Statistical and numerical optimization for speckle blind structured illumination microscopy

Penghuan Liu

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

CENTRALE NANTES

THESE DE DOCTORAT DE

L'ÉCOLE CENTRALE DE NANTES
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par

Penghuan LIU

Statistical and numerical optimization for speckle blind structured illumination microscopy

Thèse présentée et soutenue à Nantes, le 25/05/2018
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N)

Rapporteurs avant soutenance :

Laurent Mugnier	Ingénieur de recherche HDR, ONERA (Châtillon)
Jean-François Giovannelli	Professeur des Universités, Université de Bordeaux

Composition du Jury :

Président :	David Rousseau	Professeur des Universités, Université d'Angers
Examinateurs :	Perrine Paul-Gilloteaux	Ingénieure de recherche, CNRS
Dir. de thèse :	Jérôme Idier	Directeur de recherche CNRS - Laboratoire des Sciences du Numérique de Nantes
Co-encadrant :	Sébastien Bourguignon	Maître de Conférences, Ecole Centrale de Nantes

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

CENTRALE NANTES

Titre : Optimisation numérique et statistique pour la microscopie à éclairement structuré non contrôlé

Mots clés : imagerie super-résolue, optimisation numérique, éclairement de speckle, estimation à minimum de contraste, parcimonie conjointe

Résumé : La microscopie à éclairements structurés (structured illumination microscopy, SIM) permet de dépasser la limite de résolution en microscopie optique due à la diffraction, en éclairant l'objet avec un ensemble de motifs périodiques parfaitement connus. Cependant, il s'avère difficile de contrôler exactement la forme des motifs éclairants. Qui plus est, de fortes distorsions de la grille de lumière peuvent être générées par l'échantillon lui-même dans le volume d'étude, ce qui peut provoquer de forts artefacts dans les images reconstruites.
Récemment, des approches dites blind-SIM ont été proposées, où les images sont acquises à partir de motifs d'éclairement inconnus, non-périodiques, de type speckle, bien plus faciles à générer en pratique. Le pouvoir de super résolution de ces méthodes a été observé, sans forcément être bien compris théoriquement. Cette thèse présente deux nouvelles méthodes de reconstruction en microscopie à éclairements structurés inconnus (blind-speckle-SIM) : une approche conjointe et une approche marginale.
Dans l'approche conjointe, nous estimons conjointement l'objet et les motifs d'éclairement au moyen d'un modèle de type Basis Pursuit DeNoising (BPDN) avec une régularisation en norme lp,q où p=>1 et 0<q<=1. La norme lp,q est introduite afin de prendre en compte une hypothèse de parcimonie sur l'objet.
Dans l'approche marginale, nous reconstruisons uniquement l'objet et les motifs d'éclairement sont traités comme des paramètres de nuisance. Notre contribution est double. Premièrement, une analyse théorique démontre que l'exploitation des statistiques d'ordre deux des données permet d'accéder à un facteur de super-résolution de deux, lorsque le support de la densité spectrale du speckle correspond au support fréquentiel de la fonction de transfert du microscope. Ensuite, nous abordons le problème du calcul numérique de la solution. Afin de réduire à la fois le coût de calcul et les ressources en mémoire, nous proposons un estimateur marginal à base de patches. L'élément clé de cette méthode à patches est de négliger l'information de corrélation entre les pixels appartenant à différents patches.
Des résultats de simulations et en application à des données réelles démontrent la capacité de super-résolution de nos méthodes. De plus, celles-ci peuvent être appliquées aussi bien sur des problèmes de reconstruction 2D d'échantillons fins, mais également sur des problèmes d'imagerie 3D d'objets plus épais

Title : Statistical and numerical optimization for speckle blind structured illumination microscopy

Keywords : super-resolution imaging, numerical optimization, speckle illumination, minimum contrast estimation, joint sparsity.

Abstract : Conventional structured illumination microscopy (SIM) can surpass the resolution limit in optical microscopy caused by the diffraction effect, through illuminating the object with a set of perfectly known harmonic patterns. However, controlling the illumination patterns is a difficult task. Even worse, strong distortions of the light grid can be induced by the sample within the investigated volume, which may give rise to strong artifacts in SIM reconstructed images.
Recently, blind-SIM strategies were proposed, where images are acquired through unknown, non-harmonic, speckle illumination patterns, which are much easier to generate in practice. The super-resolution capacity of such approaches was observed, although it was not well understood theoretically. This thesis presents two new reconstruction methods in SIM using unknown speckle patterns (blind-speckle-SIM): one joint reconstruction approach and one marginal reconstruction approach.
In the joint reconstruction approach, we estimate the object and the speckle patterns together by considering a basis pursuit denoising (BPDN) model with lp,q-norm regularization, with p=>1 and 0<q<=1. The lp,q-norm is introduced based on the sparsity assumption of the object.
In the marginal approach, we only reconstruct the object, while the unknown speckle patterns are considered as nuisance parameters. Our contribution is twofold. First, a theoretical analysis demonstrates that using the second order statistics of the data, blind-speckle-SIM yields a super-resolution factor of two, provided that the support of the speckle spectral density equals the frequency support of the microscope point spread function. Then, numerical implementation is addressed. In order to reduce the computational burden and the memory requirement of the marginal approach, a patch-based marginal estimator is proposed. The key idea behind the patch-based estimator consists of neglecting the correlation information between pixels from different patches.
Simulation results and experiments with real data demonstrate the super-resolution capacity of our methods. Moreover, our proposed methods can not only be applied in 2D super-resolution problems with thin samples, but are also compatible with 3D imaging problems of thick samples.

Contents

List of Figures

List of Symbols

$\boldsymbol{r}, \boldsymbol{r}'$	spatial index vector
$\boldsymbol{\kappa}, \boldsymbol{\kappa}'$	index vector in Fourier domain
$\boldsymbol{u}, \boldsymbol{u}'$	index vector in Fourier domain
h	point spread function
h_a	amplitude point spread function
$\tilde{h}$	optical transfer function
$\tilde{h}_a$	modulation transfer function
$\boldsymbol{h}$	discrete form of h
$\mathbf{H}$	convolution matrix w.r.t $\boldsymbol{h}$
$\mathbf{I}_m$	m-th realization of speckle patterns
γ_{spec}	covariance function of speckle
$\tilde{\gamma}_{\text{spec}}$	spectral density function of speckle
ρ	the unknown fluorescence object
$\boldsymbol{\rho}$	discrete form of ρ
$\boldsymbol{y}_m$	the measurement w.r.t. speckle $\mathbf{I}_m$
$\boldsymbol{\Gamma}_y$	covariance matrix of corresponding random vector $\boldsymbol{y}$
$\mathbb{1}_N$	the $N \times N$ identity matrix
$\mathbf{1}_N$	the $N \times 1$ column vector with every element identified by 1
$*$	convolution
$\star$	autocorrelation
$\langle \cdot, \cdot \rangle$	inner product

Introduction

Optical microscopy is an important tool in biological science, enabling to access detailed information about living specimens. However, due to the diffraction effect of light, the conventional optical microscopy has a resolution limit (Abbe diffraction limit) about $\lambda/2NA \approx 200\,\text{nm}$, with λ denoting the wavelength of light and NA the numerical aperture of the optical system [Mer10]. It also lacks ability to perform three-dimensional (3D) imaging since the in-focus information is always blurred by the out-of-focus information in thick samples.

Far-field super-resolution fluorescence microscopy has seen its prosperity over the past twenty years, making the observation of structures below the Abbe resolution limit possible. Several techniques have been developed, such as structured illumination microscopy (SIM) [Gus00], stimulated emission depletion fluorescence scanning microscopy (STED) [HW94], stochastic optical reconstruction microscopy (STORM) [RBZ06], photo activated localization microscopy (PALM) [HGM06], and so on. Among them, SIM is a wide-field technique providing fast data acquisition and it does not require using specific fluorophores. Moreover, SIM also provides optical sectioning ability in 3D imaging [GSC+08][Wal01].

In standard SIM, the object is illuminated by a set of harmonic patterns with known spatial frequencies and phases. However, generating a perfect known harmonic illumination is a difficult task and strong distortions of the light grid can be induced within the sample volume [JTF+15]. Such uncertainty in the illumination process reduces the super-resolution capacity in SIM and may generate strong artifacts [AGJ+13]. Hence, a so-called blind-speckle-SIM method was proposed in [MBG+12], where unknown speckle patterns are used as a substitute for the harmonic illuminations. A super-resolution capacity has been observed in [MBG+12], although this super-resolution is relatively weak compared with standard SIM. In this thesis, we explore the super-resolution capacity and propose new reconstruction methods in a blind-speckle-SIM approach (that is, with unknown illuminations) in order to obtain better super-resolution capacity, by exploiting second-order statistics of the data and introducing new regularizers.

This thesis is composed of six chapters. In the first chapter, we introduce the concepts of point spread function (PSF) and optical transfer function (OTF) in microscopy. Then, we briefly present the statistical properties of speckle patterns and the existing reconstruction methods in blind-speckle-SIM.

In Chapter 2, we propose a joint reconstruction method (which jointly estimates the speckle patterns and the object) by minimizing a data fidelity term penalized with a mixed $\ell_{p,q}$-norm regularizer in order to enforce the joint sparsity of the data. The alternating direction method of multipliers (ADMM) and the primal-dual algorithm are used to solve the optimization problem.

In Chapter 3, we explore the theoretical super-resolution capacity of blind-speckle-SIM by exploiting the second-order statistics of the data with minimum prior information on the data. A marginal estimator is proposed where the estimation is based on the statistics of the nuisance parameters (the unknown speckle patterns) but not their true values. This marginal estimator is shown to be an asymptotically consistent estimator.

The computational complexity of the marginal estimator proposed in Chapter 3 is $\mathcal{O}(N^3)$, which is too high for realistic image sizes. Thus, in Chapter 4, we propose a patch-based

marginal approach in order to reduce its computational burden, with computational complexity $\mathcal{O}(N^2 \log N)$.

In chapter 5, we demonstrate the optical sectioning ability of blind-speckle-SIM with both the $\ell_{p,q}$ regularized joint estimator and the patch-based marginal estimator.

Chapter 6 is a conclusion of this thesis and introduces some possible future work.

Introduction

La microscopie optique est un outil important en sciences biologiques, permettant d'accéder à des informations détaillées au sujet de spécimens vivants. Cependant, en raison de l'effet de diffraction de la lumière, la microscopie optique conventionnelle a une limite de résolution (limite de diffraction d'Abbe) de $\lambda/2NA \approx 200$ nm, où λ est la longueur d'onde de la lumière et NA est l'ouverture numérique du système optique [Mer10]. Elle souffre également d'incapacité de réaliser une imagerie tridimensionnelle (3D) puisque l'information au plan focal est toujours perturbée par le contenu défocalisé dans des échantillons épais.

La microscopie super-résolue par fluorescence en champ lointain a connu une période de prospérité au cours des vingt dernières années, rendant possible l'observation de structures à une résolution inférieure à la limite de résolution de Abbe. Différentes techniques ont été développées, telles que la microscopie à illumination structurée (Structured Illumination Microscopy, SIM) [Gus00], la microscopie à déplétion par émission stimulée (stimulated-emission-depletion, STED) [HW94], la microscopie de reconstruction optique stochastique (stochastic optical reconstruction microscopy, STORM) [RBZ06], ou la microscopie de localisation photo-activée (photo-activated localization microscopy, PALM) [HGM06]. Parmi ces méthodes, la microscopie à illumination structurée est une technique à large champ permettant une acquisition rapide des données et ne nécessite pas de fluorophores spécifiques. De plus, cette méthode possède également une capacité de sectionnement optique en imagerie 3D [GSC+08] [Wal01].

En SIM standard, l'objet est éclairé par un ensemble de motifs harmoniques de fréquences spatiales et de phases connues. Cependant, générer une illumination harmonique parfaitement connue est une tâche difficile en pratique et de fortes distorsions de la grille de lumière peuvent être induites dans le volume étudié [JTF+15]. L'incertitude sur les motifs d'illumination réduit alors la capacité de super-résolution de cette modalité d'imagerie et peut produire de forts artéfacts [AGJ+13]. Par conséquent, une méthode dite *blind-speckle-SIM* a été proposée dans [MBG+12], où des motifs d'illumination inconnus, de type speckle, sont utilisés à la place d'illuminations harmoniques. Un pouvoir de super-résolution a ainsi été observé dans [MBG+12], cependant cette super-résolution est relativement faible comparée à la méthode SIM standard. Dans cette thèse, nous explorons la capacité de super-résolution et proposons de nouvelles méthodes de reconstruction dans une démarche aveugle (*i*.e., à illuminations inconnues) pour obtenir une meilleure capacité de super-résolution, en exploitant les statistiques au deuxième ordre des données et en proposant de nouvelles régularisations.

Cette thèse est constituée de six chapitres. Dans le premier chapitre, nous introduisons les concepts de fonction d'étalement ponctuel (point spread function, PSF) et de fonction de transfert optique (optical transfer function, OTF) de la microscopie et nous présentons brièvement les propriétés statistiques des motifs de speckle et les méthodes de reconstruction existantes de type blind-speckle-SIM.

Dans le Chapitre 2, nous proposons une méthode de reconstruction conjointe (estimation des motifs de speckle et de l'objet) en minimisant un terme de fidélité des données pénalisé par une fonction de régularisation en norme mixte $\ell_{p,q}$ pour imposer la parcimonie conjointe des

données. La méthode ADMM (Alternating Direction Method of Multipliers) et un algorithme primal-dual sont utilisés pour résoudre le problème d'optimisation.

Dans le Chapitre 3, nous explorons la capacité de super-résolution théorique d'approches de type blind-speckle-SIM en exploitant les statistiques de second ordre des données avec un minimum d'informations préalables sur les données. Un estimateur marginal est proposé, où l'estimation est basée sur les statistiques des paramètres de nuisance (profils de speckle inconnus) et non pas sur leurs vraies valeurs. Nous montrons que cet estimateur marginal est un estimateur asymptotiquement cohérent.

La complexité de calcul de l'estimateur marginal proposé au Chapitre 3 est en $\mathcal{O}(N^3)$, ce qui est trop élevé pour des tailles d'image réalistes. Aussi, dans le Chapitre 4, nous proposons une approche marginale basée sur une décomposition des images en patches afin de réduire le coût de calcul, permettant d'obtenir une complexité en $\mathcal{O}(N^2 \log N)$.

Dans le Chapitre 5, nous démontrons la capacité de sectionnement optique de blind-speckle-SIM avec l'estimateur conjoint régularisé $\ell_{p,q}$ et l'estimateur marginal basé sur la décomposition en patches précédemment introduite.

Le Chapitre 6 conclut ce travail et propose quelques pistes pour de possibles travaux futurs.

Chapter 1

Basic optical concepts

1.1 Introduction

Optical microscopy is a diffraction-limited system, where the image of a point source is not resolved as a point. Instead, a spot of finite size is created, together with a series of concentric bright rings, called the point spread function (PSF), or Airy pattern [Goo05]. When two points are too close to each other, their images overlap and may become indistinguishable, giving rise to the resolution limit of conventional microscopy.

Another way to characterize the resolution of an optical system considers its optical transfer function (OTF), which is defined as the Fourier transform of the PSF. The finer structure of the object corresponds to high-frequency components in the Fourier domain. The OTF measures how well each frequency component can be transmitted through the optical system. Only the frequency components inside the OTF support can pass the optical system, and the high-frequency part is lost during the imaging process.

My thesis focus on super-resolution microscopy induced by unknown speckle patterns (blind-speckle-SIM). In this chapter I will briefly introduce where the resolution comes from in conventional optical microscopy and present the state of the art of the super-resolution methods. I first give an analytical expression of the light field in Section 1.2. The image formation model of coherent and incoherent imaging systems are discussed in Sections 1.3. The super-resolution techniques are presented in Section 1.4, while the existing reconstruction methods proposed in blind-speckle-SIM are presented in Section 1.5.

1.2 Light field

Light can be characterized as an electromagnetic field $\boldsymbol{u}(\boldsymbol{r}, t)$ varying in space (coordinate $\boldsymbol{r}$) and time (coordinate t). The Fourier transform of $\boldsymbol{u}(\boldsymbol{r}, t)$ in both space and time variables reads:

$$\tilde{\boldsymbol{u}}(\boldsymbol{\kappa}, v) = \int \boldsymbol{u}(\boldsymbol{r}, t) e^{-j2\pi(\boldsymbol{\kappa}\cdot\boldsymbol{r} - vt)} d\boldsymbol{r} dt, \tag{1.1}$$

where $\boldsymbol{\kappa} = (\kappa_x, \kappa_y, \kappa_z)$, the Fourier conjugate variable of the position vector $\boldsymbol{r}$, is called wavevector (or k vector) and v denotes the temporal frequency. The wavenumber κ is defined as:

$$\kappa = \frac{n}{c} v, \tag{1.2}$$

in which n is the index of refraction of the medium and $c = 3 \times 10^8$ m/s denotes the speed of light in empty space. The wavelength is then given by:

$$\lambda = \frac{c}{nv} = \frac{1}{\kappa}. \tag{1.3}$$

According to the energy-momentum relation [Mer10, chapter 1]:

$$|\boldsymbol{\kappa}|^2 = \kappa^2. \tag{1.4}$$

We see from (1.2) and (1.4) that the magnitude of the wavevector $\boldsymbol{\kappa}$ is fixed by the wavelength λ.

1.2.1 Monochromatic field

Firstly, we focus our attention on a monochromatic field. For the light propagating in a dielectric medium that is linear, isotropic, homogeneous and nondispersive, the incident field can be described by a complex-valued analytic signal:

$$\boldsymbol{u}(\boldsymbol{r}, t) = U(\boldsymbol{r}) \exp(-2\pi j v t) \tag{1.5}$$

with v the frequency and $U(\boldsymbol{r})$ the complex phasor amplitude:

$$U(\boldsymbol{r}) = |U(\boldsymbol{r})| \exp\big(j\phi(\boldsymbol{r})\big) \tag{1.6}$$

1.2.2 Spatial coherence of polychromatic field

The monochromatic assumption of light is overly restrictive in realistic conditions, even illumination generated by lasers is not perfectly monochromatic. The polychromatic light field $\boldsymbol{u}(\boldsymbol{r}, t)$ is defined as *narrowband* when the bandwidth Δv is much smaller than its center frequency $\bar{v}$. For the narrowband case, the wavefield can be described as [Goo05, chapter 6]:

$$\boldsymbol{u}(\boldsymbol{r}, t) = U(\boldsymbol{r}, t) \exp(-2\pi j \bar{v} t) \tag{1.7}$$

with the time varying complex phasor amplitude:

$$U(\boldsymbol{r}, t) = |U(\boldsymbol{r}, t)| \exp\big(j\phi(\boldsymbol{r}, t)\big). \tag{1.8}$$

Since the phasor amplitude varies randomly with time, it is necessary to introduce some statistical concepts to describe the wave field. We consider two types of light here. The phasor amplitudes of the first type of light at all spatial points vary in union, called *spatially coherent*. While the second type of light is called *spatially incoherent* where the phasor amplitudes at all points vary in an uncorrelated fashion. Mathematically, this spatial coherence is described by the mutual intensity concept:

$$J_u(\boldsymbol{r}_1, \boldsymbol{r}_2) = E\big[U(\boldsymbol{r}_1, t) U^*(\boldsymbol{r}_2, t)\big], \tag{1.9}$$

where $E[\cdot]$ means expectation in time. The illumination is said to be perfectly *coherent* when the time-varying phasor amplitude $U(\boldsymbol{r}_1, t)$ and $U(\boldsymbol{r}_2, t)$ differ only by a complex constant. Equivalently, it could be written as:

$$U(\boldsymbol{r}, t) = U(\boldsymbol{r}) \frac{U(\boldsymbol{0}, t)}{\langle |U(\boldsymbol{0}, t)|^2 \rangle^{1/2}}. \tag{1.10}$$

Here the time varying phasor amplitude at origin has been arbitrarily chosen as the phase reference. Substituting (1.10) into (1.9), we obtain the mutual intensity of coherent light

$$J_u(\boldsymbol{r}_1, \boldsymbol{r}_2) = U(\boldsymbol{r}_1)U^*(\boldsymbol{r}_2). \tag{1.11}$$

For perfectly *incoherent* light, the phasor amplitudes vary in a statistically independent way, and the mutual intensity in this case is:

$$J_u(\boldsymbol{r}_1, \boldsymbol{r}_2) = \varsigma I(\boldsymbol{r}_1)\delta(\boldsymbol{r}_1 - \boldsymbol{r}_2) \tag{1.12}$$

in which ς is a real constant and $\delta(\boldsymbol{r})$ is a 2 dimensional Dirac delta function.

1.3 Image formation model in diffraction limited system

Under the narrowband assumption, the amplitude impulse response of the optical system does not change for different optical spectra v. Therefore, we can write the time-varying phasor of the image U_i as the convolution of a wavelength-independent impulse response with the time varying phasor of the object U_o [Goo05, chapter 6]:

$$U_i(\boldsymbol{r}, t) = \int h_a(\boldsymbol{r} - \boldsymbol{r}')U_o(\boldsymbol{r}', t - \delta_t)d\boldsymbol{r}' \tag{1.13}$$

where δ_t denotes the time of light propagation from the object plane to the image plane and h_a is the amplitude point spread function (APSF).

1.3.1 Coherent imaging system

In practice, we cannot measure the light field directly, instead we measure the intensity of the field, which is given by

$$I(\boldsymbol{r}) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} |\boldsymbol{u}(\boldsymbol{r}, t)|^2 dt \tag{1.14}$$

In the monochromatic case, we have $I(\boldsymbol{r}) = |U(\boldsymbol{r})|^2$ while in the narrowband non-monochromatic case, we have:

$$I(\boldsymbol{r}) = E\big[|U(\boldsymbol{r}, t)|^2\big] \tag{1.15}$$

Combining (1.13) and (1.15), we obtain:

$$I_i(\boldsymbol{r}) = \iint h_a(\boldsymbol{r} - \boldsymbol{r}_1)h_a^*(\boldsymbol{r} - \boldsymbol{r}_2)E\big[U_o(\boldsymbol{r}_1, t - \delta_{t1}), U_o(\boldsymbol{r}_2, t - \delta_{t2})\big]d\boldsymbol{r}_1 d\boldsymbol{r}_2 \tag{1.16}$$

The APSF h_a is nonzero over only a small region, so the integrand is nonzero only for points $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ that are very close to each other. Hence the difference between the time delays δ_{t1} and δ_{t2} is negligible under the narrowband assumption. Therefore, (1.16) can be equivalently written as:

$$I_i(\boldsymbol{r}) = \iint h_a(\boldsymbol{r} - \boldsymbol{r}_1)h_a^*(\boldsymbol{r} - \boldsymbol{r}_2)J_u(\boldsymbol{r}_1, \boldsymbol{r}_2)d\boldsymbol{r}_1 d\boldsymbol{r}_2 \tag{1.17}$$

When the light field is perfectly coherent, by substituting (1.11) into (1.17), we get:

$$I_i(\boldsymbol{r}) = \Big| \int h_a(\boldsymbol{r} - \boldsymbol{r}')U_o(\boldsymbol{r}')d\boldsymbol{r}' \Big|^2 = \big|(h_a * U_o)(\boldsymbol{r})\big|^2 \tag{1.18}$$

Finally, defining a time invariant phasor amplitude $U_i(\boldsymbol{r})$ in image space relative to $U_i(\mathbf{0}, t)$ as in (1.10), we can describe the coherent imaging system by:

$$U_i(\boldsymbol{r}) = \int h_a(\boldsymbol{r} - \boldsymbol{r}')U_o(\boldsymbol{r}')d\boldsymbol{r}' = (h_a * U_o)(\boldsymbol{r}) \tag{1.19}$$

The coherent imaging system is thus linear in complex amplitude. By applying the convolution theorem to (1.19), we have:

$$\tilde{U}_i(\boldsymbol{\kappa}) = \tilde{h}_a(\boldsymbol{\kappa})\tilde{U}_o(\boldsymbol{\kappa}) \tag{1.20}$$

where $\tilde{h}_a$, the Fourier transform of h_a, is called the coherent transfer function, or amplitude transfer function (ATF).

1.3.2 Incoherent imaging system

When the light emitted from the object plane is totally incoherent, we substitute (1.12) into (1.16) and get

$$I_i(\boldsymbol{r}) = \int |h_a(\boldsymbol{r} - \boldsymbol{r}')|^2 I_o(\boldsymbol{r}')d\boldsymbol{r}'. \tag{1.21}$$

We define the point spread function (PSF) h of the incoherent system as: $h = |h_a|^2$. Then, for the incoherent imaging system, the image intensity is the convolution of PSF h with the intensity of object:

$$I_i(\boldsymbol{r}) = \int h(\boldsymbol{r} - \boldsymbol{r}')I_o(\boldsymbol{r}')d\boldsymbol{r}' = (h * I_0)(\boldsymbol{r}) \tag{1.22}$$

In the frequency domain, we have:

$$\tilde{I}_i(\boldsymbol{\kappa}) = \tilde{h}(\boldsymbol{\kappa})\tilde{I}_o(\boldsymbol{\kappa}') \tag{1.23}$$

where $\tilde{h}(\boldsymbol{\kappa})$ is the optical transfer function (OTF) of the optical system. An illustration of the OTF is shown in Figure 1.1.

1.4 Super-resolution fluorescence microscopy

Fluorescence microscopy is an optical microscopy method that uses fluorescence to generate an image. The fluorophores absorbe energy from the incident light at a specific wavelength and then emit light at longer wavelengths. There is a time delay of about several nanoseconds between absorption and emission, which can be neglected in the imaging process. The emission light of the fluorophores is spatially incoherent.

The emission intensity is proportional to the excitation intensity when the excitation intensity is low, which can be written as: $I_{\text{out}} = \sigma I_{\text{in}}$, where the I_{out} is the intensity emitted and the I_{in} is the intensity of the excitation light. Normally, the fluorophores are much smaller than the resolution limit of the optical microscopy, thus they can be considered as a continuous density. One defines the function ρ such that:

$$\rho(\boldsymbol{r})d\boldsymbol{r} = \sum_{l=1}^{L} \sigma_l \tag{1.24}$$

where L is the number of fluorophores in the volume $d\boldsymbol{r}$ and σ_l is the emission coefficient of the l-th fluorophore. For the excitation intensity I, we can thus write the emission light intensity as ρI. In the linear regime, the following effects are neglected:

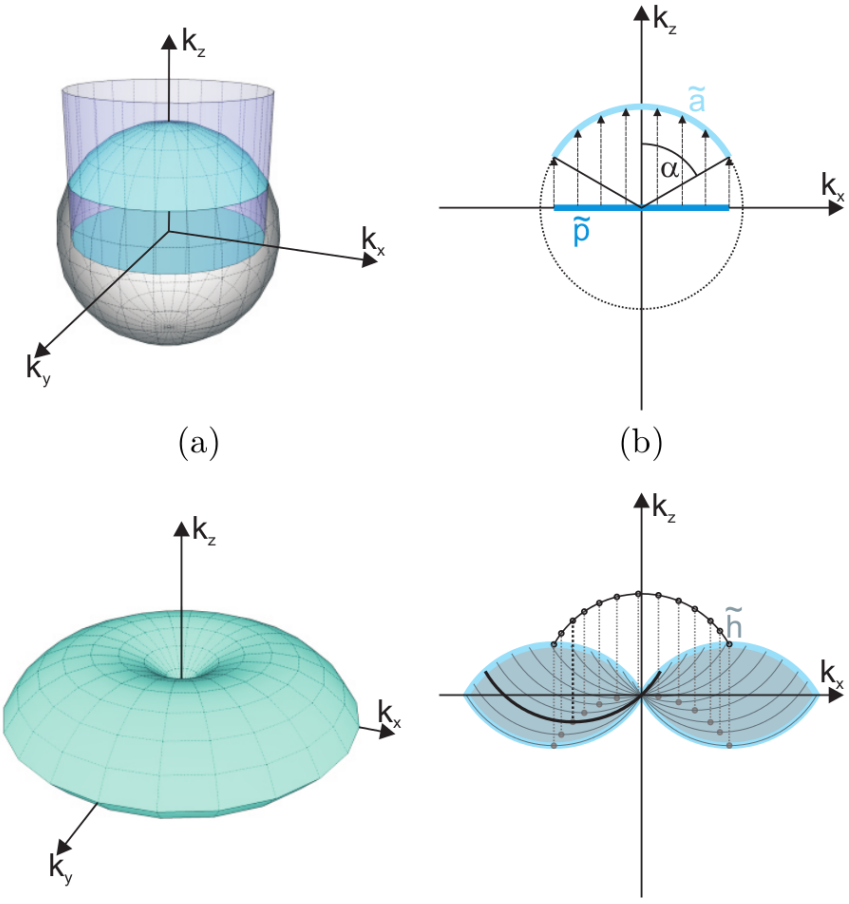


Figure 1.1 – **ATF and OTF of the system**. (Figure courtesy of [Wic10].) . All k-vectors of waves in a single wavelength λ lie on a spherical shell. Projecting the system's pupil function onto the sphere yielding the k-vectors available for imaging, which is called amplitude transfer function (ATF). The autocorrelation of ATF gives the optical transfer function (OTF) of the system.

— Photobleaching: The coefficient σ decreases with time and gradually, the sample cannot fluoresce any more. More precisely, the fluorescence decay is proportional to the emitted energy [SHYT95].
— Photoblinking: The intensity of fluorescence molecules repeatedly drops to zero and then comes back to normal. This fluctuation is the basis of the super-resolution technique SOFI [DCI+09].

1.4.1 Super resolution based on single-molecule fluorescence localization

In conventional microscopy, there is a hidden assumption that two close points emit light at the same moment. If two points do not emit light synchronously, we can locate them separately by finding the centroid position of the PSF and thus break through the resolution limit. In the past twenty years, several super-resolution techniques based on single-molecule fluorescence localization were proposed.

STORM/PALM – Stochastic optical reconstruction microscopy (STORM) [RBZ06] and photo activated localization microscopy (PALM) [HGM06] utilize sequential activation of photo switchable fluorophores to create high resolution images. In each imaging cycle, only a fraction of fluorophores are activated at any given moment, so that the position of each fluorophore can be determined with high precision. The fluorophore is subsequently deactivated, and another subset is activated and imaged. We can reconstruct a super-resolution image by repeating this process. The drawback of this technique is that it requires special fluorescent probes and the on-off cycles take a lot of time.

SOFI – Super-resolution optical fluctuation imaging (SOFI) produces highly resolved images based on fluorescence blinking by exploiting high order statistical analysis of the temporal fluctuations recorded in a sequence of images [DCI+09]. It requires that the fluorescence molecules fluctuate repeatedly and independently from each other. A second-order SOFI computes the empirical variance and this variance corresponds to the square of the original PSF. Higher resolution is possible by using high order statistics of the data. The disadvantage of SOFI is that the blinking rate of fluorescent molecules could be faster than the frame rate in video microscopy.

SPoD/ExPAN – The SPoD technique measures the orientation of the fluorophores by rotating the polarization of the excitation beam and detecting the periodic signals emitted with different phases [HGvdH+14]. This approach is not suitable for highly dynamic structures.

1.4.2 Super-resolution based on PSF engineering

STED – Stimulated emission depletion fluorescence scanning microscopy (STED) creates super-resolution images using two laser pulses. The object is illuminated by a point source through the objective lens, stimulating the fluorophores inside the region of PSF to their fluorescent state in the focal plane. With the help of an additional STED pulse, fluorophores in the outer regions of the excitation focus are quenched [HW94]. In the ideal case, the STED pulse creates a concentric annulus around the focal point, overlapping with the outer region of the Airy disk. By scanning this focal spot in the object, one retrieves the image. Since the super-resolution image is retrieved by scanning, this technique is time-consuming for large size objects, even with parallelizing STED [BAS+15].

1.4.3 Super-resolution based on structured illumination

SIM – Structured illumination microscopy (SIM) retrieves the super-resolution image by illuminating the sample with periodic light patterns, thus transfers otherwise unobservable high-frequency information about the sample into a lower-frequency region [Gus00]. Suppose the maximum spatial frequency can pass the microscopy is specified by k_0 and the frequency information contained in illumination patterns is k_1, as shown in Figure 1.2, then the frequencies will be observable as long as $|k - k_1| < k_0$, that is, the highest observable frequency is increased from k_0 to $k_0 + k_1$. This high frequency information can be extracted by some post-processing methods. As a wide-field super-resolution technique, the structured illumination microscopy (SIM) allows us to acquire images much faster than other super-resolution (SR) techniques, such as STED or STORM. Super-resolution imaging in living samples has been demonstrated in SIM [KCG+09].

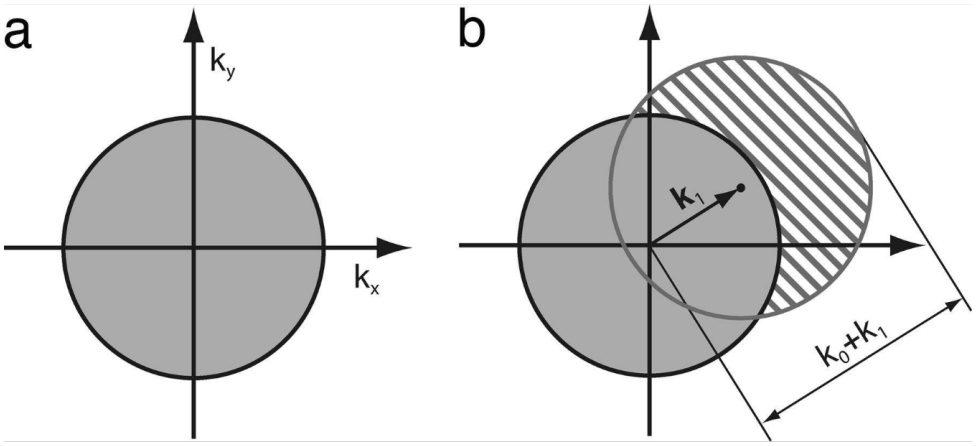


Figure 1.2 – **Concept of resolution enhancement by structured illumination.** (Figure courtesy of [Gus05].) (a.) The spatial frequencies that can pass the optical system defines a circular of radius k_0 in frequency space. (b.) If the illumination patterns contains a spatial frequency k_1, then the maximum spatial frequency that can be detected (in this direction) becomes $k_0 + k_1$.

Normally $k_1 \leq k_0$ since the illumination patterns are also limited by the diffraction effect in the same way as the object. So conventional SIM can improve the resolution by a factor of two at most. Saturated structured-illumination microscopy (SSIM) (also called non-linear SIM) [Gus05] generates illumination patterns containing high spatial frequencies taking advantage of the nonlinear properties of the fluorescent response in saturation status, thus making $k_1 > k_0$. The resolution limit of this technique is only bounded by the signal-to-noise ratio and photostability of dyes.

Blind-speckle-SIM – In standard SIM, the object is illuminated by a set of harmonic patterns with known spatial frequencies and phases. However, generating a perfectly known harmonic illumination is a difficult task and the blurring in the illumination will reduce the SR capacity in SIM [AGJ+13]. Moreover, the application of SIM is restricted to thin samples or samples with small refraction indices [MBG+12].

Blind-speckle-SIM achieves super-resolution using unknown speckle patterns as a substitute for harmonic illuminations. Compared with harmonic illuminations, the speckle patterns are easier to generate while the super-resolution is still attainable[MBG+12]. We briefly review the reconstruction methods proposed in speckle-SIM in Section 1.5.

1.5 Existing reconstruction procedures in blind-speckle-SIM

In SIM using speckle illumination, the sample is successively illuminated by M independent speckle patterns $I_m, m = 1, \cdots, M$. For each speckle pattern, the recorded data y_m can be modeled as the convolution of the emitted density from the object ρ with the point spread function (PSF) h:

$$y_m = h * (\rho I_m) + \epsilon_m, \tag{1.25}$$

where $*$ represents the convolution operator and ϵ stands for noise and model errors. Since we record digital signals physically, we rewrite Eq. (1.25) in matrix form as:

$$\boldsymbol{y}_m = \mathbf{H}\mathbf{R}\mathbf{I}_m + \boldsymbol{\epsilon}_m \tag{1.26}$$

with $\mathbf{H} \in \mathbb{R}^{N\times N}$ the convolution matrix of PSF h and $\mathbf{R} \in \mathbb{R}^{N\times N}$ the diagonal matrix whose diagonal values identify with the discretized object $\boldsymbol{\rho}$. $\mathbf{I}_m \in \mathbb{R}^N$ is the m-th realization of speckle with homogeneous intensity mean I_0. Now the problem is to estimate the object $\boldsymbol{\rho}$ from a series of low resolution images $\boldsymbol{y}_m$.

1.5.1 Blind-SIM with positivity constraint

The idea of obtaining super-resolution with speckle patterns was first proposed in [MBG+12], with a reconstruction algorithm called blind-SIM simultaneous inversion (blind-SIM-SI). It assumes that the sum of speckle patterns is homogeneous and that both the object and speckle patterns are positive. The reconstruction procedure is based on minimizing the mismatch between the acquired data and the model:

$$\begin{aligned} &\min_{\boldsymbol{\rho},\{\mathbf{I}_m\}} \quad \sum_{m=1}^{M} \|\boldsymbol{y}_m - \mathbf{H}\mathbf{R}\mathbf{I}_m\|^2 \\ &\text{s.t.} \quad \boldsymbol{\rho}, \mathbf{I}_m \geq 0, \quad \frac{1}{M}\sum_m \mathbf{I}_m = \mathbf{I}_0 \end{aligned} \tag{1.27}$$

To incorporate the positivity constraint, the auxiliary variables $\boldsymbol{\xi}$ and $\boldsymbol{\eta}_m$ are introduced with $\boldsymbol{\rho} = \boldsymbol{\xi}^2$ and $\mathbf{I}_m = \boldsymbol{\eta}_m^2$. Then the variables are updated with respect to $\boldsymbol{\xi}$ and $\boldsymbol{\eta}_m$ alternately using conjugate gradient algorithm [NW06, chapter 5]. Though we retrieve partial super-resolution information using this method, the modulation contrast in super-resolution part is relatively weaker than standard SIM.

1.5.2 Blind-SIM with sparsity and positivity constraint

Our team propose to solve the problem (1.27) in [LNI+17] by breaking it into M separate deconvolution sub-problems. The auxiliary variables $\boldsymbol{q}_m = \boldsymbol{\rho} \circ \mathbf{I}_m$ are introduced, with $\circ$ denoting the element-wise product:

$$\min_{\boldsymbol{q}_m} \quad \|\boldsymbol{y}_m - \mathbf{H}\boldsymbol{q}_m\|^2 \qquad \text{s.t.} \quad \boldsymbol{q}_m \geq 0 \tag{1.28}$$

After all $\{\boldsymbol{q}_m\}_{m=1}^M$ are obtained, the estimated object is given by $\hat{\boldsymbol{\rho}} = \sum_m \boldsymbol{q}_m/(MI_0)$. With the assumption that the product images $\boldsymbol{q}_m$ tends to be a nearly black object [DJHS92], a strictly convex penalty term is added to the formula (1.28):

$$\varphi(\boldsymbol{q}_m; \alpha, \beta) = \beta\|\boldsymbol{q}_m\|_2^2 + \alpha\|\boldsymbol{q}_m\|_1 \tag{1.29}$$

with $\|\cdot\|_p$ denoting the ℓ_p norm and $\alpha \geq 0,\ \beta > 0$. Now we are led to the optimization problem:

$$\min_{\boldsymbol{q}_m} \ f(\boldsymbol{q}_m) := \|\boldsymbol{y}_m - \mathbf{H}\boldsymbol{q}_m\|^2 + \varphi(\boldsymbol{q}_m) \qquad \text{s.t.} \qquad \boldsymbol{q}_m \geq 0 \tag{1.30}$$

The problem (1.30) can be expressed by the sum of two functions: $f(\boldsymbol{q}_m) = f_1(\boldsymbol{q}_m) + f_2(\boldsymbol{q}_m)$ with f_1 a smooth function and f_2 a non-differentiable one:

$$\begin{aligned} f_1(\boldsymbol{q}_m) &= \|\boldsymbol{y}_m - \mathbf{H}\boldsymbol{q}_m\|^2 + \beta\|\boldsymbol{q}_m\|_2^2 \\ f_2(\boldsymbol{q}_m) &= \alpha \sum_n \phi(q_{m,n}) \end{aligned} \tag{1.31}$$

where ϕ is defined as:

$$\phi(u) = \begin{cases} u, & u \geq 0 \\ +\infty, & u < 0 \end{cases}$$

Though function f_2 is nondifferentiable, its proximal operator can be written analytically:

$$\text{prox}_{\lambda f_2}(\boldsymbol{x}) = \textbf{vect}(\max\{x_n - \lambda\alpha, 0\}) \tag{1.32}$$

with x_n the n-th component of vector $\boldsymbol{x}$. Now the optimization problem (1.30) can be solved using a proximal gradient descent algorithm, such as FISTA [BT09]. However, the convergence speed of FISTA in this problem is relatively slow and a preconditioned Primal-Dual splitting algorithm has been show in [LNI+17] to behave better in terms of convergence rate. The primal-dual update reads:

$$\boldsymbol{q}_{k+1} = \boldsymbol{q}_k - \rho\tau\mathbf{B}\boldsymbol{\zeta}_k \tag{1.33a}$$

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \rho\big[\mathcal{P}_\star(\boldsymbol{\delta}_k) - \boldsymbol{w}_k\big] \tag{1.33b}$$

with $\boldsymbol{\zeta}_k = \nabla h_1(\boldsymbol{q}_k) + \boldsymbol{w}_k$ and the preconditioning matrix $\mathbf{B}$ is chosen from the Geman and Yang semi-quadratic construction:

$$\mathbf{B} = (2\mathbf{H}^t\mathbf{H} + 2\beta\mathbb{1}_N/a)^{-1} \tag{1.34}$$

where $a > 0$ is a free parameter of the preconditioner. In the dual update (1.33b), we have $\boldsymbol{\delta}_k = \boldsymbol{w}_k + \sigma(\boldsymbol{q}_k - 2\tau\mathbf{B}\boldsymbol{\zeta}_k)$ and

$$\mathcal{P}_\star(\boldsymbol{x}) = \textbf{vect}(\min\{x_n, \alpha\}). \tag{1.35}$$

When some conditions on parameters $\{\rho, \tau, \sigma, a\}$ are met [Con13], the convergence of PPDS algorithm is granted. One drawback of this method is that the hyperparameters α and β are not easy to tune and they will influence the quality of the reconstructed images.

1.5.3 Joint support recovery

In [MJK+13] the object is obtained by assuming that the set of auxiliary vectors $\{\boldsymbol{q}_m\}_{m=1}^M$ share a common non-zero support, where $\boldsymbol{q}_m$ is given by the same definition as in former Section 1.5.2. The observation model in matrix form reads as:

$$\mathbf{Y} = \mathbf{H}\mathbf{Q} + \mathbf{E} \tag{1.36}$$

with $\mathbf{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M] \in \mathbb{R}^{N\times M}$, $\mathbf{Q} = [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_M] \in \mathbb{R}^{N\times M}$ and $\mathbf{E} = [\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_M]$. Then, the reconstruction approach is based on considering the following optimization problem:

$$\mathbf{Q}_{JS} = \arg\min_{\mathbf{Q}} \|\mathbf{Y} - \mathbf{H}\mathbf{Q}\|_F^2 + \lambda\|\mathbf{Q}\|_{2,0} \tag{1.37}$$

in which $\|\cdot\|_F$ denotes the Frobenius norm and $\|\cdot\|_{2,0}$ is the $\ell_{2,0}$ norm defined by:

$$\|\mathbf{Q}\|_{2,0} = \sum_{n=1}^{N} \chi(\|\boldsymbol{q}^n\|_2) \tag{1.38}$$

where $\boldsymbol{q}^n$ denotes the n-th row of $\mathbf{Q}$ and $\chi(u)$ is the indicator function given by:

$$\chi(u) = \begin{cases} 1, & u \neq 0 \\ 0, & u = 0 \end{cases} \tag{1.39}$$

The optimization problem of joint sparsity has received great attention recently and several methods have been proposed, see [MCW05, TGS06, Tro06]. A multiple sparse Bayesian learning (M-SBL) algorithm was used in [MJK+13] to approximately solve problem (1.37), where $\mathbf{Q}_{JS}$ was viewed as a type II maximum likelihood or evidence maximization problem The conditional probability density $p(\boldsymbol{y}_m|\boldsymbol{q}_m)$ is assumed to be Gaussian with noise variance σ^2:

$$p(\boldsymbol{y}_m|\boldsymbol{q}_m) = (2\pi\sigma^2)^{-N/2} \exp\Big(-\frac{1}{2\sigma^2}\|\boldsymbol{y}_m - \mathbf{H}\boldsymbol{q}_m\|_2^2\Big) \tag{1.40}$$

Next, each row of $\mathbf{Q}$ is assigned an M-dimensional Gaussian prior with hyperparameter γ_n:

$$p(\boldsymbol{q}^n; \gamma_n) := \mathcal{N}(\mathbf{0}, \gamma_n \mathbb{1}_M) \tag{1.41}$$

where γ_n is the variance parameter and $\mathbb{1}_M$ is the $M \times M$ identity matrix. By combining all these row priors, we have:

$$p(\mathbf{Q}; \boldsymbol{\gamma}) = \Pi_{n=1}^{N}\, p(\boldsymbol{q}^n; \gamma_n) \tag{1.42}$$

where $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_N]^T \in \mathbb{R}^+$. Then joint sparsity is achieved whenever a γ_n equals zero since $p(\boldsymbol{q}^n = \mathbf{0}|\mathbf{Y}; \gamma_n = 0) = 1$. Thus minimizing the $\|\mathbf{Q}\|_{20}$ norm is shifted to estimating the hyperparameter $\boldsymbol{\gamma}$ with the correct location of nonzero elements. The M-SBL method solves this problem by treating $\mathbf{Q}$ as nuisance parameters, integrates them out and then maximizes the marginal likelihood function with respect to $\boldsymbol{\gamma}$:

$$\min_{\boldsymbol{\gamma}} \quad -2\log \int p(\mathbf{Y}|\mathbf{Q})p(\mathbf{Q}|\boldsymbol{\gamma})d\mathbf{Q} \propto \frac{1}{M}\mathrm{Tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Y}\mathbf{Y}^T) + \log|\boldsymbol{\Sigma}| \tag{1.43}$$

with $\boldsymbol{\Sigma} = \mathbf{H}\boldsymbol{\Gamma}\mathbf{H}^T + \sigma^2\mathbb{1}_N$, $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$ and σ^2 the noise variance. It is demonstrated in [WRN11] that the minimization problem in (1.43) is equivalent to the following regularized least-squares problem:

$$\min_{\boldsymbol{\gamma},\mathbf{Q}} \quad \|\mathbf{Y} - \mathbf{H}\mathbf{Q}\|_F^2 + \sigma^2\mathrm{Tr}(\mathbf{Q}^T\boldsymbol{\Gamma}^{-1}\mathbf{Q}) + \sigma M\log|\boldsymbol{\Sigma}|. \tag{1.44}$$

Then (1.44) is solved by updating $\boldsymbol{\gamma}$ and $\mathbf{Q}$ alternately. Here, we use the subscript k to indicate the variable values at k-th iteration:

1. Minimization with respect to $\mathbf{Q}$:

$$\mathbf{Q}_k = \boldsymbol{\Gamma}_k\mathbf{H}^T\boldsymbol{\Sigma}_k^{-1}\mathbf{Y} \tag{1.45}$$

2. Minimization with respect to $\boldsymbol{\gamma}$:

$$\gamma_{n;k+1} = \frac{\frac{1}{M}\|\boldsymbol{q}_k^n\|_2^2}{1 - \Sigma_{nn;k}/\gamma_{n;k}}, \qquad \text{for} \quad n \in 1, \cdots, N \tag{1.46}$$

with $\boldsymbol{q}_k^n$ the n-th row in $\mathbf{Q}_k$ and $\Sigma_{nn;k}$ the (n,n) element in $\boldsymbol{\Sigma}_k$.

3. Estimate the noise variance σ^2:

$$\sigma_{k+1}^2 = \frac{\frac{1}{M}\|\mathbf{Y} - \mathbf{HQ}\|_F^2}{\sum_{n=1}^{N} \frac{\Sigma_{nn;k+1}}{\gamma_{n;k+1}}} \tag{1.47}$$

Once the matrix $\mathbf{Q}$ is obtained, the author choose to estimate the object $\boldsymbol{\rho}$ based on second order statistics:

$$\boldsymbol{\rho}(\boldsymbol{r}) = \sqrt{Cov(\boldsymbol{q})(\boldsymbol{r})} \tag{1.48}$$

The advantage of using second order statistics is that it produce an estimator less sensitive to the background noise. One drawback of this method is that the computational burden and memory requirement of the M-SBL algorithm is too high for realistic size images due to it involves a matrix inverse operator of matrix $\boldsymbol{\Sigma}$ in each iteration, whose size is $N \times N$.

1.5.4 S-SOFI

The super-resolution optical fluctuation imaging (SOFI) achieves super-resolution taking advantage of the intrinsic blinking properties of protein fluorophores. An S-SOFI method was proposed in [KPR+15] where the blinking is induced by speckle patterns. The speckle patterns are supposed to form a second-order stationary random process with modified correlation function given by $\boldsymbol{\gamma}_{\text{spec}} \approx \delta(\boldsymbol{r})$, then the second-order correlation of the measured image can be expressed as:

$$\begin{aligned}\boldsymbol{\gamma}_y(\boldsymbol{r}) = E[\Delta \boldsymbol{y}_m(\boldsymbol{r})^2] &= \int_{\boldsymbol{r}_2}\int_{\boldsymbol{r}_1} \boldsymbol{h}(\boldsymbol{r}-\boldsymbol{r}_1)\boldsymbol{h}(\boldsymbol{r}-\boldsymbol{r}_2)\boldsymbol{\rho}(\boldsymbol{r}_1)\boldsymbol{\rho}(\boldsymbol{r}_2)\boldsymbol{\gamma}_{\text{spec}}(\boldsymbol{r}_1-\boldsymbol{r}_2)d\boldsymbol{r}_1 d\boldsymbol{r}_2 \\ &\approx \boldsymbol{h}^2 * \boldsymbol{\rho}^2\end{aligned} \tag{1.49}$$

where $\Delta \boldsymbol{y}_m = \boldsymbol{y}_m - \bar{\boldsymbol{y}}$. Then the object is estimated from $\boldsymbol{\gamma}_y(\boldsymbol{r})$ with the Fourier reweighting (FRW) method [DCV+10]. With a simulated object including two points and a Gaussian PSF, a resolution enhancement factor 1.6 is observed in [KPR+15]. The resolution improvement in S-SOFI is demonstrated for ρ^2, no general conclusion about the super-resolution capacity for ρ could we make.

1.5.5 PE-SIMS

A so-called PE-SIMS (pattern estimation structured illumination microscopy with a statistical prior) method was proposed in [YTW17]. This method includes two parts. Firstly the speckle patterns are estimated based on the deconvolution of the wide field image. Then the object is estimated from the covariance between the measured images $\boldsymbol{y}_m$ and the speckle patterns $\mathbf{I}_m$.

1.5.5.1 Pattern estimation

A regularized deconvolution of the wide-field image is given by:

$$\boldsymbol{\rho}_{wd} = \mathcal{F}^{-1}\Big\{\frac{\tilde{\bar{\boldsymbol{y}}}\cdot\tilde{\boldsymbol{h}}}{|\tilde{\boldsymbol{h}}|^2+\beta}\Big\} \tag{1.50}$$

where $\bar{\boldsymbol{y}}$ represents the average of the measured images and $\mathcal{F}^{-1}$ denote the inverse Fourier transform and β is a small regularization parameter to make the solution robust with noise.

Then, the speckle patterns are estimated from the low-resolution deconvolved wide-field image $\boldsymbol{\rho}_{wd}$. To avoid the reconstruction artifacts, the correct Fourier support constraint is added to the original optimization problem via an indicator function $\chi_{\mathcal{C}}$:

$$\arg\min_{\mathbf{I}_m} \quad \|\boldsymbol{y}_m - \mathbf{H}\mathbf{R}_{wd}\mathbf{I}_m\|_2^2 + \chi_{\mathcal{C}}(\mathbf{I}_m), \tag{1.51}$$

where $\mathbf{R}_{wd} = \mathrm{diag}(\boldsymbol{\rho}_{wd})$ and $\chi_C(\mathbf{I}_m)$ is given by:

$$\chi_{\mathcal{C}}(\mathbf{I}_m) = \begin{cases} 0, & \mathbf{I}_m \in \mathcal{C} \\ \infty, & \mathbf{I}_m \notin \mathcal{C} \end{cases}, \quad \mathcal{C} = \Big\{\mathbf{I}_m \mid \tilde{\mathbf{I}}_m(\boldsymbol{u}) = 0,\ \forall \boldsymbol{u} > \frac{2NA}{\lambda_{\text{illu}}}\Big\}, \tag{1.52}$$

where λ_{illu} is the wavelength of the excitation light. The indicator function $\chi_{\mathcal{C}}$ ensures that the frequency support of each speckle pattern is confined within the OTF of the system. The problem (1.51) is solved by a proximal gradient descent algorithm, which is designed to solve convex optimization problem with two parts: one being a differentiable term and the other being a nondifferentiable one. When the nondifferentiable part is an indicator function as in Eq. (1.51), the method is also named as projected gradient method. The gradient of the differentiable part in (1.51) with respect to $\mathbf{I}_m$ in k-th iteration is given by:

$$g_{\mathbf{I}_m,k} = -\boldsymbol{\rho}_{wd} \circ \Big[\boldsymbol{h} * \big(\boldsymbol{y}_m - \boldsymbol{h} * (\boldsymbol{\rho}_{wd} \circ \mathbf{I}_{m,k})\big)\Big] \tag{1.53}$$

Then the projection operator P_c is used to force the frequency outside the OTF support be zero at each iteration. To reduce the high-frequency artifacts, the following soft-edge filter is chosen:

$$P_c(\mathbf{I}_m) = \mathcal{F}^{-1}\Big\{\frac{\mathcal{F}\{\mathbf{I}_m\}\cdot|\tilde{\boldsymbol{\gamma}}_{\text{spec}}|}{|\tilde{\boldsymbol{\gamma}}_{\text{spec}}|+\delta}\Big\} \tag{1.54}$$

where $\boldsymbol{\gamma}_{\text{spec}}$ is the covariance of speckle patterns as shown in Eq. (A.29) and δ determines how much the high frequency information is suppressed.

1.5.5.2 Object estimation

In the second part of the PE-SIMS method, the super-resolution object is reconstructed from the covariance between measured images $\boldsymbol{y}_m$ and the estimated speckle patterns $\mathbf{I}_m$:

$$\begin{aligned} \mathbf{C}_{\text{cov}}(\boldsymbol{r}) &= E[\Delta\boldsymbol{y}_m(\boldsymbol{r})\Delta\mathbf{I}_m(\boldsymbol{r})] \\ &= \int \boldsymbol{\rho}(\boldsymbol{r}')E[\Delta\mathbf{I}_m(\boldsymbol{r})\Delta\mathbf{I}_m(\boldsymbol{r}')]\boldsymbol{h}(\boldsymbol{r}-\boldsymbol{r}')d\boldsymbol{r}' \\ &= \int \boldsymbol{\rho}(\boldsymbol{r}')\boldsymbol{\gamma}_{\text{spec}}(\boldsymbol{r}-\boldsymbol{r}')\boldsymbol{h}(\boldsymbol{r}-\boldsymbol{r}')d\boldsymbol{r}' \end{aligned} \tag{1.55}$$

with $\Delta\boldsymbol{y}_m = \boldsymbol{y}_m - \bar{\boldsymbol{y}}$ and $\Delta\mathbf{I}_m = \mathbf{I}_m - \mathbf{I}_0$. Therefore, we have:

$$\mathbf{C}_{\text{cov}} = (\boldsymbol{h} \circ \boldsymbol{\gamma}_{\text{spec}}) * \boldsymbol{\rho}. \tag{1.56}$$

When the fluorescent emission wavelength equals the excitation one, the frequency support of $\boldsymbol{\gamma}_{\text{spec}}$ identifies with $\boldsymbol{h}$. Then the object is recovered by a standard regularized deconvolution procedure:

$$\boldsymbol{\rho} = \mathcal{F}^{-1}\Big\{\frac{\tilde{\mathbf{C}}_{cov} \circ \mathbf{O}}{|\mathbf{O}|^2 + \varepsilon}\Big\} \tag{1.57}$$

with $\mathbf{O}$ given by:

$$\mathbf{O} = \mathcal{F}\big\{\boldsymbol{h} \circ \boldsymbol{\gamma}_{\text{spec}}\big\} \tag{1.58}$$

The super-resolution capacity analysis of this method is based on a perfect reconstruction of the unknown speckle patterns, which is impossible in practise. The influence of the imperfect speckle reconstruction on this estimator is theoretically unclear.

1.6 Conclusion

In this chapter, I have introduced the spatial coherence of the narrowband polychromatic field and PSF, OTF concepts of a diffraction-limited system, which determine the resolution limit of conventional optical microscopy. The state of the art super-resolution techniques in optical fluorescence microscopy are presented including STED, STORM/PALM, SOFI, and SIM. Specifically, I focused on the blind-speckle-SIM technique and the associated reconstruction methods that have been proposed are briefly described.

Chapter 2

A new joint reconstruction approach in blind-speckle-SIM

2.1 Introduction

We note here that the first three reconstruction methods in speckle-SIM presented in Section 1.5 are quite similar, namely they try to minimize the square of the ℓ_2 norm of the error between the measured data and the model, plus different constraints: positivity, sparsity, joint sparsity or their combinations. In this chapter, I propose a similar model using $\ell_{p,q}$ regularizer. What is more, we could combine $\ell_{p,q}$ regularizer with other prior information of the object, such as total variation (TV) norm or positivity.

In Section 2.2 we solve the $\ell_{p,q}$ minimization problem using alternating direction method of multipliers (ADMM). Recently the ADMM approach has been successfully applied to several convex or non-convex optimization problems, including $\ell_{2,1}$-regularized problem [DYZ12] and total variation (TV) regularized problem [WYYZ08, ABDF10]. Two different splitting patterns are presented, one for unconstrained form and the other for constrained form for the basis pursuit denoising (BPDN) model. When the $\ell_{p,q}$ norm is combined with TV norm, we choose a primal-dual approach to solve it as shown in Section 2.3.

2.2 $\ell_{p,q}$ norm minimization

The data formation model in speckle-SIM is shown in (1.36). To reconstruct the super-resolution image, we need to incorporate some prior information on the object. Specifically, we consider the following constrained form of basis pursuit denoising models:

$$\arg\min_{\mathbf{Q}} \|\mathbf{Q}\|_{pq}^q \quad \text{s.t.} \quad \left\|\mathbf{HQ}-\mathbf{Y}\right\|_F \leq \epsilon \tag{2.1}$$

where $\ell_{p,q}$ norm $\|\mathbf{Q}\|_{pq}$ is used to measure the joint sparsity of matrix $\mathbf{Q}$ with $p \geq 1,\ 0 \leq q \leq 1$:

$$\|\mathbf{Q}\|_{pq} = \Big(\sum_n \|\boldsymbol{q}^n\|_p^q\Big)^{1/q} \tag{2.2}$$

In our model, we use $\ell_{p,q}$ norm regularizer to enforce joint sparsity of matrix $\mathbf{Q}$. Here we describe how this sparsity implicitly correspond to a prior on object ρ. For the $n-$th line of matrix $\mathbf{Q}$, we have

$$\|\boldsymbol{q}^n\|_p = \Big[|\rho_n I_{1n}|^p + \cdots + |\rho_n I_{Mn}|^p\Big]^{1/p} \tag{2.3}$$

when $p = 1$,

$$\|\boldsymbol{q}^n\|_p = \rho_n(|I_{1n}| + \cdots + |I_{Mn}|) = MI_0\rho_n \tag{2.4}$$

when $p = 2$

$$\|\boldsymbol{q}^n\|_p = \rho_n\big(I_{1n}^2 + \cdots + I_{Mn}^2\big)^{1/2} = \sqrt{Mk}\rho_n \tag{2.5}$$

where k is a constant for the fully developed speckle patterns [Goo15, chapter 7]. So the sparsity of $\|\boldsymbol{q}^n\|_p$ is equivalent with the sparsity of ρ_n when $p = 1$ or 2.

We note here that the blind-SIM-SD method presented in Section 1.5.2 is equivalent with the regularizer $\alpha\|\mathbf{Q}\|_{1,1} + \beta\|\mathbf{Q}\|_F$. To simplify the notation, we express the model in vector form. Firstly we vectorize the variables $\{\boldsymbol{q}_m\}_{m=1}^M$ and rewrite the data observation model in (1.36) as:

$$\mathfrak{y} = \mathcal{H}\mathfrak{q} + \mathfrak{e} \tag{2.6}$$

in which

$$\mathcal{H} = \mathbb{1}_M \otimes \mathbf{H} = \begin{bmatrix} \mathbf{H} & & & \\ & \mathbf{H} & & \\ & & \ddots & \\ & & & \mathbf{H} \end{bmatrix},\ \mathfrak{y} = \mathbf{vect}(\mathbf{Y}) = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix},\ \mathfrak{q} = \mathbf{vect}(\mathbf{Q}) = \begin{bmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q}_M \end{bmatrix} \tag{2.7}$$

Now the minimizing problem (2.1) in vector form is:

$$\arg\min_{\mathfrak{q}} \|\mathfrak{q}\|_{\mathcal{G}pq}^q \quad \text{s.t. } \|\mathcal{H}\mathfrak{q} - \mathfrak{y}\| \leq \epsilon \tag{2.8}$$

where $\|\mathfrak{q}\|_{\mathcal{G}pq}$ indicates the $\ell_{\mathcal{G}pq}$ norm of $\mathfrak{q}$ defined by:

$$\|\mathfrak{q}\|_{\mathcal{G}pq} = \Big(\sum_{n=1}^N \|\mathfrak{q}_{\mathcal{G}_n}\|_p^q\Big)^{1/q} \tag{2.9}$$

in which $\mathfrak{q}_{\mathcal{G}_n} = \boldsymbol{q}^n$ is the n-th row of $\mathbf{Q}$. Obviously the definition of $\|\mathfrak{q}\|_{\mathcal{G}pq}$ is consistent with $\|\mathbf{Q}\|_{pq}$.

2.2.1 Unconstrained BPDN model

Rather than solving (2.8) directly, I rather consider its unconstrained form:

$$\arg\min_{\mathfrak{q}} \|\mathfrak{q}\|_{\mathcal{G}pq}^q + \frac{1}{2\mu}\|\mathcal{H}\mathfrak{q} - \mathfrak{y}\|_2^2 \tag{2.10}$$

where $\mu \in \mathbb{R}^+$ is the regularization parameter. (2.8) and (2.10) are equivalent in the following sense: for any $\epsilon > 0$, the solution of (2.8) is either the null vector, or otherwise it is a solution of (2.10) for some $\mu > 0$ [Roc15].

There is no closed form solution for problem (2.10). To overcome this difficulty, I introduce an auxiliary variable $\mathfrak{d}$ and rewrite (2.10) as:

$$\arg\min_{\mathfrak{q}} \|\mathfrak{d}\|_{\mathcal{G}pq}^q + \frac{\beta_2}{2}\|\mathcal{H}\mathfrak{q} - \mathfrak{y}\|_2^2 \quad \text{s.t. } \mathfrak{d} = \mathfrak{q} \tag{2.11}$$

Now the variables $\mathfrak{q}$ have been split into two blocks of variables $\mathfrak{q}$ and $\mathfrak{d}$ and the objective function is separable in the form of $f(\mathfrak{q}) + g(\mathfrak{d})$. Then the augmented Lagrangian associated with (2.11) is:

$$\min_{\mathfrak{q},\mathfrak{d}} \|\mathfrak{d}\|_{\mathcal{G}pq}^q + \frac{\beta_2}{2}\|\mathcal{H}\mathfrak{q} - \mathfrak{y}\|_2^2 - \boldsymbol{\lambda}_1^T(\mathfrak{d} - \mathfrak{q}) + \frac{\beta_1}{2}\|\mathfrak{d} - \mathfrak{q}\|_2^2 \tag{2.12}$$

with $\boldsymbol{\lambda}_1 \in \mathbb{R}^{MN}$ and $\beta_1 > 0$. The ADMM method is used to solve problem (2.12) by minimization with respect to $\mathfrak{q}$ and $\mathfrak{d}$ alternately. The $\mathfrak{q}$-subproblem is given by:

$$\begin{aligned} &\min_{\mathfrak{q}\geq 0} \boldsymbol{\lambda}_1^T \mathfrak{q} + \frac{\beta_1}{2}\|\mathfrak{d} - \mathfrak{q}\|_2^2 + \frac{\beta_2}{2}\|\mathcal{H}\mathfrak{q} - \mathfrak{y}\|_2^2 \\ \Leftrightarrow\ &\min_{\mathfrak{q}\geq 0} \frac{1}{2}\mathfrak{q}^T(\beta_1 \mathbb{1}_{MN} + \beta_2 \mathcal{H}^T\mathcal{H})\mathfrak{q} - (\beta_1\mathfrak{d} - \boldsymbol{\lambda}_1 + \beta_2\mathcal{H}^T\,\mathfrak{y})^T\mathfrak{q} \end{aligned} \tag{2.13}$$

Note that (2.13) is a convex quadratic problem and it is equivalent to the following linear system:

$$(\beta_1 \mathbb{1}_{MN} + \beta_2 \mathcal{H}^T\mathcal{H})\mathfrak{q} = \beta_1\mathfrak{d} - \boldsymbol{\lambda}_1 + \beta_2\mathcal{H}^T\,\mathfrak{y} \tag{2.14}$$

An exact solution gives:

$$\mathfrak{q} = \left(\beta_1 \mathbb{1}_{MN} + \beta_2 \mathcal{H}^T\mathcal{H}\right)^{-1}\left(\beta_1\mathfrak{d} - \boldsymbol{\lambda}_1 + \beta_2\mathcal{H}^T\,\mathfrak{y}\right) \tag{2.15}$$

or we could use a gradient based iteration algorithm to update $\mathfrak{q}$:

$$\mathfrak{q}_{k+1} = \mathfrak{q}_k - \alpha_k \nabla\mathfrak{q}_k \tag{2.16}$$

where the subscript k indicates iteration number and the gradient $\nabla\mathfrak{q}_k$ is given by:

$$\nabla\mathfrak{q}_k = \beta_1(\mathfrak{q}_k - \mathfrak{d}_k) + \boldsymbol{\lambda}_1 + \beta_2\mathcal{H}^T(\mathcal{H}\mathfrak{q}_k - \mathfrak{y}) \tag{2.17}$$

The corresponding $\mathfrak{d}$-subproblem of (2.12) is given by:

$$\min_{\mathfrak{d}} \|\mathfrak{d}\|_{\mathcal{G}pq}^q - \boldsymbol{\lambda}_1^T\mathfrak{d} + \frac{\beta_1}{2}\|\mathfrak{d} - \mathfrak{q}\|_2^2 \tag{2.18}$$

According to the definition of the $\ell_{\mathcal{G},p,q}$ norm, we can reexpress (2.18) as:

$$\min_{\mathfrak{d}} \sum_{n=1}^{N}\left[\|\mathfrak{d}_{\mathcal{G}_n}\|_p^q + \frac{\beta_1}{2}\left\|\mathfrak{d}_{\mathcal{G}_n} - \mathfrak{q}_{\mathcal{G}_n} - \frac{1}{\beta_1}(\boldsymbol{\lambda}_1)_{\mathcal{G}_n}\right\|_2^2\right] \tag{2.19}$$

Let us denote:

$$\boldsymbol{\tau}_n = \mathfrak{q}_{\mathcal{G}_n} + \frac{1}{\beta_1}(\boldsymbol{\lambda}_1)_{\mathcal{G}_n} \tag{2.20}$$

and

$$f_{n,p,q}(\mathfrak{d}_{\mathcal{G}_n}) = \|\mathfrak{d}_{\mathcal{G}_n}\|_p^q + \frac{\beta_1}{2}\|\mathfrak{d}_{\mathcal{G}_n} - \boldsymbol{\tau}_n\|_2^2 \tag{2.21}$$

Then the problem (2.19) has closed form solutions for the following (p, q) pairs [HLM+17]:

— for $p = 2$ and $q = 1$

$$\mathfrak{d}_{\mathcal{G}_n} = \max\left\{1 - \frac{1}{\beta_1\|\boldsymbol{\tau}_n\|_2}, 0\right\}\boldsymbol{\tau}_n \tag{2.22}$$

— for $p = 2$ and $q = 0$

$$\mathfrak{d}_{\mathcal{G}_n} = \begin{cases} \boldsymbol{\tau}_n, & \|\boldsymbol{\tau}_n\|_2 > \sqrt{(2/\beta_1)} \\ \mathbf{0} \text{ or } \boldsymbol{\tau}_n, & \|\boldsymbol{\tau}_n\|_2 = \sqrt{(2/\beta_1)} \\ \mathbf{0}, & \|\boldsymbol{\tau}_n\|_2 < \sqrt{(2/\beta_1)} \end{cases} \tag{2.23}$$

— for $p = 2$ and $q = 1/2$

$$\mathfrak{d}_{\mathcal{G}_n} = \begin{cases} \frac{16\|\boldsymbol{\tau}_n\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi_n}{3})}{\frac{3\sqrt{3}}{\beta 1}+16\|\boldsymbol{\tau}_n\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi_n}{3})}\boldsymbol{\tau}_n, & \|\boldsymbol{\tau}_n\|_2 > \frac{3}{2}(1/\beta_1)^{2/3} \\ \mathbf{0} \text{ or } \frac{16\|\boldsymbol{\tau}_n\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi_n}{3})}{\frac{3\sqrt{3}}{\beta 1}+16\|\boldsymbol{\tau}_n\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi_n}{3})}\boldsymbol{\tau}_n, & \|\boldsymbol{\tau}_n\|_2 = \frac{3}{2}(1/\beta_1)^{2/3} \\ \mathbf{0}, & \|\boldsymbol{\tau}_n\|_2 < \frac{3}{2}(1/\beta_1)^{2/3} \end{cases} \tag{2.24}$$

with

$$\psi_n = \arccos\left(\frac{1}{4\beta_1}\Big(\frac{3}{\|\boldsymbol{\tau}_n\|_2}\Big)^{3/2}\right)$$

— for $p = 2$ and $q = 2/3$

$$\mathfrak{d}_{\mathcal{G}_n} = \begin{cases} \frac{3\eta^4}{\frac{2}{\beta_1}+3\eta^4}\boldsymbol{\tau}_n, & \|\boldsymbol{\tau}_n\|_2 > 2(\frac{2}{3\beta_1})^{3/4} \\ 0 \text{ or } \frac{3\eta^4}{\frac{2}{\beta_1}+3\eta^4}\boldsymbol{\tau}_n, & \|\boldsymbol{\tau}_n\|_2 = 2(\frac{2}{3\beta_1})^{3/4} \\ 0, & \|\boldsymbol{\tau}_n\|_2 < 2(\frac{2}{3\beta_1})^{3/4} \end{cases} \tag{2.25}$$

with

$$\eta = \frac{1}{2}\left(|a| + \sqrt{\frac{2\|\boldsymbol{\tau}_n\|_2}{|a|} - a^2}\right), \quad a = \frac{2}{\sqrt{3}}\big(\frac{2}{\beta_1}\big)^{1/4}\left(\cosh\left(\frac{\phi(\boldsymbol{\tau}_n)}{3}\right)\right)^{1/2}$$
$$\phi(\boldsymbol{\tau}_n) = \operatorname{arccosh}\left(\frac{27\|\boldsymbol{\tau}_n\|_2^2}{16\big(\frac{2}{\beta_1}\big)^{3/2}}\right) \tag{2.26}$$

— for $p = 1$ and $q = 1/2$

$$\mathfrak{d}_{\mathcal{G}_n} = \begin{cases} \ddot{\boldsymbol{\tau}}_n, & f_{n,1,1/2}(\ddot{\boldsymbol{\tau}}_n) < f_{n,1,1/2}(0) \\ 0 \text{ or } \ddot{\boldsymbol{\tau}}_n, & f_{n,1,1/2}(\ddot{\boldsymbol{\tau}}_n) = f_{n,1,1/2}(0) \\ 0, & f_{n,1,1/2}(\ddot{\boldsymbol{\tau}}_n) > f_{n,1,1/2}(0) \end{cases} \tag{2.27}$$

with

$$\ddot{\boldsymbol{\tau}}_n = \boldsymbol{\tau}_n - \frac{\sqrt{3}}{4\beta\sqrt{\|\boldsymbol{\tau}_n\|_1}\cos\left(\frac{\pi}{3}-\frac{\xi(\boldsymbol{\tau}_n)}{3}\right)}\operatorname{sign}(\boldsymbol{\tau}_n)$$
$$\xi(\boldsymbol{\tau}_n) = \arccos\left(\frac{M}{4\beta}\Big(\frac{3}{\|\boldsymbol{\tau}_n\|_1}\Big)^{3/2}\right) \tag{2.28}$$

— for $p = 1$ and $q = 2/3$

$$\mathfrak{d}_{\mathcal{G}_n} = \begin{cases} \breve{\boldsymbol{\tau}}_n, & f_{n,1,2/3}(\breve{\boldsymbol{\tau}}_n) < f_{n,1,2/3}(0) \\ 0 \text{ or } \breve{\boldsymbol{\tau}}_n, & f_{n,1,2/3}(\breve{\boldsymbol{\tau}}_n) = f_{n,1,2/3}(0) \\ 0, & f_{n,1,2/3}(\breve{\boldsymbol{\tau}}_n) > f_{n,1,2/3}(0) \end{cases} \tag{2.29}$$

with

$$\breve{\boldsymbol{\tau}}_n = \boldsymbol{\tau}_n - \frac{4\sqrt{\bar{a}}}{3\beta\big(\bar{a}^{3/2} + \sqrt{2|\boldsymbol{\tau}_n|_1 - \bar{a}^3}\big)}\operatorname{sign}(\boldsymbol{\tau}_n) \tag{2.30}$$

and

$$\bar{a} = \frac{2}{\sqrt{3}}\Big(\frac{2M}{\beta}\Big)^{1/4}\left(\cosh\left(\frac{\zeta(\boldsymbol{\tau}_n)}{3}\right)\right)^{1/2}$$
$$\zeta(\boldsymbol{\tau}_n) = \operatorname{arccosh}\left(\frac{27|\boldsymbol{\tau}_n|_1^2}{16\big(\frac{2M}{\beta}\big)^{3/2}}\right) \tag{2.31}$$

Finaly, we update the multiplier $\boldsymbol{\lambda}_1$ by:

$$\boldsymbol{\lambda}_1 \leftarrow \boldsymbol{\lambda}_1 - \alpha_1\beta_1(\mathfrak{d} - \mathfrak{q}) \tag{2.32}$$

where $\alpha_1 > 0$ denotes the step length. Briefly, the ADMM approach for the constrained BPDN model is shown in Algorithm 1:

Algorithm 1: ADMM algorithm for $\ell_{p,q}$ minimization of unconstrained BPDN model

```
Initialize λ_1, λ_2 ∈ R^{MN}, β_1 > 0 and α_1 > 0 ;
while stopping criterion is not met do
    Update q by (2.15) or (2.16) ;
    Update d according the number of (p, q) pair;
    λ_1 ← λ_1 − α_1 β_1 (d − q) ;
end
```

2.2.2 Constrained BPDN model

Although the unconstrained form of BPDN model (2.10) and its constrained form (2.8) are equivalent, it is not easy to set the parameter μ in (2.10). However, the parameter ϵ has a clear meaning related to the energy of the noise. Inspired by the so-called C-SALSA algorithm [ABDF11], we present an ADMM approach to solve (2.8) directly.

Firstly, we define the ellipsoid feasible set $\mathbf{E}(\epsilon, \mathbf{H}, \boldsymbol{y})$ as:

$$\mathbf{E}(\epsilon, \mathbf{H}, \boldsymbol{y}) = \left\{\boldsymbol{\rho} \in \mathbb{R}^N \mid \|\mathbf{H}\boldsymbol{\rho} - \boldsymbol{y}\|_2 \leq \epsilon\right\} \tag{2.33}$$

Then the constrained problem (2.8) can be written as an unconstrained problem according to:

$$\arg\min_{\mathfrak{q}} \|\mathfrak{q}\|_{\mathcal{G}pq}^q + \imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}(\mathcal{H}\mathfrak{q}) \tag{2.34}$$

To solve the problem (2.34), we introduce two auxiliary variables $\mathfrak{o}$ and $\mathfrak{z}$ by:

$$\mathfrak{o} = \begin{bmatrix} \mathfrak{o}^{(1)} \\ \mathfrak{o}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathfrak{q} \\ \mathcal{H}\mathfrak{q} \end{bmatrix}, \qquad \mathfrak{z} = \begin{bmatrix} \mathfrak{q} \\ \mathfrak{o} \end{bmatrix} \tag{2.35}$$

Then we can write (2.34) as:

$$\arg\min_{\mathfrak{z}} \; h(\mathfrak{z}), \qquad \text{s.t. } \mathcal{A}\mathfrak{z} = 0 \tag{2.36}$$

where

$$\begin{aligned} h(\mathfrak{z}) = f(\mathfrak{q}) + g(\mathfrak{o}); \qquad f(\mathfrak{q}) = 0; \quad g(\mathfrak{o}) = \|\mathfrak{o}^{(1)}\|_{\mathcal{G}pq}^q + \imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}(\mathfrak{o}^{(2)}) \\ \mathcal{A} = \left(\mathcal{B}, -\mathbb{1}_{2MN}\right), \qquad \mathcal{B} = \begin{pmatrix} \mathbb{1}_{MN} \\ \mathcal{H} \end{pmatrix} \end{aligned} \tag{2.37}$$

The augmented Lagrangian associated with (2.36) is:

$$\begin{aligned} &\arg\min_{\mathfrak{q}} \; f(\mathfrak{q}) + g(\mathfrak{o}) - \boldsymbol{\lambda}\mathcal{A}\mathfrak{z} + \frac{\beta}{2}\|\mathcal{A}\mathfrak{z}\|^2 \\ \Leftrightarrow \; &\arg\min_{\mathfrak{q}} \; f(\mathfrak{q}) + g(\mathfrak{o}) + \frac{\beta}{2}\|\mathcal{B}\mathfrak{q} - \mathfrak{o} - \frac{\boldsymbol{\lambda}}{\beta}\|^2 \end{aligned} \tag{2.38}$$

The $\mathbf{q}$-subproblem is given by:

$$\begin{aligned}\arg\min_{\mathbf{q}} \|\mathcal{B}\mathbf{q} - \boldsymbol{\zeta}_k\|^2 &= (\mathcal{B}^T\mathcal{B})^{-1}\mathcal{B}^T\boldsymbol{\zeta}_k \\ &= \left(\mathbb{1}_{MN} + \mathcal{H}^T\mathcal{H}\right)^{-1}\left(\boldsymbol{\zeta}_k^{(1)} + \mathcal{H}^T\boldsymbol{\zeta}_k^{(2)}\right)\end{aligned}$$

where k indicates the number of iterations and $\boldsymbol{\zeta}_k = \mathbf{o}_k + \frac{\boldsymbol{\lambda}_k}{\beta}$.

The $\mathbf{o}$-subproblem of (2.38) is given by:

$$\begin{aligned}&\arg\min_{\mathbf{o}}\ g(\mathbf{o}) + \frac{\beta}{2}\|\mathcal{B}\mathbf{q}_{k+1} - \mathbf{o} - \frac{\boldsymbol{\lambda}_k}{\beta}\|^2 \\ \Leftrightarrow\ &\arg\min_{\mathbf{o}}\ \|\mathbf{o}^{(1)}\|_{\mathcal{G}pq}^q + \imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}(\mathbf{o}^{(2)}) + \frac{\beta}{2}\|\mathbf{o} - \mathfrak{s}_k\|^2\end{aligned} \tag{2.39}$$

where $\mathfrak{s}_k = \mathcal{B}\mathbf{q}_{k+1} - \frac{\boldsymbol{\lambda}_k}{\beta}$. Clearly, the minimization $\mathbf{o}$-subproblem could be decoupled, leading to

$$\mathbf{o}_{k+1}^{(1)} = \arg\min_{\mathbf{o}^{(1)}}\ \|\mathbf{o}^{(1)}\|_{\mathcal{G}pq}^q + \frac{\beta}{2}\|\mathbf{o}^{(1)} - \mathfrak{s}_k^{(1)}\|^2 \tag{2.40a}$$

$$\mathbf{o}_{k+1}^{(2)} = \arg\min_{\mathbf{o}^{(2)}}\ \imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}(\mathbf{o}^{(2)}) + \frac{\beta}{2}\|\mathbf{o}^{(2)} - \mathfrak{s}_k^{(2)}\|^2 \tag{2.40b}$$

which are the proximal operators of the corresponding functions. The proximal operator of $\ell_{p,q}$ norm has been shown in previous Section 2.2. The proximal operator of indicator function $\imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}$ is [ABDF11]:

$$\text{prox}_{\imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}/\beta}(\boldsymbol{s}) = \mathfrak{y} + \begin{cases}\epsilon\frac{\boldsymbol{s}-\mathfrak{y}}{\|\boldsymbol{s}-\mathfrak{y}\|_2}, & \|\boldsymbol{s}-\mathfrak{y}\| \geq \epsilon \\ \boldsymbol{s}-\mathfrak{y}, & \|\boldsymbol{s}-\mathfrak{y}\| \leq \epsilon\end{cases} \tag{2.41}$$

Finally, the multipliers $\boldsymbol{\lambda}$ are updated by:

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \beta(\mathcal{A}\boldsymbol{\mathfrak{z}}_{k+1}) = \boldsymbol{\lambda}_k + \beta(\mathcal{B}\mathbf{q}_{k+1} - \mathbf{o}_{k+1}) \tag{2.42}$$

The resulting ADMM algorithm to solve the unconstrained BPDN model is as follows:

Algorithm 2: ADMM algorithm for $\ell_{p,q}$ minimization of constrained BPDN model

Initialize $\boldsymbol{\lambda} \in \mathbb{R}^{2MN}$, $\beta > 0$;
while *stopping criterion is not met* **do**
  $\boldsymbol{\zeta}_k = \mathbf{o}_k + \frac{\boldsymbol{\lambda}_k}{\beta}$;
  $\mathbf{q}_{k+1} = \left(\mathbb{1}_{MN} + \mathcal{H}^T\mathcal{H}\right)^{-1}\left(\boldsymbol{\zeta}_k^{(1)} + \mathcal{H}^T\boldsymbol{\zeta}_k^{(2)}\right)$;
  $\mathfrak{s}_k = \mathcal{B}\mathbf{q}_{k+1} - \frac{\boldsymbol{\lambda}_k}{\beta}$;
  $\mathbf{o}_{k+1}^{(1)} = \text{prox}_{\ell_{p,q}/\beta}(\mathfrak{s}_k^{(1)})$;
  $\mathbf{o}_{k+1}^{(2)} = \text{prox}_{\imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}/\beta}(\mathfrak{s}_k^{(2)})$;
  $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \beta(\mathcal{B}\mathbf{q}_{k+1} - \mathbf{o}_{k+1})$;
  $k \leftarrow k + 1$;
end

2.3 $\ell_{p,q}$ norm plus TV norm minimization

In this section we incorporate more prior information on the object to the image reconstruction model, including positivity and the total variation (TV) regularizer:

$$\arg\min_{\mathbf{Q}} \|\mathbf{Q}\|_{pq}^{q} + \mu\|\boldsymbol{\rho}\|_{TV} \quad \text{s.t. } \|\mathbf{Y} - \mathbf{H}\mathbf{Q}\| \leq \epsilon,\ \mathbf{Q} \geq 0 \tag{2.43}$$

where the *isotropic* TV is adopted. For a $N_1 \times N_2$ object ρ, it is defined as:

$$\|\rho\|_{TV} = \sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2}\sqrt{(\rho[n_1+1,n_2]-\rho[n_1,n_2])^2 + (\rho[n_1,n_2+1]-\rho[n_1,n_2])^2} \tag{2.44}$$

Let us denote $\mathbf{C}_1, \mathbf{C}_2 \in \mathbb{R}^{N\times N}$ the two first-order forward finite difference operators (see their structures in Appendix C). For the vectorized object $\boldsymbol{\rho} \in \mathbb{R}^N$, we have:

$$\|\boldsymbol{\rho}\|_{TV} = \|\mathbf{C}\boldsymbol{\rho}\|_{\mathcal{G},2,1} = \sum_{n=1}^{N}\|(\mathbf{C}\boldsymbol{\rho})_{\mathcal{G}_n}\|_2 \tag{2.45}$$

with $\mathbf{C} = (\mathbf{C}_1;\mathbf{C}_2) \in \mathbb{R}^{2N\times N}$ denoting the total finite difference operator and $(\mathbf{C}\boldsymbol{\rho})_{\mathcal{G}_n} = [(\mathbf{C}_1\boldsymbol{\rho})_n; (\mathbf{C}_2\boldsymbol{\rho})_n] \in \mathbb{R}^2$. According to the property $\boldsymbol{\rho} = \frac{1}{MI_0}\sum_m \boldsymbol{q}_m$, we could write $\boldsymbol{\rho} = \mathcal{A}\mathbf{q}$, with matrix $\mathcal{A} \in \mathbb{R}^{N\times MN}$ given by:

$$\mathcal{A} = \left[\ \tfrac{1}{MI_0}\mathbb{1}_N, \cdots, \tfrac{1}{MI_0}\mathbb{1}_N\ \right] \tag{2.46}$$

Let us denote $\mathcal{D} = \mathbf{C}\mathcal{A}$, then the corresponding vectorized form of (2.43) is:

$$\arg\min_{\mathfrak{q}} \|\mathfrak{q}\|_{\mathcal{G}pq}^{q} + \imath_{\mathbb{R}^+}(\mathfrak{q}) + \mu\|\mathcal{D}\mathfrak{q}\|_{\mathcal{G}21} \quad \text{s.t. } \|\mathcal{H}\mathfrak{q} - \mathfrak{y}\| \leq \epsilon \tag{2.47}$$

with

$$\imath_{\mathbb{R}^+}(\mathfrak{q}) = \begin{cases} 0, & \mathfrak{q} \geq 0 \\ +\infty, & \text{otherwise} \end{cases}, \qquad \|\mathcal{D}\mathfrak{q}\|_{\mathcal{G}21} = \sum_{n=1}^{N}\|(\mathcal{D}\mathfrak{q})_{\mathcal{G}_n}\|_2 \tag{2.48}$$

2.3.1 Primal-dual algorithm for unconstrained BPDN model

Problem (2.47) can be solved by transforming it to its equivalent unconstrained form:

$$\arg\min_{\mathfrak{q}} \frac{1}{2}\|\mathcal{H}\mathfrak{q} - \mathfrak{y}\|^2 + \beta_1\imath_{\mathbb{R}^+}(\mathfrak{q}) + \beta_2\|\mathfrak{q}\|_{\mathcal{G}pq}^{q} + \beta_3\|\mathcal{D}\mathfrak{q}\|_{\mathcal{G}21} \tag{2.49}$$

in which $\beta_2, \beta_3 \geq 0$. Here we add parameter $\beta_1 \in \{0,1\}$, so that the model can be solved with or without the positivity constraint. When $\beta_1 = 1$ and $\beta_2 = \beta_3 = 0$, Eq. (2.49) reduces to the model presented in Section 1.5.1. Here we choose the primal-dual splitting method proposed in [Con13] to solve the optimization problem in (2.49). Firstly, we introduce two auxiliary variables $\mathfrak{d}, \mathfrak{p}$ by $\mathfrak{d} = \mathfrak{q}$ and $\mathfrak{p} = \mathcal{D}\mathfrak{q}$ and we rewrite (2.49) according to:

$$\begin{aligned} &\arg\min_{\mathfrak{q}}\ f(\mathfrak{q}) + g(\mathfrak{q}) + h_1(\mathfrak{d}) + h_2(\mathfrak{p}) \\ \text{with} \qquad &f(\mathfrak{q}) = \frac{1}{2}\|\mathcal{H}\mathfrak{q} - \mathfrak{y}\|^2, \quad g(\mathfrak{q}) = \beta_1\imath_{\mathbb{R}^+}(\mathfrak{q}) \\ &h_1(\mathfrak{d}) = \beta_2\|\mathfrak{d}\|_{\mathcal{G}pq}^{q} \quad h_2(\mathfrak{p}) = \beta_3\|\mathfrak{q}\|_{\mathcal{G}21} \end{aligned} \tag{2.50}$$

The dual problem associated with (2.50) is:

$$\begin{aligned} &\arg\min_{\mathfrak{d},\mathfrak{p}} \ \Big[(f+g)^*(-\mathfrak{d}-\mathcal{D}^*\mathfrak{p})+h_1^*(\mathfrak{d})+h_2^*(\mathfrak{p})\Big] \\ \equiv\ &\arg\min_{\mathfrak{d},\mathfrak{p}}\min_{\mathfrak{q}} \ \Big[f^*(-\mathfrak{d}-\mathcal{D}^*\mathfrak{p}-\mathfrak{q})+g^*(\mathfrak{q})+h_1^*(\mathfrak{d})+h_2^*(\mathfrak{p})\Big] \end{aligned} \tag{2.51}$$

where we have used the following property:

$$(f+g)^*(-\mathfrak{d}-\mathcal{D}^*\mathfrak{p})=\min_{\mathfrak{q}} \ f^*(-\mathfrak{d}-\mathcal{D}^*\mathfrak{p}-\mathfrak{q})+g^*(\mathfrak{q}) \tag{2.52}$$

in which $\mathcal{D}^*$ denotes the conjugate transpose of matrix $\mathcal{D}$ and f^* is the convex conjugate of function f defined as:

$$f^*(x^*)=\sup\big\{\langle x^*,x\rangle-f(x)\big\} \tag{2.53}$$

According to classical Karush-Kuhn-Tucker theory, if $(\hat{\mathfrak{q}},\hat{\mathfrak{d}},\hat{\mathfrak{p}})$ is a solution of the monotone variational inclusion:

$$\begin{pmatrix} 0\\0\\0 \end{pmatrix} \in \begin{pmatrix} \partial g(\hat{\mathfrak{q}})+\hat{\mathfrak{d}}+\mathcal{D}^*\hat{\mathfrak{p}}+\nabla f(\hat{\mathfrak{q}}) \\ -\hat{\mathfrak{q}}+\partial h_1^*(\hat{\mathfrak{d}}) \\ -\mathcal{D}\hat{\mathfrak{q}}+\partial h_2^*(\hat{\mathfrak{p}}) \end{pmatrix} \tag{2.54}$$

then $\hat{\mathfrak{q}}$ is solution to the primal problem (2.50) and $\hat{\mathfrak{d}},\hat{\mathfrak{p}}$ is the solution to the dual problem (2.51). The primal-dual splitting method to solve (2.54) is shown in Algorithm 3:

Algorithm 3: Primal-dual splitting method to solve $\ell_{p,q}$ plus TV norm minimization

Initialize the parameter $\tau,\sigma,\theta>0$ and $\mathfrak{q}_0,\mathfrak{d}_0,\mathfrak{p}_0$
while *stopping criterion is not met* **do**
  if $\beta_1=1$ **then**
    $\bar{\mathfrak{q}}_{k+1}=\max\Big(\mathfrak{q}_k-\tau\big(\nabla f(\mathfrak{q}_k)+\mathfrak{d}_k+\mathcal{D}^*\mathfrak{p}_k\big),0\Big)$;
  else
    $\bar{\mathfrak{q}}_{k+1}=\mathfrak{q}_k-\tau\big(\nabla f(\mathfrak{q}_k)+\mathfrak{d}_k+\mathcal{D}^*\mathfrak{p}_k\big)$;
  end
  $\mathfrak{q}_{k+1}=\theta\bar{\mathfrak{q}}_{k+1}+(1-\theta)\mathfrak{q}_k$;
  $\bar{\mathfrak{d}}_{k+1}=\text{prox}_{\sigma h_1^*}\big(\mathfrak{d}_k+\sigma(2\bar{\mathfrak{q}}_{k+1}-\mathfrak{q}_k)\big)$;
  $\bar{\mathfrak{p}}_{k+1}=\text{prox}_{\sigma h_2^*}\big(\mathfrak{p}_k+\sigma\mathcal{D}(2\bar{\mathfrak{q}}_{k+1}-\mathfrak{q}_k)\big)$;
  $\mathfrak{d}_{k+1}=\theta\bar{\mathfrak{d}}_{k+1}+(1-\theta)\mathfrak{d}_k$;
  $\mathfrak{p}_{k+1}=\theta\bar{\mathfrak{p}}_{k+1}+(1-\theta)\mathfrak{p}_k$;
end

The convergence results of algorithm 3 is as follows [Con13]:

Property 2.1. Let L be the Lipschitz constant of the gradient ∇f (or function f is L-smooth) and $\tau,\sigma>0$. When $p\geq 1$ and $q=1$, the sequence $\mathfrak{q}_k$ generated by Algorithm 3 converges to a solution of problem (2.50) as long as the following holds:

— $\tau\Big(\frac{L}{2}+\sigma\big\|\mathbb{1}_{MN}+\mathcal{D}^*\mathcal{D}\big\|_{op}\Big)<1$, where $\|\cdot\|_{op}$ is the *operator norm*.

— $\theta\in(0,1]$

When $q<1$, ℓ_{pq} is a nonconvex function and we cannot assure Algorithm 3 converges to its global minimum.

The Lipschitz constant of $\nabla f(\mathfrak{q}) = \mathcal{H}^*\big(\mathcal{H}\mathfrak{q} - \mathfrak{y}\big)$ is: $L = \|\mathcal{H}\|_{op}^2 = \|\mathbb{1}_M\|_{op}^2\|\mathbf{H}\|_{op}^2$. Since $\mathbf{H}$ is a lowpass convolution operator with symmetric boundary conditions, we have $\|\mathbf{H}\|_{op} = 1$, so that $L = 1$.

According of the definition of norm,

$$\big\|\mathbb{1}_{MN} + \mathcal{D}^*\mathcal{D}\big\|_{op} \leq \|\mathbb{1}_{MN}\|_{op} + \|\mathcal{D}^*\mathcal{D}\|_{op} \tag{2.55}$$

with

$$\|\mathcal{D}^*\mathcal{D}\|_{op} = \|\mathcal{D}\|_{op}^2 \leq \|\mathcal{A}\|_{op}^2\|\mathbf{C}\|_{op}^2 \tag{2.56}$$

According to the inequality between root-mean square and arithmetic mean, we have:

$$\|\mathcal{A}\mathfrak{q}\|_2^2 \leq \frac{1}{MI_0}\|\mathfrak{q}\|_2^2 \tag{2.57}$$

Since $\|\mathbf{C}\|_{op}^2 \leq 8$ [CP11], combining Eqs. (2.55)–(2.57):

$$\big\|\mathbb{1}_{MN} + \mathcal{D}^*\mathcal{D}\big\|_{op} \leq 1 + \frac{8}{MI_0} \tag{2.58}$$

The proximal operator for h_1^* could be found through the relation:

$$\text{prox}_{\sigma h_1^*}(\boldsymbol{x}) = \boldsymbol{x} - \sigma\text{prox}_{h_1/\sigma}(\boldsymbol{x}/\sigma) \tag{2.59}$$

2.3.2 Primal-dual algorithm for constrained BPDN model

The constrained form of (2.47) is:

$$\arg\min_{\mathfrak{q}} \ \imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}(\mathcal{H}\mathfrak{q}) + \beta_1\imath_{\mathbb{R}^+}(\mathfrak{q}) + \beta_2\|\mathfrak{q}\|_{\mathcal{G}pq}^q + \beta_3\|\mathcal{D}\mathfrak{q}\|_{\mathcal{G}21} \tag{2.60}$$

To solve (2.60) with primal-dual algorithm, we introduce the auxiliary variables $\mathfrak{d}, \mathfrak{p}, \mathfrak{r}$ with $\mathfrak{d} = \mathfrak{q}$ and $\mathfrak{p} = \mathcal{D}\mathfrak{q}$ and $\mathfrak{r} = \mathcal{H}\mathfrak{q}$. Then (2.60) can be rewritten as:

$$\begin{aligned} &\arg\min_{\mathfrak{q}} \ g(\mathfrak{q}) + h_1(\mathfrak{d}) + h_2(\mathfrak{p}) + h_3(\mathfrak{r}) \\ \text{with} \quad &g(\mathfrak{q}) = \beta_1\imath_{\mathbb{R}^+}(\mathfrak{q}), \quad h_1(\mathfrak{d}) = \beta_2\|\mathfrak{d}\|_{\mathcal{G}pq}^q \\ &h_2(\mathfrak{p}) = \beta_3\|\mathfrak{p}\|_{\mathcal{G}21}, \quad h_3(\mathfrak{r}) = \imath_{\mathbf{E}(\epsilon,\mathbb{1},\mathfrak{y})}(\mathfrak{r}) \end{aligned} \tag{2.61}$$

The primal-dual algorithm to solve (2.60) is as follows:

2.4 Simulation results

To study the numerical performance of the blind-SIM-JSP model, a 2D 'star-like' simulated target is used as the true object, with a fluorescence density given by $\rho(r,\theta) \propto 1 + \cos(40\theta)$ in polar coordinates, as shown in Figure 2.1. Its spatial frequencies are higher closer to the star center, making it easy to visualize the resolution improvement. The point spread function is chosen as:

$$h(r,\theta) = \left(\frac{J_1(NAk_0r)}{k_0r}\right)^2\frac{k_0^2}{\pi} \tag{2.62}$$

where J_1 is the first order Bessel function of the first kind, *NA* is the objective numerical aperture set to 1.49 and $k_0 = \frac{2\pi}{\lambda}$ is the free-space wavenumber with λ the emission and the excitation

Algorithm 4: Primal-dual splitting method to solve $\ell_{p,q}$ plus TV norm minimization of form 2

Initialize the parameter $\tau, \sigma, \theta > 0$ and $\mathbf{q}_0, \mathfrak{d}_0, \mathfrak{p}_0, \mathfrak{r}_0$
Set $k = 0$
while *stopping criterion is not met* **do**
  if $\beta_1 = 1$ **then**
    $\bar{\mathbf{q}}_{k+1} = \max\Big(\mathbf{q}_k - \tau\big(\mathfrak{d}_k + \mathcal{D}^*\mathfrak{p}_k + \mathcal{H}^*\mathfrak{r}_k\big), 0\Big)$;
  else
    $\bar{\mathbf{q}}_{k+1} = \mathbf{q}_k - \tau\big(\mathfrak{d}_k + \mathcal{D}^*\mathfrak{p}_k + \mathcal{H}^*\mathfrak{r}_k\big)$;
  end
  $\mathbf{q}_{k+1} = \theta\bar{\mathbf{q}}_{k+1} + (1-\theta)\mathbf{q}_k$;
  $\bar{\mathfrak{d}}_{k+1} = \text{prox}_{\sigma h_1^*}\big(\mathfrak{d}_k + \sigma(2\bar{\mathbf{q}}_{k+1} - \mathbf{q}_k)\big)$;
  $\bar{\mathfrak{p}}_{k+1} = \text{prox}_{\sigma h_2^*}\big(\mathfrak{p}_k + \sigma\mathcal{D}(2\bar{\mathbf{q}}_{k+1} - \mathbf{q}_k)\big)$;
  $\bar{\mathfrak{r}}_{k+1} = \text{prox}_{\sigma h_3^*}\big(\mathfrak{r}_k + \sigma\mathcal{H}(2\bar{\mathbf{q}}_{k+1} - \mathbf{q}_k)\big)$;
  $\mathfrak{d}_{k+1} = \theta\bar{\mathfrak{d}}_{k+1} + (1-\theta)\mathfrak{d}_k$;
  $\mathfrak{p}_{k+1} = \theta\bar{\mathfrak{p}}_{k+1} + (1-\theta)\mathfrak{p}_k$;
  $\mathfrak{r}_{k+1} = \theta\bar{\mathfrak{r}}_{k+1} + (1-\theta)\mathfrak{r}_k$;
  $k \leftarrow k+1$;
end

wavelengths. The speckle patterns are generated through the same optical device as the recorded images. The sampling step in the object should be finer than $\lambda/8NA$ to observe an SR factor of two. In our simulations, a larger SR factor is expected in some situations. Therefore, a sampling step of $\lambda/20$ is adopted so that aliasing does not destroy the attainable SR. For the sampling rate in the raw images, we will not lose the SR capacity as long as it is higher than the Nyquist rate $4NA/\lambda$. In the simulations presented in this chapter, we use the same sampling rate for the raw images and the object.

Firstly, we do simulations under 40dB Gaussian white noise and 300 speckle patterns with different regularizers. The average of the raw images (wide-field image) $\bar{\boldsymbol{y}}$ and its Wiener deconvolution by Eq. (1.50) are shown in Figure 2.2(b,c). It is clear that we see no super-resolution in the wide-field image and its deconvolution. The reconstructed image with positivity constraint is shown in Figure 2.2(d). We retrieve partial super-resolution with weak modulation contrast after introducing positivity constraint, which is consistent with the result shown in [MBG+12]. Figure 2.2(e,f) are obtained with $\ell_{2,0}$ norm regularizer with M-SBL algorithm as in [MJK+13] and $\ell_1 + \ell_2$ norm plus positivity regularizer with PPDS algorithm presented in [LNI+17], respectively. We can see that the image reconstructed by M-SBL algorithm as in [MJK+13] do not scale well and there are some artifacts in low resolution part.

The reconstructed objects obtained by using $\ell_{p,q}$ norms with different (p, q) pairs are shown in Figure 2.3. The positivity constraint and the TV norm are not considered. The images in first line are obtained by minimizing the unconstrained BPDN model (2.12) with ADMM algorithm and $\beta_1 = \beta_2 = 2$. The results in second line are obtained with primal-dual algorithm by minimizing unconstrained BPDN model (2.49) where we set $\beta_1 = \beta_3 = 0$ and $\beta_2 = 0.5$. Here $\sigma = \theta = 1$ and $\tau = 0.6$ satisfy the convergence conditions of primal-dual algorithm when $p \geq 1$ and $q = 1$.

While the results shown in third line in Figure 2.3 are obtained by minimizing the constrained BPDN model (2.60) with $\beta_1 = \beta_3 = 0$ and $\beta_2 > 0$. For the hyperparameter ϵ, we

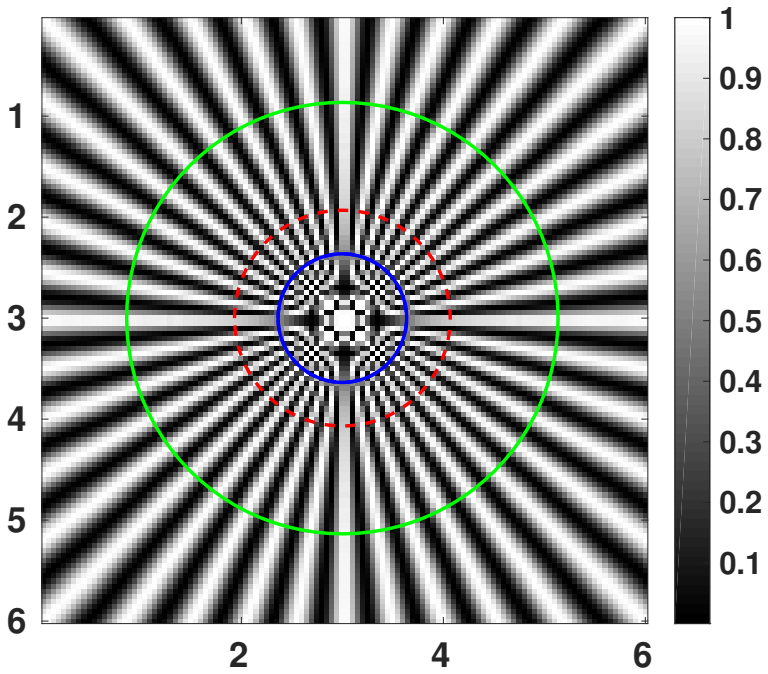


Figure 2.1 – **The true object.** The green solid circle corresponds to spatial frequencies transmitted by OTF support, and red dashed circle corresponds to two times of the OTF support. The Nyquist limit introduced by discretization is marked by the blue circle.

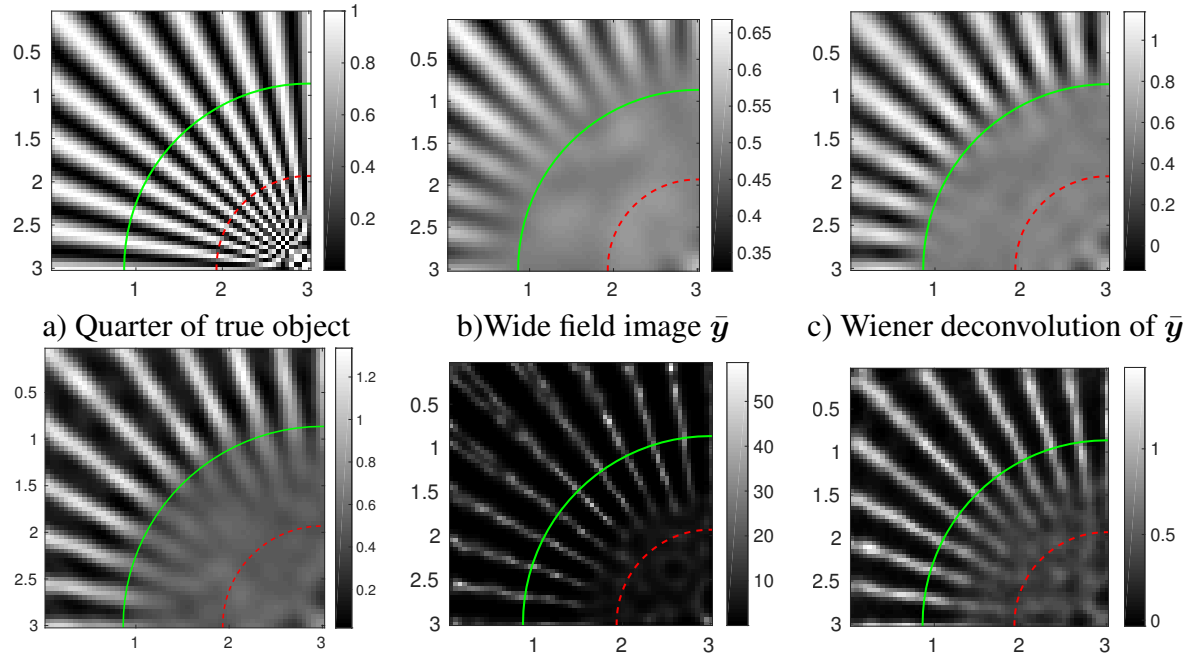
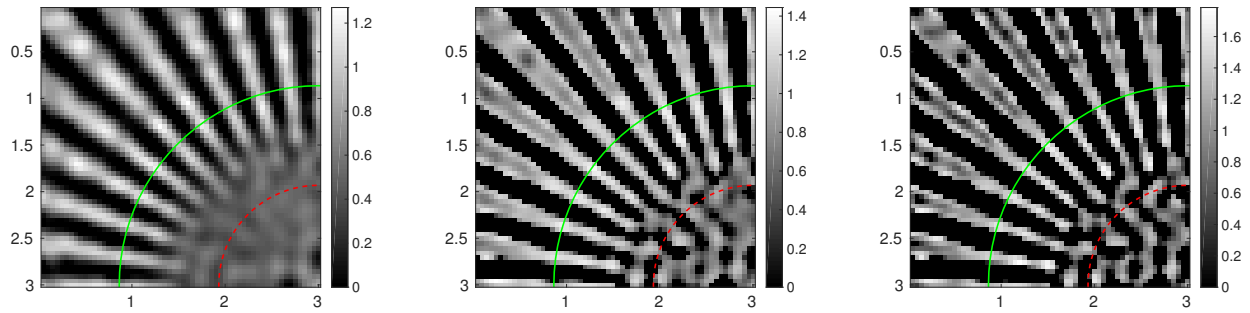
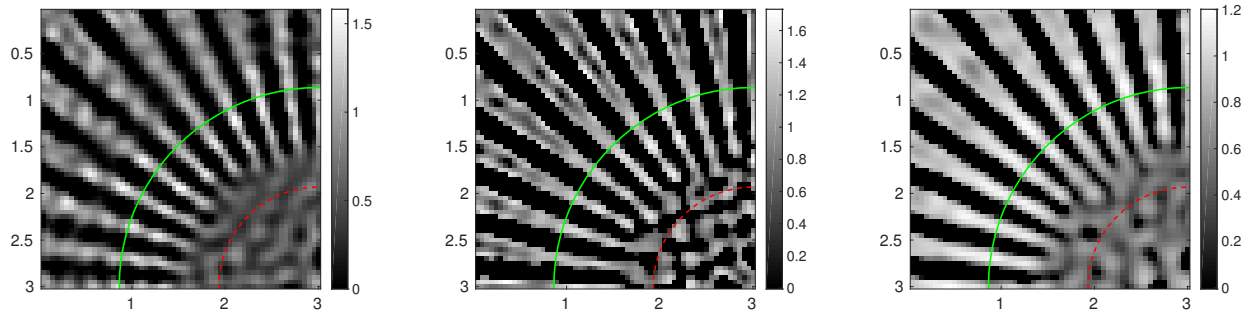


Figure 2.2 – **Joint reconstruction approach in blind-speckle-SIM using different regularizers with 300 speckle patterns and 40dB Gaussian noise.** d) is obtained using positivity constraint as shown in [MBG+12], e) is obtained with $\ell_{2,0}$ norm regularizer using M-SBL algorithm as in [MJK+13] and f) is obtained with $\ell_1 + \ell_2$ norm regularizer plus positivity constraint using PPDS algorithm shown in [LNI+17].

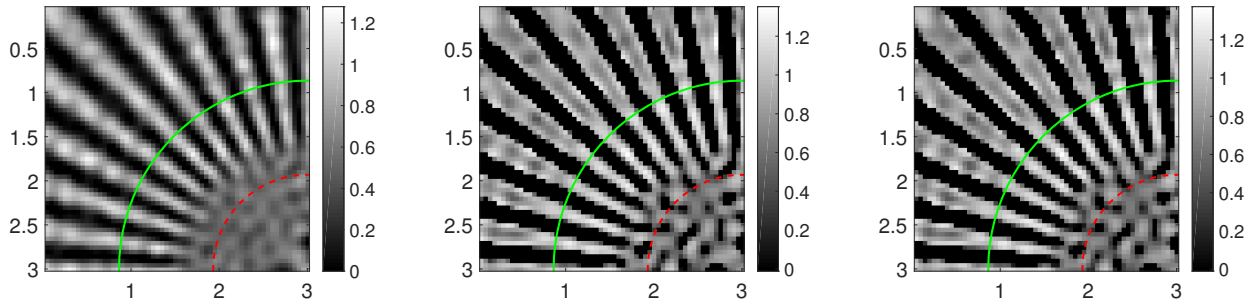
set it equal to its true value $\epsilon_{\text{real}} = \sqrt{MN}\nu$ with ν denoting the standard deviation of the noise, unless otherwise stated. The primal-dual algorithm converges if $\theta \in]0, 2[$ and $\sigma\tau \leq \frac{1}{2}$ [Con13]. It shows that the $\ell_{p,q}$ norm regularizer can induce good super-resolution results. Particularly, for the constrained BPDN model, we retrieve the super-resolution information twice better than the conventional wide-field image, as good as standard SIM. However, let us remark that prior information introduces bias to the joint estimator.



ADMM for unconstrained BPDN



Primal-Dual for unconstrained BPDN



Primal-Dual for constrained BPDN

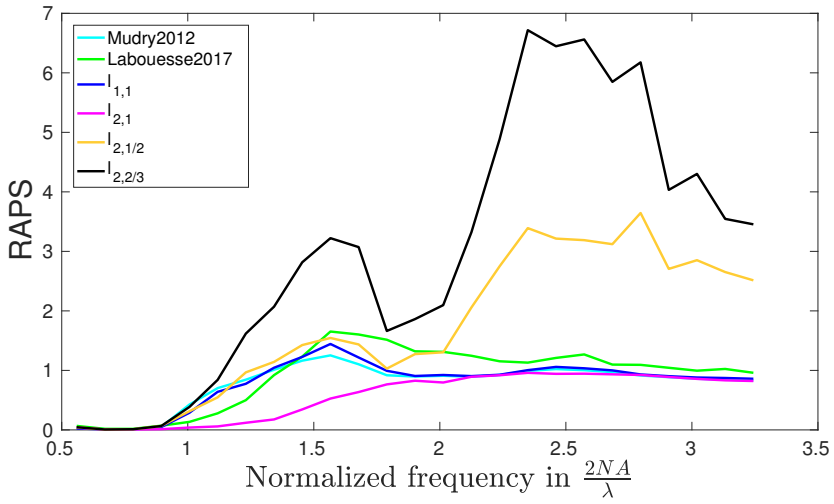
(col 1.) $\ell_{2,1}$ norm (col 2.) $\ell_{2,1/2}$ norm (col 3.) $\ell_{2,2/3}$ norm

Figure 2.3 – **Reconstructed object with 300 speckle patterns and 40dB Gaussian noise by minimizing $\ell_{p,q}$ norm.**

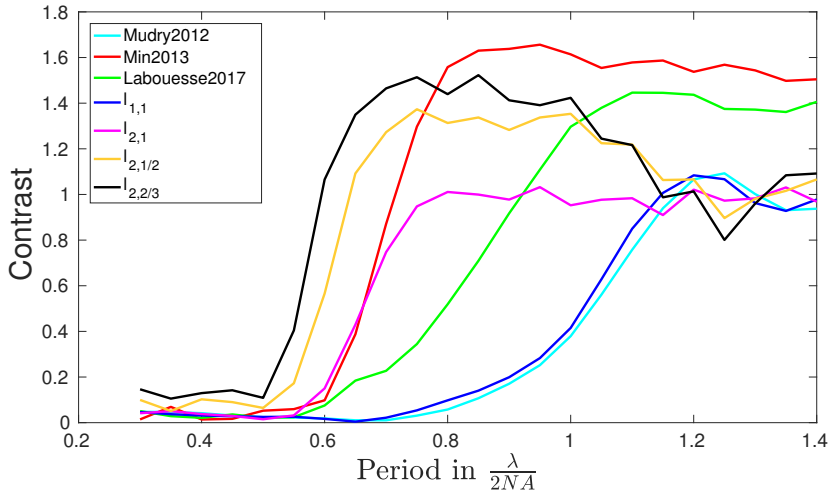
To measure the quality of the reconstructed images, we use the normalized radially averaged power spectrum (RAPS) of the error image and the modulation contrast function of the reconstructed object.

The normalized RAPS function is defined as:

$$f(r) = \frac{\int_{-\pi}^{\pi} \left|\widetilde{\widehat{\rho}}(\boldsymbol{u}) - \widetilde{\rho^*}(\boldsymbol{u})\right|^2 d\theta}{\int_{-\pi}^{\pi} |\widetilde{\rho^*}(\boldsymbol{u})|^2 d\theta}, \qquad \text{with} \quad \boldsymbol{u} = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}, \; r > 0, \; \theta \in (0, 2\pi) \tag{2.63}$$



(a) Normalized RAPS curves



(b) MCF curves

Figure 2.4 – **The normalized RAPS and MCF curves of the reconstructed object by joint estimators.**

The period of the pattern taken on a circle of radius R is $L(R) = 2\pi R/40$ and the modulation contrast function (MCF) is defined as:

$$C(R) = 2\tilde{f}_R(1/L(R))/\tilde{f}_R(0) \tag{2.64}$$

where $\tilde{f}_R$ is the 1D Fourier transform of $f_R(s) \propto 1+\cos(2\pi s/L(R))$ with s the arclength along the circle. We can check that for the true object $\boldsymbol{\rho}^*$, $C(R) = 1$ for all radius R.

The normalized RAPS of errors and the MCF of the reconstructed object by constrained BPDN model are shown in Figure 2.4. In the MCF curves, we see that the constrained $\ell_{2,1/2}$ or $\ell_{2,2/3}$ regularizer give better contrast in high frequency part, however, their error energy in high frequency part is also stronger, probably caused by the binary effect in the reconstructed images. The reconstructed object by [MJK+13] is not well scaled, so we do not plot its RAPS curve. Generally, we can conclude that the constrained $\ell_{2,1}$ regularizer gives a balanced result with respect to the constrast and error energy.

2.4.1 Influence of the hyperparameter in constrained BPDN model

In this section, we explore how the parameter ϵ influence the estimator when it is not correctly set in constrained BPDN form. The reconstruction results with 300 speckle patterns and 40dB white noise using different ϵ value are shown in Fig. 2.5. When ϵ is much lower than its true value as shown in the second line of Figure 2.5 ($\epsilon = 0.2\epsilon_{\text{real}}$), we still retrieve good super-resolution, nevertheless, some artifacts (random dots) appears in low-resolution part when using $\ell_{2,1/2}$ and $\ell_{2,2/3}$ norm. When $\epsilon = 5\epsilon_{\text{real}}$ as shown in first line of Figure 2.5, we lost partial super-resolution compared to the case when it is correctly set, however, it is still better than the results obtained with only positivity constraint (Figure 2.2(c)).

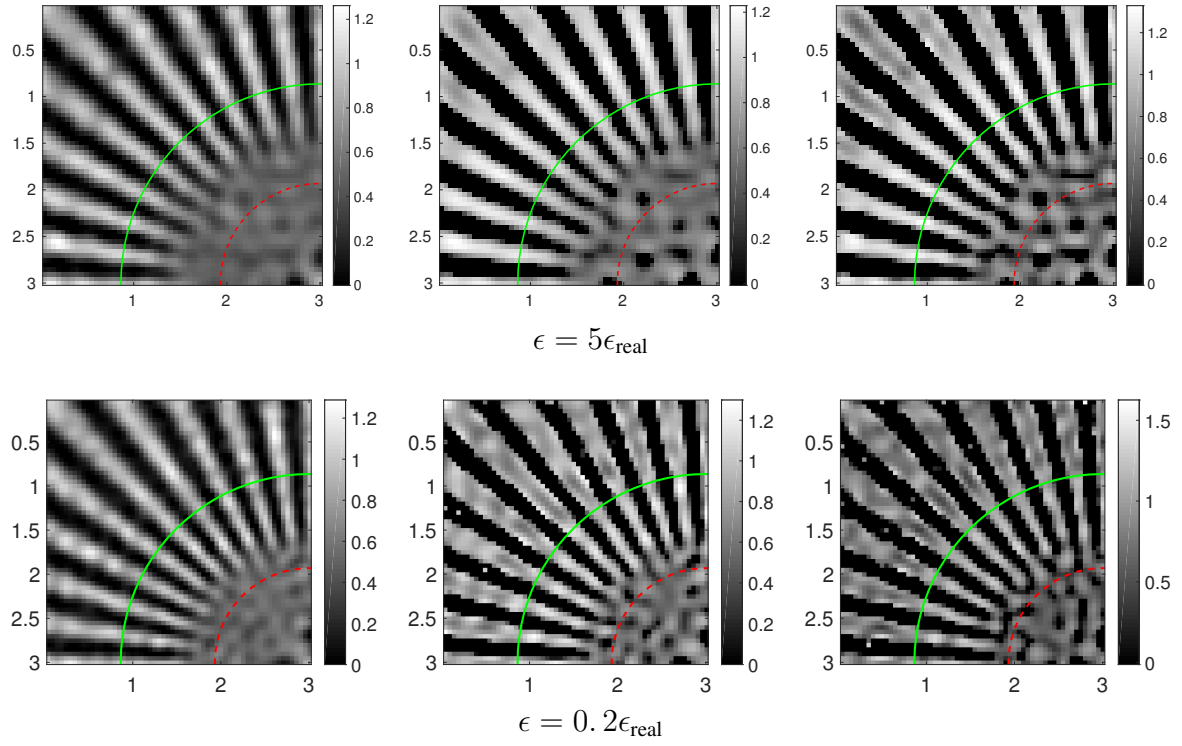
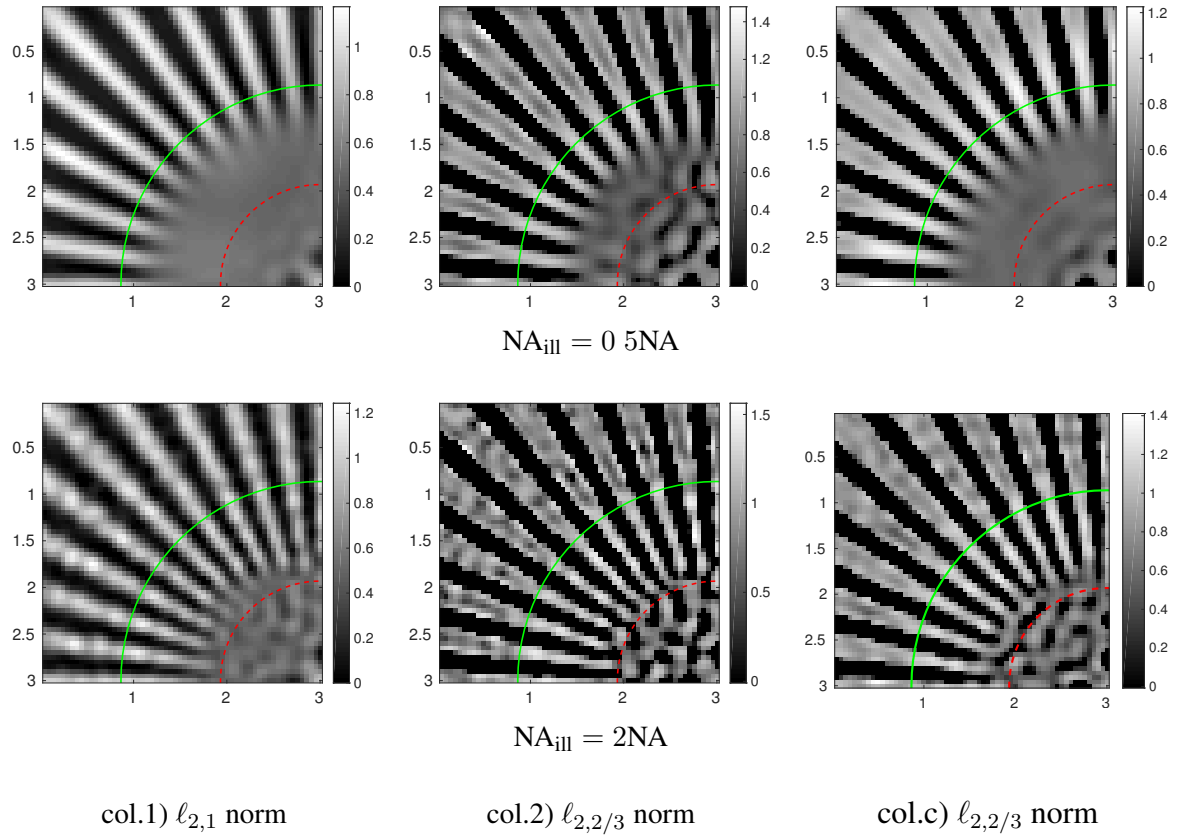


Figure 2.5 – **Reconstructed object with 300 speckle patterns and 40dB Gaussian noise by minimizing $\ell_{p,q}$ norm of constrained form.**

2.4.2 Resolution under different frequency support of speckle patterns

Surpassing the resolution limit more than a factor of two is possible in saturated structured illumination microscopy (nonlinear SIM) [Gus05]. In this section the influence by Fourier support of speckle on the super-resolution imaging is explored. The reconstruction results of the proposed method with speckle generated with different numerical aperture (NA_{ill}) under 300 illumination and 40dB SNR are shown in Fig. 2.6. When the support of power spectral density of the speckle patterns becomes smaller, we lost partial super-resolution, as shown in the first column in Fig. 2.6. The super-resolution information beyond a factor of two is still inaccessible under the situation the support of speckle spectral density is enlarged in the constrained BPDN model.



$\mathrm{NA_{ill}} = 0\ 5\mathrm{NA}$

$\mathrm{NA_{ill}} = 2\mathrm{NA}$

col.1) $\ell_{2,1}$ norm col.2) $\ell_{2,2/3}$ norm col.c) $\ell_{2,2/3}$ norm

Figure 2.6 – **Reconstructed object with 300 speckle patterns and 40dB Gaussian noise with different frequency support of speckle patterns.**

2.4.3 Resolution under Poisson noise

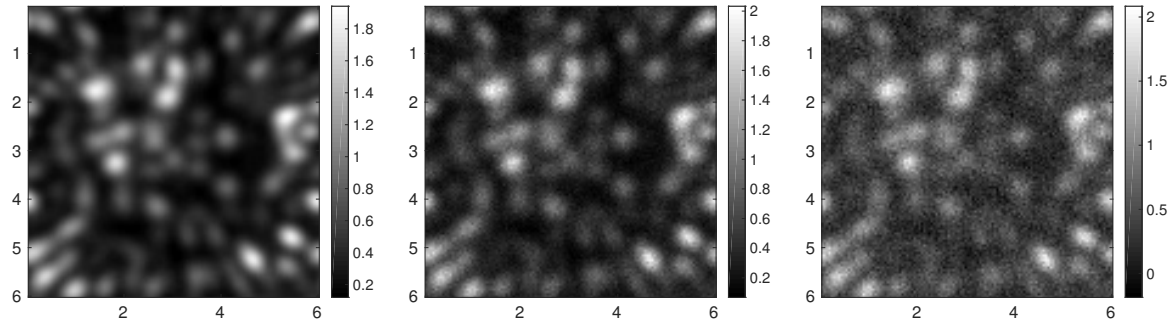
In the previous image formation model in Eq. (1.26), we does not consider the shot noise of CCD caused by the random arrival of photons. For a given photon, the probability of its arrival within a given time period is governed by Poisson distribution. In Figure 2.7, we show the image formation model where a noise-free raw image is corrupted with the mixture of Poisson and Gaussian noise. The number of photons per pixel on average is assumed to be 100 and the SNR of Gaussian noise added is 15dB.

The reconstruction results by minimizing Eq. (2.60) with mixture of Poisson and Gaussian noise using 300 speckle patterns are presented in Figure 2.8. The results shown in the first line are obtained by considering $\ell_{p,q}$ norm regularizer and the TV regularizer, with the hyperparameters $\beta_2 = 1, \beta_3 = 0.1$, while the results shown in the second line are obtained by only considering $\ell_{p,q}$ norm regularizer with $\beta_2 = 1, \beta_3 = 0$. We see that the super-resolution can be retrieved with only $\ell_{p,q}$ norm regularizer under mixture of Poisson and Gaussian noise, and after introducing TV norm regularizer, the results become smoother as expected.

2.4.4 The comparison of two estimators

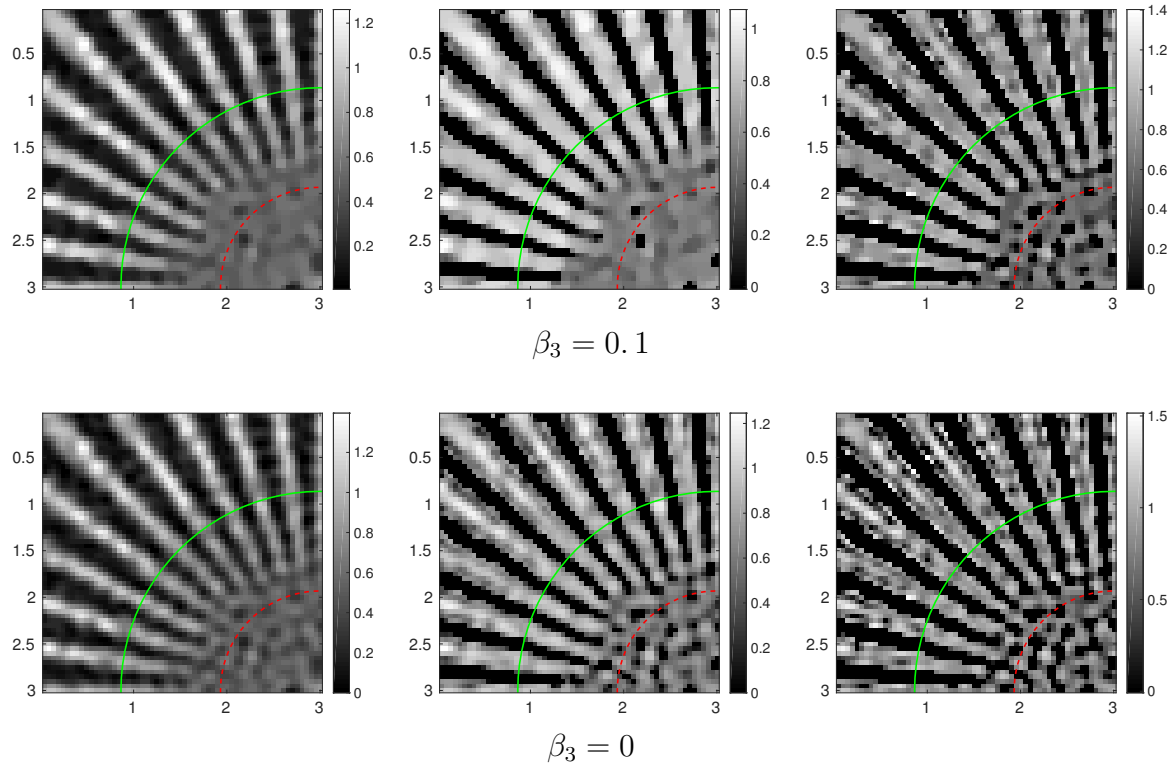
As has been indicated in Section 1.5, once the matrix $\mathbf{Q}$ is obtained, we could estimate the object $\boldsymbol{\rho}$ by either the mean:

$$\hat{\boldsymbol{\rho}} = \frac{1}{I_0}\bar{\boldsymbol{q}}_m = \frac{1}{MI_0}\sum_m \boldsymbol{q}_m \tag{2.65}$$



(a) (b) (c)

Figure 2.7 – **Illustration of Gaussion and Poisson noise.** (a) A noise-free raw image. (b) The noise-free image corrupted with Poisson noise. (c) The noise-free image corrupted with the mixture of Poisson and Gaussian noise.



(col 1.) $\ell_{2,1}$ norm (col 2.) $\ell_{2,1/2}$ norm (col 3.) $\ell_{2,2/3}$ norm

Figure 2.8 – **Reconstructed object with 300 speckle patterns under mixture of Poisson and Gaussian noise by minimizing** (2.60)**.** The SNR of Gaussian noise is 15dB and the number of photons per pixel per measurement is set to 100.

or by its covariance:

$$\hat{\boldsymbol{\rho}} = \sqrt{\frac{1}{M}\sum_{m}\left(\boldsymbol{q}_m - \bar{\boldsymbol{q}}_m\right)^2} \tag{2.66}$$

The measured fluorescence images are always blurred by the non-uniform background comprising of detector offset, scattered light and the out-of-focus parts in sample. If we assume the background keep fixed in the imaging process, then the image formation model could be described by:

$$\boldsymbol{y}_m = \mathbf{H}\boldsymbol{q}_m + \boldsymbol{\epsilon}_m + \boldsymbol{b} \tag{2.67}$$

with $\boldsymbol{b} \in \mathbb{R}^N$ denoting the background. So what we reconstructed in the $m-$th column of $\mathbf{Q}$ is in fact $\hat{\boldsymbol{q}}_m = \boldsymbol{q}_m + \mathbf{H}^+(\boldsymbol{\epsilon}_m + \boldsymbol{b})$ instead of $\boldsymbol{q}_m$, with $\mathbf{H}^+$ the pesudo inverse of $\mathbf{H}$. Then the expectation of $\hat{\boldsymbol{q}}_m$ will be blurred by $\mathbf{H}^+\boldsymbol{b}$. However, the covariance of $\hat{\boldsymbol{q}}_m$ is not influenced by the presentation of the background. To verify this point, simulation results using 300 speckle patterns and 40dB Gaussian noise with a fixed background (lena) are shown in figure 2.9. The second line and the third line are obtained by minimizing the constrained ℓ_{21} and ℓ_{11} regularizer, respectively. As we expected, the mean of $\mathbf{Q}$ are blured by the background while the square root of covariance of $\mathbf{Q}$ are rather clear. So in the reconstruction problem by the experimental data as shown in next section, we choose Eq. (2.66) to estimate the object.

2.5 Experiments with real data

The raw images are obtained with an objective of 1.49 NA and $100\times$ magnification. The wavelength of excitation is 488 nm and 520 nm for the collection. The PSF used is simulated using a ICY plug-in called PSF Generator with Gibson & Lanni 3D Optical Model. The reconstructed objects by minimizing the constrained $\ell_{p,q}$ regularizer with 80 speckle patterns of the Podosome and Actin sample are shown in Figure 2.10. Fourier transforms of the Podosome images are displayed in Figure 2.11. The zoom of a small part of these images are shown in Figure 2.12. Line sections extracted from the Podosome sample are plotted in Fig. 2.13. Clearly we see better details using our proposed method compared with the Wiener deconvolution of the average raw images in Podosome sample. The improvement in Actin sample is not easy to visualize.

2.6 Conclusion

In this chapter I have presented a super-resolution image reconstruction approach based on the prior information on the object. Specifically, we use $\ell_{p,q}$ norm to enforce the joint sparsity of matrix $\mathbf{Q}$, which is equivalent with the sparsity of object $\boldsymbol{\rho}$. Both the constraint form and the unconstraint form of the BNDN model are presented. The hyperparameter in constraint form is easier to tune than the unconstraint form since it has a clear meaning, which is proportional to the standard deviation of the noise. The ADMM algorithm and the primal-dual algorithm are chosen to solve the corresponding BPDN optimization problems. Other prior information of the object can be incorporated into our method easily without big changes of the optimization algorithm, such as TV regularizer or the positivity constraint. Simulation results show that our proposed methods could retrieve good super-resolution information, even under the mixture of Poisson and Gaussian noise.

Since we reconstruct the unknown speckle patterns and the object together, the proposed method belongs to a joint reconstruction family. In the following chapters, we will present a marginal reconstruction approach to speckle SIM, where *marginal* means that the reconstruction

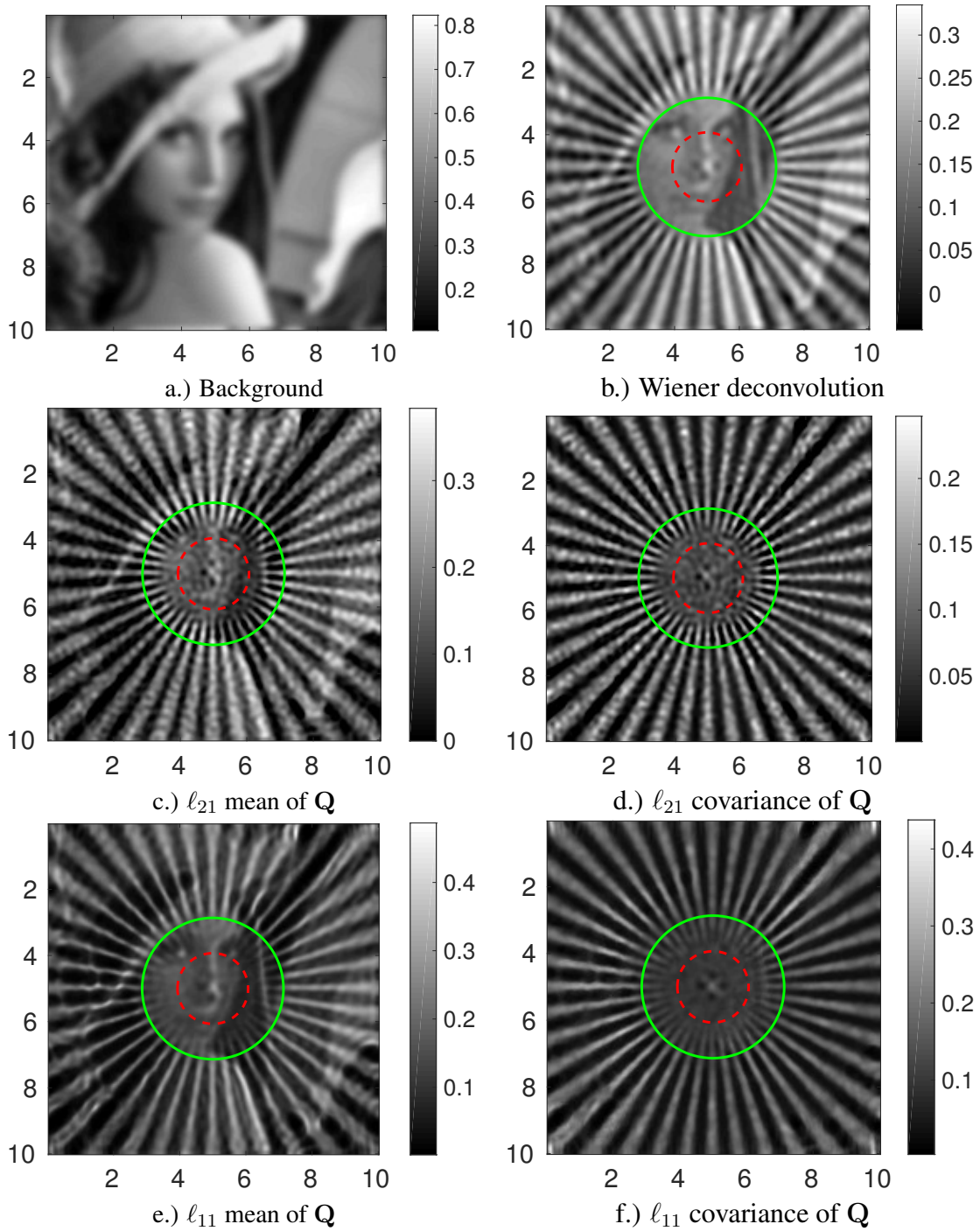


Figure 2.9 – **Reconstruction results by 300 speckle patterns and 40dB Gaussian noise with a fixed background.**

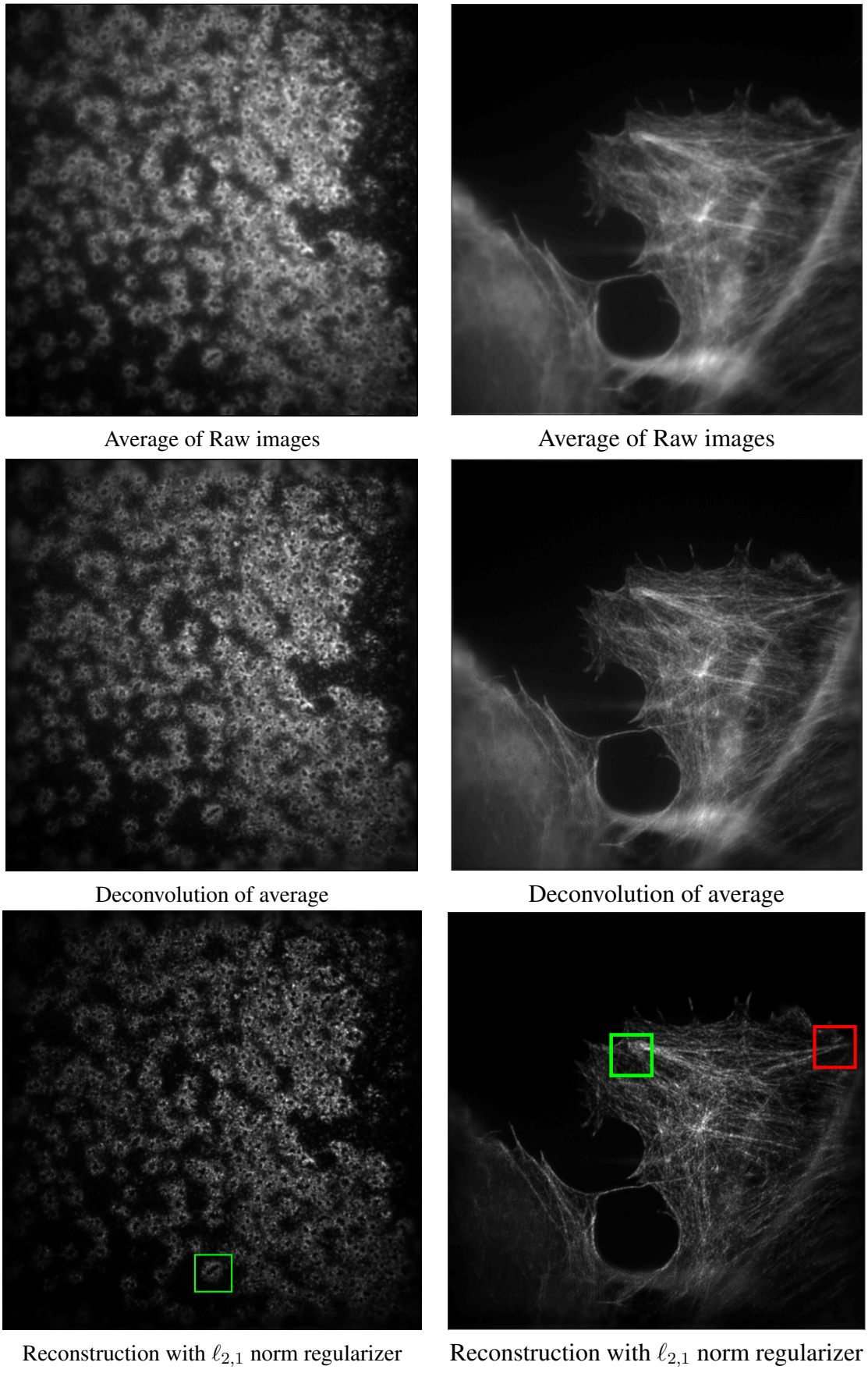
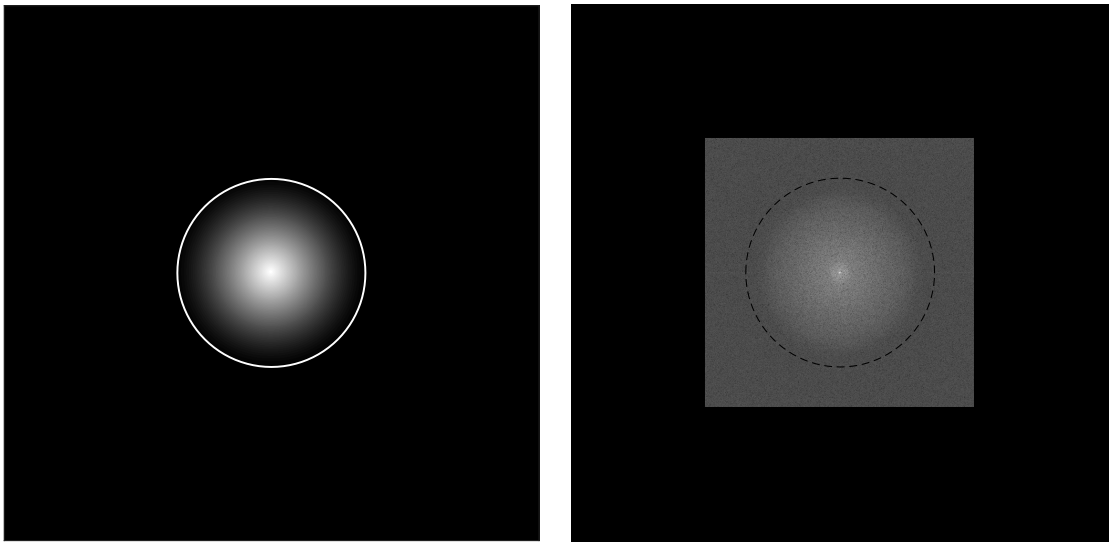
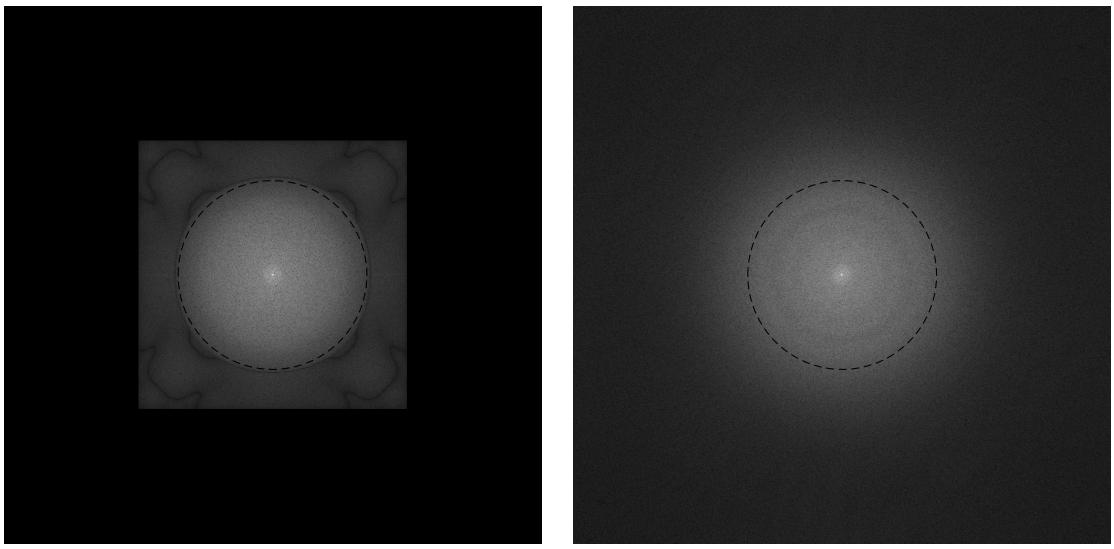


Figure 2.10 – **Reconstructed object with 80 speckle patterns of Podosome sample (first column) and Actin (second column) sample.**



OTF of system

Average of Raw images



Deconvolution of average

Reconstruction with $\ell_{2,1}$ norm regularizer

Figure 2.11 – **Fourier domain plot of Podosome images with 80 speckle patterns.**

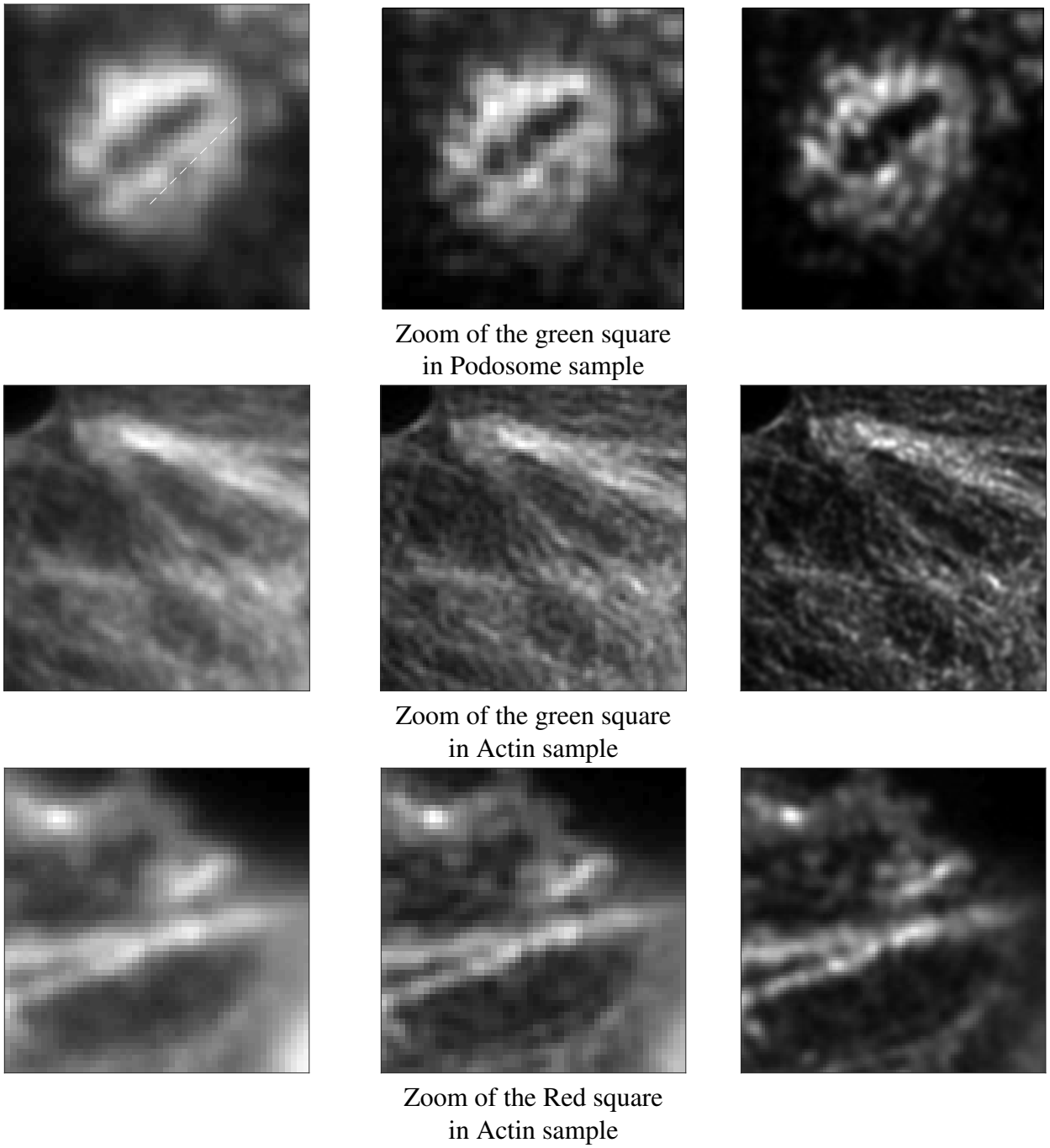


Figure 2.12 – **Partial enlargement of the reconstructed object with 80 speckle patterns of Podosome and Actin sample.**

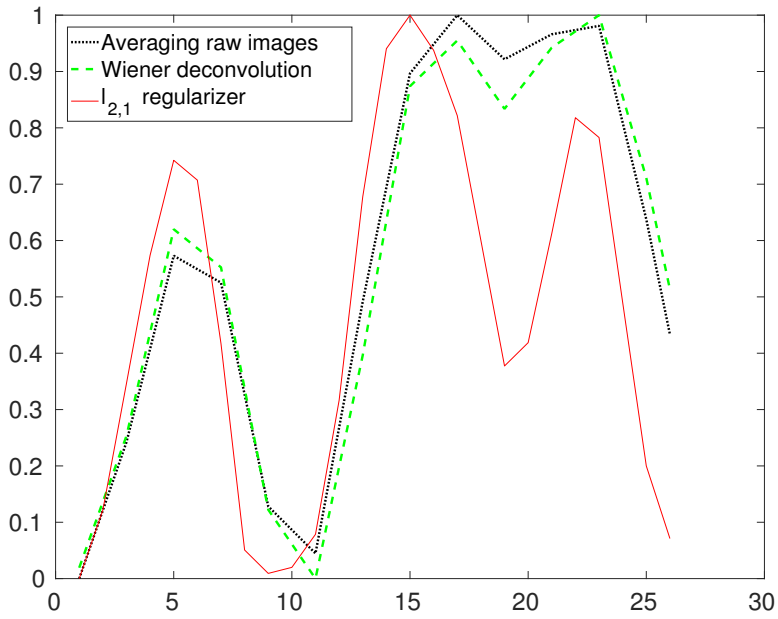


Figure 2.13 – **Line section plot extracted from the reconstructions of Podosome sample shown in Fig. 2.12.**

of speckle patterns is not necessary. Contrary to the methods presented in this chapter, only minimal prior assumptions on the object are needed to produce a super-resolution factor equal to two. What is more, the parameter of interest does not grow with the increasing number of measurements in the marginal approach, and we will demonstrate that it is asymptotically consistent.

Chapter 3

Super-resolution capacity using second-order statistics

3.1 Introduction

In Section 1.5 and Chapter 2, I have presented some reconstruction methods in *blind-speckle-SIM*. As it has been stated, there are no theoretical guarantees about the super-resolution (SR) capacity in these methods except S-SOFI. In particular, since the number of unknowns (which include both the unknown object and the speckle patterns) increases with the number of observations, these methods are not generally reliable from the statistical viewpoint [LR83].

In this chapter we study the SR capacity in blind-speckle-SIM, taking advantage of second-order statistics. Let us denote the radii of the support of OTF and the speckle power spectral density by ν_{PSF} and ν_{spec}. When $\nu_{\mathrm{spec}} \leq \nu_{\mathrm{PSF}}$, our theoretical analysis demonstrates that the SR capacity obtained by blind-speckle-SIM can be $2\nu_{\mathrm{spec}}/\nu_{\mathrm{PSF}}$ (asymptotically). When $\nu_{\mathrm{spec}} = \infty$, a SR factor of two could be obtained for the square of the object ρ. No general conclusion could we make in the case $\nu_{\mathrm{PSF}} \leq \nu_{\mathrm{spec}} < \infty$. A marginal estimation approach based on such second-order statistics is proposed in this chapter. Contrary to the reconstruction methods presented previously, the marginal estimator is shown to be unbiased and consistent.

Section 3.2 gives the expression of the first two moments of the observation data. We explore the SR capacity under assumptions of uncorrelated and correlated speckle cases in Section 3.3. A marginal estimation approach is proposed in Section 3.4, together with its asymptotic statistical analysis, and a possible numerical implementation is discussed. The influence of using Poisson statistics is considered in Section 3.5. Finally, simulation results are presented in Section 3.7. It is deserved to be mentioned here that most of the contents in this chapter has been published in [ILA+17] where I am a co-author.

3.2 First and second-order statistics of the data

i As shown in Eq. (1.26), the observation model in the speckle-SIM framework is given by:

$$y_m(\boldsymbol{r}) = k_m(\boldsymbol{r}) + \epsilon_m(\boldsymbol{r}) \tag{3.1a}$$

$$\text{with} \qquad k_m(\boldsymbol{r}) = \int h(\boldsymbol{r}-\boldsymbol{r}')\rho(\boldsymbol{r}')I_m(\boldsymbol{r}')d\boldsymbol{r}' \tag{3.1b}$$

in which I_m is the m-th realization of speckle and ϵ_m stands for electronic noise and model errors. In the analysis below, the following assumptions are adopted:

1. The PSF h is both integrable and square-integrable, and it is positive semidefinite (*i.e.,* its Fourier transform is real and nonnegative). Moreover, its Fourier transform $\tilde{h}$ takes finite values and vanishes outside a bounded set that we note $\mathcal{D}_{\text{PSF}} = \{\boldsymbol{u} \mid \tilde{h}(\boldsymbol{u}) > 0\}$.
2. The data grid is fine enough to sample the PSF without discretization error.
3. The object ρ is integrable and takes finite, nonnegative values over $\mathbb{R}^d$.
4. The noise ϵ_m and speckle patterns I_m are second-order stationary and mutually uncorrelated random processes. Without loss of generality, we further assume that the noise is centered, that is, $E[\epsilon_m(\boldsymbol{r})] = 0$.
5. The noise and illuminations are second-order stationary, mutually decorrelated random processes.
6. The first two moments of speckle patterns are assumed to be known with homogeneous mean I_0 over the sample and covariance $\gamma_{\text{spec}}(\boldsymbol{r})$. This is true for fully developed speckle patterns according to our analysis in appendix A.

Our analysis is confined to the second-order statistics, *i.e.,* only the mean and covariance of the data are concerned. To characterize the SR capacity of second-order methods, we assign each component of $\tilde{\rho}(\boldsymbol{u})$, the Fourier transform of the object $\rho(\boldsymbol{r})$, to one of the three classes:

— *Non-identifiable* components are those for which the second-order data statistics bring no information.
— *Partially identifiable* components are those for which the second-order data statistics bring some information, but for which some ambiguity remains.
— *Identifiable* spectral components are those which are uniquely determined given the second-order data statistics.

3.2.1 First-order statistics of data

As can be seen from Eq. (3.1), we can obtain the statistical mean of $\boldsymbol{y}_m$ by:

$$\mu_y(\boldsymbol{r}) = I_0 \int h(\boldsymbol{r} - \boldsymbol{r}')\rho(\boldsymbol{r}')d\boldsymbol{r}'. \tag{3.2}$$

According to the *convolution theorem*, the Fourier transform of μ_y gives:

$$\tilde{\mu}_y(\boldsymbol{u}) = I_0\tilde{h}(\boldsymbol{u})\tilde{\rho}(\boldsymbol{u}). \tag{3.3}$$

Therefore, we know that any spectral component of ρ inside the support $\mathcal{D}_{\text{PSF}}$ is identifiable from μ_y provided that $I_0 \neq 0$. So, we have the following property:

Property 3.1. **The frequency component $\tilde{\rho}(\boldsymbol{u})$ is identifiable from $\tilde{\mu}_y$ if and only if $\boldsymbol{u} \in \mathcal{D}_1$ with**

$$\mathcal{D}_1 = \begin{cases} \mathcal{D}_{\text{PSF}} & \textbf{if } I_0 \neq 0 \\ \emptyset & \textbf{otherwise} \end{cases} \tag{3.4}$$

In any case, the first-order statistics of the data bring no information of spectral components outside $\mathcal{D}_{\text{PSF}}$.

3.2.2 Second-order statistics of data

Since the noise is centered and both the speckle patterns and noise are second-order stationary random processes, we write their correlation functions as:

$$\gamma_{\text{spec}}(\boldsymbol{r}) = E[I(\boldsymbol{x})I(\boldsymbol{x} - \boldsymbol{r})] - I_0^2, \tag{3.5a}$$

$$\gamma_{\epsilon}(\boldsymbol{r}) = E[\epsilon(\boldsymbol{x})\epsilon(\boldsymbol{x} - \boldsymbol{r})]. \tag{3.5b}$$

Then the correlation function of the measured image y_m is given by:

$$\gamma_y(\boldsymbol{r},\boldsymbol{r}') = E[y_m(\boldsymbol{r})y_m(\boldsymbol{r}')] - \mu_y(\boldsymbol{r})\mu_y(\boldsymbol{r}'). \tag{3.6}$$

Since we assume that the speckle patterns and noise are mutually uncorrelated, we have:

$$\gamma_y(\boldsymbol{r},\boldsymbol{r}') = \gamma_k(\boldsymbol{r},\boldsymbol{r}') + \gamma_\epsilon(\boldsymbol{r}-\boldsymbol{r}'), \tag{3.7}$$

where

$$\gamma_k(\boldsymbol{r},\boldsymbol{r}') = \iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')h(\boldsymbol{r}-\boldsymbol{x})h(\boldsymbol{r}'-\boldsymbol{x}')\gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')d\boldsymbol{x}d\boldsymbol{x}'. \tag{3.8}$$

A double Fourier transform of (3.8) with respect to space variables $\boldsymbol{r}$ and $\boldsymbol{r}'$ gives:

$$\tilde{\gamma}_k(\boldsymbol{u},\boldsymbol{u}') = \iiiint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')h(\boldsymbol{r}-\boldsymbol{x})h(\boldsymbol{r}'-\boldsymbol{x}')e^{-2\pi i\boldsymbol{r}\cdot\boldsymbol{u}}e^{-2\pi i\boldsymbol{r}'\cdot\boldsymbol{u}'}\gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')d\boldsymbol{x}d\boldsymbol{x}'d\boldsymbol{r}d\boldsymbol{r}' \tag{3.9}$$

According to Fubini's theorem [Fre00], we have

$$\tilde{\gamma}_k(\boldsymbol{u},\boldsymbol{u}') = \tag{3.10}$$

$$\iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')\gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')\Big(\int h(\boldsymbol{r}-\boldsymbol{x})e^{-2\pi i\boldsymbol{r}\cdot\boldsymbol{u}}d\boldsymbol{r}\int h(\boldsymbol{r}'-\boldsymbol{x}')e^{-2\pi i\boldsymbol{r}'\cdot\boldsymbol{u}'}d\boldsymbol{r}'\Big)d\boldsymbol{x}d\boldsymbol{x}' \tag{3.11}$$

Substitute $\boldsymbol{z}=\boldsymbol{r}-\boldsymbol{x}, \boldsymbol{z}'=\boldsymbol{r}'-\boldsymbol{x}'$, then $d\boldsymbol{z}=d\boldsymbol{r}, d\boldsymbol{z}'=d\boldsymbol{r}'$, so:

$$\begin{aligned}
&\tilde{\gamma}_k(\boldsymbol{u},\boldsymbol{u}') = \\
&\iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')\gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')\Big(\int h(\boldsymbol{z})e^{-2\pi i(\boldsymbol{z}+\boldsymbol{x})\cdot\boldsymbol{u}}d\boldsymbol{z}\int h(\boldsymbol{z}')e^{-2\pi i(\boldsymbol{z}'+\boldsymbol{x}')\cdot\boldsymbol{u}'}d\boldsymbol{z}'\Big)d\boldsymbol{x}d\boldsymbol{x}' \\
&= \iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')\gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')e^{-2\pi i\boldsymbol{x}\cdot\boldsymbol{u}}e^{-2\pi i\boldsymbol{x}'\cdot\boldsymbol{u}'}\Big(\int h(\boldsymbol{z})e^{-2\pi i\boldsymbol{z}\cdot\boldsymbol{u}}d\boldsymbol{z}\int h(\boldsymbol{z}')e^{-2\pi i\boldsymbol{z}'\cdot\boldsymbol{u}'}d\boldsymbol{z}'\Big)d\boldsymbol{x}d\boldsymbol{x}' \\
&= \tilde{h}(\boldsymbol{u})\tilde{h}(\boldsymbol{u}')\iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')\gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')e^{-2\pi i\boldsymbol{x}\cdot\boldsymbol{u}}e^{-2\pi i\boldsymbol{x}'\cdot\boldsymbol{u}'}d\boldsymbol{x}d\boldsymbol{x}'
\end{aligned} \tag{3.12}$$

Let us define function $\tilde{g}(\boldsymbol{u},\boldsymbol{u}')$ as:

$$\begin{aligned}
\tilde{g}(\boldsymbol{u},\boldsymbol{u}') &= \iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')\gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')e^{-2\pi i\boldsymbol{x}\cdot\boldsymbol{u}}e^{-2\pi i\boldsymbol{x}'\cdot\boldsymbol{u}'}d\boldsymbol{x}d\boldsymbol{x}' \\
&= \iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')e^{-2\pi i\boldsymbol{x}\cdot\boldsymbol{u}}e^{-2\pi i\boldsymbol{x}'\cdot\boldsymbol{u}'}\Big(\int \tilde{\gamma}_{\text{spec}}(\boldsymbol{u}'')e^{2\pi i\boldsymbol{u}''\cdot(\boldsymbol{x}-\boldsymbol{x}')}d\boldsymbol{u}''\Big)d\boldsymbol{x}d\boldsymbol{x}' \\
&= \int\Big(\int \rho(\boldsymbol{x}')e^{-2\pi i\boldsymbol{x}'\cdot(\boldsymbol{u}'+\boldsymbol{u}'')}\Big(\int \rho(\boldsymbol{x})e^{-2\pi i\boldsymbol{x}\cdot(\boldsymbol{u}-\boldsymbol{u}'')}d\boldsymbol{x}\Big)d\boldsymbol{x}'\Big)\tilde{\gamma}_{\text{spec}}(\boldsymbol{u}'')d\boldsymbol{u}'' \\
&= \int \tilde{\rho}(\boldsymbol{u}-\boldsymbol{u}'')\tilde{\rho}(\boldsymbol{u}+\boldsymbol{u}'')\tilde{\gamma}_{\text{spec}}(\boldsymbol{u}'')d\boldsymbol{u}''
\end{aligned} \tag{3.13}$$

where $\tilde{\gamma}_{\text{spec}}$ is called the power spectral density function of the speckle pattern, which is identified by the Fourier transform of the correlation function of γ_{spec} according to the *Wiener-Khinchin theorem* [Goo15, Section 3.4]. The SR capacity analysis will be studied from the above equations, which link in the Fourier domain "observable" quantities $\tilde{\gamma}_y$ to the available information of the object ρ.

3.3 SR capacity using second-order statistics

The noise covariance conveys no information about the object ρ, so the knowledge of γ_y is equivalent to that of γ_k. According to Eqs. (3.12)–(3.13), the spectral density $\tilde{\gamma}_{\text{spec}}$ plays a central role in the identification of the spectral components of the sample. However, there is no analytical expression to estimate the object from this set of Eqs. (3.12)–(3.13) in particular because of the quadratic functional dependence of the data correlation on the unknown sample ρ. Two special cases lend themselves to a deeper analysis, though. The first one corresponds to "sufficiently" correlated speckles, by saying which we mean the support of the spectral power density of speckle is contained in the frequency support of the PSF. The second one corresponds to uncorrelated speckle patterns. For brevity of notations, we introduce the Minkowski difference of two sets:

$$A \ominus B = \{x - y \mid x \in A, y \in B\} \tag{3.14}$$

3.3.1 "Sufficiently" correlated speckle case

First we assume that the frequency support of the speckle correlation is confined to some frequency domain that we note $\mathcal{D}_{\text{spec}}$. According to expression (3.12), $\tilde{\gamma}_k(\boldsymbol{u}, \boldsymbol{u}')$ vanishes when either $\boldsymbol{u}$ or $\boldsymbol{u}'$ is outside $\mathcal{D}_{\text{PSF}}$. On the other hand, according to (3.13), $\tilde{g}(\boldsymbol{u}, \boldsymbol{u}')$ conveys no information on the frequency components $\tilde{\rho}(\boldsymbol{v})$ such that $\boldsymbol{u} - \boldsymbol{v} \notin \mathcal{D}_{\text{spec}}$ or $\boldsymbol{v} - \boldsymbol{u}' \notin \mathcal{D}_{\text{spec}}$. Then, we conclude that the following property holds.

Property 3.2. **Any spectral component $\tilde{\rho}(\boldsymbol{u})$ where $\boldsymbol{u} \notin \mathcal{D}_{\text{PSF}} \ominus \mathcal{D}_{\text{spec}}$ is non-identifiable from the mean $\mu_y(\boldsymbol{r})$ and the correlation function $\gamma_k(\boldsymbol{r}, \boldsymbol{r}')$.**

Property 3.2 defines a boundary that cannot be surpassed from standard statistical information. A further analysis allows us to draw a positive conclusion about what is accessible when the support of the spectral power density of speckle is not larger than that of the PSF.

Property 3.3. **Provided that γ_{spec} is such that $\mathcal{D}_{\text{spec}} \subseteq \mathcal{D}_{\text{PSF}}$, any spectral component $\tilde{\rho}(\boldsymbol{u})$ is identifiable from the mean $\mu_y(\boldsymbol{r})$ and the correlation function $\gamma_k(\boldsymbol{r}, \boldsymbol{r}')$ if $\boldsymbol{u} \in \mathcal{D}_1 \cup \mathcal{D}_2$ with $\mathcal{D}_2 = \mathcal{D}_{\text{spec}} \ominus \mathcal{D}_{\text{spec}}$.**

Proof. Let q denote the impulse response of the filter defined in the Fourier domain by:

$$\tilde{q}(\boldsymbol{u}) = \begin{cases} \left(\tilde{\gamma}_{\text{spec}}(\boldsymbol{u})\right)^{1/2}, & \boldsymbol{u} \in \mathcal{D}_{\text{spec}} \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

Similarly to γ_{spec}, q is positive semi-definite, and it is a Hermitian symmetric function. So we have $\tilde{\gamma}_{\text{spec}} = \tilde{q}^2$, and hence

$$\gamma_{\text{spec}} = q * q. \tag{3.16}$$

Let us define the following kernels:

$$f(\boldsymbol{r}, \boldsymbol{r}') = \int q(\boldsymbol{r} - \boldsymbol{x}) q^*(\boldsymbol{r}' - \boldsymbol{x}) \rho(\boldsymbol{x}) d\boldsymbol{x} \tag{3.17}$$

$$F(\boldsymbol{r}, \boldsymbol{r}') = \int f(\boldsymbol{r}, \boldsymbol{r}'') f^*(\boldsymbol{r}, \boldsymbol{r}'') d\boldsymbol{r}'' \tag{3.18}$$

and the induced integral operators K_f and K_F:

$$\begin{aligned} (K_f \phi)(\cdot) &= \int f(\cdot, \boldsymbol{r}) \phi(\boldsymbol{r}) d\boldsymbol{r}, \\ (K_F \phi)(\cdot) &= \int F(\cdot, \boldsymbol{r}) \phi(\boldsymbol{r}) d\boldsymbol{r}. \end{aligned} \tag{3.19}$$

Now let us review Eq. (3.17). f can be seen as the inner product of $q(\boldsymbol{r}-\boldsymbol{x})$ and $q(\boldsymbol{r}'-\boldsymbol{x})$ with the reference measure induced by ρ. According to the Cauchy-Schwarz inequality, we have:

$$|f(\boldsymbol{r},\boldsymbol{r}')|^2 \leq \int |q(\boldsymbol{r}-\boldsymbol{x})|^2\rho(\boldsymbol{x})d\boldsymbol{x} \int |q(\boldsymbol{r}'-\boldsymbol{x})|^2\rho(\boldsymbol{x})d\boldsymbol{x}. \tag{3.20}$$

As a consequence,

$$\iint |f(\boldsymbol{r},\boldsymbol{r}')|^2 d\boldsymbol{r}d\boldsymbol{r}' \leq \iint |q(\boldsymbol{r}-\boldsymbol{x})|^2\rho(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{r} \iint |q(\boldsymbol{r}'-\boldsymbol{x})|^2\rho(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{r}'. \tag{3.21}$$

Now notice that

$$\begin{aligned}\iint |q(\boldsymbol{r}-\boldsymbol{x})|^2\rho(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{r} &= \int \rho(\boldsymbol{x}) \int |q(\boldsymbol{r}-\boldsymbol{x})|^2 d\boldsymbol{r}d\boldsymbol{x} \\ &= \int |q(\boldsymbol{r})|^2 d\boldsymbol{r} \int \rho(\boldsymbol{x})d\boldsymbol{x}.\end{aligned} \tag{3.22}$$

Combining (3.21) and (3.22), we obtain:

$$\iint |f(\boldsymbol{r},\boldsymbol{r}')|^2 d\boldsymbol{r}d\boldsymbol{r}' \leq \Big(\int |q(\boldsymbol{r})|^2 d\boldsymbol{r} \int \rho(\boldsymbol{x})d\boldsymbol{x}\Big)^2. \tag{3.23}$$

Since ρ is integrable, and

$$\int |q(\boldsymbol{r})|^2 d\boldsymbol{r} = \int \tilde{\gamma}_{\text{spec}}(\boldsymbol{u})d\boldsymbol{u} = \gamma_{\text{spec}}(0) < \infty, \tag{3.24}$$

So, we have

$$\iint |f(\boldsymbol{r},\boldsymbol{r}')|^2 d\boldsymbol{r}d\boldsymbol{r}' < \infty \tag{3.25}$$

i.e., $f(\boldsymbol{r},\boldsymbol{r}') \in L^2(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{C})$, and consequently, K_f is a Hilbert-Schmidt integral operator [Ped12]. Now let us go to the heart of the proof, which is threefold. The first step allows us to show that kernel F is uniquely defined from γ_k. In a second step, we establish that f is uniquely defined from F given (3.18). At this point, we conclude that the knowledge of γ_k implies that of f, which is a linear functional of ρ (whereas the dependency of γ_k in ρ is quadratic). The last step consists in a Fourier analysis of f, in order to determine which spectral components of ρ are identifiable from the knowledge of f.

Step 1) Given (3.17) and (3.16), we have the following alternate expression for (3.18):

$$F(\boldsymbol{r},\boldsymbol{r}') = \iint \rho(\boldsymbol{x})\rho(\boldsymbol{x}')\, q(\boldsymbol{r}-\boldsymbol{x})q^*(\boldsymbol{r}'-\boldsymbol{x}')\, \gamma_{\text{spec}}(\boldsymbol{x}-\boldsymbol{x}')d\boldsymbol{x}d\boldsymbol{x}'. \tag{3.26}$$

Comparing it to (3.8), we can see that $F = \gamma_k$ when $q = h$, *i.e.*, when the correlation function of speckle is $h * h$. More generally, the double Fourier transform of (3.26) gives:

$$\begin{aligned}F(\boldsymbol{u},\boldsymbol{u}') &= \tilde{q}(\boldsymbol{u})\tilde{q}^*(-\boldsymbol{u}')\tilde{g}(\boldsymbol{u},\boldsymbol{u}') \\ &= \begin{cases} \frac{\tilde{q}(\boldsymbol{u})\tilde{q}(-\boldsymbol{u}')}{\tilde{h}(\boldsymbol{u})\tilde{h}^*(-\boldsymbol{u}')}\tilde{\gamma}_k(\boldsymbol{u},\boldsymbol{u}'), & (\boldsymbol{u},\boldsymbol{u}') \in \mathcal{D}_{\text{spec}}, \\ 0, & \text{otherwise} \end{cases}\end{aligned} \tag{3.27}$$

We note here that $\tilde{h}(\boldsymbol{u}) \neq 0$ since we have assumed $\mathcal{D}_{\text{spec}} \subseteq \mathcal{D}_{\text{PSF}}$.

Step 2) Kernel f is obviously symmetric. Moreover, it is positive semi-definite, since for any square integrable function ϕ,

$$\iint f(\boldsymbol{r},\boldsymbol{r}')\phi(\boldsymbol{r})\phi^*(\boldsymbol{r}')d\boldsymbol{r}d\boldsymbol{r}' = \int |q * \phi|^2(\boldsymbol{x})\rho(\boldsymbol{x})d\boldsymbol{x} \geq 0. \tag{3.28}$$

It is easy to check that kernel F is also positive semi-definite. Moreover K_F is bounded, since it is a Hilbert-Schmidt operator. Being bounded and positive semi-definite, K_F admits a unique square root [Ped12]. In other words, K_f is uniquely defined given K_F. Thus, given the kernel F, there exists a unique kernel f that fulfills (3.18).

Step 3) The Fourier transform of (3.17) gives

$$\tilde{f}(\boldsymbol{u}, \boldsymbol{u}') = \tilde{q}(\boldsymbol{u})\tilde{q^*}(-\boldsymbol{u}')\tilde{\rho}(\boldsymbol{u} + \boldsymbol{u}'). \tag{3.29}$$

The latter identity shows that $\tilde{\rho}(\boldsymbol{u}' + \boldsymbol{u}'')$ is identifiable for all couples $(\boldsymbol{u}', \boldsymbol{u}'')$ such that $\boldsymbol{u}'$ and $-\boldsymbol{u}''$ belong to $\mathcal{D}_{\text{PSF}}$. We thus conclude that the frequency components $\tilde{\rho}(\boldsymbol{u})$ are identifiable from kernel f, and thus from the data correlation γ_k, for all $\boldsymbol{u} \in \mathcal{D}_{\text{spec}} \ominus \mathcal{D}_{\text{spec}}$.

□

In 2D blind-speckle-SIM, $\mathcal{D}_{\text{PSF}}$ and $\mathcal{D}_{\text{spec}}$ are centered disks of radii ν_{PSF} and ν_{spec}, respectively. Under these conditions, $\mathcal{D}_{\text{spec}} \ominus \mathcal{D}_{\text{spec}}$ is a disk with radius $2\nu_{\text{spec}}$. As a consequence, when $\frac{1}{2}\nu_{\text{PSF}} \leq \nu_{\text{spec}} \leq \nu_{\text{PSF}}$, a SR factor of $2\nu_{\text{spec}}/\nu_{\text{PSF}}$ is obtained. Considering the case $\mathcal{D}_{\text{spec}} = \mathcal{D}_{\text{PSF}}$, which occurs when the speckle illuminations and observations are performed *via* the same optical device, the second-order data statistics are sufficient to identify the frequency components corresponding to a SR factor of two. So we retrieve the same SR factor as with classical SIM [Gus00] with perfectly controlled illuminations in fluorescence microscopy.

3.3.2 Uncorrelated speckle case

In the uncorrelated speckle case, the correlation function of speckle patterns can be written as: $\gamma_{\text{spec}}(\boldsymbol{r}) = \gamma_{\text{spec}}(\boldsymbol{0})\delta(\boldsymbol{r})$, where $\delta(\boldsymbol{r})$ is the Dirac delta function. Then the correlation function of the data becomes:

$$\gamma_k(\boldsymbol{r}, \boldsymbol{r}') = \gamma_{\text{spec}}(\boldsymbol{0}) \int \rho^2(\boldsymbol{x})h(\boldsymbol{r} - \boldsymbol{x})h(\boldsymbol{r}' - \boldsymbol{x})d\boldsymbol{x}. \tag{3.30}$$

The Fourier transform of (3.30) gives:

$$\tilde{\gamma}_k(\boldsymbol{u}, \boldsymbol{u}') = \gamma_{\text{spec}}(\boldsymbol{0}) \iint \int \rho^2(\boldsymbol{x})h(\boldsymbol{r} - \boldsymbol{x})h(\boldsymbol{r}' - \boldsymbol{x})e^{-2\pi i\boldsymbol{r}\cdot\boldsymbol{u}}e^{-2\pi i\boldsymbol{r}'\cdot\boldsymbol{u}'}d\boldsymbol{x}d\boldsymbol{r}d\boldsymbol{r}' \tag{3.31}$$

Similarly as in (3.12), we have

$$\begin{aligned}\tilde{\gamma}_k(\boldsymbol{u}, \boldsymbol{u}') &= \gamma_{\text{spec}}(\boldsymbol{0}) \int \rho^2(\boldsymbol{x})e^{-2\pi i\boldsymbol{x}\cdot(\boldsymbol{u}+\boldsymbol{u}')}\Big(\iint h(\boldsymbol{z})h(\boldsymbol{z}')e^{-2\pi i\boldsymbol{z}\cdot\boldsymbol{u}}e^{-2\pi i\boldsymbol{z}'\cdot\boldsymbol{u}'}d\boldsymbol{z}d\boldsymbol{z}'\Big)d\boldsymbol{x} \\ &= \gamma_{\text{spec}}(\boldsymbol{0})\tilde{h}(\boldsymbol{u})\tilde{h}(\boldsymbol{u}') \int \rho^2(\boldsymbol{x})e^{-2\pi i\boldsymbol{x}\cdot(\boldsymbol{u}+\boldsymbol{u}')}d\boldsymbol{x} \\ &= \gamma_{\text{spec}}(\boldsymbol{0})\tilde{h}(\boldsymbol{u})\tilde{h}(\boldsymbol{u}')\widetilde{\rho^2}(\boldsymbol{u} + \boldsymbol{u}')\end{aligned} \tag{3.32}$$

From expression (3.32) we can see that the frequency component at $\boldsymbol{u} + \boldsymbol{u}'$ of ρ^2 is accessible when both $\boldsymbol{u}$ and $\boldsymbol{u}'$ are restricted to the support of PSF $\mathcal{D}_{\text{PSF}}$. So, the following property holds.

Property 3.4. **The frequency component $\widetilde{\rho^2}(\boldsymbol{u})$ is identifiable from the correlation function $\gamma_k(\boldsymbol{r}, \boldsymbol{r}')$ under the uncorrelated speckle assumption if and only if $\boldsymbol{u} \in \mathcal{D}_{\text{PSF}} \ominus \mathcal{D}_{\text{PSF}}$.**

We stress here that the spectral components $\widetilde{\rho^2} = \tilde{\rho} * \tilde{\rho}$ can be retrieved on $\mathcal{D}_{\text{PSF}} \ominus \mathcal{D}_{\text{PSF}}$ does not mean that ρ can be retrieved on the same domain, nor in any other domain.

Remark 3.1. We remark here that Property 3.4 still holds if only the variance $\gamma_k(\boldsymbol{r},\boldsymbol{r})$ are considered instead of all cross-correlation values $\gamma_k(\boldsymbol{r},\boldsymbol{r}')$. Let us denote $\nu(\boldsymbol{r}) = \gamma_k(\boldsymbol{r},\boldsymbol{r})$, then we can deduce from (3.30):

$$\nu(\boldsymbol{r}) = \gamma_{\text{spec}}(\mathbf{0})\ (\rho^2 * h^2)(\boldsymbol{r}). \tag{3.33}$$

So the spectral components of $\rho^2(\boldsymbol{u})$ can be identified from $\nu(\boldsymbol{r})$ as long as $\boldsymbol{u} \in \mathcal{D}_{\text{PSF}} \ominus \mathcal{D}_{\text{PSF}}$.

3.4 A marginal reconstruction procedure

Let us recall that the observation model in discretized form is (1.26):

$$\boldsymbol{y}_m = \mathbf{H}\mathbf{R}\mathbf{I}_m + \boldsymbol{\epsilon}_m \tag{3.34}$$

with $\mathbf{H} \in \mathbb{R}^{N\times N}$ the convolution matrix and $\mathbf{R} = \text{diag}(\boldsymbol{\rho})$. $\mathbf{I}_m \in \mathbb{R}^N$ is the m-th realization of speckle illumination, with homogeneous intensity mean I_0. According to Eqs. (3.3)–(3.8), we have:

$$\boldsymbol{\mu}_y = I_0\mathbf{H}\boldsymbol{\rho}, \qquad \boldsymbol{\Gamma}_y = \mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s\mathbf{R}\mathbf{H} + \sigma^2\mathbb{1}_N \tag{3.35}$$

where $\boldsymbol{\mu}_y, \boldsymbol{\Gamma}_y$ are respectively the mean and the covariance matrix of each measured image $\boldsymbol{y}_m$, while $\boldsymbol{\Gamma}_s$ denotes the covariance matrix of speckle patterns. $\mathbb{1}_N$ is the $N \times N$ identity matrix. The principle of marginal estimation is to infer the sample $\boldsymbol{\rho}$ from the statistical characteristics of the collected data. A typical marginal estimation procedure would consist in maximizing the likelihood of the data as a function of $\boldsymbol{\rho}$, the unknown speckle illuminations being integrated out of the likelihood. However, because of the complexity of the speckle statistics, it is hard or even impossible to express such a likelihood in closed form.

A preferable procedure is to estimate $\boldsymbol{\rho}$ by minimizing the mismatch between the theoretical second-order data statistics (3.35) and the empirical moments

$$\hat{\boldsymbol{\mu}}_y = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{y}_m, \qquad \hat{\boldsymbol{\Gamma}}_y = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{y}_m\boldsymbol{y}_m^{\text{t}} - \hat{\boldsymbol{\mu}}_y\hat{\boldsymbol{\mu}}_y^{\text{t}}. \tag{3.36}$$

One possible choice of dissimilarity measure is the Kullback-Leibler (KL) divergence $D_M(\boldsymbol{\rho}) = D_{\text{KL}}(\mathcal{N}(\hat{\boldsymbol{\mu}}_y, \hat{\boldsymbol{\Gamma}}_y)\|\mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Gamma}_y))$, in which $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ is the normal distribution of mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Gamma}$. An explicit expression for $D_M(\boldsymbol{\rho})$ is [Kul97, chapter 9]:

$$D_M(\boldsymbol{\rho}) = \frac{1}{2}\text{Tr}\big(\boldsymbol{\Gamma}_y^{-1}\widehat{\boldsymbol{\Gamma}}_y\big) + (\boldsymbol{\mu}_y - \widehat{\boldsymbol{\mu}}_y)^{\text{t}}\frac{\boldsymbol{\Gamma}_y^{-1}}{2}(\boldsymbol{\mu}_y - \widehat{\boldsymbol{\mu}}_y) + \frac{1}{2}\log\frac{|\boldsymbol{\Gamma}_y|}{|\widehat{\boldsymbol{\Gamma}}_y|} - \frac{N}{2} \tag{3.37}$$

where $|\cdot|$ and $\text{Tr}(\cdot)$ are the determinant and the trace of a square matrix, respectively. According to the (weak) law of large numbers, we have:

$$\hat{\boldsymbol{\mu}}_y \xrightarrow[M\to\infty]{\mathcal{P}} \boldsymbol{\mu}_y^*, \qquad \hat{\boldsymbol{\Gamma}}_y \xrightarrow[M\to\infty]{\mathcal{P}} \boldsymbol{\Gamma}_y^* \tag{3.38}$$

in which $\xrightarrow{\mathcal{P}}$ means convergence in probability [1] and $\boldsymbol{\mu}_y^*, \boldsymbol{\Gamma}_y^*$ denote the true values of the corresponding quantities, that are obtained by Eq. (3.35) for $\boldsymbol{\rho}$ equal to the true object. So, as $M \to \infty$, we have:

$$D_M(\boldsymbol{\rho}) \xrightarrow{\mathcal{P}} D(\boldsymbol{\rho}), \qquad \text{for every } \boldsymbol{\rho} \tag{3.39}$$

1. Convergence in probability: $\hat{\boldsymbol{\mu}}_y \xrightarrow[M\to\infty]{\mathcal{P}} \boldsymbol{\mu}_y^* \iff \forall\epsilon > 0, \lim_{M\to\infty}\text{Pr}\left(|\widehat{\boldsymbol{\mu}}_y - \boldsymbol{\mu}_y^*|) > \epsilon\right) = 0.$

with

$$D(\boldsymbol{\rho}) = D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{\mu}_y^*, \boldsymbol{\Gamma}_y^*) \| \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Gamma}_y)) \tag{3.40}$$

Let us denote the minimizers of $D_M(\boldsymbol{\rho})$ and $D(\boldsymbol{\rho})$ by $\hat{\boldsymbol{\rho}}_M$ and $\hat{\boldsymbol{\rho}}$, respectively. According to the property 3.3, $\widetilde{\hat{\boldsymbol{\rho}}}(\boldsymbol{u})$ is identifiable from $D(\boldsymbol{\rho})$ as long as $\boldsymbol{u} \in \mathcal{D}_1 \cup \mathcal{D}_2$. However, the pointwise convergence of (3.39) is not strong enough to assure the convergence: $\hat{\boldsymbol{\rho}}_M \xrightarrow{\mathcal{P}} \hat{\boldsymbol{\rho}}$. Because the value $\hat{\boldsymbol{\rho}}_M$ depends on the whole function $D_M(\boldsymbol{\rho})$. Section 3.4.1 addresses the consistency analysis of $\hat{\boldsymbol{\rho}}_M$.

3.4.1 Asymptotic analysis

The statistical principle behind such an inferential principle is called *minimum contrast estimation* [DCD12, BD15a], or alternatively, *M-estimation* [VdV00, Chap. 5]. Given a set of observations $(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_M)$ and unknown parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^N$, the theory of minimum contrast estimation relies on the following definition.

Definition 1. **Let $\boldsymbol{y}_m$, $m = 1, \ldots, M$, be independent, identically distributed data vectors, each taking its values in $\mathcal{Y}$, with a common probability distribution depending on a parameter vector $\boldsymbol{\theta}^*$ in $\boldsymbol{\Theta}$. A function $C : \mathcal{Y} \times \boldsymbol{\Theta} \to \mathbb{R}$ is a *contrast function* for $\boldsymbol{\theta}^*$ if its statistical expectation $J(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = \mathrm{E}[C(\boldsymbol{y}, \boldsymbol{\theta})]$ has a strict minimum at $\boldsymbol{\theta}^*$. Then, the minimizer $\hat{\boldsymbol{\theta}}_M$ of the *contrast process*,**

$$\hat{\boldsymbol{\theta}}_M = \arg\min_{\boldsymbol{\theta}} \frac{1}{M} \sum_{m=1}^{M} C(\boldsymbol{y}_m, \boldsymbol{\theta}), \tag{3.41}$$

is called a *minimum contrast estimate*.

The theory of minimum contrast estimation ensures the consistency (and asymptotic normality) of $\hat{\boldsymbol{\theta}}_M$, under appropriate technical conditions [BD15b]. In particular, $\hat{\boldsymbol{\theta}}_M$ converges in probability towards $\boldsymbol{\theta}^*$ if the contrast process (3.41) converges uniformly in $\boldsymbol{\Theta}$ towards $D(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ when $M \to \infty$, $\boldsymbol{\Theta}$ being a bounded set. The bounded constraint of $\boldsymbol{\Theta}$ may be relaxed, while still ensuring the consistency of $\hat{\boldsymbol{\theta}}_M$ [VdV00]. Such technical aspects fall outside the scope of the thesis.

To continue our asymptotic analysis, in order to get rid of the high-frequency parts in ρ that cannot be identified (see Section 3.3), we define the variable $\boldsymbol{\theta} = \mathcal{S}\boldsymbol{\rho}$, in which $\mathcal{S}$ is an ideal low-pass filter with frequency support given by $\mathcal{D}_1 \cup \mathcal{D}_2$, and we note $\mathcal{S}^+$ the pseudo-inverse of $\mathcal{S}$. Now, consider the related cost function:

$$J_M(\boldsymbol{\theta}) = D_M(\mathcal{S}^+\boldsymbol{\theta}), \text{ with } D_M \text{ defined in Eq. (3.37).} \tag{3.42}$$

Property 3.5. **The minimizer $\hat{\boldsymbol{\theta}}_M$ of Eq. (3.42) is a minimum contrast estimate.**

Proof. Let

$$C(\boldsymbol{y}, \boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu}_y)^{\mathrm{t}} \boldsymbol{\Gamma}_y^{-1} (\boldsymbol{y} - \boldsymbol{\mu}_y)^{\mathrm{t}} + \frac{1}{2} \log |\boldsymbol{\Gamma}_y| . \tag{3.43}$$

Comparing (3.37) and (3.43), straightforward manipulations yield

$$J_M(\boldsymbol{\theta}) = \frac{1}{M} \sum_{m=1}^{M} C(\boldsymbol{y}_m, \boldsymbol{\theta}) + K,$$

where K is a term that does not depend on $\boldsymbol{\theta}$. Moreover, we could define the constrast function $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ by

$$J(\boldsymbol{\theta}^*, \boldsymbol{\theta}) := \mathrm{E}\left[C(\boldsymbol{y}, \boldsymbol{\theta})\right] = D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{\mu}_y^*, \boldsymbol{\Gamma}_y^*) \| \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Gamma}_y))$$

Compared with Eq. (3.40), we can easily verify that $J(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = D(\mathcal{S}^+\boldsymbol{\theta})$. Since $D_{\mathrm{KL}}(P_1 \| P_2) \geq 0$ for all probability measures P_1, P_2, and $D_{\mathrm{KL}}(P_1 \| P_2) = 0$ if and only if $P_1 = P_2$ almost everywhere, we deduce that $D(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ is minimum if and only if $\boldsymbol{\mu}_y = \boldsymbol{\mu}_y^*$ and $\boldsymbol{\Gamma}_y = \boldsymbol{\Gamma}_y^*$. According to Property 3.3, we conclude that $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ has a strict minimum at $\boldsymbol{\theta}^*$. So $\hat{\boldsymbol{\theta}}_M$, the minimizer of Eq. (3.42), is a minimum contrast estimate. □

Within the minimum contrast estimation framework, we can further prove that $\hat{\boldsymbol{\theta}}_m$ is consistent.

Property 3.6. **$\hat{\boldsymbol{\theta}}_m$ is a consistent estimator for $\boldsymbol{\theta}^*$.**

Proof. To prove that $\hat{\boldsymbol{\theta}}_m$ is consistent, we need two properties of function $J_M(\boldsymbol{\theta})$ in (3.42):

— $J_M(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$

— $J_M(\boldsymbol{\theta})$ converges uniformly to $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, i.e.

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |J_M(\boldsymbol{\theta}) - J(\boldsymbol{\theta}^*, \boldsymbol{\theta})| \xrightarrow{\mathcal{P}} 0$$

Demonstrating the continuity property of $J_M(\boldsymbol{\theta})$ is straightforward. We see in Eq. (3.37) that every component in $D_m(\boldsymbol{\rho})$ is continuous. According to the rules for constructing continuous functions [Mun00], it is obvious that $D_M(\boldsymbol{\rho})$ is a continuous function, therefore so is $J_M(\boldsymbol{\theta})$.

As for the uniform convergence of $J_M(\boldsymbol{\theta})$ to $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, we note that:

$$\begin{aligned}
D_M(\boldsymbol{\rho}) - D(\boldsymbol{\rho}) = {} & \frac{1}{2}\mathrm{Tr}\big(\boldsymbol{\Gamma}_y^{-1}\widehat{\boldsymbol{\Gamma}}_y\big) + (\boldsymbol{\mu}_y - \widehat{\boldsymbol{\mu}}_y)^{\mathrm{t}}\frac{\boldsymbol{\Gamma}_y^{-1}}{2}(\boldsymbol{\mu}_y - \widehat{\boldsymbol{\mu}}_y) + \frac{1}{2}\log\frac{|\boldsymbol{\Gamma}_y|}{|\widehat{\boldsymbol{\Gamma}}_y|} \\
& - \frac{1}{2}\mathrm{Tr}\big(\boldsymbol{\Gamma}_y^{-1}\boldsymbol{\Gamma}_y^*\big) - (\boldsymbol{\mu}_y - \boldsymbol{\mu}_y^*)^{\mathrm{t}}\frac{\boldsymbol{\Gamma}_y^{-1}}{2}(\boldsymbol{\mu}_y - \boldsymbol{\mu}_y^*) - \frac{1}{2}\log\frac{|\boldsymbol{\Gamma}_y|}{|\boldsymbol{\Gamma}_y^*|} \\
= {} & \frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\Gamma}_y^{-1}(\hat{\boldsymbol{\Gamma}}_y - \boldsymbol{\Gamma}_y^*)\Big) + \boldsymbol{\mu}_y^{\mathrm{t}}\boldsymbol{\Gamma}_y^{-1}(\boldsymbol{\mu}_y^* - \widehat{\boldsymbol{\mu}}_y) + \frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\Gamma}_y^{-1}(\widehat{\boldsymbol{\mu}}_y\widehat{\boldsymbol{\mu}}_y^{\mathrm{t}} - \boldsymbol{\mu}_y^*\boldsymbol{\mu}_y^{*t})\Big) + \frac{1}{2}\log\frac{|\boldsymbol{\Gamma}_y^*|}{|\hat{\boldsymbol{\Gamma}}_y|} \\
= {} & \frac{1}{2}\mathrm{Tr}\Big(\boldsymbol{\Gamma}_y^{-1}\big(\hat{\boldsymbol{\Gamma}}_y - \boldsymbol{\Gamma}_y^* + \widehat{\boldsymbol{\mu}}_y\widehat{\boldsymbol{\mu}}_y^{\mathrm{t}} - \boldsymbol{\mu}_y^*\boldsymbol{\mu}_y^{*t}\big)\Big) + \boldsymbol{\mu}_y^{\mathrm{t}}\boldsymbol{\Gamma}_y^{-1}(\boldsymbol{\mu}_y^* - \widehat{\boldsymbol{\mu}}_y) + \frac{1}{2}\log\frac{|\boldsymbol{\Gamma}_y^*|}{|\hat{\boldsymbol{\Gamma}}_y|}
\end{aligned} \tag{3.44}$$

Since the covariance matrix $\boldsymbol{\Gamma}_k = \mathbf{HR}\boldsymbol{\Gamma}_s\mathbf{RH}$ is a positive semi-definite matrix, applying Weyl's inequality [Ser02] to the definition of $\boldsymbol{\Gamma}_y$ in Eq. (3.35), we have:

$$\lambda_{\min}(\boldsymbol{\Gamma}_y) \geq \lambda_{\min}(\boldsymbol{\Gamma}_\varepsilon), \tag{3.45}$$

where $\lambda_{\min}(\mathbf{A})$ denotes the smallest eigenvalue of the matrix $\mathbf{A}$ and $\boldsymbol{\Gamma}_\varepsilon = \sigma^2\,\mathbf{Id}_N$ is the noise covariance. Since $\boldsymbol{\Gamma}_y^{-1}$ is a symmetric, positive definite, matrix, its ℓ_2-norm equals its maximum eigenvalue and:

$$\left\|\boldsymbol{\Gamma}_y^{-1}\right\| = \lambda_{\max}(\boldsymbol{\Gamma}_y^{-1}) = \frac{1}{\lambda_{\min}(\boldsymbol{\Gamma}_y)} \leq \frac{1}{\sigma^2}. \tag{3.46}$$

Combining Eqs. (3.38), (3.44) and (3.46), we have:

$$\sup_{\boldsymbol{\rho} \in \mathbb{R}^N} |D_M(\boldsymbol{\rho}) - D(\boldsymbol{\rho})| \xrightarrow{\mathcal{P}} 0 \tag{3.47}$$

Restricting the domain of $D_M(\boldsymbol{\rho})$ to $\{\boldsymbol{\rho} | \boldsymbol{\rho} = \mathcal{S}^+\boldsymbol{\theta}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, we obtain:

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |D_M(\mathcal{S}^+\boldsymbol{\theta}) - D(\mathcal{S}^+\boldsymbol{\theta})| \xrightarrow{\mathcal{P}} 0 \tag{3.48}$$

Now we have that $J_M(\boldsymbol{\theta})$ converges uniformly to $J(\boldsymbol{\theta}^*, \boldsymbol{\theta})$. □

3.5 Applying Poisson noise

The recorded data of detectors in imaging obeys Poisson statistics [BM13]. A more accurate observation model than Eq. (3.1) is:

$$\boldsymbol{y}_m = \mathcal{P}(\boldsymbol{k}_m) + \boldsymbol{\epsilon}_m, \tag{3.49}$$

where $\boldsymbol{k}_m$ is given in Eq. (3.1) and $\mathcal{P}(\boldsymbol{k}_m)$ denotes a realization of the Poisson process with mean $\boldsymbol{k}_m$. For brevity of expression, we remove the subscript m and write $g_r = g(\boldsymbol{r}),\ k_r = k(\boldsymbol{r})$, in which $\boldsymbol{r} \in \{1, \cdots, N\}$ denotes the spatial index, and $g_r = \mathcal{P}(k_r)$. Since k_r are still random variables, the Poisson variables g_r are called *doubly stochastic Poisson random variables* [BM13]. The impact of Poisson statistics on the former developments amounts to how the second-order statistics are modified, *i.e.*, we are considering the knowledge in $E[\boldsymbol{g}]$ and γ_g instead of $E[\boldsymbol{k}]$ and γ_k.

3.5.1 Mean of the Poisson vector

Since $E[g_r|k_r] = k_r$, it comes directly that

$$\begin{aligned} E[g_r] &= \int E[g_r|k_r]\, p(k_r)\, dk_r \\ &= \int k_r\, p(k_r)\, dk_r \\ &= E[k_r]. \end{aligned}$$

Therefore, the mean of the Poisson vector $\boldsymbol{g}$ equals the mean of its parameters, *i.e.,* $E[\boldsymbol{g}] = E[\boldsymbol{k}]$, with $E[\boldsymbol{k}] = E[\boldsymbol{y}_m] = I_0 \mathbf{H}\boldsymbol{\rho}$.

3.5.2 Covariance of the Poisson vector

Similarly, one can obtain the second-order statistics of $\boldsymbol{g}$:

— for $\boldsymbol{r} \neq \boldsymbol{r}'$:

$$E[g_r g_{r'}\ |k_r, k_{r'}] = E[g_r|k_r]\ E[g_{r'}|k_{r'}] = k_r k_{r'}.$$

Therefore,

$$E[g_r g_{r'}] = \iint k_r k_{r'} p(k_r, k_{r'}) dk_r dk_{r'} = E[k_r k_{r'}].$$

Since from Section 3.5.1 we have $E[g_r] = E[k_r]$, we conclude that $\gamma_g(\boldsymbol{r}, \boldsymbol{r}') = \gamma_k(\boldsymbol{r}, \boldsymbol{r}')$.

— for $\boldsymbol{r} = \boldsymbol{r}'$:

$$E[g_r^2|k_r] = \text{Var}(g_r^2|k_r) + (E[g_r^2|k_r])^2 = k_r + k_r^2.$$

Therefore,

$$E[g_r^2] = \int (k_r + k_r^2) p(k_r) dk_r = E[k_r] + E[k_r^2].$$

We finally get $\gamma_g(\boldsymbol{r}, \boldsymbol{r}) = E[g_r^2] - E[g_r]^2 = E[k_r] + E[k_r^2] - E[k_r]^2 = \gamma_k(\boldsymbol{r}, \boldsymbol{r}) + E[k_r]$. Consequently, we can see that only the diagonal of the covariance matrix is modified when the Poisson statistics are taken into consideration. So, the super-resolution analysis in 3.3 still holds. As the second-order moments have a quadratic dependence in the random vector $\boldsymbol{k}$, the difference of the covariance matrix after introducing Poisson statistics can be neglected except when $\boldsymbol{k}$ is small in low photon counting cases.

3.6 Numerical optimization

Up to an additive constant K, Eq. (3.37) admits the following alternative expression:

$$D_M(\boldsymbol{\rho}) = \frac{1}{2}\log|\boldsymbol{\Gamma}_y| + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Gamma}_y^{-1}\mathbf{W}\right) + K \tag{3.50}$$

where $\mathbf{W} = \frac{1}{M}\mathbf{V}\mathbf{V}^{\mathrm{t}}$, with $\mathbf{V} = (\boldsymbol{y}_1 - \boldsymbol{\mu}_y|\cdots|\boldsymbol{y}_M - \boldsymbol{\mu}_y)$. In practice, the minimizer $\hat{\boldsymbol{\rho}}$ of D_M has no closed-form expression. Local optimization methods relying on gradient-based iterations can be used to minimize D_M, whose gradient reads (see appendix E):

$$\nabla D_M(\boldsymbol{\rho}) = \left(\left(\boldsymbol{\Omega}^{\mathrm{t}}\left(\boldsymbol{\Gamma}_y - \mathbf{W}\right)\boldsymbol{\Omega}\right)\circ\mathbf{C}\right)\boldsymbol{\rho} - \frac{I_0}{M}\boldsymbol{\Omega}^{\mathrm{t}}\mathbf{V}\mathbf{1}_N \tag{3.51}$$

where $\circ$ denotes the Hadamard (*i.e.,* element-wise) product and $\boldsymbol{\Omega} = \boldsymbol{\Gamma}_y^{-1}\mathbf{H}$. Then we choose a quasi-Newton method to optimize $D_M(\boldsymbol{\rho})$. Given an approximation $\mathbf{H}_e^{(k)}$ of the Hessian matrix $\nabla^2 D_M(\boldsymbol{\rho}_k)$, a quasi-Newton iteration is given by:

i) Solve $\mathbf{H}_e^{(k)}\boldsymbol{d}_k = -\nabla D_M(\boldsymbol{\rho}_k)$

ii) Update $\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k + \alpha_k\boldsymbol{d}_k$

where $\boldsymbol{d}_k$ is the descent direction at the k-th iteration and α_k is step length obtained by a line search algorithm to satisfy the Wolfe conditions (see [NW06, Chapter 3]). If we set $\mathbf{H}_e^{(k)} = \nabla^2 D_M(\boldsymbol{\rho}_k)$, it becomes a classical Newton method. In quasi-Newton methods, we approximate the Hessian matrix $\mathbf{H}_e^{(k)}$ using only the gradient of the objective function at each iteration. Specifically, in our simulations, we use the L-BFGS [NW06, Chapter 6] to optimize $D_M(\boldsymbol{\rho})$ in Eq. (3.50) and the L-BFGS-B [BLNZ95] algorithm to incorporate positivity constraint of $\boldsymbol{\rho}$. A brief discussion about the two algorithms can be found in Appendix D.

The gradient of current value $\boldsymbol{\rho}_k$ involves the inverse of the $N \times N$ covariance matrix $\boldsymbol{\Gamma}_y$, where the compuputational complexity is $\mathcal{O}(N^3)$. Thus this method can only be applied for moderate size problems. A method to reduce the memory and computational burden is proposed in next chapter.

3.7 Simulation results

We now show simulation results under the same conditions as in Chapter 2 in order to validate the theoretical developments and the estimator properties of the marginal approach. Both the asymptotic case ($M = \infty$) and finite M are considered. The asymptotic case is simulated by assigning $\hat{\boldsymbol{u}}_y$ and $\hat{\boldsymbol{\Gamma}}_y$ their true values $I_0\mathbf{H}\boldsymbol{\rho}^*$ and $\mathbf{H}\mathrm{diag}(\boldsymbol{\rho}^*)\boldsymbol{\Gamma}_s\mathrm{diag}(\boldsymbol{\rho}^*)\mathbf{H}^T$, with $\boldsymbol{\rho}^*$ the true value of object. Small-size object (60×60) and data are generated so that the estimator defined in Eq. (3.50) can be computed. The correlation function of the speckle patterns was set to $\boldsymbol{\gamma}_{\text{spec}}(\boldsymbol{r}) = \frac{I_0^2}{h(0)}h(\boldsymbol{r})$, unless stated otherwise.

The L-BFGS algorithm is chosen to optimize the marginal criterion. If the positivity constraint on $\boldsymbol{\rho}$ is considered, we use the L-BFGS-B algorithm. All simulations were performed with the same initialization, at the image obtained by averaging all measured images $\frac{1}{M}\sum_m \boldsymbol{y}_m$. The number of corrections stored in the limited memory matrix is 20 (parameter L in L-BFGS). The algorithm will terminate when either one of the following conditions is satisfied:

— The relative descends $\left|\frac{D_{MP}(\boldsymbol{\rho}_{k+1}) - D_{MP}(\boldsymbol{\rho}_k)}{D_{MP}(\boldsymbol{\rho}_k)}\right| \leq 10^{-5}$ for finite number of speckle patterns and $\left|\frac{D_{MP}(\boldsymbol{\rho}_{k+1}) - D_{MP}(\boldsymbol{\rho}_k)}{D_{MP}(\boldsymbol{\rho}_k)}\right| \leq 10^{-8}$ for infinite speckle patterns

— The ℓ_∞-norm of the gradient $\|\nabla D_{MP}(\boldsymbol{\rho}_k)\|_\infty \leq 10^{-3}$.

The reconstructed objects by our marginal approach, using different numbers of speckle patterns ($M = \infty$, $M = 300$ and $M = 100$), are shown in Figure 3.1. It turned out that the positivity constraint can not further increase the super-resolution. In real experiments, the raw images are blured by background noise. However, the reconstruction is still robust if we only take the covariance information into consideration (see Section 3.8). The reconstructed results by minimizing Eq. (3.53) are shown in the third line in Fig. 3.1.

The normalized RAPS of error and MCF curves (respectively defined by (2.63) and (2.64)) of the reconstructed objects with 300 speckle patterns and SNR 40dB are shown in Fig. 3.2. We see that the quality of the reconstructions by marginal approach are similar to that obtained by the constrained $\ell_{2,1}$ regularizer shown in Chapter 2.

The evolution of the normalized RAPS of errors and of the MCF during iterations in the asymptotic case are shown in Figure 3.3. It shows that the information in low resolution part is retrieved first, and gradually, the information in super-resolution part is retrieved up to a factor of two, which confirm the property 3.3 since the radius of $\mathcal{D}_{\text{spec}} \ominus \mathcal{D}_{\text{spec}}$ is twice of the radius of OTF in this case.

3.8 Experiments with real data

In real data obtained from microscopy, there are three noise sources: photon noise, dark noise and readout noise. Photon noise and dark noise are independent Poisson processes, therefore so is their sum. The readout noise usually follows a Gaussian distribution. A more legitimate data formation model of real data is then:

$$\boldsymbol{y}_m = \mathcal{P}(\mathbf{HRI}_m + \boldsymbol{b}) + \boldsymbol{\epsilon}_m \tag{3.52}$$

where $\mathcal{P}(\cdot)$ denotes the realization of a Poisson process, $\boldsymbol{\epsilon}_m \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_N)$ denotes readout noise, and $\boldsymbol{b} \in \mathbb{R}^N$ is the mean intensity of the background caused by out-of-focus light. Due to the presence of background $\boldsymbol{b}$, our previous objective function $D_M(\rho)$ as shown in (3.37) is not accurate, since the empirical mean of the data is not $I_0\mathbf{H}\boldsymbol{\rho}$ any more. Instead of estimating $\boldsymbol{b}$ and then subtracting it from the data, we notice that the empirical covariance of the data is not affected by $\boldsymbol{b}$. Therefore, we slightly adapt the objective function introduced in (3.50) and reconstruct ρ with only the second-order statistics (that is, we do not consider the mean values):

$$D_R(\boldsymbol{\rho}) = D_{\text{KL}}(\mathcal{N}(0, \hat{\boldsymbol{\Gamma}}_y) \| \mathcal{N}(0, \boldsymbol{\Gamma}_y)) = \frac{1}{2}\log|\boldsymbol{\Gamma}_y| + \frac{1}{2}\text{Tr}\big(\boldsymbol{\Gamma}_y^{-1}\widehat{\boldsymbol{\Gamma}}_y\big) + K \tag{3.53}$$

where K is a constant number. The gradient of $D_R(\boldsymbol{\rho})$ is:

$$\nabla D_R(\boldsymbol{\rho}) = \left(\left(\boldsymbol{\Omega}^{\text{t}}\left(\boldsymbol{\Gamma}_y - \widehat{\boldsymbol{\Gamma}}_y\right)\boldsymbol{\Omega}\right) \circ \boldsymbol{\Gamma}_{\text{s}}\right)\boldsymbol{\rho} \tag{3.54}$$

The poisson noise is neglected. To make the reconstruction possible in normal computer, we select only a small part from the raw images of Podosome sample and actin sample. The reconstructions are shown in Figures 3.4,3.5,3.6. The spatial sampling rate of raw images is equal or slightly above the Nyquist rate $\frac{\lambda}{4NA}$. For the final reconstructed object, an up-sampling factor of two with respect to the camera acquisitions was applied so that the sampling rate reaches that of the expected super-resolution. The wide field image is plotted by averaging 1000 raw images, and the deconvolution is obtained by:

$$\boldsymbol{\rho}_{\text{decon}} = \arg\min_{\boldsymbol{\rho}} \|\bar{\boldsymbol{y}} - \mathbf{H}\boldsymbol{\rho}\|^2 + \beta\|\boldsymbol{\rho}\|^2 \tag{3.55}$$

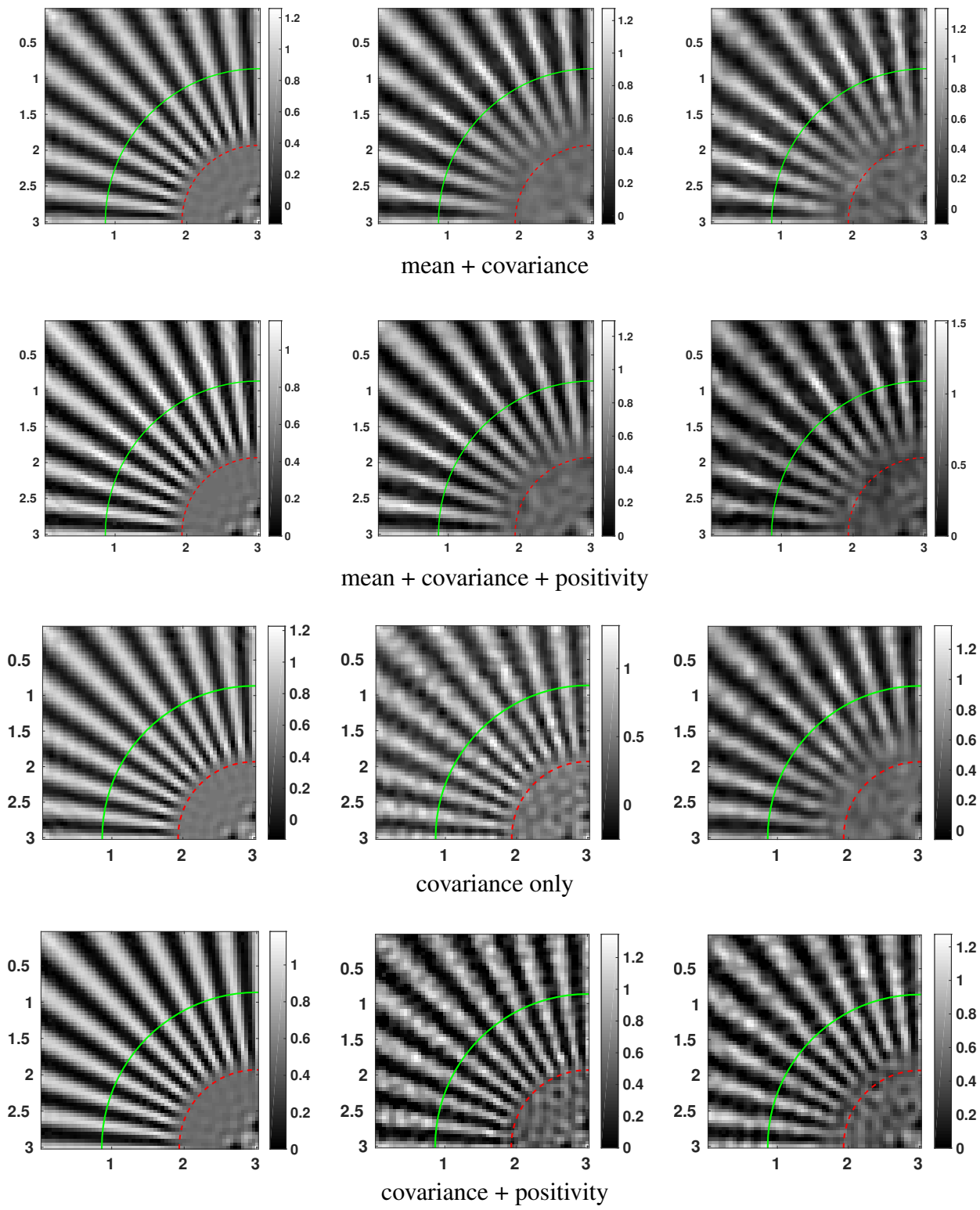
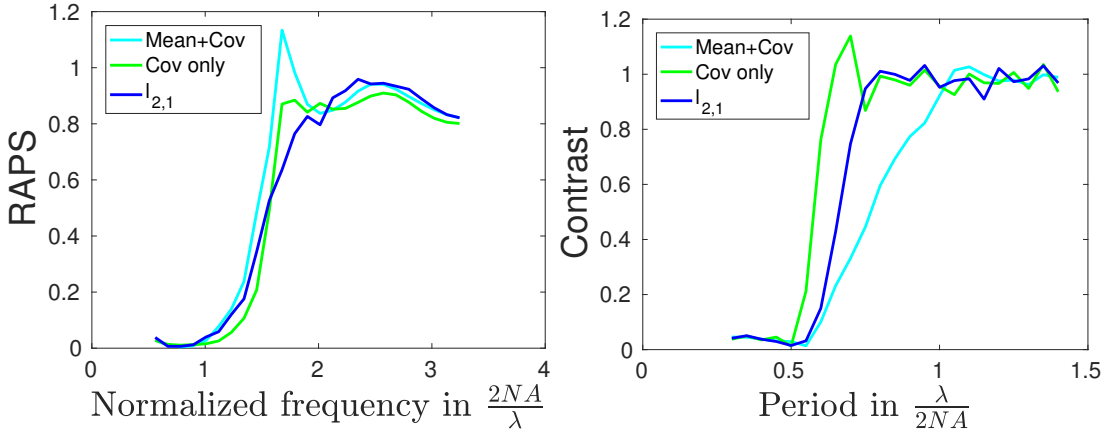


Figure 3.1 – **Reconstruction by marginal approach with different number of speckle patterns and 40dB Gaussian white noise.** The first row corresponds to the L-BFGS algorithm, while the second line is obtained with the L-BFGS-B algorithm including the positivity constraint. The third row shows the reconstructed objects with only the covariance information taken into considered and the images shown in last row are obtained after adding positivity constraint to the KL divergence of covariance .



(a) The normalized RAPS curves.

(b) The Contrast function.

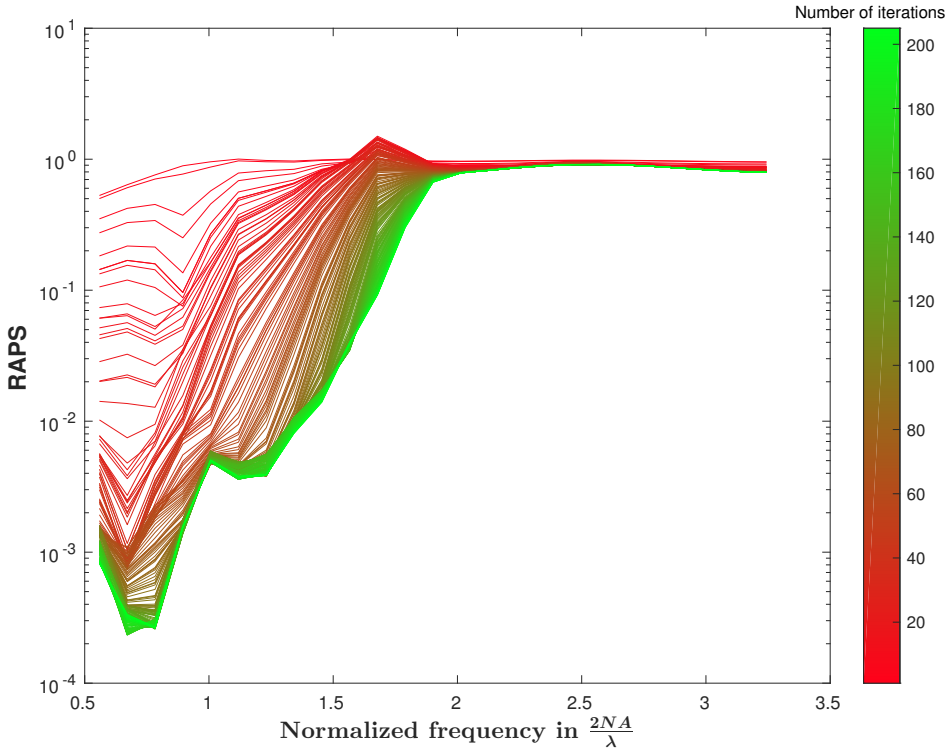
Figure 3.2 – **Measurements of error (left panel) and of contrast (right panel) as a function of radial frequency / period of different reconstruction methods with 300 speckle patterns and SNR 40dB.**

with the regularizer parameter $\beta = 0\,01$. We see that the structure details in images reconstructed by the marginal approach are clearer than those of the wide-field images and their deconvolution, even if we use only 50 speckle patterns. In raw images obtained with the Podosome sample, we can see a strong photobleaching (section 1.4) effect (see Figure 3.7a), that is why the reconstruction with more speckle patterns does not appear better. On the contrary, the fluorophores emitted relatively stable light in experiments with Actin sample during the imaging process (see Figure 3.7b).

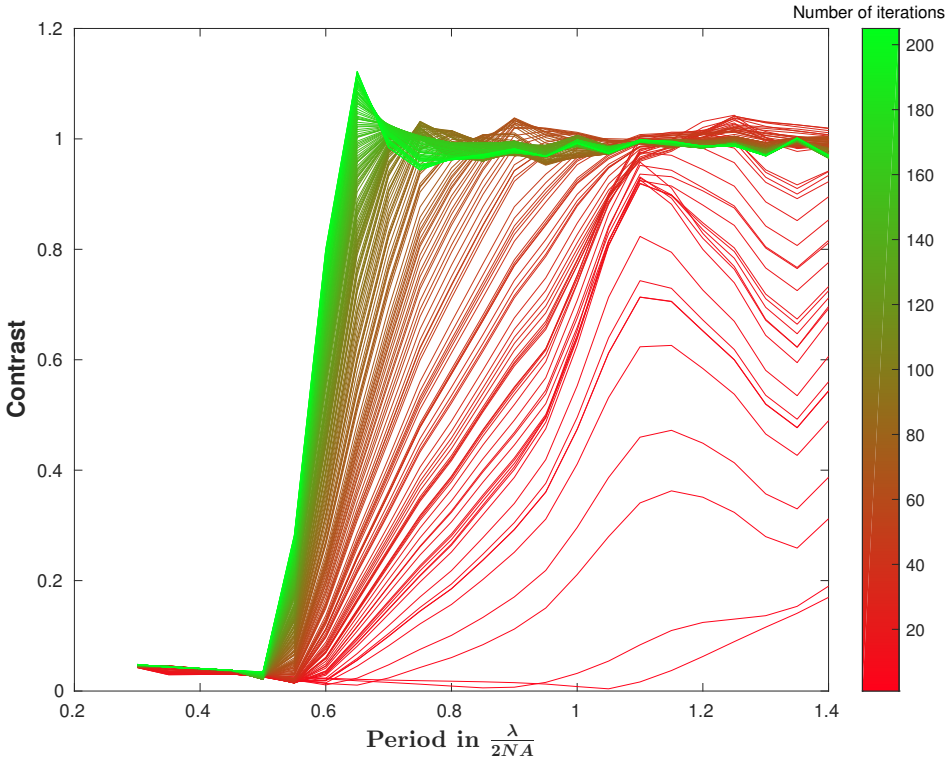
3.9 Conclusion

In this chapter, we analyzed the SR capacity of blind-speckle-SIM reconstruction using second-order moments fitting. Our analysis demonstrates that we can (at most) double the frequency support by using speckle patterns in asymptotic sense, *i.e.*, with infinite number of speckle patterns. A marginal estimator is proposed by minimizing the Kullback-Leibler divergence between the empirical moments and the theoretical ones. A theoretical analysis showed that this estimator is a minimum contrast estimator and that it is asymptotically consistent. Our analysis demonstrates that the SR capacity of the marginal approach also holds under Poisson statistics. These theoretical analysis are demonstrated on both simulated and experimental data.

Unfortunately, the computations of the objective function and of its gradient require the explicit inversion of the $N \times N$ covariance matrix $\mathbf{\Gamma}_y$ at each iteration of the optimization procedure, the cost of which being $\mathcal{O}(N^3)$, with N the number of pixels. This is clearly an unaffordable computing task for realistic size problems. Even worse, simply computing and storing the matrix $\mathbf{\Gamma}_y$ is impossible in realistic conditions. For example, for a 512×512 image, storing $\mathbf{\Gamma}_y$ with double-precision requires 512 Gigabytes (see appendix B.2). Some possible variations of the second-order method address this problem in Chapter 4.



(a) Normalized RAPS curves of each iteration.



(b) Contrast function of each iteration.

Figure 3.3 – **Behavior of L-BFGS iterations in the asymptotic case (infinite number of speckle patterns). Measurements of error (top panel) and of quality (bottom panel) as a function of radial frequency / period.**

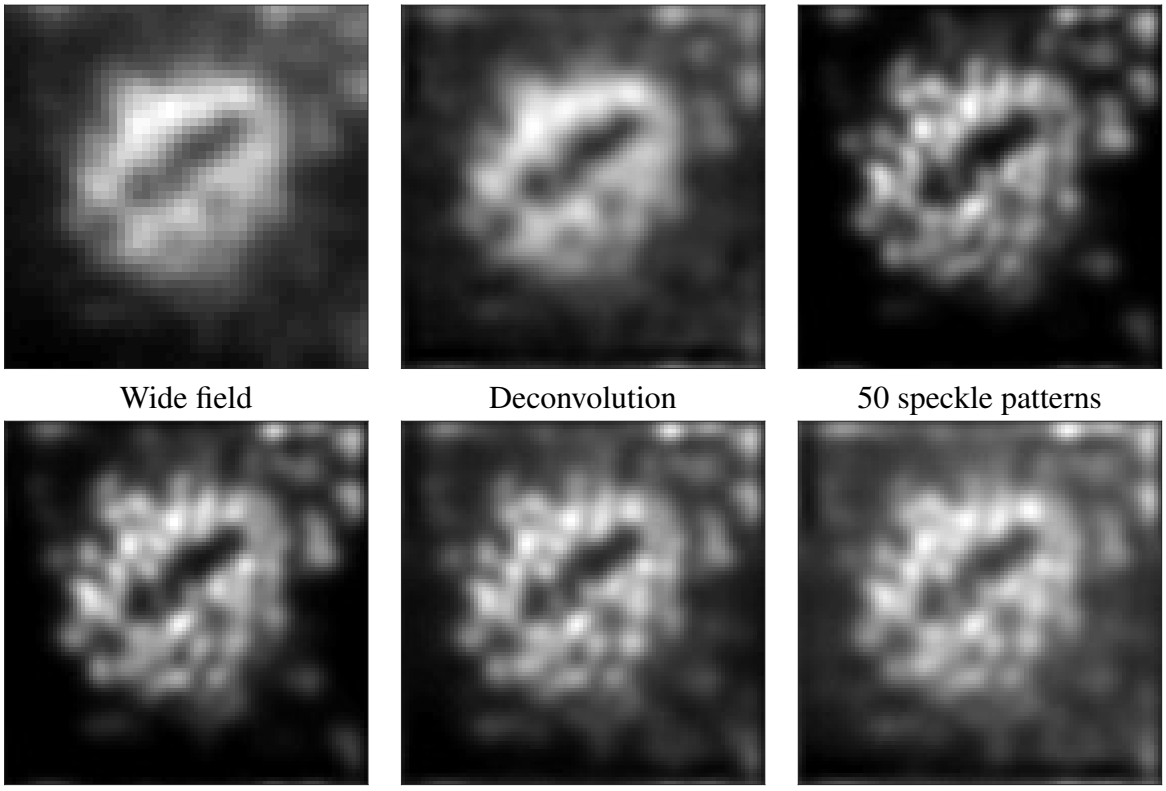


Figure 3.4 – **Marginal reconstruction of a small patch in Posodome sample**.

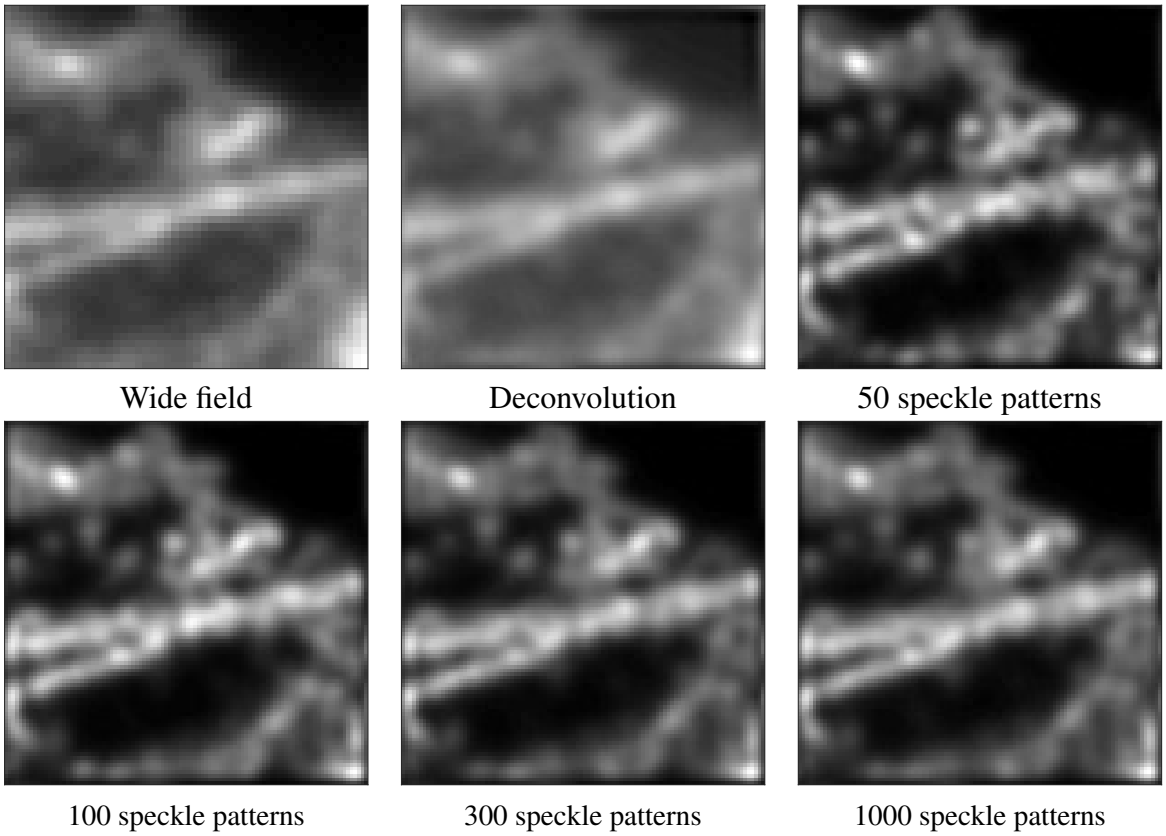


Figure 3.5 – **Marginal reconstruction of a small sparse patch in actin sample**.

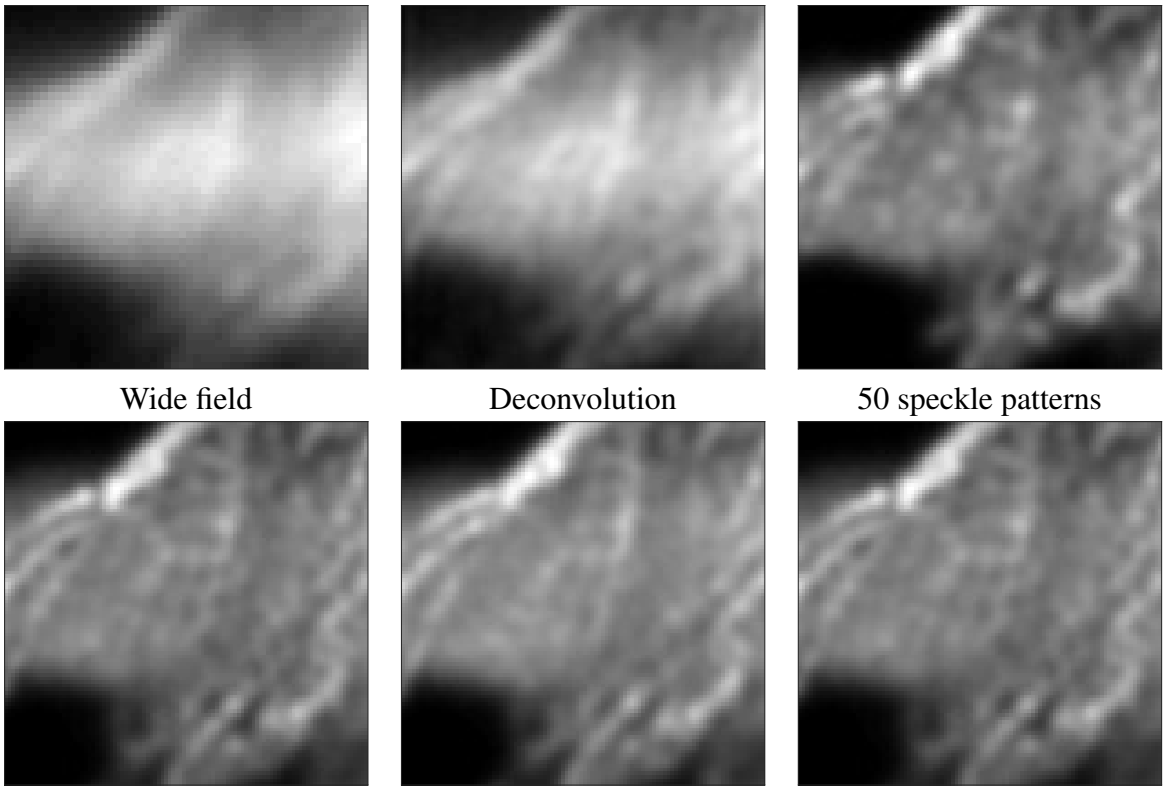
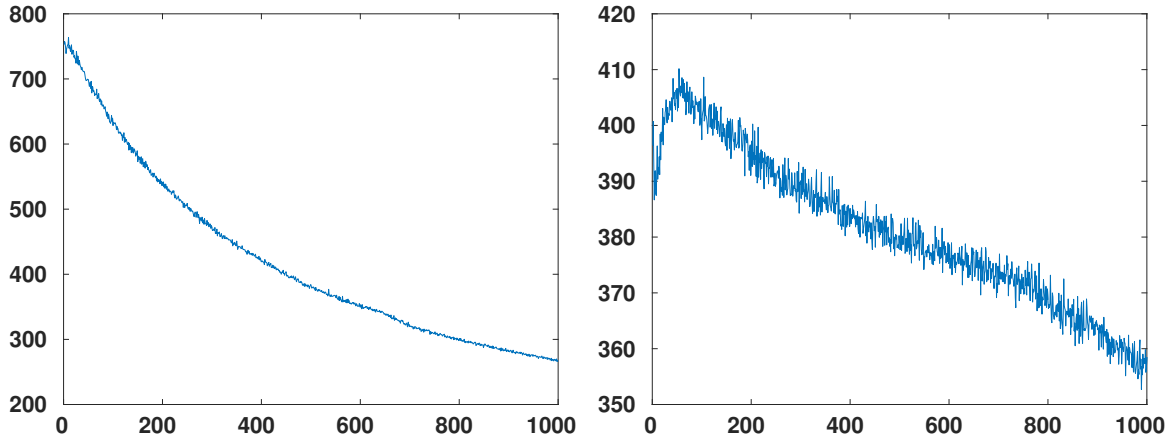


Figure 3.6 – **Marginal reconstruction of a small dense patch in actin sample**.



(a) Podosome sample (b) Actin sample

Figure 3.7 – **Mean value of raw images for 1000 speckle patterns.** We can see strong photobleaching effect in the Podosome sample, while the fluorophores emitted relatively stable with Actin sample.

Chapter 4

A patch-based approach to marginal estimation

4.1 Introduction

As it has been presented in Chapter 3, the statistical analysis of *blind-speckleSIM* shows that a super-resolution (SR) factor equal to two can be reached based on the second-order data statistics, provided that sufficiently many speckle images are acquired. The marginal algorithm proposed in Chapter 3 was validated on small size simulated blind-SIM examples. However, the associated computing cost and storage requirement are in $\mathcal{O}(N^3)$, with N the number of pixels in each image, which makes it an unaffordable task for realistic image sizes.

In this chapter, we propose a new marginal blind-speckleSIM algorithm that scales to real-size two-dimensional SIM. Our key idea is to neglect some of the long-distance correlations between pixels in the acquired images, so that the resulting approximated covariance becomes block-diagonal. Equivalently, each image is cut into a set of patches, which contribute independently to the image formation model. Then, the required inverse matrix computations and storage are performed in lower dimension, thus reducing the computational burden. In Section 4.2 we introduce the patch-based marginal algorithm. The computational analysis is presented in Section 4.3.

4.2 The patch-based marginal approach

Our main proposition consists in dividing the observed image grid into P subgrids. The pixels belonging to each subgrid are not necessarily contiguous, and the subgrids are possibly overlapping. For the sake of simplicity, let us consider that all subgrids contain the same number L of pixels. Each observed image $\boldsymbol{y}_m$ then generates P subimages $\boldsymbol{y}_m = (\boldsymbol{y}_{m_1}, \cdots, \boldsymbol{y}_{m_P})$. In what follows, each subimage will be called a patch.

Since each patch is extracted from $\boldsymbol{y}_m$, whose data formation model is shown in Eq. (3.1), the observation model for each patch reads:

$$\boldsymbol{y}_{m_p} = \mathbf{H}_p \mathbf{R} \mathbf{I}_m + \boldsymbol{\epsilon}_{m_p} \tag{4.1}$$

where the lines of $\mathbf{H}_p \in \mathbb{R}^{L \times N}$ are extracted from $\mathbf{H}$ according to the geometry of the p-th patch. The mean vector and the covariance matrix of $\boldsymbol{y}_{mp}$ take the following form:

$$\boldsymbol{\mu}_{y_p} = I_0 \mathbf{H}_p \boldsymbol{\rho}, \qquad \boldsymbol{\Gamma}_{y_p} = \mathbf{H}_p \mathbf{R} \boldsymbol{\Gamma}_s \mathbf{R} \mathbf{H}_p^{\mathrm{t}} + \sigma^2 \mathbb{1}_L \tag{4.2}$$

where $\mathbb{1}_L$ denotes the $L \times L$ identity matrix. Our proposition consists in modifying the previously introduced minimum contrast estimator shown in Section 3.4, so that the data are taken into account independently from one patch to another. For this purpose, we simply replace the KL divergence criterion (3.37) by a sum of KL divergence terms, each one being specific to one patch:

$$\begin{aligned} D_{MP}(\boldsymbol{\rho}) &= \sum_{p=1}^{P} D_{\mathrm{KL}}(\mathcal{N}(\hat{\boldsymbol{\mu}}_{y_p}, \hat{\boldsymbol{\Gamma}}_{y_p}) \| \mathcal{N}(\boldsymbol{\mu}_{y_p}, \boldsymbol{\Gamma}_{y_p})) + \beta \|\boldsymbol{\rho}\|^2 \\ &= \frac{1}{2} \sum_{p=1}^{P} \left(\log \left| \boldsymbol{\Gamma}_{y_p} \right| + \mathrm{Tr} \left(\boldsymbol{\Gamma}_{y_p}^{-1} \mathbf{W}_p \right) \right) + \beta \|\boldsymbol{\rho}\|^2 + K' \end{aligned} \tag{4.3}$$

where K' is an additive constant and $\mathbf{W}_p = \frac{1}{M} \mathbf{V}_p \mathbf{V}_p^{\mathrm{t}}$, with $\mathbf{V}_p = (\boldsymbol{y}_{1p} - \boldsymbol{\mu}_p, \cdots, \boldsymbol{y}_{Mp} - \boldsymbol{\mu}_p)$. The corresponding gradient expression is:

$$\nabla D_{MP}(\boldsymbol{\rho}) = \sum_{p=1}^{P} \left(\left(\left(\boldsymbol{\Omega}_p^{\mathrm{t}} \left(\boldsymbol{\Gamma}_{y_p} - \mathbf{W}_p \right) \boldsymbol{\Omega}_p \right) \circ \boldsymbol{\Gamma}_s \right) \boldsymbol{\rho} - \frac{I_0}{M} \boldsymbol{\Omega}_p^{\mathrm{t}} \mathbf{V}_p \mathbf{1}_L \right) + 2\beta \boldsymbol{\rho} \tag{4.4}$$

with $\boldsymbol{\Omega}_p = \boldsymbol{\Gamma}_{y_p}^{-1} \mathbf{H}_p \in \mathbb{R}^{L \times N}$. Now the values of the elements in $\boldsymbol{\Gamma}_y$ are assigned to 0 when the corresponding two pixels are not from the same patch, thus the $N \times N$ matrix inversion $\boldsymbol{\Gamma}_y^{-1}$ is replaced by P matrix inversions $\boldsymbol{\Gamma}_{y_p}^{-1}$ of size $L \times L$.

The geometry shape of the patch could be rectangle, triangle or even irregular shape. The pixels belonging to each patch are not necessarily contiguous. For the sake of simplicity, let us consider that all subgrids contain the same number L of pixels. Considering the feasibility of implementation, we choose rectangle patches in our simulation. The patches could overlap each other or they could simply correspond to a partition of the plane.

4.3 Algebraic tricks

Now let us review the expression of the gradient in (4.4). We find that, although most computations involved in the evaluation of the cost function (4.3) and its gradient (4.4) are matrix operations in size L, the matrix $\left(\boldsymbol{\Omega}_p^{\mathrm{t}} \left(\boldsymbol{\Gamma}_p - \mathbf{W}_p \right) \boldsymbol{\Omega}_p \right) \circ \boldsymbol{\Gamma}_s$ is still a $N \times N$ matrix, which is difficult to deal with for realistic size images. This problem can be solved by re-expressing the gradient with its equivalent form.

First, we introduce the auxiliary matrix $\mathbf{Z}_p = \boldsymbol{\Gamma}_{y_p} - \mathbf{W}_p$. For brevity, we abbreviate the subscript p here. Let us denote $\boldsymbol{z}_\ell^{\mathrm{t}} \in \mathbb{R}^L$ and $\boldsymbol{\omega}_\ell \in \mathbb{R}^N$ the ℓ-th line and the ℓ-th column of

matrices $\mathbf{Z}$ and $\mathbf{\Omega}^t$, respectively. Then, we have:

$$\begin{aligned}
\Big(\big(\mathbf{\Omega}^t\mathbf{Z}\mathbf{\Omega}\big)\circ\mathbf{\Gamma}_s\Big)\boldsymbol{\rho} &= \Big(\big(\sum_{u=1}^{L}\boldsymbol{\omega}_u\boldsymbol{z}_u^t\mathbf{\Omega}\big)\circ\mathbf{\Gamma}_s\Big)\boldsymbol{\rho} \\
&= \Big(\big(\sum_{u=1}^{L}\sum_{v=1}^{L}\boldsymbol{\omega}_u(z_{uv})\boldsymbol{\omega}_v^t\big)\circ\mathbf{\Gamma}_s\Big)\boldsymbol{\rho} \\
&= \sum_{u=1}^{L}\sum_{v=1}^{L}\boldsymbol{\omega}_u\circ\Big(z_{uv}\big[\mathbf{\Gamma}_s(\boldsymbol{\omega}_v\circ\boldsymbol{\rho})\big]\Big) \\
&= \Big(\mathbf{\Omega}^t\circ\Big[\mathbf{\Gamma}_s(\mathbf{\Omega}^t\circ\boldsymbol{\rho}_{\text{rep}})\mathbf{Z}^t\Big]\Big)\mathbf{1}_L \\
&= \Big(\mathbf{\Omega}^t\circ\Big[\mathbf{\Gamma}_s\mathbf{R}\mathbf{\Omega}^t(\mathbf{\Gamma}-\mathbf{W})^t\Big]\Big)\mathbf{1}_L \\
&= \Big(\mathbf{\Omega}^t\circ\Big[\mathbf{\Gamma}_s\mathbf{R}\mathbf{H}_p^t(\mathbb{1}_L-\mathbf{\Gamma}^{-1}\mathbf{W})\Big]\Big)\mathbf{1}_L
\end{aligned} \tag{4.5}$$

where $\boldsymbol{\rho}_{\text{rep}} = \boldsymbol{\rho}\mathbf{1}_L^t \in \mathbb{R}^{N\times L}$ is obtained by replicating L times the $N\times 1$ column vector $\boldsymbol{\rho}$ corresponding to the object. Then, an equivalent expression for the gradient $\nabla D_{MP}(\boldsymbol{\rho})$ is:

$$\nabla D_{MP}(\boldsymbol{\rho}) = \sum_{p=1}^{P}\left(\Big(\mathbf{\Omega}_p^t\circ\Big[\mathbf{\Gamma}_s\mathbf{R}\mathbf{H}_p^t(\mathbb{1}_L-\mathbf{\Gamma}_{y_p}^{-1}\mathbf{W}_p)\Big]\Big)\mathbf{1}-\frac{I_0}{M}\mathbf{\Omega}_p^{\mathrm{t}}\mathbf{V}_p\mathbf{1}_L\right)+2\beta\boldsymbol{\rho} \tag{4.6}$$

One possible implementation for both the objective function and its gradient is given in algorithm 5. The main costs that are written as comments in the pseudo-code. In line 11 when we compute $\mathbf{\Gamma}_s\mathbf{A}_p^{\mathrm{t}}$, we choose the FFT algorithm since $\mathbf{\Gamma}_s$ has the same structure as matrix $\mathbf{H}$ according to our analysis in Appendix A. The number of flops for one patch at each iteration is therefore (reminding that we only take the multiplications into account and that the addition operations are neglected):

$$f_p(N) = LN\log N + 3L^2N + 3LN + ML^2 + 2L^3$$

If we consider non-overlapping patches, we have $P = \frac{N}{L}$ patches, and the total number of flops in one iteration is:

$$f(N) = N^2\log N + (3L+3)N^2 + (ML+2L^2)N + N\log N$$

Now the computational complexity for each iteration has reduced to $\mathcal{O}(N^2(\log N + L))$. What is more, this algorithm can be highly parallelized since each term in the sum of Eq. (4.3) for one patch does not rely on the others and thus they can be processed by different units in parallel. Once the expression of the gradient is obtained, L-BFGS and L-BFGS-B algorithms are choosen to solve the optimization problem $\min_{\boldsymbol{\rho}} D_{MP}(\boldsymbol{\rho})$ without and with positivity constraint, respectively.

4.4 Statistical analysis

Property 4.1. **The minimizer of D_{MP} is a minimum contrast estimate, provided that every pixel pair $\{u, v\}$ belongs at least to one patch.**

Algorithm 5: Pseudo-code to compute criterion (4.3) and its gradient (4.4).

ObjectGradient $(\boldsymbol{\rho}_k)$;
Input : The current value $\boldsymbol{\rho}_k$, β, I_0
Output: $D_{MP}(\boldsymbol{\rho}_k)$ and $\nabla D_{MP}(\boldsymbol{\rho}_k)$
Initialize $D_{MP}(\boldsymbol{\rho}_k)$ and $\nabla D_{MP}(\boldsymbol{\rho}_k)$;
$\boldsymbol{\rho}_{\text{rep}} = \boldsymbol{\rho}_k \mathbf{1}_L^t$;
$\bar{\mathbf{Y}} = I_0 \mathbf{H} \boldsymbol{\rho}_k$; // $N \log N$ flops
for $p \leftarrow 1$ **to** P **do**
  generate the matrix $\mathbf{H}_p$;
  $\mathbf{Y}_p = (\mathbf{y}_{1_p}, \cdots, \mathbf{y}_{M_p})$;
  $\mathbf{A}_p = \mathbf{H}_p \circ \boldsymbol{\rho}_{\text{rep}}^t$; // $\mathbf{A}_p = \mathbf{H}_p \mathbf{R}$, LN flops
  $\boldsymbol{\mu}_p = \bar{\mathbf{Y}}_p$;
  $\mathbf{V}_p = \mathbf{Y}_p - \boldsymbol{\mu}_p \mathbf{1}_L^t$;
  $\mathbf{B}_p = \boldsymbol{\Gamma}_s \mathbf{A}_p^{\mathrm{t}}$; // $LN \log N$ flops
  $\boldsymbol{\Gamma}_{y_p} = \mathbf{A}_p \mathbf{B}_p$; // $\boldsymbol{\Gamma}_{y_p} = \mathbf{H}_p \mathbf{R} \boldsymbol{\Gamma}_s \mathbf{R} \mathbf{H}_p^{\mathrm{t}}$, $L^2 N$ flops
  $\boldsymbol{\Gamma}_{y_p} = \boldsymbol{\Gamma}_{y_p} + \sigma^2 \mathbb{1}_L$;
  $\mathbf{T}_p = \frac{1}{M} \boldsymbol{\Gamma}_{yp}^{-1} (\mathbf{V}_p \mathbf{V}_p^{\mathrm{t}})$; // $ML^2 + 2L^3$ flops
  $D_p(\boldsymbol{\rho}_k) = \frac{1}{2} \log |\boldsymbol{\Gamma}_{y_p}| + \frac{1}{2} \mathrm{Tr}(\mathbf{T}_p)$;
  $\boldsymbol{\Omega}_p = \boldsymbol{\Gamma}_{y_p}^{-1} \mathbf{H}_p$; // $L^2 N$ flops
  $\nabla D_p(\boldsymbol{\rho}_k) = \Big(\boldsymbol{\Omega}_p^{\mathrm{t}} \circ \big[\mathbf{B}_p (\mathbb{1}_L - \mathbf{T}_p) \big] \Big) \mathbf{1}_L - \frac{I_0}{M} \boldsymbol{\Omega}_p^{\mathrm{t}} (\mathbf{V}_p \mathbf{1}_L)$; // $L^2 N + 2LN$ flops
end
$D_{MP}(\boldsymbol{\rho}_k) = \sum_p D_p(\boldsymbol{\rho}_k) + \beta \|\boldsymbol{\rho}\|^2$;
$\nabla D_{MP}(\boldsymbol{\rho}_k) = \sum_p \nabla D_p(\boldsymbol{\rho}_k) + 2\beta \boldsymbol{\rho}$;

Proof. Let $\boldsymbol{\mu}_p$ and $\boldsymbol{\Gamma}_p$ denote the mean and the covariance of the collected data in the p-th patch. Indeed, one can extract $\boldsymbol{\mu}_p$ and $\boldsymbol{\Gamma}_p$ from $\boldsymbol{\mu}_y$ and $\boldsymbol{\Gamma}_y$, as a subvector and a submatrix, respectively. Moreover, we have $D_{MP}(\rho) = \frac{1}{M}\sum_{m=1}^{M}(\sum_p C_p(\boldsymbol{y}_{m,p}, \boldsymbol{\rho}))$ up to an additive constant, with

$$C_p(\boldsymbol{y}_p, \boldsymbol{\rho}) = \frac{1}{2}\sum_{p=1}^{P}(\boldsymbol{y}_p - \boldsymbol{\mu}_p)^t \boldsymbol{\Gamma}_p^{-1}(\boldsymbol{y}_p - \boldsymbol{\mu}_p)^t + \frac{1}{2}\left|\boldsymbol{\Gamma}_p\right|, \tag{4.7}$$

and

$$D_p(\boldsymbol{\rho}^*, \boldsymbol{\rho}) = \mathrm{E}\left[\sum_p C_p(\boldsymbol{y}_p, \boldsymbol{\rho})\right] = \sum_p D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{\mu}_p^*, \boldsymbol{\Gamma}_p^*)\|\mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Gamma}_p)). \tag{4.8}$$

It is clear from (4.8) that each $E[C_p]$ is minimal if and only if $\boldsymbol{\mu}_p = \boldsymbol{\mu}_p^*$ and $\boldsymbol{\Gamma}_p = \boldsymbol{\Gamma}_p^*$. Moreover, D_p is minimal when all $E[C_p]$ are minimal. If every pixel pair $\{u, v\}$ belongs at least to one patch, this is equivalent to $\boldsymbol{\mu}_y = \boldsymbol{\mu}_y^*$ and $\boldsymbol{\Gamma}_y = \boldsymbol{\Gamma}_y^*$, and we can terminate the proof as the one of Property 3.5. □

4.5 Simulation results

We now evaluate the patch-base approach on simulations performed under similar conditions as in Chapters 3. The same simulated object $\rho(r, \theta) \propto [1 + \cos(40\theta)]$ is used as shown in Figure 2.1. The point spread function h is given by Eq. (2.62). The PSF and the reconstructed objects were discretized with sampling step $\lambda/20$. Raw images were also simulated at the same sampling rate, unless otherwise stated.

4.5.1 Resolution under different patch sizes

First, we consider simulations under "theoretical" conditions (*i.e.,* an infinite number of speckle patterns and infinite SNR) to test the behavior of the patch-based marginal approach. The L-BFGS algorithm is used to find the minimizer of function (4.3). The simulations are run under different patch sizes with the same initialization, at the average of the measured images $\frac{1}{M}\sum_m \boldsymbol{y}_m$. The number of corrections stored in the limited memory matrix is 20. Termination tolerances are adopted when either of the following conditions is met:

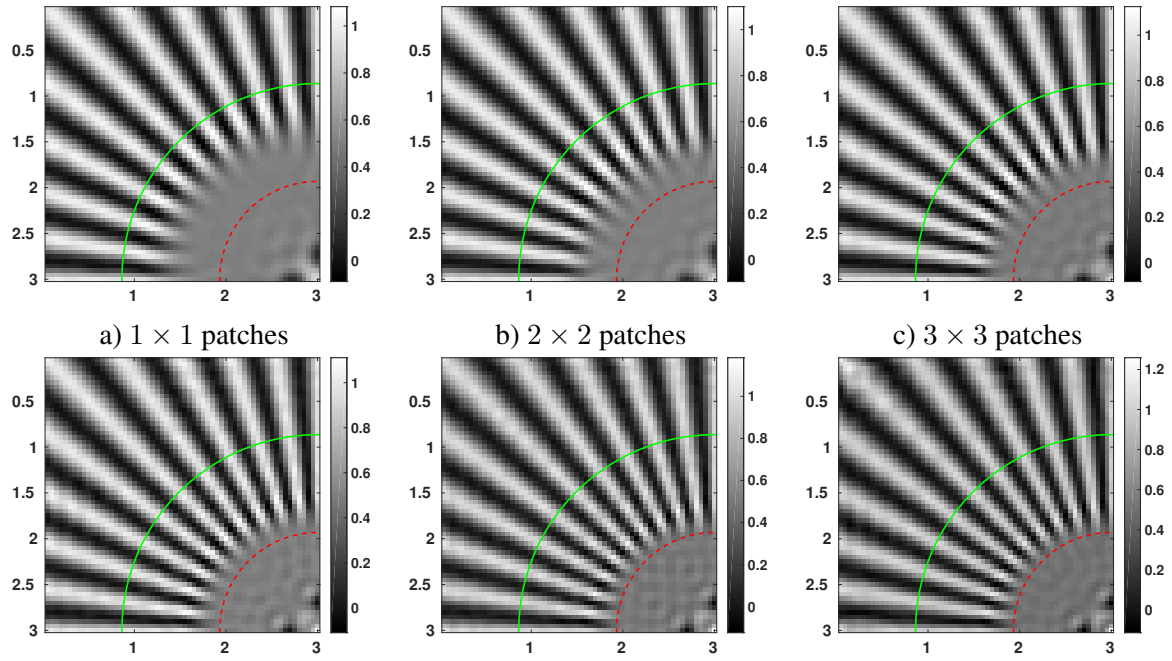
— The relative descends $\left|\frac{D_{MP}(\boldsymbol{\rho}_{k+1}) - D_{MP}(\boldsymbol{\rho}_k)}{D_{MP}(\boldsymbol{\rho}_k)}\right| \leq 10^{-8}$

— The ℓ_∞-norm of the gradient $\|\nabla D_{MP}(\boldsymbol{\rho}_k)\|_\infty \leq 10^{-5}$.

The reconstructed objects are shown in Figure 4.1, and the corresponding normalized RAPS curves (definition given in (2.63)) for the error images and the MCF (defined by (2.64)) are shown in Figure 4.2.

It shows that even in the extreme case with one-pixel patches, i.e. only only the diagonal of the covariance $\boldsymbol{\Gamma}_y$ is considere, we still retrieve some super-resolution. As the patch size increases, the SR effect increases fast and becomes as good as the original marginal estimator when the patch size reaches 5×5.

The time elapsed using different patch sizes for a 80×80 object on a normal computer with intel CORE i5 CPU implemented by Matlab is plotted in Fig. 4.3. We see that except when the patch size is very small (due to the inefficienct for-loop processing in Matlab and more $\mathbf{H}_p$ need to be generated), the time elapsed for one iteration increases almost linearly as the patch size grows.



a) 1×1 patches b) 2×2 patches c) 3×3 patches

d) 5×5 patches e) 15×15 patches f) Without patch (full image)

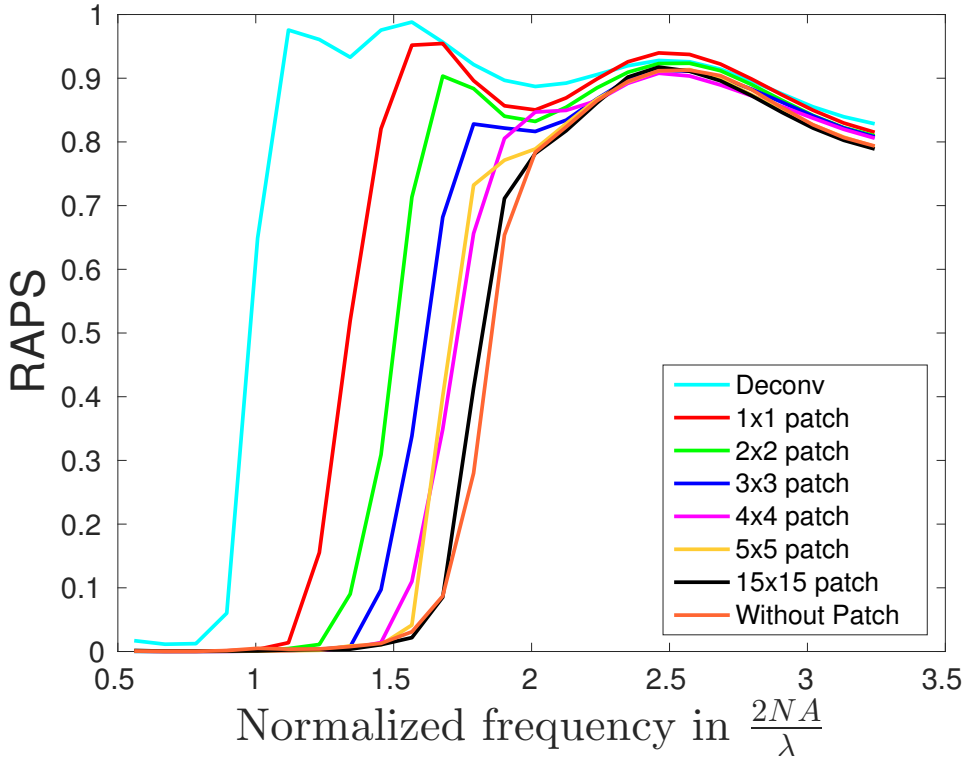
Figure 4.1 – **Reconstructed objects under theoretical conditions (infinite number of speckle patterns and infinite SNR) with different patch sizes**. Graduations in x and y are in wavelength λ, and the frequency in green circle (red circle) corresponds to cut-off frequency of OTF (twice the cut-off frequency, respectively).

4.5.2 Positivity constraint

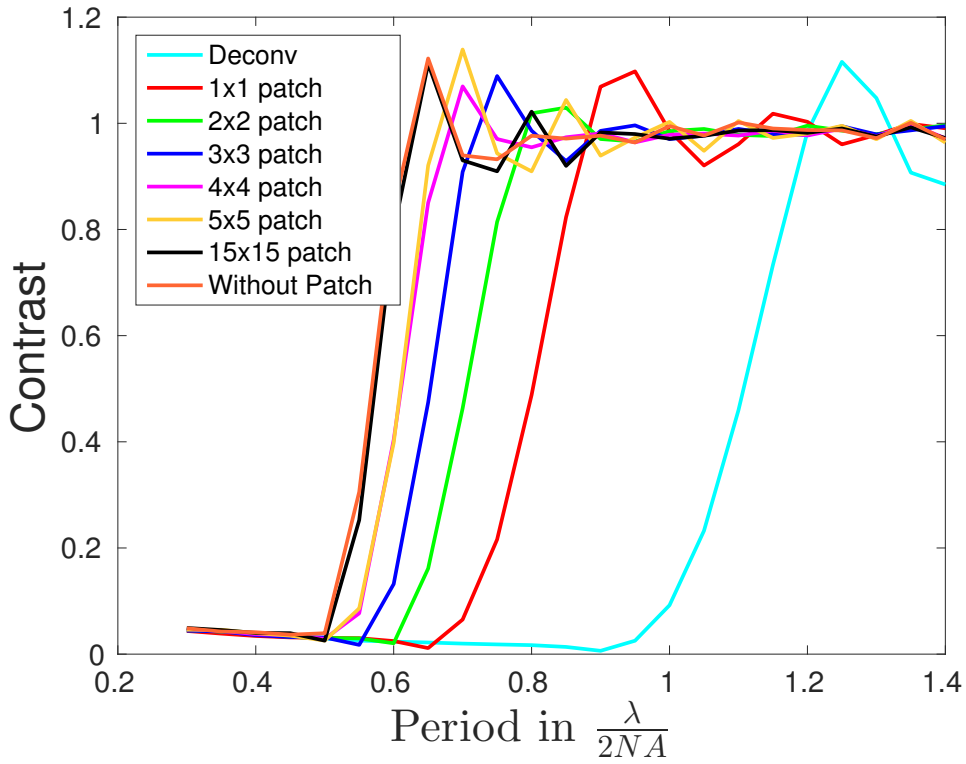
From Figure 3.1, we have known that the positivity constraint cannot further improve the super-resolution capacity of marginal estimator. Since the super-resolution of the patch-based marginal approach can not reach twice the resolution limit when the patch size is small, we explore the influence of the positivity constraint on patch-based marginal extimator by adding the constraint $\boldsymbol{\rho} \geq 0$ and then solve the minimization problem of Eq. (4.3) with the L-BFGS-B algorithm (see Appendix D.2). The simulation results performed under theoretical conditions are shown in Figure 4.4. Compared with the results shown in Figure 4.1, we conclude that the positivity constraint can not further improve the super-resolution capacity of patch based marginal estimator under same patch sizes.

4.5.3 Overlapping patches

The former simulations were performed with adjacent square patches. When the patch size is small (smaller than 3×3 for example), we have sacrificed partial super-resolution information to accelerate the computational time and reduce the memory required. If the patches overlapped with each other, equivalently more information in covariance matrix $\boldsymbol{\Gamma}_y$ is used and thus we have the possibility to obtain better super-resolution. Since we know that the overlapping strategy can not give more informatin than the marginal estimator without patch, i.e. it can not surpass more than twice the resolution limit, therefore we focus on overlapping small patches in simulations. Simulation results under asymptotic conditions with overlapping patches are shown in Figure 4.5. Each pair neighbour patches (left, right, up, down) overlap with each other by one column or one line. Compared with the results shown in Figure 4.1, it shows that the overlap patch strategy does not improve the super-resolution capacity compared with a

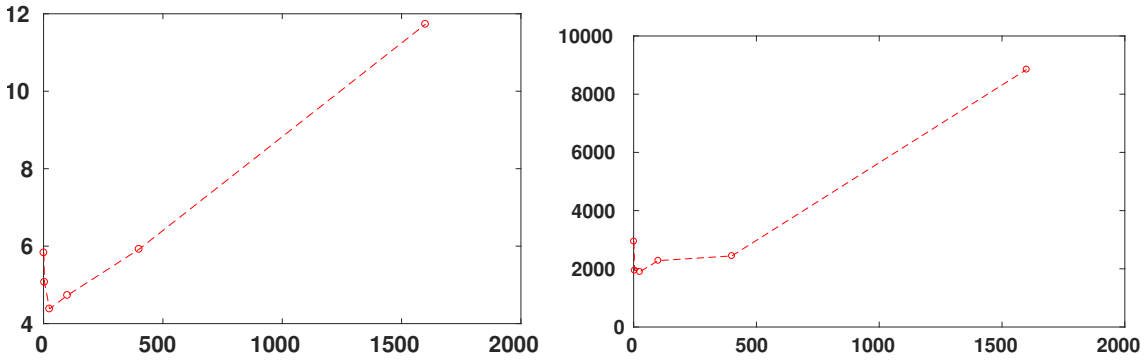


(a) Normalized RAPS curves for error images with different patch sizes.



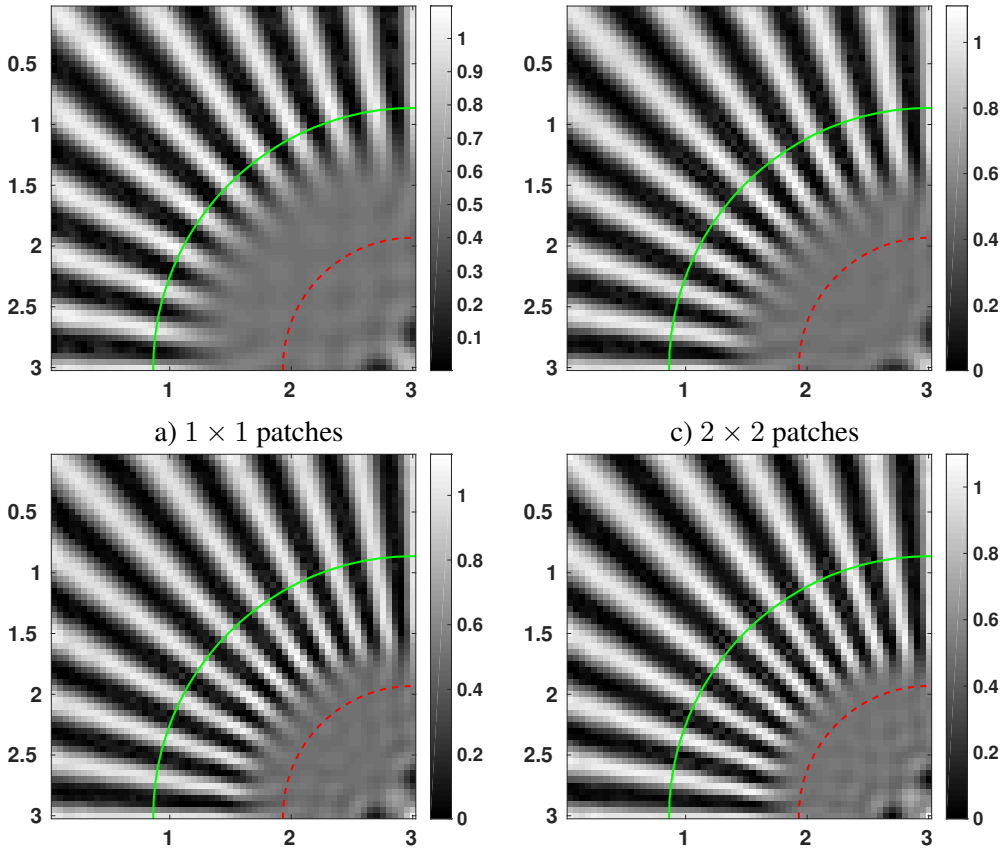
(b) Contrast function of reconstructed object with different patch sizes.

Figure 4.2 – **The RAPS curves and contrast function of reconstructed object with different patch sizes under theoretical condition (infinite number of speckle patterns and infinite SNR).**



(a) Time elapsed per iteration (b) Total time elapsed

Figure 4.3 – **The time elapsed for different Patch sizes for a** 80×80 **object.** X-axis represent number of pixels in one patch, and Y-axis represents time with units seconds.

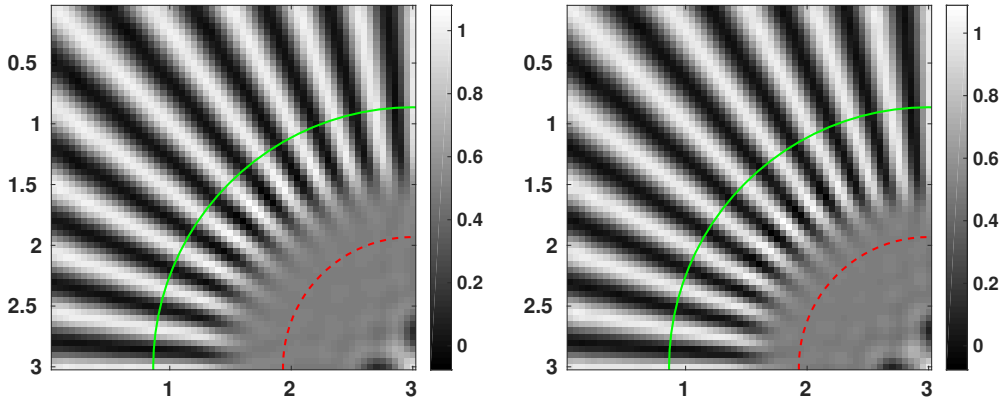


a) 1×1 patches c) 2×2 patches

d) 3×3 patches e) 5×5 patches

Figure 4.4 – **Reconstructed objects under theoretical conditions and different patch sizes (similar to Figure 4.1), by incorporating the positivity constraint**.

non-overlap adjacent patches under same patch sizes.



a) Overlapping 2×2 patches b) Overlapping 3×3 patches

Figure 4.5 – **Reconstructed objects under theoretical conditions with overlapping patches**. The overlapping area of neighbour patches (left, right, up, down) is one column or one line.

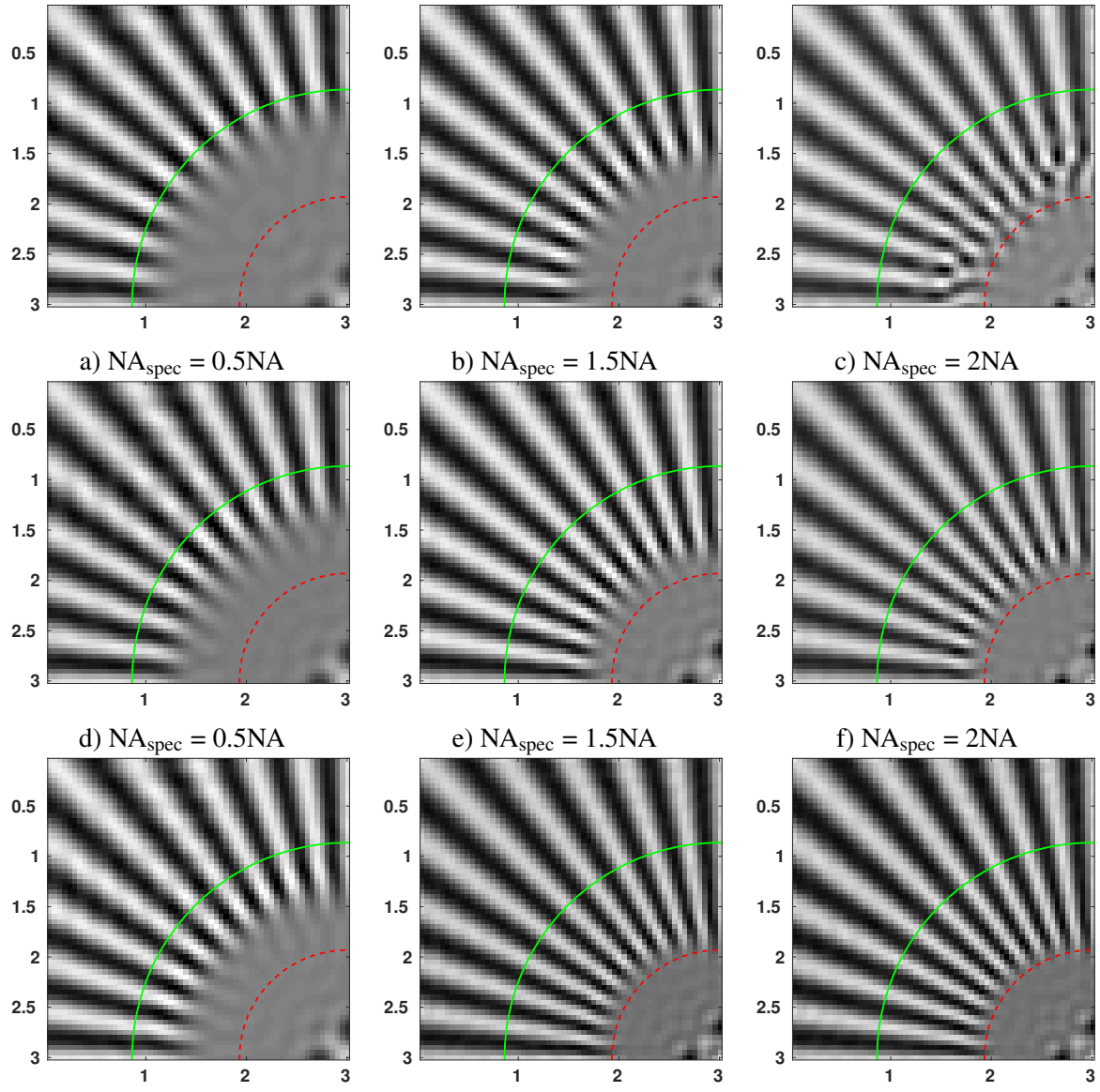
4.5.4 Resolution as a function of the numerical aperture

In realistic experimental conditions, the support of spectral density of speckle patterns may differ from the frequency support of PSF due to the Stokes-shift effect in fluorescence [Alb11], or in total internal reflection fluorescence microscopy (TIRF). We can simulate this effect by generating the speckle patterns with different numerical aperture, and the corresponding reconstructed objects are shown in Figure 4.6. Clearly, the resolution improves as the support of the speckle spectral density increases when the patch size is small (e.g. pixel and 2×2 patches). The normalized RAPS and MCF curves of the reconstructed objects (Figure 4.7) show that we can retrieve not only $2\nu_{\text{spec}}$ when $\nu_{\text{spec}} < \nu_{\text{PSF}}$ as demonstrated in Chapter 3, but also part of information between $2\nu_{\text{spec}}$ and $\nu_{\text{spec}} + \nu_{\text{PSF}}$. For example, when $\nu_{\text{spec}} = 0\ 5\nu_{\text{PSF}}$, we can always retrieve some high frequency information higher than $2\nu_{\text{spec}} = \nu_{\text{spec}} = \frac{2NA}{\lambda}$. And when $\nu_{\text{spec}} > \nu_{\text{PSF}}$, the resolution beyond $2\nu_{\text{PSF}}$ is still inaccessible, unlike the nonlinear-SIM which has theoretically unlimited resolution [Gus05].

In conclusion, enlarging the support of spectral density of the speckle patterns allow us to retrieve good super-resolution with very small patch sizes (e.g. 2×2), however, we do not observe nonlinear-SIM effect in blind-speckle-SIM using marginal approach.

4.5.5 Resolution as a function of the sampling step

In the previous simulations, the sampling step of the measured images $\boldsymbol{y}_m$ was set to the same value as the oen of the object, that is, $\frac{\lambda}{20}$. Since we know that there are no useful information in frequencies outside the support of OTF in data, the sampling length finer than Nyquist rate $\frac{\lambda}{4NA} \approx 0.168\ \lambda$ is enough to assure that we do not lose useful information in the raw images. The simulation results performed under asymptotic conditions (infinite number of speckle patterns and infinite SNR) with different sampling lengthes of the raw images are shown in Figure 4.8. It shows that even the sampling rate of raw images is lower than the Nyquist rate, for example in the second column in Figure 4.8, the marginal approach retrieves super-resolution information as good as the simulation results performed with raw images sampled at a finer rate than the Nyquist rate.

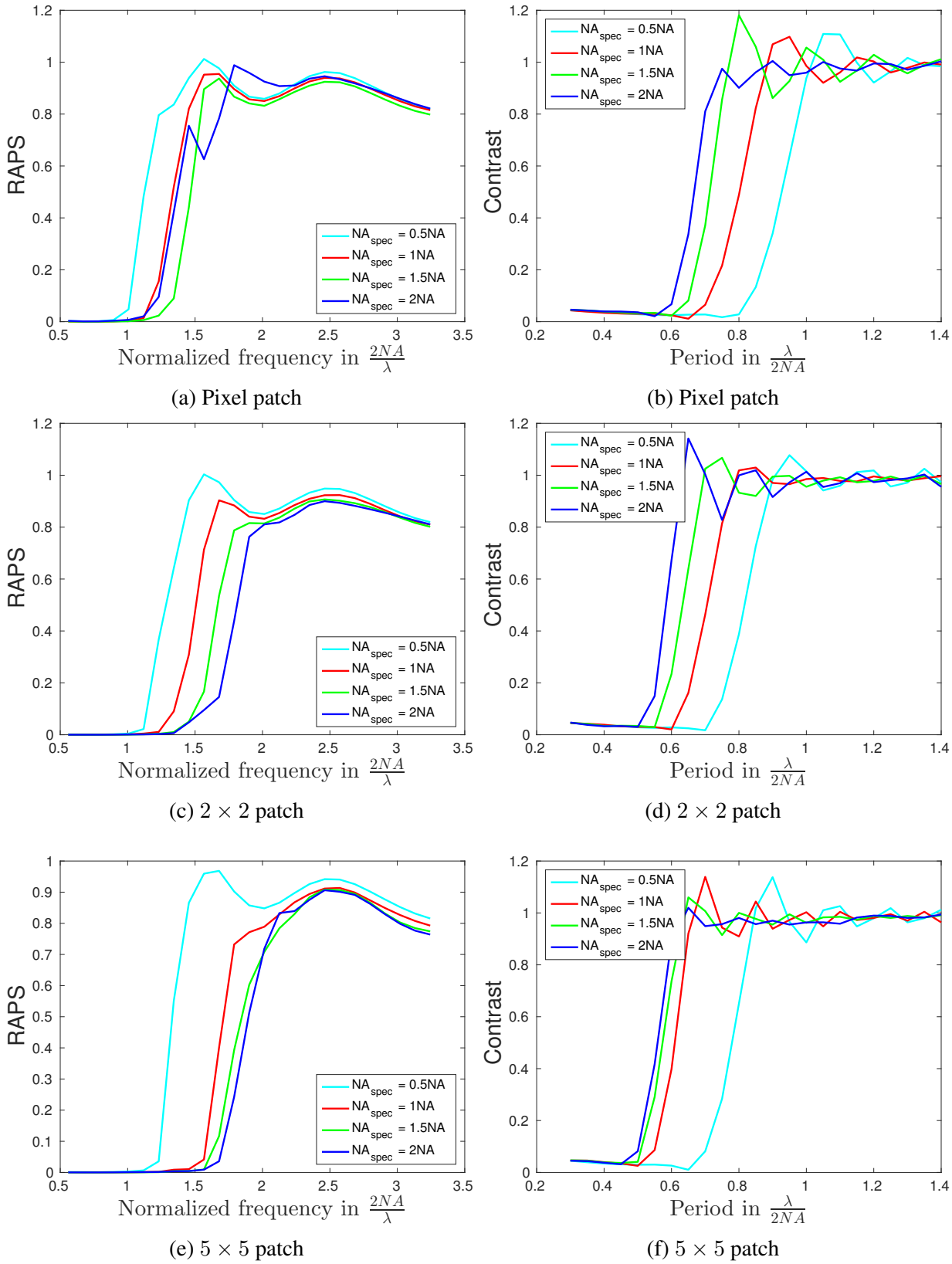


a) NA_{spec} = 0.5NA b) NA_{spec} = 1.5NA c) NA_{spec} = 2NA

d) NA_{spec} = 0.5NA e) NA_{spec} = 1.5NA f) NA_{spec} = 2NA

g) NA_{spec} = 0.5NA h) NA_{spec} = 1.5NA i) NA_{spec} = 2NA

Figure 4.6 – **Reconstructed objects under theoretical conditions and varying support of speckle patterns**. The first row corresponds to use 1-pixel patches, the second row is obtained with 2×2 patches and the last row with 5×5 patches.



(a) Pixel patch (b) Pixel patch

(c) 2×2 patch (d) 2×2 patch

(e) 5×5 patch (f) 5×5 patch

Figure 4.7 – **Normalized RAPS curves for error images as a function of the frequency (first column) and MCF of the reconstructed objects as a function of the period (second column) with varying support of speckle patterns**.

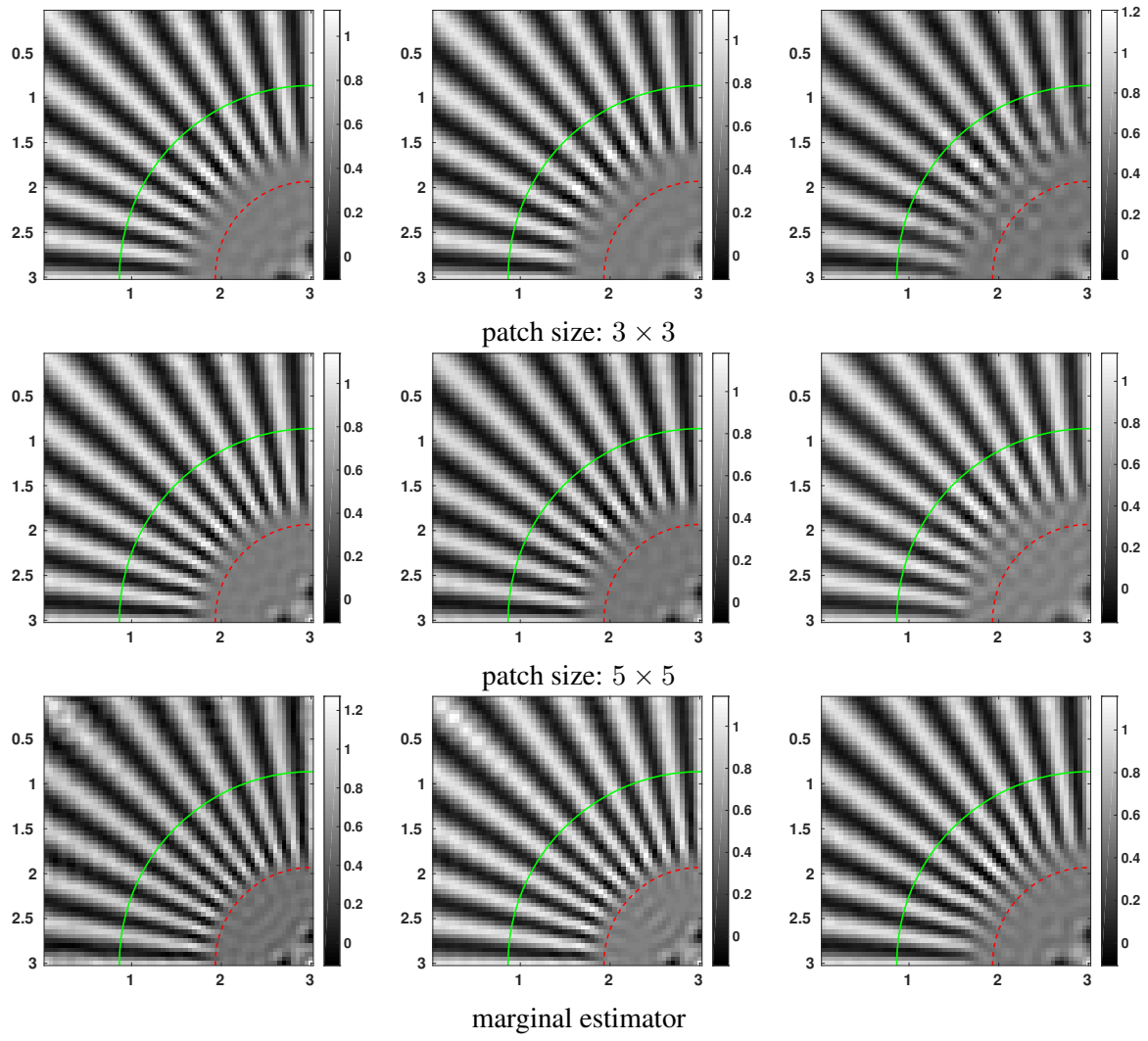


Figure 4.8 – **Reconstructed objects under asymptotic conditions with different image sampling rates**. The image sampling step is 0.15 λ in the first column, 0.2 λ in the second column and 0.25 λ in the third column. The Nyquist limit is 0.168 λ.

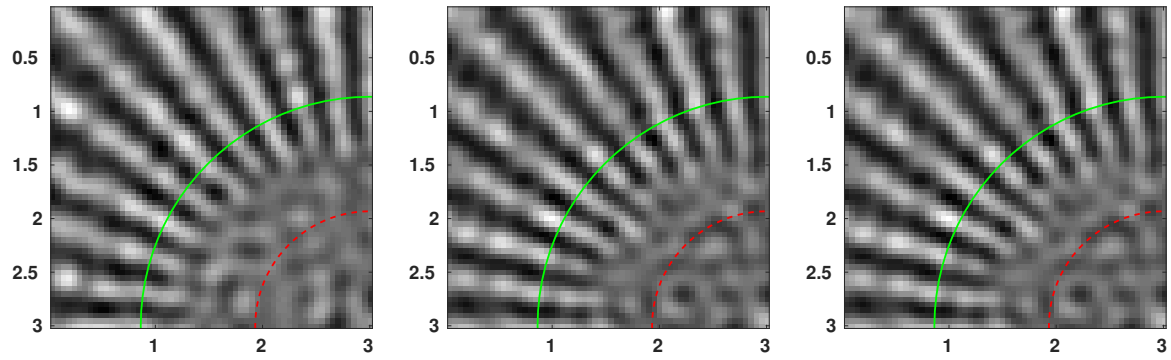
4.5.6 Simulations under realistic imaging conditions

In this section, we simulate realistic imaging conditions by applying Poisson noise. The total photon budget detected on one Nyquist pixel (pixel size corresponds to $\lambda/4NA$) is fixed, denoting as N_{tp}. As the number of speckle patterns increases, the photons per Nyguist pixel per measurement will be reduced. This can be achieved by reducing the exposure time or by lowering the laser energy. The electronic noise is assumed to be 3-photon standard deviation per pixel and per measurement. The noise level in this case is $3M/N_{tp}$, corresponding to SNR = 20dB when 100 speckle patterns are used for 30 000 photons per Nyguist pixel. In our simulations the regularizer parameter β is set to 10^{-6}. Optimization is performed by the L-BFGS algorithm with same configuration as in previous problems. The reconstructed objects under different number of speckle patterns are shown in Figure 4.9, and the RAPS and MCF curves of the reconstructed objects are ploted in Figure 4.10. We see that once the total number of photons per Nyquist pixel is fixed, increasing the number of speckle patterns cannot improve the resolution, since the SNR decreases at the same time.

4.6 Conclusion

In this chapter, we introduced a patch-based marginal estimator in order to reduce the computational complexity and memory requirements of the marginal approach proposed in Chapter 3. By cutting the each image into a set of patches and neglecting the correlation information of pixels from different patches, we are able to retrieve a super-resolution image with computational complexity $\mathcal{O}(N^2(\log N + L))$ with L the number of pixels in each patch, instead of $\mathcal{O}(N^3)$ while the memory requirement reduced from $\mathcal{O}(N^2)$ to $\mathcal{O}(NL)$. Moreover, the patch-based method enables most computations to be parallelized. We also explored the influence of the patch size, positivity constraint, overlapping strategy, correlation of speckle patterns and the sampling rate of raw images on the marginal estimator.

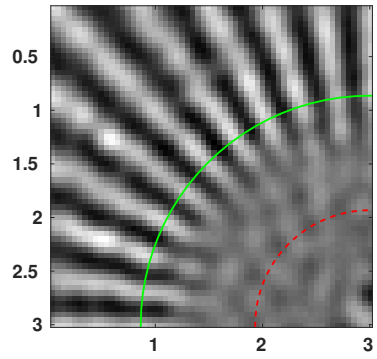
Our simulations show that the positivity constraint and the overlapping strategy cannot enforce the super-resolution capacity of the patch-based marginal estimator for a given patch size. When the patch size is small, increasing the support of the speckle spectral density function could help us retrieve better super-resolution information. Finally, even if the data sampling rate is lower than the Nyquist limit, we could still reconstruct the object with a super-resolution factor of two using the marginal approach.



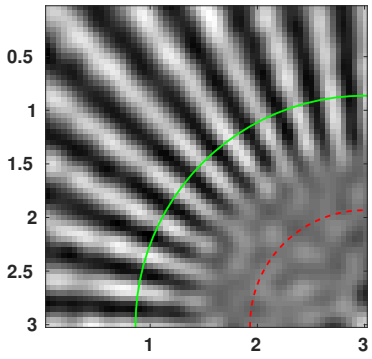
a) 100 speckle
SNR_e = 15.2 dB

b) 100 speckle
SNR_e = 20 dB

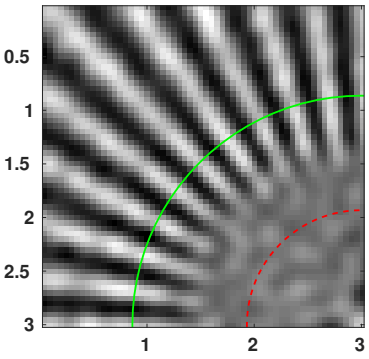
c) 100 speckle
SNR_e = 30 dB



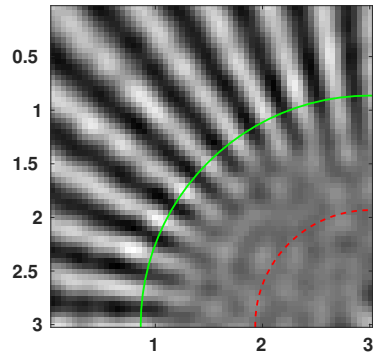
d) 300 speckle
SNR_e = 10.4 dB



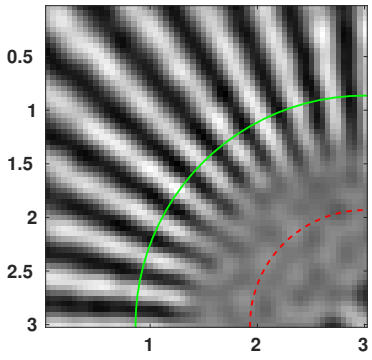
e) 300 speckle
SNR_e = 15.2 dB



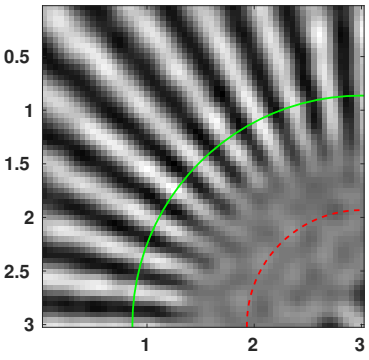
f) 300 speckle
SNR_e = 25 dB



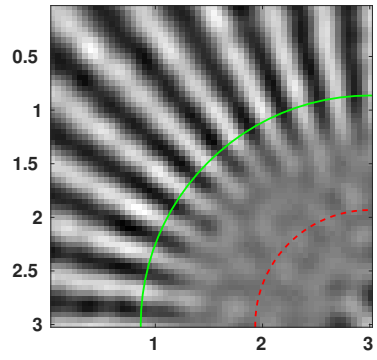
g) 500 speckle
SNR_e = 8.2 dB



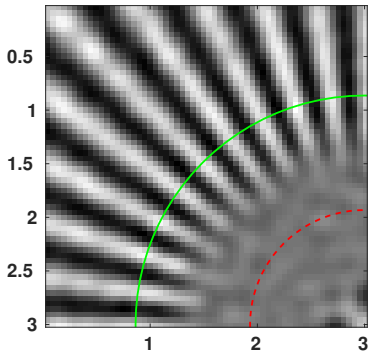
h) 500 speckle
SNR_e = 13 dB



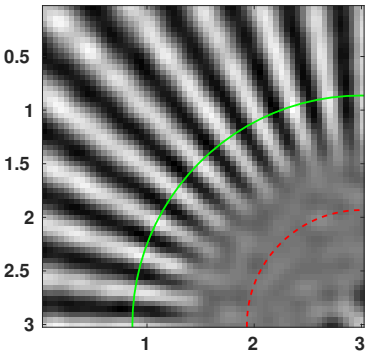
i) 500 speckle
SNR_e = 23 dB



j) 1000 speckle
SNR_e = 5.2 dB

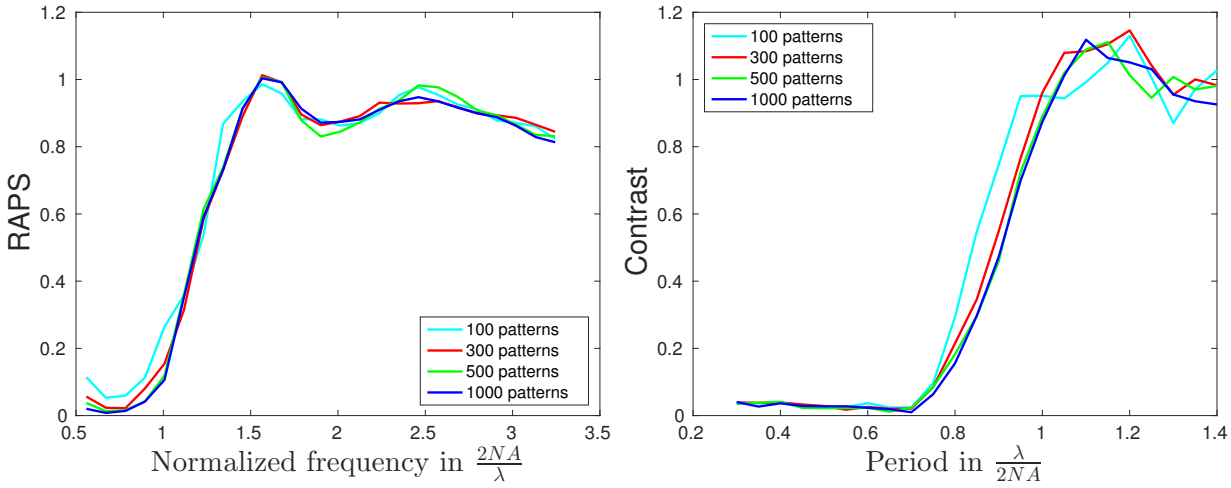


k) 1000 speckle
SNR_e = 10 dB

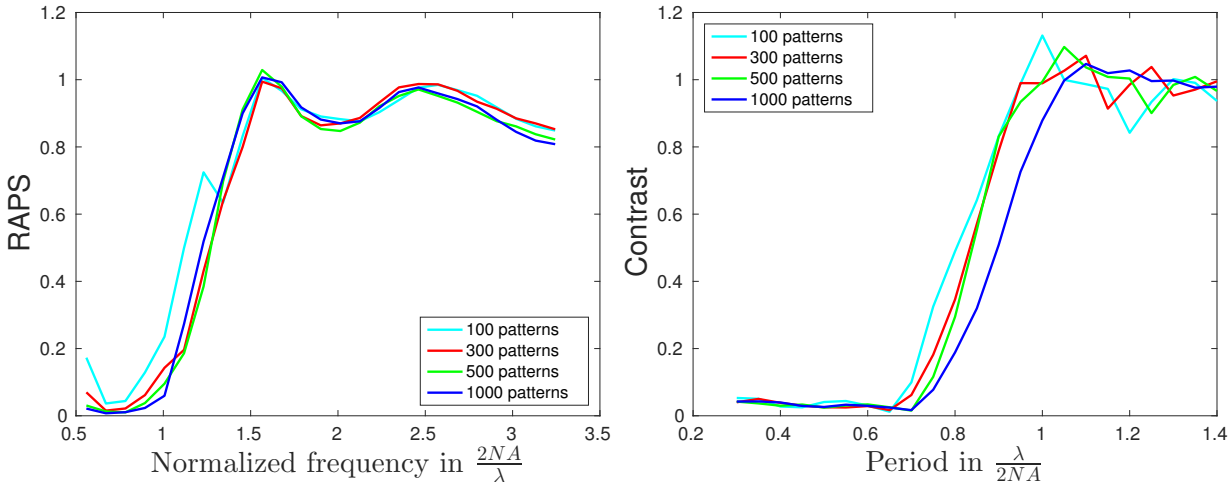


l) 1000 speckle
SNR_e = 20 dB

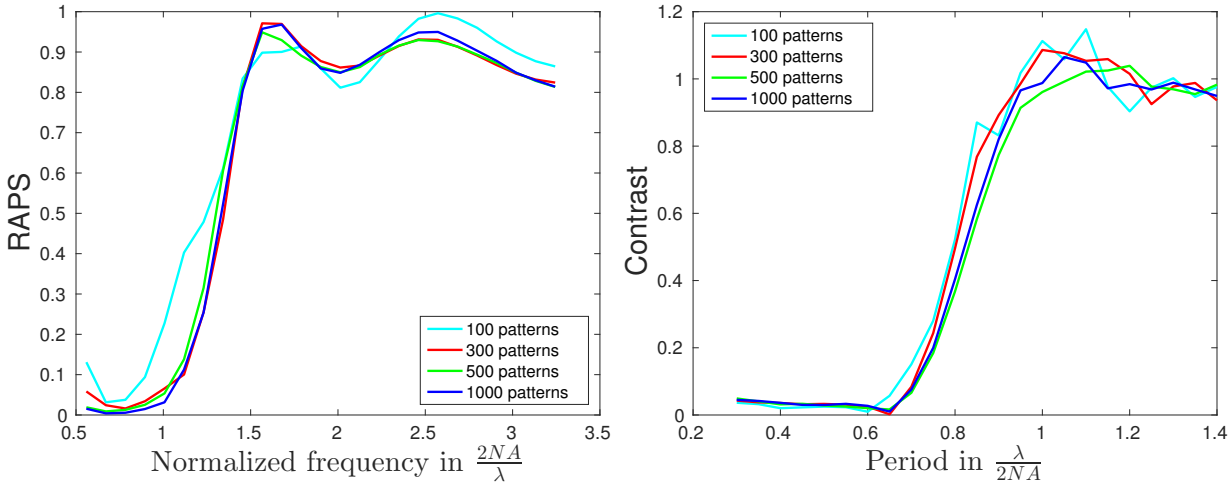
Figure 4.9 – **Reconstructed objects with Poisson statistics and Gaussian noise with different numbers of speckle patterns, with 5×5 patches**. The first column corresponds to a total photon budget of 10^4 photons per pixel on average (3×10^4 and 3×10^5 for the second column and third column, respectively). Electronic noise with 3-photon standard deviation per pixel and per measurement is added. SNR_e denotes the signal-to-noise ratio of electronic noise.



(a) Total photon budget of 10^4 per pixel on average



(b) Total photon budget of 3×10^4 per pixel on average



(c) Total photon budget of 3×10^5 per pixel on average

Figure 4.10 – **Normalized RAPS curves for error images as a function of the frequency (first column) and the modulation contrast of the reconstructed objects as a function of the period (second column) for results shown in Figure 4.9**.

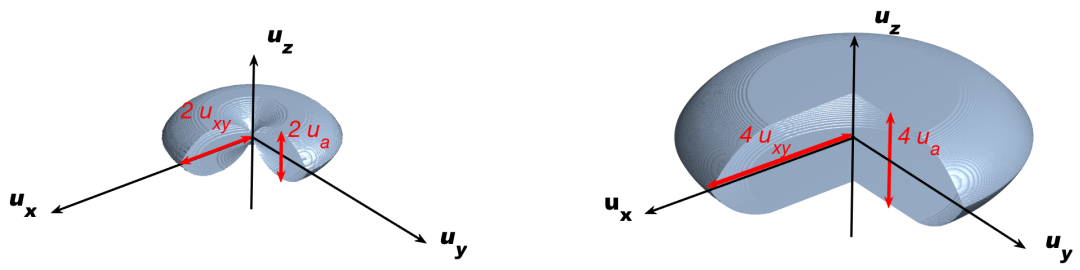
Chapter 5

Optical sectioning and super-resolution in 3D imaging

5.1 Introduction

The super-resolution methods presented in previous chapters are 2D imaging techniques restricted to thin objects. For thick objects, the targeted in-focus image is always blurred with the out-of-focus information. In frequency domain, the optical transfer function (OTF) in axial direction $\boldsymbol{u}_z$ is highly dependent on lateral frequencies $\boldsymbol{u}_\perp$, as shown in Figure 5.1(a). In particular, for a uniform plane object, the intensity spectrum contains only zero frequency component. Because $\tilde{h}(\boldsymbol{u}_z) = 0$ when $\boldsymbol{u}_\perp = 0$, the axial structure is lost during the imaging process. In other words, a uniform plane can not be axially resolved in conventional microscopy. This is called the "missing cone" problem [Str84]. Optical sectioning is defined as the capacity to provide axial resolution for all lateral frequencies within the OTF support to fulfill the "missing cone".

In section 5.2 we introduce the existing 3D super-resolution techniques. Simulation results in Section 5.3 demonstrate the 3D sectioning ability using unknown speckle patterns by constrained $\ell_{p,q}$ norm minimization and the patch based marginal approach.



(a) OTF for h in 3D imaging

(b) OTF for h^2 in 3D imaging

Figure 5.1 – **The Fourier transform for h and h^2 in 3D imaging.** Figure courtesy in [ILA+17].

5.2 Existing super-resolution 3D imaging techniques

5.2.1 Confocal microscopy

The principle of confocal microscopy to achieve optical sectioning is by illumination-beam scanning and image detection through a pinhole, as shown in Fig. 5.2. Unlike the widefield illumination technique, the sample is illuminated in confocal microscopy with a tightly focused light beam. Correspondingly, only the signal emitted from the focused illuminated region is detected. The illuminated 3D volume is called a probe volume and a pinhole is put in front of the detector to reject the our-of-focus background. To obtain the whole 3D image of the sample, either the probe volume is scanned in three dimension while keep the sample fixed, or the probe volume is held fixed while the sample is scanned in three dimension. Then the final 3D image is synthesized from the detected signals.

One requirement of the light source in confocal microscopy is that the light should be spatially coherent. This can be achieved either by using a laser or adding a pinhole after the lamp beam. In fluorescence confocal microscopy, the equivalent PSF is h^2 under the infinitely small pinhole assumption where h denotes the PSF of wide-field microscopy [Mer10]. The difference between the frequency support of h and h^2 is shown in Figure 5.1. For more thorough discussions about confocal microscopy, please refer to [KC96][Dia01].

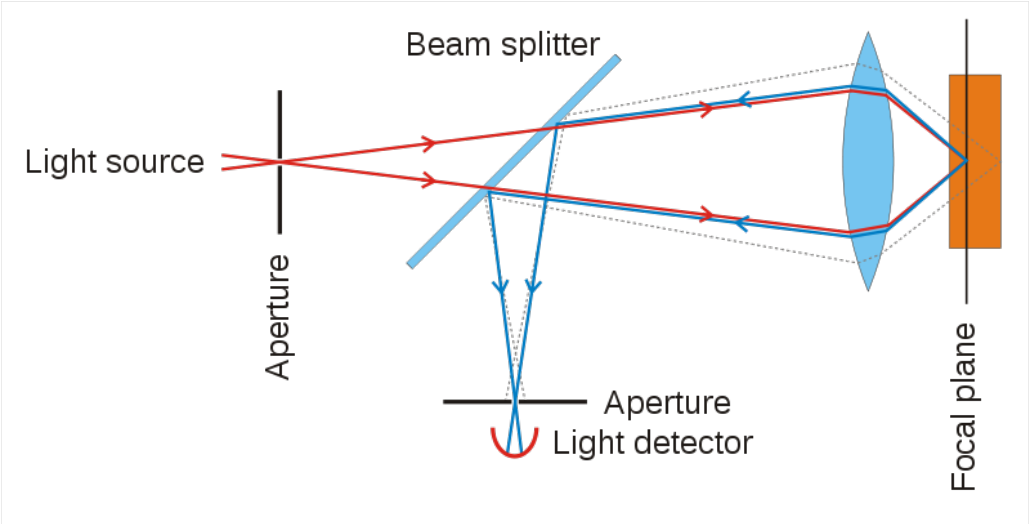


Figure 5.2 – The geometries for confocal microscopy. (Figure courtesy of wikipedia.)

5.2.2 Two-photon microscopy

Two-photon fluorescence microscopy takes advantage of the fluorescence property that a fluorophore is excited by simultaneous absorption of two photons. It is an alternative optical sectioning technique in 3D imaging. Similar as the confocal microscopy, the 3D image is synthesized by scanning the two-photon excitation volume throughout the sample. Since two-photon microscopy depends on the nonlinear fluorescence excitation, the size of the scanning probe volume in two-photon microscopy is governed only by the illumination optics, while in confocal microscopy it is governed by both illumination and detection optics.

One typical optics geometry of two-photon microscopy is shown in Fig. 5.3, compared with that of confocal microscopy, the pinhole in front of the detector is disappeared. The optical sectioning capacity of two-photon microscopy is identical to a confocal microscopy, with an

equivalent PSF of h^2. For a more detailed description of two-photon microscopy, please see [Dia01],[XW02].

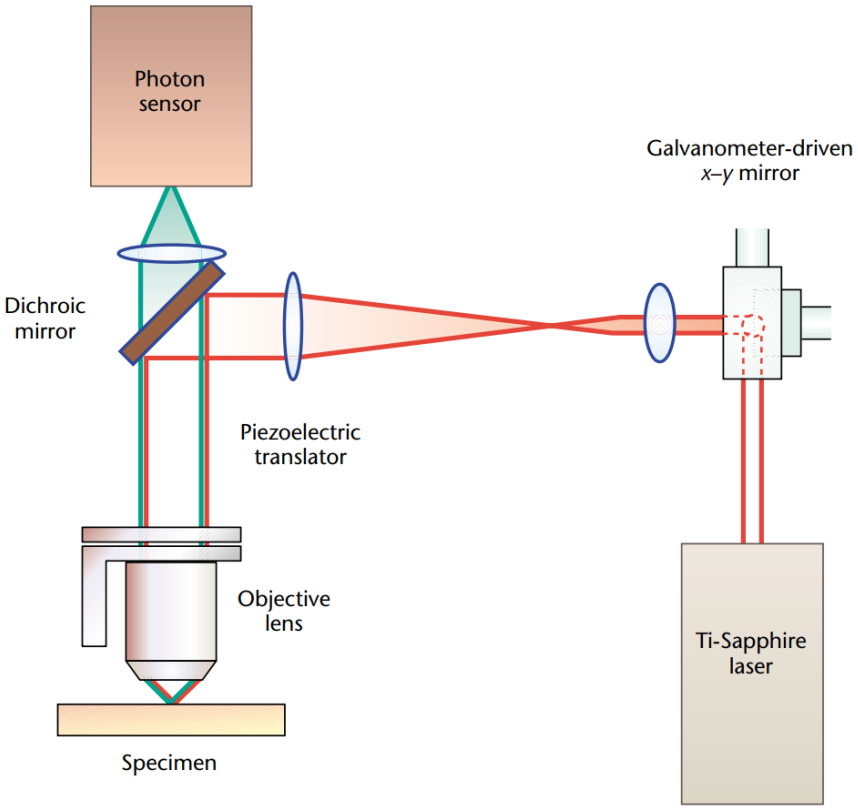


Figure 5.3 – The geometry for typical two-photon microscopy. (Figure courtesy of Peter TC So, Massachusetts Institute of Technology.)

5.2.3 3D STED

STED technique can be applied to 3D imaging with a 4Pi illumination geometry, where the sample is placed in the common focus of two opposing lenses [DH02]. 33 nm axial resolution was obtained in STED-4π microscopy, however, the lateral resolution does not improve in STED-4π. STED microscopy with super-resolution in all three dimensions are proposed in [WMKH09]. By introducing a two phase plate, a PSF with reduced length in all three dimensions is obtained.

5.2.4 3D STORM/PALM

5.2.4.1 Astigmatism-based 3D STORM

The super-resolution 3D STORM was introduced in [HWBZ08] by adding a weak cylindrical lens into the imaging path as shown in figure 5.4. As a result, the PSF has an ellipticity shape and its orientation varies as the position changes along the axial direction. By fitting the measured image with an elliptical Gaussian function, we can locate the position of fluorophore in lateral and axial coordinates. Since the axial location is based on the shape of the PSF, the axial scanning is unnecessary in 3D STORM. One drawback of astigmatism-based axial location is that the lateral resolution varies as a function of depth due to the changes of the lateral shape in astigmatic PSF. A similar 3D super-resolution technique was presented in [PTB+09]

where a double-helix PSF (DH-PSF) is generated through a spatial light modulator (SLM). The individual fluorophore is identified by the angular orientation of the 2 lobes of the DH-PSF, as shown in figure 5.5. The DH-PSF has deeper imaging depth than astigmatism-based 3D STORM, which can be applied to thick samples $> 2\,\mu$m with a resolution $10\,$nm laterally and $> 20\,$nm axially.

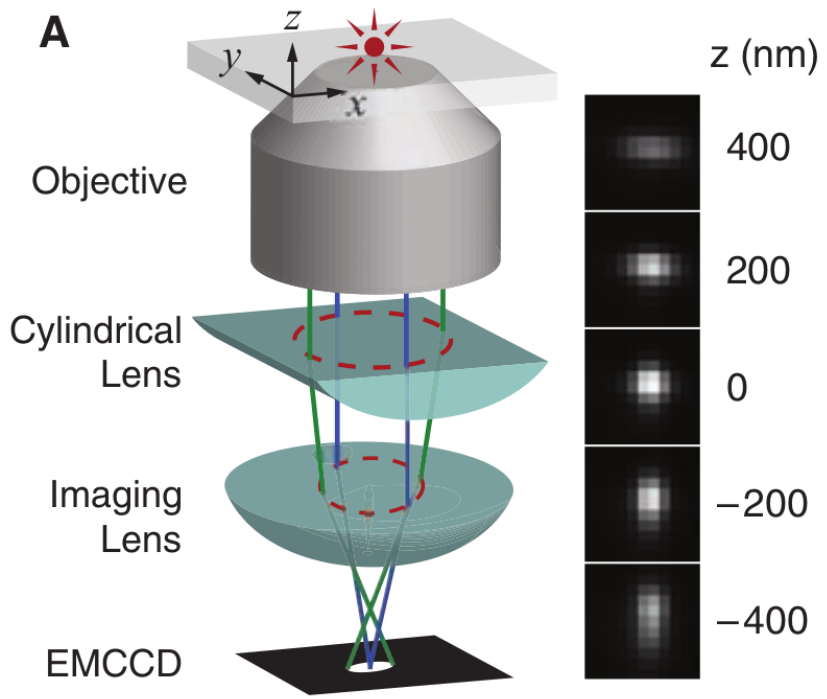


Figure 5.4 – The geometries of 3D STORM. (Figure courtesy of [HWBZ08])

5.2.4.2 Biplane FPALM

Another 3D super-resolution imaging technique is called biplane (BP) FPALM [JGL+08]. Two axially separated planes are simultaneously detected by a 50:50 beam splitter to accelerate the imaging process. Once the 3D data stack is obtained, the particles are identified in axial coordinate by iteratively searching for the brightest pixels and eliminating regions within a fixed distance. The lateral resolution does not change as the image depth varies in BP FPALM scheme.

5.2.5 3D SIM

The wide-field fluorescence microscopy using speckle patterns proposed in [Wal01][JW04] has an optical sectioning ability in theory as good as the confocal microscopy by recording a sequence of reference speckle patterns I_m together with the images of speckle-illuminated object y_m. Taking advantage of the equality :

$$\mathrm{Cov}(y_m, I_m) = h^2 * \rho \tag{5.1}$$

it is possible to retrieve the object with the equivalent point spread function h^2, as good as the ideal confocal microscopy.

SIM using harmonic patterns has been demonstrated to be able to enhance the spatial resolution in all three dimensions [GSC+08]. Although fewer illumination patterns are used

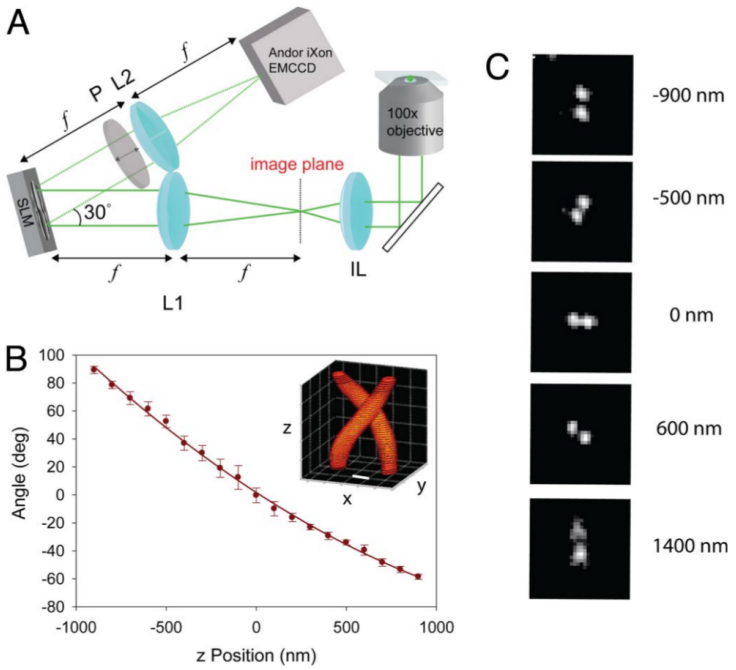


Figure 5.5 – DH-PSF imaging system. (Figure courtesy of [PTB+09])

in [GSC+08] than speckle-SIM, the illumination patterns are not maintained fixed in relation to the object in the imaging process, making the design of the devices quite complex. Moreover, this technique and the speckle-SIM proposed in [Wal01] could not be applied in scattering media, just as the standard 2D-SIM.

Recently, a 3D super-resolution technique using saturated-speckle patterns is proposed in [PGK+17] by illuminating the object with saturated speckle patterns. Super-resolution is observed after a direct deconvolution of the recorded data with the 3D saturated speckle PSF (hI_s).

A technique called dynamic speckle illumination microscopy (DSI) produce optical sectioning with totally unknown speckle patterns [VM05][VM06][VHM07][LCM08][LFCM11]. The basic idea behind this technique is that the changing of the speckle patterns will cause a large variation of the in-focus light but a small variation in the out-of-focus light. So the variation in a sequence of images y_m preferentially extracts the information from the focal plane.

5.3 3D super-resolution in blind-speckle-SIM

To check the optical sectioning ability of our proposed methods in blind-speckle-SIM, we use the same 2D star-like object and put it in the $y = 0$ plane. The fluorescence density is given by:

$$\rho(x, y, z) \propto [1 + \cos(40\theta)]\delta(y) \tag{5.2}$$

with $\tan\theta = z/x$. The normalized bandwidth limit of h is set to $\frac{f_{\max}}{f_S/2} = 0.5$ in lateral direction and 0.25 in axial direction, with $f_{\max}$ the cut-off frequency of the microscopy and f_S the spatial sampling frequency. The PSF and OTF used in simulation are shown in Fig. 5.7 and the reconstruction results by constrained $\ell_{p,q}$ norm minimization and patch based marginal estimator are displayed in figure 5.8. We can see that the resolution in both the lateral and axial direction are improved compared with the average wide-field image $\bar{\boldsymbol{y}}$ and its deconvolution. The ideal confocal image is simulated by convolving the actual fluorescence distribution of the sample with the square of the point spread function h^2 and then deconvolve it.

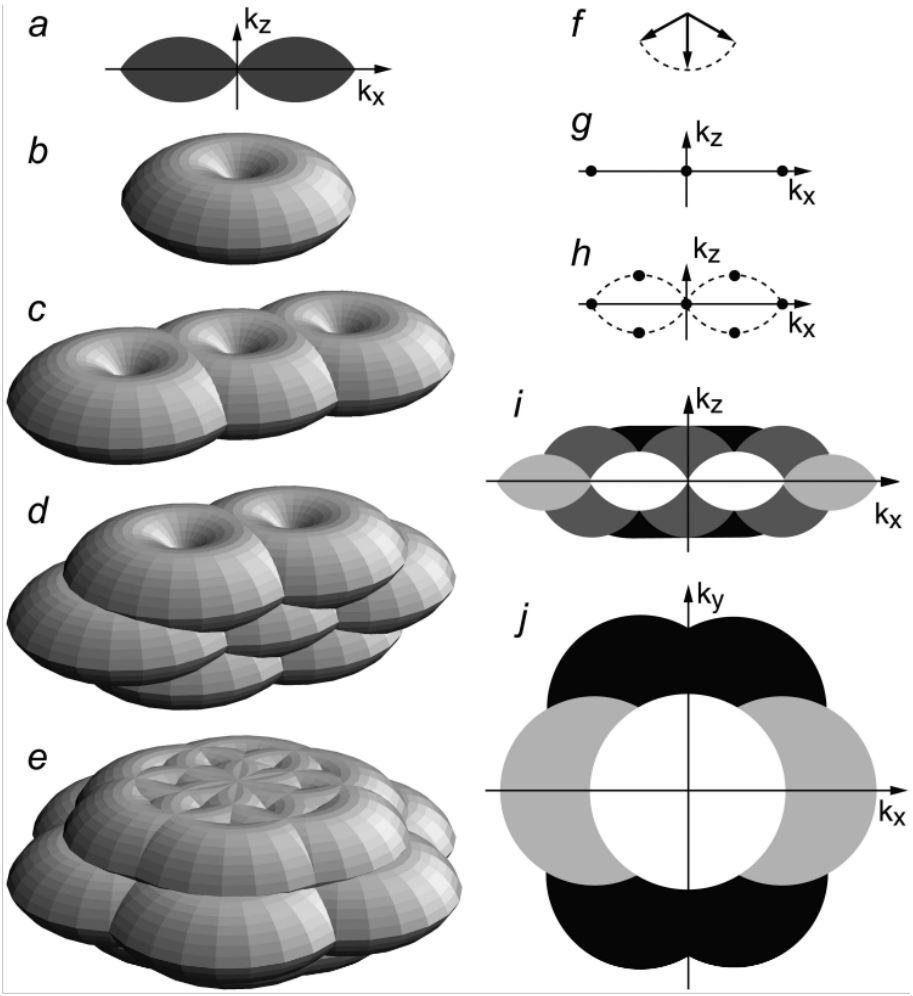
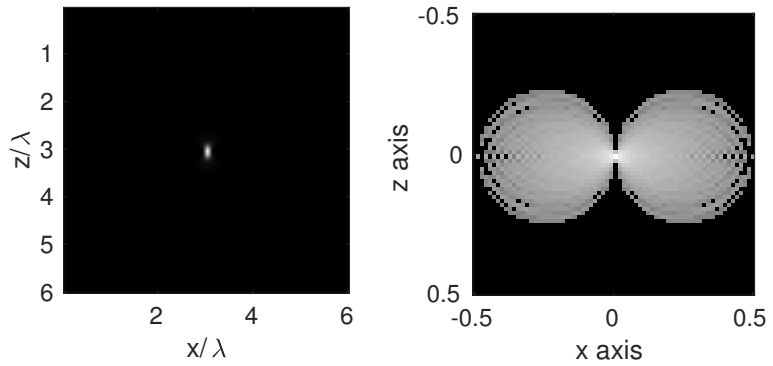


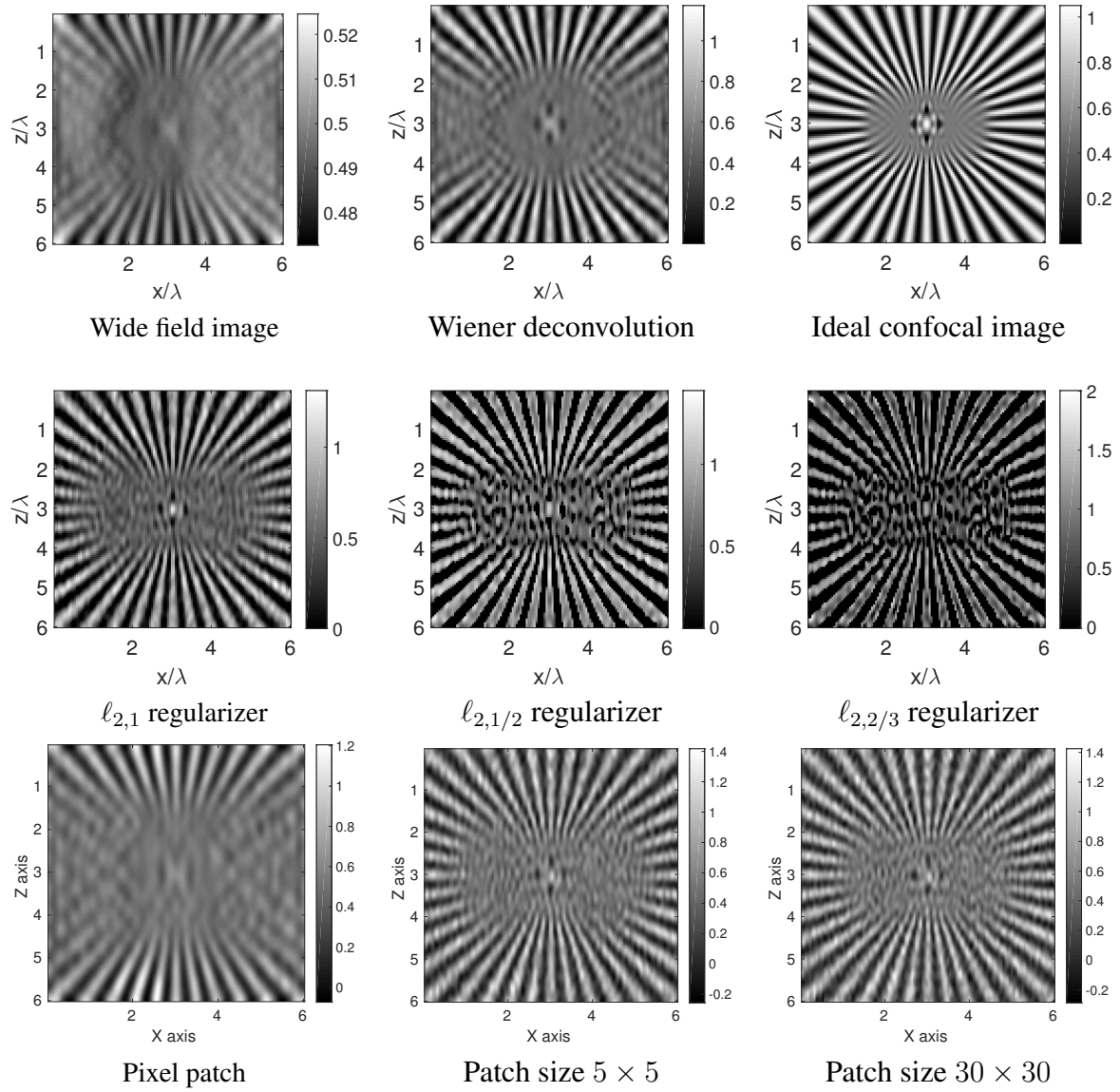
Figure 5.6 – The concept of resolution improvement in 3D SIM with harmonic patterns. (Figure courtesy of [GSC+08].) (a and b) Observable regions for the conventional microscopy. Resolution enhancement using two illumination beams (c), and three illumination beams in one (d,e). (f) The three wave vectors corresponding to the three illumination beam directions to generate 3D harmonic pattern. (g–h) The resulting spatial frequency components of the illumination intensity for the two-beam (g) and three-beam (h) case. (i, j): xz (i) and xy (j) sections through the OTF supports.

The other 3D simulations are performed with the synthetic volume $128 \times 128 \times 128$ of spherical beads and hollow bars as shown in Fig. 5.9, which are available online [3D 17]. The beads are getting closer and closer in the Y and Z directions and the 3D bars data consist of six parallel hollow bars. The normalized bandwidth limit of PSF is set to $\frac{f_{\max}}{f_S/2} = 0\,4$ in lateral direction and $0\,2$ in axial direction. The deconvolution with Tikhonov regularizator and reconstruction with the $\ell_{2,1}$ regularizer of the proposed method using 90 speckle patterns are shown in Fig. 5.11 and 5.12. In the Tikhonov regularized deconvolution, we see no structure information in axial direction with 3D beads sample, while the beads in axial direction can be well separated in low frequency part using our proposed method. The improvement in axial resolution with hollow bars is also observed as shown in Fig. 5.12.



a.) PSF b.) OTF

Figure 5.7 – **PSF and OTF in XZ axis.**



Wide field image | Wiener deconvolution | Ideal confocal image

$\ell_{2,1}$ regularizer | $\ell_{2,1/2}$ regularizer | $\ell_{2,2/3}$ regularizer

Pixel patch | Patch size 5×5 | Patch size 30×30

Figure 5.8 – **Reconstruction results with 300 speckle patterns and 40dB Gaussian white noise of 2D object in $y = 0$ plane.** The second line are obtained by joint reconstruction approach with $\ell_{p,q}$ norm regularizer, and the third line are obtained by patch based marginal estimator.

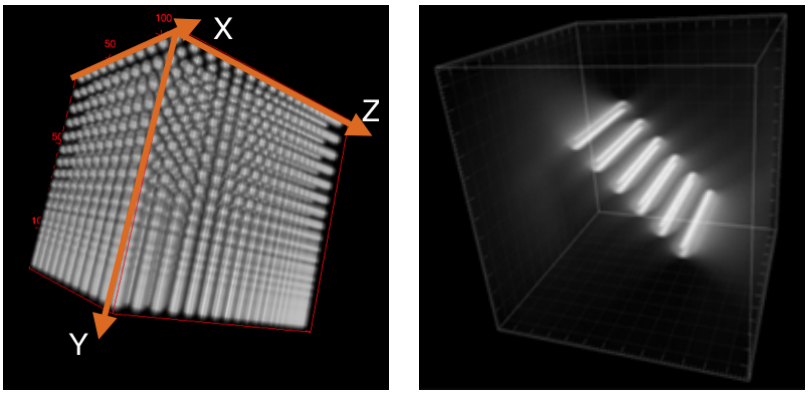


Figure 5.9 – **Synthetic data of** $128 \times 128 \times 128$ **cube of spherical beads and hollow bars.**

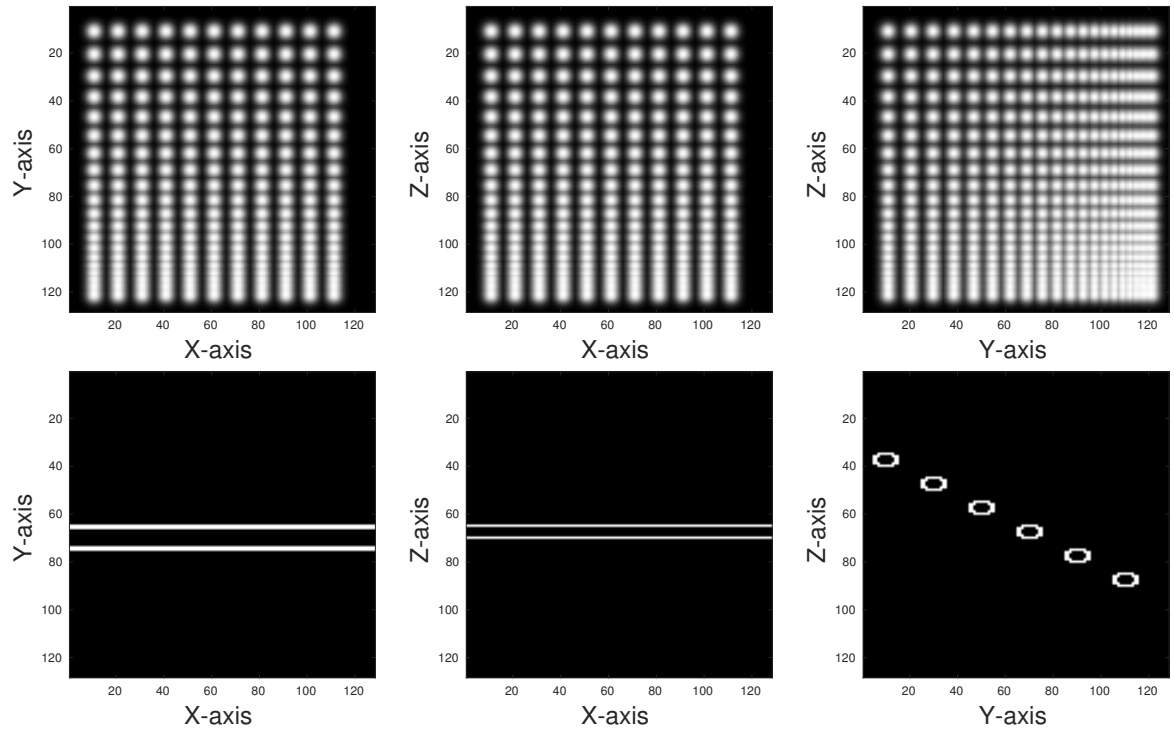


Figure 5.10 – **The 64th slice of spherical beads (first line) and the 68th slice of hollow bars (second line) in the XY, XZ, YZ dimension shown in figure 5.9.**

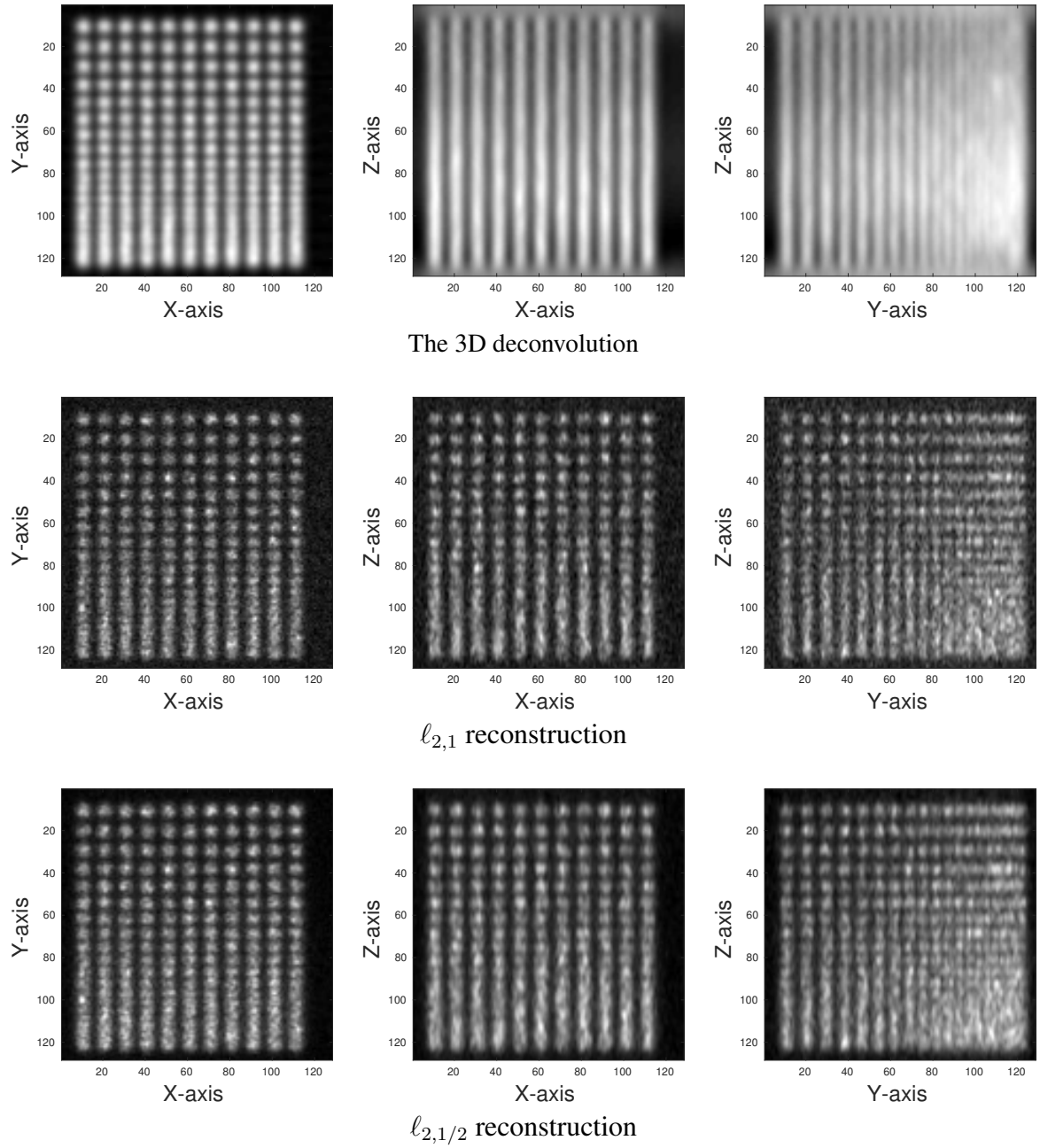


Figure 5.11 – **Reconstruction of cube of spherical beads with 90 3D speckles.** The first line corresponds to certain slides in XY, XZ, YZ directions about the 3D deconvolution of the mean, while the second and third lines show the slides in the same positions of the reconstructed object using constraint $\ell_{2,1}$ and $\ell_{2,1/2}$ regularizer.

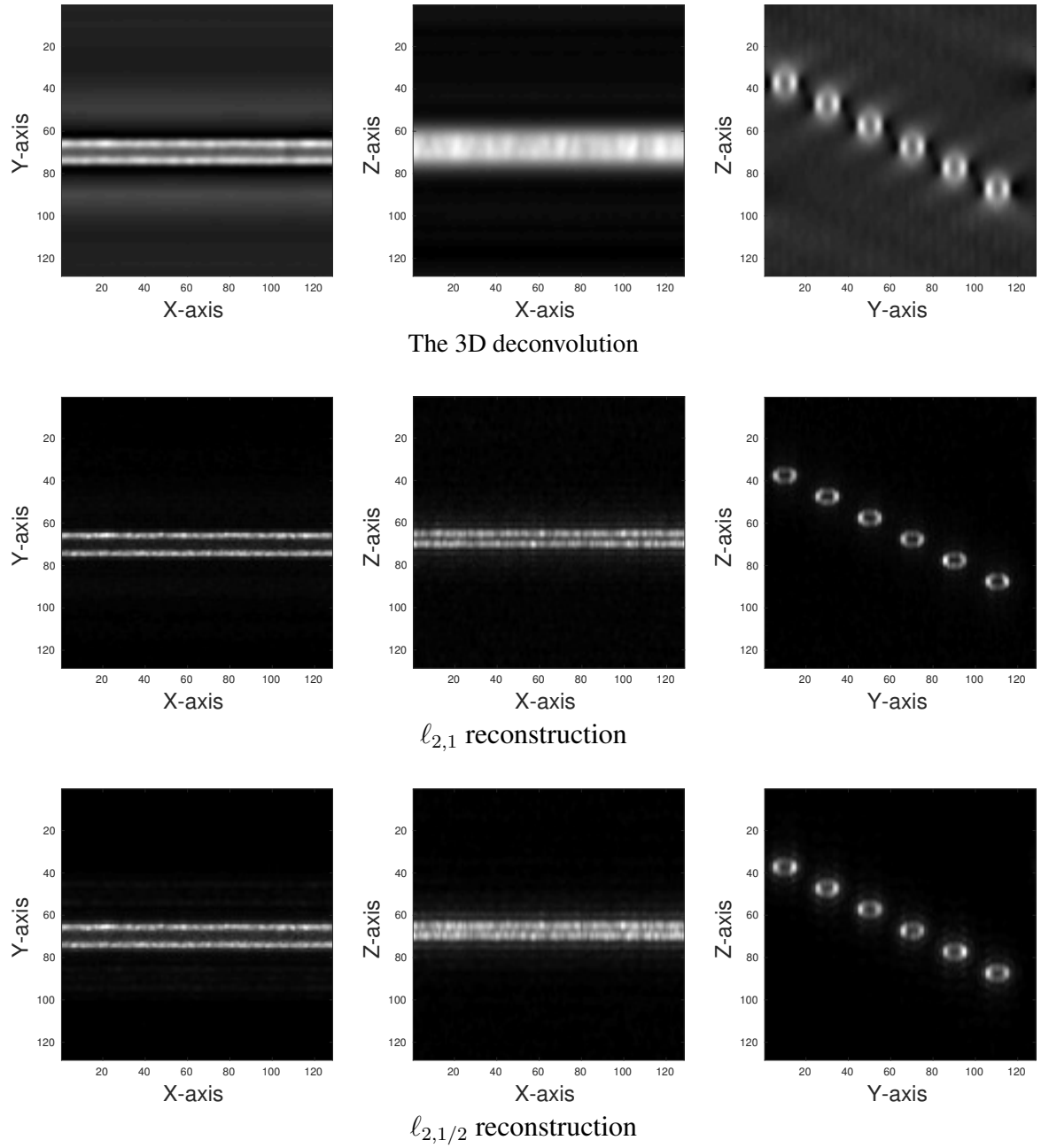


Figure 5.12 – **Reconstruction of hollow bars with 90 3D speckles.** The first line corresponds to certain slides in XY, XZ, YZ directions about the 3D deconvolution of the mean, while the second line and third line show the slides in the same positions of the reconstructed object using constraint $\ell_{2,1}$ and $\ell_{2,1/2}$ regularizers.

5.4 Conclusion

In this chapter we briefly introduce the optical sectioning techniques in microscopy and demonstrate the 3D imaging ability of blind-speckle-SIM using the constrained $\ell_{p,q}$ regularizer. The simulation results show an evident improvement in axial resolution compared with the 3D Wiener deconvolution.

Chapter 6

Conclusion and perspectives

This thesis proposes two new reconstruction methods in SIM using unknown speckle patterns (blind-speckle-SIM). One joint reconstruction approach uses $\ell_{p,q}$ regularizer with $p \geq 1$ and $0 < q \leq 1$. Both the constrained form and the unconstrained form of the $\ell_{p,q}$ regularizer are presented in Chapter 2. For the constrained form, the hyperparameter is easy to tune and its associated optimization algorithm is assured to converge. The simulation results show that we could retrieve better super-resolution using $\ell_{p,q}$ norm regularizer than the reconstruction method shown in [MBG+12] with positivity constraint regularizer. Moreover, the other prior information about the object can be easily incorporated into our method, such as TV norm or positivity constraint, without big changes of the associated primal-dual algorithm.

The theoretical analysis of the super-resolution capacity using second-order statistics of data are presented in Chapter 3. Our analysis shows that it allows us to retrieve super-resolution with equivalent PSF h^2 by using fully developed speckle patterns with covariance identifying with PSF h. Together with this theoretical analysis, a consistent marginal estimator is proposed by minimizing the dissimilarity between the empirical and theoretical second-order statistics of the data. The computational complexity of the marginal estimator is $\mathcal{O}(N^3)$ and the memory requirement is $\mathcal{O}(N^2)$, which is too high to apply it in realistic size imaging problems.

A patch-based marginal estimator is proposed in Chapter 4 by neglecting the correlation information of pixels from different patches. Thus the computational complexity and the memory request can be reduced to $\mathcal{O}\big(N^2(\log N+L)\big)$ and $\mathcal{O}(LN)$, respectively. The simulation results show that we could reach good super-resolution using relatively small patch sizes. Under asymptotic conditions, even with pixel patch we can retrieve partial super-resolution. Besides the theoretical assurance of the super-resolution capacity, the marginal estimator is more robust to the background noise, since the covariance of data does not change when a stable background noise is presented.

The conventional SIM and blind-SIM is a 2D super-resolution technique compatible with very thin samples. In 3D imaging problems with thick samples, the image of a given focal plane is plagued by out-of-focus fluorescence. The optical sectioning ability of blind-speckle-SIM is demonstrated in Chapter 5. Simulation results show that our proposed methods can improve both the lateral and the axial resolution in 3D imaging.

Surpassing the resolution limit beyond a factor of two is possible in saturated structured illumination microscopy (nonlinear SIM) [Gus05]. However, we do not find method to reach resolution more than a factor of two in blind-speckle-SIM using speckle patterns with frequency support higher than that of PSF h as shown in Section 4.5.4. Whether we could surpass the traditional resolution limit more than a factor of two in blind-speckle-SIM is still unclear.

When the support of the speckle spectral density is lower than the frequency support of PSF $\nu_{\text{spec}} < \nu_{\text{PSF}}$, the simulations show that we can not only retrieve frequency information interior

of the circle with radius $2\nu_{\text{spec}}$ as predicted by the theoretical analysis in Chapter 4 using the marginal estimator, but also the information between $2\nu_{\text{spec}}$ and $\nu_{\text{spec}} + \nu_{\text{PSF}}$. The theoretical reason behind this phenomena deserve to be better understood.

Both the joint reconstruction approach with $\ell p, q$ norm regularizer and the patch based marginal approach allow parallel computing in optimization part. In the joint reconstruction approach presented in Chapter 2, the variable $\mathbf{Q}$ and the auxiliary variables can be decoulpled column-wise or line-wise, so they can be processed by different cores in parallel. In the patch based marginal approach, the gradient for the objective function of each patch could be processed independently.

For the numerical optimization algorithms we used in the manuscript, the preconditioning technique has the possibility to accelerate the ADMM and primal-dual algorithm presented in Chapter 2. And the Majorize-Minimize algorithm for subspace optimization normally converges faster than the L-BFGS algorithm used in Chapter 3 and Chapter 4 [CIM11].

Finally, the speckle patterns can not only be used in fluorescence microscopy, its application in other imaging system, such as tomographic diffraction microscopy [HBGS10], photoacoustic imaging [CGA+16], should be explored in future.

Appendix A

Statistical properties of speckle patterns

It was noted in the early 1960s that when an object is illuminated by *coherent* light (see section 1.2.2), a high-contrast, fine-scale granular pattern would be observed [Goo07]. This type of granularity pattern is known as "speckle". Speckle also appears when coherent light is transmitted through diffusers.

The origin of the speckle phenomenon is due to the roughness of the object surface at the scale of optical wavelengths. The observed field is composed of multiple randomly phased elementary contributions, and these contributions interfere with each other to produce the observed granular pattern.

Speckle is also found in other coherent imaging systems, including synthetic-aperture radar [ALBA13], ultrasound medical imaging of organs [HAB13], and coherent X-ray imaging [SMG+91]. Initially, speckle was considered mainly due to its negative impact and was treated as a phenomenon that had to be suppressed. Recently, the positive impact of speckle has been considered and it turned out that speckle can help us obtain better resolution in diffraction limited systems [ILA+17]. Chapter 3 of the current manuscript is dedicated to such analysis.

Since the micro-structure of the surface roughness which generates speckle is unknown, we cannot describe the speckle information exactly. However, it can be dscribed through its statistical properties. Here we explore the statistical properties of the polarized fully developed speckle patterns. It shows that for fully developed speckle patterns, the intensity is homogeneous in average and that its image covariance has the same shape as the PSF of an unaberrated system (where the pupil function is either 0 or 1). For more detailed description, please see [Goo15, Section 7.7] and [Goo07, Section 4.4].

A.1 Statistical properties of speckle at a single point

As has been discussed in section 1.2.2, the complex phasor amplitude of coherent light can be written as $U(\boldsymbol{r})$. The optical field at a single point in speckle is the superposition of multiple randomly phased "elementary" complex components:

$$U(\boldsymbol{r}) = \sum_n U_n(\boldsymbol{r}) = \sum_{n=1}^{N_c} a_n(\boldsymbol{r}) e^{j\phi_n(\boldsymbol{r})} \tag{A.1}$$

with N_c the number of components and $a_n(\boldsymbol{r}) = |U_n(\boldsymbol{r})|$. For brevity of expression, the dependence on $(\boldsymbol{r})$ of U_n is dropped and we write:

$$U = \sum_n a_n e^{j\phi_n} \tag{A.2}$$

Throughout our discussions, the *fully developed* speckle patterns are concerned, that is, the following assumptions about $\boldsymbol{a}_n$ are adopted [Goo07]:

1. The amplitudes a_n and phases ϕ_n are statistically independent of a_m and ϕ_m provided that $n \neq m$.
2. For any n, a_n and ϕ_n are statistically independent.
3. For any n, ϕ_n are uniformly distributed on $(-\pi, \pi)$.

A.1.1 First and second moments of real and imaginary parts of speckle

Given the assumptions above, we can write the mean of the real ($u_{\mathcal{R}}$) and imaginary ($u_{\mathcal{I}}$) parts of U as:

$$E[u_{\mathcal{R}}] = \sum_n E[a_n \cos\phi_n] = \sum_n E[a_n]E[\cos\phi_n] = 0, \tag{A.3}$$

$$E[u_{\mathcal{I}}] = \sum_n E[a_n \sin\phi_n] = \sum_n E[a_n]E[\sin\phi_n] = 0, \tag{A.4}$$

where both the expectations of $\cos\phi_n$ and $\sin\phi_n$ are zero due to the uniform distribution of ϕ_n. Since the means are zero, the variance of $u_{\mathcal{R}}$ and $u_{\mathcal{I}}$ are given by their second moments:

$$\begin{aligned}\sigma_{\mathcal{R}}^2 = E[u_{\mathcal{R}}^2] &= \sum_{n=1}^{N}\sum_{m=1}^{N} E[a_n a_m]E[\cos\phi_n \cos\phi_m] \\ &= \sum_{n=1}^{N} E[a_n^2]E[\cos^2\phi_n] = \sum_n E[a_n^2]E[\frac{1}{2} + \frac{1}{2}\cos 2\phi_n] \\ &= \sum_n \frac{E[a_n^2]}{2}\end{aligned} \tag{A.5}$$

in which we use the fact that $E[\cos\phi_n \cos\phi_m] = E[\cos\phi_n]E[\cos\phi_m] = 0$ when $n \neq m$. Similarly, we could obtain that:

$$\sigma_{\mathcal{I}}^2 = \sum_n \frac{E[a_n^2]}{2} \tag{A.6}$$

The correlation between the real and imaginary parts of the speckle are given by:

$$\begin{aligned}\Gamma_{\mathcal{R},\mathcal{I}} = E[u_{\mathcal{R}} u_{\mathcal{I}}] &= \sum_n E[a_n^2]E[\cos\phi_n \sin\phi_n] \\ &= \sum_n \frac{1}{2}E[a_n^2]E[\sin 2\phi_n] = 0.\end{aligned} \tag{A.7}$$

A.1.2 Circular complex Gaussian variable

When the number of elementary components N_c is very large, $u_{\mathcal{R}}$ and $u_{\mathcal{I}}$ are given by sums of a large number of independent random variables. According to the Central Limit Theorem [Haj15], $u_{\mathcal{R}}$ and $u_{\mathcal{I}}$ are asymptotically Gaussian as $N_c \to \infty$. Given the mean, variance and correlation of $u_{\mathcal{R}}$ and $u_{\mathcal{I}}$, we can write the joint probability density function of the speckle $\boldsymbol{u}$ by:

$$p(u_{\mathcal{R}}, u_{\mathcal{I}}) = \frac{1}{2\pi\sigma^2}\exp(-\frac{u_{\mathcal{R}}^2 + u_{\mathcal{I}}^2}{2\sigma^2}) \tag{A.8}$$

with $\sigma^2 = \sigma_{\mathcal{R}}^2 = \sigma_{\mathcal{I}}^2$. This complex phasor U is called circular complex Gaussian variable. Concerning the amplitude a and phase θ of the resultant phasor $\boldsymbol{u}$, we have:

$$\begin{aligned} a &= \sqrt{u_{\mathcal{R}}^2 + u_{\mathcal{I}}^2} \\ \theta &= \arctan\{\frac{u_{\mathcal{I}}}{u_{\mathcal{R}}}\} \end{aligned} \tag{A.9}$$

The joint probability density function with respect to a and θ is related to that of $u_{\mathcal{R}}, u_{\mathcal{I}}$ by:

$$p(a,\theta) = p\big(u_{\mathcal{R}}(a,\theta), u_{\mathcal{I}}(a,\theta)\big)|J| \tag{A.10}$$

in which $|J|$ represents the Jacobian matrix given by:

$$|J| = \begin{vmatrix} \frac{\partial u_{\mathcal{R}}}{\partial a} & \frac{\partial u_{\mathcal{R}}}{\partial \theta} \\ \frac{\partial u_{\mathcal{I}}}{\partial a} & \frac{\partial u_{\mathcal{R}}}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -a\sin\theta \\ \sin\theta & a\cos\theta \end{vmatrix} = a. \tag{A.11}$$

Therefore, the joint probability density function of u and θ is:

$$p(a,\theta) = \frac{a}{2\pi\sigma^2}\exp(-\frac{a^2}{2\sigma^2}) \tag{A.12}$$

for $a \geq 0$ and $-\pi \leq \theta \leq \pi$, zero otherwise. The marginal probability density function of amplitude a is:

$$p(a) = \int_{-\pi}^{\pi} p(a,\theta)d\theta = \frac{a}{\sigma^2}\exp(-\frac{a^2}{2\sigma^2}) \tag{A.13}$$

for $a \geq 0$ and $p(a)$ is called the Rayleigh probability density function [Haj15].

A.1.3 Statistical properties of speckle intensity

Here we focus on intensity of speckle patterns. For the coherent illumination, we have $I = U^2 = a^2$. Combining with the probability density function $p(a)$ in Eq. (A.13), we have:

$$p(I) = p(\sqrt{a})|\frac{da}{dI}| = \frac{1}{2\sigma^2}\exp\big(-\frac{I}{2\sigma^2}\big) \tag{A.14}$$

for $I \geq 0$. The i-th order moment of this distribution is given by [Goo07]:

$$\overline{I^q} = (2\sigma^2)^q q! \tag{A.15}$$

Then we could express the first and second moments of speckle intensity by:

$$\bar{I} = 2\sigma^2, \qquad \overline{I^2} = 8\sigma^4 = 2\bar{I}^{\,2} \tag{A.16}$$

Now we can conclude that the intensity for fully developed speckle patterns is homogeneous in average with $\bar{I}(\boldsymbol{r}) = 2\sigma^2$.

A.2 Joint statistical properties of speckle at two points

After the discussion of statistical properties of one point in speckle, we explore the joint properties of speckle patterns at two points in this subsection.

A.2.1 Multivariate Gaussian statistics

The mathematical model to describe a fully developed speckle is a circular complex Gaussian random process, with real and imaginary parts being real-valued Gaussian random processes. Let $\mathbf{U} = [U_1, \cdots, U_N]^t$ be an N-dimensional random vector subject to circular complex Gaussian distribution, its probability density function is given by [Goo15]:

$$p(\mathbf{U}) = \frac{1}{(2\pi)^{N/2}|\mathcal{C}|^{1/2}} \exp\big[-\frac{1}{2}(\mathbf{U}-\bar{\mathbf{U}})^t\mathcal{C}^{-1}(\mathbf{U}-\bar{\mathbf{U}})\big] \tag{A.17}$$

with $\mathcal{C}$ denoting the covariance matrix and $\bar{\mathbf{U}}$ the expectation of $\mathbf{U}$. The joint moments of a zero-mean complex Gaussian random process can be found by the *complex Gaussian moment theorem* [Ree62]:

$$E\big[U_1^*U_2^*\cdots U_k^*U_{k+1}U_{k+2}\cdots U_{2k}\big] = \sum_{\Pi} E[U_1^*U_{p_1}]E[U_2^*U_{p_2}]\cdots E[U_k^*U_{p_k}] \tag{A.18}$$

where the symbol $\sum_{\Pi}$ represents the sum over all possible $k!$ permutations $(p_1, \cdots, p_k)$ of $(1, \cdots, k)$.

A.2.2 Correlation of speckle through the imaging system

Now we turn attention to the speckle passing through an imaging system as shown in Fig. A.1. A rough object is illuminated by coherent light. The light passes through the imaging system and is then focused in the image plane.

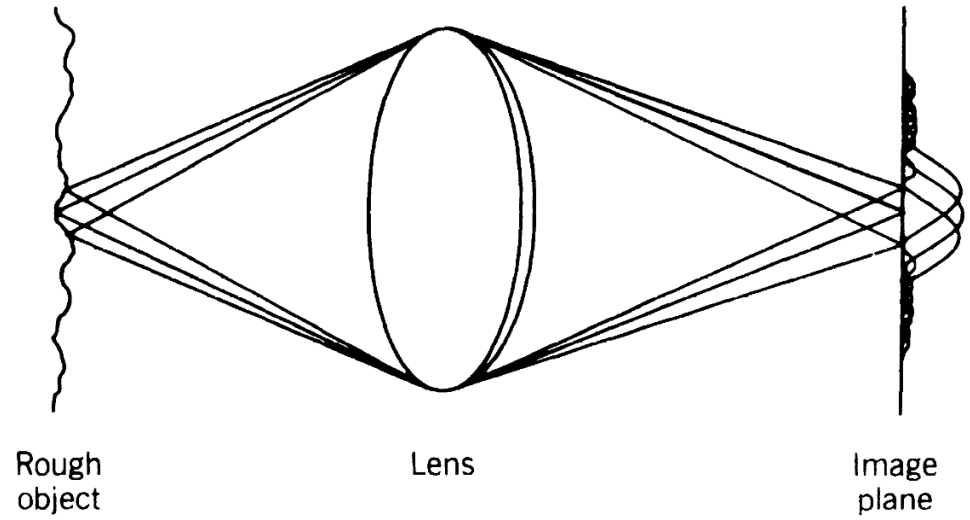


Figure A.1 – **Speckle formation in imaging system**. Figure courtesy from [Goo15, chapter 7]

Let h_a denote the amplitude point spread function (APSF) of the system , then the complex amplitude in the image plane is given by:

$$S(\boldsymbol{r}) = (h_a * U)(\boldsymbol{r}) = \int h_a(\boldsymbol{r}-\boldsymbol{r}')U(\boldsymbol{r}')d\boldsymbol{r}' \tag{A.19}$$

The correlation of complex phasor amplitude in image plane, sometimes also called ensemble-averaged mutual intensity to distinguish from the time-averaged mutual intensity shown in

Eq. (1.9) , is given by:

$$\begin{aligned} E[S_k S_r^*] &= E\Big[\int h_a(k-x)U_x dx \int h_a^*(r-y)U_y^* dy\Big] \\ &= \int\int h_a^*(r-y)h_a(k-x)E[U_x U_y^*]dxdy \end{aligned} \tag{A.20}$$

The correlation of the light in the source plane is determined by the surface-height correlation function and the wavelength of incident light [Goo15, chapter 8]. For the moment, we assume it can be represented by delta function:

$$E[U_x U_y^*] = \bar{I}\delta(\boldsymbol{x}-\boldsymbol{y}) \tag{A.21}$$

where $\bar{I}$ is the ensemble average of the light intensity. Then we have:

$$\begin{aligned} E[S_k S_r^*] &= \bar{I}\int h_a^*(r-x)h_a(k-x)dx \\ &= \bar{I}\,\Gamma_{h_a}(r-k) \end{aligned} \tag{A.22}$$

in which Γ_{h_a} denotes the autocorrelation of h_a:

$$\Gamma_{h_a}(\tau) = \int h_a(x)h_a^*(x-\tau)dx. \tag{A.23}$$

The intensity of the speckle in the image plane is given by $I(\boldsymbol{r}) = |S(\boldsymbol{r})|^2$. Then the mean intensity of speckle patterns in image plane is:

$$E[I(\boldsymbol{r})] = E[S_r S_r^*] = \bar{I}\,\Gamma_{h_a}(0) \tag{A.24}$$

It is clear from (A.24) that the average intensity does not depend on the position $\boldsymbol{r}$, so the intensity mean of speckle in image plane is still homogeneous and we denote it as I_0, to distinguish from the intensity mean in source plane $\bar{I}$. The covariance of the speckle intensity in the image plane is :

$$\begin{aligned} E[(I_k-I_0)(I_r-I_0)] =& E[I_k I_r] - I_0^2 && \text{(A.25a)}\\ =& E[S_k S_k^* S_r S_r^*] - I_0^2 && \text{(A.25b)}\\ =& E[S_k S_k^*]E[S_r S_r^*] + E[S_k S_r^*]E[S_r S_k^*] - I_0^2 && \text{(A.25c)}\\ =& E[S_k S_r^*]E[S_r S_k^*] && \text{(A.25d)}\\ =& |E(S_k S_r^*)|^2 && \text{(A.25e)}\\ =& \bar{I}^{\,2}\left|\Gamma_{h_a}(r-k)\right|^2 && \text{(A.25f)}\\ =& \frac{I_0^2}{|\Gamma_{h_a}(0)|^2}\left|\Gamma_{h_a}(r-k)\right|^2 && \text{(A.25g)} \end{aligned}$$

where (A.25c) is obtained by applying the *complex Gaussian moment theorem* to (A.25b). We see from(A.25) that the covariance of the intensity of speckle patterns can be totally characterized by the difference of the position of two points. Let's denote the covariance of speckle intensity by γ_{spec}, we have :

$$\gamma_{\text{spec}}(r) = \frac{I_0^2}{|\Gamma_{h_a}(0)|^2}\left|\Gamma_{h_a}(r)\right|^2 \tag{A.26}$$

The Fourier transform of γ_{spec} gives:

$$\begin{aligned}\tilde{\gamma}_{\text{spec}}(u) &= \frac{I_0^2}{|\Gamma_{h_a}(0)|^2}\Big(\tilde{\Gamma}_{h_a} \star \tilde{\Gamma}_{h_a}\Big)(u) \\ &= \frac{I_0^2}{|\Gamma_{h_a}(0)|^2}\Big(|\tilde{h}_a|^2 \star |\tilde{h}_a|^2\Big)(u)\end{aligned} \tag{A.27}$$

in which $\star$ denotes the autocorrelation operator and $\tilde{h}_a$ is the amplitude transfer function of the system, which is the projection of the scaled pupil function on the Ewald sphere [Goo05]. When the lens is not apodized, the pupil function equals either 1 or 0, so that we have $|\tilde{h}_a|^2 = \tilde{h}_a$. Thus the shape of the Fourier transform of the covariance of speckle intensity identifies with the OTF of the unaberrated system, up to a normalizing factor:

$$\tilde{\gamma}_{\text{spec}}(u) = \frac{I_0^2}{h(0)}\tilde{h}(u) \tag{A.28}$$

According to the *Wiener-Khinchin theorem*, $\tilde{\gamma}_{\text{spec}}$ is identified by the power spectral density of the speckle patterns [Goo15, Section 3.4]. Taking the inverse Fourier transform of (A.28), we have:

$$\gamma_{\text{spec}}(r) = \frac{I_0^2}{h(0)}h(r) \tag{A.29}$$

A.3 Simulation of speckle patterns

To simulate the speckle patterns with given mean i_0 and covariance $\gamma_{\text{spec}}(\boldsymbol{r})$, a simple way is to pass a independent and identically distributed (i.i.d) gaussian field U with a specific filter ϕ:

$$S(\boldsymbol{r}) = (U * \phi)(\boldsymbol{r}) \tag{A.30}$$

with

$$\begin{aligned}U(\boldsymbol{r}) &= a(\boldsymbol{r})e^{j\theta(\boldsymbol{r})} \\ a(\boldsymbol{r}) &\sim \mathcal{N}(0, \sigma^2) \\ \theta(\boldsymbol{r}) &= \mathcal{U}(0, 2\pi)\end{aligned} \tag{A.31}$$

where $\mathcal{U}$ denotes uniform distribution. According to our previous analysis, let

$$I(\boldsymbol{r}) = \frac{i_0}{2\sigma^2\phi(0)}|S|^2(\boldsymbol{r}) \tag{A.32}$$

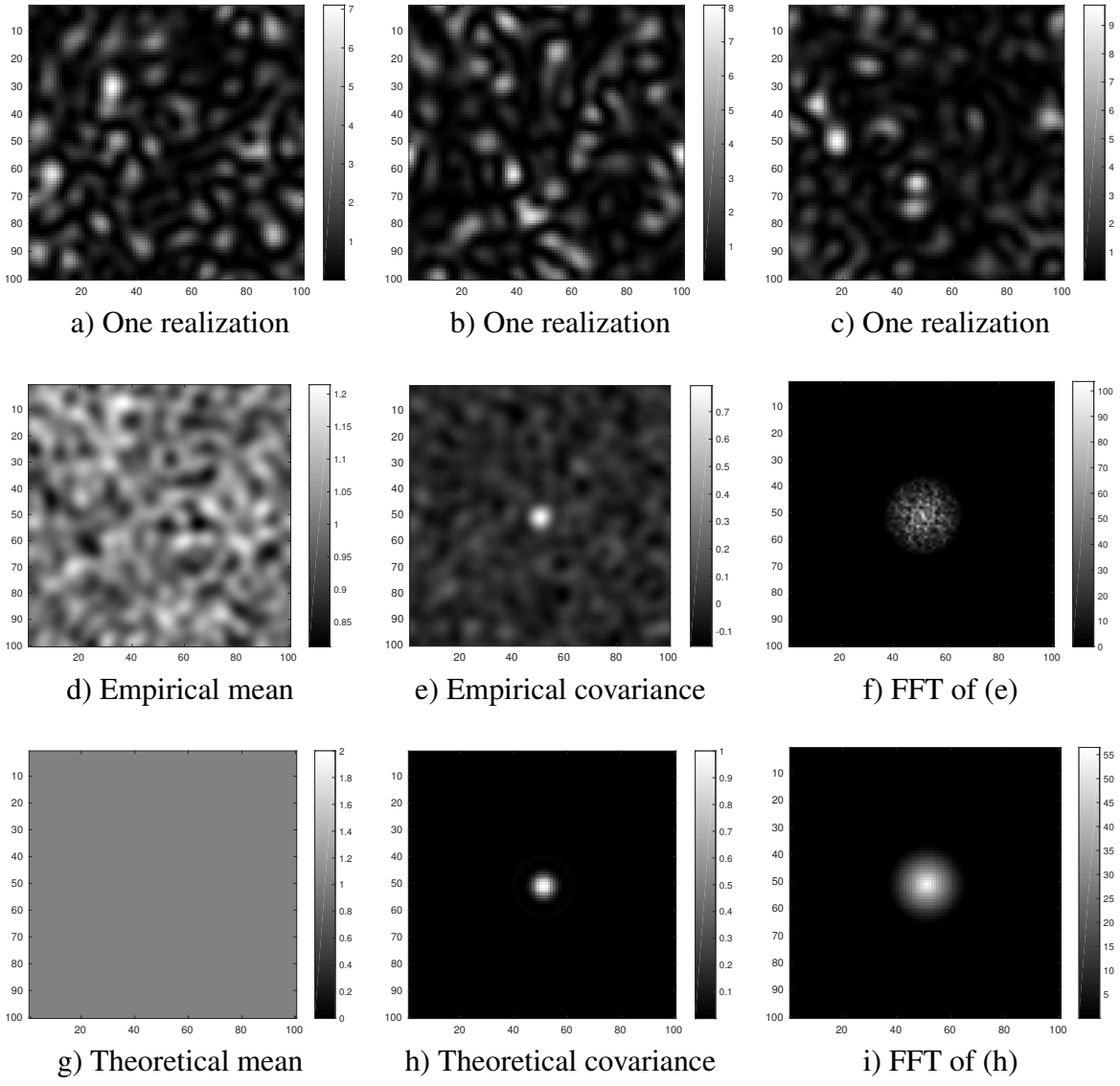
We have:

$$\begin{aligned}E[I(\boldsymbol{r})] &= i_0 \\ \text{Cov}\big(I(\boldsymbol{r}), I(\boldsymbol{k})\big) &= \frac{i_0^2}{|\phi(0)|^2}|\phi(\boldsymbol{r}-\boldsymbol{k})|^2\end{aligned} \tag{A.33}$$

For non-apodized lens, the covariance of $I(\boldsymbol{r})$ is identified with γ_{spec} as long as:

$$\phi(\boldsymbol{r}) = \frac{\gamma_{\text{spec}}(\mathbf{0})}{i_0}\sqrt{\gamma_{\text{spec}}(\boldsymbol{r})} \tag{A.34}$$

Three specific speckle realizations are shown in the first line of Fig. A.2, and the empirical mean and covariance of 300 speckle patterns are shown in the second line of Fig. A.2. We can see that they match their theorical values (shown in the third line of Fig. A.2) very well.



a) One realization b) One realization c) One realization

d) Empirical mean e) Empirical covariance f) FFT of (e)

g) Theoretical mean h) Theoretical covariance i) FFT of (h)

Figure A.2 – **The speckle realizations (first line) and its empirical and theoretical mean and covariance of 300 speckle patterns .**

Appendix B

The structure of 2D convolution matrix

B.1 The 2D convolution methods

Suppose we have an object $\boldsymbol{\rho} \in \mathbb{R}^{\sqrt{n}\times\sqrt{n}}$ and the bluring kernel $\boldsymbol{h} \in \mathbb{R}^{\sqrt{m}\times\sqrt{m}}$, where $m \leq n$, there are several ways to evaluate discrete convolution $\boldsymbol{z} = \boldsymbol{h} * \boldsymbol{\rho}$. According to the definition of discrete convolution:

$$\boldsymbol{z}(r_1, r_2) = \sum_{r_1'=1}^{\sqrt{m}} \sum_{r_2'=1}^{\sqrt{m}} \boldsymbol{\rho}(r_1 - r_1', r_2 - r_2')\boldsymbol{h}(r_1', r_2') \tag{B.1}$$

A straightforward computation requires $\sqrt{m} \times \sqrt{m}$ multiplications and addations for each pixel of $\boldsymbol{z}$. Here the circular convolution is considered, i.e. the signal is supposed to be periodic. The size of $\boldsymbol{z}$ is the same as the object $\boldsymbol{\rho}$, the computation complexity of convolution by definition is $\mathcal{O}(mn)$.

B.1.1 Convolution by FFT

Circular convolution could be implemented with the help of discrete Fourier transform:

$$\boldsymbol{z} = \mathcal{F}^{-1}(\tilde{\boldsymbol{\rho}} \circ \tilde{\boldsymbol{h}}) \tag{B.2}$$

Taking advantage of the fast Fourier transform (FFT) algorithm, the computation complexity to evaluate 2D convolution is $\mathcal{O}(n \log n)$. Linear convolution could also be implemented with FFT by padding zeros to the original object and kernel to avoid overlap between one period and the next. In the case where $m \ll n$, separate convolution and overlap convolution are strong candidates for efficient implementation.

B.1.2 Fast 2D convolution by SVD

The 2D fast convolution is based on the fact that when the convolution kernel has a product decomposition, i.e. $\boldsymbol{h} = \boldsymbol{u}\boldsymbol{v}^T$, with $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{m\times 1}$, we have more efficient ways to do the convolution [HS91, 298-299].

$$\begin{bmatrix} h_{11} & h_{12} & \dots & h_{1m} \\ h_{21} & h_{22} & \dots & h_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ h_{m1} & h_{m2} & \dots & h_{mm} \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix}$$

Simple manipulation of (B.1) gives:

$$\begin{aligned}\boldsymbol{z}(r_1, r_2) &= \sum_{r_1'=1}^{\sqrt{m}} \sum_{r_2'=1}^{\sqrt{m}} \boldsymbol{\rho}(r_1 - r_1', r_2 - r_2')\boldsymbol{h}(r_1', r_2') \\ &= \sum_{r_1'=1}^{\sqrt{m}} \sum_{r_2'=1}^{\sqrt{m}} \boldsymbol{\rho}(r_1 - r_1', r_2 - r_2')\boldsymbol{u}(r_1')\boldsymbol{v}(r_2') \\ &= \sum_{r_2'=1}^{\sqrt{m}} \left[\sum_{r_1'=1}^{\sqrt{m}} \boldsymbol{\rho}(r_1 - r_1', r_2 - r_2')\boldsymbol{u}(r_1') \right] \boldsymbol{v}(r_2')\end{aligned} \tag{B.3}$$

The bracketed term is the object convolving with column vector $\boldsymbol{u}$, then the result convolve again with the line vector $\boldsymbol{v}^T$. As a result, the original 2D convolution are separated into 2 convolutions:

$$\boldsymbol{z} = (\boldsymbol{\rho} * \boldsymbol{u}) * \boldsymbol{v}^T \tag{B.4}$$

The number of multiplications and additions required is $\mathcal{O}(2n\sqrt{m})$. Next support the kernel $\boldsymbol{h}$ can be expressed as the sum of K separable kernels:

$$\boldsymbol{h} = \sum_{k=1}^{K} \boldsymbol{h}_k = \sum_{k=1}^{K} \boldsymbol{u}_k \boldsymbol{v}_k^T \tag{B.5}$$

We have:

$$\boldsymbol{z} = \sum_{k=1}^{K} (\boldsymbol{h}_k * \boldsymbol{\rho}) = \sum_{k=1}^{K} (\boldsymbol{\rho} * \boldsymbol{u}_k) * \boldsymbol{v}_k \tag{B.6}$$

Such a separable representation is possible for any kernel $\boldsymbol{h}$ by using singular value decomposition (SVD):

$$\boldsymbol{h} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T = \sum_{i=1}^{\sqrt{m}} \lambda_i \boldsymbol{u}_i \boldsymbol{v}_i^T \tag{B.7}$$

Let the singular values λ_i arrange in a decreasing order. Since λ_i decay to zero rapidly, only the first few large values need to be kept. Suppose κ singular values are retained, the computational complexity for 2D convolution by using separate convolution is $\mathcal{O}(\kappa n\sqrt{m})$.

B.1.3 Overlap methods

When we implement convolution by FFT, zeros are appended to the kernel $\boldsymbol{h}$ to get the same size as the object. For large size object and small size kernels, a more efficient way to implement it is using blockwise linear convolution, which is the basic idea behind the overlap methods.

In overlap add method, the blocks are non-overlapping and the linear convolution for each block is achieved with circulant convolution by adding zeros to both the block data and the kernel.

While in overlap save method, no zero padding is performed on the input data. The object is cut into overlapping blocks, and then convolve each block with the kernel using circulant convolution. The portions that "wrap around" are discarded and the left equivalent with linear convolution is saved.

Suppose each block has l pixels ($l \geq m$), the computational complexity for overlap methods is $\mathcal{O}(n \log l)$. In practice, l is often chosen to be an integer power of 2 to achieve good performance of FFT.

B.2 Generating the convolution matrix for specific patch

For the 2D convolution $y = h * \rho$, with $y, h, \rho \in \mathbb{R}^{m\times n}$, the associated matrix form is:

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{\rho} \tag{B.8}$$

where $\boldsymbol{y}, \boldsymbol{\rho} \in \mathbb{R}^{mn}$ and $\mathbf{H} \in \mathbb{R}^{mn\times mn}$. If only partial part $\boldsymbol{y}_p$ of $\boldsymbol{y}$ is concerned, we have:

$$\boldsymbol{y}_p = \mathbf{H}_p\boldsymbol{\rho} \tag{B.9}$$

In the patch based marginal approach as presented in chapter 4, we need to generate the convolution matrix $\mathbf{H}_p$ given the location index of the corresponding patch. One possible method is to construct the whole convolution matrix $\mathbf{H} \in \mathbb{R}^{N\times N}$ first, then extract $\mathbf{H}_p$ according to the patch locations. However, when the object size is big, storing matrix $\mathbf{H}$ is very memory consuming. For example, for a 512×512 image, if we store $\mathbf{H}$ with double-precision (8 bytes), the memory required is:

$$\frac{512^2 \times 512^2 \times 8}{2^{10} \times 2^{10} \times 2^{10}} = 512(GB) \tag{B.10}$$

which is unaffordable for normal computers, even for supercomputers. Next, we will introduce a generation approach for patch convolution matrix $\mathbf{H}_p$, based on the natural structure of $\mathbf{H}$, where constructing $\mathbf{H}$ is not necessary.

B.2.1 Construct $\mathbf{H}_p$ with BCCB $\mathbf{H}$

With periodic boundary conditions of the object $\boldsymbol{\rho}$, $\mathbf{H}$ is a block circulant with circulant blocks (BCCB) matrix. Otherwise with zero boundary assumption, $\mathbf{H}$ has a block Toeplitz with Toeplitz blocks (BTTB) structure [HNO06]. Firstly we focus on BCCB $\mathbf{H}$. A $mn \times mn$ BCCB matrix has the following form:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_0 & \mathbf{B}_{m-1} & \dots & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{B}_0 & \dots & \mathbf{B}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{m-1} & \mathbf{B}_{m-2} & \dots & \mathbf{B}_0 \end{bmatrix}$$

in which each $\mathbf{B}_j$, $j = (0, 1, \cdots, m-1)$, is a $n\times n$ circulant matrix. To generate the convolution matrix $\mathbf{H}_p$ for p-th patch, we firstly construct the convolution matrix $\mathbf{H}_1$ explicitly for the first patch. Then we use $\mathbf{H}_1$ as reference and construct $\mathbf{H}_p$ by circulating the lines and columns of $\mathbf{H}_1$ according to their relatively position difference of the patch p and the first patch.

Appendix C

Implementing TV regularizer by FFT

For a discrete 2D object $\rho_{2D} \in \mathbb{R}^{N_1 \times N_2}$, the *anisotropic* TV is defined as:

$$TV_a(\rho_{2D}) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \left|\rho[n_1+1, n_2] - \rho[n_1, n_2]\right| + \left|\rho[n_1, n_2+1] - \rho[n_1, n_2]\right| \tag{C.1}$$

and the so-called *isotropic* TV is defined as:

$$TV_i(\rho_{2D}) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \sqrt{(\rho[n_1+1, n_2] - \rho[n_1, n_2])^2 + (\rho[n_1, n_2+1] - \rho[n_1, n_2])^2} \tag{C.2}$$

It is well known that the anisotropic TV is a poor definition of discrete TV since it favors horizontal and vertical structures, while the isotropic TV has good isotropic property in practical application as its name implies. Some other formulation of the discrete TV has also been proposed, such as *upwind* TV [CLL11], Shannon TV [AM16] to enforce the isotropic property. In the method presented in section 2.3, we use isotropic TV.

C.1 Mixed norm representation of TV

First we vectorize the object ρ_{2D} in lexicological order and write it as $\boldsymbol{\rho} \in \mathbb{R}^N$ with $N = N_1 N_2$. Then we note that the $TV_i(\rho_{2D})$ can be expressed as the mixed norm $\ell_{\mathcal{G},2,1}$ norm of $\mathbf{C}\boldsymbol{\rho}$:

$$TV_i(\rho_{2D}) = \|\mathbf{C}\boldsymbol{\rho}\|_{\mathcal{G},2,1} = \sum_{n=1}^{N} \|(\mathbf{C}\boldsymbol{\rho})_{\mathcal{G}_n}\|_2 \tag{C.3}$$

with $\mathbf{C} = (\mathbf{C}_1; \mathbf{C}_2)$ and $\mathbf{C}_1, \mathbf{C}_2$ the first-order horizontal and vertical finite difference operators. The n-th group $(\mathbf{C}\boldsymbol{\rho})_{\mathcal{G}_n} = [(\mathbf{C}_1\boldsymbol{\rho})_n; (\mathbf{C}_2\boldsymbol{\rho})_n] \in \mathbb{R}^2$.

C.2 Implementing the finite difference operator by FFT

Now let us zoom in the structure of $\mathbf{C}_1$ and $\mathbf{C}_2$. Under the periodic boundary conditions for ρ_{2D}, $\mathbf{C}_1, \mathbf{C}_2$ have a block Circulant with Circulant blocks (BCCB) structure:

$$\mathbf{C}_1 = \begin{bmatrix} -\mathbb{1}_{N_1} & \mathbb{1}_{N_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & -\mathbb{1}_{N_1} & \mathbb{1}_{N_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbb{1}_{N_1} & \mathbf{0} & \mathbf{0} & \dots & -\mathbb{1}_{N_1} \end{bmatrix}$$

where $\mathbb{1}_{N_1}$ is a $N_1 \times N_1$ identity matrix. Similarly,

$$\mathbf{C}_2 = \begin{bmatrix} \mathbf{A}^0 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^0 & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}^0 \end{bmatrix}$$

where $\mathbf{A}^0$ is a $N_1 \times N_1$ circulant matrix:

$$\mathbf{A}^0 = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & -1 \end{bmatrix}$$

A BCCB matrix can be diagonalized by 2D discrete Fourier transform [Dav12]. Let the first column of $\mathbf{C}_1$ be $\boldsymbol{c}_1$, we can write:

$$\mathbf{C}_1 = (\mathbb{F}_{N_1} \otimes \mathbb{F}_{N_2})^{-1} \text{diag}(\tilde{\boldsymbol{c}}_1)(\mathbb{F}_{N_1} \otimes \mathbb{F}_{N_2}) \tag{C.4}$$

with $\mathbb{F}$ the discrete Fourier transform matrix and $\tilde{\boldsymbol{c}}_1 = (\mathbb{F}_{N_1} \otimes \mathbb{F}_{N_2})\boldsymbol{c}_1$. Then $\mathbf{C}_1\boldsymbol{\rho}$ could be implemented by:

$$\mathbf{C}_1\boldsymbol{\rho} = (\mathbb{F}_{N_1} \otimes \mathbb{F}_{N_2})^{-1} \text{diag}(\tilde{\boldsymbol{c}}_1)(\mathbb{F}_{N_1} \otimes \mathbb{F}_{N_2})\boldsymbol{\rho} = \mathcal{F}^{-1}\left(\tilde{\boldsymbol{c}}_1 \circ \tilde{\boldsymbol{\rho}}\right) \tag{C.5}$$

The implementation of $\mathbf{C}_2\boldsymbol{\rho}$ can be treated similarly. On the other hand, under the Neumann boundary condition of the object ρ_{2D}, $\mathbf{C}_1$ and $\mathbf{C}_2$ has a block Toeplitz-plus-Hankel structure, and $\mathbf{C}_1\boldsymbol{\rho}$ could be implemented with the help of discrete cosine transforms (DCTs) [HNO06].

C.3 The transpose of finite difference operator

Let us denote $\boldsymbol{p} = \mathbf{C}\boldsymbol{\rho} \in \mathbb{R}^{2N}$. The operation $\mathbf{C}^T\boldsymbol{p}$ is required sometimes. Instead of implementing it directly, we could split $\boldsymbol{p}$ into two parts $[\boldsymbol{p}_1; \boldsymbol{p}_2]$ where $\boldsymbol{p}_1 = \mathbf{C}_1\boldsymbol{\rho}$ and $\boldsymbol{p}_2 = \mathbf{C}_2\boldsymbol{\rho}$. Then $\mathbf{C}^T\boldsymbol{p}$ could be expressed as the sum of two items:

$$\mathbf{C}^T\boldsymbol{p} = \mathbf{C}_1^T\boldsymbol{p}_1 + \mathbf{C}_2^T\boldsymbol{p}_2 \tag{C.6}$$

where each item could be implemented with FFT:

$$\mathbf{C}_1^T\boldsymbol{p}_1 = \mathcal{F}^{-1}\left(\tilde{\boldsymbol{c}}_1^* \circ \tilde{\boldsymbol{p}}_1\right) \tag{C.7}$$

Appendix D

Numerical optimization methods

D.1 L-BFGS method

In this section we briefly present the L-BFGS algorithm to solve the optimization problem:

$$\arg\min_{\boldsymbol{\rho}} \; D_M(\boldsymbol{\rho}) \tag{D.1}$$

In each iteration of BFGS algorithm, we require the approximation of inverse of Hessian matrix $\mathbf{B}^{(k)}$ to satisfy the secant equation:

$$\mathbf{B}^{(k+1)}\boldsymbol{z}_k = \boldsymbol{s}_k, \quad \text{where} \;\; \boldsymbol{s}_k = \boldsymbol{\rho}_{k+1} - \boldsymbol{\rho}_k, \;\; \boldsymbol{z}_k = \nabla D_M(\boldsymbol{\rho}_{k+1}) - \nabla D_M(\boldsymbol{\rho}_k) \tag{D.2}$$

Only the secant equation is not enough to determine a unique solution of $\mathbf{B}^{(k+1)}$. Thus we impose additional conditions as well as the secant equation to $\mathbf{B}^{(k+1)}$, that $\mathbf{B}^{(k+1)}$ should be symmetric and the difference between successive $\mathbf{B}^{(k)}$ and $\mathbf{B}^{(k+1)}$ should have low rank. In particular, $\mathbf{B}^{(k+1)}$ is specified by the following conditions:

$$\mathbf{B}^{(k+1)} = \arg\min_{\mathbf{B}} \|\mathbf{B} - \mathbf{B}^{(k)}\|_F \quad \text{s.t.} \quad \mathbf{B} = \mathbf{B}^T, \;\; \mathbf{B}\boldsymbol{z}_k = \boldsymbol{s}_k \tag{D.3}$$

Then the BFGS updating formula is given as:

$$\mathbf{B}^{(k+1)} = (\mathbb{1}_N - t_k \boldsymbol{s}_k \boldsymbol{z}_k^T)\mathbf{B}^{(k)}(\mathbb{1}_N - t_k \boldsymbol{z}_k \boldsymbol{s}_k^T) + t_k \boldsymbol{s}_k \boldsymbol{s}_k^T \tag{D.4}$$

where

$$t_k = \frac{1}{\boldsymbol{z}_k^T \boldsymbol{s}_k} \tag{D.5}$$

Given an initial inverse Hessian approximation $\mathbf{B}^{(0)}$, by repeating equation (D.4), the BFGS updating formula reads:

$$\begin{aligned}
\mathbf{B}^{(k)} =& \big(\mathbf{S}_{k-1}^T \cdots \mathbf{S}_{k-m(k)}^T\big)\mathbf{B}^{(0)}\big(\mathbf{S}_{k-m(k)} \cdots \mathbf{S}_{k-1}\big) \\
&+ t_{k-m(k)}\big(\mathbf{S}_{k-1}^T \cdots \mathbf{S}_{k-m(k)+1}^T\big)\boldsymbol{s}_{k-m(k)}\boldsymbol{s}_{k-m(k)}^T\big(\mathbf{S}_{k-m(k)+1} \cdots \mathbf{S}_{k-1}\big) \\
&+ t_{k-m(k)+1}\big(\mathbf{S}_{k-1}^T \cdots \mathbf{S}_{k-m(k)+2}^T\big)\boldsymbol{s}_{k-m(k)+1}\boldsymbol{s}_{k-m(k)+1}^T\big(\mathbf{S}_{k-m(k)+2} \cdots \mathbf{S}_{k-1}\big) \\
&+ \cdots \\
&+ t_{k-1}\boldsymbol{s}_{k-1}\boldsymbol{s}_{k-1}^T
\end{aligned} \tag{D.6}$$

in which

$$\mathbf{S}_k = \mathbb{1}_N - t_k \boldsymbol{s}_k \boldsymbol{z}_k^T \tag{D.7}$$

In the BFGS updating scheme, $m(k) = k$ and the initial inverse Hessian approximation $\mathbf{B}^{(0)}$ is often set a constant multiply the identify matrix.

When the dimension of $\boldsymbol{\rho}$ is large, storing and manipulating $\mathbf{B}^{(k)}$ is prohibitive. To circumvent this problem, we set a number M_2 and only the recent M_2 pairs of $\{\boldsymbol{s}_k, \boldsymbol{z}_k\}$ are stored, which gives limited-memory BFGS (L-BFGS) algorithm. The updating of $\mathbf{B}^{(k)}$ in L-BFGS is same as shown in (D.6) with $m(k) = \min(k, M_2)$ and initial approximation $\mathbf{B}^{(0)}$ could vary from iteration to iteration. For instance, $\mathbf{B}^{(0)}$ is typically replaced by $\mathbf{B}_k^{(0)}$ where:

$$\mathbf{B}_k^{(0)} = \gamma_k \mathbb{1}_N, \quad \text{with} \quad \gamma_k = \frac{\boldsymbol{z}_{k-1}^T \boldsymbol{s}_{k-1}}{\boldsymbol{z}_{k-1}^T \boldsymbol{z}_{k-1}} \tag{D.8}$$

The L-BFGS algorithm does not require to construct $\mathbf{B}^{(k)}$ explicitly. In practice, the product $\mathbf{B}^{(k)} \nabla D_M(\boldsymbol{\rho}_k)$, rather than $\mathbf{B}^{(k)}$ is updated. Specially, this is done by a two-loop recursion algorithm as shown in algorithm 6. A complete statement of L-BFGS algorithm is presented in algorithm 7.

Algorithm 6: L-BFGS two-loop recursion

$\boldsymbol{g} \leftarrow \nabla D_M(\boldsymbol{\rho}_k)$;
for $i \in \{k-1, \cdots, k-M_2\}$ **do**
  $\mu_i \leftarrow t_i \boldsymbol{s}_i^T \boldsymbol{g}$; `// store` μ_i
  $\boldsymbol{g} \leftarrow \boldsymbol{g} - \mu_i \boldsymbol{z}_i$;
end
$r = \mathbf{B}_k^{(0)} \boldsymbol{g}$;
for $i \in \{k-M_2, \cdots, k-1\}$ **do**
  $\beta \leftarrow t_i \boldsymbol{z}_i^T \boldsymbol{r}$;
  $\boldsymbol{r} \leftarrow \boldsymbol{r} + \boldsymbol{s}_i(\mu_i - \beta)$;
end
$\mathbf{B}^{(k)} \nabla D_M(\boldsymbol{\rho}_k) = \boldsymbol{r}$;

Algorithm 7: L-BFGS algorithm

Initialize starting point $\boldsymbol{\rho}_0$, integer $M_2 > 0$;
$k \leftarrow 0$;
while *stopping criterion is not met* **do**
  Choose $\mathbf{B}_k^{(0)}$; `// for example using (D.8)`
  Compute $\boldsymbol{d}_k = -\mathbf{B}^{(k)} \nabla D_M(\boldsymbol{\rho}_k)$ by two-loop recursion shown in algorithm 6 ;
  Update $\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k + \alpha_k \boldsymbol{d}_k$ where α_k is chosen to satisfy Wolfe conditions;
  if $k > M_2$ **then**
    Discard the pairs $\{\boldsymbol{s}_{k-m}, \boldsymbol{z}_{k-m}\}$;
    Compute and store $\boldsymbol{s}_k = \boldsymbol{\rho}_{k+1} - \boldsymbol{\rho}_k, \boldsymbol{z}_k = \nabla D_M(\boldsymbol{\rho}_{k+1}) - \nabla D_M(\boldsymbol{\rho}_k)$;
    $k \leftarrow k + 1$;
  end
end

D.2 L-BFGS-B algorithm

Since we know that the fluorescence object is non-negative, we can add positivity constraint $\boldsymbol{\rho} \geq 0$ to the marginal approach (3.50) and then solve the optimization problem with bound constrained limited memory BFGS algorithm (L-BFGS-B) [BLNZ95]. The L-BFGS-B algorithm is designed to iteratively solve optimization problem $\min f(\boldsymbol{x})$ with bound constraints $\boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}$, where $\boldsymbol{l}, \boldsymbol{u}, \boldsymbol{x} \in \mathbb{R}^N$. At the beginning of each iteration, the variable value $\boldsymbol{x}_k$, the function value $f(\boldsymbol{x}_k)$, the gradient $\boldsymbol{g}_k$ and an approximation of Hessian matrix $\mathbf{H}_k$ obtained as in L-BFGS method are given. Then the function can be approximated by the quadratic model:

$$m_k(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^T\mathbf{H}_k(\boldsymbol{x} - \boldsymbol{x}_k) \tag{D.9}$$

The L-BFGS-B algorithm approximately minimize $m_k(\boldsymbol{x})$ subject to the bound constraints $\boldsymbol{l} \leq \boldsymbol{x} \leq \boldsymbol{u}$. This is done by firstly find a generalized Cauchy point $\boldsymbol{x}^c$, and then do a subspace minimization along the free variables.

Generalized Cauchy point The generalized Cauchy point is defined as the first local minimizer of the piece-wise quadratic $m_k(x(t))$ where $x(t)$ is a piece-wise linear path obtained by projecting the steepest descent direction onto the feasible region:

$$x(t) = \text{proj}(\boldsymbol{x}_k - t\boldsymbol{g}_k, \boldsymbol{l}, \boldsymbol{u}) \tag{D.10}$$

The projection operation is defined as:

$$\text{proj}(\boldsymbol{x}_k - t\boldsymbol{g}_k, \boldsymbol{l}, \boldsymbol{u})_i = \begin{cases} l_i, & x_i < l_i \\ x_i, & l_i \leq x_i \leq u_i \\ u_i, & x_i > u_i \end{cases} \tag{D.11}$$

To simplify notation, we use $\boldsymbol{x}^0$ to denote $\boldsymbol{x}_K$ and $\boldsymbol{g}$ for $\boldsymbol{g}_k$ in this subsection. Each coordinate $x_i(t)$ of the piece-wise linear path is given by:

$$x_i(t) = x_i^0 - tg_i, \quad t \in [0, t_i] \tag{D.12}$$

The breakpoint t_i is:

$$t_i = \begin{cases} (x_i^0 - u_i)/g_i, & g_i < 0 \\ (x_i^0 - l_i)/g_i, & g_i > 0 \\ \infty, & otherwise \end{cases} \tag{D.13}$$

Then we sort $\{t_i \mid i = 1, \cdots, N\}$ in a ordered set $\{t^j \mid t^j \leq t^{j+1}, j = 1, \cdots, N\}$. The piece-wise linear path $x(t)$ can be reexpressed using t_i as:

$$x_i(t) = \begin{cases} x_i^0 - tg_i, & t \leq t_i \\ x_i^0 - t_i g_i, & otherwise \end{cases} \tag{D.14}$$

In the line interval $[x(t^{j-1}), x(t^j)]$, the quadratic model $m_k(x(t))$ can be written in $\Delta t = t - t^{j-1}$ as:

$$m_k^j(\Delta t) = f_{j-1} + f'_{j-1}\Delta t + \frac{1}{2}f''_{j-1}\Delta t^2 \tag{D.15}$$

where

$$\begin{aligned} f_{j-1} &= f(\boldsymbol{x}^0) + \boldsymbol{g}^T(x(t^{j-1}) - \boldsymbol{x}^0) + \frac{1}{2}(x(t^{j-1}) - \boldsymbol{x}^0)^T \mathbf{H}_k (x(t^{j-1}) - \boldsymbol{x}^0) \\ f'_{j-1} &= \boldsymbol{g}^T \boldsymbol{d}^{j-1} + (\boldsymbol{d}^{j-1})^T \mathbf{H}_k (x(t^{j-1}) - \boldsymbol{x}^0) \\ f''_{j-1} &= (\boldsymbol{d}^{j-1})^T \mathbf{H}_k \boldsymbol{d}^{j-1} \end{aligned} \tag{D.16}$$

The definition of $\boldsymbol{d}^{j-1}$ is:

$$d_i^{j-1} = \begin{cases} -g_i, & t^{j-1} < t_i \\ 0, & otherwise \end{cases} \tag{D.17}$$

Setting the gradient of $m_k^j(\Delta t)$ to zero, we obtain $\Delta t = -f'_{j-1}/f''_{j-1}$.

Subspace minimization Once the generalized Cauchy point $\boldsymbol{x}^c$ is obtained, the variables whose value are at the lower or upper bounds, comprising the active set $\mathcal{A}(\boldsymbol{x}^c)$, are held fixed. The quadratic model is minimized over the subspace of free variables in $\boldsymbol{x}^c$. A few methods could be used to solve the minimization problem. For example, a direct primal method based on the Sherman-Morrison-Woodbury formula gives a unconstrained solution $\bar{\boldsymbol{x}}_{k+1}$ in subspace. We then find the next iteration value by a line search method along the direction $\boldsymbol{d}_k = \bar{\boldsymbol{x}}_{k+1} - \boldsymbol{x}_k$ that satisfy the strong Wolfe conditions [NW06]. To prevent the line search method from generating infeasible points, we define a maximum step length as the step to the closest bounds alone the search direction.

This procedure is repeated until the convergence condition is reached.

Appendix E

Gradient of Kullback-Leibler divergence

Here we derive the expression of the gradient of Kullback-Leibler divergence as shown in (3.50)

$$D_M(\boldsymbol{\rho}) = \frac{1}{2}\log|\boldsymbol{\Gamma}_y| + \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\Gamma}_y^{-1}\mathbf{W}\right) + K \tag{E.1}$$

with $\mathbf{W} = \frac{1}{M}\mathbf{V}\mathbf{V}^{\mathrm{t}}$ and $\mathbf{V} = (\boldsymbol{y}_1 - \boldsymbol{\mu}_y|\cdots|\boldsymbol{y}_M - \boldsymbol{\mu}_y)$. Our derivation of the gradient are based on the following equations [PP+08, Section 2] :

$$\begin{aligned}
\nabla_\theta \log|\mathbf{A}| &= \mathrm{Tr}(\mathbf{A}^{-1}(\nabla_\theta \mathbf{A})) \\
\nabla_\theta(\mathbf{A}^{-1}) &= -\mathbf{A}^{-1}(\nabla_\theta \mathbf{A})\mathbf{A}^{-1} \\
\nabla_\theta(\mathbf{A}\mathbf{B}) &= (\nabla_\theta \mathbf{A})\mathbf{B} + \mathbf{A}(\nabla_\theta \mathbf{B}) \\
\nabla_\theta \mathrm{Tr}(\mathbf{A}) &= \mathrm{Tr}(\nabla_\theta(\mathbf{A})) \\
\nabla_\theta(\mathbf{A}^T) &= (\nabla_\theta \mathbf{A})^T
\end{aligned} \tag{E.2}$$

where $\mathbf{A}$ and $\mathbf{B}$ are two matrix depending on a real scalar parameter θ. From these relations, we have:

$$\partial_n D_M(\boldsymbol{\rho}) = \frac{1}{2}\mathrm{Tr}(\boldsymbol{\Gamma}_y^{-1}(\partial_n \boldsymbol{\Gamma}_y)) + \frac{1}{2}\mathrm{Tr}\Big((\partial_n \boldsymbol{\Gamma}_y^{-1})\mathbf{W} + \boldsymbol{\Gamma}_y^{-1}\partial_n \mathbf{W}\Big) \tag{E.3}$$

with $\partial_n = \nabla_{\rho_n}$ and

$$\nabla D_M(\boldsymbol{\rho}) = \mathrm{vec}\big\{\partial_n D_M(\boldsymbol{\rho})\big\} \tag{E.4}$$

Where $\mathrm{vec}\{v_n\} = [v_1, \cdots, v_N]^T$. According to (E.2,E.3), the expression of $\partial_n \boldsymbol{\Gamma}_y$ and $\partial_n \mathbf{W}$ are required to derive the expression of $\nabla D_M(\boldsymbol{\rho})$. Let $\boldsymbol{e}_n$ be the n-th canonical vector, $\boldsymbol{h}_n$ be the n-th column of $\mathbf{H}$ and $\mathbf{1}_M = (1, \cdots, 1)^T \in \mathbb{R}^M$. From (3.35) we have:

$$\begin{aligned}
\partial_n \boldsymbol{\Gamma}_y &= \mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s \partial_n(\mathbf{R}\mathbf{H}^T) + \partial_n(\mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s)\mathbf{R}\mathbf{H} \\
&= \mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s \partial_n(\mathbf{R}\mathbf{H}^T) + \big(\mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s \partial_n(\mathbf{R}\mathbf{H}^T)\big)^T \\
&= \mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s \boldsymbol{e}_n \boldsymbol{h}_n^T + (\mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s \boldsymbol{e}_n \boldsymbol{h}_n^T)^T
\end{aligned} \tag{E.5}$$

and similarly,

$$\partial_n \mathbf{W} = -\frac{I_0}{M}\big(\mathbf{V}\mathbf{1}_M \boldsymbol{h}_n^T + (\mathbf{V}\mathbf{1}_M \boldsymbol{h}_n^T)^T\big) \tag{E.6}$$

The derivative of the three terms in (E.3) could now be obtained. According to the identity:

$$\mathrm{Tr}(\mathbf{B}\boldsymbol{a}\boldsymbol{a}^T) = \boldsymbol{a}^T\mathbf{B}\boldsymbol{a} \tag{E.7}$$

with matrix $\mathbf{B} \in \mathbb{R}^{N\times N}$, $\boldsymbol{a} \in \mathbb{R}^N$, we have:

$$\begin{aligned}
\mathrm{Tr}\big(\boldsymbol{\Gamma}_y^{-1}(\partial_n \boldsymbol{\Gamma}_y)\big) &= 2\mathrm{Tr}(\boldsymbol{\Gamma}_y^{-1}\mathbf{H}\mathbf{R}\boldsymbol{\Gamma}_s \boldsymbol{e}_n \boldsymbol{h}_n^T) \\
&= 2\boldsymbol{e}_n^T \mathbf{X}\boldsymbol{h}_n
\end{aligned} \tag{E.8}$$

where we define $\mathbf{X} = \boldsymbol{\Gamma}_s \mathbf{R} \mathbf{H}^T \boldsymbol{\Gamma}_y^{-1}$. For the second and third term in (E.3)

$$\mathrm{Tr}\Big((\partial_n \boldsymbol{\Gamma}_y^{-1})\mathbf{W}\Big) = -2\boldsymbol{e}_n^T(\boldsymbol{\Gamma}_y^{-1}\mathbf{W}\mathbf{X})\boldsymbol{h}_n \tag{E.9}$$

and

$$\mathrm{Tr}(\boldsymbol{\Gamma}_y^{-1}\partial_n \mathbf{W}) = -\frac{2I_0}{M}\boldsymbol{h}_n^T \boldsymbol{\Gamma}_y^{-1}\mathbf{V}\mathbf{1}_M \tag{E.10}$$

To obtain the full gradient of $D_M(\boldsymbol{\rho})$, we deduce from (E.8) that

$$\begin{aligned}\mathrm{vec}\Big\{\mathrm{Tr}\big(\boldsymbol{\Gamma}_y^{-1}(\partial_n \boldsymbol{\Gamma}_y)\big)\Big\} &= 2\big((\mathbf{X}\mathbf{H}) \circ \mathbb{I}_N\big)\mathbf{1}_N \\ &= 2\big((\mathbf{H}^T\boldsymbol{\Gamma}_y^{-1}\mathbf{H}) \circ \boldsymbol{\Gamma}_s\big)\boldsymbol{\rho}\end{aligned} \tag{E.11}$$

where in the second step we have used the identity:

$$\big((\mathbf{A}\mathrm{diag}(\boldsymbol{\rho})\mathbf{B}^T) \circ \mathbb{I}_N\big)\mathbf{1}_N = (\mathbf{A} \circ \mathbf{B})\boldsymbol{\rho} \tag{E.12}$$

Similarly, we have

$$\mathrm{vec}\Big\{\mathrm{Tr}\Big((\partial_n \boldsymbol{\Gamma}_y^{-1})\mathbf{W}\Big)\Big\} = -2\Big((\boldsymbol{\Omega}^T\mathbf{W}\boldsymbol{\Omega}) \circ \boldsymbol{\Gamma}_s\Big)\boldsymbol{\rho} \tag{E.13}$$

and

$$\mathrm{vec}\Big\{\mathrm{Tr}(\boldsymbol{\Gamma}_y^{-1}\partial_n \mathbf{W})\Big\} = -\frac{2}{M}I_0\boldsymbol{\Omega}^T\mathbf{V}\mathbf{1}_M \tag{E.14}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Gamma}_y^{-1}\mathbf{H}$. Plugging (E.3,E.11,E.13,E.14) into (E.4) reads:

$$\nabla D_M(\boldsymbol{\rho}) = \Big(\big(\boldsymbol{\Omega}^{\mathrm{t}}\left(\boldsymbol{\Gamma}_y - \mathbf{W}\right)\boldsymbol{\Omega}\big) \circ \boldsymbol{\Gamma}_s\Big)\,\boldsymbol{\rho} - \frac{I_0}{M}\boldsymbol{\Omega}^{\mathrm{t}}\mathbf{V}\mathbf{1}_M \tag{E.15}$$

Bibliography

[3D 17] 3D Deconvolution Microscopy EPFL. 3d reference datasets, 2017. [Online; accessed 19-April-2018].

[ABDF10] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356, 2010.

[ABDF11] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2011.

[AGJ+13] R. Ayuk, H. Giovannini, A. Jost, E. Mudry, J. Girard, T. Mangeat, N. Sandeau, R. Heintzmann, K. Wicker, K. Belkebir, and A. Sentenac. Structured illumination fluorescence microscopy with distorted excitations using a filtered blind-SIM algorithm. *Optics Letters*, 38(22):4723–4726, Nov 2013.

[Alb11] J. R. Albani. *Structure and dynamics of macromolecules: absorption and fluorescence studies*. Elsevier, 2011.

[ALBA13] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone. A tutorial on speckle reduction in synthetic aperture radar images. *IEEE Geoscience and remote sensing magazine*, 1(3):6–35, 2013.

[AM16] R. Abergel and L. Moisan. The shannon total variation. *Journal of Mathematical Imaging and Vision*, pages 1–30, 2016.

[BAS+15] F. Bergermann, L. Alber, S. J. Sahl, J. Engelhardt, and S. W. Hell. 2000-fold parallelized dual-color sted fluorescence nanoscopy. *Optics express*, 23(1):211–223, 2015.

[BD15a] P. J. Bickel and K. A. Doksum. *Mathematical statistics: basic ideas and selected topics*, volume 2. CRC Press, 2015.

[BD15b] P. J. Bickel and K. A. Doksum. *Mathematical Statistics: Basic Ideas and Selected Topics, volume I*, volume 117. CRC Press, 2015.

[BLNZ95] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[BM13] H. H. Barrett and K. J. Myers. *Foundations of image science*. John Wiley & Sons, 2013.

[BT09] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[CGA+16] T. Chaigne, J. Gateau, M. Allain, O. Katz, S. Gigan, A. Sentenac, and E. Bossy. Super-resolution photoacoustic fluctuation imaging with multiple speckle illumination. *Optica*, 3(1):54–57, 2016.

[CIM11] E. Chouzenoux, J. Idier, and S. Moussaoui. A majorize–minimize strategy for subspace optimization applied to image restoration. *IEEE Transactions on image processing*, 20(6):1517 – 1528, 2011.

[CLL11] A. Chambolle, S. E. Levine, and B. J. Lucier. An upwind finite-difference method for total variation–based image smoothing. *SIAM Journal on Imaging Sciences*, 4(1):277–299, 2011.

[Con13] L. Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013.

[CP11] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

[Dav12] P. J. Davis. *Circulant matrices*. American Mathematical Soc., 2012.

[DCD12] D. Dacunha-Castelle and M. Duflo. *Probability and statistics*, volume 2. Springer Science & Business Media, 2012.

[DCI+09] T. Dertinger, R. 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*, 106(52):22287–22292, 2009.

[DCV+10] T. Dertinger, R. Colyer, R. Vogel, J. Enderlein, and S. Weiss. Achieving increased resolution and more pixels with superresolution optical fluctuation imaging (sofi). *Optics express*, 18(18):18875–18885, 2010.

[DH02] M. Dyba and S. W. Hell. Focal spots of size $\lambda/23$ open up far-field florescence microscopy at 33 nm axial resolution. *Physical review letters*, 88(16):163901, 2002.

[Dia01] A. Diaspro. Confocal and two-photon microscopy: foundations, applications and advances. *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances, by Alberto Diaspro (Editor), pp. 576. ISBN 0-471-40920-0. Wiley-VCH, November 2001.*, page 576, 2001.

[DJHS92] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern. Maximum entropy and the nearly black object. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 41–81, 1992.

[DYZ12] W. Deng, W. Yin, and Y. Zhang. Group sparse optimization by alternating direction method. Technical report, RICE UNIV HOUSTON TX DEPT OF COMPUTATIONAL AND APPLIED MATHEMATICS, 2012.

[Fre00] D. H. Fremlin. *Measure theory*, volume 4. Torres Fremlin, 2000.

[Goo05] J. Goodman. *Introduction to Fourier Optics*. Roberts & Company Publishers, 2005.

[Goo07] J. W. Goodman. *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2007.

[Goo15] J. W. Goodman. *Statistical optics*. John Wiley & Sons, 2015.

[GSC+08] M. G. Gustafsson, L. Shao, P. M. Carlton, C. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical journal*, 94(12):4957–4970, 2008.

[Gus00] M. G. L. Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy*, 2000.

[Gus05] M. G. Gustafsson. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37):13081–13086, 2005.

[HAB13] P. Hiremath, P. T. Akkasaligar, and S. Badiger. Speckle noise reduction in medical ultrasound images. In *Advancements and Breakthroughs in Ultrasound Imaging*. InTech, 2013.

[Haj15] B. Hajek. *An exploration of random processes for engineers*. Cambridge University Press, 2015.

[HBGS10] O. Haeberlé, K. Belkebir, H. Giovaninni, and A. Sentenac. Tomographic diffractive microscopy: basics, techniques and perspectives. *Journal of Modern Optics*, 57(9):686–699, 2010.

[HGM06] S. T. Hess, T. P. K. Girirajan, and M. D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272, 2006.

[HGvdH+14] N. Hafi, M. Grunwald, L. S. van den Heuvel, T. Aspelmeier, J. Chen, M. Zagrebelsky, O. M. Schütte, C. Steinem, M. Korte, A. Munk, et al. Fluorescence nanoscopy by polarization modulation and polarization angle narrowing. *Nature methods*, 11(5):579–584, 2014.

[HLM+17] Y. Hu, C. Li, K. Meng, J. Qin, and X. Yang. Group sparse optimization via lp, q regularization. *Journal of Machine Learning Research*, 18(30):1–52, 2017.

[HNO06] P. C. Hansen, J. G. Nagy, and D. P. O'leary. *Deblurring images: matrices, spectra, and filtering*. SIAM, 2006.

[HS91] R. M. Haralock and L. G. Shapiro. *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc., 1991.

[HW94] S. W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782, 1994.

[HWBZ08] B. Huang, W. Wang, M. Bates, and X. Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 319(5864):810–813, 2008.

[ILA+17] J. Idier, S. Labouesse, M. Allain, P. Liu, S. Bourguignon, and A. Sentenac. On the super-resolution capacity of imagers using unknown speckle illuminations. *IEEE Transactions on Computational Imaging*, 2017.

[JGL+08] M. F. Juette, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf. Three-dimensional sub–100 nm resolution fluorescence microscopy of thick samples. *Nature methods*, 5(6):527–529, 2008.

[JTF+15] A. Jost, E. Tolstik, P. Feldmann, K. Wicker, A. Sentenac, and R. Heintzmann. Optical sectioning and high resolution in single-slice structured illumination microscopy by thick slice blind-sim reconstruction. *PloS one*, 10(7):e0132174, 2015.

[JW04] S.-h. Jiang and J. G. Walker. Experimental confirmation of non-scanning fluorescence confocal microscopy using speckle illumination. *Optics communications*, 238(1):1–12, 2004.

[KC96] G. S. Kino and T. R. Corle. *Confocal scanning optical microscopy and related imaging systems*. Academic Press, 1996.

[KCG+09] P. Kner, B. B. Chhun, E. R. Griffis, L. Winoto, and M. G. Gustafsson. Super-resolution video microscopy of live cells by structured illumination. *Nature methods*, 6(5):339–342, 2009.

[KPR+15] M. Kim, C. Park, C. Rodriguez, Y. Park, and Y.-H. Cho. Superresolution imaging with optical fluctuation using speckle patterns illumination. *Scientific reports*, 5, 2015.

[Kul97] S. Kullback. *Information theory and statistics*. Courier Corporation, 1997.

[LCM08] D. Lim, K. K. Chu, and J. Mertz. Wide-field fluorescence sectioning with hybrid speckle and uniform-illumination microscopy. *Optics letters*, 33(16):1819–1821, 2008.

[LFCM11] D. Lim, T. N. Ford, K. K. Chu, and J. Mertz. Optically sectioned in vivo imaging with speckle illumination hilo microscopy. *Journal of biomedical optics*, 16(1):016014–016014, 2011.

[LNI+17] S. Labouesse, A. Negash, J. Idier, S. Bourguignon, T. Mangeat, P. Liu, A. Sentenac, and M. Allain. Joint reconstruction strategy for structured illumination microscopy with unknown illuminations. *IEEE Transactions on Image Processing*, 2017.

[LR83] R. J. Little and D. B. Rubin. On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, 37(3):218–220, 1983.

[MBG+12] E. Mudry, K. Belkebir, J. Girard, J. Savatier, E. Le Moal, C. Nicoletti, M. Allain, and A. Sentenac. Structured illumination microscopy using unknown speckle patterns. *Nature Photonics*, 6(5):312–315, 2012.

[MCW05] D. Malioutov, M. Cetin, and A. S. Willsky. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE transactions on signal processing*, 53(8):3010–3022, 2005.

[Mer10] J. Mertz. *Introduction to optical microscopy*, volume 138. Roberts, 2010.

[MJK+13] J. Min, J. Jang, D. Keum, S.-W. Ryu, C. Choi, K.-H. Jeong, and J. C. Ye. Fluorescent microscopy beyond diffraction limits using speckle illumination and joint support recovery. *Scientific reports*, 3:2075, 2013.

[Mun00] J. R. Munkres. *Topology*. Prentice Hall, 2000.

[NW06] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[Ped12] G. K. Pedersen. *Analysis now*, volume 118. Springer Science & Business Media, 2012.

[PGK+17] M. Pascucci, S. Ganesan, O. Katz, V. Emiliani, and M. Guillon. Three-dimensional super-resolution microscopy with speckle-saturated fluorescence excitation. *arXiv preprint arXiv:1710.05056*, 2017.

[PP+08] K. B. Petersen, M. S. Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7:15, 2008.

[PTB+09] S. R. P. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. Moerner. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences*, 106(9):2995–2999, 2009.

[RBZ06] M. J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793–796, 2006.

[Ree62] I. Reed. On a moment theorem for complex gaussian processes. *IRE Transactions on Information Theory*, 8(3):194–195, 1962.

[Roc15] R. T. Rockafellar. *Convex analysis*. Princeton university press, 2015.

[Ser02] D. Serre. Matrices: Theory and applications, 2002.

[SHYT95] L. Song, E. Hennink, I. T. Young, and H. J. Tanke. Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophysical journal*, 68(6):2588–2600, 1995.

[SMG+91] M. Sutton, S. Mochrie, T. Greytak, S. Nagler, L. Berman, G. Held, and G. Stephenson. Observation of speckle by diffraction with coherent x-rays. *Nature*, 352(6336):608, 1991.

[Str84] N. Streibl. Depth transfer by an imaging system. *Journal of Modern Optics*, 31(11):1233–1241, 1984.

[TGS06] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[Tro06] J. A. Tropp. Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. *Signal Processing*, 86(3):589–602, 2006.

[VdV00] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[VHM07] C. Ventalon, R. Heintzmann, and J. Mertz. Dynamic speckle illumination microscopy with wavelet prefiltering. *Optics letters*, 32(11):1417–1419, 2007.

[VM05] C. Ventalon and J. Mertz. Quasi-confocal fluorescence sectioning with dynamic speckle illumination. *Optics letters*, 30(24):3350–3352, 2005.

[VM06] C. Ventalon and J. Mertz. Dynamic speckle illumination microscopy with translated versus randomized speckle patterns. *Optics express*, 14(16):7198–7209, 2006.

[Wal01] J. G. Walker. Non-scanning confocal fluorescence microscopy using speckle illumination. *Optics communications*, 189(4):221–226, 2001.

[Wic10] K. Wicker. *Increasing resolution and light efficiency in fluorescence microscopy*. PhD thesis, King's College London, 2010.

[WMKH09] D. Wildanger, R. Medda, L. Kastrup, and S. Hell. A compact sted microscope providing 3d nanoscale resolution. *Journal of microscopy*, 236(1):35–43, 2009.

[WRN11] D. P. Wipf, B. D. Rao, and S. Nagarajan. Latent variable bayesian models for promoting sparsity. *IEEE Transactions on Information Theory*, 57(9):6236–6255, 2011.

[WYYZ08] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[XW02] C. Xu and W. W. Webb. Multiphoton excitation of molecular fluorophores and nonlinear laser microscopy. In *Topics in fluorescence spectroscopy*, pages 471–540. Springer, 2002.

[YTW17] L.-H. Yeh, L. Tian, and L. Waller. Structured illumination microscopy with unknown patterns and a statistical prior. *Biomedical optics express*, 8(2):695–711, 2017.

List of publication

Journal Articles

— Jérôme Idier, Simon Labouesse, Marc Allain, Penghuan Liu, Sebastien Bourguignon and Anne Sentenac: **On the super-resolution capacity of imagers using unknown speckle illuminations**. – *IEEE Transactions on Computational Imaging*, 2017

— Simon Labouesse, Awoke Negash, Jérôme Idier, Sebastien Bourguignon,Thomas Mangeat, Penghuan Liu, Anne Sentenac, Marc Allain: **Joint reconstruction strategy for structured illumination microscopy with unknown illuminations**. – *IEEE Transactions on Image Processing*, 2017

Conference Papers

— Penghuan Liu, Sebastien Bourguignon, Jérôme Idier, Marc Allain, Simon Labouesse and Anne Sentenac: **A marginal image reconstruction approach in fluorescence microscopy with pseudo-random illumination patterns**. – *Mathematics in Imaging*, 2016, Heidelberg, Germany

— Penghuan Liu, Sebastien Bourguignon, Jérôme Idier, Marc Allain, Simon Labouesse and Anne Sentenac: **Minimum contrast estimation for super-resolution fluorescence microscopy using speckle patterns**. – *GRETSI*, 2017, Juan-Les-Pins, France

Journal Articles in prepare

— Penghuan Liu, Sebastien Bourguignon, Marc Allain, Simon Labouesse, Anne Sentenac and Jérôme Idier: **Super-resolution image reconstruction in fluorescence microscopy with unknown speckle patterns via constrained $\ell_{p,q}$ norm minimization**.

— Penghuan Liu, Sebastien Bourguignon, Marc Allain, Simon Labouesse, Anne Sentenac, Jérôme Idier: **A practical marginal estimation algorithm for structured illumination microscopy using unknown speckle patterns**.