ($\Lambda_i = 0$), because, in fact, only one source is needed to determine the distance $d/2$.

In the following section we present an alternative derivation of FREM, which resolves these problems. In contrast to the original FREM$^{\text{orig}}$ our version gives diverging FREM for $d \to 0$ even for sources with different intensities. It also diverges in the situation when one of the sources is missing. For sources with equal intensities $\Lambda_1 = \Lambda_2$ our version and the original version of FREM$^{\text{orig}}$ give identical results.

3.4 An alternative derivation of FREM (FREM$^{\text{static}}$)

We assume two sources located along a line at positions c_1 and c_2 with intensities Λ_1 and Λ_2, respectively. If both sources have identical PSFs (here denoted as $q(x)$) we can express the intensity as:

$$\lambda(\boldsymbol{c}) = \Lambda_1 q(x - c_1) + \Lambda_2 q(x - c_2). \tag{3.6}$$

The distance between the two sources is $d = c_1 - c_2$, which is a linear combination $\boldsymbol{a}^T \cdot \boldsymbol{c}$ of the variable $\boldsymbol{c} = (c_1, c_2)^T$, where $\boldsymbol{a} = (1, -1)^T$. The variance of d is therefore

given by

$$\begin{aligned} \mathrm{var}(d) &= \mathrm{var}(\boldsymbol{a}^T \cdot \boldsymbol{c}) \\ &= \boldsymbol{a}^T \cdot \boldsymbol{Q} \cdot \boldsymbol{a}, \end{aligned} \tag{3.7}$$

where $\boldsymbol{Q}$ is the covariance matrix with lower bound given by the inverse of the Fisher information matrix (see Eq. (3.1) and Eq. (3.2)):

$$\boldsymbol{Q} \geq \boldsymbol{I}^{-1}(\boldsymbol{c}) = \frac{1}{I_{11}I_{22} - I_{12}^2} \begin{pmatrix} I_{22} & -I_{12} \\ -I_{12} & I_{11} \end{pmatrix}. \tag{3.8}$$

Expressing the elements of the covariance matrix $\boldsymbol{Q}$ from Eq. (3.8) and substitution to Eq. (3.7) gives the expression for $\mathrm{var}(d)$ from the elements of the Fisher information matrix

$$\begin{aligned} \mathrm{var}(d) &= Q_{11} + Q_{22} - 2Q_{12} \\ &\geq \frac{I_{11} + I_{22} + 2I_{12}}{I_{11}I_{22} - I_{12}^2}. \end{aligned} \tag{3.9}$$

We assume that the recorded images are corrupted with Poisson noise only (denoted here as $\mathrm{Po}(n;\lambda)$, or sometimes in a shorter version $\mathrm{Po}(\lambda)$, leaving only the expectation value λ as an argument). The probability distribution of n_k photons detection in the kth pixel is therefore

$$p(n_k|\boldsymbol{c}) = \mathrm{Po}\left(n_k; \lambda_k(\boldsymbol{c})\right), \tag{3.10}$$

where λ_k is the expected intensity in pixel k. It is obtained by integration of the intensity distribution $\lambda(x)$ from Eq. (3.6) over the area of a pixel Γ_k:

$$\lambda_k(\boldsymbol{c}) = \int_{\Gamma_k} \Lambda_1 q(x - c_1) + \Lambda_2 q(x - c_2) dx + b. \tag{3.11}$$

The constant b is a homogeneous background in each pixel.

If we suppose uncorrelated noise between pixels, we get the log-likelihood function for N pixels:

$$\mathcal{L} = \sum_{k=1}^{N} \log p(n_k|\boldsymbol{c}) = \sum_{k=1}^{N} \log\left[\mathrm{Po}\left(n_k; \lambda_k(\boldsymbol{c})\right)\right]. \tag{3.12}$$

Inserting $\mathcal{L}$ into Eq. (3.2), the elements of the Fisher information matrix become (see Eq. (B.32) in Appendix B for details)

$$I_{ij}(\boldsymbol{c}) = \sum_{k=1}^{N} \frac{1}{\lambda_k} \frac{\partial \lambda_k}{\partial c_i} \frac{\partial \lambda_k}{\partial c_j}; \;\; i,j \in \{1,2\}. \tag{3.13}$$

By substitution from Eq. (3.11) we get for the individual elements of the Fisher information matrix (see Eq. (B.32) in Appendix B for details):

$$I_{ij} = \Lambda_i \Lambda_j \sum_{k=1}^{N} \frac{q_k'(c_i) q_k'(c_j)}{\Lambda_1 q_k(c_1) + \Lambda_2 q_k(c_2) + b}; \;\; i,j \in \{1,2\}, \tag{3.14}$$

where $q_k(c_i)$ and $q_k'(c_i)$ are the pixelated versions (pixel area Γ_k) of the PSF and the derivative, respectively:

$$q_k(c_i) = \int_{\Gamma_k} q(x - c_i) dx$$
$$q_k'(c_i) = \int_{\Gamma_k} \frac{\partial q(x - c_i)}{\partial x} dx.$$

For equally strong sources ($\Lambda_1 = \Lambda_2 = \Lambda$) we get a compact expression for the entries of the Fisher information:

$$I_{ij} = \Lambda \sum_{k=1}^{N} \frac{q_k'(c_i) q_k'(c_j)}{q_k(c_1) + q_k(c_2) + b/\Lambda}; \;\; i,j \in \{1,2\}, \tag{3.15}$$

and due to the symmetry of the entries ($I_{11} = I_{22}$ and $I_{12} = I_{21}$) the variance Eq. (3.9) can be expressed as

$$\mathrm{var}(d) \geq \frac{2}{I_{11} - I_{12}}. \tag{3.16}$$

Inserting the matrix elements Eq. (3.15) into Eq. (3.16) shows that for situations where the background level is considerably smaller than the intensity $b/\Lambda \ll 1$, the lower bound on variance scales with the sources' intensity Λ as

$$\mathrm{var}(d) \propto \frac{1}{\Lambda}. \tag{3.17}$$

However, the exact value depends on the shape of the PSF $q(x)$.

In Appendix B we show the equivalence of the original FREM$^{\mathrm{orig}}$ Eq. (3.3) and our version Eq. (3.14) for sources with equal strength ($\Lambda_1 = \Lambda_2$). However, as we demonstrate in Sect. 3.7.1, the expression gives very different results for sources of unequal intensity.

FREM$^{\text{static}}$ computed from Eq. (3.14) have reasonable behaviour in the limits $d \to 0$ and $d \to \infty$ (see Sect. B.4 for details). The limit of very close sources ($d \to 0$) gives FREM$^{\text{static}} \to \infty$ for any value of Λ_i and Λ_j. Also, in contrast to the original FREM$^{\text{orig}}$ expression, our FREM$^{\text{static}}$ diverges if one of the sources is missing $\Lambda_i = 0$, because we do not make any assumption about the symmetry with respect to the origin.

For well-separated sources ($d \to \infty$) the off-diagonal elements of the Fisher information matrix vanish ($I_{ij} = 0$ for $i \neq j$) and from Eq. (3.9) we get

$$\text{var}(d) \geq \frac{1}{I_{11}} + \frac{1}{I_{22}}. \tag{3.18}$$

Fraction $1/I_{ii}$ is the lower bound on the variance of the single source location c_i estimator. This can be obtained from the general expression of the Fisher information for Poisson distributed data (see Eq. (B.16) in Appendix B). The bound on the total variance is therefore composed from the sum of bounds on variances for localisation of individual sources, as we expect.

3.5 FREM for blinking sources (FREM$^{\text{blink}}$)

Fundamental resolution measure discussed in the previous section considers only the total number of photons Λ_i emitted by each source s_i. In this section we derive FREM$^{\text{blink}}$ for sources with intermittent intensity and compare it to the "static" FREM$^{\text{static}}$ derived above.

To address this question we assume a simple model of Poisson distributed data with expected pixel values λ_k (Eq. (3.11)). To account for the intermittent behaviour of the intensity, we turn the intensity vector $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$ into a random variable distributed over four distinctive states (indexed with a superscript α):

$$\left\{ \mathbf{\Lambda}^{\alpha=1} = (\Lambda_1, 0), \mathbf{\Lambda}^{\alpha=2} = (0, \Lambda_2), \mathbf{\Lambda}^{\alpha=3} = (\Lambda_1, \Lambda_2), \mathbf{\Lambda}^{\alpha=4} = (0,0) \right\}, \tag{3.19}$$

which is a simple model of, for example, two blinking quantum dots. The expected

intensity in the kth pixel when $\mathbf{\Lambda}$ is in the state $\mathbf{\Lambda}^{\alpha}$ is then $\lambda_k^{\alpha} = \lambda_k(\mathbf{\Lambda}^{\alpha})$:

$$\begin{aligned}
\lambda_k^{\alpha=1} &= \Lambda_1 q_k(x-c_1) &&+ b,\\
\lambda_k^{\alpha=2} &= &\Lambda_2 q_k(x-c_2) &+ b,\\
\lambda_k^{\alpha=3} &= \Lambda_1 q_k(x-c_1) &+\Lambda_2 q_k(x-c_2) &+ b,\\
\lambda_k^{\alpha=4} &= &&+ b,
\end{aligned} \tag{3.20}$$

where homogeneous background b was added to each pixel.

3.5.1 Averaging the Fisher information

The "averaging" of the Fisher Information matrix presented in this section assumes knowledge of the intensity state (ON/OFF) of each source in every acquired frame. This information is not accessible in the real situation. However, we show the derivation to emphasise the difference between this approach and the more realistic situation, where the intensity states are described by probability distribution (Sect. 3.5.2).

If the intensity states $\mathbf{\Lambda}$ *were known*, we would write the log-likelihood function as

$$\mathcal{L}(\theta, \Lambda) = \sum_{k=1}^{N} \log\left(l_k(\theta, \mathbf{\Lambda})\right). \tag{3.21}$$

and the expected Fisher information matrix would become an average over the individual known states (see Appendix B for details)

$$I(\theta) = \int_{\mathbf{\Lambda}} p(\mathbf{\Lambda}) I(\theta, \mathbf{\Lambda}) d\mathbf{\Lambda},$$

where $p(\mathbf{\Lambda})$ is the distribution of the $\mathbf{\Lambda}$ states and $I(\theta, \mathbf{\Lambda})$ is the Fisher information computed for a specific value of $\mathbf{\Lambda}$ (see Eq. (3.11) and Eq. (3.13)). For discrete states of $\mathbf{\Lambda}^{\alpha}$ shown in Eq. (3.20) we get

$$I(\theta) = \sum_{\alpha} p(\mathbf{\Lambda}^{\alpha}) I(\theta, \mathbf{\Lambda}^{\alpha}), \tag{3.22}$$

where the Fisher Information for every configuration of $\mathbf{\Lambda}^{\alpha}$ is averaged with weights $p(\mathbf{\Lambda}^{\alpha})$.

3.5.2 Integrating over the intensity states

However, we assume that the variable $\mathbf{\Lambda}$ is fully described by the probability $p(\mathbf{\Lambda})$ over the states. The exact state in time frame is unknown. Therefore we have to integrate over $\mathbf{\Lambda}$. The likelihood function is then

$$\begin{aligned} l(\theta) &= \prod_{k=1}^{N} p(n_k|\theta) \\ &= \prod_{k=1}^{N} \int_{\mathbf{\Lambda}} p(n_k, \mathbf{\Lambda}|\theta) \\ &= \prod_{k=1}^{N} \sum_{\alpha=1}^{4} p(n_k|\theta, \mathbf{\Lambda}^{\alpha}) p(\mathbf{\Lambda}^{\alpha}). \end{aligned} \tag{3.23}$$

This complicates the evaluation of the Fisher information matrix Eq. (3.2) because of the summation within the logarithm in the log-likelihood

$$\mathcal{L}(\theta) = \log l(\theta) = \sum_{k} \log \left(\sum_{\alpha=1}^{4} p(n_k|\theta, \mathbf{\Lambda}^{\alpha}) p(\mathbf{\Lambda}^{\alpha}) \right). \tag{3.24}$$

In Appendix B we show that the Fisher information matrix for uniform distribution $p(\mathbf{\Lambda}^{\alpha}) = \frac{1}{4}$ over the four intensity states Eq. (3.19) is given by

$$I_{rs}(\theta) = \sum_{k=1}^{N} \mathbb{E}_k \left[\frac{\left(\sum_{\alpha=1}^{4} \frac{\partial \mathrm{Po}(\lambda_k^{\alpha})}{\partial c_r} \right) \left(\sum_{\alpha=1}^{4} \frac{\partial \mathrm{Po}(\lambda_k^{\alpha})}{\partial c_s} \right)}{\left(\sum_{\alpha=1}^{4} \mathrm{Po}(\lambda_k^{\alpha}) \right)^2} \right], \tag{3.25}$$

where $\mathbb{E}_k[.]$ represents the expectation value with respect to $p(n_k, \mathbf{\Lambda}|\theta)$ (see Eq. (3.24)). Expressing the derivatives and the expectation value gives

$$\begin{aligned} I_{rs}(\theta) = \frac{1}{4} \sum_{k=1}^{N} \left(\frac{\partial \lambda_k^{\alpha=r}}{\partial c_r} \right) \left(\frac{\partial \lambda_k^{\alpha=s}}{\partial c_s} \right) \times \\ \times \sum_{n_k \geq 0} \left[\frac{\left(\sum_{\alpha \in \{r,3\}} \mathrm{Po}(n_k; \lambda_k^{\alpha}) \frac{(n_k - \lambda_k^{\alpha})}{\lambda_k^{\alpha}} \right) \left(\sum_{\alpha \in \{s,3\}} \mathrm{Po}(n_k; \lambda_k^{\alpha}) \frac{(n_k - \lambda_k^{\alpha})}{\lambda_k^{\alpha}} \right)}{\sum_{\alpha=1}^{4} \mathrm{Po}(n_k; \lambda_k^{\alpha})} \right]. \end{aligned} \tag{3.26}$$

In Appendix B we show that the limit $d \to 0$ gives $\mathrm{var}(d) \to \infty$ and the limit $d \to \infty$ gives $\mathrm{var}(d) \geq \frac{1}{I_{11}} + \frac{1}{I_{22}}$. We also show, that for well-separated sources ($d \to \infty$) and negligible background ($b \ll \Lambda$) the variance $\mathrm{var}(d)$ is identical for both blinking and static situation, if the total number of emitted photons is kept constant.

3.6 Experimental parameters and numerical evaluations

We made a comparison of the original FREM$^{\text{orig}}$ formula computed from the Fisher information Eq. (3.3) with our proposed fixed FREM$^{\text{static}}$ formula computed from Eq. (3.14), for sources with static intensity. We also compared the static situation with FREM$^{\text{blink}}$ for sources with intermittent intensity computed from Eq. (3.26).

We considered 625 nm emission light wavelength and 1.2 NA objective. Images were pixelated with 80×80 nm pixels. Various intensities of the emitters Λ_i and pixel background levels b were considered.

The pixelated version $q_k(c_i)$ of the continuous PSF $q(x - c_i)$ and the corresponding derivatives $q'_k(c_i)$ from Eq. (3.3) and Eq. (3.14) were computed by summing 10×10 pixels of $10\times$ oversampled images (approximation of the continuous PSF $q(x)$ on the 8×8 nm grid). The pixelated λ_k^α in Eq. (3.25) was computed in a similar manner.

The expectation values in Eq. (3.25) were evaluated using the expression Eq. (3.26). The set of images for a range $n_k = [0..n_{max}]$ was computed to perform the summation $\sum_{n_k \geq 0}$. The value of n_{max} was set such that the Poisson cumulative distribution function $F(n)$ for the pixel with the maximum intensity satisfies $F(n > n_{max}) > 1 - t$, with $t = 10^{-6}$.

3.7 Results

We computed the FREM for simulated datasets corresponding to different experimental settings, such as the separation of the sources d, the total number of emitted photons by each source Λ and the background offset b in the recorded frames. FREM gives us the lower bound on the standard deviation ($\sqrt{\text{var}(d)}$) for the measurement of the source separation d. The source separation equivalent to FREM (d=FREM) can be considered as a "natural resolution limit", which takes the statistical nature of the photon detection into account.

3.7.1 Comparison of the original FREM$^{\text{orig}}$ and proposed FREM$^{\text{static}}$ formula

We compared FREM$^{\text{orig}}$ computed from the original Eq. (3.3) and our proposed FREM$^{\text{static}}$ Eq. (3.14) formula of the Fisher information for two static sources. It can be shown (Appendix B), that if the sources have equal strength ($\Lambda_1 = \Lambda_2$), both formulas give identical results. However, for unequal sources $\Lambda_1 \neq \Lambda_2$ the FREM values differ significantly.

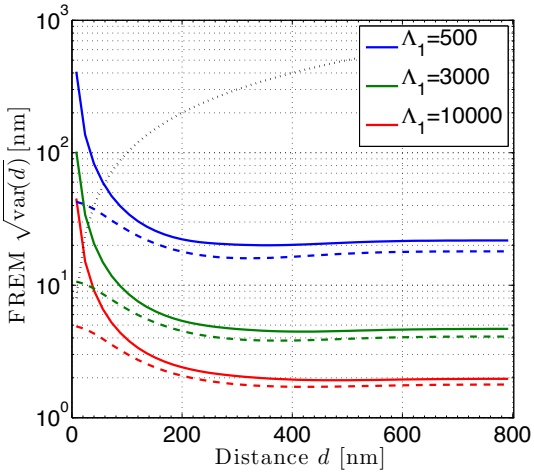


Figure 3.1: Comparison of the original FREM$^{\text{orig}}$ formula computed from Eq. (3.3) (dashed line) and our proposed FREM$^{\text{static}}$ formula Eq. (3.14) (solid line) for two sources with unequal intensities $\Lambda_2 = 2\Lambda_1$ and background $b = 100$ photons/pixel. The black dotted curve corresponds to FREM=d ("natural resolution limit"). This would be a straight line with unit gradient in a linear plot.

The sources s_i were represented with an in-focus PSF centred at c_i. The intensity of s_2 was set to double of the intensity of s_1: $\Lambda_2 = 2\Lambda_1$. Three different intensity levels $\Lambda_1 = 500, 3000$ and 10^4 photons with homogeneous background $b = 100$ photons/pixel were considered. Figure 3.1 shows FREM (lower bound on $\sqrt{\text{var}(d)}$) for a range of sources separations d evaluated with the original FREM$^{\text{orig}}$, computed from Eq. (3.3) (dashed line) and our proposed FREM$^{\text{static}}$ Eq. (3.14) (solid line).

The FREM$^{\text{orig}}$ formula gives consistently lower values then FREM$^{\text{static}}$ (dashed curves

are under the solid lines for the whole range of d in Figure 3.1). The original FREM$^{\text{orig}}$ (dashed curves) also tends to finite values even for $d \to 0$. We discuss this behaviour further in Sect. 3.8.1.

3.7.2 FREM for static and blinking sources

In order to compare the blinking situation Eq. (3.25) with the static case Eq. (3.14) we evaluated FREM$^{\text{static}}$ and FREM$^{\text{blink}}$ as a function of the source separation d. For the blinking situation we considered equal strength of the sources

$$\Lambda_1^{blink} = \Lambda_2^{blink} = 2\Lambda \tag{3.27}$$

and the homogeneous background b^{blink} in each pixel of each frame. Because the sources are "ON" only in 50% of the cases (see Eq. (3.20)), the total number of emitted photons per source per frame is Λ on average.

For the static case we considered the situation of two sources emitting with equal intensities. To keep the total number of emitted photons per frame equal to the blinking case, we set the intensity

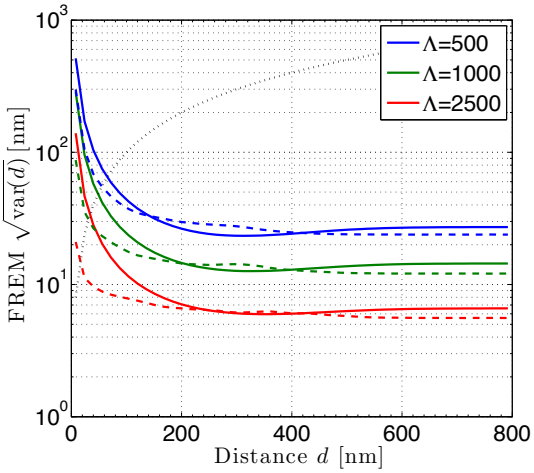
$$\Lambda_1^{static} = \Lambda_2^{static} = \Lambda. \tag{3.28}$$

The background values are equal for the blinking and the static case $b^{blink} = b^{static}$. Comparison of FREM as a function of the separation d for the blinking and the static case is shown in Fig. 3.2. Three different values of the total number of photons Λ were considered in the semi-logarithmic plot Fig. 3.2a. All curves are computed for a fixed background level $b = 100$ photons/pixel.

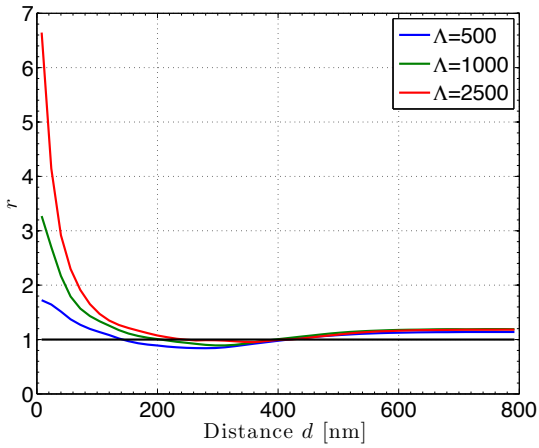
The ratio of the FREM curves

$$r = \frac{FREM^{\text{static}}(d)}{FREM^{\text{blink}}(d)} \tag{3.29}$$

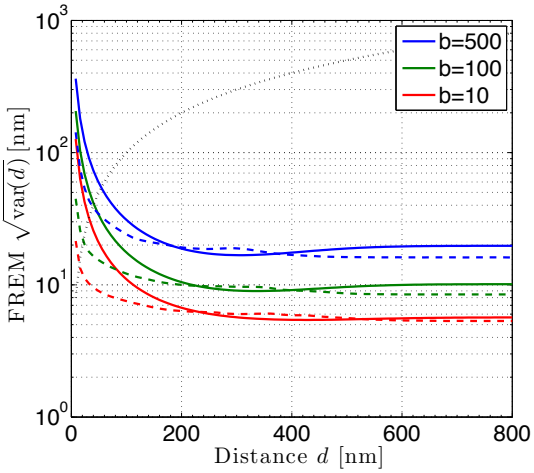
for the blinking and the static case are shown in Fig. 3.2b. The plot shows how many times is the FREM$^{\text{blink}}$ lower when compared to the FREM$^{\text{static}}$. Figure 3.2c shows the



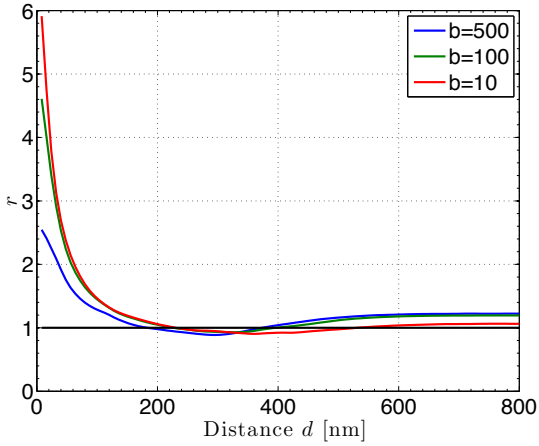
(a) FREM (fixed b =100 phot/pixel)



(b) Ratio of the curves form (a)



(c) FREM (fixed $\Lambda = 1500$ photons)



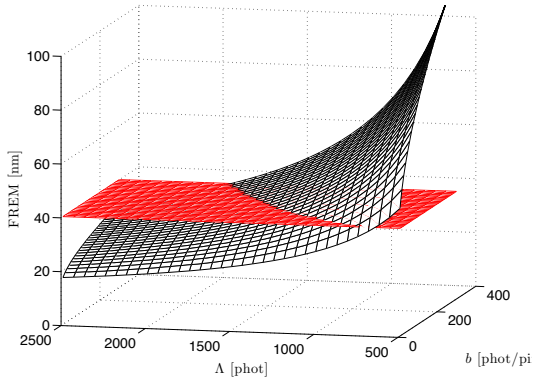
(d) Ratio of the curves form (c)

Figure 3.2: *Left:* FREM (a) for fixed background $b = 100$ photons and three different intensities Λ of the sources and (c) for fixed total number of emitted photons $\Lambda = 1500$ and three different values of background b. Dashed lines correspond to the blinking situation $\mathrm{FREM}^{\mathrm{blink}}$ Eq. (3.25), solid lines correspond the static situation $\mathrm{FREM}^{\mathrm{static}}$ Eq. (3.14). *Right:* Ratio r of the static (solid) to blinking (dashed) curves showing how many times is $\mathrm{FREM}^{\mathrm{blink}}$ lower compared to the $\mathrm{FREM}^{\mathrm{static}}$. The classical resolution limit δ (radius of an Airy disk) corresponds to $\delta = 320$ nm.

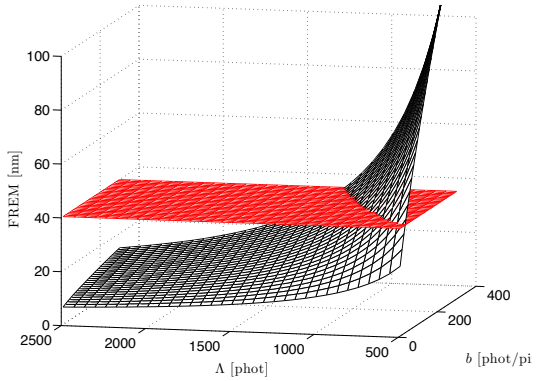
FREM curves for three different background values b. The total number of emitted photons per source was set to $\Lambda = 1.5 \cdot 10^3$ photons.

$\mathrm{FREM}^{\mathrm{blink}}$ (dashed curves) is in general lower than $\mathrm{FREM}^{\mathrm{static}}$. The exception is a

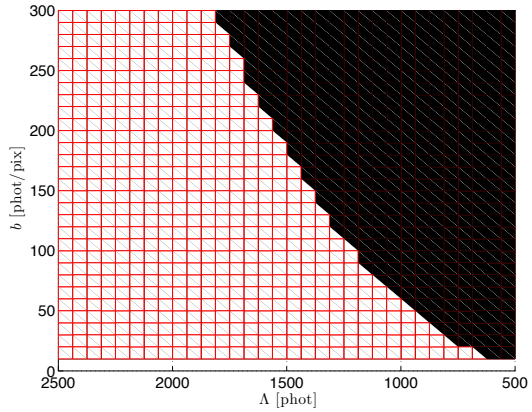
small region centred at around 300 nm (see Sect. 3.8.3 for more discussion). The difference between the curves is most pronounced for closely spaced sources ($d < 100$ nm) and data with high signal-to-noise ratio (red curves - bright sources with low background).



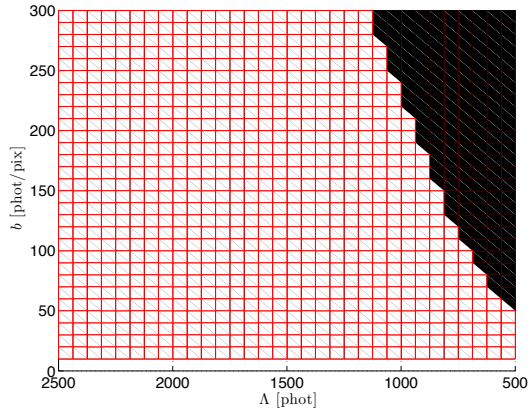
(a) FREM$^{\text{static}}$



(b) FREM$^{\text{blink}}$



(c) FREM$^{\text{static}}$ - top view



(d) FREM$^{\text{blink}}$ - top view

Figure 3.3: *Top:* FREM for (a) static and (b) blinking situation for two sources separated by $d = 40$ nm. *Bottom:* Top view on the surfaces. Red plane corresponds to situation when FREM is equal to the separation of the sources $d = 40$ nm. The region, where the surface is above the red plane (in black) does not allow precise estimation of d (FREM $> d$). Examples of the data frames corresponding to the points in the $\Lambda \times b$ plane are shown in Fig. 3.4.

For further comparison of the static and the blinking FREM, we fixed the separation between the two sources to $d = 40$ nm and computed FREM for a range of background

b and intensity Λ values. Figure 3.3 compares the static (left) with the blinking situation (right). The red plane corresponds to the "natural" resolution limit FREM $= d$ (see Sect. 3.1), where the lower bound on standard deviation of the distance estimation $\sqrt{\mathrm{var}(d)}$ is equal to the separation d. In the region, where the black surface is above the red plane, the distance estimation is very imprecise. This region corresponds to the sources closer than the "natural resolution limit". These (black) regions can be easily observed from the top view shown in the bottom plots of Fig. 3.3 demonstrating the increase of the "resolution region" for the blinking case.

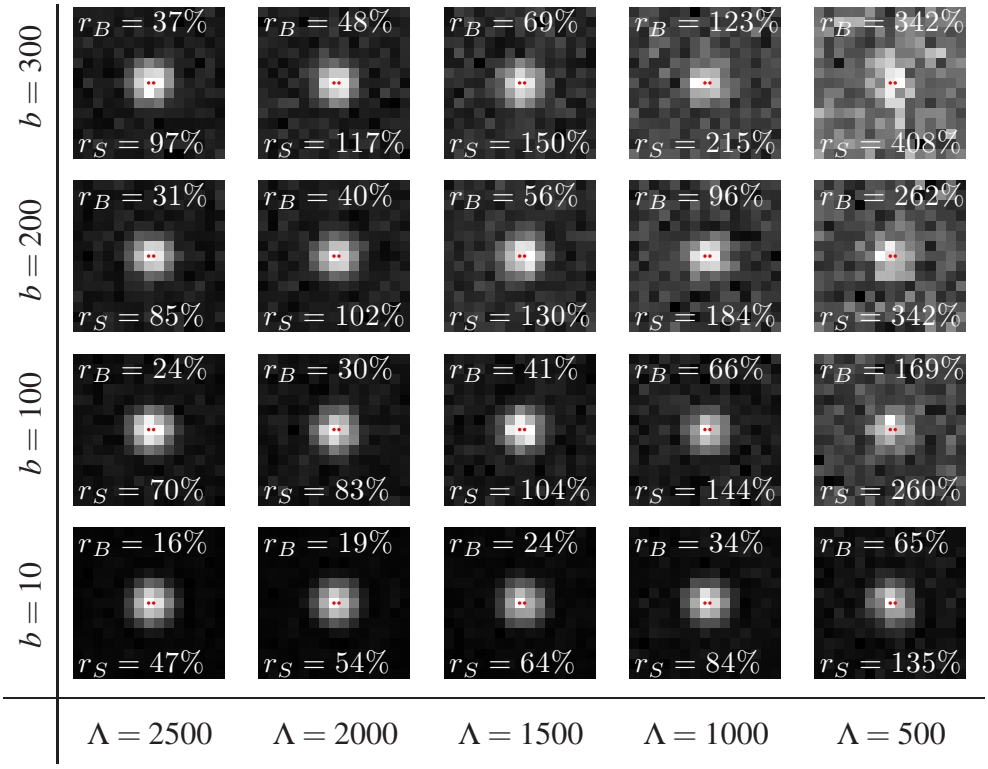
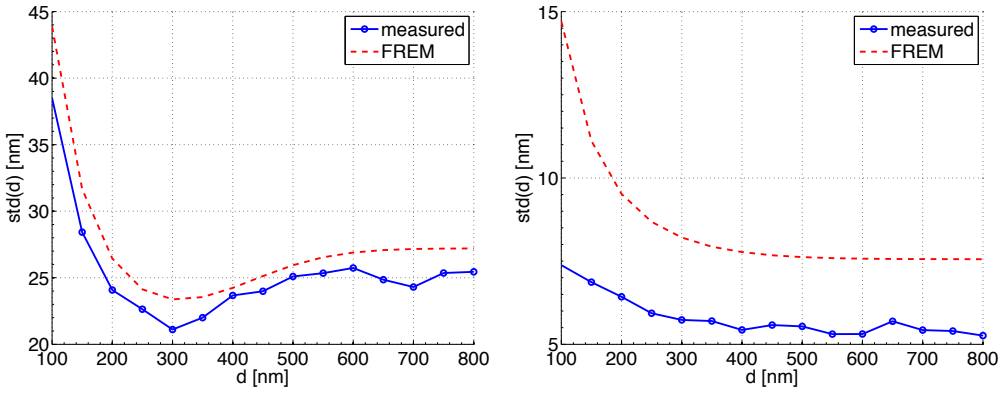


Figure 3.4: Illustration of two simulated sources separated by $d = 40$ nm with intensity Λ (total number of emitted photons per source per frame) and the background b. Data were corrupted with Poisson noise. The red dots indicate the positions of the sources. The layout corresponds to Fig. 3.3. Numbers at the top and the bottom of each figure state the ratio (in percent) $r_B = FREM^{\mathrm{blink}}/d \times 100$ and $r_S = FREM^{\mathrm{static}}/d \times 100$ for the blinking and the static situation, respectively, indicating the how many percent of the separation d represents the FREM value. The smaller the values, the higher the precision of the distance estimator.

Figure 3.4 shows the noisy images of two sources with parameters b and Λ with layout similar to the graphs in Fig. 3.3. The black regions from Fig. 3.3c and Fig. 3.3d correspond to extremely noisy data (top right corner) and the high FREM values are not surprising.

For high signal-to-noise ratio data (high Λ and low b, see bottom-left corner of Fig. 3.3 and Fig. 3.4) the lower bound on the standard deviation of the separation $d = 40\,\text{nm}$ estimation can be as low as 7 nm for the blinking case $\text{FREM}^{\text{blink}}$ (Fig. 3.3b). $\text{FREM}^{\text{static}}$ are approximately three times higher ($\sim 20\,\text{nm}$, Fig. 3.3a). Note that the $\text{FREM}^{\text{blink}}$ surface in Fig. 3.3b for the blinking situation has much steeper increase from the sub 40 nm region than the surface for the static case Fig. 3.3a.



(a) b=100 **(b)** b=0

Figure 3.5: $\text{FREM}^{\text{static}}$ and measured standard deviation ($\sqrt{\text{var}(d)}$) for the estimation of the separation between two simulated noisy static sources ($\Lambda_1 = \Lambda_2 = 500$ photons). A homogeneous background of (a) $b = 100$ photons/pixel and (b) $b = 0$ photons/pixel was added to each simulated image before realisation of Poisson noise. $\text{FREM}^{\text{static}}$ is shown as a red dashed line. Estimated standard deviation from 1000 different realisation of Poisson noise is plotted with blue circles. (a) $\text{FREM}^{\text{static}}$ curve shows a distinct "dip" at 300 nm when background noise is present in the recorded images. (b) For background-free data $\text{FREM}^{\text{static}}$ curves are monotonically decreasing. The measured values are under the $\text{FREM}^{\text{static}}$ curves due to the initialisation of the maximum-likelihood fitting with $\boldsymbol{c}_{true}$ (see discussion in the text).

The $\text{FREM}^{\text{static}}$ formula for static sources derived from Eq. (3.14) shows an interesting

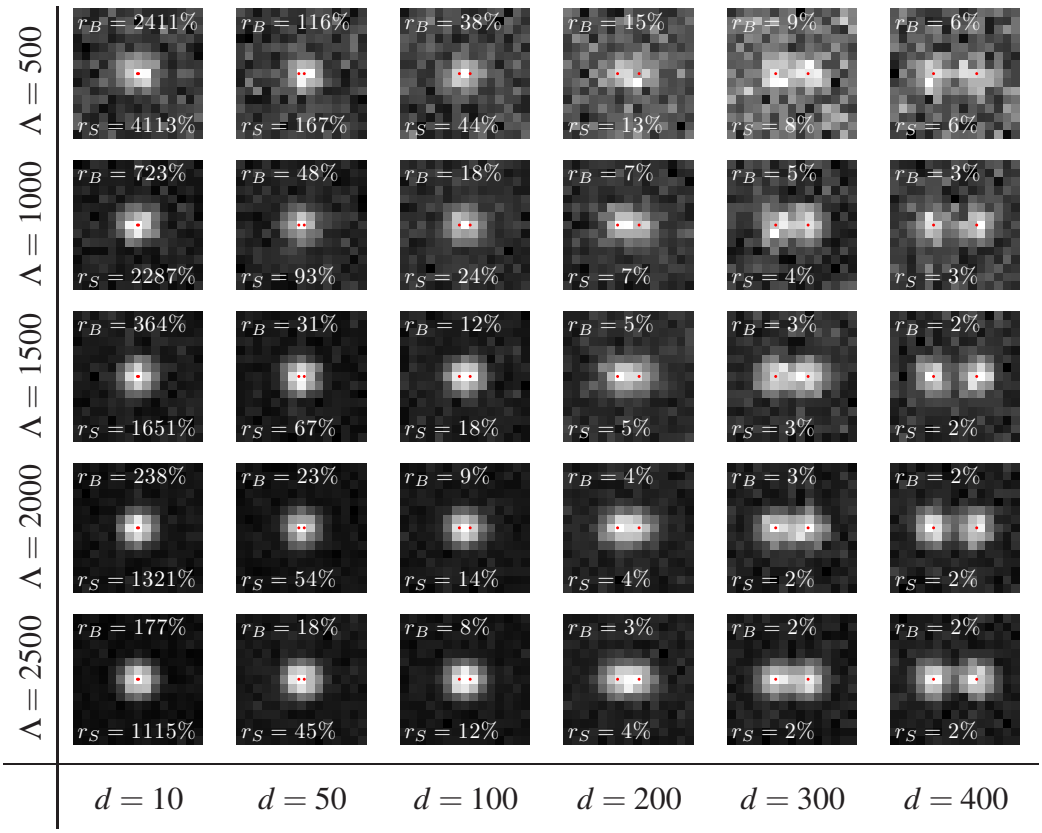


Figure 3.6: Two simulated sources separated by distance d [nm] with intensity Λ [total number of emitted photons per source]. The background was set to $b = 100$ photons/pixel and the images were corrupted with Poisson noise. The red dots indicate the positions of the sources. Numbers at the top and the bottom of each figure state the ratio (in percent) $r_B = FREM^{\text{blink}}/d \times 100$ and $r_S = FREM^{\text{static}}/d \times 100$ for the blinking and the static situation, respectively, indicating the how many percent of the separation d represents the FREM value. The smaller the values, the higher the precision of the distance estimator. The classical resolution limit δ (radius of an Airy disk) corresponds to $\delta = 320$ nm.

behaviour for weak sources with large background values. The red dashed curve in Fig. 3.5a shows $\text{FREM}^{\text{static}}$ for two sources of equal intensity $\Lambda_1 = \Lambda_2 = 500$ photons with background $b = 100$ photons/pixel. This parameter settings corresponds to the top line in Fig. 3.6. Contrary to our intuition, $\text{FREM}^{\text{static}}$ is not monotonic with separation d, and the $\text{FREM}^{\text{static}}$ curve shows a "dip" at $d \approx 300$ nm.

After careful checking of the derivation and the numerical calculations of the curves

(see Sect. B.8 in Appendix B for details), we interpret the dip to be a result of two competing factors. The first factor is the decrease of $\mathrm{FREM}^{\mathrm{static}}$ with increasing separation d. This is in accordance with our intuition, that the separation between the sources becomes progressively easier to estimate with increasing distance d between the sources (reducing their mutual overlap). While this is the case for the background-free data Fig. 3.5b, the situation is more complicated for data with noisy background. Weak sources can "disappear" in the strong background noise if they are well separated (cf. right of the top line in Fig. 3.6). The same sources are easier to detect if they are close, because their overlapping PSFs create a bright object in the noisy background (left of the top line in Fig. 3.6). The "visibility" of the sources is the second competing factor, which decreases with d. The dip in the $\mathrm{FREM}^{\mathrm{static}}$ curves therefore represents an "optimal separation" d, where the sources are already sufficiently separated to be localised with good precision but still "visible" due to their overlapping PSFs. Further increase of the localisation precision with d is not sufficient to compensate for the fact that sources "disappear" in noise. Note, that for our simulation ($\lambda_{em} = 625\,\mathrm{nm}$, 1.2NA) the Abbe resolution criterion is 260 nm.

A qualitative confirmation of this counter-intuitive behaviour is shown on simulated data in Fig. 3.5. We simulated two sources separated with a distance d. The intensity and the background was set to the same values as for the theoretical $\mathrm{FREM}^{\mathrm{static}}$ curves ($\Lambda_1 = \Lambda_2 = 500$ photons, $b = 100$ photons/pixel, see top row in Fig. 3.6). For each separation d we created $M = 1000$ images with different realisation of Poisson noise. Using conjugate gradient optimisation (NETLAB function `conjgrad` [82]), we found the maximum-likelihood estimator of the positions $\boldsymbol{c}_{ML} = (c_1^{ML}, c_2^{ML})$ of two PSFs (see Eq. (3.11)). The initial position for the fitting procedure was set to the true values $\boldsymbol{c}_{true}$. The standard deviation of the M estimators $d^{ML} = |c_1^{ML} - c_2^{ML}|$ is plotted with blue circles in Fig. 3.5a and shows a "dip" similar to the one in the theoretical $\mathrm{FREM}^{\mathrm{static}}$ curve. For zero background ($b = 0$) $\mathrm{FREM}^{\mathrm{static}}$ is monotonically decreasing (the dashed red line in Fig. 3.5b). Standard deviation of the M estimators is plotted

with blue circles in Fig. 3.5b. No dip is observed in this case.

Note that the measured values of the standard deviation (blue circles) are lower than the theoretical lower bound (red dashed curve). This is likely due to the initialisation of the maximum-likelihood fitting with $\boldsymbol{c}_{true}$. Random initialisation would be more appropriate, but the measured standard deviation does not provide curves smooth enough to show the "dip" clearly.

3.8 Discussion

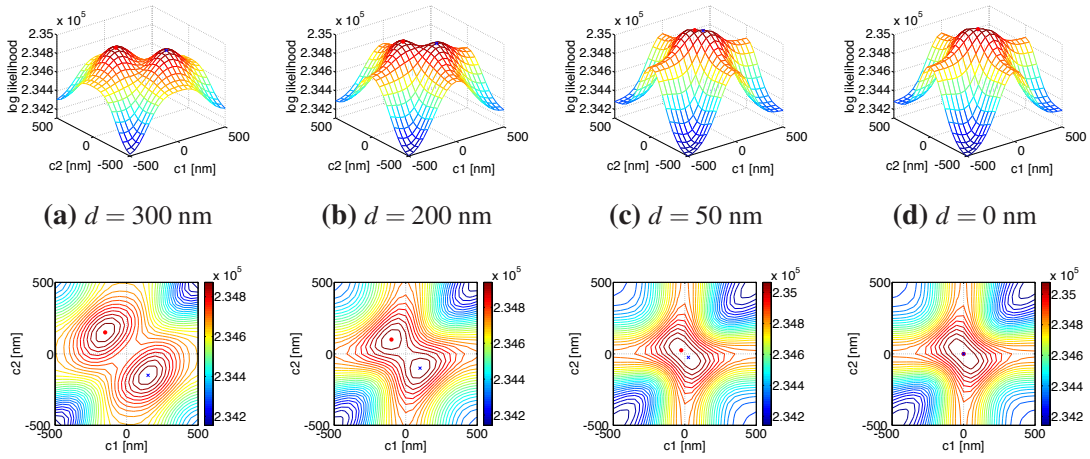
This section contains a general discussion about FREM and the results presented above. We also discuss and explain the strange behaviour of the original FREM$^{\text{orig}}$ curves.

In Sect. 3.8.1 we give some insight into the behaviour of FREM$^{\text{static}}$ by visualisation of the expected log-likelihood surface. We consider the Fisher information as a measure of the surface's curvature. Section 3.8.3 compares the difference between the averaging and integrating over the intensity states $\boldsymbol{\Lambda}^{\boldsymbol{\alpha}}$ in the Fisher information matrix for blinking sources. In Sect. 3.8.4 we comment on the scaling of FREM with background and the sources' intensity and in Sect. 3.8.5 we shortly discuss noise in the recorded images and in Sect. 3.8.6 we make a link to the iNMF results.

3.8.1 Visualisation of the expected log-likelihood surface

In order to understand the behaviour of the Fisher information matrix, we visualised the surface of the expected log-likelihood Eq. (3.12) as a function of the parameter $\boldsymbol{c} = (c_1, c_2)$ in Fig. 3.7:

$$\begin{aligned}\mathbb{E}_{p(n|\lambda^{true})}\left[\mathcal{L}(\boldsymbol{c})\right] &= \mathbb{E}_{p(n|\lambda^{true})}\left[\sum_{k=1}^{N}\log p\left(n_k|\lambda_k(\boldsymbol{c})\right)\right] \\ &= \sum_{k=1}^{N}\left(\lambda_k^{true}\log\lambda_k(\boldsymbol{c}) - \lambda_k(\boldsymbol{c})\right) + A. \end{aligned} \tag{3.30}$$



(a) $d = 300$ nm **(b)** $d = 200$ nm **(c)** $d = 50$ nm **(d)** $d = 0$ nm

(e) $d = 300$ nm **(f)** $d = 200$ nm **(g)** $d = 50$ nm **(h)** $d = 0$ nm

Figure 3.7: Surface of the expected log-likelihood Eq. (3.30) as a function of $\boldsymbol{c} = (c_1, c_2)$ for different separation d between the two sources, located at $\boldsymbol{c}^{true}$ (marked with red dot). The point where the sources exchange their locations is marked with blue cross. The classical resolution limit corresponds to $\delta = 320$ nm. Movement along the "top-left to bottom-right" diagonal represents moving the points apart (see Fig. 3.8 for details).

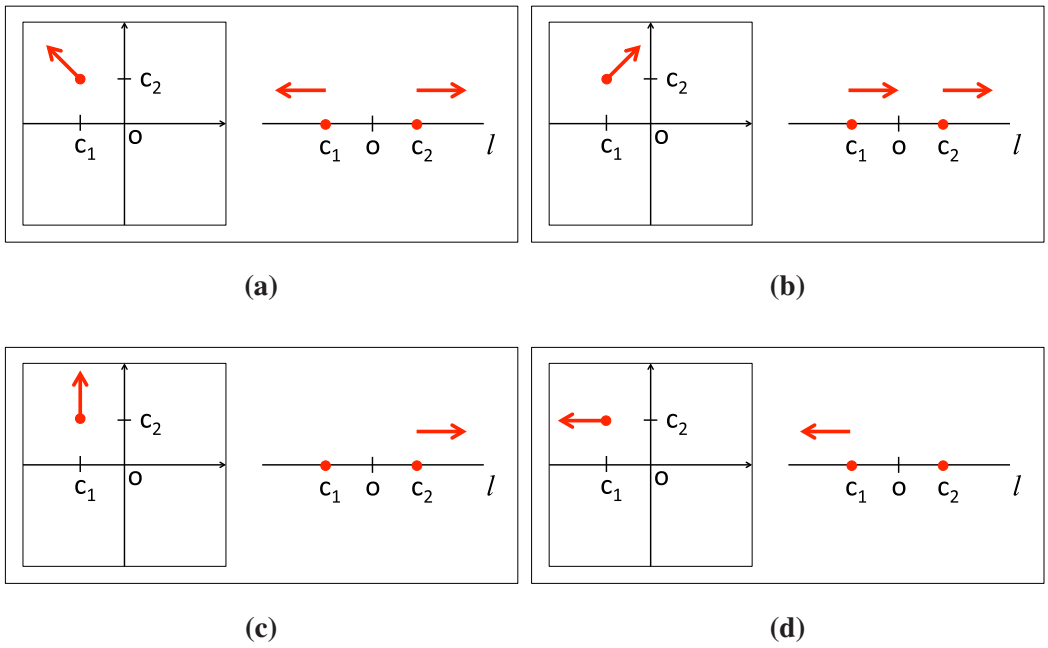


Figure 3.8: Illustration of the translation of sources s_1 and s_2 along line l and the corresponding movement in the parameter space from Fig. 3.7.

Note that the expectation is taken with respect to the "true" distribution $\lambda^{true} = \lambda(\boldsymbol{c}^{true})$, while the log-likelihood $\mathcal{L}$ is a function of $\boldsymbol{c}$. A is independent of $\boldsymbol{c}$.

The surface in Fig. 3.7 shows the average log-likelihood for a model with two sources s_1 and s_2 located at c_1 and c_2, respectively, for data generated from a model consisting of two sources s_1^{true} and s_2^{true} located at c_1^{true} and c_2^{true}, respectively, corrupted with Poisson noise. Parameters of the simulation were $\Lambda = 10^3$, $b = 100$ and wavelength 625 nm.

The correspondence between the (c_1, c_2) space of Fig. 3.7 and the physical movement of the sources is illustrated in Fig. 3.8. The coordinates (c_1, c_2) represent the positions of two sources on a line l intersecting both sources s_1 and s_2. The origin $o = (0,0)$ corresponds to the geometric centre between c_1^{true} and c_2^{true}. Moving along the top-left to bottom-right diagonal (Fig. 3.8a) represents a symmetrical movement of s_1 and s_2 in opposite directions with respect to o, while moving parallel to the top-right to bottom-left diagonal (Fig. 3.8b) represents the translation of s_1 and s_2 together along

l, while keeping their distance from each other constant. Moving along a vertical line corresponds to the situation, where the position of s_1 fixed while s_2 is moving (Fig. 3.8c) and vice versa for horizontal lines (Fig. 3.8d).

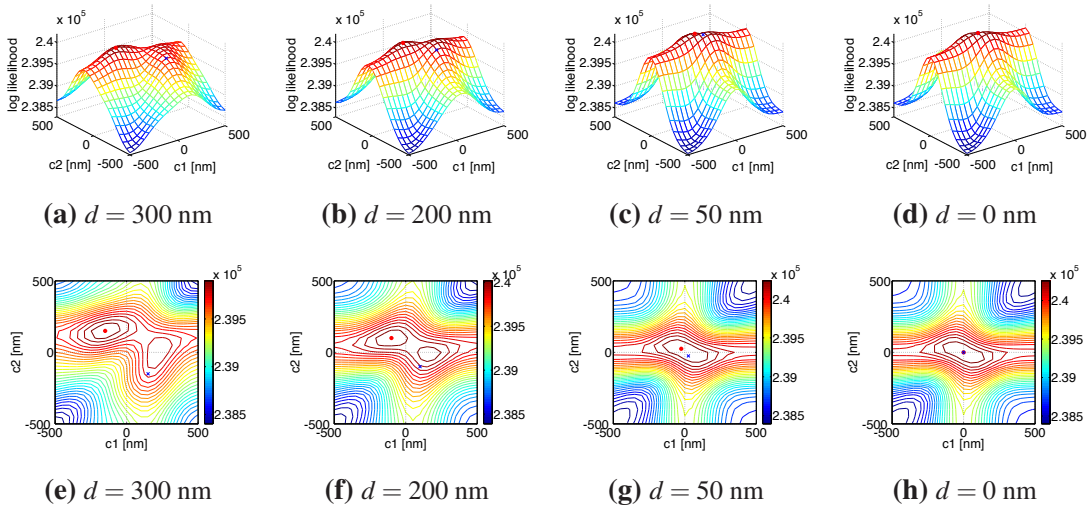
For well-separated sources the surface (Fig. 3.7a) has a sharp maximum at $\boldsymbol{c}^{true} = (c_1^{true}, c_2^{true})$ (red dot in Fig. 3.7a). In fact, there is another equivalent maximum (blue cross in Fig. 3.7a) as the points are interchangeable and the surface is symmetrical along top-right to bottom-left diagonal. It is important to note that the surface falls sharply in all directions around the maximum. In other words, the likelihood of a model s_1 and s_2 for data generated from s_1^{true} and s_2^{true} drops quickly once the s_1 and s_2 move anywhere from the "true" locations $\boldsymbol{c}^{true}$.

Once the true sources s_1^{true} and s_2^{true} come closer together (Fig. 3.7b,c), the maximum of the surface becomes less pronounced, especially along the top-left to bottom-right diagonal. The likelihood of a model s_1 and s_2 is not very sensitive to small symmetrical movement of s_1 and s_2 with respect to $o = (0,0)$.

Once the sources s_1^{true} and s_2^{true} get very close, the saddle point in o disappears and turns into a flat crest (Fig. 3.7d). The likelihood becomes insensitive to small variations of s_1 and s_2.

The Fisher information matrix Eq. (3.2) describes the curvature (Hessian) at $\boldsymbol{c}^{true}$ (red dot in Fig. 3.7a). For well-separated sources Fig. 3.7a, the curvature is very high in all directions, resulting in a large determinant of the Hessian matrix, which in turn results in a small variance $\mathrm{var}(d)$ of the distance $d = \left|c_1^{true} - c_2^{true}\right|$ estimation (see Eq. (3.8)). Once the "true" sources get closer, the curvature at the surface maximum decreases leading to larger $\mathrm{var}(d)$. For very close "true" sources, the determinant of Hessian becomes zero, and the lower bound on $\mathrm{var}(d)$ diverges.

The situation of infinitely close sources $c_1^{true} = c_2^{true}$, shown in Fig. 3.7d is equivalent to the situation with one source of double intensity and the second source missing, resulting in a divergence of $\mathrm{var}(d)$ in accordance with our discussion of Eq. (3.14) of the limit $d \to 0$ (see Sect. B.4 in Appendix B).



(a) $d = 300$ nm **(b)** $d = 200$ nm **(c)** $d = 50$ nm **(d)** $d = 0$ nm

(e) $d = 300$ nm **(f)** $d = 200$ nm **(g)** $d = 50$ nm **(h)** $d = 0$ nm

Figure 3.9: Surface of the expected log-likelihood Eq. (3.30) as a function of $\boldsymbol{c} = (c_1, c_2)$ for different separations d between two sources of unequal intensity $\Lambda_2 = 2\Lambda_1$, located at $\boldsymbol{c}^{true}$ (marked with a red dot). The position where the sources exchange their true locations is marked with a blue cross.

The symmetry of the surface breaks when we consider two sources with unequal intensity ($\Lambda_1 \neq \Lambda_2$), because such sources are no longer interchangeable. The situation for $\Lambda_2 = 2\Lambda_1$ is shown in Fig. 3.9. The displacement of the stronger source (s_2), which corresponds to the movement along the vertical lines in Fig. 3.9e-h (see Fig. 3.8 for explanation), has a dramatic effect on the likelihood of the model. The surface drops steeply in the horizontal direction, while it decreases rather slowly along the horizontal line (displacement of the weaker source s_1). For the limit $d \to 0$, shown in Fig. 3.9d,h, the flat crest in the origin still exists (which results in a divergence of var(d)), however it is not aligned with the top-left to bottom-right diagonal as for the equal sources (see Fig. 3.7d,h). There is a non-zero curvature along this diagonal.

The Fisher information for the original FREM$^{\text{orig}}$ formula Eq. (3.3) is derived from the curvature of the surface along the top-left to bottom-right diagonal (symmetrical displacement of the sources with respect to the origin cf. Fig. 3.8). For the symmetrical situation $\Lambda_1 = \Lambda_2$ the original FREM$^{\text{orig}}$ gives the correct results (see Appendix B for a mathematical explanation), however, for the asymmetrical case $\Lambda_1 \neq \Lambda_2$ the non-zero

curvature along the diagonal results in finite FREM$^{\text{orig}}$ even for the limit $d \to 0$. Our proposed derivation of FREM$^{\text{static}}$ from the Fisher information matrix (see Eq. (3.7)) accommodates for the unequal sources correctly and gives diverging FREM$^{\text{static}}$ for this limit.

3.8.2 Blinking vs static sources

Figure 3.2 and Fig. 3.3 suggest that the intensity blinking can facilitate localisation of closely separated sources when compared to the static situation. The difference is more pronounced for data with high signal-to-noise ratio (low b/Λ). Quantum dots with intermittent intensity and an order of magnitude higher brightness than the organic fluorophores are therefore interesting for localisation microscopy even from a theoretical point of view.

3.8.3 Integrating out Λ vs averaging

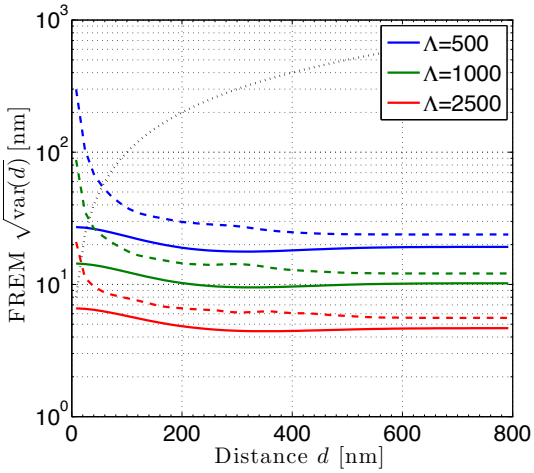


Figure 3.10: Comparison of FREM$^{\text{blink}}$ computed from the Fisher information with integration over the states within the log-likelihood function Eq. (3.23) (dashed lines) and the averaging of the Fisher information over different intensity states Λ^{α} Eq. (3.22) (solid lines). Background was set to $b = 100$ phot/pixel. Dotted black line corresponds to FREM$^{\text{blink}} = d$.

As we pointed out in Sect. 3.5, in the real situation we do not know the intensity states

$\mathbf{\Lambda}^{\alpha}$ of the individual emitters in each frame (see Eq. (3.19)). We have to therefore integrate (sum) over these states within the likelihood function Eq. (3.23), rather than average the Fisher information over different configurations of $\mathbf{\Lambda}^{\alpha}$ as in Eq. (3.22).

To further emphasise the difference between the "integrating over states in" and "averaging of" the Fisher information we plot $\mathrm{FREM}^{\mathrm{blink}}$ as a function of the separation d for both concepts in Fig. 3.10.

$\mathrm{FREM}^{\mathrm{blink}}$ computed from the averaged Fisher information is consistently lower for the whole range of d. It is also lower than the $\mathrm{FREM}^{\mathrm{static}}$ curves for static sources (solid lines in Fig. 3.2b), which in certain region cross the $\mathrm{FREM}^{\mathrm{blink}}$ corresponding to the blinking sources (dashed lines in Fig. 3.2b). In other words, if we *knew* the intensity configurations $\mathbf{\Lambda}^{\alpha}$ of the blinking sources in each recorded frame, we would be able to reach the highest estimation precision. In the blinking situation with the probabilistic description of the intensity states, the integration (summation) over all possible states is required and the estimation precision is lower.

Note, that the averaging approach does not results in divergence of the $\mathrm{FREM}^{\mathrm{blink}}$ for the $d \to 0$ limit (see Fig. 3.10a). This is due to the fact that we assume the configuration of the intensity state in each frame to be *known*. We can therefore determine the position of each source individually from the frames, where only one source is emitting ($\mathbf{\Lambda}^{\alpha=1}$ and $\mathbf{\Lambda}^{\alpha=2}$ in Eq. (3.19)). The averaging Eq. (3.22) fills in the (otherwise zero) diagonal entries of the Fisher information matrix with non-zero values and we get a finite precision for the separation estimation even when $d = 0$.

3.8.4 Scaling of FREM for different levels of intensity and background

The Fisher information matrix Eq. (3.15) for two sources with equal (static) intensity $\Lambda_1 = \Lambda_2 = \Lambda$ suggests that increasing the intensity of the sources by a factor of M, leads to approximately M times higher Fisher information matrix entries (and therefore a $\sqrt{M}$ times lower FREM). More precisely, the dependency of the entries on the intensity

and the background is as follows

$$\left\{ I^{static}_{\Lambda,b} \right\}_{ij} = \Lambda \frac{A}{B + b/\Lambda}, \tag{3.31}$$

where A and B are independent of Λ and b. The subscripts Λ, b in $I_{\Lambda,b}$ express the parametric dependency of the Fisher information on the intensity and background values. Note that the background and the intensity appear as a ratio b/Λ in Eq. (3.31), but the whole expression is multiplied by Λ. An M fold increase of the acquisition time leads to an M fold increase of both the total number of emitted photons and the background b. In this case we get a $\sqrt{M}$ fold decrease of FREM$^{\text{static}}$. For zero background $b = 0$

$$FREM^{\text{static}} \propto \frac{1}{\sqrt{\Lambda}}. \tag{3.32}$$

The scaling with Λ and b of the Fisher information entries for the blinking case is more complicated. In Appendix B we show, that the blinking situation gives results equal to the static situation (up to a factor of 1/2, accounting for half of the total number of emitted photons) for the limit of well-separated sources ($d \to \infty$) and zero background (see Eq. (B.89) in Appendix B). In the non-zero background situation, b appears within the arguments of Poisson terms organised in a complicated fraction (see Eq. (B.84) in Appendix B) and the dependence of the Fisher information matrix entries on the background is therefore highly non-linear.

3.8.5 A note on noise

The Poisson distribution models noise associated with the photon detection [9]. This noise, derived from the nature of the signal itself, is often called "shot noise" [83] and is present even under ideal imaging conditions free of any noise introduced by the sensor. Noise introduced by the sensor is usually divided into two components: the dark noise (or dark current), which represents the electrons thermally generated in the detector, and the read-out noise (or read noise) associated with analogue-to-digital conversion of the signal. The dark noise (Poisson distributed) is efficiently eliminated by the cooling

of the sensor and is negligible in high-performance cameras. The read-out noise is assumed to follow a Gaussian distribution and is characterised by a standard deviation (often called r.m.s.). The read-out noise depends on the read-out frequency and can be reduced by optimising the design of the detector's electrical circuits. The standard deviation of the read-out noise for a typical scientific CCD camera is 5-10 electrons. We assume Poisson distributed data throughout this chapter. We therefore neglect the read-out noise of the sensor and assume only the shot noise to be present in the recorded images. Ram et al. [78] shows a modification of the Fisher information for a model combining both Poisson (shot) and Gaussian (read-out) noise.

3.8.6 Comparison to iNMF results

In Sect. 3.1 I mentioned that the standard deviation lower bound can be achieved only with the "optimal" algorithm. In Sect. 2.9 we discussed on simulated data that iNMF can separate two sources as close as 40 nm for sources with intensity 500 photons/source/frame and background 100 photons/pixel/frame. These parameters correspond to the dashed green curve in Fig. 3.2a. The "natural" resolution limit (intersection with the blacked dotted line - $\mathrm{FREM}^{\mathrm{blink}}$=$d$) for this parameters is ~ 20 nm. We see that iNMF is a somewhat sub-optimal algorithm from this point of view.

3.9 Conclusions

The alternative derivation of the fundamental resolution measure ($\mathrm{FREM}^{\mathrm{static}}$) provides correction to the original formula published by Ram et al. [78]. The results presented in Fig. 3.2 suggest that the blinking sources can significantly increase localisation precision compared to the static situation if the total number of emitted photons is kept equal. The increase of the localisation precision for blinking sources is stronger for close ($d < 50\,\mathrm{nm}$) and bright ($\Lambda > 1000$ photons/source/frame) sources with lower background levels ($b < 100$ photons/pixel/frame). For well-separated sources ($d \to \infty$)

the static and the blinking situations give identical results.

Background has a large impact on the localisation precision. It is desirable to keep background as low as possible. In practice a large proportion of the background intensity comes from the out-of-focus light. Typically TIRF (Total Internal Reflection Fluorescent microscopy) illumination of the sources is used to reduce the out-of-focus blur. However, different techniques such as sparse illumination of the sample can be used.

Note that the Cramér – Rao lower bound approach gives a lower bound for the variance of the distance d estimator. The actual "localisation precision" depends on the specific algorithm we use for localising the individual sources. For example, our proposed LM algorithm discussed in Chapter 2 computationally separates the overlapping sources before localisation. The ability to separate the individual emitters is the limiting factor in this case. It is also important to note that the "localisation" becomes much harder with increasing number of emitters in the sub-diffraction area. The pixellation is taken into account through Eq. (3.11).

In summary bright sources with low background levels (high signal-to-noise ratio) are desirable for the LM techniques. In this setting the blinking provides lower FREM values for closely spaced sources. This makes bright sources with intermittent intensity, such as quantum dots, an interesting candidate for localisation microscopy.

Chapter 4

Line Scan - Structured Illumination Microscopy

In this chapter we discuss a combination of structured illumination microscopy with line scanning. The work has been done in collaboration with Institute of photonic technology (IPHT) Jena in Germany. The work has been recently published in Optics Express [84]:

O. Mandula, M. Kielhorn, K. Wicker, G. Krampert, I. Kleppe, and R. Heintzmann, "Line scan - structured illumination microscopy super-resolution imaging in thick fluorescent samples," Optics Express **20**, 24167 (2012).

G. Krampert, I. Kleppe, and R. Heintzmann has designed the ELYRA setup with scanning head. Martin Kielhorn and I developed the electronic part of the setup synchronising the line scanning with switching of the laser light. I have taken the experimental data and optimised the acquisition procedure. I evaluated the data together with Kai Wicker and Rainer Heintzmann, and I wrote the manuscript. All the authors discussed the manuscript before submission.

4.1 Structured illumination microscopy

Structured illumination microscopy (SIM) is a fluorescence microscopy technique providing images of biological samples with resolution surpassing the classical diffraction limit [85, 25]. SIM requires modification of the illumination part of a wide-field fluorescent microscope such that the sample is illuminated with a spatially varying intensity pattern. The most common pattern used in SIM consists of dense stripes with a sinusoidal profile [86]. This pattern is typically generated by laser light passing through an optical grating (we denote this as the SIM grating) and focused with an objective into a sample [87]. The SIM grating is placed in the plane conjugate to the sample plane and therefore translation and rotation of the optical grating result in translation and rotation of the illumination pattern. High-resolution information is extracted by processing images with different translation (henceforth referred to as phase) and rotation of the illumination pattern [85, 25]. The comprehensive review of the SIM methods can found in [88].

Despite the capability of producing optically sectioned images [89], SIM becomes increasingly challenging when applied to thick ($> 20\,\mu$m) densely labelled fluorescent samples. The wide-field-like (WF) illumination of the SIM system generates a high intensity of out-of- focus fluorescent light. This homogeneous background is added to the spatially modulated fluorescent emission and reduces the pattern modulation in the recorded images Fig. 4.1b. The additive background also increases the noise level, which further deteriorates the quality of the pattern. As a result, the reconstructed images are corrupted by strong noise artifacts (Fig. 4.2c,d and Fig. 4.3b).

SIM with sparse illumination patterns [90] in order to reduce the out-of-focus light has been demonstrated recently [91]. The method was used in conjunction with assigning detected light (pixel reassignment) to the most likely position of the emitter [92, 93]. The sparse patterns were generated by a digital micromirror device (DMD). In this chapter, we propose generation of the sparse illumination patterns by combining the structured illumination with line scanning (LS) microscopy.

In the LS microscopy, the excitation light is focused into a thin line and swept across a fluorescent sample. The out-of-focus light is discarded either by a physical confocal slit (line confocal microscope) or computationally post acquisition [90, 94]. The axial response of a line confocal system is discussed in [95].

The combination of the LS and SIM methods, which is the focus of this manuscript, merges the ability of a line scanning system to physically suppress out-of-focus light with the resolution enhancement of structured illumination. LS-SIM therefore enables high-resolution imaging in thick fluorescent samples. This idea has been demonstrated on simulated data [96] and in this article we show LS-SIM reconstructed images of thick fluorescent sample.

4.2 Experimental methods

4.2.1 Setup

We used a pre-commercial prototype of the ZEISS ELYRA-S inverted microscope (63× /1.4 N.A. objective) with a line-scanning module (ZEISS LSM DuoScan SL). A schematic illustration of the setup is shown in Fig. 4.1. A cylindrical lens was used to focus the laser light (488 nm) into a thin scanning line. The line was focused onto a SIM grating, with bars perpendicular to the scanning line. This produced a sinusoidal intensity modulation along the scanning line in the sample plane (Fig. 4.1g). The phase of the fine sinusoidal SIM pattern was controlled by translation of the SIM grating while the orientation was changed with an image rotator (integral part of the ZEISS ELYRA-S system) positioned between the SIM grating and the objective. This setup therefore avoids the physical rotation of the cylindrical lens and the SIM grating.

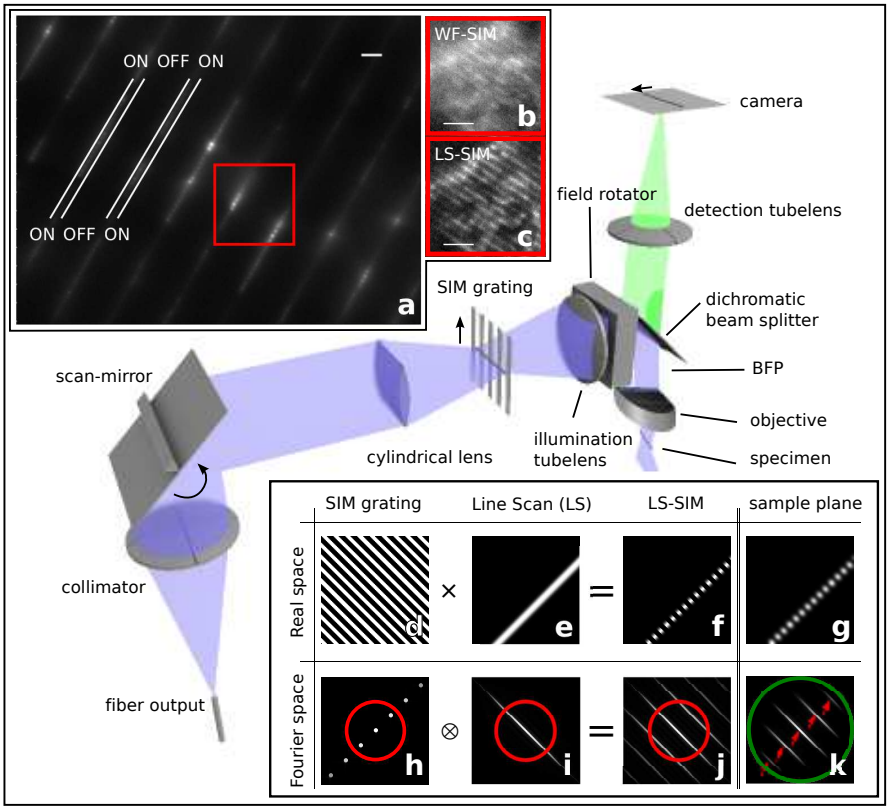


Figure 4.1: Illustration of the LS-SIM setup. The black arrows indicate the movement of the line scan. Upper left inset: (a) one frame of the LS-SIM raw data $I_{n,o,p}$ with indicated ON (illuminated) and OFF (not illuminated) regions (see further explanation in the main text). A close-up of the red box region is shown as (b) a wide-field SIM image I^{WF-SIM} and (c) a line-scan SIM $I_{o,p}^{LS-SIM}$ image. Scale bar (a-c) $2\,\mu$m. Bottom right inset: Illustration of the LS-SIM pattern formation. Real space (top row) corresponds to a SIM grating plane. Fourier space (bottom row) represents the distribution of the intensity in the back focal plane (BFP) of the objective (h, i, j). The aperture of the BFP is indicated as a red circle. The multiplication ($\times$) of the images in real space corresponds to the convolution operation ($\otimes$) in Fourier space. The intensity distribution in the sample plane is shown in (g). The Fourier transform of (g) is shown in (k) with the border of the optical transfer function indicated as a green circle and the position of the Fourier transformed intensity orders of the grating are indicated with red arrows.

4.2.2 Data acquisition

During each camera (Andor iXon DU-885K) acquisition frame of 120 ms, the line was swept across the sample. The scanning galvo-mirror was synchronised with an AOM-based modulation of the excitation laser light such that a sparse periodic pattern consisting of thin bright (ON) lines separated by wide dark (OFF) areas was generated in the sample (Fig. 4.1a). The ratio of the ON area to the total area (ON + OFF), the mark/area ratio (MAR) [86], was set to approximately 1/26 with 32 partially overlapping scan positions of the line. The fine sinusoidal (SIM) modulation was superimposed on each bright (ON) line creating a pattern resembling beads on a string (Fig. 4.1a). The synchronisation was controlled with an Arduino microcontroller (MCNOVE, MultiComp). An illustration of the trigger signals in the microscope is shown in supplementary figure (Appendix C).

For each orientation (o) and phase (p) of the SIM grating we acquired $N = 32$ images $I_{n,o,p}$ with different positions ($n = 1..N$) of the bright (ON) lines (Fig. 4.1a). The ON lines were shifted in each image $I_{n,o,p}$, such that when summed over the N images, they fill the dark areas between the lines. The sum of acquired images therefore corresponds to a standard wide-field SIM (WF-SIM) raw image with slightly reduced modulation of the SIM pattern:

$$I_{o,p}^{WF-SIM} = \sum_{n=1}^{N} I_{n,o,p}. \tag{4.1}$$

A conventional WF image can be computed by summing all WF-SIM frames $I^{WF} \approx \sum_{o,p} I_{o,p}^{WF-SIM}$. However, the instrument can also be switched between the line scanning illumination and the wide-field mode. We therefore acquired genuine WF-SIM data for comparison (Fig. 4.1b) with approximately the same total number of photons per pixel over all necessary images.

We treat the raw data $I_{n,o,p}$ in two distinct steps. In the first step, we take advantage of the line scan to produce background-reduced, optically sectioned images I^{LS-SIM} (Eq. (4.2)) with improved SIM pattern visibility as compared to conventional WF-SIM. In the consecutive step, we extract the high-resolution information from the structured

illumination using conventional WF-SIM treatment of the data.

4.2.3 Data evaluation

A background reduced image with a SIM pattern of superior quality and high modulation (Fig. 4.1c) can be computed from the raw data $I_{n,o,p}$ using [90, 97]

$$I_{o,p}^{LS-SIM} = \max_n(I_{n,o,p}) + \min_n(I_{n,o,p}) - 2\,\mathrm{mean}_n(I_{n,o,p}), \tag{4.2}$$

which is known to yield good results for sparse illumination patterns [97] (low MAR). Equation (4.2) is one of many ways to obtain optical sectioning from the raw data [89, 90], and this equation has been chosen for simplicity and robustness. Images $I_{o,p}^{LS-SIM}$ with different phase translations (p) and orientations (o) of the SIM pattern are passed to a SIM reconstruction algorithm [98] to produce a final reconstructed image shown in Fig. 4.2 and Fig. 4.3a.

We used five phases ($p = 1..5$) and three rotations ($o = 1..3$) of the SIM pattern ($5 \times 3 = 15$ I^{LS-SIM} images). Each I^{LS-SIM} image requires $n = 1..32$ individual scan images $I_{n,o,p}$. Therefore we captured $5 \times 3 \times 32 = 480$ images for one reconstructed plane. The total acquisition time was approximately 75 s with 120 ms acquisition time for each frame $I_{n,o,p}$.

Corresponding line-confocal images (without SIM) can be computed by summing over orientations (o) and phases (p) of the SIM pattern using the preprocessed sectioned data $I^{LS} = \sum_{o,p} I_{o,p}^{LS-SIM}$. The LS image is shown in Fig. 4.2e,f.

4.3 Results

A LS-SIM reconstruction of a Calliphora salivary gland stained with Alexa488-Phalloidin (prepared by Otto Baumann and Eva Simbürger) is shown in Fig. 4.2i. The sample was $\sim 30\,\mu$m thick and the section was taken $\sim 5\,\mu$m below the surface. We chose this specimen because it is relatively thick and densely stained throughout the vol-

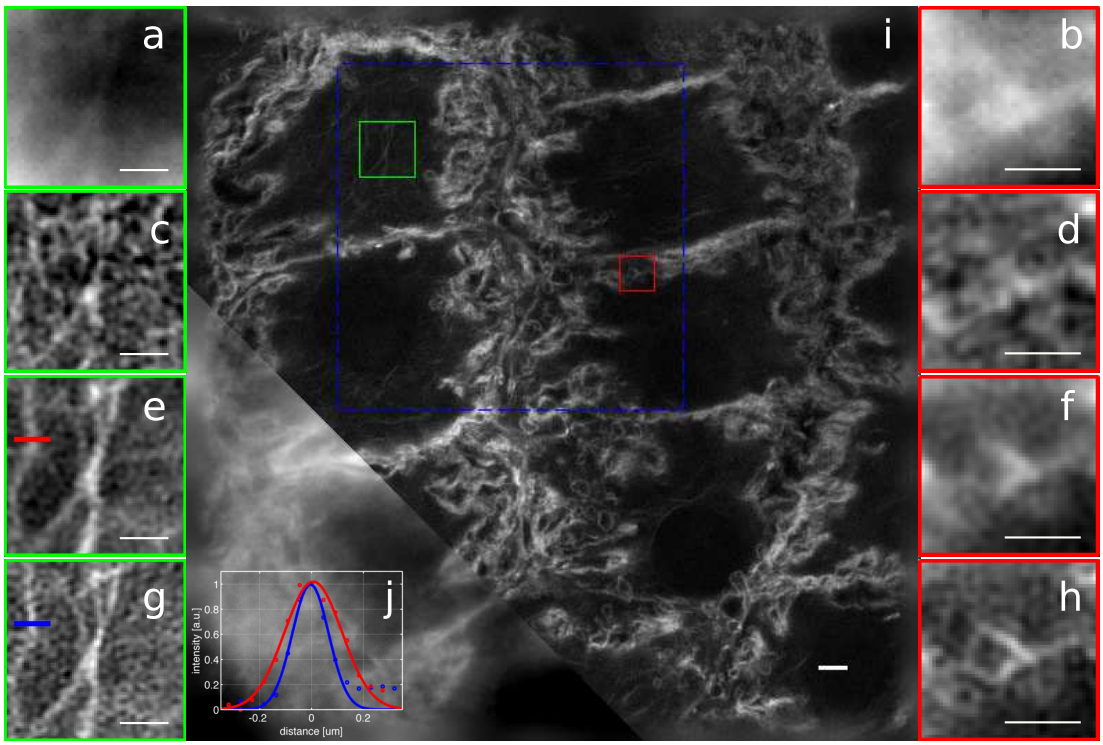


Figure 4.2: LS-SIM reconstructed image of a Calliphora salivary gland (i). WF image (bottom-left corner of (i)) shown for comparison. Scale bar $2\,\mu$m. Selected regions (red, green) reveal details of the actin structures: (a, b) WF image, (c, d) WF-SIM reconstruction, (e, f) LS image, (g, h) LS-SIM reconstruction. Intensity profiles along red line in (e) and blue line in (g) is plotted in (j) in corresponding colours. Scale bar (a-h) $1\,\mu$m.

ume. This gives lot of out-of-focus fluorescence and makes the standard SIM reconstruction difficult. The bottom left corner of Fig. 4.2i shows a wide-field (WF) image for comparison. The out-of-focus light is dramatically reduced in the LS-SIM reconstructed images, revealing complex structural detail within the sample. The fine actin structure remains completely unresolved in the WF image Fig. 4.2a,b. Reconstructed conventional WF-SIM data reveal some structural detail, but the image is severely corrupted by noise artifacts (Fig. 4.2c,d and Fig. 4.3b). The WF-SIM data were taken in the wide-field mode of the microscope prior to the LS-SIM data acquisition with the same number of phases and orientations of the SIM pattern. The acquisition time was $120\,\mathrm{ms/frame}$ and the laser power was adjusted to achieve an approximately similar number of photons/pixels as for the whole series of LS-SIM data. The appropriate

adjustment of the laser power was determined from the previous measurement using a different region of the same sample.

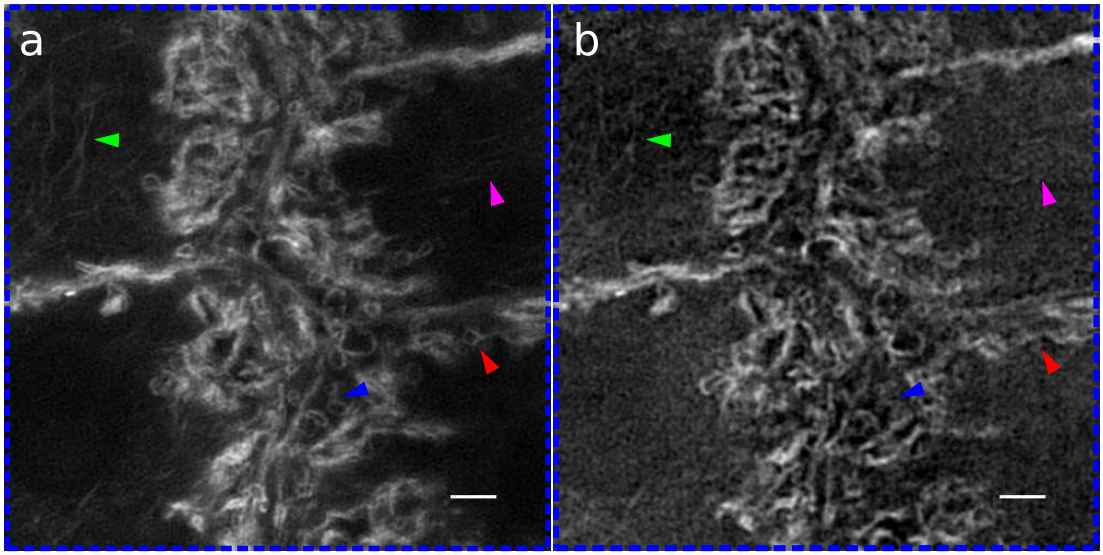


Figure 4.3: Comparison of (a) LS-SIM and (b) a conventional WF-SIM reconstruction of the blue-framed region from Fig. 4.2. LS-SIM image is less affected with noise artifacts, which results in cleaner image. Arrows are pointing to the structures revealed in LS-SIM image. Scale bar 2μm.

The line scan image I^{LS}, corresponds to an image from a line confocal microscope. It provides optical sectioning with strongly reduced out-of-focus light (Fig. 4.2e,f). However, when compared with a wide field image, resolution in the lateral direction has not been improved. In LS-SIM the optical sectioning capability of the line scanning and the resolution improvement of the structured illumination combine to show very fine details of the specimen's inner structure (see Fig. 4.2h and arrows in Fig. 4.3a), unresolved in WF (Fig. 4.2b) and barely visible in WF-SIM (Fig. 4.2d and arrows in Fig. 4.3b) and LS mode (Fig. 4.2f). Due to the higher quality of the SIM pattern in the $I_{o,p}^{LS-SIM}$ images (Fig. 4.1c), the reconstructed images are less affected by noise artifacts compared with the conventional WF-SIM reconstruction. Some of the very dim structures in LS-SIM are completely concealed in the WF-SIM image (magenta arrow in Fig. 4.3 best visible directly on the screen).

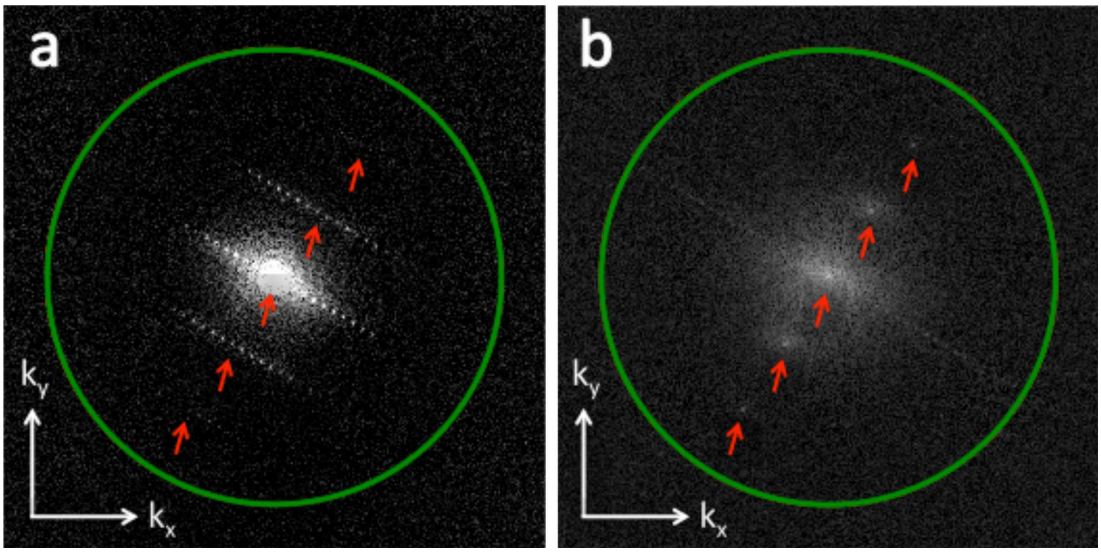


Figure 4.4: Fourier transforms (a) of a single raw scan frame $I_{n,o,p}$ and (b) of a single $I_{n,o,p}^{LS-SIM}$ image computed from Eq. (4.2). Red arrows point at five peaks of the illumination pattern (see Fig. 4.1k). The second diffraction peaks are located at 70% of the cut-off frequency (green circle). The division of the diffraction lines into several points in (a) stems from the illumination of a multitude of lines per exposure.

Intensity profiles measured along a cross-section of a fine, vertically oriented structure revealed in the LS (blue line in Fig. 4.2e) and the LS-SIM image (red line in Fig. 4.2g) are shown in Fig. 4.2j. The cross-section was measured as an average of a stripe five pixels ($\sim 200\,\text{nm}$) in width to provide smoother curves and is plotted with circular marks of corresponding colour in Fig. 4.2j. Gaussian fits are plotted as smooth curves. The estimated full width in half maximum (FWHM) of the LS-SIM profile was 1.6 times smaller than the one extracted from the LS image. This ratio is a rough estimate for the resolution improvement achieved with structured illumination. In our setup, the second diffraction orders were located at about 70% of the BFP aperture radius (Fig. 4.4), which corresponds to the expansion of the cut-off frequency border (i.e. the OTF support) by a factor of 1.7 after image reconstruction Fig. 4.5. This is consistent with the above stated narrowing of the line by a factor of 1.6.

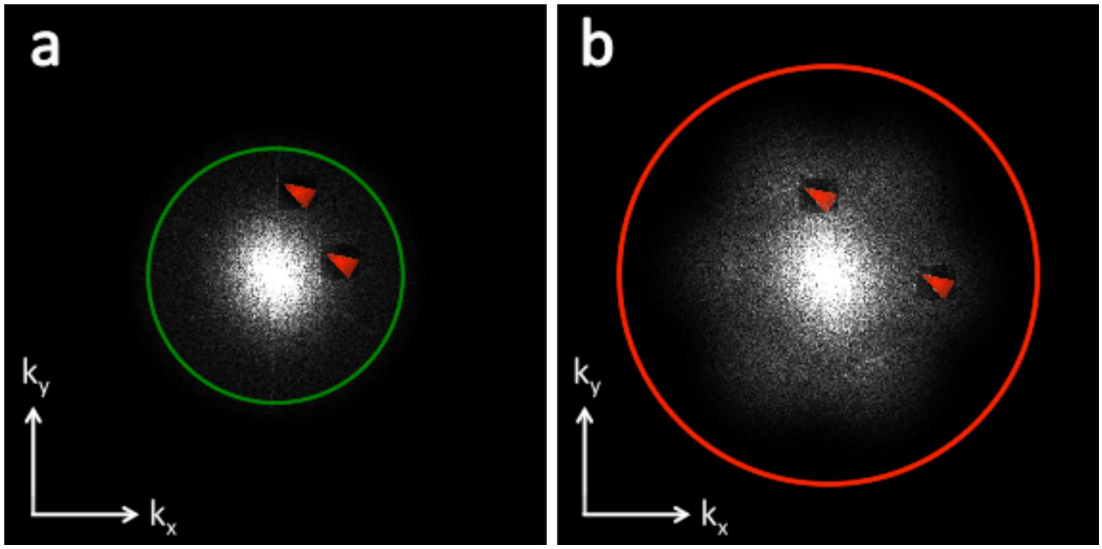


Figure 4.5: Fourier transform of the LS image (a) with a green circle marking the cut-off frequency region. The Fourier transform of the reconstructed LS-SIM image (b) shows the extension of the transferred frequencies. The extended cut-off border is shown as a red circle. The arrows point at suspicious peaks in the spectrum possibly giving rise to the artifacts in the reconstructed image.

4.4 Discussion

The sectioning capability and reconstruction artifact reduction demonstrated in this article are significant advantages of LS-SIM over conventional structured illumination (WF-SIM). Slit scanning systems are known to be well capable of suppressing out-of-focus light [95] and can be operated at quite high speed [99]. Even though, when imaging a fluorescent sheet, the asymptotic intensity decay of slit scan systems is typically inferior to that of pinhole based systems ($1/u$ vs. $1/u^2$ for axial distance u), axial resolutions below a micrometer are achievable as seen in the images presented in [99]. The sectioning performance depends on the size of the illuminated volume - in our case on the width of the ON lines - and on the mark/area ratio (sparsity) of the line illumination pattern (Fig. 4.1a). The pattern must be sufficiently sparse to avoid cross talk between adjacent ON regions, and the ON lines should be sufficiently thin to reduce the illuminated volume in each frame. However, sparser patterns with thinner ON lines require more scans, increasing the total acquisition time. There is therefore a trade-off between the sparseness of the illumination pattern and the acquisition time. In LS-SIM we have to capture 32 times more frames than in standard WF-SIM. This requires a highly stable setup as movement of the sample during acquisition (75 seconds for a single slice) may impair the reconstruction. The LS-SIM reconstruction takes about 3 mins on my laptop computer (Intel® Core™2 Duo @ 2GHz processor with 2GB of RAM).

Noise artifact reduction is associated with an improvement of the signal-to-noise ratio in the line-scanned data. Assuming Poisson noise in the CCD camera images, the variance of the noise is proportional to the signal intensity. Only a small in-focus fraction of the sample volume is illuminated at any given time during the line scan. The fluorophores in the out-of- focus regions of the specimen are excited less, which reduces their contribution to the pixels imaging the fluorescence signal collected from the line-illuminated in-focus region. This physical reduction of the out-of-focus background decreases the total intensity and hence the noise variance in the detected data, render-

ing the SIM reconstruction more accurate. A further refinement of optical sectioning is achieved by application of Eq. (4.2), which has the effect of essentially subtracting the out-of-focus background estimated from the OFF (not illuminated) regions in the raw data (Fig. 4.1a).

Despite the image quality improvement, LS-SIM images (like SIM images) are not always free of artifacts. Arrows in Fig. 4.5b point to suspicious peaks in the reconstructed Fourier transform of the LS-SIM image. These regions are a likely source of the reconstruction artifacts and might be caused by non-uniform bleaching of the sample during the line scanning process as suggested by peaks in the Fourier transform of the LS image (arrows in Fig. 4.4a) or by fluctuations in laser intensity. We tried to reduce the bleaching artifacts by shifting the ON lines (Fig. 4.1a) in consecutive scans not in a natural consecutive order, but in an order maximising the distance between each two consecutive ON regions.

The illustration of the LS-SIM pattern generation process in Fig. 4.1d-k highlights specific features and limitations of the LS-SIM method. The diffraction peaks of the SIM pattern, shown Fig. 4.1h, are smeared into line structures (Fig. 4.1j) due to the line focusing of the illumination (Fig. 4.1e,i). While the first and zero diffraction orders of the SIM grating lie well within the back focal plane (BFP) aperture, as shown in Fig. 4.1h (BFP aperture shown as red circle), their line-illumination versions are spread out along an orthogonally oriented line and then partially blocked (Fig. 4.1j). The strength of the diffracted orders is therefore reduced, which in turn reduces the modulation of the illumination SIM pattern in the sample plane. In other words, the line scanning improves the modulation of the detected pattern (because of optical sectioning) but can reduce the modulation of the actual illumination pattern. This effect becomes stronger for finer SIM gratings, where the diffracted orders are near the border of the BFP aperture. Finer gratings are required for further resolution improvement, but the advantages of LS-SIM become less prominent when the line-shaped diffracted orders (Fig. 4.1j) are nearly at the edge of the BFP aperture.

Note that the whole LS-SIM data can be treated by a SIM reconstruction algorithm assuming a series of 2D illumination patterns [100]. This might provide better sectioning and superior performance for noisy data. However, this requires an excessive amount of computation and therefore we opted for two-step treatment of the data.

The LS and WF images have been up-sampled to the size of the reconstructed images. This has been achieved by padding the corresponding Fourier transform by zeros. To reduce hexagonal "knitting artifacts", a circular window smoothly decaying outside the support of the original optical transfer function was applied in Fourier space prior to padding (see, for example, Fig. 4.5a).

In a recent publication [91] the sample was illuminated with a multitude of point focuses. In this case, the contrast for high illumination frequencies follows the transfer curve of a wide field system and thus almost no contrast is present for the very high frequencies. In contrast, the line-scan SIM illumination used in our work conserves the high-frequency contrast very well up to only shortly below the cut-off frequency. Therefore an enhanced super-resolution capability can be expected in our case. The image-processing scheme in our work is also significantly different from the photon reassignment [93] used in [91].

4.5 Conclusion

LS-SIM provides optically sectioned images of thick fluorescent samples with a significant resolution improvement (1.6 times as measured by a line scan) in the lateral plane. The LS-SIM reconstructed images suffer significantly less from reconstruction artifacts than conventional structured illumination. LS-SIM reveals fine details of a biological specimen's inner structure with higher resolution than line-confocal microscopy and with image quality superior to conventional structure illumination.

Chapter 5

Conclusions and Future Work

5.1 Conclusion

In this thesis we have made three contributions to super-resolution methods for fluorescence microscopy:

We have shown that non-negative matrix factorisation with iterative restarts (iNMF) can separate highly overlapping intermittent sources with arbitrary shape. iNMF is comparable in performance to other recently published methods (CSSTORM and 3B analysis). We introduced average precision (AP) as a quantitative measure for comparing the performance of the iNMF algorithm. AP can be used for data, with known true locatins of the sources (e.g. simulated data). We compared iNMF with CSSTORM and the 3B analysis and demonstrated superior performance of iNMF on simulated data of highly overlapping sources. We described a pipeline for evaluating and visualising realistic datasets, and used iNMF to show super-resolution images of experimental data consisting of tubulin structures labelled with quantum dots. iNMF is a promising and very accessible technique with the potential to deliver super-resolution images of three-dimensional samples.

The combination of structured illumination with line scanning (LS-SIM) presented in this thesis provides images of thick fluorescent samples with resolution improvement

in the lateral plane. Line scanning reduces the out-of-focus background and the LS-SIM images suffer less from reconstruction artefacts when compared to conventional structured illumination. LS-SIM reveals the fine details of biological specimens' inner structures with higher resolution than line-confocal microscopy and with the image quality superior to conventional structure illumination.

In addition we discuss the theoretical resolution limit for noisy and pixelated datasets. We present an alternative derivation of fundamental resolution measure (FREM), correcting the original formula published by Ram et al. [78]. We show that fluorescence intermittency (such as quantum dots blinking) can be beneficial for resolution when compared to the sources with static intensity.

5.2 Future Work

The unique ability of the iNMF algorithm to recover sources with different shapes discussed in Sect. 2.8.7 allows extension of the super-resolution imaging to three dimensional samples. The axial position of a fluorophore can be determined from the shape of the recovered out-of-focus PSF by, for example, determination of the diameter of the outmost ring [74]. However, the conventional out-of-focus PSF decreases quickly in brightness when compared to the in-focus PSF (see Fig. 2.1). This makes it difficult to separate overlapping sources located in different focal planes. The brightness of a tailored PSF, such as the double helix PSF [75] or the PSF with introduced astigmatism [76] is less sensitive to defocus. The out-of-focus PSF remains compact over defocus of several micrometres. On the other hand, the in-focus PSF is less bright than the one in the system without aberrations. The axial position is determined from the specific changes of the PSF shape. Testing the iNMF algorithm on data with a tailored PSF is a logical extension of the current work.

We also want to apply iNMF to specimens labelled with standard organic fluorophores dyes. For example, dSTORM [73] exploits the repetitive transfer of conventional

fluorescent probes between bright ON states and stable and reversible dark OFF states. This results in blinking of the fluorescent sources. The determination of the overlapping sources with iNMF can significantly speed up the data acquisition.

Separate publications from Chapter 2 and Chapter 3 are in preparation.

Appendix A

NMF Algorithm

The classic NMF updates [28] minimise the generalised Kullback – Leibler (KL) divergence between the data matrix $\boldsymbol{D}$ and its factorised version $\boldsymbol{WH}$ (see Eq. (2.7))

$$\mathrm{KL}(\boldsymbol{D} \parallel \boldsymbol{WH}) = -\sum_{xt} \left(d_{xt} \log \sum_{k=1}^{K} w_{xk} h_{kt} - \sum_{k=1}^{K} w_{xk} h_{kt} \right) + C, \tag{A.1}$$

where C is a constant independent on $\boldsymbol{W}$ or $\boldsymbol{H}$. In Sect. 2.3 we show that minimising the KL divergence is equivalent to maximising the likelihood of the model under assumption of Poisson noise (see Eq. (2.6) and Eq. (2.7)). The optimisation can be solved by a scaled gradient descent method:

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{W} - \alpha^{W} \frac{\partial f(\boldsymbol{W},\boldsymbol{H})}{\partial \boldsymbol{W}} \\ \boldsymbol{H} &= \boldsymbol{H} - \alpha^{H} \frac{\partial f(\boldsymbol{W},\boldsymbol{H})}{\partial \boldsymbol{H}}, \end{aligned} \tag{A.2}$$

with respect to the objective function $f(\boldsymbol{W},\boldsymbol{H}) = \mathrm{KL}(\boldsymbol{D} \parallel \boldsymbol{WH})$. The explicit derivation of the objective function f gives

$$\begin{aligned} \frac{\partial f(\boldsymbol{W},\boldsymbol{H})}{\partial w_{xk}} &= \sum_{t=1}^{T} h_{kt} - \left[(\boldsymbol{D} \oslash \boldsymbol{WH}) \boldsymbol{H}^{\top} \right]_{xk} \\ \frac{\partial f(\boldsymbol{W},\boldsymbol{H})}{\partial h_{kt}} &= \sum_{x=1}^{N} w_{xk} - \left[\boldsymbol{W}^{\top} (\boldsymbol{D} \oslash \boldsymbol{WH}) \right]_{xt}, \end{aligned} \tag{A.3}$$

where the symbol "$\oslash$" denotes the element-wise division of matrices.

Lee and Seung [28] proposed

$$\begin{aligned} \alpha_{xk}^{W} &= \frac{w_{xk}}{\sum_{t=1}^{T} h_{kt}} \\ \alpha_{kt}^{H} &= \frac{h_{kt}}{\sum_{x=1}^{N} w_{xk}}, \end{aligned} \tag{A.4}$$

which leads to compact multiplicative updates

$$\begin{aligned} w_{xk} &= \frac{w_{xk}}{\sum_{t=1}^{T} h_{kt}} \left[(\boldsymbol{D} \oslash \boldsymbol{W}\boldsymbol{H})\boldsymbol{H}^{\top} \right]_{xk} \\ h_{kt} &= \frac{h_{kt}}{\sum_{x=1}^{N} w_{xk}} \left[\boldsymbol{W}^{\top} (\boldsymbol{D} \oslash \boldsymbol{W}\boldsymbol{H}) \right]_{kt}. \end{aligned} \tag{A.5}$$

Penalty terms $J^{W}(\boldsymbol{W})$, $J^{H}(\boldsymbol{H})$ can be added to the objective function $f(\boldsymbol{W},\boldsymbol{H})$ to enforce auxiliary constraints [45]:

$$f(\boldsymbol{W},\boldsymbol{H}) = f(\boldsymbol{W},\boldsymbol{H}) + \beta^{W} J^{W}(\boldsymbol{W}) + \beta^{H} J^{H}(\boldsymbol{H}). \tag{A.6}$$

With a choice of

$$\begin{aligned} \alpha_{xk}^{W} &= \frac{w_{xk}}{\sum_{t=1}^{T} h_{kt} + \beta^{W} \frac{\partial J^{W}}{\partial w_{xk}}} \\ \alpha_{kt}^{H} &= \frac{h_{kt}}{\sum_{x=1}^{N} w_{xk} + \beta^{H} \frac{\partial J^{H}}{\partial h_{kt}}}, \end{aligned} \tag{A.7}$$

the multiplicative updates change to

$$\begin{aligned} w_{xk} &= \frac{w_{xk}}{\sum_{t=1}^{T} h_{kt} + \beta^{W} \frac{\partial J^{W}}{\partial w_{xk}}} \left[(\boldsymbol{D} \oslash \boldsymbol{W}\boldsymbol{H})\boldsymbol{H}^{\top} \right]_{xk} \\ h_{kt} &= \frac{h_{kt}}{\sum_{x=1}^{N} w_{xk} + \beta^{H} \frac{\partial J^{H}}{\partial h_{kt}}} \left[\boldsymbol{W}^{\top} (\boldsymbol{D} \oslash \boldsymbol{W}\boldsymbol{H}) \right]_{kt}. \end{aligned} \tag{A.8}$$

Appendix B

Resolution Limit for Blinking QDs

This is derivation of the Fisher information for Poisson distributed variable X with mean $\lambda(\theta)$:

$$X \sim \mathrm{Po}(n;\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}. \tag{B.1}$$

B.1 Likelihood

Likelihood of the Poisson distributed variable with detection n_k in K pixels:

$$l(\theta) = \prod_{k=1}^{K} l_k = \prod_{k=1}^{K} \frac{\lambda_k^{n_k} e^{-\lambda_k}}{n_k!}, \tag{B.2}$$

where $l_k(\theta) = p(n_k|\theta)$ to emphasise the dependency on the parameter θ.

Log-Likelihood:

$$\mathcal{L} = \sum_{k=1}^{K} \mathcal{L}_k, \tag{B.3}$$

where

$$\mathcal{L}_k = n_k \log \lambda_k - \lambda_k - \log n_k!. \tag{B.4}$$

B.2 Fisher Information

The Fisher information can be expression in these equivalent forms:

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}}{\partial \theta^2}\right] = \mathbb{E}\left[\left(\frac{\partial \mathcal{L}}{\partial \theta}\right)^2\right] = \mathbb{E}\left[\left(\sum_k \frac{\partial \log(l_k)}{\partial \theta}\right)^2\right] = \mathbb{E}\left[\left(\sum_k \frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\right)^2\right]. \tag{B.5}$$

Therefore

$$I(\theta) = \mathbb{E}\left[\left(\sum_k \frac{\partial \mathcal{L}_k}{\partial \theta}\right)\left(\sum_m \frac{\partial \mathcal{L}_m}{\partial \theta}\right)\right] \tag{B.6}$$

$$= \mathbb{E}\left[\sum_k \left(\frac{\partial \mathcal{L}_k}{\partial \theta}\right)^2\right] + \mathbb{E}\left[\sum_k \sum_{m\neq k} \frac{\partial \mathcal{L}_k}{\partial \theta}\frac{\partial \mathcal{L}_m}{\partial \theta}\right]. \tag{B.7}$$

Because n_k are independent identically distributed (i.i.d.), the second term can be expressed as

$$\mathbb{E}\left[\sum_k \sum_{m\neq k} \frac{\partial \mathcal{L}_k}{\partial \theta}\frac{\partial \mathcal{L}_m}{\partial \theta}\right] = \sum_k \sum_{m\neq k} \mathbb{E}_k\left[\frac{\partial \mathcal{L}_k}{\partial \theta}\right]\mathbb{E}_m\left[\frac{\partial \mathcal{L}_m}{\partial \theta}\right], \tag{B.8}$$

where

$$\mathbb{E}_k\left[f(n_k)\right] = \sum_{n_k \geq 0} p(n_k|\theta) f(n_k). \tag{B.9}$$

But

$$\mathbb{E}_k\left[\frac{\partial \mathcal{L}_k}{\partial \theta}\right] = \mathbb{E}_k\left[\frac{1}{l_k}\frac{\partial l_k}{\partial \theta}\right] = \sum_{n_k} l_k \frac{1}{l_k}\frac{\partial l_k}{\partial \theta} = \sum_{n_k}\frac{\partial l_k}{\partial \theta} = \frac{\partial}{\partial \theta}\sum_{n_k} p(n_k|\theta) = 0, \tag{B.10}$$

as $\sum_{n_k} p(n_k|\theta) = 1$.

The Fisher Information can then be written as

$$I(\theta) = \mathbb{E}\left[\sum_k \left(\frac{\partial \mathcal{L}_k}{\partial \theta}\right)^2\right] \tag{B.11}$$

and expressing the derivatives from Eq. (B.4)

$$\frac{\partial \mathcal{L}_k}{\partial \theta} = \left(\frac{n_k - \lambda_k}{\lambda_k}\right)\frac{\partial \lambda_k}{\partial \theta}, \tag{B.12}$$

we get for the Fisher information

$$I(\theta) = \mathbb{E}\left[\sum_k \frac{(n_k - \lambda_k)^2}{\lambda_k^2}\left(\frac{\partial \lambda_k}{\partial \theta}\right)^2\right] \tag{B.13}$$

$$= \sum_k \frac{1}{\lambda_k^2}\left(\frac{\partial \lambda_k}{\partial \theta}\right)^2 \mathbb{E}\left[(n_k - \lambda_k)^2\right]. \tag{B.14}$$

We recognise the variance $\mathrm{var}(n_k) = \mathbb{E}\left[(n_k - \lambda_k)^2\right]$. For Poisson variable

$$\mathrm{var}(n_k) = \mathrm{mean}(n_k) = \lambda_k, \tag{B.15}$$

and therefore the Fisher Information for Poisson distributed data becomes

$$I(\theta) = \sum_{k=1}^{K} \frac{1}{\lambda_k} \left(\frac{\partial \lambda_k}{\partial \theta} \right)^2. \tag{B.16}$$

B.3 Two sources separated by a distance d

This section comments on the Fisher Information estimation as described by Ram et al. in [78].

For two sources separated by a distance d we have the mean value of the intensity:

$$\lambda(x) = \Lambda_1 f_1(x) + \Lambda_2 f_2(x), \tag{B.17}$$

where f_i and Λ_i is the response function and the intensity, respectively, of the source i. For translationally invariant PSF and in-focus sources: $f_1 = q(x - \frac{d}{2})$ and $f_2 = q(x + \frac{d}{2})$

$$\lambda(d) = \Lambda_1 q(x - \frac{d}{2}) + \Lambda_2 q(x + \frac{d}{2}), \tag{B.18}$$

where q is the PSF of the system. For pixelated version (integral over pixel area Γ_k) with homogeneous background b in each pixel the intensity can be expressed as:

$$\lambda_k(d) = \Lambda_1 \int_{\Gamma_k} q(x - \frac{d}{2}) dx + \Lambda_2 \int_{\Gamma_k} q(x + \frac{d}{2}) dx + b. \tag{B.19}$$

By plugging into Eq. (B.16) we obtain expression for the Fisher Information:

$$I(d) = \frac{1}{4} \sum_{k=1}^{K} \frac{\left(\Lambda_1 \int_{\Gamma_k} \partial_x q(x - \frac{d}{2}) dx - \Lambda_2 \int_{\Gamma_k} \partial_x q(x + \frac{d}{2}) dx \right)^2}{\Lambda_1 \int_{\Gamma_k} q(x - \frac{d}{2}) dx + \Lambda_2 \int_{\Gamma_k} q(x + \frac{d}{2}) dx + b} \tag{B.20}$$

$$= \frac{1}{4} \sum_{k=1}^{N} \frac{\left[\Lambda_1 q_k'(-\frac{d}{2}) - \Lambda_2 q_k'(\frac{d}{2}) \right]^2}{\Lambda_1 q_k(-\frac{d}{2}) + \Lambda_2 q_k(\frac{d}{2}) + b}, \tag{B.21}$$

where we have set $q_k(z) = \int_{\Gamma_k} q(x - z) dx$ as the pixelated version of the point spread function translated by z. Γ_k is the area of the kth pixel, and $q_k'(z) = \int_{\Gamma_k} \frac{\partial q(x-z)}{\partial x} dx$ is the corresponding pixelated derivative.

FREM is defined as a lower bound on the standard deviation ($\sqrt{\text{var}(d)}$) of the source separation d estimator

$$\sqrt{\text{var}(d)} \geq \text{FREM} = \sqrt{I^{-1}(\theta)}. \tag{B.22}$$

Limit $d = 0$:

If $\Lambda_1 = \Lambda_2$ then $I(d = 0) = 0$ which means $\text{var}(d = 0) \to \infty$. However, for unequal sources $\Lambda_1 \neq \Lambda_2$ this does not hold and the variance remains finite.

Limit $d \to \infty$:

When the sources are far apart, the mixing term in the nominator in Eq. (B.21) vanishes: $\Lambda_1\Lambda_2\partial_x q(x - \frac{d}{2})\partial q(d + \frac{d}{2}) = 0$ because $\partial_x q(x - \frac{d}{2})$ and $\partial_x q(x + \frac{d}{2})$ do not have common overlap. Equation (B.21) then decomposes into two individual terms

$$I(d) = \frac{1}{4}\sum_{k=1}^{K}\left[\frac{\left(\Lambda_1 q_k'(x - \frac{d}{2})\right)^2}{\Lambda_1 q_k(x - \frac{d}{2}) + b} + \frac{\left(\Lambda_2 q_k'(x + \frac{d}{2})\right)^2}{\Lambda_2 q_k(x + \frac{d}{2}) + b}\right] \tag{B.23}$$

This corresponds to the sum of the Fisher Information for localisation of individual sources.

Situation with missing source $\Lambda_i = 0,\ \Lambda_j \neq 0$:

Even if one of the source is missing the Fisher information is strangely non-zero $I(d) \neq 0$. The variance remains finite even if one of the sources is not present! This is a consequence of the assumption that the sources are located symmetrically at $\pm d/2$ with respect to the origin. Only one source is then needed to determine the distance $d/2$.

B.4 An alternative way to derive FREM

Below we show how to fix the problems with the original FREM. Our FREM gives infinite variance when one of the sources is not present and fixes the strange behaviour

of the unequal sources for the limit $d \to 0$.

We consider two sources located at c_1 and c_2, respectively. The expectation of the intensity is therefore expressed as (cf. Eq. (B.18)):

$$\lambda_k(\boldsymbol{c}) = \Lambda_1 q_k(x - c_1) + \Lambda_2 q_k(x - c_2) + b. \tag{B.24}$$

The distance between the sources is $d = c_1 - c_2$. This is a linear combination $\boldsymbol{a}^T \cdot \boldsymbol{c}$ of the variable $\boldsymbol{c} = (c_1, c_2)^T$, where $\boldsymbol{a} = (1, -1)^T$. The variance of d is given by

$$\mathrm{var}(d) = \mathrm{var}(\boldsymbol{a}^T \cdot \boldsymbol{c}) = \boldsymbol{a}^T \cdot \boldsymbol{Q} \cdot \boldsymbol{a} = Q_{11} + Q_{22} - 2Q_{12}, \tag{B.25}$$

where $\boldsymbol{Q}$ is the covariance matrix. The lower bound on $\boldsymbol{Q}$ is expressed as the inverse of the Fisher information matrix

$$\boldsymbol{Q} \geq \boldsymbol{I}^{-1}(\theta), \tag{B.26}$$

where the Fisher information matrix

$$\boldsymbol{I}(\theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22}, \end{pmatrix} \tag{B.27}$$

is given by generalisation of Eq. (B.16)

$$I_{ij}(\theta) = \sum_{k=1}^{K} \frac{1}{\lambda_k} \frac{\partial \lambda_k}{\partial c_i} \frac{\partial \lambda_k}{\partial c_j}. \tag{B.28}$$

Note that the Fisher information matrix is symmetrical $I_{12} = I_{21}$ due to exchangeability of the derivatives.

The covariance matrix $\boldsymbol{Q}$ is therefore

$$\boldsymbol{Q} \geq \boldsymbol{I}^{-1}(\theta) = \frac{1}{I_{11}I_{22} - I_{12}^2} \begin{pmatrix} I_{22} & -I_{12} \\ -I_{12} & I_{11} \end{pmatrix} \tag{B.29}$$

and the variance of $d = c_1 - c_2$ estimator

$$\mathrm{var}(d) = (1, -1)^T \cdot \boldsymbol{Q} \cdot (1, -1) \tag{B.30}$$

$$\geq \frac{I_{11} + I_{22} + 2I_{12}}{I_{11}I_{22} - I_{12}^2} = \frac{p}{r}. \tag{B.31}$$

The individual terms of the Fisher Information matrix by using Eq. (B.24) in Eq. (B.28)

$$I_{ij} = \Lambda_i \Lambda_j \sum_{k=1}^{K} \frac{q'_k(c_i) q'_k(c_j)}{f_k(c_1, c_2)}, \tag{B.32}$$

where $q_k(c_i)$ is the pixelated version (pixel area Γ_k) of the PSF

$$q_k(c_i) = \int_{\Gamma_k} q(x - c_i) dx \tag{B.33}$$

$$q'_k(c_i) = \int_{\Gamma_k} \frac{\partial q(x - c_i)}{\partial x} dx \tag{B.34}$$

and $f_k(c_1, c_2) = \Lambda_1 q_k(c_1) + \Lambda_2 q_k(c_2) + b$.

Then the numerator $p = I_{11} + I_{22} + 2I_{12}$ in Eq. (B.31) is given by

$$p = \sum_{k=1}^{K} \frac{1}{f_k(c_1, c_2)} \left[\Lambda_1^2 q'^2{}_k(c_1) + \Lambda_2^2 q'^2_k(c_2) + 2\Lambda_1 \Lambda_2 q'_k(c_1) q'_k(c_2) \right]. \tag{B.35}$$

The terms in the denominator $r = I_{11} I_{22} - I_{12}^2$ in Eq. (B.31) are given by

$$I_{11} I_{22} = \Lambda_1^2 \Lambda_2^2 \sum_{k,l}^{K} \frac{\left(q'_k(c_1) q'_l(c_2) \right)^2}{f_k(c_1, c_2) f_l(c_1, c_2)} \tag{B.36}$$

$$I_{12}^2 = \Lambda_1^2 \Lambda_2^2 \sum_{k,l}^{K} \frac{q'_k(c_1) q'_k(c_2) q'_l(c_1) q'_l(c_2)}{f_k(c_1, c_2) f_l(c_1, c_2)} \tag{B.37}$$

We now consider the limits of very close sources ($d \to 0$) and well-separated emitters ($d \to \infty$) for FREM.

Close sources limit: $d \to 0 \Rightarrow c_1 \to c_2 \Rightarrow q_k(c_1) \to q_k(c_2)$**:**

$$p = (\Lambda_1^2 + \Lambda_2^2 + 2\Lambda_1 \Lambda_2) \sum_{k=1}^{K} \frac{q'^2_k(c)}{f_k(c, c)}, \tag{B.38}$$

which can be further simplified by explicitly substituting $f_k(c_1, c_2)$

$$p = (\Lambda_1 + \Lambda_2) \sum_{k=1}^{K} \frac{q'^2_k(c)}{q_k(c) + b/(\Lambda_1 + \Lambda_2)}. \tag{B.39}$$

q_k and $(q'_k)^2$ are strictly positive functions, therefore the sum is not zero and p is non-zero for any Λ_1, Λ_2.

The two terms in the denominator in Eq. (B.31) are identical for $c_1 = c_2$

$$I_{11} I_{22} = I_{12}^2 \tag{B.40}$$

and therefore

$$r = I_{11}I_{22} - I_{12}^2 = \det[\boldsymbol{I}(\theta)] \equiv 0 \tag{B.41}$$

for any Λ_i. $\boldsymbol{I}(\theta)$ is therefore a singular matrix for $d = 0$ and inverse $\boldsymbol{I}^{-1}(\theta)$ does not exist for $c_1 = c_2$. However, the limit $c_1 \to c_2, (d \to 0)$ gives $p \neq 0, r \to 0$ and FREM$= \frac{p}{r} \to \infty$. FREM therefore diverges for sources with any combination of source intensities Λ_1 and Λ_2. This is in contrast to the original FREM formula Eq. (B.21) which gives diverging FREM only for equally strong sources $\Lambda_1 = \Lambda_2$.

Well-separated sources limit $d \to \infty$:

The cross term I_{ij} in Eq. (B.32) vanishes ($I_{ij} = 0$, $i \neq j$) because of the multiplication $q'_k(c_1)q'_k(c_2)$, which is very close to zero for well-separated PSFs. This assumes that the PSF (and its first derivative) decreases to negligible values for distance far from the centre of the PSF. Then from Eq. (B.31)

$$\mathrm{var}(d) \geq \frac{1}{I_{11}} + \frac{1}{I_{22}}, \tag{B.42}$$

which is the sum of bounds on variances for localisation of two individual sources:

$$I_{ii} = \Lambda_i^2 \sum_{k=1}^{K} \frac{q_k'^2(c_i)}{\Lambda_1 q_k(c_1) + \Lambda_2 q_k(c_2) + b} \tag{B.43}$$

$$= \Lambda_i \sum_{k=1}^{K} \frac{q_k'^2(c_i)}{q_k(c_i) + b/\Lambda_i}. \tag{B.44}$$

We used the fact that the PSFs $q(c_1)$ and $q(c_2)$ are well-separated and decrease (with its first derivatives) towards zero in the regions far from their centre. Therefore the $q(c_j)$ is negligible in the region where $q'(c_i)$ $(j \neq i)$ have any significant values. The term $\Lambda_j q(c_j)$ in the denominator can be therefore neglected.

If we use Gaussian approximation of the PSF (see Zhang et al. [8])

$$q(x - c_i) = \frac{1}{Z} \exp\left(-\frac{(x - c_i)^2}{2\sigma^2}\right) \tag{B.45}$$

with derivatives with respect to c_i

$$q'(x-c_i) = \frac{x-c_i}{\sigma^2}\frac{1}{Z}\exp\left(-\frac{(x-c_i)^2}{2\sigma^2}\right) \tag{B.46}$$

$$= \frac{x-c_i}{\sigma^2}q(x-c_i), \tag{B.47}$$

then from Eq. (B.44) for the situation with negligible background $b/\Lambda \ll 1$

$$I_{ii} = \frac{\Lambda_i}{\sigma^4}\sum_k \int_{\Gamma_k} (x-c_i)^2 q(x-c_i)dx. \tag{B.48}$$

Now using

$$\sum_k \int_{\Gamma_k} (x-c_i)^2 q(x-c_i)dx = \int_{\mathbb{R}^2} (x-c_i)^2 q(x-c_i)dx \tag{B.49}$$

$$= \sigma^2, \tag{B.50}$$

we obtain the terms of the Fisher information matrix

$$I_{ii} = \frac{\Lambda_i}{\sigma^2}. \tag{B.51}$$

These terms correspond to the localisation of the individual sources. The "localisation precision" for one sources s_i is then bounded by $\sigma/\sqrt{\Lambda_i}$. This corresponds to the "squeezing" of the initial " localisation uncertainty" σ by the square root of the source's intensity.

For well-separated sources is then the lower bound on the variance give by

$$\mathrm{var}(d \to \infty) \geq \sigma^2\left(\frac{1}{\Lambda_1}+\frac{1}{\Lambda_2}\right). \tag{B.52}$$

One sources missing: $\Lambda_i = 0,\ \Lambda_j \neq 0$**:**

$I_{ii} \equiv 0$ and $I_{ij} \equiv 0$ and so $\det(\boldsymbol{I}(\theta)) \equiv 0$, and the matrix is singular. In the limit $\Lambda_i \to 0$ the variance Eq. (B.31) $\mathrm{var}(d) \to \infty$. This is again in contrast to the original FREM computed from Eq. (B.21), which gives finite FREM even for one source missing (see discussion of limits in Sect. B.3).

B.5 Comparison of the original FREM with our version

For equally strong sources ($\Lambda_1 = \Lambda_2 = \Lambda$) the original FREM formula gives identical results as ours. In this situation Eq. (B.32) gives equality of the diagonal terms $I_{11} = I_{22}$ for any c_1 and c_2. From Eq. (B.31)

$$\mathrm{var}(d) \geq \frac{2(I_{11}+I_{12})}{I_{11}^2 - I_{12}^2} \tag{B.53}$$

$$\geq \frac{2}{I_{11}-I_{12}}. \tag{B.54}$$

Using $q_k(c_1) = q_k(-d/2)$ and $q_k(c_2) = q_k(+d/2)$ we can rewrite the original FREM expression Eq. (B.21) using Eq. (B.32)

$$\mathrm{var}^{ORIG}(d) \geq \frac{4}{I_{11}-2I_{12}+I_{22}}, \tag{B.55}$$

which for $\Lambda_1 = \Lambda_2$ ($I_{11} = I_{22}$) reduces to

$$\mathrm{var}^{ORIG}(d) \geq \frac{2}{I_{11}-I_{12}}. \tag{B.56}$$

Comparison with Eq. (B.54) shows the equality of the both formulas for the situation of equally strong sources $\Lambda_1 = \Lambda_2$.

The expressions leads to different results for sources with unequal intensity. If $\Lambda_2 = \alpha\Lambda_1$, then can be shown, that for negligible values of b/Λ_2 the ratio between the original FREM and our proposed formula for $d \to \infty$ is $2\sqrt{2}/3$.

B.6 Time distribution of the intensities - averaging

We assume the likelihood dependent on parameter $\mathbf{\Lambda}$ (for example, $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2)$ - intensity of two sources in the recorded frame). If we *knew the configuration* of $\mathbf{\Lambda}$ we would write the log-likelihood

$$\mathcal{L}(\theta, \Lambda) = \sum_{k=1}^{K} \log\left(l_k(\theta, \mathbf{\Lambda})\right). \tag{B.57}$$

Derivatives with respect to the parameter θ:

$$\frac{\partial \mathcal{L}(\theta,\mathbf{\Lambda})}{\partial \theta} = \sum_{k=1}^{K} \frac{\partial \log\left(l_k(\theta,\mathbf{\Lambda})\right)}{\partial \theta} \tag{B.58}$$

$$= \frac{\partial \mathcal{L}(\theta,\mathbf{\Lambda})}{\partial \theta}. \tag{B.59}$$

If we assume a probability distribution $p(\mathbf{\Lambda})$ of the intensity states $\mathbf{\Lambda}$ we can express the Fisher information (see Eq. (B.5)):

$$I(\theta) = \int_{\mathbf{\Lambda}} p(\mathbf{\Lambda}) I_{\mathbf{\Lambda}}(\theta) d\mathbf{\Lambda}, \tag{B.60}$$

where $I_{\mathbf{\Lambda}}(\theta)$ is the Fisher information computed for a specific value of $\mathbf{\Lambda}$ (see Eq. (B.5)). For discrete states of $\mathbf{\Lambda}$, for example

$$\mathbf{\Lambda} = \left\{ \mathbf{\Lambda}^{\alpha=1}, \mathbf{\Lambda}^{\alpha=2}, \mathbf{\Lambda}^{\alpha=3}, \mathbf{\Lambda}^{\alpha=4} \right\} \tag{B.61}$$

$$= \{[\Lambda_1, 0], [\Lambda_2, 0], [\Lambda_1, \Lambda_2], [0,0]\} \tag{B.62}$$

we get

$$I(\theta) = \sum_{\alpha} p(\mathbf{\Lambda}^{\boldsymbol{\alpha}}) I_{\mathbf{\Lambda}^{\boldsymbol{\alpha}}}(\theta), \tag{B.63}$$

where the Fisher Information for every configuration of $\mathbf{\Lambda}^{\boldsymbol{\alpha}}$ is averaged with weights $p(\mathbf{\Lambda}^{\boldsymbol{\alpha}})$.

B.7 Time distribution of the intensities - integrating out Λ

If we *do not know the configuration* of $\mathbf{\Lambda}^{\boldsymbol{\alpha}}$ in each frame, then we have to rely only on the distribution $p(\Lambda)$ and integrate over it within the likelihood function:

$$l_k(\theta) = \int_{\Lambda} l_k(\theta,\Lambda) d\Lambda = \int_{\Lambda} p(n_k|\theta,\Lambda) p(\Lambda) d\Lambda \tag{B.64}$$

We assume four state model of two sources: $\{(\Lambda_1,0),(0,\Lambda_2),(\Lambda_1,\Lambda_2),(0,0)\}$:

$$\lambda_k^{\alpha=1} = \Lambda_1 q_k(x-c_1) + b, \tag{B.65}$$

$$\lambda_k^{\alpha=2} = \Lambda_2 q_k(x-c_2) + b, \tag{B.66}$$

$$\lambda_k^{\alpha=3} = \Lambda_1 q_k(x-c_1) + \Lambda_2 q_k(x-c_2) + b, \tag{B.67}$$

$$\lambda_k^{\alpha=4} = + b, \tag{B.68}$$

with uniform distribution over these states. We used uniform background intensity b in each pixel of each frame.

Assuming $p(\mathbf{\Lambda^{\alpha}}) = 1/4$ for $\alpha = 1,..4$, then from Eq. (B.64):

$$l_k(\theta) = \frac{1}{4}\sum_{\alpha=1}^{4} \mathrm{Po}(\lambda_k^{\alpha}), \tag{B.69}$$

with derivatives

$$\frac{\partial l_k}{\partial c_p} = \frac{1}{4}\sum_{\alpha}\frac{\partial \mathrm{Po}(\lambda_k^{\alpha})}{\partial c_p} = \frac{1}{4}\sum_{\alpha}\left(\mathrm{Po}(\lambda_k^{\alpha})\frac{(n_k-\lambda_k^{\alpha})}{\lambda_k^{\alpha}}\frac{\partial \lambda_k^{\alpha}}{\partial c_p}\right). \tag{B.70}$$

The diagonal entries of the Fisher information matrix are give by:

$$I_{pp}(\boldsymbol{c}) = \mathbb{E}\left[\left(\sum_{k=1}^{N}\frac{1}{l_k}\frac{\partial l_k}{\partial c_p}\right)^2\right] \tag{B.71}$$

$$= \mathbb{E}\left[\left\{\sum_{k=1}^{N}\left(\frac{1}{\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_k^{\alpha})}\frac{\partial\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_k^{\alpha})}{\partial c_p}\right)\right\}\left\{\sum_{l=1}^{N}\left(\frac{1}{\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_l^{\alpha})}\frac{\partial\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_l^{\alpha})}{\partial c_p}\right)\right\}\right] \tag{B.72}$$

$$= \sum_{k=1}^{N}\mathbb{E}_k\left[\frac{\left(\sum_{\alpha=1}^{4}\frac{\partial \mathrm{Po}(\lambda_k^{\alpha})}{\partial c_p}\right)^2}{\left(\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_k^{\alpha})\right)^2}\right], \tag{B.73}$$

because the cross terms (k, l) in the sum (2nd row) are zeros:

$$\mathbb{E}\left[\left(\frac{\sum_{\alpha=1}^{4}\frac{\partial \mathrm{Po}(\lambda_k^{\alpha})}{\partial c_p}}{\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_k^{\alpha})}\right)\left(\frac{\sum_{\alpha=1}^{4}\frac{\partial \mathrm{Po}(\lambda_l^{\alpha})}{\partial c_p}}{\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_l^{\alpha})}\right)\right] = \mathbb{E}_k\left[\frac{\sum_{\alpha=1}^{4}\frac{\partial \mathrm{Po}(\lambda_k^{\alpha})}{\partial c_p}}{\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_k^{\alpha})}\right]\mathbb{E}_l\left[\frac{\sum_{\alpha=1}^{4}\frac{\partial \mathrm{Po}(\lambda_l^{\alpha})}{\partial c_p}}{\sum_{\alpha=1}^{4}\mathrm{Po}(\lambda_l^{\alpha})}\right] \tag{B.74}$$

$$= \sum_{\alpha=1}^{4}\frac{\partial}{\partial c_p}\left(\sum_{n_k\geq 0}\mathrm{Po}(\lambda_k^{\alpha})\right)\sum_{\alpha=1}^{4}\frac{\partial}{\partial c_p}\left(\sum_{n_k\geq 0}\mathrm{Po}(\lambda_l^{\alpha})\right) \tag{B.75}$$

$$= 0 \tag{B.76}$$

Expressing the derivatives and the expectation from Eq. (B.73) we can write for the diagonal entries of the Fisher information matrix:

$$I_{pp}(\boldsymbol{c}) = \sum_{k=1}^{N} \mathbb{E}_k \left[\left\{ \frac{\sum_{\alpha=1}^{4} \left(\mathrm{Po}(n_k;\lambda_k^\alpha) \frac{(n_k-\lambda_k^\alpha)}{\lambda_k^\alpha} \frac{\partial \lambda_k^\alpha}{\partial c_p} \right)}{\sum_{\alpha=1}^{4} \mathrm{Po}(n_k;\lambda_k^\alpha)} \right\}^2 \right] \tag{B.77}$$

$$= \frac{1}{4} \sum_{k=1}^{N} \sum_{n_k \geq 0} \frac{\left\{ \sum_{\alpha=1}^{4} \left(\mathrm{Po}(n_k;\lambda_k^\alpha) \frac{(n_k-\lambda_k^\alpha)}{\lambda_k^\alpha} \frac{\partial \lambda_k^\alpha}{\partial c_p} \right) \right\}^2}{\sum_{\alpha=1}^{4} \mathrm{Po}(n_k;\lambda_k^\alpha)} \tag{B.78}$$

For the four states model we have $\lambda^{\alpha=3}(c_1,c_2) = \lambda^{\alpha=1}(c_1) + \lambda^{\alpha=2}(c_2) - b$ and so $\frac{\partial \lambda^{\alpha=3}}{\partial c_p} = \frac{\partial \lambda^{\alpha=p}}{\partial c_p}$ and $\frac{\partial \lambda^{\alpha=j}}{\partial c_p} = 0$, $j \neq p$ for $p = \{1,2\}$, $j = \{1,2,4\}$; so

$$I_{pp}(\boldsymbol{c}) = \frac{1}{4} \sum_{k=1}^{N} \left(\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p} \right)^2 \sum_{n_k \geq 0} \frac{\left\{ \sum_{\alpha=\{p,3\}} \left(\mathrm{Po}(n_k;\lambda_k^\alpha) \frac{(n_k-\lambda_k^\alpha)}{\lambda_k^\alpha} \right) \right\}^2}{\sum_{\alpha=1}^{4} \mathrm{Po}(n_k;\lambda_k^\alpha)} \tag{B.79}$$

The off-diagonal entries of the Fisher information matrix are given by:

$$I_{pq}(\boldsymbol{c}) = \sum_{k=1}^{N} \mathbb{E}_k \left[\frac{\left(\sum_{\alpha=1}^{4} \frac{\partial \mathrm{Po}(\lambda_k^\alpha)}{\partial c_p} \right) \left(\sum_{\alpha=1}^{4} \frac{\partial \mathrm{Po}(\lambda_k^\alpha)}{\partial c_q} \right)}{\left(\sum_{\alpha=1}^{4} \mathrm{Po}(\lambda_k^\alpha) \right)^2} \right] \tag{B.80}$$

$$= \frac{1}{4} \sum_{k=1}^{N} \left(\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p} \right) \left(\frac{\partial \lambda_k^{\alpha=q}}{\partial c_q} \right) \tag{B.81}$$

$$\times \sum_{n_k \geq 0} \frac{\left(\sum_{\alpha=\{p,3\}} \mathrm{Po}(n_k;\lambda_k^\alpha) \frac{(n_k-\lambda_k^\alpha)}{\lambda_k^\alpha} \right) \left(\sum_{\alpha=\{q,3\}} \mathrm{Po}(n_k;\lambda_k^\alpha) \frac{(n_k-\lambda_k^\alpha)}{\lambda_k^\alpha} \right)}{\sum_{\alpha=1}^{4} \mathrm{Po}(n_k;\lambda_k^\alpha)} \tag{B.82}$$

Limit $d \to 0$:

If $c^1 = c^2$ then $\lambda^{\alpha=1} = \lambda^{\alpha=2}$ and $\frac{\partial \mathrm{Po}(\lambda^{\alpha=1})}{\partial c^1} = \frac{\partial \mathrm{Po}(\lambda^{\alpha=2})}{\partial c^2}$. Then all entries in I_{pq} are equal and the matrix is singular. For the limit $d \to 0$ the determinant $\det(\boldsymbol{I}) \to 0$ and the variance $\mathrm{var}(d) \to \infty$.

Limit $d \to \infty$:

Sources are far apart and $\lambda^{\alpha=1}$ and $\lambda^{\alpha=2}$ do not have a common overlap. For k' where $\frac{\partial \lambda_{k'}^{\alpha=p}}{\partial c_p} \neq 0$, $\frac{\partial \lambda_{k'}^{\alpha=q}}{\partial c_p} \equiv 0$ and $\mathrm{Po}(n_{k'},\lambda_{k'}^{\alpha=3}) = \mathrm{Po}(n_{k'},\lambda_{k'}^{\alpha=1})$. Also $\sum_\alpha \mathrm{Po}(\lambda_k^\alpha) = 2\mathrm{Po}(\lambda_k^{\alpha=p}) + 2\mathrm{Po}(b)$ in the region where $\frac{\partial \lambda^{\alpha=p}}{\partial c_p} \neq 0$.

From Eq. (B.73) the cross terms vanish ($I_{pq} = 0$ because $\frac{\partial \mathrm{Po}(\lambda^{\alpha=p})}{\partial c_p}\frac{\partial \mathrm{Po}(\lambda^{\alpha=q})}{\partial c_q} = 0$). The diagonal elements

$$I_{pp} = \sum_{k=1}^{N} \mathbb{E}_k \left[\frac{\left(2\frac{\partial \mathrm{Po}(\lambda_k^{\alpha=p})}{\partial c_p}\right)^2}{\left(2\mathrm{Po}(\lambda_k^{\alpha=p}) + 2\mathrm{Po}(b)\right)^2} \right] \tag{B.83}$$

$$= \sum_{k=1}^{N} \mathbb{E}_k \left[\frac{\left(\mathrm{Po}(\lambda_k^{\alpha=p})\frac{\left(n_k - \lambda_k^{\alpha=p}\right)}{\lambda_k^{\alpha=p}}\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p}\right)^2}{\left(\mathrm{Po}(\lambda_k^{\alpha=p}) + \mathrm{Po}(b)\right)^2} \right] \tag{B.84}$$

for $b = 0$ ($\mathrm{Po}(b) = 0$):

$$I_{pp} = \sum_{k=1}^{N} \left(\frac{1}{\lambda_k^{\alpha=p}}\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p}\right)^2 \mathbb{E}_k\left[\left(n_k - \lambda_k^{\alpha=p}\right)^2\right] \tag{B.85}$$

$$= \sum_{k=1}^{N} \left(\frac{1}{\lambda_k^{\alpha=p}}\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p}\right)^2 \frac{1}{4}\sum_{n_k \geq 0}\left(\sum_{i=1}^{4} \mathrm{Po}(\lambda_k^{\alpha})\left(n_k - \lambda_k^{\alpha=p}\right)^2\right) \tag{B.86}$$

$$= \sum_{k=1}^{N} \left(\frac{1}{\lambda_k^{\alpha=p}}\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p}\right)^2 \frac{1}{4}\sum_{n_k \geq 0}\left(2\mathrm{Po}(\lambda_k^{\alpha=p})\left(n_k - \lambda_k^{\alpha=p}\right)^2\right) \tag{B.87}$$

$$= \sum_{k=1}^{N} \left(\frac{1}{\lambda_k^{\alpha=p}}\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p}\right)^2 \frac{1}{2}\lambda_k^{\alpha=p} \tag{B.88}$$

$$= \frac{1}{2}\sum_{k=1}^{N} \frac{1}{\lambda_k^{\alpha=p}}\left(\frac{\partial \lambda_k^{\alpha=p}}{\partial c_p}\right)^2 \tag{B.89}$$

which is the expression for the static sources Eq. (B.16), up to a factor $1/2$. The factor 1/2 comes from the fact that the source appears only in 50% of the observations. If we keep the number of photons constant in both blinking and the static case (by reducing the intensity of the static sources by factor of two) we get identical value of $\mathrm{var}(d)$ for the $d \to \infty$.

For non zero background $b > 0$ we cannot simplify Eq. (B.84) due to the background term $\mathrm{Po}(b)$ in the denominator. However, as the term is positive, the element I_{pp} will be decreasing with increasing background. The background makes therefore the variance $\mathrm{var}(d)$ bigger as we would expect.

B.8 Comments on the "dip" in FREM curves for static sources

These are several comments about the strange "dip" in the FREM curves for static sources (see Fig. 3.5a). The expression for computing FREM for two sources with equal intensity is from Eq. (3.16)

$$\mathrm{var}(d) = \left[\frac{I_{11} - I_{22}}{2}\right]^{-1}, \tag{B.90}$$

and the individual entries of the Fisher information matrix is from Eq. (B.32):

$$I_{ij} = \Lambda \sum_{k=1}^{K} \frac{q_k'(c_i) q_k'(c_j)}{q_k(c_1) + q_k(c_2) + b/\Lambda}; \;\; i,j = \{1,2\}, \tag{B.91}$$

where q_k and q_k' are pixelated PSF and derivative, respectively.

For large background values ($b/\Lambda \gg \max[q_k(c_i)]$) the bottom term is nearly constant and therefore

$$I_{ij} \approx C \sum_{k=1}^{K} q_k'(c_i) q_k'(c_j), \tag{B.92}$$

where C is a constant.

From Eq. (B.90)

$$\mathrm{var}(d) \approx \left[C/2 \sum_{k} \left(q_k'(c_1)^2 - q_k'(c_1) q_k'(c_2)\right)\right]^{-1} \tag{B.93}$$

For Gaussian approximation of the PSF, the expression can be integrated analytically. The resulting curve sows the identical "dip" in the curves.

Appendix C

LS-SIM Electronics

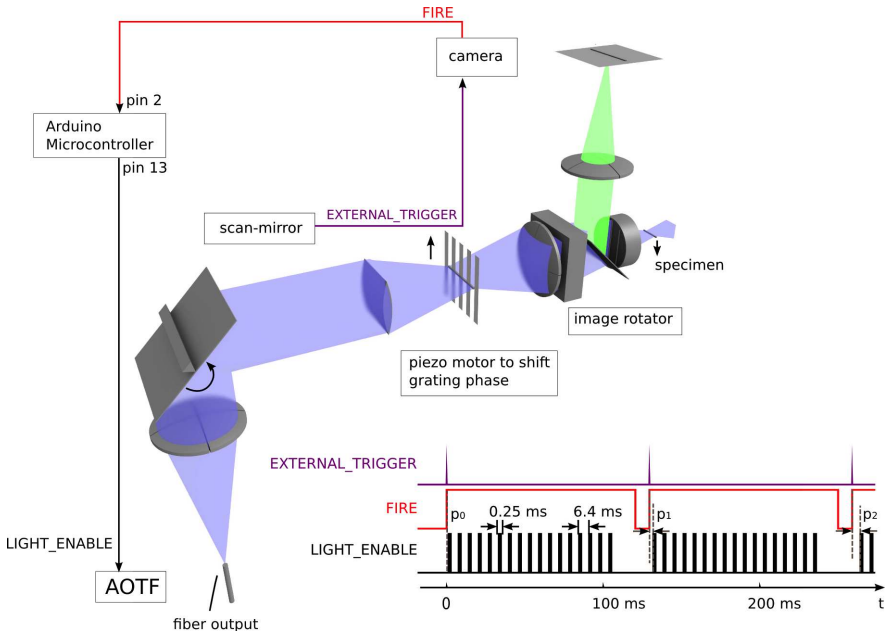


Figure C.1: Schematic of the trigger signals in the microscope. The scan-mirror is the master and triggers the camera when it starts moving in one edge of the field. Upon receiving EXTERNAL_TRIGGER the camera starts integrating in less than a microsecond and sets FIRE to high. This triggers an interrupt in the micro-controller and it generates LIGHT_ENABLE to modulate the laser. The initial p_i delays determine the position of the scanning lines. The values in the timing diagram (bottom right) were measured with a logic analyser (Logic16, Salea). These values correspond to an optimised timing of the LS-SIM system, such that the sum of the 32 line positions imprinted on the sample produced a homogeneous wide-field illumination.

Bibliography

[1] Ernst Abbe. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv fur mikroskopische Anatomie*, 9(1):413–418, 1873. (Not cited.)

[2] Frits Zernike. Phase contrast, a new method for the microscopic observation of transparent objects. *Physica*, 9(7):686–698, 1942. (Not cited.)

[3] Alexander Egner, Claudia Geisler, Claas von Middendorff, Hannes Bock, Dirk Wenzel, Rebecca Medda, Martin Andresen, Andre C Stiel, Stefan Jakobs, Christian Eggeling, Andreas Schönle, and Stefan W Hell. Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophysical journal*, 93(9):3285–90, November 2007. (Not cited.)

[4] Stefan W Hell. Far-field optical nanoscopy. *Science (New York, N.Y.)*, 316(5828):1153–8, May 2007. (Not cited.)

[5] Stefan W. Hell, Roman Schmidt, and Alexander Egner. Diffraction-unlimited three-dimensional optical nanoscopy with opposing lenses. *Nature Photonics*, 3(7):381–387, July 2009. (Not cited.)

[6] M Born and E Wolf. *Principles of optics*, volume 1. Cambridge University Press, 1999. (Not cited.)

[7] B. A. Scalettar, J. R. Swedlow, J. W. Sedat, and D. A. Agard. Dispersion, aberration and deconvolution in multi-wavelength fluorescence images. *Journal of Microscopy*, 182(1):50–60, April 1996. (Not cited.)

[8] Bo Zhang, Josiane Zerubia, and Jean-Christophe Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10):1819–1829, 2007. (Not cited.)

[9] James B. Pawley, editor. *Handbook Of Biological Confocal Microscopy*, volume 13. Springer US, Boston, MA, third edition, 2006. (Not cited.)

[10] Marta Fernández-Suárez and Alice Y Ting. Fluorescent probes for super-resolution imaging in living cells. *Nature reviews. Molecular cell biology*, 9(12):929–43, December 2008. (Not cited.)

[11] Roger Y. Tsien. The green fluorescent protein. *Annual review of biochemistry*, 67:509–44, January 1998. (Not cited.)

[12] H. Bock, C. Geisler, C.a. Wurm, C. von Middendorff, S. Jakobs, a. Schönle, a. Egner, S.W. Hell, and C. Eggeling. Two-color far-field fluorescence nanoscopy based on photoswitchable emitters. *Applied Physics B*, 88(2):161–165, July 2007. (Not cited.)

[13] Liisa Hirvonen, Ondrej Mandula, Kai Wicker, and Rainer Heintzmann. Structured illumination microscopy using photoswitchable fluorescent proteins. In *Proceedings of SPIE*, volume 6861, pages 68610L–68610L–8. SPIE, 2008. (Not cited.)

[14] E. H. Rego, L. Shao, J. J. Macklin, L. Winoto, G. a. Johansson, N. Kamps-Hughes, M. W. Davidson, and M. G. L. Gustafsson. PNAS Plus: Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution. *Proceedings of the National Academy of Sciences*, 1:1–9, December 2011. (Not cited.)

[15] Armand Paul Alivisatos. Semiconductor Clusters, Nanocrystals, and Quantum Dots. *Science*, 271(5251):933–937, February 1996. (Not cited.)

[16] Jyoti K. Jaiswal and Sanford M. Simon. Potentials and pitfalls of fluorescent quantum dots for biological imaging. *Trends in cell biology*, 14(9):497–504, September 2004. (Not cited.)

[17] Ute Resch-Genger, Markus Grabolle, Sara Cavaliere-Jaricot, Roland Nitschke, and Thomas Nann. Quantum dots versus organic dyes as fluorescent labels. *Nature methods*, 5(9):763–75, September 2008. (Not cited.)

[18] Maureen Walling, Jennifer Novak, and Jason R. E. Shepard. Quantum dots for live cell and in vivo imaging. *International journal of molecular sciences*, 10(2):441–91, February 2009. (Not cited.)

[19] M. Kuno, D. P. Fromm, H. F. Hamann, A. Gallagher, and D. J. Nesbitt. ”On”/”off” fluorescence intermittency of single semiconductor quantum dots. *The Journal of Chemical Physics*, 115(2):1028, 2001. (Not cited.)

[20] Fernando D. Stefani, Jacob P. Hoogenboom, and Eli Barkai. Beyond quantum jumps: Blinking nanoscale light emitters. *Physics Today*, 62(2):34, 2009. (Not cited.)

[21] Eric Betzig, George H. Patterson, Rachid Sougrat, O. Wolf Lindwasser, Scott Olenych, Juan S. Bonifacino, Michael W. Davidson, Jennifer Lippincott-Schwartz, and Harald F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science (New York, N.Y.)*, 313(5793):1642–5, September 2006. (Not cited.)

[22] Samuel T. Hess, Thanu P. K. Girirajan, and Michael D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–72, December 2006. (Not cited.)

[23] Michael J. Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793–5, October 2006. (Not cited.)

[24] Thomas Dertinger, Ryan Colyer, G. Iyer, S. Weiss, and J. Enderlein. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proceedings of the National Academy of Sciences of the United States of America*, 106(52):22287–92, December 2009. (Not cited.)

[25] Mats G. L. Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy*, 198(Pt 2):82–7, May 2000. (Not cited.)

[26] Rainer Heintzmann, Thomas M. Jovin, and Christoph Cremer. Saturated patterned excitation microscopy-a concept for optical resolution improvement. *Journal of the Optical Society of America A*, 19(8):1599, 2002. (Not cited.)

[27] Stefan W. Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–2, June 1994. (Not cited.)

[28] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13, 2001. (Not cited.)

[29] Raimund J. Ober, Sripad Ram, and E. Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical journal*, 86(2):1185–1200, February 2004. (Not cited.)

[30] Keith A. Lidke, Bernd Rieger, Thomas M. Jovin, and Rainer Heintzmann. Superresolution by localization of quantum dots using blinking statistics. *Optics Express*, 13(18):7052, 2005. (Not cited.)

[31] Alexander R. Small. Theoretical limits on errors and acquisition rates in localizing switchable fluorophores. *Biophysical journal*, 96(2):L16–8, January 2009. (Not cited.)

[32] Keith A. Lidke and Rainer Heintzmann. Localization fluorescence microscopy using quantum dot blinking. In *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 936–939. IEEE, 2007. (Not cited.)

[33] Paul Harrington, Jonas Anderson, Bernd Rieger, Diane Lidke, and Keith A Lidke. Poster: A Bayesian Approach to Fluorescence Intermittency Based Localization Microscopy. *Supplement of Biophysical Journal*, 96:20–20, 2008. (Not cited.)

[34] Paul D Simonson, Eli Rothenberg, and Paul R Selvin. Single-molecule-based super-resolution images in the presence of multiple fluorophores. *Nano letters*, 11(11):5090–6, November 2011. (Not cited.)

[35] Fang Huang, Samantha L Schwartz, Jason M Byars, and Keith A. Lidke. Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomedical optics express*, 2(5):1377–93, January 2011. (Not cited.)

[36] Seamus J Holden, Stephan Uphoff, and Achillefs N Kapanidis. DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature methods*, 8(4):279–80, April 2011. (Not cited.)

[37] Susan Cox, Edward Rosten, James Monypenny, Tijana Jovanovic-Talisman, Dylan T Burnette, Jennifer Lippincott-Schwartz, Gareth E Jones, and Rainer Heintzmann. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature methods*, 9(2):195–200, January 2011. (Not cited.)

[38] Lei Zhu, Wei Zhang, Daniel Elnatan, and Bo Huang. Faster STORM using compressed sensing. *Nature methods*, 9(7):721–3, April 2012. (Not cited.)

[39] K. Shimizu, R. Neuhauser, C. Leatherdale, S. Empedocles, W. Woo, and M. Bawendi. Blinking statistics in single semiconductor nanocrystal quantum dots. *Physical Review B*, 63(20):1–5, May 2001. (Not cited.)

[40] Thomas Dertinger, Ryan Colyer, Robert Vogel, Jörg Enderlein, and Shimon Weiss. Achieving increased resolution and more pixels with Superresolution Optical Fluctuation Imaging (SOFI). *Optics Express*, 18(18):18875, August 2010. (Not cited.)

[41] Stefan Geissbuehler, Noelia L Bocchio, Claudio Dellagiacoma, Corinne Berclaz, Marcel Leutenegger, and Theo Lasser. Mapping molecular statistics with balanced super-resolution optical fluctuation imaging (bSOFI). *Optical Nanoscopy*, 1(1):4, 2012. (Not cited.)

[42] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994. (Not cited.)

[43] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–91, October 1999. (Not cited.)

[44] David Donoho. When does non-negative matrix factorization give a correct decomposition into parts. *Advances in neural information processing*, 2004. (Not cited.)

[45] Michael W. Berry, Murray Browne, Amy N. Langville, V. Paul Pauca, and Robert J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics \& Data Analysis*, 52(1):155–173, September 2007. (Not cited.)

[46] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. *The Journal of Machine Learning Research*, 5:1457–1469, 2004. (Not cited.)

[47] Zhe Chen and Andrzej Cichocki. Nonnegative Matrix Factorization with Temporal Smoothness and / or Spatial Decorrelation Constraints. *Signal Processing*, 2005. (Not cited.)

[48] V. Paul Pauca, J. Piper, and Robert J. Plemmons. Nonnegative matrix factorization for spectral data analysis. *Linear Algebra and its Applications*, 416(1):29–47, 2006. (Not cited.)

[49] John Canny. GaP: a factor model for discrete data. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 122–129. ACM, 2004. (Not cited.)

[50] Karthik Devarajan. Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. *PLoS computational biology*, 4(7):e1000029, January 2008. (Not cited.)

[51] Philip M Kim and Bruce Tidor. Subsystem identification through dimensionality reduction of large-scale gene expression data. *Genome research*, 13(7):1706–18, July 2003. (Not cited.)

[52] A. Heger and L. Holm. Sensitive pattern discovery with 'fuzzy' alignments of distantly related proteins. *Bioinformatics*, 19(Suppl 1):i130–i137, July 2003. (Not cited.)

[53] Jean-Philippe Brunet, Pablo Tamayo, Todd R Golub, and Jill P Mesirov. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America*, 101(12):4164–9, March 2004. (Not cited.)

[54] A Rabinovich, S Agarwal, and CA Laris. Unsupervised color decomposition of histologically stained tissue samples. *Advances in Neural Information Processing Systems*, 2003. (Not cited.)

[55] Richard A Neher, Miso Mitkovski, Frank Kirchhoff, Erwin Neher, Fabian J Theis, and André Zeug. Blind source separation techniques for the decomposition of multiply labeled fluorescence images. *Biophysical journal*, 96(9):3791–800, May 2009. (Not cited.)

[56] Paritosh Pande, Brian E Applegate, and Javier a Jo. Application of non-negative matrix factorization to multispectral FLIM data analysis. *Biomedical optics express*, 3(9):2244–62, September 2012. (Not cited.)

[57] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks : the official journal of the International Neural Network Society*, 13(4-5):411–30, 2000. (Not cited.)

[58] William Hadley Richardson. Bayesian-Based Iterative Method of Image Restoration. *Journal of the Optical Society of America*, 62(1):55, January 1972. (Not cited.)

[59] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745, June 1974. (Not cited.)

[60] Timothy J Holmes. Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach: errata. *J. Opt. Soc. Am. A*, 9(11):2097, 1992. (Not cited.)

[61] Sarang Joshi and Michael I Miller. Maximum α posteriori estimation with {G}ood's roughness for three-dimensional optical-sectioning microscopy. *J. Opt. Soc. Am. A*, 10(5):1078–1085, 1993. (Not cited.)

[62] D A Fish, A M Brinicombe, E. R. Pike, and J. G. Walker. Blind deconvolution by means of the {R}ichardson-{L}ucy algorithm. *Journal of the Optical Society of America A*, 12(1):58, January 1995. (Not cited.)

[63] G. M. P. van Kempen, L. J. van Vlie, P. J. Verveer, and H. T. M. van der Voort. A quantitative comparison of image restoration methods for confocal microscopy. *Journal of Microscopy*, 185(3):354–365, March 1997. (Not cited.)

[64] P. J. Verveer, M. J. Gemkow, and T. M. Jovin. A comparison of image restoration approaches applied to three-dimensional confocal and wide-field fluorescence microscopy. *Journal of Microscopy*, 193(1):50–61, January 1999. (Not cited.)

[65] B.D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (methodological)*, 39(2):172–212, 1977. (Not cited.)

[66] Gerard Salton and Michael J McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., 1986. (Not cited.)

[67] Mark Everingham, Luc Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338, September 2009. (Not cited.)

[68] Jingu Kim and Haesun Park. Toward Faster Nonnegative Matrix Factorization: A New Algorithm and Comparisons. *2008 Eighth IEEE International Conference on Data Mining*, pages 353–362, December 2008. (Not cited.)

[69] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. (Not cited.)

[70] Wray Buntine and Aleks Jakulin. Discrete component analysis. In C. Saunders, M. Grobelnik, S. Gunn, and J. Shawe-Taylor, editors, *Subspace, Latent Structure and Feature Selection*, pages 1–33. Springer, 2006. (Not cited.)

[71] Ondrej Mandula, Rainer Heintzmann, and Christopher K. I. Williams. Poster: Localisation Microscopy with Quantum Dots using Non-negative Matrix Factorisation. *Society for Neuroscience (SfN)*, 2010. (Not cited.)

[72] B Roy Frieden. *Probability, statistical optics, and data testing: a problem solving approach*. Springer series in information sciences ;10. Springer-Verlag, Berlin, 1991. (Not cited.)

[73] Sebastian van de Linde, Anna Löschberger, Teresa Klein, Meike Heidbreder, Steve Wolter, Mike Heilemann, and Markus Sauer. Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nature protocols*, 6(7):991–1009, July 2011. (Not cited.)

[74] Michael Speidel, Alexandr Jonás, and Ernst-Ludwig Florin. Three-dimensional tracking of fluorescent nanoparticles with subnanometer precision by use of off-focus imaging. *Optics letters*, 28(2):69–71, January 2003. (Not cited.)

[75] Sean Quirin, Sri Rama Prasanna Pavani, and Rafael Piestun. Optimal 3D single-molecule localization for superresolution microscopy with aberrations and engineered point spread functions. Supporting Information. *Proceedings of the National Academy of Sciences of the United States of America*, 109(3):675–9, January 2012. (Not cited.)

[76] Bo Huang, Wenqin Wang, Mark Bates, and Xiaowei Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science (New York, N.Y.)*, 319(5864):810–3, February 2008. (Not cited.)

[77] Hari Shroff, C.G. Galbraith, J.A. Galbraith, and Eric Betzig. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nature Methods*, 5(5):417–423, 2008. (Not cited.)

[78] Sripad Ram, E. Sally Ward, and Raimund J. Ober. Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4457–62, March 2006. (Not cited.)

[79] Sripad Ram, E. Sally Ward, and Raimund J. Ober. A Stochastic Analysis of Performance Limits for Optical Microscopes. *Multidimensional Systems and Signal Processing*, 17(1):27–57, January 2006. (Not cited.)

[80] C. Radhakrishna Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull Calcutta Math Soc*, 37:81–91, 1945. (Not cited.)

[81] Thomas M. Cover and Joy A. Thomas. Elements of Information Theory. *Book*, (Wiley Series in Telecommunications):542, 1991. (Not cited.)

[82] Ian Nabney and Christopher M. Bishop. Netlab Neural Network Software \url{www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/}. (Not cited.)

[83] W Schottky. Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. *Annalen der Physik*, 362(23):541–567, 1918. (Not cited.)

[84] Ondrej Mandula, Martin Kielhorn, Kai Wicker, Gerhard Krampert, Ingo Kleppe, and Rainer Heintzmann. Line scan - structured illumination microscopy super-resolution imaging in thick fluorescent samples. *Optics Express*, 20(22):24167, October 2012. (Not cited.)

[85] Rainer Heintzmann and Christoph Cremer. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. In *Proceedings of SPIE*, volume 3568, pages 185–196. SPIE, 1999. (Not cited.)

[86] Rainer Heintzmann. Structured Illumination Methods. In James B. Pawley, editor, *Handbook of biological confocal microscopy*, chapter 13, pages 265–279. Springer, New York, third edition, 2006. (Not cited.)

[87] Rainer Heintzmann and Mats G. L. Gustafsson. Subdiffraction resolution in continuous samples. *Nature Photonics*, 3(7):362–364, July 2009. (Not cited.)

[88] Aurélie Jost and Rainer Heintzmann. Superresolution Multidimensional Imaging with Structured Illumination Microscopy. *Annual Review of Materials Research*, 43(1), September 2012. (Not cited.)

[89] Lothar Schermelleh, Peter M Carlton, Sebastian Haase, Lin Shao, Lukman Winoto, Peter Kner, Brian Burke, M Cristina Cardoso, David A. Agard, Mats G L Gustafsson, Heinrich Leonhardt, and John W Sedat. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science (New York, N.Y.)*, 320(5881):1332–6, June 2008. (Not cited.)

[90] Pier A Benedetti, V. Evangelista, D. Guidarini, and S. Vestri. U.S. Patent 6016367, 1997. (Not cited.)

[91] Andrew G York, Sapun H Parekh, Damian Dalle Nogare, Robert S Fischer, Kelsey Temprine, Marina Mione, Ajay B Chitnis, Christian A Combs, and Hari Shroff. Resolution doubling in live, multicellular organisms via multifocal structured illumination microscopy - supplementary. *Nature methods*, 9(7):749–54, May 2012. (Not cited.)

[92] I. J. Cox, Colin J. R. Sheppard, and Tony Wilson. Improvement in resolution by nearly confocal microscopy. *Applied Optics*, 21(5):778, March 1982. (Not cited.)

[93] I. J. Cox, Colin J. R. Sheppard, and Tony Wilson. Super-resolution by confocal fluorescent microscopy. *Optik*, 60(4):391–396, 1982. (Not cited.)

[94] Mark A. Neil, Rimas Juskaitis, and Tony Wilson. Method of obtaining optical sectioning by using structured light in a conventional microscope. *Optics letters*, 22(24):1905–7, December 1997. (Not cited.)

[95] Vincent Poher, Gordon T. Kennedy, Hugh B. Manning, Dylan M. Owen, Haoxiang X. Zhang, Erdan Gu, Martin D. Dawson, Paul M W French, and Mark A A Neil. Improved sectioning in a slit scanning confocal microscope. *Optics letters*, 33(16):1813–5, August 2008. (Not cited.)

[96] Taejoong Kim, DaeGab Gweon, and Jun-Hee Lee. Enhancement of fluorescence confocal scanning microscopy lateral resolution by use of structured illumination. *Measurement Science and Technology*, 20(5):055501, May 2009. (Not cited.)

[97] Rainer Heintzmann and Pier A Benedetti. High-resolution image reconstruction in fluorescence microscopy with patterned excitation. *Applied optics*, 45(20):5037–45, July 2006. (Not cited.)

[98] Kai Wicker. *Increasing resolution and light efficiency in fluorescence microscopy*. PhD thesis, King's College London, 2010. (Not cited.)

[99] Edward J Botcherby, Martin J Booth, Rimas Juskaitis, and Tony Wilson. Real-time slit scanning microscopy in the meridional plane. *Optics letters*, 34(10):1504–6, May 2009. (Not cited.)

[100] Rainer Heintzmann. Saturated patterned excitation microscopy with two-dimensional excitation patterns. *Micron*, 34(6-7):283–291, October 2003. (Not cited.)