

Contributions à la microscopie à fluorescence en imagerie biologique : modélisation de la PSF, restauration d'images et détection super-résolutive

Bo Zhang

▶ **To cite this version:**

Bo Zhang. Contributions à la microscopie à fluorescence en imagerie biologique : modélisation de la PSF, restauration d'images et détection super-résolutive. Mathématiques [math]. Télécom ParisTech, 2007. Français. NNT : . pastel-00003273

HAL Id: pastel-00003273
https://pastel.hal.science/pastel-00003273

Submitted on 11 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





École Doctorale
d'Informatique,
Télécommunications et
Électronique de Paris

THÈSE

présentée pour obtenir le grade de docteur
de l'École Nationale Supérieure des Télécommunications

Spécialité : **Signal et Images**

ZHANG Bo

Contributions à la microscopie à fluorescence en imagerie biologique : modélisation de la PSF, restauration d'images et détection super-résolutive

Soutenue le 30 novembre 2007 devant le jury composé de

Jean-Christophe	OLIVO-MARIN	Directeur
Henri	MAITRE	Président et co-directeur
Raimund	OBER	Rapporteur
Claude	BOCCARA	Rapporteur
Dimitri	VAN DE VILLE	Examinateur
Josiane	ZERUBIA	Invitée
Jean-Luc	STARCK	Invité





École Doctorale
d'Informatique,
Télécommunications et
Électronique de Paris

DISSERTATION

presented in fulfillment of the requirements
for the degree of Doctor of Philosophy
of l'École Nationale Supérieure des Télécommunications

Domain : **Signal and Image Processing**

ZHANG Bo

Contributions to fluorescence microscopy in biological imaging: PSF modeling, image restoration, and super-resolution detection

Defensed on November 30, 2007

Jury

Jean-Christophe	OLIVO-MARIN	Advisor
Henri	MAITRE	President and co-advisor
Raimund	OBER	Reviewer
Claude	BOCCARA	Reviewer
Dimitri	VAN DE VILLE	Examiner
Josiane	ZERUBIA	Invited
Jean-Luc	STARCK	Invited

Acknowledgments

First, I would like to express my sincere gratitude to my thesis advisor Jean-Christophe Olivo-Marin, head of Quantitative Image Analysis Group of Institut Pasteur. I have been in his laboratory during my engineer intern and he has proposed me to commence this Ph.D. project. His understanding, encouragement, support and guidance have made the present thesis possible.

I am deeply grateful to my co-adviser, Henri Maître, professor of the department of signal and image processing of ENST Paris and deputy director of ENST Paris. He has given me many impressive courses in image processing during my study at ENST. His intelligence and passion motivated me to pursuit my research in the signal and image processing domain. He has given me his valuable comments, suggestions, criticisms, and encouragement all along my work.

My sincere thanks are due to the jury members, Jean-Christophe Olivo-Marin, Henri Maître, Raimund Ober, professor at the department of Electrical Engineering of University of Texas at Dallas, Claude Boccara, head of Physical Optics Laboratory and scientific director of ESPCI, Dimitri Van De Ville, responsible of the Signal Processing Antenna of CIBM at the University Hospital of Geneva and researcher at Biomedical Imaging Group of EPFL, Josiane Zerubia, head of the Ariana project at INRIA, and Jean-Luc Starck, researcher at Astrophysics service of CEA-Saclay, who have taken their time for reviewing this work in details and providing valuable advices, comments and criticisms.

I am greatly indebted to my collaborators Josiane Zerubia, Jean-Luc Starck, Jalal Fadili, associated professor at GREYC of ENSICAEN, Christophe Zimmer, researcher at UAIQ of Institut Pasteur, Claude Boccara, Benoît Dubertret and Frédéric Chasles, researcher and Ph.D. student of Physical Optics Laboratory respectively, Musa M. Mhlanga, researcher at UBCN of Institut Pasteur, and Jost Enninga, researcher at UPMM of Institut Pasteur. They have worked with me in some of my research subjects, shared with me their expert knowledge, and provided me with their great help, detailed and constructive comments, and valuable ideas.

I am very grateful to Jean-Luc Starck, Musa M. Mhlanga, Jost Enninga, Pierre Thérizols, Ph.D. student at UGML of Institut Pasteur for providing most of the data sets in this thesis.

I am very glad to participate into the Franco-Israeli P2R project, the GLAST project,

and the ANR DETECFINE project, of which the subjects have partly contributed to this work. Thus, I wish to express my great obligations to my collaborators Jean-Luc Starck, Jalal Fadili, Seth W. Digel, Jim Chiang, Josiane Zerubia, Laure Blanc-Féraud, Caroline Chaux, Praveen Pankajakshan, Zvi Kam, Arie Feuer, Daniel Duclos, Yann Le Guilloux, Gilles Aubert, Alexis Baudour, Luis Lopes Neves De Almeida, David Chiron, Jean-François Aujol for their valuable discussions, criticisms and suggestions, and their support.

I wish to express my warm thanks to Jie Huang, researcher at Fejer group of Stanford University, James Pawley, researcher at Wisconsin Madison University, Gaudenz Danuser, head of Laboratory for Computational Cell Biology at the Scripps Research Institute, Sheng Yang, doctor at the Department of Communications and Electronics of ENST Paris, Bin Luo, doctor at Signal and Image Processing Department of ENST Paris, and Tieyong Zeng, doctor at the University of Paris 13 for their valuable discussions.

I would like to thank all my colleagues in the laboratory, Vannary Meas-Yedid Hardy, Elena Spirina, Fabrice Coutier, Tarn Duong, Nicolas Chenouard, Sylvain Berlemont, Fabrice De Chaumont, Christophe Zimmer, Marcio De Moraes Marim, Marie-Anne Lin, and Auguste Genovesio for their encouragement, warm helps and support during my work.

This research has been supported and funded by various organizations including CNRS, Institut Pasteur, GREYC-ENSICAEN, the France-Israel Research Networks Program, and the ANR DETECFINE Project, and I thank them all for their confidence in me.

Finally, I owe my loving thanks to my parents Xiaoxiang Gong and Siming Zhang, and my wife Xiaofei Wang, for their continuous support, encouragement and love.

Résumé

Cette thèse propose trois contributions principales pour l'imagerie en microscopie à fluorescence. **(1) Modélisation du système optique :** nous avons étudié les approximations gaussiennes des moindres carrés pour les réponses impulsionnelles optiques (PSFs) limitées par la diffraction du microscope en champ large (WFFM), du microscope confocal (LSCM), et du microscope confocal à disque rotatif (DSCM). Les situations paraxiales/non-paraxiales et 2D/3D sont toutes considérées. Les PSFs sont décrites par les intégrales de diffraction de Debye. Nous avons dérivé les paramètres gaussiens optimaux pour la PSF 2D paraxiale du WFFM, sous les normalisations L^∞ et L^1. Pour les autres PSFs, avec la normalisation L^∞, des paramètres quasi-optimaux sont explicitement dérivés par l'appariement des séries de Maclaurin. Ces modèles approximatifs gaussiens peuvent être calculés rapidement, et facilitent considérablement la modélisation des objets biologiques. **(2) Débruitage des images de fluorescence :** les images issues des LSCM et DSCM ont des statistiques purement poissoniennes ou poissoniennes et gaussiennes mélangées (MPG) selon les différents modes d'acquisition du microscope. Deux approches sont proposées pour restaurer une image poissonienne. La première méthode, basée sur les tests d'hypothèses dans le domaine de Haar biorthogonale, est particulièrement appropriée à estimer rapidement les intensités poissoniennes régulières dans des données de grandes tailles. Notre deuxième méthode, basée sur une transformée stabilisatrice de variance (VST), permet de gaussianiser et stabiliser un processus poissonien filtré. Cette VST peut être combinée avec de nombreuses transformées multi-échelles, ce qui conduit aux VSTs multi-échelles (MS-VSTs). Nous montrons que les MS-VSTs permettent de restaurer efficacement des structures saillantes de diverses formes (isotropes, linéiques et curvilignes) avec un (très) faible flux photonique. La MS-VST est également généralisée pour débruiter les données MPG et pour extraire les taches fluorescentes dans les données MPG. **(3) Détection super-résolutive :** nous avons revu et étendu les résultats des limites de la résolution pour les sources ponctuelles issus des théories de la détection, de l'estimation, et de l'information. En particulier, nous avons proposé d'appliquer la VST pour étudier les limites de la résolution dans le cas d'observations poissoniennes ou MPG. Les résultats sont asymptotiquement consistants et les expressions sont de formes closes. Nous avons également généralisé une approche de super-résolution qui est basée sur l'ajustement de modèle paramétrique et la sélection de l'ordre de modèle pour localiser un nombre inconnu de spots ou de bâtonnets. Cette méthode permet non seulement de localiser les sources ayant des configurations spatiales complexes, mais aussi d'extraire les objets séparés par des distances inférieures à la résolution optique de Rayleigh (super-résolution).

Mots-clés: PSF microscopique, approximation gaussienne de PSF, ondelette de Haar biorthogonale, transformée stabilisatrice multi-échelle de variance, processus poissonien, processus poissonien et gaussien mélangé, détection de spots, détection super-résolutive de spots, détection super-résolutive de bâtonnets.

Abstract

This thesis mainly contributes to three aspects in fluorescence microscopy imaging. **(i) Optical system modeling:** we have comprehensively studied the least squares Gaussian approximations of the diffraction-limited 2D/3D paraxial/non-paraxial point spread functions (PSFs) of wide-field fluorescence microscope (WFFM), laser-scanning confocal microscope (LSCM) and disk-scanning confocal microscope (DSCM) described using the Debye diffraction integrals. Optimal Gaussian parameters are derived for the 2D paraxial WFFM PSF, under both the L^∞ and L^1 normalizations. For the other PSFs, with the L^∞ normalization, near-optimal parameters in explicit forms are derived using Maclaurin series matching. These Gaussian approximative PSF models allow fast computation and greatly simplify the modeling of biological objects under these microscopes. **(ii) Fluorescence image denoising:** images produced by LSCM and DSCM have either a Poisson or a mixed-Poisson-Gaussian (MPG) statistical nature according to different function modes of the microscope. We have proposed two approaches for Poisson noise removal. One method is based on biorthogonal Haar-domain hypothesis tests, which is particularly suitable for fast estimating smooth intensities from large datasets. Our second method makes use of a well designed variance stabilizing transform (VST) allowing to Gaussianize and stabilize a filtered Poisson process. This VST can be combined with most multi-scale transforms yielding multi-scale VSTs (MS-VST). We show that this MS-VST approach provides a very effective denoiser capable of recovering important structures of various (isotropic, line-like and curvilinear) shapes in (very) low-count images. This MS-VST method has also been extended to remove MPG noise, and to extract fluorescent spots within MPG noisy data. **(iii) Super-resolution object detection:** we have reviewed and extended the results of resolution limits for point-like sources under detection-theoretic, estimation-theoretic and information-theoretic points of view. In particular, we propose to apply the VST to study the limiting resolution with Poisson or MPG data, leading to asymptotically consistent results with closed-form expressions. We have also generalized an existing super-resolution approach, which is based on parametric model fitting and model-order selection, to localize an unknown number of spots or rods. This method allows not only to localize sources having complex spatial configurations, but also to detect objects separated with distances smaller than Rayleigh optical resolution (super-resolution).

Keywords: Fluorescence microscope PSF, Gaussian PSF approximation, biorthogonal Haar wavelet, multi-scale variance stabilizing transform, Poisson process, mixed-Poisson-Gaussian process, spot detection, super-resolution spot detection, super-resolution rod detection.

Résumé de la thèse « Contributions à la microscopie à fluorescence en imagerie biologique : modélisation de la PSF, restauration d'images et détection super-résolutive »

Bo ZHANG

1 Introduction

Cette thèse propose trois contributions principales pour l'imagerie en microscope à fluorescence. **(1) Modélisation du système optique :** nous avons étudié les approximations gaussiennes des moindres carrés pour les réponses impulsionnelles optiques (PSFs) limitées par la diffraction du microscope en champ large (WFFM), du microscope confocal (LSCM), et du microscope confocal à disque rotatif (DSCM). Les situations paraxiales/non-paraxiales et 2D/3D sont toutes considérées. Les PSFs sont décrites par les intégrales de diffraction de Debye. Nous avons dérivé les paramètres gaussiens optimaux pour la PSF 2D paraxiale du WFFM, sous les normalisations L^∞ et L^1. Pour les autres PSFs, avec la normalisation L^∞, des paramètres quasi-optimaux sont explicitement dérivés par l'appariement des séries de Maclaurin. Ces modèles approximatifs gaussiens peuvent être calculés rapidement, et facilitent considérablement la modélisation des objets biologiques. **(2) Débruitage des images de fluorescence :** les images issues des LSCM et DSCM ont des statistiques purement poissoniennes ou poissoniennes et gaussiennes mélangées (MPG) selon les différents modes d'acquisition du microscope. Deux approches sont proposées pour restaurer une image poissonienne. La première méthode, basée sur les tests d'hypothèses dans le domaine de Haar biorthogonale, est particulièrement appropriée pour estimer rapidement les intensités poissoniennes régulières dans des données de grandes tailles. Notre deuxième méthode, basée sur une transformée stabilisatrice de variance (VST), permet de « gaussianiser » et stabiliser un processus poissonien filtré. Cette VST peut être combinée avec de nombreuses transformées multi-échelles, ce qui conduit aux VSTs multi-échelles (MS-VSTs). Nous montrons que les MS-VSTs permettent de restaurer efficacement les structures saillantes de diverses formes (isotropes, linéiques et curvilignes) avec un (très) faible flux photonique. La MS-VST est également généralisée pour débruiter des données MPG et pour extraire les taches fluorescentes. **(3) Détection super-résolutive :** nous avons revu et étendu les résultats des limites de la résolution pour les sources ponctuelles issus des théories de la détection, de l'estimation, et de l'information. En particulier, nous avons proposé d'appliquer la VST pour étudier les limites de la résolution dans le cas d'observations poissoniennes ou MPG. Les résultats sont asymptotiquement consistants et les expressions sont de formes closes. Nous avons également généralisé une approche de super-résolution

pour localiser un nombre inconnu de spots ou de bâtonnets. Cette méthode permet non seulement de localiser les sources ayant des configurations spatiales complexes, mais aussi d'extraire les objets séparés par des distances inférieures à la résolution optique de Rayleigh (super-résolution).

2 Approximations gaussiennes des PSFs microscopiques

Les PSFs des microscopes à fluorescence (WFFM, LSCM et DSCM) jouent un rôle important dans les études de la performance d'imagerie. De nombreux efforts ont été apportés pour dériver les modèles physiques de PSFs. Malgré l'existence des modèles physiques rigoureux, les PSFs approximatives notamment les approximations gaussiennes sont largement préférées dans les applications qui requièrent des traitements rapides des données. En effet, comparé aux modèles physiques, une PSF gaussienne a une expression beaucoup plus simple et peut être calculée beaucoup plus rapidement. De plus, grâce à la simplicité, une PSF gaussienne permet de faciliter des analyses théoriques et des modélisations.

Nous étudions les approximations gaussiennes des moindres carrés pour les PSFs limitées par la diffraction du WFFM, LSCM et DSCM. Les PSFs sont décrites par les intégrales de diffraction de Debye (Section 2.1). Nous dérivons sous les normalisations L^∞ et L^1 les paramètres gaussiens optimaux et quasi-optimaux pour les PSFs (Section 2.2). Les simulation numériques montrent que : (i) les approximations 2D sont très précises ; (ii) il n'existe pas d'approximations précises pour la PSF 3D du WFFM ; (iii) pour les sténopés de tailles standards, les approximations 3D sont précises pour DSCM et quasi-parfaites pour LSCM. Tous les paramètres gaussiens dérivés ont des formes closes qui sont présentés dans les Tableaux 1, 2 et 3.

2.1 Modèles physiques des PSFs limitées par la diffraction

Nous supposons que les modèles des PSFs sont limités par la diffraction et que les aberrations optiques sont ignorées. Considérons un objectif microsco-

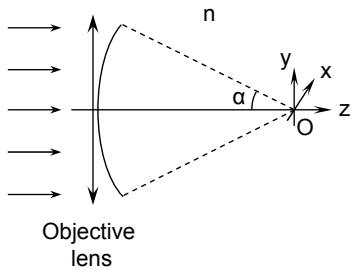


FIG. 1: Un faisceau de lumière focalisé par un objectif.

pique uniformément éclairé (Fig. 1). La distribution d'amplitude proche-focale est donnée par l'intégral de Debye :

$$\mathbf{h}(x, y, z; \lambda) = C_0 \int_0^\alpha \sqrt{\cos\theta} J_0(k\rho \sin\theta) e^{-ikz\cos\theta} \sin\theta \, d\theta \qquad (1)$$

où λ est la longueur d'onde d'excitation, n l'indice de réfraction de l'espace d'objet, $k = n\frac{2\pi}{\lambda}$ le nombre d'onde, C_0 une constante complexe, J_0 la fonction Bessel d'ordre zéro de première espèce, $\rho = \sqrt{x^2 + y^2}$, et α le semi-angle maximum de la convergence par l'objectif. Il est connu que ce modèle de Debye est bien adapté pour décrire la distribution de lumière en situation non-paraxiale, i.e., pour un objectif d'une ouverture numérique (NA) importante. Si NA diminue, (1) tend asymptotiquement vers le modèle paraxial :

$$\mathbf{h}(x, y, z; \lambda) = C_1 e^{-ikz} \int_0^1 J_0(k\rho t \sin\alpha) e^{\frac{i}{2}kzt^2 \sin^2\alpha} t \, dt \tag{2}$$

où $C_1 = \alpha^2 C_0$.

2.1.1 PSFs du WFFM, LSCM et DSCM

La PSF du WFFM est donné par (3).

$$\mathrm{PSF}_{WFFM}(x, y, z) = |\mathbf{h}(x, y, z; \lambda_{em})|^2 \tag{3}$$

où λ_{em} est la longueur d'onde d'émission.

Pour le LSCM, nous supposons que le sténopé est circulaire de rayon $r = D/2$. La longueur d'onde d'excitation est notée par λ_{ex}. La PSF du LSCM est exprimée par :

$$\mathrm{PSF}_{LSCM}(x, y, z) = |\mathbf{h}(x, y, z; \lambda_{ex})|^2 \cdot \int_{\{x_1^2+y_1^2 \le r^2\}} |\mathbf{h}(x-x_1, y-y_1, z; \lambda_{em})|^2 \, dx_1 dy_1 \tag{4}$$

Pour le DSCM, la distance entre deux sténopés adjacents est notée par d. La PSF est :

$$\begin{aligned} \mathrm{PSF}_{DSCM}(x, y, z) \quad = \quad & \left| \sum_{(n_x, n_y) \in \mathcal{D}} \mathbf{h}\left(x - \frac{d}{2}(n_x + n_y), y - \frac{\sqrt{3}}{2} d(n_x - n_y), z; \lambda_{ex} \right) \right|^2 \cdot \\ & \int_{\{x_1^2+y_1^2 \le r^2\}} |\mathbf{h}(x - x_1, y - y_1, z; \lambda_{em})|^2 \, dx_1 dy_1 \end{aligned} \tag{5}$$

où l'ensemble des indices des sténopés d'illumination est noté par $\mathcal{D} \subset \mathbb{Z}^2$.

Finalement, les expressions des PSFs paraxiales des trois microscopes sont obtenues en insérant l'intégrale paraxiale (2) dans (3), (4) et (5). Les expressions des PSFs non-paraxiales sont dérivées de la même façon, sauf que l'intégrale paraxiale (1) doit être utilisée. On peut vérifier à partir de ces expressions que toutes les PSFs sont symétriques par rapport au plan xy. De plus, PSF_{WFFM} et PSF_{LSCM} ont une symétrie cylindrique par rapport à l'axe z, et PSF_{DSCM} est presque cylindriquement symétrique si la distance d est suffisamment large.

2.2 Approximations Gaussiennes des modèles de PSFs

Pour dériver les approximations gaussiennes des PSFs (3), (4) et (5), nous supposons que la fonction gaussienne est centrée à l'origine de la PSF considérée et est séparable. Les gaussiennes centrées et séparables sont les seules fonctions

gaussiennes qui préservent les symétries des PSFs. Par conséquent, nous notons les gaussiennes 2D et 3D g_σ :

$$g_{\sigma_\rho}(x,y) := A_1 \exp\left(-\frac{x^2+y^2}{2\sigma_\rho^2}\right) = A_1 \exp\left(-\frac{\rho^2}{2\sigma_\rho^2}\right) \quad (6)$$

$$g_{\sigma_\rho,\sigma_z}(x,y,z) := A_2 \exp\left(-\frac{x^2+y^2}{2\sigma_\rho^2} - \frac{z^2}{2\sigma_z^2}\right) = A_2 \exp\left(-\frac{\rho^2}{2\sigma_\rho^2} - \frac{z^2}{2\sigma_z^2}\right) \quad (7)$$

Nous souhaitons trouver les paramètres gaussiens optimaux, i.e. $\sigma^* = \sigma_\rho^*$ pour le cas 2D, et $\sigma^* = \{\sigma_\rho^*, \sigma_z^*\}$ pour 3D, qui minimisent le critère des moindres carrés (LSQ), i.e.

$$\sigma^* = \underset{\sigma>0}{\operatorname{argmin}} \|\mathrm{PSF} - g_\sigma\|_2^2 \quad (8)$$

2.2.1 Exemple : paramètres gaussiens pour les PSFs 2D du WFFM

Nous décrivons les approximations pour les PSFs 2D du WFFM. La PSF 2D paraxial du WFFM, connue sous le nom « disque d'Airy », est dérivée par (2) et (3) où z est fixé à 0, i.e.

$$\mathrm{PSF}_{WFFM}(\rho) = \left[2\frac{J_1(k_{em}\mathrm{NA}\rho)}{k_{em}\mathrm{NA}\rho}\right]^2 \quad (9)$$

La solution de (8) est donnée par

$$\sigma_\rho^* \approx 0.21\frac{\lambda_{em}}{\mathrm{NA}} \quad (10)$$

Contrairement au cas paraxial où le paramètre optimal peut être déduit exactement, la minimisation de (8) est difficile pour la situation non-paraxiale puisque la PSF non-paraxiale n'a pas de forme close. Toutefois, nous pouvons constater que l'énergie L^2 de la PSF ($\|\mathrm{PSF}_{WFFM}\|_2^2$) se concentre essentiellement dans le lobe principal qui se situe dans un petit voisinage de l'origine. Ainsi, une approximation près de l'origine capturant la décroissance du lobe principal, e.g. $g_\sigma(\rho) \to \mathrm{PSF}_{WFFM}(\rho)$ quand $\rho \to 0$, pourrait donner une approximation précise. Cela peut être facilement atteint en appariant la série Maclaurin de la PSF et celle de gaussienne. Les deux séries diffèrent à partir de leurs termes d'ordre second. Ainsi, en imposant l'égalité de leurs termes du second ordre, nous obtenons :

$$\hat{\sigma}_\rho^* = \frac{1}{nk_{em}}\left[\frac{4 - 7\cos^{\frac{3}{2}}\alpha + 3\cos^{\frac{7}{2}}\alpha}{7(1-\cos^{\frac{3}{2}}\alpha)}\right]^{-\frac{1}{2}} \quad (11)$$

D'un point de vue géométrique, cette approche apparie les courbures principales des deux fonctions à l'origine.

La solution exacte de (8) peut être aussi déduite pour la PSF 2D du WFFM sous la normalisation L^1. Pour les PSFs restantes, l'appariement des séries Maclaurin est utilisé pour trouver les paramètres approximatifs. Ces paramètres sont montrés dans les Tableaux 1, 2 et 3.

2.2.2 Consistances des paramètres gaussiens

Système paraxial : Si NA devient petit, les paramètres gaussiens non-paraxiaux 2D/3D du WFFM, LSCM et DSCM se rapprochent des paramètres paraxiaux ;

Confocalité idéal : Si le rayon des sténopés tend vers zéro, les paramètres gaussiens 2D/3D du LSCM et DSCM convergent vers les paramètres dérivés pour des sténopés nuls ;

Sténopé entièrement ouvert : Si le rayon des sténopés tend vers l'infini, les paramètres gaussiens 2D/3D du LSCM et DSCM se rapprochent de ceux du WFFM où λ_{em} est remplacée by λ_{ex}, *sauf* pour le paramètre non-paraxial $\hat{\sigma}_z^*$. La dernière exception implique que les approximations gaussiennes non-paraxiales 3D pour LSCM et DSCM ne peuvent pas être appliquées aux cas de grands sténopés.

2.2.3 Résultats

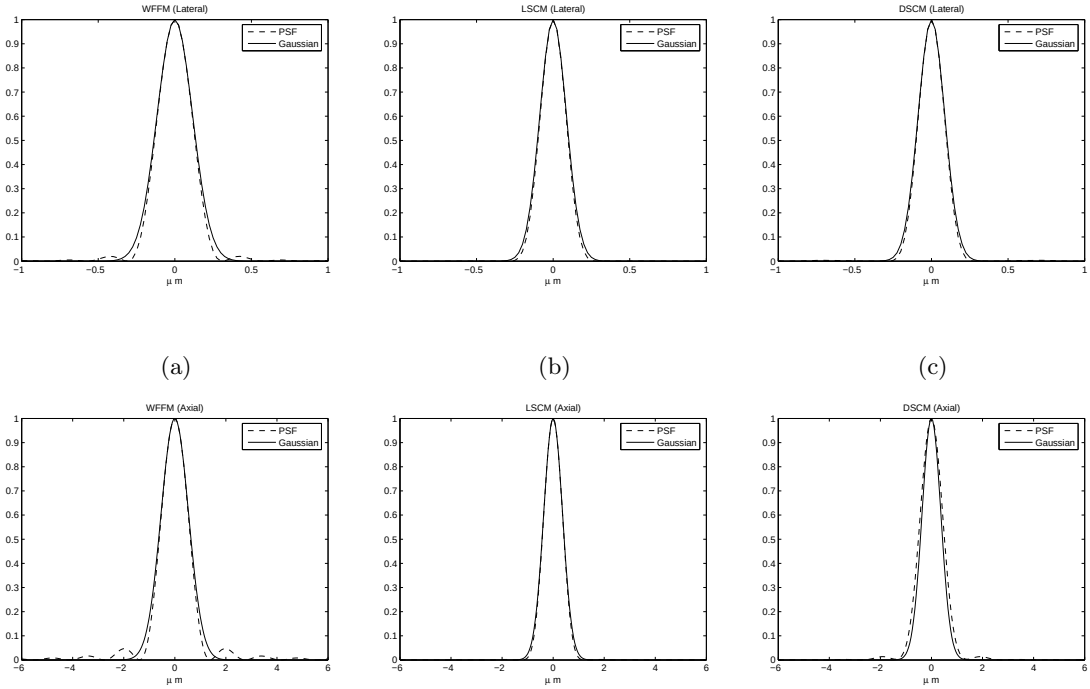
La Figure 2 montre des exemples d'approximations gaussiennes pour les PSFs. Nous pouvons constater visuellement que les approximations sont précises sauf celle pour la PSF axiale du WFFM où les lobes secondaires de la PSF ne peuvent pas être bien approximés.

Quantitativement, les erreurs des approximations ont été évaluées en utilisant le critère de l'erreur quadratique relative (RSE) définie par

$$\mathrm{RSE} := \frac{\|\mathrm{PSF} - g_{\hat{\sigma}^*}\|_2^2}{\|\mathrm{PSF}\|_2^2} \tag{12}$$

où $g_{\hat{\sigma}^*}$ est la fonction gaussienne approximative. Le critère de l'erreur relative du paramètre (PRE), i.e. $|\hat{\sigma}^* - \sigma^*|/\sigma^*$, a été aussi utilisé pour comparer le paramètre proposé $\hat{\sigma}^*$ et celui de vérité terrain σ^* déduit numériquement.

Les simulation ont montré que : (i) les approximations 2D sont très précises ; (ii) il n'existe pas d'approximations précises pour la PSF 3D du WFFM ; (iii) pour les sténopés de tailles standards, les approximations 3D sont précises pour DSCM et quasi-parfaites pour LSCM.



(a) (b) (c)

(d) (e) (f)

FIG. 2: Approximations gaussiennes des PSFs du WFFM, LSCM et DSCM avec la normalisation L^∞. Cas non-paraxiaux, $\lambda_{ex} = 488\text{nm}$, $\lambda_{em} = 520\text{nm}$, $n = 1.515$, NA $= 1.0$, $d_0 := \frac{1}{2}\left(\frac{\lambda_{em}}{\lambda_{ex}} + D + 1\right)$ AU, et le diamètre des sténopés $D = 0.5\text{AU}$ pour LSCM et DSCM. (a) PSF latérale du WFFM ; (b) PSF latérale du LSCM ; (c) PSF latérale du DSCM ($d = d_0$) ; (d) PSF axiale du WFFM ; (e) PSF axiale du LSCM ; (f) PSF axiale du DSCM ($d = d_0$).

TAB. 1: Paramètres gaussiens pour les PSFs 2D

Microscope	Paramètre gaussien[1]
WFFM paraxial (normalisation L^∞)	$\hat{\sigma}^*_\rho = 0.21\lambda_{em}/\mathrm{NA}$
WFFM paraxial (normalisation L^1)	$\hat{\sigma}^*_\rho = 0.22\lambda_{em}/\mathrm{NA}$
WFFM non-paraxial (normalisation L^∞)	$\hat{\sigma}^*_\rho = \frac{1}{nk_{em}}\left[\frac{4-7\cos^{\frac{3}{2}}\alpha+3\cos^{\frac{7}{2}}\alpha}{7(1-\cos^{\frac{3}{2}}\alpha)}\right]^{-\frac{1}{2}}$
LSCM et DSCM paraxiaux ($d \geq d_0$, normalisation L^∞)	$\hat{\sigma}^*_\rho = \sqrt{2}\left[\frac{c_1^2}{r^2} + \frac{4c_2J_0(c_2)J_1(c_2)-8J_1^2(c_2)}{r^2[J_0^2(c_2)+J_1^2(c_2)-1]}\right]^{-\frac{1}{2}}$
LSCM et DSCM non-paraxiaux ($d \geq d_0$, normalisation L^∞)	$\hat{\sigma}^*_\rho = \sqrt{2}\left[\frac{2\sigma^4_{em,\rho}\left[\exp\left(\frac{r^2}{2\sigma^2_{em,\rho}}\right)-1\right]+r^2\sigma^2_{ex,\rho}}{\sigma^2_{ex,\rho}\sigma^4_{em,\rho}\left[\exp\left(\frac{r^2}{2\sigma^2_{em,\rho}}\right)-1\right]}\right]^{-\frac{1}{2}}$

[1] $k_{ex} := \frac{2\pi}{\lambda_{ex}}$, $k_{em} := \frac{2\pi}{\lambda_{em}}$, $c_1 := k_{ex}r\mathrm{NA}$, $c_2 := k_{em}r\mathrm{NA}$, $d_0 := \frac{1}{2}\left(\frac{\lambda_{em}}{\lambda_{ex}} + D + 1\right)$ AU, $\sigma_{em,\rho}$ est donné par l'expression de $\hat{\sigma}^*_\rho$ du WFFM non-paraxial (normalisation L^∞), et $\sigma_{ex,\rho}$ par la même expression avec k_{em} replacé par k_{ex}.

TAB. 2: Paramètres gaussiens latéraux pour les PSFs 3D (normalisation L^∞)

Microscope	Paramètre gaussien latéral[1]
WFFM paraxial	$\hat{\sigma}^*_\rho = \sqrt{2}/(k_{em}\mathrm{NA})$
WFFM non-paraxial	$\hat{\sigma}^*_\rho = \frac{1}{nk_{em}}\left[\frac{4-7\cos^{\frac{3}{2}}\alpha+3\cos^{\frac{7}{2}}\alpha}{7(1-\cos^{\frac{3}{2}}\alpha)}\right]^{-\frac{1}{2}}$
LSCM et DSCM paraxiaux ($d \geq d_0$)	$\hat{\sigma}^*_\rho = \sqrt{2}\left[\frac{c_1^2}{r^2} + \frac{4c_2J_0(c_2)J_1(c_2)-8J_1^2(c_2)}{r^2[J_0^2(c_2)+J_1^2(c_2)-1]}\right]^{-\frac{1}{2}}$
LSCM et DSCM non-paraxiaux ($d \geq d_0$)	$\hat{\sigma}^*_\rho = \sqrt{2}\left[\frac{2\sigma^4_{em,\rho}\left[\exp\left(\frac{r^2}{2\sigma^2_{em,\rho}}\right)-1\right]+r^2\sigma^2_{ex,\rho}}{\sigma^2_{ex,\rho}\sigma^4_{em,\rho}\left[\exp\left(\frac{r^2}{2\sigma^2_{em,\rho}}\right)-1\right]}\right]^{-\frac{1}{2}}$

[1] $k_{ex} := \frac{2\pi}{\lambda_{ex}}$, $k_{em} := \frac{2\pi}{\lambda_{em}}$, $c_1 := k_{ex}r\mathrm{NA}$, $c_2 := k_{em}r\mathrm{NA}$, $d_0 := \frac{1}{2}\left(\frac{\lambda_{em}}{\lambda_{ex}} + D + 1\right)$ AU, $\sigma_{em,\rho}$ est donné par l'expression de $\hat{\sigma}^*_\rho$ du WFFM non-paraxial, et $\sigma_{ex,\rho}$ par la même expression avec k_{em} replacé par k_{ex}.

TAB. 3: Paramètres gaussiens axiaux pour les PSFs 3D (normalisation L^∞)

Microscope	Paramètre gaussien axial[1]
WFFM paraxial	$\hat{\sigma}_z^* = 2\sqrt{6} \cdot n/(k_{em}\mathrm{NA}^2)$
WFFM non-paraxial	$\hat{\sigma}_z^* = \frac{5\sqrt{7}(1-\cos^{\frac{3}{2}}\alpha)}{\sqrt{6}\cdot n k_{em}\left[4\cos^5\alpha - 25\cos^{\frac{7}{2}}\alpha + 42\cos^{\frac{5}{2}}\alpha - 25\cos^{\frac{3}{2}}\alpha + 4\right]^{\frac{1}{2}}}$
LSCM et DSCM paraxiaux ($d \geq d_0$)	$\hat{\sigma}_z^* = 2\sqrt{6}\left[\frac{c_1^2\mathrm{NA}^2}{r^2n^2} - \frac{48c_2^2[J_0^2(c_2)+J_1^2(c_2)]-192J_1^2(c_2)}{n^2k_{em}^2r^4[J_0^2(c_2)+J_1^2(c_2)-1]}\right]^{-\frac{1}{2}}$
LSCM et DSCM paraxiaux ($d \geq d_0$)	$\hat{\sigma}_z^* = \frac{\sigma_{ex,z}\sigma_{em,z}}{\left[\sigma_{ex,z}^2+\sigma_{em,z}^2\right]^{\frac{1}{2}}}$

[1] $k_{ex} := \frac{2\pi}{\lambda_{ex}}$, $k_{em} := \frac{2\pi}{\lambda_{em}}$, $c_1 := k_{ex}r\mathrm{NA}$, $c_2 := k_{em}r\mathrm{NA}$, $d_0 := \frac{1}{2}\left(\frac{\lambda_{em}}{\lambda_{ex}} + D + 1\right)$ AU, $\sigma_{em,z}$ est donné par l'expression de $\hat{\sigma}_z^*$ du WFFM non-paraxial, et $\sigma_{ex,z}$ par la même expression avec k_{em} replacé par k_{ex}.

3 Débruitage d'images de fluorescence

Nous avons étudié les différentes sources du bruits qui dégradent la qualité d'images de fluorescence (bruit photonique, bruit de courant noir, bruit de lecture, et bruit de quantification). Les modèles finaux des signaux du WFFM, LSCM et DSCM sont résumés comme suit.

WFFM Le signal S est contaminé par un bruit gaussien :

$$S \approx \alpha\lambda + \mu + B, \quad B \sim \mathcal{N}(0, \sigma^2) \tag{13}$$

où α est le gain du microscope, μ la moyenne et σ^2 la variance du bruit.

LSCM Nous distinguons les deux modes d'acquisition :

- Pour le mode analogique, le signal est un processus poissonien et gaussien mélangé (MPG) :

$$S = \alpha N_s + N_r, \quad N_s \sim \mathcal{P}(\lambda), \quad N_r \sim \mathcal{N}(\mu, \sigma^2) \tag{14}$$

  où N_s, une variable poissonienne, modélise le comptage de photons, et N_r, une variable gaussienne, représente le bruit de lecture.
- Pour le mode de comptage de photons, le signal est Poissonien :

$$S \sim \mathcal{P}(\lambda) \tag{15}$$

DSCM Le signal est un processus MPG :

$$S = \alpha N_s + N_r, \quad N_s \sim \mathcal{P}(\lambda), \quad N_r \sim \mathcal{N}(\mu, \sigma^2) \tag{16}$$

3.1 Suppression du bruit poissonien

Nous formulons d'abord le problème de débruitage poissonien. Nous observons des données discrètes de comptage de photons $\mathbf{X} = (X_i)_{i\in\mathbb{Z}^q}$ où X_i est une

variable aléatoire de Poisson d'intensité λ_i, i.e., $X_i \sim \mathcal{P}(\lambda_i)$. Ici, nous supposons que les X_i's sont mutuellement indépendantes. Le débruitage vise à estimer les intensités sous-jacentes $\Lambda = (\lambda_i)_{i \in \mathbb{Z}^q}$ des observations $\mathbf{X}$.

Concernant ce sujet, notre contribution a été de proposer deux nouvelles méthodes de débruitage poissoniens. Notre première approche est basée sur les tests d'hypothèses dans le domaine des ondelettes Haar biorthogonales (Bi-Haar) (Section 3.1.1). Cette approche est particulièrement adaptée pour estimer rapidement les intensités régulières des données volumineuses. Notre deuxième méthode utilise une transformée stabilisatrice de la variance (VST) bien conçue. Elle peut être associée à la plupart des transformées multi-échelles pour adapter les diverses morphologies des sources à restaurer (Section 3.1.2).

3.1.1 Débruitage poissonien par les tests d'hypothèses dans le domaine de Bi-Haar

Le débruitage par ondelettes peut être achevé par réduction à zéro des coefficients non-significatifs tout en préservant ceux qui sont significatifs. Nous détectons les coefficients significatifs par l'application d'un test binaire pour chaque coefficient d'ondelette d :

$$H_0 : d = 0 \text{ vs. } H_1 : d \neq 0$$

Notons que comme toute ondelette a une moyenne zéro, si d provient d'un signal d'intensité constante dans le support d'ondelette, nous aurons $d \in H_0$. Les tests d'hypothèses individuels sont largement utilisés pour contrôler un taux de faux positifs (FPR) préfixé dans le domaine d'ondelettes, disons α.

Les tests requièrent le calcul de la p-valeur de chaque coefficient d'ondelette sous H_0. Pour obtenir des distributions de formes maniables pour les coefficients, des ondelettes simples sont privilégiées, comme Haar. A notre connaissance, Haar est la seule ondelette dont la distribution de probabilité (*pdf*) des coefficients a une forme close. Mais, à cause des filtres discontinus de Haar, les intensités estimées pourraient être très irrégulières avec des artefacts d'escaliers lorsque la décimation est impliquée.

Pour résoudre ce dilemme entre la maniabilité de la distribution et la régularité de reconstruction, nous proposons d'utiliser l'ondelette Bi-Haar. Son banc de filtre est donné par :

$$\begin{array}{ll} h = 2^{-c}[1,1], & g = 2^{-c} r[\frac{1}{8}, \frac{1}{8}, -1, 1, -\frac{1}{8}, -\frac{1}{8}]; \\ \tilde{h} = 2^{c-1} r[-\frac{1}{8}, \frac{1}{8}, 1, 1, \frac{1}{8}, -\frac{1}{8}], & \tilde{g} = 2^{c-1}[1,-1] \end{array}$$

où c et $r = (1 + 2^{-5})^{-1/2}$ sont deux facteurs normalisateurs, (h, g) et $(\tilde{h}, \tilde{g})$ sont respectivement le banc de filtres d'analyse et celui de synthèse. Rappelons que le banc de filtres de Haar est ($h = 2^{-c}[1,1]$, $g = 2^{-c}[-1,1]$, $\tilde{h} = 2^{c-1}[1,1]$, $\tilde{g} = 2^{c-1}[1,-1]$). Il s'ensuit que la fonction d'échelle synthétique de Haar est discontinue tandis que celle de Bi-Haar est quasi-Lipschitzienne. Ainsi, la reconstruction de Bi-Haar sera plus régulière.

Nous notons respectivement un coefficient de Haar et celui de Bi-Haar par d_j^h et d_j^{bh}. Notons aussi que $p_H := \Pr(d_j^h \geq 2^{-cj} k_0 | H_0)$ la p-valeur d'un coefficient de Haar où $k_0 = 1, 2, \cdots$, et $p_{BH} := \Pr(d_j^{bh} \geq 2^{-cj} k_0 | H_0)$ la p-valeur d'un coefficient Bi-Haar. Pour des intensités élevées ou de grandes échelles, d_j^h et d_j^{bh}

convergent asymptotiquement vers une même loi (gaussienne). Ainsi, ils auront les mêmes p-valeurs, i.e., $p_H \approx p_{BH}$.

Pour de faibles intensités ($\lambda \ll 1$) et petites échelles, nous avons montré que p_{BH} est essentiellement bornée supérieurement par p_H sous H_0. Par exemple, pour les données 1D, nous avons $p_{BH} \leq p_H + A(\lambda)(1-2p_H)$ où $A(\lambda) = \frac{2^{9j-7}}{2835}\lambda^9 + o(\lambda^9)$ et j l'indice d'échelle.

Les tests individuels peuvent être mis en œuvre par les opérateurs de seuillage. Autrement dit, nous pouvons trouver $\tilde{t}_j$ tel que $\Pr(|d_j^{bh}| \geq \tilde{t}_j | H_0) \leq \alpha$ où α représente le FPR contrôlé. Supposons que dans le cas de Haar nous avons dérivé le seuil t_j. Alors, en fixant $\tilde{t}_j := 2^{-cjq}\lceil 2^{cjq} t_j \rceil$ la relation entre p_H et p_{BH} nous permet de conclure que le FPR d'un test Bi-Haar sera toujours borné supérieurement par α. Ainsi, nous bénéficions de la régularité des filtres Bi-Haar pour avoir une reconstruction régulière tout en maintenant une complexité faible de calcul (celle de Haar décimée). Il reste donc à trouver t_j. Pour cela, nous écrivons :

$$\Pr(d_j^h \geq t_j | H_0) = \Pr\left(\chi^2_{(2m_j)}(\lambda_j) < \lambda_j\right) \approx \Pr(\gamma \chi^2_{(f)} < \lambda_j) \quad (17)$$

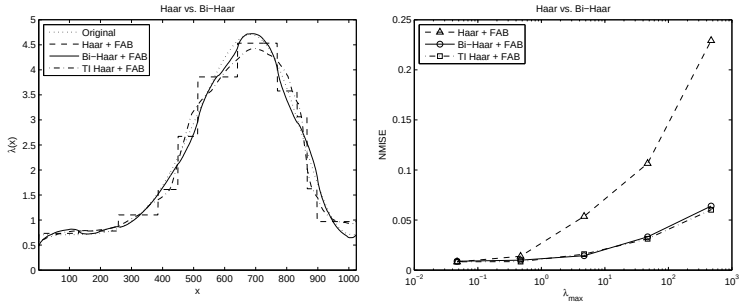
$$\approx \Pr\left(Z > \sqrt{2f-1} - \sqrt{\frac{2\lambda_j}{\gamma}}\right) \quad (18)$$

où $m_j = 2^{cjq} t_j$, $\gamma = (2m_j + 2\lambda_j)/(2m_j + \lambda_j)$, $f = (2m_j + \lambda_j)^2/(2m_j + 2\lambda_j)$, et $Z \sim \mathcal{N}(0,1)$. Ici, deux étapes d'approximation sont utilisées : (1) la distribution chi-deux non-centrale est approximée par une distribution chi-deux centrale (17) ; (2) cette dernière est encore approximée par la loi normale (approximation de Fisher (18)). Nous avons dérivé le seuil t_j sous cette approximation de Fisher (noté par le seuillage FAB).

Résultats La Figure 3 présente les restaurations de la fonction « Smooth » par la méthode Haar et celle de Bi-Haar. La Figure 3(a) montre que notre approche permet de restaurer les intensités sous-jacentes tout en préservant la régularité des intensités. Nous pouvons aussi voir que l'approche Haar conduit à des artefacts d'escaliers. La Figure 3(b) compare les erreurs quadratiques moyennes intégrées normalisées des débruitages. Nous pouvons constater dans cet exemple que la performance de l'estimation de Bi-Haar est tout à fait comparable à celle de Haar non-décimée (TI Haar) qui a une complexité temporelle/spatiale beaucoup plus importante.

3.1.2 Débruitage poissonien par MS-VST

Le débruitage par Bi-Haar est rapide, et produit des estimations régulières. Toutefois, pour les applications ne nécessitant pas de traitements en temps réel, les transformées non-décimées sont toujours préférables que celles décimées. Cependant, le gain offert par Bi-Haar non-décimée plutôt que Haar est marginal. Une autre limite de cette approche est qu'elle est conçue pour les ondelettes Bi-Haar seulement. Pour les sources linéiques ou curvilignes, ce filtre d'ondelettes est clairement inadapté. Pour les sources isotropes q-dimensionelles ($q \geq 2$), la transformée en ondelettes biorthogonales (plus généralement pour toute transformée en ondelettes séparables) n'offre pas une analyse isotrope. De ce fait, cette méthode conduit à une estimation non-optimale de ces sources.



(a) (b)

FIG. 3: Débruitage de la fonction « Smooth » (longueur = 1024). Comparaison des estimations par Haar, Bi-Haar et TI Haar (non-decimée) avec $\alpha = 10^{-3}$ et nombre d'échelles $J = 7$. (a) résultats des débruitages ; (b) erreurs quadratiques moyennes intégrées normalisées (NMISEs).

Ainsi, nous avons proposé un autre estimateur des intensités poissoniennes qui est plus souple pour adapter de différents types de sources. Il est basé sur une VST conçue pour d'une part stabiliser la variance d'un processus poissonien filtré et d'autre part pour le gaussieniser. Cette nouvelle transformée, qui étend celle d'Anscombe, est simple, rapide et efficace pour des données à (très) faible flux de comptage.

Etant donné un processus poissonien $\mathbf{X} := (X_i)_i$, $Y_j := \sum_i h[i]X_{j-i}$ est le processus filtré en convoluant $\mathbf{X}$ avec un filtre h. Y_j pour j quelconque sera noté par Y. Nous définissons $\tau_k := \sum_i (h[i])^k$ pour $k = 1, 2, \cdots$. De plus, nous adoptons une hypothèse d'homogénéité locale que $\lambda_{j-i} = \lambda$ pour tout i dans le support de h.

Nous pouvons montrer que si h a une moyenne non-nulle, la transformée de racine carrée sous forme (19) est une VST.

$$T(Y) := b \cdot \mathrm{sgn}(Y + c)|Y + c|^{1/2} \tag{19}$$

Ici, b est un facteur normalisateur et $c \in \mathbb{R}$, dont la valeur contrôle la vitesse de convergence vers la loi normale. Nous avons montré que la valeur asymptotiquement optimale est donnée par

$$c = \frac{7\tau_2}{8\tau_1} - \frac{\tau_3}{2\tau_2} \tag{20}$$

En conséquence, nous avons $T(Y) - b\sqrt{\tau_1 \lambda} \xrightarrow[\lambda \to +\infty]{\mathcal{D}} \mathcal{N}(0, 1)$ où $b = 2\sqrt{\frac{\tau_1}{\tau_2}}$.

En reconnaissant qu'une grande famille des transformées multi-échelles est implémentée par des filtres (e.g. ondelettes, ridgelets et curvelets), cette VST peut être combinée avec les banques de filtres des transformées, ce qui conduit aux VSTs multi-échelles (MS-VSTs). En vue du débruitage poissonien, nous sommes autorisés à choisir ou à concevoir la transformée la plus adaptative pour les sources à restaurées selon leurs morphologies. Nous montrerons que notre VST peut facilement être associée à ces différentes décompositions géométriques, ce qui produit des coefficients asymptotiquement normalement distribués dont la variation est connue. Les tests d'hypothèses classiques sont ensuite adoptés

pour détecter les coefficients significatifs, et une approche itérative est proposée pour reconstruire l'estimation finale.

Par exemple, (21) montre comment la VST est combinée avec la transformée en ondelettes isotropes (IUWT). En effet, comme les coefficients d'approximation (a_j) sont itérativement filtrés par les filtres passe-bas $(\bar{h}^{\uparrow j})_j$, nous sommes autorisés à appliquer les VSTs sur a_j pour les stabiliser et gaussieniser.

$$\text{IUWT} \left\{ \begin{array}{lcl} a_j & = & \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j & = & a_{j-1} - a_j \end{array} \right. \Longrightarrow \begin{array}{c} \text{MS-VST} \\ + \\ \text{IUWT} \end{array} \left\{ \begin{array}{lcl} a_j & = & \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j & = & T_{j-1}(a_{j-1}) - T_j(a_j) \end{array} \right. \tag{21}$$

où $h^{\uparrow j}[l] = h[l]$ si $l/2^j \in \mathbb{Z}$ et 0 sinon, et $\bar{h}[n] := h[-n]$. Notons que le filtrage sur a_{j-1} peut être ré-écrit comme un filtrage sur les données originales $a_0 := \mathbf{X}$, i.e., $a_j = h^{(j)} \star a_0$, où $h^{(j)} = \bar{h}^{\uparrow j-1} \star \cdots \star \bar{h}^{\uparrow 1} \star \bar{h}$ pour $j \geq 1$ et $h^{(0)} = \delta$. T_j est la VST à l'échelle j :

$$T_j(a_j) = b^{(j)} \operatorname{sgn}(a_j + c^{(j)}) \sqrt{|a_j + c^{(j)}|} \tag{22}$$

Nous définissons aussi $\tau_k^{(j)} := \sum_i \left(h^{(j)}[i]\right)^k$. Selon (20), la constante $c^{(j)}$ associée à $h^{(j)}$ est donnée par

$$c^{(j)} := \frac{7\tau_2^{(j)}}{8\tau_1^{(j)}} - \frac{\tau_3^{(j)}}{2\tau_2^{(j)}} \tag{23}$$

Avec ce schéma MS-VST, nous pouvons montrer que les coefficients d'ondelettes sont stabilisés et gaussienisés :

$$d_j[k] \xrightarrow[\lambda \to +\infty]{\mathcal{D}} \mathcal{N}\left(0, \quad \frac{\tau_2^{(j-1)}}{4{\tau_1^{(j-1)}}^2} + \frac{\tau_2^{(j)}}{4{\tau_1^{(j)}}^2} - \frac{\langle h^{(j-1)}, h^{(j)} \rangle}{2\tau_1^{(j-1)}\tau_1^{(j)}}\right) \tag{24}$$

Ce résultat nous permet d'appliquer les tests d'hypothèses sur d_j pour détecter les coefficients significatifs. Notons le support de multi-résolution $\mathcal{M} := \{(j,k) \mid \text{si } d_j[k] \text{ est significatif}\}$.

Finalement, la reconstruction est formulée par un problème d'optimisation convexe (25).

$$\begin{array}{l} \min_{\mathbf{d} \in \mathcal{C}} J(\mathbf{d}), \quad J(\mathbf{d}) := \|\mathbf{d}\|_1 \\ \text{où } \mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2, \ \mathcal{S}_1 := \{\mathbf{d} | d_j[k] = (\mathcal{W}\mathbf{X})_j[k], (j,k) \in \mathcal{M}\}, \ \mathcal{S}_2 := \{\mathbf{d} | \mathcal{R}\mathbf{d} \geq 0\} \end{array} \tag{25}$$

où $\mathcal{W}$ représente l'opérateur de la transformée en ondelettes, et $\mathcal{R}$ celui de synthèse. Nous cherchons une solution qui peut être représentée dans le domaine d'ondelettes de façon clairsemée. Ainsi, nous minimisons la norme ℓ^1 des coefficients d'ondelette $\mathbf{d}$, sous deux contraintes $\mathcal{S}_1$ et $\mathcal{S}_2$. $\mathcal{S}_1$ requiert que les coefficients significatifs de $\mathbf{d}$ préservent ceux des données $\mathbf{X}$; $\mathcal{S}_2$ assure une estimation positive puisque les intensités poissoniennes sont toujours non-négative. Ce problème est résolu par la méthode itérative de la descente rapide hybride (HSD).

Outre cette combinaison de la VST avec IUWT, nous avons aussi combiné la VST avec la transformée en ondelettes séparables, la transformée en ridgelets et celle en curvelets.

Résultats La Figure 4 permet de comparer notre approche à plusieurs débruiteurs poissoniens proposés dans la littérature. Nous pouvons voir que visuellement et quantitativement le meilleur résultat est donné par celui de MS-VST combinée avec les curvelets. Le dictionnaire de curvelets est adapté puisque les tubulines ont des formes filamenteuses. Nous trouvons donc que notre approche est très compétitive par rapport à l'état de l'art.

En résumé, cette approche MS-VST présente les avantages suivants :

- Elle est efficace et sensible pour l'estimation dans les données à (très) faible flux de comptage ;
- Nous avons le choix d'intégrer les VST avec la transformées multi-échelle que nous croyons la plus appropriée pour restaurer des sources ayant certaine caractéristique morphologique (isotrope, linéique, courviligne, etc) ;
- Le temps de calcul est similaire à celui d'un débruitage gaussien.

3.2 Suppression du bruit MPG

Pour un LSCM travaillant dans le mode analogique, ou un DSCM, les images sont modélisés par des processus MPG :

$$X_i = \alpha U_i + V_i, \quad U_i \sim \mathcal{P}(\lambda_i), \quad V_i \sim \mathcal{N}(\mu, \sigma^2) \tag{26}$$

où $\alpha > 0$ est le gain du système, U_i une variable poissonienne, V_i une variable normale, et toutes $(U_i)_i$ et $(V_i)_i$ sont supposées mutuellement indépendantes. Etant donné un filtre h, le processus MPG filtré est défini par :

$$Y_i := \sum_{j \in \mathbb{Z}^q} h[j] X_{i-j} \tag{27}$$

Nous pouvons montrer que (19) est encore une VST pour (27) sauf que la valeur optimale du paramètre c est donnée par

$$c^* := -\mu\tau_1 + \frac{\tau_1 \sigma^2}{\alpha} + \frac{7\alpha\tau_2}{8\tau_1} - \frac{\alpha\tau_3}{2\tau_2} \tag{28}$$

La MS-VST est obtenu de la même manière comme dans le cas poissonien.

3.2.1 Résultats

Un exemple de débruitage d'une image de télomères sous DSCM est montré par la Figure 5. Les résultats de débruitage sous différentes hypothèses du bruit sont comparés. Nous pouvons voir qu'en supposant seulement un bruit MPG on peut supprimer tout le bruit dans l'image.

3.3 Détection de spots fluorescents

Nous avons aussi proposé une méthode de détection de spots fluorescents basée sur la MS-VST. Nous dérivons premièrement le support de multi-résolution en appliquant la MS-VST combinée avec IUWT. Ensuite, les coefficients positifs et significatifs d'IUWT à travers plusieurs échelles sont corrélés (i.e. multipliés). En effet, les maxima locaux des coefficients d'ondelettes ont tendance à se propager à travers plusieurs échelles s'ils sont dus aux spots, alors qu'ils ne le sont pas s'ils sont issus du bruit. Les échelles sont choisies en fonction des tailles des

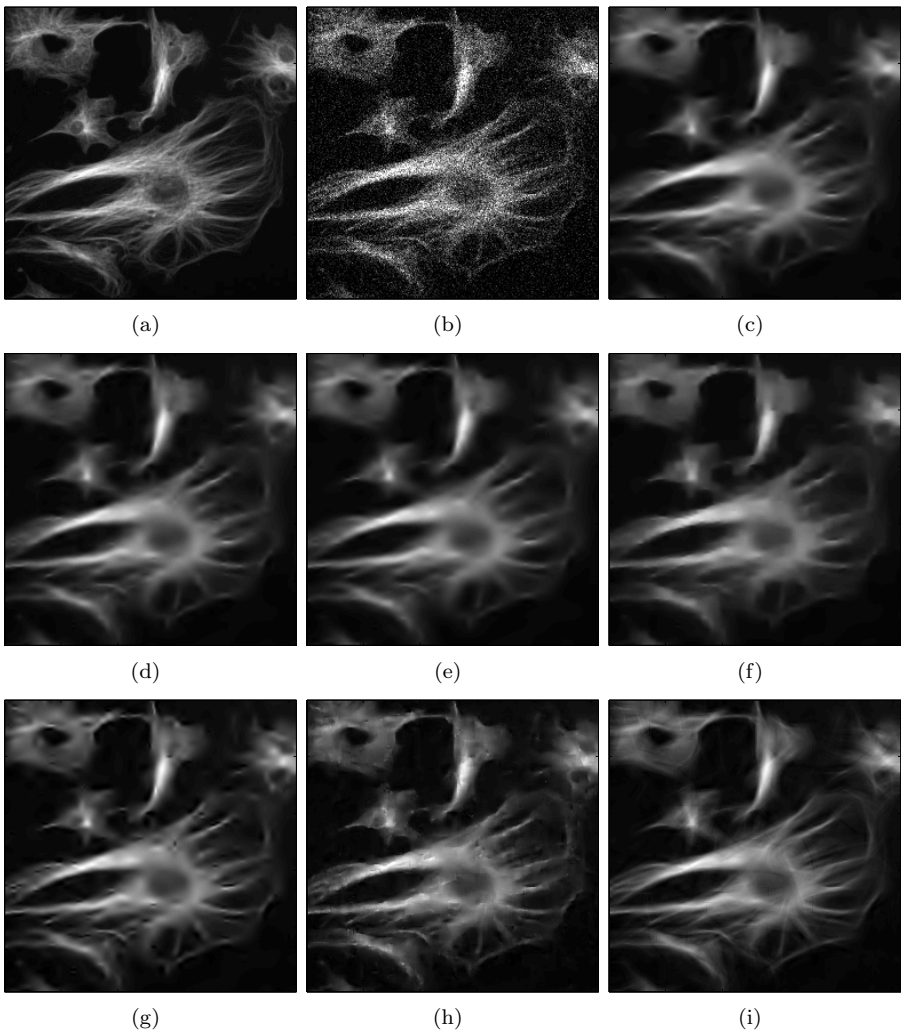
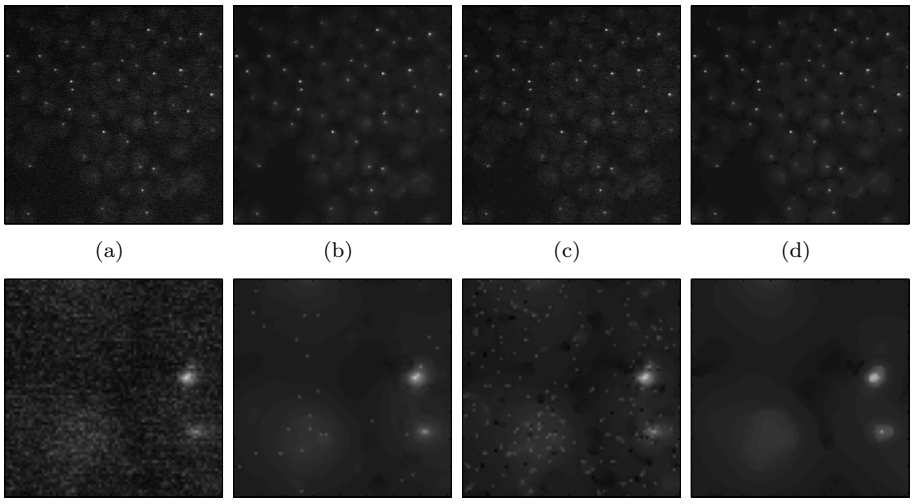


FIG. 4: Débruitages poissoniens des tubulines en fluorescence (taille d'image : 256×256). (a) image d'intensités poissoniennes (intensité $\in [0.53, 16.93]$) ; (b) image bruitée ; image débruitée par (c) Anscombe (ondelettes non-décimées (UWT), banc de filtres 7/9, $J = 4$, FDR $= 0.1$, NMISE $= 0.095$) ; (d) Haar-Fisz (UWT, banc de filtres 7/9, $J = 4$, FDR $= 0.1$, 25 déplacements cycliques où 5 sur chaque axe, NMISE $= 0.096$) ; (e) CVS (UWT, banc de filtres 7/9, $J = 4$, FDR $= 0.1$, NMISE $= 0.10$) ; (f) Haar+FDR (Haar UWT, $J = 4$, FDR $= 0.1$, NMISE $= 0.10$) ; (g) MS-VST+UWT (UWT, banc de filtres 7/9, $J = 4$, $N_{\max} = 5$ itérations HSD, FDR $= 0.1$, NMISE $= 0.090$) ; (h) platelets ($\gamma = 1/3$, 25 déplacements aléatoires cycliques, NMISE $= 0.079$) ; (i) MS-VST+Curvelets ($J = 4$, $N_{\max} = 5$ itérations HSD, FDR $= 0.1$, NMISE $= 0.078$).



(a) (b) (c) (d)

(e) (f) (g) (h)

FIG. 5: Débruitage avec IUWT d'une image des télomères de cellules saccharomyces cerevisiae. h = 2D B_3-Spline filtre, nombre d'échelles $J = 5$, et FDR = 10^{-2}. (a) image observée ; (b) image débruitée en supposant un bruit gaussien ; (c) image débruitée en supposant un bruit poissonien ($N_{max} = 10$ itérations HSD ; (d) image débruitée en supposant un bruit MPG ($N_{max} = 10$ itérations HSD) ; (e) zoom local de (a) ; (f) zoom local de (b) ; (g) zoom local de (c) ; (h) zoom local de (d).

spots. Le fond régulier peut être facilement éliminé grâce aux moments nuls des ondelettes. L'image corrélée est seuillée et binarisée, et finalement les composants connexes sont déterminés comme les spots putatifs. En faisant cela, nous pouvons montrer que le taux de fausses détections d'une image de bruit MPG homogène ($\lambda_i \equiv \lambda$) est borné par le FDR contrôlé dans le domaine d'ondelettes.

3.3.1 Résultats

La Figure. 6 permet de comparer notre méthode de détection avec celle basée sur la transformée Anscombe généralisée (GAT) pour une grille de spots simulée. Nous avons trouvé que pour tous les FDRs testés, notre approche est toujours plus sensible et le taux de vrais positifs reste environs 20% plus élevé que celui de GAT.

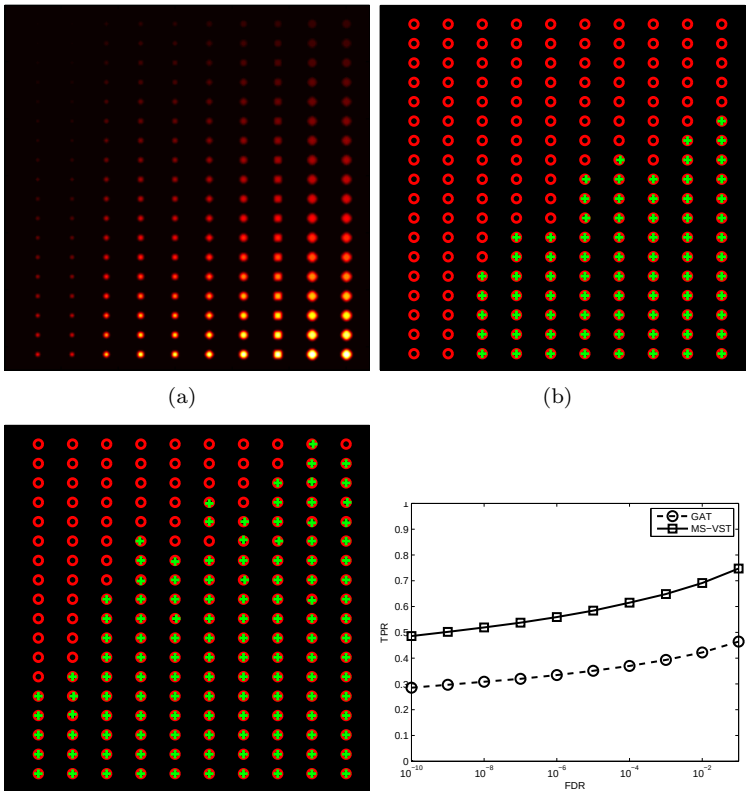
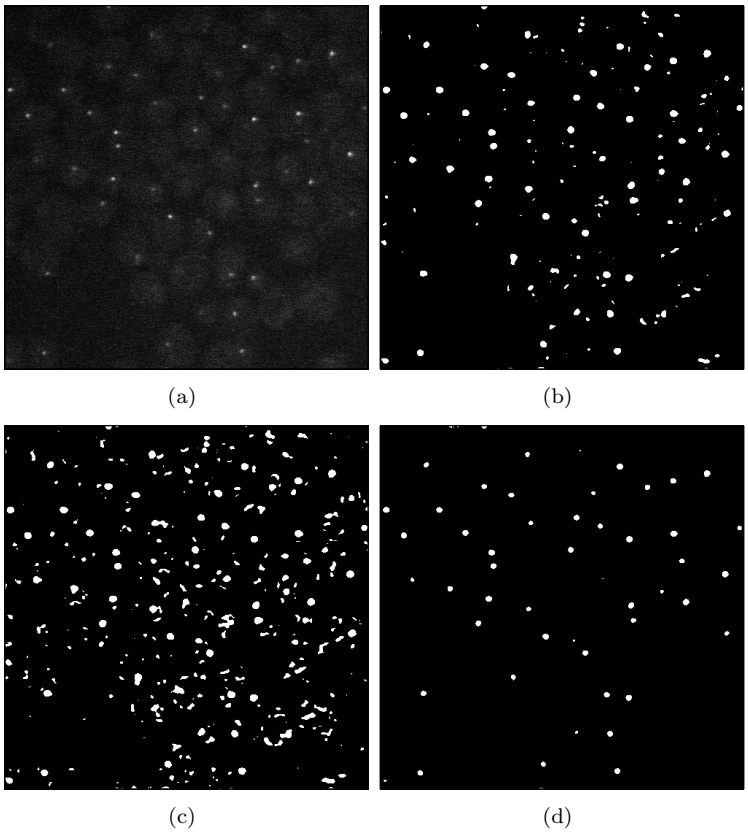


FIG. 6: Détection des spots simulés. Paramètres MPG : $\alpha = 20, \mu = 10, \sigma = 1$; h = 2D B_3-spline filtre, et les indices d'échelles corrélées $I_s = \{3, 4\}$. (a) sources simulées (amplitudes $\lambda_A \in [0.05, 30]$; fond = 0.05) ; (b) un exemple de la détection basée sur GAT (FDR= 10^{-2} ; 75 bonnes détections (croix vertes) parmi 180 (cercles rouges), 0 fausse détection) ; (c) un exemple de la détection basée sur la MS-VST (FDR= 10^{-2} ; 125 bonnes détections (croix vertes) parmi 180 (cercles rouges), 0 fausse détection) ; (d) Les taux de vrais positifs en fonction du FDR$\in [10^{-9}, 10^{-1}]$.

La Figure 7 présente la détection des spots (télomères) vus sous DSCM en supposant différentes hypothèses du bruit. La préparation du spécimen biologique garantit que chaque cellule dans l'image contient au plus un seul télomère. Par conséquent, des détections multiples dans une même cellule sont les faux positifs. Les cellules peuvent être vues dans la Figure 7(a) qui ont des formes de bulbes. Encore une fois, la bonne hypothèse du bruit (MPG) conduit au résultat avec le moins de faux positifs.



(a) (b)

(c) (d)

FIG. 7: Détection des télomères de cellules saccharomyces cerevisiae. h = 2D B_3-Spline filtre, FDR = 10^{-2}, et $I_s = \{3, 4\}$. (a) image observée ; (b) spots détectés en supposant un bruit gaussien ; (c) spots détectés en supposant un bruit poissonien ; (d) spots détectés en supposant un bruit MPG.

4 Détection super-résolutive des objets biologiques

La résolution classique du WFFM 2D est prédite par le critère de Rayleigh, qui est donnée par $d_R := 0.61\lambda_{em}/\text{NA}$. Dans le sens de Rayleigh, deux sources ponctuelles séparées par une distance latérale de moins de d_R ne sont pas distinguables. Il est clair que cette limite de résolution d_R est formulée de façon

déterministe. Il ne prend pas en compte la nature statistique des images de fluorescence.

Nous avons donc ré-étudié les limites de la résolution à l'aide des théories de la détection, de l'estimation et de l'information. Notamment, nous avons proposé d'utiliser la VST pour étudier la résolution pour les observations poissoniennes ou MPG. Cette approche simplifie considérablement l'analyse en gaussianisant le problème, et conduit à des résultats asymptotiquement consistants qui s'expriment sous formes closes. La conclusion principale est que la résolution n'est pas limitée, et la limite de Rayleigh peut effectivement être dépassée si le rapport signal sur bruit (SNR) est suffisamment élevé. Nous avons aussi étendu une méthode de super-résolution pour détecter des spots ou des bâtonnets séparés par des distances inférieures à la résolution de Rayleigh (super-résolution).

4.1 Limites de la résolution

Premièrement, nous avons ré-étudié la limite de résolution par la théorie de la détection. Nous considérons la situation idéale où nous observons une image $\mathbf{y} = (y(\mathbf{x}_n))_n$ composée de deux sources ponctuelles placées sur l'axe x séparées latéralement d'une distance d, i.e.,

$$\begin{aligned} y(\mathbf{x}_n) &= f(\mathbf{x}_n; d) + w(\mathbf{x}_n),\ n = 1, \dots, N && (29) \\ f(\mathbf{x}_n; d) &= A \cdot h(x_n - d/2, y_n, z_n) + A \cdot h(x_n + d/2, y_n, z_n) && (30) \end{aligned}$$

où $\mathbf{x}_n := [x_n, y_n, z_n]^T$ est un échantillon, N le nombre de pixels, $w(\mathbf{x}_n) \sim \mathcal{N}(0, \sigma^2)$ le processus du bruit gaussien, A l'amplitude des sources, et h la PSF. Pour décider s'il y a une ou deux sources, nous testons les hypothèses binaires $H_0 : d = 0$ vs. $H_1 : d > 0$. Etant donnés p_f et p_d respectivement le FPR et le taux minimum de vrais positifs à atteindre, nous pouvons déduire que

$$d \geq \frac{2}{\sqrt{A/\sigma}(\mathbf{h}_{11}^T \mathbf{h}_{11})^{1/4}} \left[\Phi^{-1}(1 - p_f) - \Phi^{-1}(1 - p_d)\right]^{1/2} \quad (31)$$

où Φ est la distribution cumulée (*cdf*) de la loi normale standard. Si le SNR est défini par $\mathrm{SNR} := A/\sigma$, nous avons $d \propto \mathrm{SNR}^{-1/2}$.

Pour les observations MPG, le modèle d'image est :

$$\begin{aligned} y(\mathbf{x}_n) &= \alpha F(\mathbf{x}_n; d) + w(\mathbf{x}_n), \quad F(\mathbf{x}_n; d) \sim \mathcal{P}(f(\mathbf{x}_n; d)), w(\mathbf{x}_n) \sim \mathcal{N}(\mu, \sigma^2) && (32) \\ f(\mathbf{x}_n; d) &= A \cdot h(x_n - d/2, y_n, z_n) + A \cdot h(x_n + d/2, y_n, z_n) \end{aligned}$$

(32) se réduit au modèle poissonien si $\alpha = 1$ et $\mu = \sigma = 0$. Nous appliquons la VST pour gaussieniser le problème, i.e.,

$$T(y(\mathbf{x}_n)) = 2\sqrt{f(\mathbf{x}_n; d)} + w_T(\mathbf{x}_n) \quad (33)$$

où $T(x) = 2/\sqrt{\alpha}\,\mathrm{sgn}(x + c^*)|x + c^*|^{1/2}$, et $w_T(\mathbf{x}_n) \sim \mathcal{N}(0, 1)$. Aucune information n'est perdue puisque T est inversible. Remarquons que cette gaussienisation par la VST est seulement valable pour des intensités poissoniennes élevées.

Grâce à cette gaussienisation, nous pouvons montrer que

$$d \geq \frac{2 \cdot 2^{1/4}}{A^{1/4} \left[\sum_{n=1}^N h_{11}(\mathbf{x}_n)^2 / h(\mathbf{x}_n)\right]^{1/4}} \left[\Phi^{-1}(1 - p_f) - \Phi^{-1}(1 - p_d)\right]^{1/2} \quad (34)$$

Selon la statistique MPG, nous pouvons définir SNR $:= \alpha A/\sqrt{\alpha^2 A + \sigma^2}$. (34) montre encore que $d \propto \mathrm{SNR}^{-1/2}$.

Deuxièmement, la limite résolutive peut être également calculée par la théorie de l'estimation. Sous ce point de vue, la distance entre les sources d dans le modèle (29) est un paramètre à estimer. Ensuite, la borne de Cramér-Rao (CRLB) peut être appliquée, ce qui donne une limite inférieure de la précision (l'écart type) de l'estimation parmi tous estimateurs sans biais. Par conséquent, d est considérée résolvable si d est supérieure à l'écart type donné par CRLB. Pour des observations gaussiennes, nous avons la limite résolutive :

$$d \geq \frac{12\sigma_{xy}}{\left[36\sqrt{3}\sigma_{xy}\frac{A}{\sigma}\sqrt{f_{xy}^2 f_z \pi^{3/2}\sigma_z} - 15\right]^{1/2}} \tag{35}$$

où σ_{xy} et σ_z sont les écart-types de la PSF gaussienne, et f_{xy} et f_z sont respectivement les fréquences latérales et axiales d'échantillonnage de l'image. (35) montre que $d_0 \propto \mathrm{SNR}^{-1/2}$.

Pour des observations MPG, en appliquant la VST qui gaussienise notre problème nous pouvons montrer que :

$$d \geq \frac{2\sqrt{2}\sigma_{xy}}{\left[4\sqrt{2A}\sigma_{xy}\sqrt{f_{xy}^2 f_z \pi^{3/2}\sigma_z\sqrt{2}} - 2\right]^{1/2}} \tag{36}$$

(36) montre encore que $d_0 \propto \mathrm{SNR}^{-1/2}$.

Troisièmement, la limite résolutive a été étudiée par la théorie de l'information. Nous considérons une image y avec N pixels au total :

$$y(\mathbf{x}_n) = f(\mathbf{x}_n) + w(\mathbf{x}_n), \quad 1 \leq n \leq N \tag{37}$$

(37) peut être considéré comme l'utilisation de N canaux gaussiens parallèles. Ces canaux ont une capacité, c'est-à-dire, la limite supérieure du nombre de bits d'information par l'utilisation qui peut être transmis de façon fiable. Ceci est prédit par le théorème de la capacité de Shannon. Nous considérons que f est composée de K sources ponctuelles où le nombre K est supérieurement borné. De plus, nous supposons que K suive une distribution dont la moyenne est μ_K. Nous avons calculé la capacité de nos canaux qui borne la quantité maximale d'information dans notre image, à partir de laquelle nous avons dérivé la limite inférieure des séparations résolvables entre ces K sources :

$$d \geq \frac{V^{\frac{1}{3}}}{\left(\frac{A}{\sigma}\right)^{\frac{2}{3}}\left[\frac{1}{2}\sigma_{xy}^2\sigma_z\pi^{3/2}\mu_K f_{xy}^2 f_z \log_2 e\right]^{\frac{1}{3}}} \tag{38}$$

où V est la volume d'image. Pour des observations MPG, nous avons :

$$d \geq \frac{V^{\frac{1}{3}}}{A^{\frac{1}{3}}\left(4\sqrt{2}\pi^{3/2}\mu_K\sigma_{xy}^2\sigma_z f_{xy}^2 f_z \log_2 e\right)^{\frac{1}{3}}} \tag{39}$$

(38) et (39) montrent que $d \propto \mathrm{SNR}^{-2/3}$. Clairement, ces bornes basées sur la théorie de l'information sont plus lâches que celles issues des théories de la détection et de l'estimation.

4.2 Détection super-résolutive de spots

Nous avons montré qu'il est théoriquement possible d'identifier des sources ponctuelles séparées par des distances plus petites que celle de Rayleigh, à condition que nous ayons un SNR suffisamment élevé. Pratiquement, nous avons conçu une approche de détection qui est composée d'une première pré-détection suivie par l'ajustement itératif de modèle.

La pré-détection vise à offrir une première estimation du nombre, des positions et des amplitudes des spots dans l'image. Nous supposons une image MPG y décrite par :

$$y(\mathbf{x}_n) = \alpha F(\mathbf{x}_n) + w(\mathbf{x}_n) \tag{40}$$

$$F(\mathbf{x}_n) \sim \mathcal{P}(f(\mathbf{x}_n)), \quad f(\mathbf{x}_n) = b + \sum_{i=1}^{K} A_i h(\mathbf{x}_n - \mathbf{p}_i) \tag{41}$$

$$w(\mathbf{x}_n) \sim \mathcal{N}(\mu, \sigma^2) \tag{42}$$

où $b \geq 0$ représente le fond, A_i et $\mathbf{p}_i$ représentent respectivement l'amplitude et la position du i-ième spot, K le nombre total des spots, et h le profil d'une source avec une amplitude normalisée à 1.

Notre pré-détection est réalisée comme suit. Nous appliquons le schéma MS-VST+IUWT pour obtenir le support de multi-résolution. Ensuite, nous reconstruisons l'image par les coefficients significatifs d'IUWT en mettant à zéro la bande d'approximation. Cela permet d'éliminer le bruit et le fond régulier de l'image. Puis, dans l'image reconstruite, les pixels de valeurs négatives sont mis à zéro. Enfin, tous les maxima locaux sont extraites comme spots putatifs.

Les paramètres des spots calculés à ce stade doivent être améliorés pour plusieurs raisons. La pré-détection pourrait produire de faux positifs à cause du bruit. Si deux ou plusieurs spots sont très proches, i.e. la fluorescence se chevauche, la pré-détection pourrait produire une seule réponse. Nous cherchons donc à améliorer la pré-détection en recourant à l'estimation par un modèle paramétrique.

En supposant que la distribution d'intensité d'une source est décrite par une fonction gaussienne de l'écart-type σ_0, et que la PSF est également approximée par une gaussienne, le profil observé pour une source est donné par :

$$h(x, y, z) = \exp\left[-\frac{x^2 + y^2}{2(\sigma_0^2 + \sigma_{xy}^2)} - \frac{z^2}{2(\sigma_0^2 + \sigma_z^2)}\right]$$

(41) est donc un mélange gaussien et l'ensemble des paramètres du signal f est donné par $\Theta_K = \{(\mathbf{p}_i)_{1 \leq i \leq K}, (A_i)_{1 \leq i \leq K}, \sigma_0, b\}$ où K indique l'ordre du modèle (nombre de spots). Pour spécifier que f dépend de Θ_K, nous écrivons $f(\mathbf{x}; \Theta_K)$.

Etant donné une valeur K, Θ_K peut être estimé par la méthode des moindres carrés appliquée sur les observations gaussienisées par la VST (notée par T, voir (43)), sachant que l'estimateur de maximum de vraisemblance pour les observations MPG n'a pas de formes closes.

$$\hat{\theta}_K = \arg\min_{\Theta_K} \sum_{n=1}^{N} \left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n; \Theta_K)}\right]^2 \tag{43}$$

Des contraintes sur les paramètres peuvent aussi être imposées selon les connaissances *a priori*. Outre cette méthode, on pourrait aussi considérer l'approche

des moindres carrés généralisés :

$$\hat{\theta}_K = \arg\min_{\Theta_K} \sum_{n=1}^{N} \frac{(y(\mathbf{x}_n) - \alpha f(\mathbf{x}_n; \Theta_K) - \mu)^2}{\alpha^2 f(\mathbf{x}_n; \Theta_K) + \sigma^2} \tag{44}$$

Il nous reste donc à déterminer l'ordre du modèle K. Ceci est effectué par les tests d'hypothèses itératifs. Etant donné un modèle de l'ordre M (Θ_M), nous ajustons le nouveau modèle Θ_{M+1} dans l'image pour savoir si Θ_{M+1} donne une approximation de l'image significativement meilleure que Θ_M. Ceci est réalisé en testant $H_0 : K = M$ vs. $H_1 : K > M$. Sous H_0, la statistique à tester est celle de Fisher :

$$F := \frac{\sum_{n=1}^{N} \left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n; \hat{\theta}_M)} \right]^2 / r_M}{\sum_{n=1}^{N} \left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n; \hat{\theta}_{M+1})} \right]^2 / r_{M+1}} \tag{45}$$

La distribution de F est supposée celle de Fisher avec r_M et r_{M+1} degrés de liberté. Ici $r_M = N - p_M$, et $p_M = 4M + 2$ est le nombre des paramètres libres du modèle Θ_M.

Nous ajoutons itérativement un nouveau spot et le nouveau modèle est accepté lorsque le test ci-dessus reste toujours significatif. Nous appelons cette étape « bottom-up » puisque l'ordre du modèle testé se croît de façon monotone. La « bottom-up » nous permet d'identifier les sources proches que la pré-détection n'a pas arrivé à résoudre. A la suite de la « bottom-up », nous diminuons itérativement l'ordre en supprimant chaque fois un spot dans le modèle. Les tests sont donc inversés : $H_0 : K = M$ vs. $H_1 : K < M$. Cette étape, qu'on nomme « top-down », nous permet d'enlever les faux positifs de la pré-détection.

4.2.1 Résultats

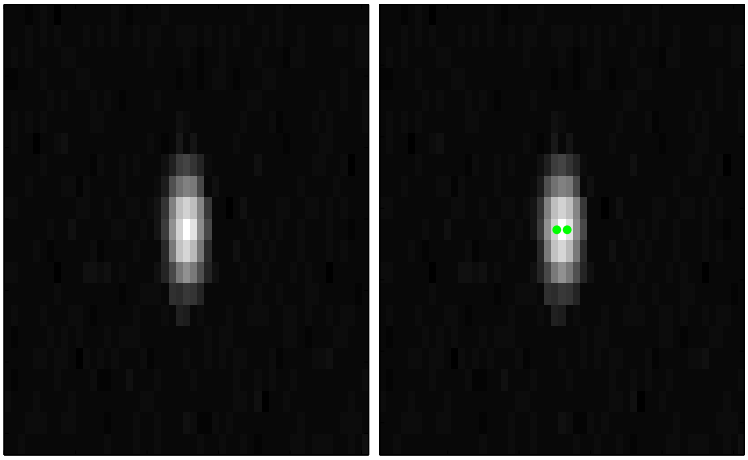
Nous avons d'abord testé la détection pour un spot individuel. Les résultats pour le cas SNR = 15 sont résumés dans le Tableau 4. Toutes les erreurs de localisation sont de l'ordre de quelques nanomètres. La précision de l'estimation de l'amplitude est également élevée, avec des erreurs de quelques pour cent seulement.

Cette approche permet de s'affranchir de la limite de Rayleigh dans la détection lorsque le SNR est suffisant. Par exemple, la Figure 8 montre une configuration de deux spots séparés par une distance deux fois plus petite que celle de Rayleigh. Pour un SNR = 15, notre méthode est capable de les identifier de façon fiable.

TAB. 4: Performance de la localisation (SNR = 15)

	CDR%	RMSE$_x$ (nm)	RMSE$_y$ (nm)	RMSE$_z$ (nm)	PRE (A)%
F-test*	(100, 100)	(1.65, 1.53)	(1.77, 1.82)	(4.66, 5.74)	(4.50, 5.54)
F-test	(100, 100)	(1.65, 1.53)	(1.77, 1.82)	(4.66, 5.74)	(4.50, 5.54)

Dans les parenthèses, les premières valeurs sont issues de la méthode des moindres carrés basée sur la VST, et les deuxièmes sont données par l'approche des moindres carrés généralisés. La p-valeur pour F-test* est donnée par $\alpha = 0.05$; celle pour F-test est fixée à $\alpha = 0.1$. 50 réplications sont effectuées pour chaque méthode d'ajustement. CDR (le taux des détections correctes) ; RMSE (la racine de l'erreur quadratique moyenne) ; PRE (l'erreur relative du paramètre).



(a) (b)

FIG. 8: Deux spots simulés séparés par une distance latérale $d_{xy} = d_R/2$ et SNR = 15. Ils peuvent être résolus de façon fiable par les tests de Fisher (p-valeur $\alpha = 0.05$) avec la méthode des moindres carrés généralisés, ou par les mêmes tests (p-valeur $\alpha = 0.1$) avec la méthode des moindres carrés basée sur la VST. (a) le plan xz ; (b) les positions des spots sont indiquées en vert pour (a).

4.3 Détection super-résolutive de bâtonnets

La détection super-résolutive des spots a été étendue pour la détection des bâtonnets. Nous avons conçu un modèle paramétrique pour un bâtonnet individuel qui est décrit par la convolution de la PSF avec un segment de longueur $L = 2l$. Le segment est supposé avoir une section transversale gaussienne d'un

écart-type σ_0. Le modèle d'un bâtonnet est donné par :

$$
\begin{aligned}
R(x,y,z) &= C\int_{-1}^{1} l\exp\left[-\frac{(x-tl\sin(\phi)\cos(\theta))^2+(y-tl\sin(\phi)\sin(\theta))^2}{2\sigma_1^2}-\frac{(z-tl\cos(\phi))^2}{2\sigma_2^2}\right]dt \\
&= \frac{A}{2}\exp\left[-\frac{\sigma_2^2(\sin(\theta)\sin(\phi)x-\cos(\theta)\sin(\phi)y)^2}{2\sigma_1^2(\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2)}\right. \\
&\left.-\frac{x^2\cos(\phi)^2+y^2\cos(\phi)^2+z^2\sin(\phi)^2-\cos(\theta)\sin(2\phi)xz-\sin(\theta)\sin(2\phi)yz}{2(\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2)}\right] \\
&\cdot\left(\operatorname{erf}\left[\frac{\sqrt{2}(l\sigma_2^2\sin(\phi)^2+l\cos(\phi)^2\sigma_1^2+\cos(\theta)\sin(\phi)\sigma_2^2x+\sin(\theta)\sin(\phi)\sigma_2^2y+\cos(\phi)\sigma_1^2z)}{2\sigma_1\sigma_2\sqrt{\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2}}\right]\right. \\
&\left.+\operatorname{erf}\left[\frac{\sqrt{2}(l\sigma_2^2\sin(\phi)^2+l\cos(\phi)^2\sigma_1^2-\cos(\theta)\sin(\phi)\sigma_2^2x-\sin(\theta)\sin(\phi)\sigma_2^2y-\cos(\phi)\sigma_1^2z)}{2\sigma_1\sigma_2\sqrt{\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2}}\right]\right) \\
&\cdot\left(\operatorname{erf}\left[\frac{\sqrt{2}l(\sigma_2^2\sin(\phi)^2+\sigma_1^2\cos(\phi)^2)}{2\sigma_1\sigma_2\sqrt{\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2}}\right]\right)^{-1}
\end{aligned}
\tag{46}
$$

où A est l'amplitude, C une constante telle que R a une valeur maximum de A, ϕ et θ sont respectivement les angles formés entre le bâtonnet et l'axe z et l'axe x (voir Fig. 9), $\sigma_1 = [\sigma_{xy}^2+\sigma_0^2]^{1/2}$, $\sigma_1 = [\sigma_z^2+\sigma_0^2]^{1/2}$, et $\operatorname{erf}(x) := 2\Phi(\sqrt{2}x)-1$ est la fonction d'erreur.

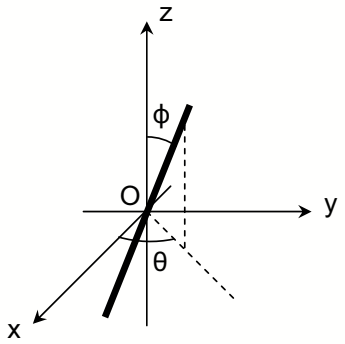


FIG. 9: Le modèle d'un bâtonnet.

Ainsi, notre image est modélisée par un mélange de bâtonnets où chacun est décrit par (46). Les mêmes étapes « bottom-up » et « top-down » comme dans le cas de spots sont effectuées pour localiser un nombre inconnu des bâtonnets en super-résolution.

4.3.1 Résultats

Cette approche a été appliquée avec succès pour détecter les bactéries Shigella sous DSCM. Un exemple est montré par la Figure 10 où l'algorithme a réussi à résoudre les bactéries très proches qui sont difficile à distinguer visuellement.

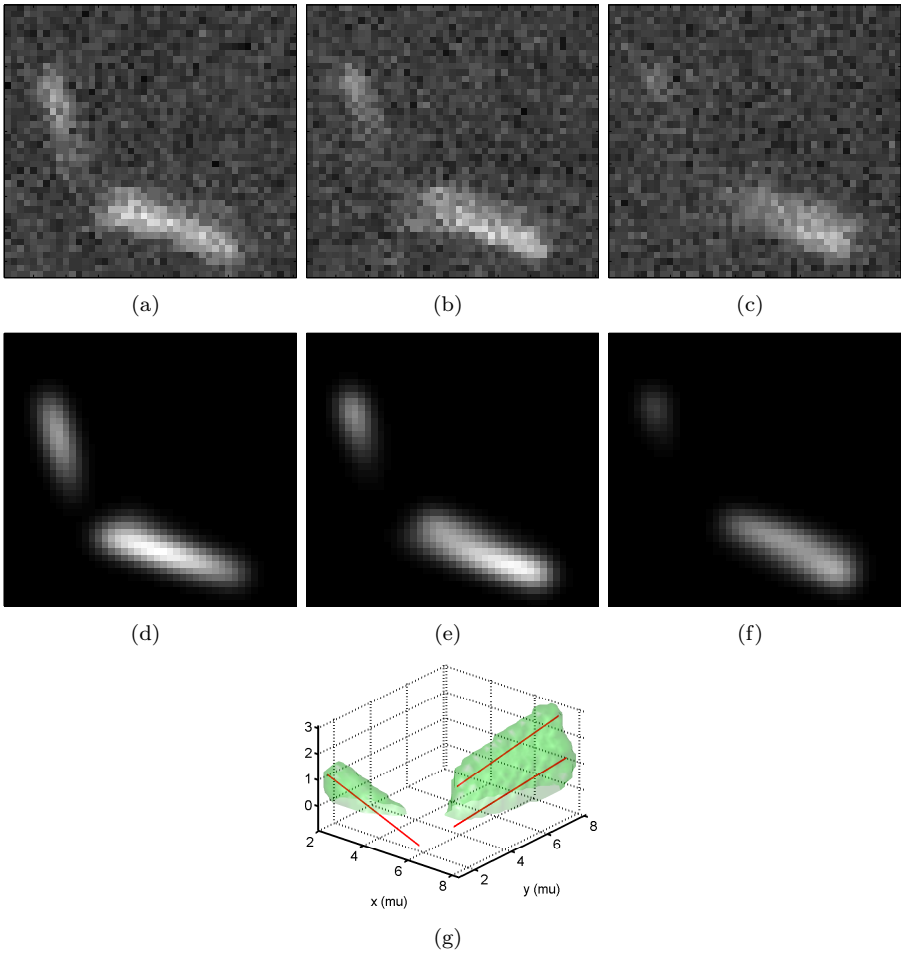


FIG. 10: Détection des bactéries Shigella. Trois bactéries existent dans l'image dont deux se situent en bas qui sont très proches et difficile à distinguer visuellement. L'existence des deux bactéries à cet endroit-là a été confirmée par leur séparation dans les images à la suite dans la séquence de vidéo. Notre algorithme a réussi à identifier les trois bactéries. La p-valeur dans les tests de Fisher a été fixée à $\alpha = 0.05$. Coupe optique observée à (a) $z = 0.2\mu m$; (b) $z = 0.8\mu m$; (c) $z = 1.4\mu m$; Modèle généré par les paramètres de la détection observé à (d) $z = 0.2\mu m$; (e) $z = 0.8\mu m$; (f) $z = 1.4\mu m$; (g) la reconstruction 3D des bactéries détectées.

5 Conclusion

Cette thèse apporte une contribution dans trois sujets de l'imagerie microscopique à fluorescence. Nous avons premièrement étudié les approximations gaussiennes des moindres carrés pour les PSFs du WFFM, LSCM et DSCM limitées par la diffraction. Les paramètres (quasi-)optimaux ont été dérivés et ont des formes explicites. Nous avons constaté que les approximations 2D sont très précises ; aucune approximation précise n'existe pour la PSF 3D du WFFM ; pour les sténopés de tailles standards, les approximations sont précises pour la PSF 3D du DSCM et quasiment parfaite pour celle du LSCM. Ces modèles gaussiens permettent de simplifier significativement la modélisation des objets biologiques.

En perspectives, les approximations pourraient être améliorées en considérant les modèles vectoriels de diffraction. Nous pourrions aussi envisager d'étudier les approximations en présence d'aberrations optiques. Pourtant, ce ne serait pas trivial parce que les aberrations détruisent les symétries de la PSF. De ce fait, les approximations par les fonctions autres que celles gaussiennes pourraient être plus appropriées.

Le second problème que nous avons abordé est le débruitage des images de fluorescence qui ont des statistiques poissoniennes ou MPG. Deux approches sont proposées pour la suppression du bruit poissonien. L'une repose sur les tests d'hypothèses dans le domaine Bi-Haar, adaptée pour estimer rapidement les intensités régulières dans les données de grandes tailles. L'autre fait appel à une VST bien conçue permettant de stabiliser et gaussieniser un processus poissonien filtré. La combinaison de cette VST avec des dictionnaires multi-échelles tels que les ondelettes, les ridgelets et les curvelets conduit à un estimateur MS-VST très efficace pour restaurer les sources de diverses morphologies observées avec un faible flux de photons. Nous avons aussi étendu l'approche MS-VST pour débruiter les données MPG.

Concernant les travaux futurs, nous pouvons envisager la combinaison simultanée de la VST avec plusieurs dictionnaires pour adapter simultanément de différentes structures morphologiques. Cela pourrait être fait de façon similaire comme dans le cas du débruitage gaussien. De plus, ce serait aussi intéressant d'appliquer cette MS-VST aux problèmes de déconvolution.

La troisième partie de notre travail considère la détection super-résolutive des objets. Nous avons revu et étendu les limites de la résolution en utilisant les théories de la détection, de l'estimation et de l'information. Nous avons aussi établi un algorithme capable de localiser très précisément les spots ou les bâtonnets dans une image et d'identifier les sources séparées par des distances inférieures à celle de Rayleigh (super-résolution). Nous envisageons de combiner cette méthode de détection avec des approches de séparation de sources, ce qui permet de détecter les objets avec un fond complexe (e.g. un fond structuré).

Contents

Acknowledgments v
Résumé vii
Tables v
Figures vi
Abbreviations viii

Introduction **1**

I Optical System Modeling for Fluorescence Microscopes **3**

1 Diffraction of a Lens System **5**
1.1 Scalar Diffraction Integral for a Nonparaxial Lens 5
1.1.1 Integral theorem of Helmholtz and Kirchhoff 6
1.1.2 Classical diffraction theories 7
1.1.3 Diffraction model for a nonparaxial lens 11
1.2 Scalar Debye integrals 13
1.3 Aberrations and Vectorial Diffractions 15
1.3.1 Aberrations 15
1.3.2 Vectorial diffractions 16
1.4 Conclusion 16

2 Fluorescence Microscope PSFs **19**
2.1 Optical Models of WFFM, LSCM and DSCM 20
2.1.1 WFFM optical model 20
2.1.2 LSCM optical model 20
2.1.3 DSCM optical model 21
2.2 Photodynamics of Fluorophores 24
2.3 Diffraction-limited PSFs of WFFM, LSCM and DSCM 27
2.3.1 WFFM PSF 28
2.3.2 LSCM PSF 29
2.3.3 DSCM PSF 30
2.4 Conclusion 31

3 Gaussian Approximations of Diffraction-limited Microscope PSFs 33
3.1 Gaussian Approximations of the PSF Models with an L^∞ Constraint 34
3.1.1 Gaussian approximations of 2D WFFM PSFs 35
3.1.2 Gaussian approximations of 2D LSCM PSFs 36
3.1.3 Gaussian approximations of 2D DSCM PSFs 38
3.1.4 Gaussian approximations of 3D PSFs 38
3.2 Gaussian Approximations of the PSF Models with an L^1 Constraint 39
3.3 Numerical Evaluations of the Approximations 40
3.3.1 Numerical results 40
3.3.2 Discussion 41
3.4 Conclusion 42

II Fluorescence Image Restoration 51

4 Noise Sources in Fluorescence Microscope 53
4.1 CCD Noise Sources 53
4.1.1 Noise in wide-field microscopes 54
4.1.2 Noise in disc-scanning confocal microscopes 55
4.2 PMT Noises 55
4.2.1 Noise in laser-scanning confocal microscopes 56
4.3 Conclusion 56

5 Poisson Noise Removal 59
5.1 Existing Methods for Denoising Poisson Data 60
5.1.1 Wavelet background 60
5.1.2 VST-based denoisers 63
5.1.3 Wavelet-domain filters 65
5.1.4 Hypothesis tests 66
5.1.5 Bayesian and penalized ML estimators 68
5.2 Denoising by Hypothesis Tests in the Biorthogonal Haar Domain 70
5.2.1 Bi-Haar domain testing 71
5.2.2 Thresholds for individual tests 72
5.2.3 Summary of the denoising algorithm 74
5.2.4 Results 76
5.3 Denoising by Multi-scale Variance Stabilizing Transform 78
5.3.1 VST of a filtered Poisson process 79
5.3.2 Denoising by MS-VST+wavelets 82
5.3.3 Denoising by MS-VST+ridgelets 91
5.3.4 Denoising by MS-VST+curvelets 95
5.4 Conclusion 98

6 Mixed-Poisson-Gaussian Noise Removal **101**
6.1 VST for a Filtered MPG Process . . . 102
6.1.1 Optimal parameter of the VST . . . 103
6.2 Fluorescence Image Denoising Using MS-VST . . . 104
6.2.1 IUWT-based MS-VST . . . 106
6.2.2 Detection of significant coefficients by FDR . . . 106
6.2.3 Sparsity-driven iterative reconstruction . . . 107
6.2.4 Results . . . 108
6.3 Detection of Fluorescent Spots Using MS-VST . . . 110
6.3.1 Results . . . 113
6.4 Conclusion . . . 114

III Super-resolution Object Detection **119**

7 Super-resolution Spot Detection **121**
7.1 Resolution Limit . . . 122
7.1.1 Detection-theoretic resolution limit . . . 122
7.1.2 Estimation-theoretic resolution limit . . . 125
7.1.3 Information-theoretic resolution limit . . . 129
7.2 Super-resolution Spot Detection . . . 134
7.2.1 Pre-detection . . . 135
7.2.2 Model fitting and model-order selection . . . 136
7.2.3 Detection within Gaussian data . . . 140
7.2.4 Experiments . . . 141
7.2.5 Fast super-resolution detection . . . 148
7.3 Conclusion . . . 149

8 Super-resolution Rod Detection **151**
8.1 Parametric Rod Model . . . 151
8.1.1 Asymptotic models . . . 153
8.2 Rod Pre-detection . . . 153
8.3 Rod Mixture Model Fitting . . . 154
8.4 Experiments . . . 155
8.4.1 Localizing an individual rod . . . 155
8.4.2 Resolving close rods . . . 157
8.4.3 Detecting Shigella flexneri . . . 158
8.5 Conclusion . . . 158

Conclusion and Perspectives **161**

Appendix **167**

A Proofs of the Results **167**
A.1 Proof of Lemma 1 . . . 167
A.2 Proof of Theorem 1 . . . 167
A.3 Proof of Theorem 2 . . . 168
A.4 Proof of Proposition 1 . . . 169
A.5 Proofs of Proposition 2 and Proposition 3 . . . 169
A.6 Proof of Proposition 4 . . . 170
A.7 List of Maclaurin Expansions of the PSFs . . . 171
A.8 Proof of Proposition 5 . . . 172
A.9 Proof of Proposition 6 . . . 173
A.10 Proof of Lemma 2 . . . 173
A.11 Proof of Proposition 7 . . . 174
A.11.1 Proof – stage I . . . 174
A.11.2 Proof – stage II . . . 175
A.11.3 Proof – stage III . . . 176
A.11.4 Proof – stage IV . . . 177
A.12 Proofs of Theorem 3 and Theorem 4 . . . 177
A.13 Proof of Theorem 5 . . . 179
A.14 Proof of Lemma 3 . . . 180
A.15 Proof of Theorem 6 . . . 181

B MPG Parameter Estimation by Cumulant Matching **183**

C MSE Lower Bounds **185**
C.1 Cramér-Rao Lower Bound . . . 185
C.2 Bhattacharyya Lower Bound . . . 187
C.3 Barankin Lower Bound . . . 187

D Capacity of Real Gaussian Channels **189**

Publications and Softwares **193**

References **195**

Tables

3.1 Gaussian parameters for 2D PSFs 43
3.2 Lateral Gaussian parameters for 3D PSFs (L^∞ constraint) 44
3.3 Axial Gaussian parameters for 3D PSFs (L^∞ constraint) 44
3.4 Approximation errors on WFFM PSFs (L^∞ constraint) 45
3.5 Approximation errors on the 2D paraxial WFFM PSF (L^1 constraint) . . . 45
3.6 Approximation errors on LSCM PSFs (L^∞ constraint) 46
3.7 Approximation errors on DSCM PSFs (L^∞ constraint) 47

5.1 p_H and p_{BH} 72
5.2 $C_{\mathbb{E}}$ and C_{Var} of different filters 82

7.1 Localization performance (SNR = 5) 143
7.2 Localization performance (SNR = 10) 143
7.3 Localization performance (SNR = 15) 144
7.4 Lateral resolution performance 146
7.5 Axial resolution performance 148

8.1 Localization performance (Rod) 155
8.2 Lateral resolution performance (Rods) 157
8.3 Axial resolution performance (Rods) 157

Figures

1.1 Integral theorem of Helmholtz and Kirchhoff 7
1.2 Diffraction by a planar screen 7
1.3 Paraxial diffraction model of a lens. 11
1.4 Fresnel-Kirchhoff diffraction of a nonparaxial lens. 12
1.5 General aberration model. 16

2.1 WFFM optical path 21
2.2 LSCM optical path 22
2.3 Nipkow disk (W is the viewport) 23
2.4 Tandem disk 24
2.5 Pinhole array on a Tandem disk 25
2.6 DSCM optical path 26
2.7 Simplified model of the energy level system for fluorescence molecules . . . 27
2.8 The emission intensity I_{em} as a function the ratio I_{ex}/I_{sat} 28
2.9 Coordinate systems for WFFM PSF modeling. 29
2.10 Coordinate systems for LSCM PSF modeling. 30

3.1 Examples of the Gaussian approximations of PSFs with an L^∞ constraint . 48
3.2 xz-planes of PSFs and of the Gaussian approximations with an L^∞ constraint 49
3.3 Examples of the Gaussian approximations of 2D Paraxial WFFM PSF . . . 50

5.1 Wavelet transform implemented by an iterative filter bank. 61
5.2 Undecimated wavelet transform implemented by an iterative filter bank. . . 62
5.3 Comparison of the thresholds as functions of λ_j 75
5.4 Denoising the "Smooth" function 76
5.5 Denoising a mixture of 15 simulated Gaussian sources 77
5.6 Stabilization performance comparison 82
5.7 Diagrams of the MS-VST in 1D for (a) the IUWT and (b) the standard UWT. 83
5.8 Denoising an image of simulated spots of different radii 92
5.9 Denoising a galaxy image 93
5.10 Poisson denoising of smooth ridges 96
5.11 Poisson denoising of fluorescent tubulins 99

6.1 Stabilization performance comparison (MPG) 105

6.2 Simulated source denoising using IUWT (MPG) 109
6.3 IUWT denoising of a drosophila melanogaster ovary image 110
6.4 IUWT denoising of an image of saccharomyces cerevisiae cells 111
6.5 Detection of simulated spots (MPG) 115
6.6 Detection of subtelomeres of saccharomyces cerevisiae cells 116

7.1 Relation between λ and σ 125
7.2 d_0 as functions of the source amplitude and of NA 128
7.3 Simulated individual spot at different SNRs 142
7.4 Resolving two close spots 147

8.1 The coordinate system for a rod. 152
8.2 Simulated individual rod at different SNRs 156
8.3 Detection of Shigella bacilli I 159
8.4 Detection of Shigella bacilli II 160

A.1 Proof of the integral theorem of Helmholtz and Kirchhoff 168

Abbreviations

ADC	Analog-to-Digital Converter	**AU**	Airy Unit
Bi-Haar	Biorthogonal Haar	**CCD**	Charge-Coupled Device
CLT	Central Limit Theorem	**CVS**	Conditional Variance Stabilization
DSCM	Disc-Scanning Confocal Microscope	**EM**	Expectation Maximization
EM-CCD	Electron-Multiplying CCD	**FDR**	False Discovery Rate
FPR	False Positive Rate	**FRAP**	Fluorescence Recovery After Photobleaching
FWER	Family-Wise Error Rate	**GAT**	Generalized Anscombe Transform
GFP	Green Fluorescent Protein	**HSD**	Hybrid Steepest Descent
KKT	Karush-Kuhn-Tucker	**LRT**	Likelihood Ratio Test
LSCM	Laser-Scanning Confocal Microscope	**MAD**	Median Absolute Deviation
MAP	Maximum A Posteriori	**MDL**	Minimum Description Length
ML	Maximum Likelihood	**MPG**	Mixed-Poisson-Gaussian
NMISE	Normalized Mean Integrated Squared Error	**PMT**	Photomultiplier
PSF	Point-Spread Function	**RMSE**	Root Mean Squared Error
SNR	Signal-to-Noise Ratio	**TPR**	True Positive Rate
UWT	Undecimated Wavelet Transform	**VST**	Variance-Stabilizing Transform
WFFM	Wide-Field Fluorescence Microscope	**a.e.**	almost everywhere
a.s.	almost surely	***cdf***	cumulative distribution function
iff	if and only if	***pdf***	probability distribution function
q**D**	q-Dimensional	**w.r.t.**	with respect to

Introduction

Fluorescence microscopes have become important and widely used tools in investigating underlying biological processes by imaging fluorescence emission from the objects of interest. Modern microscopes, such as the wide-field fluorescence microscope (WFFM), the laser-scanning confocal microscope (LSCM) and the disc-scanning confocal microscope (DSCM), allow monitoring of the specimen in 3D space and in time. These imaging systems provide us with unprecedented large quantity of data such that automated approaches of image processing and analysis are indispensable in order to extract pertinent quantitative information, based on which faithful interpretations can be carried out.

This thesis contributes principally to three aspects of fluorescence microscopic imaging and fluorescence image processing. The three contributions have been organized into the three parts of the thesis.

The first presents the modeling of the optical PSFs of fluorescence microscopes. Chapter 1 and 2 review the classical diffraction theory and derive the general PSF forms for fluorescence microscopes, respectively. In chapter 3, we comprehensively study the least squares Gaussian approximations of the diffraction-limited 2D/3D paraxial/non-paraxial PSFs of WFFM, LSCM and DSCM described using the Debye diffraction integrals. Optimal Gaussian parameters with closed forms are derived for the 2D paraxial WFFM PSF, under both the L^∞ normalization imposing peak matching and the L^1 normalization imposing energy conservation. For the other PSFs, with the L^∞ normalization, near-optimal parameters in explicit forms are derived using Maclaurin series matching. These Gaussian approximative PSF models can be computed fast and greatly simplify the modeling of biological objects under these microscopes.

Our second contribution is for fluorescence image denoising. Noise is one of the major factors that degrade image quality. Chapter 4 points out that the images produced

by LSCM and DSCM have either a Poisson or a MPG statistical nature according to different function modes of the microscope. Regarding Poisson noise removal, after discussing the state of the art we have proposed two approaches in chapter 5. One method is based on biorthogonal Haar-domain hypothesis tests, which is fast in estimating smooth intensities from large datasets. Our second method makes use of a well designed VST allowing to Gaussianize and stabilize a filtered Poisson process. This VST can be combined with most multi-scale transforms yielding MS-VSTs. We show that this MS-VST approach provides a very effective denoiser capable of recovering important structures of various (isotropic, line-like and curvilinear) shapes in (very) low-count images. Furthermore, in chapter 6 this MS-VST method has been extended to remove MPG noise and to extract fluorescent spots embedded in MPG noisy data.

Our third part considers the super-resolution object detection problem. In chapter 7, we have reviewed and extended the results of resolution limits for point-like sources under detection-theoretic, estimation-theoretic and information-theoretic points of view. In particular, we propose to apply the VST to study the limiting resolution with Poisson or MPG data, leading to asymptotically consistent results with closed-form expressions. We have also generalized an existing super-resolution method, which is based on parametric model fitting and model-order selection, to localize an unknown number of point-like sources. Chapter 8 extends this framework to super-resolution detection of rods. This approach allows not only to localize sources having complex spatial configurations, but also to detect objects separated with distances smaller than Rayleigh optical resolution (super-resolution).

At the end of every chapter, a conclusion is drawn to summarize the important points in the chapter. At the end of this thesis, we present a general conclusion and suggest future directions of our work. Finally, all mathematical details are deferred to appendices.

Part I

Optical System Modeling for Fluorescence Microscopes

Chapter 1

Diffraction of a Lens System

For a microscope operating at the far-field region, the behavior of the light distribution is dominated by the diffraction effect of the instrument. The goal of this chapter is to derive Debye's scalar diffraction models of a lens system, which will serve as basis for our analysis and modeling in the sequel. Toward this aim, we first review the classical scalar diffraction theories (Section 1.1.2) which are useful in describing fields of a paraxial lens. Then, we will derive a nonparaxial diffraction integral using Kirchhoff's theory (Section 1.1.3). Based on this integral, Debye's diffraction formulas are re-derived in Section 1.2, which will be useful in describing theoretical microscope point-spread functions (PSFs) (Chapter 2). For self-completeness, diffraction in the presence of aberrations and vectorial diffraction theories will also be briefly introduced in Section 1.3. We conclude this chapter in Section 1.4.

1.1 Scalar Diffraction Integral for a Nonparaxial Lens

In this section, we consider the 3D diffraction field of a nonparaxial lens focusing a beam of light. We will modify the classical planar diffraction model such that the new model adapts to the nonparaxial situation. Then, by making use of the Helmholtz-Kirchhoff integral theorem, we derive the general diffraction formula for a nonparaxial lens, which can be deemed as an extension of the Fresnel-Kirchhoff diffraction formula.

We point out that our treatment is based on the scalar diffraction theory, which neglects the vectorial nature of the EM-field. Fortunately, for the diffraction within a homogeneous medium by an aperture, a scalar approach yields accurate results if 1) the diffracting aperture is much larger than a wavelength; 2) the diffraction field is observed at

a distance from the aperture much larger than a wavelength [1][2][3]. These two conditions are satisfied in our problems.

Let us first recall some fundamental definitions and results. A time-harmonic scalar wave of frequency ω is defined by

$$V(\mathbf{x}, t) := U(\mathbf{x})e^{-i\omega t} \tag{1.1}$$

where $\mathbf{x} := [x, y, z]^T$, and $U(\mathbf{x})$ is usually termed as complex amplitude. The wavenumber is defined as $k := 2\pi/\lambda$, where λ is the wavelength. We denote the refractive index by n.

In a homogeneous medium containing no currents or charges, every scalar component of the EM-field satisfies the following homogeneous wave equation [1]

$$\triangle V - \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2} = 0 \tag{1.2}$$

where v represents the light velocity in the medium, and $\triangle$ is the Laplacian w.r.t. the spatial coordinates. For any time-harmonic scalar wave V satisfying (1.2), U satisfies the Helmholtz equation

$$(\triangle + k^2)U = 0 \tag{1.3}$$

1.1.1 Integral theorem of Helmholtz and Kirchhoff

The integral theorem of Helmholtz and Kirchhoff [1] serves as a fundamental tool for our following analysis. We first establish Lemma 1 whose proof is given in Appendix A.1.

Lemma 1 *Let $\mathcal{S} = \partial\mathcal{W}$ be an arbitrary closed surface and $\mathcal{W}$ be its enclosed volume (see Fig. 1.1). Let V be a time-harmonic scalar wave satisfying (1.2) with $U \in \mathcal{C}^2(\mathcal{W})$. Then, for any $U' \in \mathcal{C}^2(\mathcal{W})$ and U' satisfying (1.3), we have*

$$\int_{\mathcal{S}} U\frac{\partial U'}{\partial \mathbf{n}} - U'\frac{\partial U}{\partial \mathbf{n}}\, dS = 0 \tag{1.4}$$

where $\mathbf{n}$ is the inward normal to $\mathcal{S}$.

If the auxiliary function U' in Lemma 1 is chosen to be a spherical wave emitted from an arbitrary inner point P of $\mathcal{W}$, we can show the following integral theorem of Helmholtz and Kirchhoff (Theorem 1, proof in Appendix A.2). This theorem expresses the solution of (1.2) at P in terms of the values of the solution and its first derivatives at all points on the closed surface $\mathcal{S} = \partial\mathcal{W}$ surrounding that point (see Fig. 1.1).

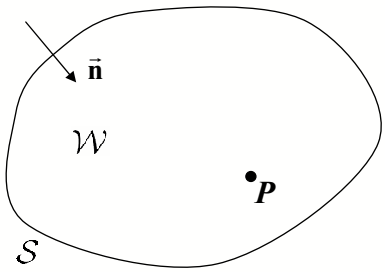


Fig. 1.1: Integral theorem of Helmholtz and Kirchhoff

Theorem 1 (Integral theorem of Helmholtz and Kirchoff [1]) *Suppose a wave* $V(\mathbf{x},t) = U(\mathbf{x})e^{-i\omega t}$ *satisfying (1.2). For a point* P *inside the closed surface* $\mathcal{S}$*, let* $s = \|P\mathbf{x}\|$*, and* $U'(\mathbf{x}) = \frac{e^{iks}}{s}$*. Then we have*

$$U(P) = \frac{1}{4\pi}\int_{\mathcal{S}} U\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial \mathbf{n}}\,dS = \frac{1}{4\pi}\int_{\mathcal{S}} U\frac{\partial U'}{\partial \mathbf{n}} - U'\frac{\partial U}{\partial \mathbf{n}}\,dS \tag{1.5}$$

1.1.2 Classical diffraction theories

In this section, we will briefly review the classical diffraction theories, i.e., Fresnel-Kirchhoff theory and Rayleigh-Sommerfeld theory [1][4]. They are constructed to describe the diffraction field given by a planar aperture, and are applicable for expressing diffraction field of a paraxial lens.

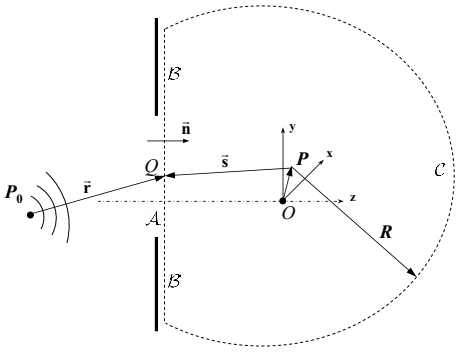


Fig. 1.2: Diffraction by a planar screen

Consider the diffraction model depicted in Fig. 1.2. A point source P_0 emits a spherical wave which propagates through an opening $\mathcal{A}$ in a plane opaque screen. We intend to derive the disturbance at the point P due to the diffraction of the screen. For

this, we surround the point P by a closed surface $\mathcal{S} = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ such that $\mathcal{B}$ is the surface behind the opaque screen, and $\mathcal{C}$ is the large spherical surface shown in Fig. 1.2. We note $r := \|\mathbf{r}\|$ and $s := \|\mathbf{s}\|$, and we assume $r \gg \lambda$ and $s \gg \lambda$. Now, Theorem 1 shows that the disturbance at P can be determined by the disturbance on $\mathcal{S}$, i.e.,

$$U(P) = \frac{1}{4\pi}\left[\int_{\mathcal{A}} + \int_{\mathcal{B}} + \int_{\mathcal{C}}\right]\left\{U\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial \mathbf{n}}\right\} dS \tag{1.6}$$

It can be shown that the contribution from the surface $\mathcal{C}$ tends to zero as $R \to +\infty$ [2] (see also Section 1.1.3). Now, for the values of U and $\partial U/\partial \mathbf{n}$ on $\mathcal{A}$ and $\mathcal{B}$, which are not known exactly, Kirchhoff supposes the following boundary conditions:

1. on $\mathcal{A}$: $U = U^{(i)}$, $\frac{\partial U}{\partial \mathbf{n}} = \frac{\partial U^{(i)}}{\partial \mathbf{n}}$;
2. on $\mathcal{B}$: $U = 0$, $\frac{\partial U}{\partial \mathbf{n}} = 0$.

where $U^{(i)} = e^{ikr}/r$ is the incident wave. The hypothesis on $\mathcal{A}$ means that U and $\partial U/\partial \mathbf{n}$ do not differ substantially from the values obtained in the absence of the screen. This assumption clearly ignores the modification of the EM-field in the immediate neighborhood of the rim of the aperture. The error introduced by ignoring this rim-effect is negligible only if the aperture is large compared with a wavelength. The hypothesis on $\mathcal{B}$ supposes that no wave propagates on that part. Eventually, under Kirchhoff's boundary conditions, $U(P)$ can be expressed as a function of the disturbance on $\mathcal{A}$

$$U(P) = -\frac{i}{2\lambda}\int_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs}[\cos(\mathbf{n},\mathbf{r}) - \cos(\mathbf{n},\mathbf{s})]\, dS \tag{1.7}$$

which is known as the Fresnel-Kirchhoff formula. Kirchhoff's theory have been found experimentally to be remarkably accurate [2]. However, this theory contains some internal inconsistencies. First, the boundary condition on $\mathcal{B}$ imposes zero on both the field and its normal derivative. This will imply that the field vanishes in the entire space behind the aperture [2], which is not true. Second, Kirchhoff's diffraction integral does not recover the assumed boundary values as P approaches the aperture [1].

The inconsistencies of Kirchhoff's theory have been removed by Rayleigh-Sommerfeld diffraction theory [1][4][2], which eliminates the necessity of imposing boundary values on both the disturbance and its normal derivative simultaneously.

Indeed, thanks to the planar diffraction interface ($\mathcal{A} \cup \mathcal{B}$), we can make a particular choice of the auxiliary function U' in Lemma 1, i.e.,

$$U'_I(\mathbf{x}) = \frac{e^{iks}}{s} - \frac{e^{iks'}}{s'} \tag{1.8}$$

where $s' := \|P'\mathbf{x}\|$, and P' is the mirror point of P w.r.t. the plane $\mathcal{A} \cup \mathcal{B}$. Note that P' is outside the surface $\mathcal{S}$. With this choice, we see immediately that $U_I'(\mathcal{A} \cup \mathcal{B}) \equiv 0$. Then, one can verify that $(R \to +\infty)$ [1][2]

$$\begin{aligned} U(P) &= \frac{1}{4\pi}\int_{\mathcal{S}} U\frac{\partial U_I'}{\partial \mathbf{n}} - U_I'\frac{\partial U}{\partial \mathbf{n}}\,dS = \frac{1}{4\pi}\int_{\mathcal{A}\cup\mathcal{B}} U\frac{\partial U_I'}{\partial \mathbf{n}} - U_I'\frac{\partial U}{\partial \mathbf{n}}\,dS \\ &= \frac{1}{4\pi}\int_{\mathcal{A}\cup\mathcal{B}} U\frac{\partial U_I'}{\partial \mathbf{n}}\,dS = \frac{1}{2\pi}\int_{\mathcal{A}\cup\mathcal{B}} U\frac{e^{iks}}{s}\left(ik - \frac{1}{s}\right)\cos(\mathbf{n},\mathbf{s})\,dS \qquad (1.9) \end{aligned}$$

Now, as (1.9) suggests, we on longer need to impose boundary values on the normal derivative of U. Thus, we suppose that

1. on $\mathcal{A}$, $U = U^{(i)}$;
2. on $\mathcal{B}$, $U = 0$.

Then, we obtain from (1.9) that

$$\begin{aligned} (1.9) &= \frac{1}{2\pi}\int_{\mathcal{A}} U^{(i)}\frac{e^{iks}}{s}\left(ik - \frac{1}{s}\right)\cos(\mathbf{n},\mathbf{s})\,dS \approx \frac{i}{\lambda}\int_{\mathcal{A}} U^{(i)}\frac{e^{iks}}{s}\cos(\mathbf{n},\mathbf{s})\,dS \qquad (1.10) \\ &= \frac{i}{\lambda}\int_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs}\cos(\mathbf{n},\mathbf{s})\,dS \qquad (1.11) \end{aligned}$$

Here, we assumed $s \gg \lambda$ in the approximation. (1.11) is known as the first-kind Rayleigh-Sommerfeld diffraction formula.

Another possible choice of U' is given by

$$U_{II}'(\mathbf{x}) = \frac{e^{iks}}{s} + \frac{e^{iks'}}{s'} \qquad (1.12)$$

In this case, $\frac{\partial U_{II}'}{\partial \mathbf{n}}|_{\mathcal{A}\cup\mathcal{B}} \equiv 0$. This allows us to impose boundary values for the normal derivative only, i.e.,

1. on $\mathcal{A}$, $\frac{\partial U}{\partial \mathbf{n}} = \frac{\partial U^{(i)}}{\partial \mathbf{n}}$;
2. on $\mathcal{B}$, $\frac{\partial U}{\partial \mathbf{n}} = 0$.

The resulting integral is the second-kind Rayleigh-Sommerfeld diffraction formula $(r \gg \lambda)$

$$U(P) = -\frac{1}{2\pi}\int_{\mathcal{A}} \frac{e^{iks}}{s}\frac{\partial U^{(i)}}{\partial \mathbf{n}}\,dS \approx -\frac{i}{\lambda}\int_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs}\cos(\mathbf{n},\mathbf{r})\,dS \qquad (1.13)$$

For small incidence angles $\cos(\mathbf{n},\mathbf{r}) \approx 1$ and small diffraction angles $\cos(\mathbf{n},\mathbf{s})$, we can see that (1.7) $\approx$ (1.11) $\approx$ (1.13), i.e., under these circumstances Fresnel-Kirchhoff's formula and

Rayleigh-Sommerfeld's formulas have no essential difference. As the field point approaches the aperture plane, it can be shown that the Rayleigh-Sommerfeld integrals reproduce the assumed boundary values [1]. However, these values are not true physical field in the aperture plane as the diffraction at the rim of the aperture is ignored. Indeed, although the Rayleigh-Sommerfeld theory solves the inconsistencies of the Fresnel-Kirchhoff diffraction, it does not necessarily yield more accurate results in practice [4]. There exists even evidence that the Kirchhoff's theory gives results that are in closer agreement with observation [1].

Now, let us turn to the diffraction field of a lens system. Consider the model shown in Fig. 1.3. We denote respectively by U_1 and U_2 the light fields on the planes $\mathcal{U}_1$ and $\mathcal{U}_2$ which are immediately before and behind the lens. The fields $U_1(x, y)$ and $U_2(x, y)$ are related by the transmittance of the lens $t(x, y)$, i.e., $U_2(x, y) = t(x, y)U_1(x, y)$. The transmittance can be expressed by

$$t(x, y) = P(x, y)e^{i\phi(x,y)}$$

where $P(x, y)$ is the pupil function of the lens. $P(x, y)$ and $\phi(x, y)$ respectively represent the amplitude and phase changes for the incident light. For a typical lens having a uniform circular aperture of a radius a, $P(x, y)$ is nothing but the indicating function of the circular transparent area, i.e.,

$$P(x, y) = \begin{cases} 1 & \sqrt{x^2 + y^2} \leq a \\ 0 & \text{otherwise} \end{cases} \tag{1.14}$$

It remains to determine the phase term ϕ. For a lens with spherical surfaces working under the paraxial condition, the function t can be explicitly specified [2][4]. Thus, given an incident field U_1, the output field U_2 can be easily calculated. Then, (1.10) is used to derive the diffraction field at the point P where $U^{(i)}$ is set to U_2 [2][4]. However, the transmittance function of a practical microscope objective is generally unknown because 1) microscope objectives often work under the nonparaxial condition; 2) their surfaces are usually made aspherical to correct optical aberrations; 3) the geometric information of the optical components in an objective is unavailable. Consequently, the diffraction field has to be derived differently. Section 1.1.3 provides a general solution to the diffraction field generated by a uniformly illuminated objective working under the nonparaxial condition.

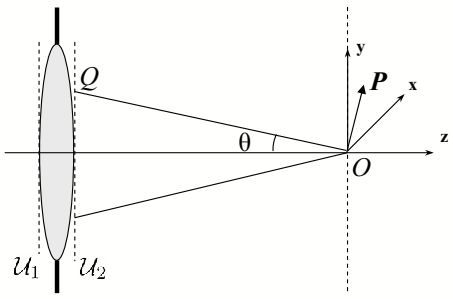


Fig. 1.3: Paraxial diffraction model of a lens.

1.1.3 Diffraction model for a nonparaxial lens

Consider a uniformly illuminated lens shown in Fig. 1.4(a). To derive its diffraction field, we make use of the fact that a well corrected objective should give out a spherical converging wave-front ($\mathcal{A}$ in Fig. 1.4(a)). We further suppose a light density $A(\mathbf{x})$ on the wave-front $\mathcal{A}$. $A(\mathbf{x})$ is usually called apodization function and its explicit form is typically known (see Section 1.1.3.1). As a result, the field on $\mathcal{A}$ is expressed by

$$U_{\mathcal{A}}(\mathbf{x}) = A(\mathbf{x})\frac{e^{-ikf}}{f} \tag{1.15}$$

where f is the focal length. Now, our problem becomes that given $U_{\mathcal{A}}$ we want to determine the field at P. Note that under this formulation, we do not need to know the transmittance of the objective. However, as the wave-front $\mathcal{A}$ is no longer planar, the Rayleigh-Sommerfeld diffraction theory cannot be applied. The Kirchhoff's formulation has no restriction on the shape of the region $\mathcal{A}$, but the classical model of the diffraction by a planar screen should be modified.

Here, we construct a more appropriate model for the light focusing of a (nonparaxial) lens, which is illustrated in Fig. 1.4(b). The surface $\mathcal{A}$ represents the spherical wave-front on the aperture, focusing at the point O. The expression of the field at P is given by Theorem 2, and its proof is provided in Appendix A.3. The derivation is completely similar to that of the Fresnel-Kirchhoff formula.

Theorem 2 *Consider the model in Fig. 1.4(b). We note* $\mathbf{n}$ *as the inward normal on the surface* $\mathcal{S} = \mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$, $f := \|OQ\|$, *and* $s := \|\mathbf{s}\| = \|PQ\|$ *where* Q *is any point on* $\mathcal{A}$. *We suppose that*

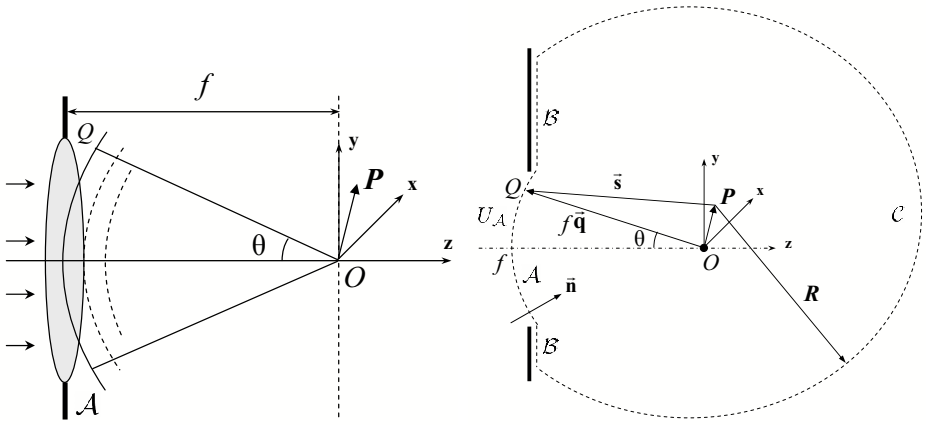


Fig. 1.4: Fresnel-Kirchhoff diffraction of a nonparaxial lens.

1. *on* $\mathcal{A}$, $\frac{\partial A}{\partial \mathbf{n}} = 0$;
2. *on* $\mathcal{B}$, $U_{\mathcal{B}} \approx 0$ *and* $\frac{\partial U_{\mathcal{B}}}{\partial \mathbf{n}} \approx 0$;
3. *on* $\mathcal{C}$, $\lim_{R\to+\infty} R\left(ikU + \frac{\partial U}{\partial \mathbf{n}}\right) = 0$;
4. $f, s \gg \lambda$.

$U_{\mathcal{A}}$ being given by (1.15), we have that

$$U(P) = \frac{i}{2\lambda}\int_{\mathcal{A}} A\frac{e^{ik(s-f)}}{sf}(\cos(\mathbf{n},\mathbf{s}) - 1)\, dS \tag{1.16}$$

The hypothesis on $\mathcal{A}$ is reasonable, as it supposes that in the immediate neighborhood of $\mathcal{A}$, the ray density is maintained in the direction of the wave propagation. The hypothesis on $\mathcal{B}$ is the same as the Kirchhoff's condition. The hypothesis on $\mathcal{C}$ is known as the Sommerfeld radiation condition. It is satisfied if the disturbance U vanishes as fast as a diverging spherical wave while R increases, so that the contribution from $\mathcal{C}$ to $U(P)$ approaches zero (see the proof in Appendix A.3). It guarantees that only outgoing waves will present on $\mathcal{C}$ rather than incoming waves. Since only outgoing waves will fall on $\mathcal{C}$ in our problem, this condition is satisfied here (see also [2]).

1.1.3.1 Apodization function

Consider a lens with a uniform circular aperture. Typically, the apodization function $A(\mathbf{x})$ is radially symmetric w.r.t. the optical axis and can be represented by $A(\theta)$. For commercial

microscopes, $A(\theta)$ is usually designed to be

$$A(\theta) = \sqrt{\cos\theta} \tag{1.17}$$

(1.17) is known as the sine condition [1][4]. This condition guarantees that a small region of the object plane in the neighborhood of the optical axis is imaged sharply by a family of rays which can have any angular divergence [4]. The optical system satisfying the sine condition is called aplanatic which exhibits 2D transverse shift invariance.

Another less used apodization function is

$$A(\theta) = 1 \tag{1.18}$$

That is, the ray density is constant over the range of the convergence angle. This is known as the Herschel condition [1][4]. When this condition is fulfilled, an element of the axis will be imaged sharply by a family of rays having any angular divergence, i.e., the axial shift invariance is satisfied [4]. Note that if a lens has a small NA, i.e., working under the paraxial situation, the difference between the two apodization functions above is negligible.

Real microscopes usually have high-NA lens (nonparaxial lens) designed to satisfy the sine condition. By the above discussion, they will not satisfy the Herschel condition. Fortunately, in biological applications, particularly when dealing with thin specimens, the axial working range of a microscope objective is very small, say, several microns. For an object undergone an axial shift within such small distances, its image will not be substantially changed except for a corresponding shift [5]. Thus, the microscope lens can still be considered to have both the transverse and the axial shift invariance.

1.2 Scalar Debye integrals

We still consider the model in Fig. 1.4(b). We note $\mathbf{p} := OP$, and $\mathbf{q} := \frac{1}{f}OQ$ which is the unit vector in the direction of OQ. Debye's integral expresses the diffraction field at near-focus points, i.e., at the points P such that $\|\mathbf{p}\|$ is small (compared with f). It can be obtained by applying the following approximations on our nonparaxial diffraction integral (1.16).

1. $s - f \approx -\mathbf{q}\cdot\mathbf{p}$. This means that the spherical wavelets originating from the wave-front $\mathcal{A}$ are approximated by plane wavelets;

2. $\cos(\mathbf{n}, \mathbf{s}) \approx -1$;

3. the distance s in the dominator in (1.16) is approximated by f.

Consequently, the Debye integral is given by

$$U(P) = -\frac{i}{\lambda}\int_{\mathcal{A}} A\frac{e^{-ik\mathbf{q}\cdot\mathbf{p}}}{f^2}\,dS = -\frac{i}{\lambda}\int_{\Omega} Ae^{-ik\mathbf{q}\cdot\mathbf{p}}\,d\Omega \tag{1.19}$$

where $d\Omega$ is an element of the solid angle subtended by $\mathcal{A}$ at O. It was known that the Debye's integral is well adapted to high NA situations [4].

To make (1.19) more explicit, let us introduce the spherical coordinates to represent Q, and the cylindric coordinates to represent P.

$$\begin{aligned} Q &= (x_a, y_a, z_a) = (f\sin\theta\cos\phi, f\sin\theta\sin\phi, -f\cos\theta), \quad 0 \le \theta \le \alpha,\ 0 \le \phi < 2\pi \\ P &= (x, y, z) = (\rho\cos\psi, \rho\sin\psi, z), \quad 0 \le \psi < 2\pi \end{aligned}$$

where α is the maximum angle of the convergence of rays. We have also $d\Omega = \sin\theta d\theta d\phi$. Suppose that $A = A(\theta)$. Then, (1.19) becomes

$$\begin{aligned} (1.19) &= -\frac{i}{\lambda}\int_0^{2\pi}\int_0^{\alpha} A(\theta)\exp\left[-ik\rho\sin\theta\cos(\phi-\psi) + ikz\cos\theta\right]\sin\theta d\theta d\phi \\ &= -ik\int_0^{\alpha} A(\theta)J_0(k\rho\sin\theta)e^{ikz\cos\theta}\sin\theta\,d\theta \qquad (1.20) \\ &= -ike^{ikz}\int_0^{\alpha} A(\theta)J_0(k\rho\sin\theta)e^{-2ikz\sin^2\frac{\theta}{2}}\sin\theta\,d\theta \qquad (1.21) \end{aligned}$$

where J_0 is the zero order Bessel function of the first kind. It can be seen that the Debye theory predicts the intensity distribution $|U(P)|^2$ to be axially symmetric.

Now, the Debye integral for a lens satisfying the sine condition is given by

$$\begin{aligned} U(P) &= -ik\int_0^{\alpha}\sqrt{\cos\theta}J_0(k\rho\sin\theta)e^{ikz\cos\theta}\sin\theta\,d\theta \qquad (1.22) \\ &= -ike^{ikz}\int_0^{\alpha}\sqrt{\cos\theta}J_0(k\rho\sin\theta)e^{-2ikz\sin^2\frac{\theta}{2}}\sin\theta\,d\theta \qquad (1.23) \end{aligned}$$

If the NA of the lens becomes small ($\alpha \to 0$), i.e., the system becomes paraxial, we will have that $\sin\theta \approx \theta$ and $\sqrt{\cos\theta} \approx 1$. Hence, $U(P)$ tends asymptotically to

$$U(P) = -ike^{ikz}\sin^2\alpha\int_0^1 J_0(k\rho t\sin\alpha)e^{-\frac{i}{2}kzt^2\sin^2\alpha}t\,dt \tag{1.24}$$

1.3 Aberrations and Vectorial Diffractions

1.3.1 Aberrations

Practical optical systems are rarely aberration-free. Aberration means that the wave-front after a lens is not a spherical surface. In the presence of aberrations, the system would lose the shift-invariance property.

General aberrations can be formulated as the phase differences from the real wave-front to a reference spherical wave-front [4]. Consider the model in Fig. 1.5. $\mathcal{S}$ is the reference spherical wave-front that is supposed to converge to the point O. The real wave-front is given by $\mathcal{A}$. Given a point P where the diffraction field is to be determined and any point Q on $\mathcal{S}$, the extension of the line PQ intersects $\mathcal{A}$ at point Q^*. The path difference $\Phi := Q^*Q$ is a function of the point Q^* and is defined as the aberration function. Thereby, $k\Phi$ stands for the phase change w.r.t. $\mathcal{S}$ caused by aberration. Now, we can rewrite (1.16) by introducing this phase term

$$U(P) = \frac{i}{2\lambda} \int_{\mathcal{S}} A e^{-ik\Phi} \frac{e^{ik(s-d)}}{sd} (\cos(\mathbf{n}, \mathbf{s}) - 1)\, dS \tag{1.25}$$

Note that the integration is performed on the reference surface $\mathcal{S}$. Then, Debye's approximations (see Section 1.2) can be used to simplify (1.25).

Traditionally, Zernike's circular polynomials $(Z_i(t, \phi))_i$ are used to expand Φ for a circular lens working under the paraxial condition [4][6]. This is appropriate because these polynomials $(Z_i(t, \phi))_i$ form a complete orthogonal set on the unit disk [1]. Here, t $(0 \leq t \leq 1)$ is the normalized radial coordinate and ϕ the polar angle on the circular aperture plane. Thus, we can write $\Phi(t, \phi) = \sum_i a_i Z_i(t, \phi)$ where the expansion coefficients $(a_i)_i$ should be experimentally measured [6].

It has been pointed out that in biological microscopes the most significant aberration may be that due to the mismatches of refractive indices [7][8]. This occurs if, for example, the investigated specimen is mounted in a medium with a refractive index different from that of the immersion medium, or the index of the immersion medium differs from that of the coverslip. In these cases, analytical expressions of Φ have been studied [7][9] as a function of the microscope setup parameters such as the indices and the thickness of the media. If the reference wave-front $\mathcal{S}$ is focused at an on-axis point, Φ under this aberration is radially symmetric and is termed spherical aberration [4]. This aberration is more pronounced as the specimen becomes thicker [4].

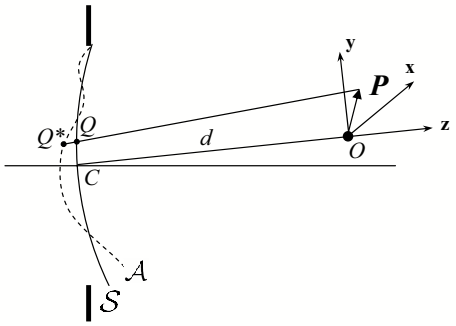


Fig. 1.5: General aberration model.

1.3.2 Vectorial diffractions

The scalar Debye diffraction integral (1.19) can be generalized into a vectorial form by replacing A with the vectorial electric field on the spherical wave-front [10][11][12][4]. The resulting vectorial Debye integral provides a more accurate description of the diffraction field than the scalar theory. In particular, it allows to explicitly take the light polarization into account. Always based on Debye's approximations, [13][14][15][16] have proposed vectorial computations of the diffraction field with spherical aberrations due to focusing a light into stratified media of mismatched indexes. For the same situation, [17] proposed to compute the field using the Huygens-Fresnel principle. This approach is shown to be equivalent to that using Debye's approximations as the Fresnel number[1] tends to infinite, which can be deemed as a good assumption for practical microscopes [18].

1.4 Conclusion

In this chapter, we first reviewed the classical scalar diffraction theory, which is useful to describe the diffraction field of a paraxial lens. Then, we constructed a diffraction model adapted to the nonparaxial imaging condition. By using Kirchhoff's theory, we derived a diffraction integral for nonparaxial lens. Based on this formulation, Debye's diffraction integrals were re-derived, which adapt well high NA conditions and are the basis of our following work. Finally, aberration theory and vectorial diffractions were briefly reviewed.

We point out that although vectorial theories give a more accurate description of the

[1]Fresnel number is defined to be $N = \mathrm{NA}^2 f/(n\lambda)$, where f is the focal length and λ is the vacuum wavelength. In practical microscope, f is much larger than λ.

lens diffraction, in our work only the scalar formulation will be used, and the microscope imaging system will be assumed diffraction-limited with negligible aberrations. This is because the scalar theory introduces substantial simplifications compared with a full vectorial theory, while the assumption of negligible aberrations is reasonable for imaging thin objects. The simplification provided by a diffraction-limited scalar model facilitates considerably our subsequent analysis and modeling.

Finally, let us rewrite the derived Debye's diffraction formulas which will serve extensively in the following chapters. The nonparaxial diffraction integral is given by

$$\mathbf{h}(P;\lambda) = -ik \int_0^{\alpha} \sqrt{\cos\theta} J_0(k\rho \sin\theta) e^{ikz\cos\theta} \sin\theta \, d\theta \tag{1.26}$$

where λ is the vacuum wavelength, $k = n\frac{2\pi}{\lambda}$, and n the refractive index. The paraxial integral is given by

$$\mathbf{h}(P;\lambda) = -ike^{ikz} \sin^2\alpha \int_0^{1} J_0(k\rho t \sin\alpha) e^{-\frac{i}{2}kzt^2\sin^2\alpha} t \, dt \tag{1.27}$$

Chapter 2

Fluorescence Microscope PSFs

Fluorescence microscope has become a widely used and essential tool in biology research to imaging cellular or sub-cellular targets with a high degree of specificity. Its imaging function relies on the principle of fluorescence emission upon light excitation. Typically, the specimen is first stained by certain fluorophores (dyes) which can target themselves to label the components of interest (genes, proteins, membranes, vesicles, etc.). The specimen is then illuminated with light at a specific wavelength which is absorbed by the fluorophores, causing them to emit light of a longer wavelength, i.e., light of a different color than the absorbed light. Thus, the much weaker emitted fluorescence can be separated from the illumination light through the use of an emission filter, and is finally detected by a charge-coupled device (CCD) or a photomultiplier (PMT) detector. In this way, only the labeled components of interest are imaginged but not the entire cellular structures. To date, more and more fluorescence dyes with different excitation and emission characteristics have been made available allowing simultaneous labeling of different cellular components. Furthermore, dynamic biological processes in live cells can be observed using less phototoxic dyes e.g. green fluorescent protein (GFP). The richness of the fluorescent probes make fluorescence microscope a powerful tool to reveal diverse biological phenomena.

The goal of this chapter is to characterize three types of fluorescence microscopes, i.e., WFFM, LSCM and DSCM, by their diffraction-limited PSFs. These PSFs have been extensively studied in the literature (e.g., see [19][20][21][11][22][23][24][25][26]). Here, we give a comprehensive review and provide a formal derivation of the PSFs.

The structure of the chapter is as follows. First, an overview of the optical models of these microscopes is introduced (Section 2.1). Then, we study the photodynamics of

fluorescence dyes (Section 2.2). Finally, using these results and the Debye's diffraction integrals (Chapter 1), we show the microscope PSFs under both the paraxial and nonparaxial conditions (Section 2.3).

2.1 Optical Models of WFFM, LSCM and DSCM

2.1.1 WFFM optical model

Fig. 2.1 provides an illustrative optical path model in a WFFM. Commercial microscope usually uses an arc lamp as light source to provide sufficiently intense illumination. The illumination light is filtered by an excitation filter, which selects the desired band of excitation wavelengths (λ_{ex}) at which the fluorophores attain a high or even a maximum absorption rate.

The specimen is evenly illuminated and the dyes emit fluorescence of an average wavelength λ_{em}. The emission wavelength is always longer than the excitation one, a phenomenon called Stokes shift [27]. This is because the fluorophores lose a portion of the absorbed energy before re-releasing the rest of the energy as fluorescence. Thanks to the Stokes shift, the emission filter can be designed to eliminate from the fluorescence any residual excitation light. The fluorescence is finally detected by a CCD camera situated at the rear focal plane of the emission beam. To scan the whole volume of the specimen, the objective is moved along the z-direction, and the planes at different depths of the specimen will be in focus and will be imaged on the camera. It can be noted that for the imaging at any in-focus z-plane, the fluorescence from the out-focus planes will also be superposed on the camera, as the whole volume is illuminated all the time.

2.1.2 LSCM optical model

In LSCM (see Fig. 2.2), each sample point of the specimen is excited by a highly focused light from a laser source. The emitted fluorescence is detected by a PMT, a single-element photon detector, immediately behind a pinhole located at the rear focal point of the emission beam. By focusing the laser sequentially at every sample point, the 3D volume can be scanned. Unlike the WFFM case, the presence of a pinhole in a LSCM prevents the detection of out-focus fluorescence. As a result, the confocal system has a better resolution than the conventional widefield system. But the fact that the LSCM has to do a pointwise sampling

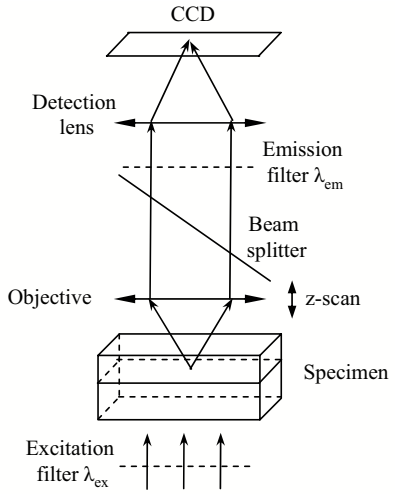


Fig. 2.1: WFFM optical path

makes the imaging duration much longer than in WFFM.

In the excitation phase shown in Fig. 2.2, a laser source is used, which provides both a sufficient excitation photon flux and a highly specific wavelength. The laser is first expanded by a beam expander. The expansion of the light is necessary to provide a uniform illumination on the rear part of the objective [28]. Indeed, the beam expander performs a spatial filtering, and the resulting intensity profile of the excitation light mainly depends on the diffraction of the excitation pinhole, and is little influenced by the profile of the original laser [28].

In the emission phase, the detection lens plays a role of light collimator which converges the emission light to the detection pinhole. This pinhole must be small enough to reject the emission light from out-focus planes in order to enhance the resolution. But a too small pinhole will reduce the number of photons arrived at the PMT so eventually reduce the signal-to-noise ratio (SNR). Thus, a trade-off exists between resolution and SNR.

2.1.3 DSCM optical model

The Nipkow disk (see Fig. 2.3) plays a central role in the design of modern DSCMs. A series of pinholes are arranged in an Archimedean spiral of constant pitch on the disk. The polar-system representation of such spirals is given by $\rho(\theta) = c \cdot \theta$ where ρ and θ are radial and angular coordinates respectively, and $c > 0$. While the Nipkow disk is rotated,

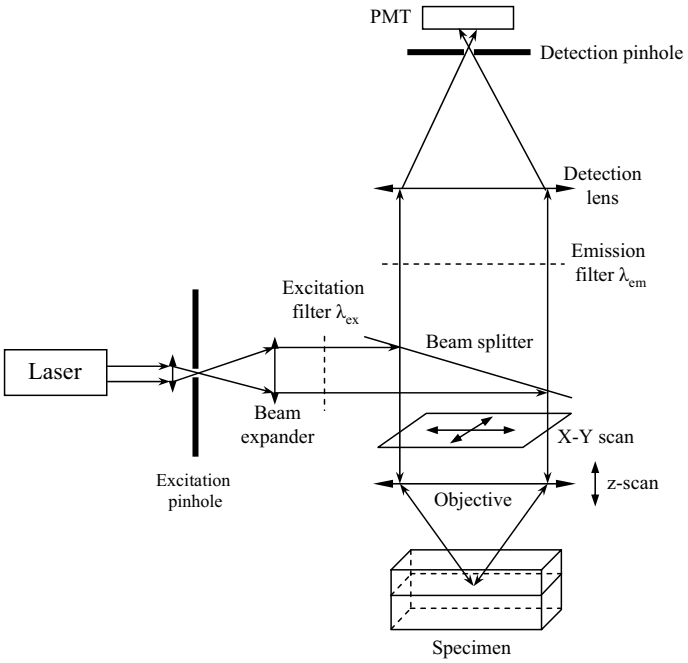


Fig. 2.2: LSCM optical path

these pinholes form a series of concentric arcs in the viewport W. If an image is present at the back side of the disk, the image will undergoes a line-by-line scanning. Since the scanlines are in effect not straight lines but rather curves, the ideal Nipkow disk should have either a very large diameter and the viewport situating at the outer sectors (to have smaller curvatures), or a very narrow angular opening of its viewport. The initial version of the disk allowed only a small number of pinholes, and at any instant, only a single pinhole is present in the viewport, which makes it inefficient for detecting image information.

Petrán redesigned the Nipkow disk, who made the implementation of a real DSCM possible. Petrán's disk [29], or tandem disk (Fig. 2.4), contains a series of interlaced Archimedean spirals of constant pitch. Every spiral is started in a sector that repeats an even number of times around the disk. One of the spirals is highlighted in green in Fig. 2.4. Note that the pattern of the holes repeats from sector to sector. In practice, tandem disk can have a large number of interlaced spirals, such that the disk contains a high density of holes (Fig. 2.5(a)). A regional zooming view of the disk is shown in (Fig. 2.5(b)). It can be seen that the holes form a quasi-hexagonal array. At any instant, thousands of pinholes are

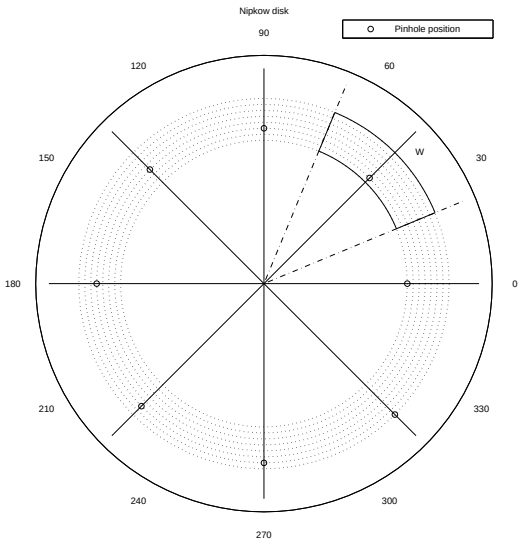


Fig. 2.3: Nipkow disk (W is the viewport)

illuminated in parallel, and their image will be projected by the objective onto the focus plane in the specimen creating the same number of excitation points. The emission light will be rerouted to the pinholes diametrically opposite to the illuminated ones. This design, despite its extreme mechanical complexity, avoids the influence of the reflected illumination light from the disk to the emission light, so that the image can be directly viewed by eyes without any reflection glare. When the disk begins to rotate, the same number of scanlines ("scancurves" in effect) as for the pinholes will be created. This kind of scanning greatly reduces the acquisition period on a plane compared with the single-pinhole scanning mechanism used in LSCM. However, this system (and also its variants [30, 31]) suffers from a low illumination efficiency ($\sim 4\%$) [32] due to the limited transparent area (pinholes) on the disk.

To enhance the illumination efficiency, modern DSCM employs the Yokogawa's microlens design where a second microlens disk is set up above the tandem disk (Fig. 2.6). The microlens disk contains about 2×10^4 microlens, each one focusing the illuminating laser beam onto its corresponding pinhole below. The light is then diffused by each pinhole to cover the rear stop of the objective. A beam splitter between the two disks directs the fluorescence light emerging from the pinholes to the CCD camera via the detection lens. Note that although the transparent area on the pinhole disk is very small, that on the

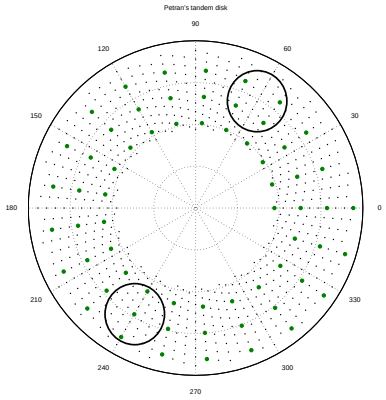


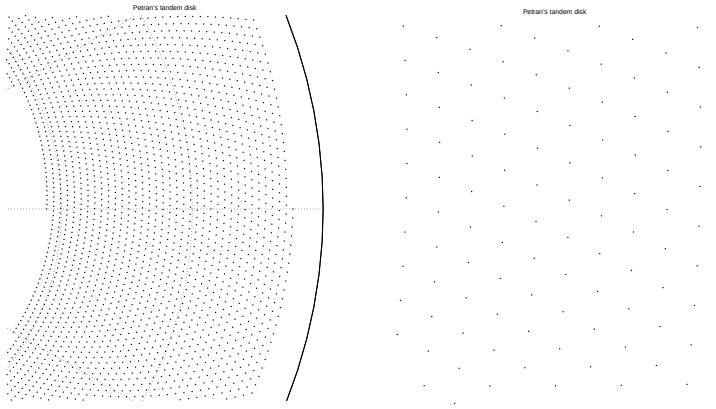
Fig. 2.4: Tandem disk. One of the circular area will be illuminated, and the emission light will be rerouted to the opposite circular area.

microlens disk can be quite large. By microlens focusing, very little light is wasted between the two disks. As a result, the illumination efficiency has been observed to be improved by an order of magnitude ($\sim 40\%$) [32] compared with the tandem disk.

2.2 Photodynamics of Fluorophores

The goal of this section is to show that the emission intensity of fluorescence is approximately linearly proportional to the excitation intensity, as long as the excitation intensity is moderate.

Fig. 2.7 sketches a simplified model of the energy level system for fluorescence molecules. The molecules are initially at the ground state S_0. Upon absorption of a photon of an energy matching the difference between the energy level S_0 and S_1 (excitation state), the molecule will be excited to the state S_1. Let us consider the upward transition ($S_0 \rightarrow S_1$). Suppose I_{ex} (W/m^2) to be the excitation intensity, σ to be the area of the absorption cross section of a single fluorescence molecule, and $h\nu_{ex}$ to be the energy of an absorbed photon where h is Planck's constant and ν_{ex} is the photon's frequence. Assume further that we have initially N_0 molecules at the ground state S_0. For a short period of time Δt, the molecules are irradiated by a total energy of $E_{ex} = I_{ex}N_0\sigma\Delta t$. Thereby, the average number of the excited molecules is $N_{ex} = E_{ex}/(h\nu_{ex})$. Consequently, the upward



(a) (b)

Fig. 2.5: Quasi-hexagonal pinhole array on a Tandem disk. (a) Tandem disk (b) local zoom

transition rate for the ground-state molecules is given by

$$\alpha = \frac{N_{ex}}{N_0 \Delta t} = \frac{I_{ex}\sigma}{h\nu_{ex}} \quad \text{percent/sec} \tag{2.1}$$

Next, let us consider the downward transition ($S_1 \rightarrow S_0$). There are several ways for a molecule at S_1 to give up energy to return to S_0, one of which is by the fluorescence emission. Suppose that we have initially N_1 molecules at S_1. As the radioactive decay is an exponential function of time (t), we have that

$$N_1(t) = N_1 e^{-\beta t}$$

where β is the decay constant for the considered dye. The downward transition rate can now be calculated for a short Δt, which actually equals to β [Eq.(2.2)].

$$\frac{N_1 - N_1(\Delta t)}{N_1 \Delta t} = \frac{1 - e^{-\beta \Delta t}}{\Delta t} \overset{\Delta t \rightarrow 0+}{\longrightarrow} \beta \quad \text{percent/sec} \tag{2.2}$$

Note that this rate includes not only the radiative rate such as the fluorescence decay rate, but also all other energy-releasing rates such as that of the nonradiative conversions [27]. Therefore, the fluorescence decay rate is always smaller than β, and is written as $Q\beta$, where $Q\,(< 1)$ is the quantum efficiency for fluorescence.

Now, it is straightforward to write out the following population equations depicting

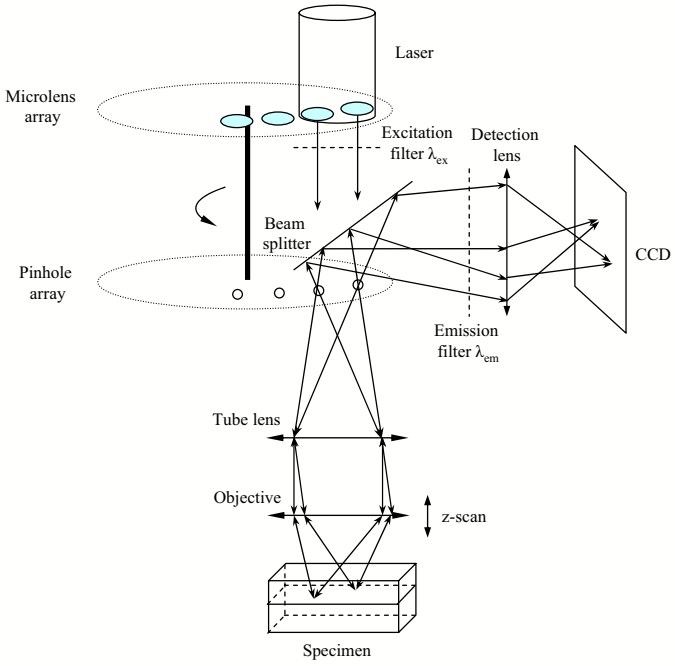


Fig. 2.6: DSCM optical path

the dynamics between S_0 and S_1.

$$\begin{cases} N_0'(t) = \beta N_1(t) - \alpha N_0(t) \\ N_1'(t) = \alpha N_0(t) - \beta N_1(t) \end{cases}$$
$$N_0(0) = N_0, N_1(0) = N_1 \tag{2.3}$$

where $N_0(t)$ and $N_1(t)$ are respectively the populations of the state S_0 and S_1 at time t. The solution of (2.3) is given by

$$N_1(t) = \frac{\alpha}{\alpha+\beta}(N_0+N_1) + \frac{\beta N_1 - \alpha N_0}{\alpha+\beta} e^{-(\alpha+\beta)t}$$

Thereby, the steady state solution (large t) N_1^s is given by

$$N_1^s = \frac{\alpha}{\alpha+\beta}(N_0+N_1) \tag{2.4}$$

When $\alpha = \beta$, half of the molecules will be in the excited state. We define the corresponding excitation intensity in this case as $I_{sat} := \beta h\nu_{ex}/\sigma$. I_{sat} characterizes the saturation of the system. We also write out the global fluorescence emission rate (w.r.t. the total number of

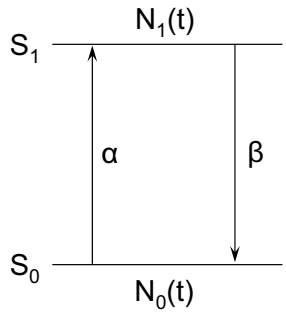


Fig. 2.7: Simplified model of the energy level system for fluorescence molecules

molecules $N_0 + N_1$)

$$r_f = \frac{Q\beta N_1^s}{N_0 + N_1} = Q\frac{\alpha\beta}{\alpha + \beta} = \frac{Q\beta\sigma I_{ex}}{\sigma I_{ex} + \beta h\nu_{ex}} \quad \text{percent/sec} \tag{2.5}$$

The emission intensity of the fluorescence is thus

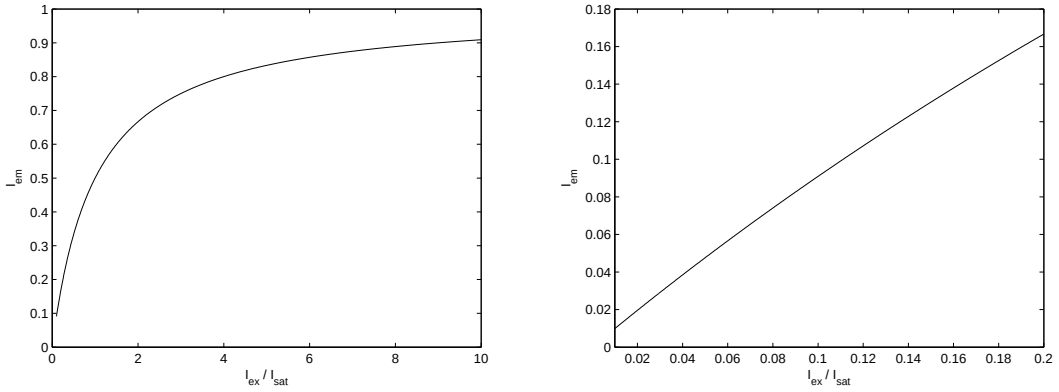
$$I_{em} = \frac{r_f(N_0 + N_1)h\nu_{em}}{S} = \frac{Q\beta(N_0 + N_1)h\nu_{em}}{S} \cdot \frac{I_{ex}}{I_{ex} + I_{sat}} \tag{2.6}$$

where ν_{em} is the frequence of emission photons, and S is the observation-surface area. (2.6) shows that the emission intensity is proportional to the underlying fluorescein density $(N_0 + N_1)$, and is a nonlinear function of the excitation intensity. But if $I_{ex} \ll I_{sat}$, I_{em} becomes approximately linear w.r.t. I_{ex}.

Fig. 2.8 illustrates the behavior of I_{em} as a function of the ratio I_{ex}/I_{sat} for the dye Rh B with an excitation vacuum wavelength $\lambda_{ex} = 500nm$. For this dye, we have that [27] $\sigma = 1.3 \times 10^{-20}m^2$ and $\beta = 3.33 \times 10^8$. The nonlinear behavior of I_{em} can be clearly seen in Fig. 2.8(a). As I_{ex} raises higher and higher than I_{sat}, the emission intensity becomes eventually saturated. Fig. 2.8(b) shows that if no saturation occurs, i.e., $I_{ex} \ll I_{sat}$, we find an almost linear emission response.

2.3 Diffraction-limited PSFs of WFFM, LSCM and DSCM

For a fluorescence microscope, the input is the spatial fluorophore distribution in the specimen, i.e., the density of the fluorophore molecules as a function of spatial locations. The output of the system is the fluorescence intensity image recorded by the camera. We suppose that the emitted fluorescence is an incoherent light. Then, by the fact that the emission



(a) (b)

Fig. 2.8: The emission intensity I_{em} as a function the ratio I_{ex}/I_{sat}. The multiplicative constant factor in (2.6) is omitted. Dye Rh B is assumed. $\lambda_{ex} = 500nm$, $\sigma = 1.3 \times 10^{-20} m^2$ and $\beta = 3.33 \times 10^8$. (a) the ratio $I_{ex}/I_{sat} \in [0.1, 10]$; (b) the ratio $I_{ex}/I_{sat} \in [0.01, 0.2]$.

intensity is proportional to the underlying fluorescein density (Section 2.2), the fluorescence microscope system can be seen to be linear. We further suppose that the system is 3D shift invariant. Then, we will characterize the microscopes by their PSFs such that the output can be obtained by convolving the input with the PSF.

In the following PSF derivation, we suppose that the excitation intensity is moderate such that the linear relationship between the emission and the excitation holds. We will also make use of the reciprocity principle [33]. It indicates that in a continuous linear medium with symmetric permittivity and magnetic permeability, the relationship between an oscillating point source and the resulting disturbance is unchanged if one interchanges the positions where the source is placed and where the disturbance is measured.

2.3.1 WFFM PSF

Fig. 2.9 shows the coordinate systems used for the WFFM model presented in Fig. 2.1. The object space is where the specimen locates. Without loss of generality, we create $O-\mathbf{xyz}$ as its 3D coordinate system where the z-axis coincides with the optical axis. The image space is where the camera situates. $O' - \mathbf{x}'\mathbf{y}'$ is a 2D coordinate system for the camera plane. We also introduce the system $O_0 - \mathbf{x}_0\mathbf{y}_0\mathbf{z}_0$ centered at the objective focus for expressing the diffraction field $\mathbf{h}(x_0, y_0, z_0; \lambda)$ using Debye's integral. We further denote the lateral

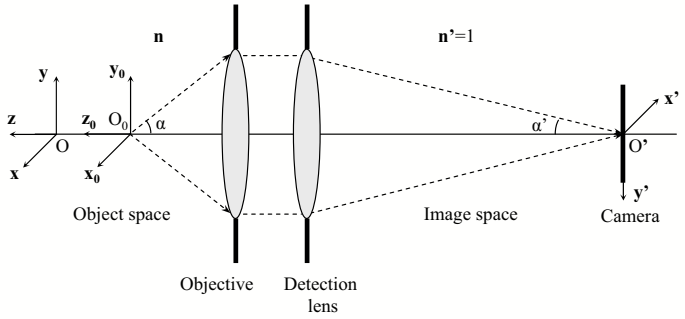


Fig. 2.9: Coordinate systems for WFFM PSF modeling.

magnification of the microscope by M_L. To avoid confusion, a subscript will be used for every coordinate to indicate the coordinate system that is used for representing that point. For example, $(a, b, c)_O$ is represented under the system $O - \mathbf{xyz}$, and $(a, b, c)_{O_0}$ is under the system $O_0 - \mathbf{x}_0\mathbf{y}_0\mathbf{z}_0$. Let the coordinate of O_0 to be $(0, 0, z)_O$.

We place a Dirac fluorescein distribution at O, and the PSF is given by the recorded 3D intensity image. As the whole specimen is evenly illuminated with a constant excitation intensity, say I_{ex}, the fluorescence emitted by the Dirac fluorophore should also have a constant intensity proportional to I_{ex} (see Section 2.2). As the imaging is shift-invariant, the intensity recorded in the image at $(x', y')_{O'} = (M_L x, M_L y)_{O'}$ for the current focus depth z is equivalent to that recorded at O' by shifting the Dirac to $(-x, -y, 0)_O$. Now let us apply the reciprocity theorem. We should calculate the intensity at $(-x, -y, 0)_O$ by placing the Dirac at O'. But this intensity is given by the Debye's integral, i.e.,

$$\mathrm{PSF}_{WFFM}(x, y, z)_O = C_0|\mathbf{h}(-x, -y, -z; \lambda_{em})_{O_0}|^2 = C_0|\mathbf{h}(x, y, z; \lambda_{em})_{O_0}|^2 \qquad (2.7)$$

where C_0 is a constant proportional to I_{ex}, $\mathbf{h}$ is given by (1.26) in the nonparaxial condition, or by (1.27) under the paraxial situation.

2.3.2 LSCM PSF

Fig. 2.10 shows the coordinate systems used for the LSCM model in Fig. 2.2. O_0 is the current scanning position, and has a coordinate $(x, y, z)_O$. O' is the pinhole location. We suppose that the pinhole is circular with a radius $r = D/2$ when projected into the object space.

The excitation of LSCM is due to focusing a beam of laser. For a Dirac fluorescein

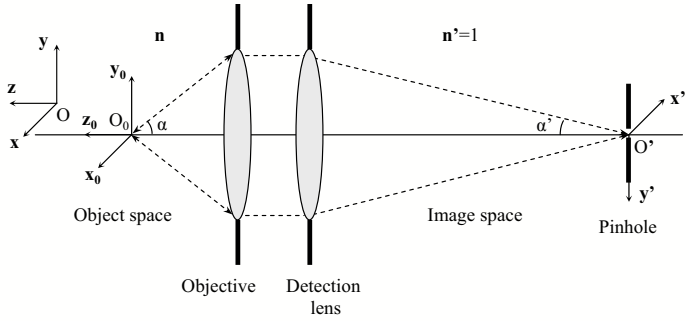


Fig. 2.10: Coordinate systems for LSCM PSF modeling.

distribution at O, the excitation intensity distribution in the object space, and particularly at O, can be derived using Debye's formula, i.e.,

$$I_{ex}(0,0,0)_O = |\mathbf{h}(-x,-y,-z;\lambda_{ex})_{O_0}|^2 = |\mathbf{h}(x,y,z;\lambda_{ex})_{O_0}|^2$$

Thereby, the Dirac fluorophore should emit fluorescence with an intensity proportional to $I_{ex}(0,0,0)_O$. It remains to calculate the integral intensity presented in the pinhole support. This can be proceeded in the same way as in the WFFM case and by applying the reciprocity theorem, we obtain

$$\mathrm{PSF}_{LSCM}(x,y,z)_O = C_1|\mathbf{h}(x,y,z;\lambda_{ex})_{O_0}|^2 \cdot \int_{\{x_1^2+y_1^2\le r^2\}} |\mathbf{h}(x-x_1,y-y_1,z;\lambda_{em})_{O_0}|^2\, dx_1 dy_1 \tag{2.8}$$

where C_1 is a constant. Note that the LSCM PSF is the product of excitation and emission intensity distributions, i.e., $|\mathbf{h}(x,y,z;\lambda_{ex})|^2$ and $|\mathbf{h}(x,y,z;\lambda_{em})|^2$. They are respectively termed to be the excitation PSF and the emission PSF.

2.3.3 DSCM PSF

We have shown that the pinholes on the disk of DSCM form a nearly periodic hexagonal pattern in the object space (see Section 2.1.3) with an adjacent pinhole distance d. For excitation from each pinhole, a local coordinate system with origin O_{n_x,n_y} at the excitation focus will be created, where (n_x, n_y) indexes the pinholes. Then, the DSCM PSF can be

derived in the same way as for LSCM PSF. The final expression is given by

$$\begin{aligned}
\mathrm{PSF}_{DSCM}(x,y,z)_O &= C_2\left|\sum_{(n_x,n_y)\in\mathcal{D}} \mathbf{h}\left(x-\frac{d}{2}(n_x+n_y), y-\frac{\sqrt{3}}{2}d(n_x-n_y), z;\lambda_{ex}\right)_{O_{n_x,n_y}}\right|^2 \cdot \\
&\quad \int_{\{x_1^2+y_1^2\le r^2\}} |\mathbf{h}(x-x_1, y-y_1, z;\lambda_{em})_{O_{0,0}}|^2\, dx_1 dy_1 \\
&= C_2\left|\sum_{(n_x,n_y)\in\mathcal{D}} \mathbf{h}\left(x-\frac{d}{2}(n_x+n_y), y-\frac{\sqrt{3}}{2}d(n_x-n_y), z;\lambda_{ex}\right)_{O_{0,0}}\right|^2 \cdot \\
&\quad \int_{\{x_1^2+y_1^2\le r^2\}} |\mathbf{h}(x-x_1, y-y_1, z;\lambda_{em})_{O_{0,0}}|^2\, dx_1 dy_1 \qquad (2.9)
\end{aligned}$$

where C_2 is a constant. Note that the total excitation distribution is the sum of the individual excitation distributions from the illuminating pinholes whose index set is denoted as $\mathcal{D} \subset \mathbb{Z}^2$. We point out that the light source of DSCM here is assumed to be a laser, since it is the most widely used source in commercial microscopes. Thus, as a coherent source is used, the first modulus sign in (2.9) is applied to the entire sum instead of to each summing term.

It can be verified from (2.7), (2.8) and (2.9) that all the PSFs have mirror symmetry about the xy-plane. Furthermore, PSF_{WFFM} and PSF_{LSCM} have circular symmetry about the z-axis, and PSF_{DSCM} is almost circularly symmetric if the distance d is sufficiently large.

2.4 Conclusion

We have shown in this chapter the diffraction-limited PSFs of WFFM, LSCM and DSCM based on Debye's integrals. The LSCM and DSCM PSFs can be written as a product of the excitation distribution and the emission distribution. As we have shown, these multiplicative forms emerge only if the excitation intensity is controlled well below the saturation level so that the fluorescence emission reacts linearly to the excitation.

Chapter 3

Gaussian Approximations of Diffraction-limited Microscope PSFs

The PSFs of WFFM, LSCM and DSCM discussed in Chapter 2 play a central role in understanding the imaging performances, such as the theoretical resolution limit and the optical sectioning capacity. Despite the availability of these physical models of PSFs, approximative PSFs and particularly separable Gaussian approximations are widely preferred in practical microscopic applications requiring fast data processing, such as single fluorescent particle tracking [34, 35, 36], fluorescent dot localization and tracking with super-resolution [37, 38], and myopic deconvolution [39]. Indeed, compared with a physical PSF model, which usually involves non-trivial terms such as integrals and infinite series, a Gaussian approximative PSF is much simpler and can be computed much faster. Furthermore, due to its special analytical form and nice properties (semi-group property, invariance under Fourier transform, etc.), a Gaussian PSF is often chosen to facilitate theoretical analysis and modeling, such as the derivation of analytical solutions to 3D Fluorescence Recovery After Photobleaching (FRAP) process [40, 41], the modeling of 3D tubular structures in confocal images [42] and the analysis of convergence properties of expectation-maximization (EM) deconvolution [43].

Despite the popularity of Gaussian approximations, most of the above mentioned works either assume the validity of the approximations or only justify them empirically on

observed data. The approximation accuracy and the selection of Gaussian parameters have rarely been rigorously investigated with any physical PSF model, leaving these approximations essentially arbitrary. To the best of our knowledge, only a few works [44, 38] considered the approximations based on physical models, but they solely covered the paraxial WFFM PSF case with an L^∞ constraint (see Section 3.1 for details on this constraint).

In this chapter, we propose a comprehensive study of the least squares Gaussian approximations of the diffraction-limited 2D/3D paraxial/non-paraxial PSFs of WFFM, LSCM and DSCM. The PSFs are expressed using the Debye integrals, as have been presented in Chapter 2. Then, in Section 3.1 we derive, under an L^∞ constraint imposing peak matching, optimal and near-optimal Gaussian parameters for the WFFM, LSCM and DSCM PSFs. Next, we consider in Section 3.2 the approximations with an L^1 constraint imposing energy conservation, where an optimal Gaussian parameter is derived for the 2D paraxial WFFM PSF. Numerical results in Section 3.3 show that: (i) the 2D approximations are all very accurate; (ii) no accurate Gaussian approximation exists for 3D WFFM PSFs; (iii) with typical pinhole sizes, the 3D approximations are accurate for DSCM and nearly perfect for LSCM. All the Gaussian parameters derived in this work are in explicit analytical forms listed in Tab. 3.1, 3.2 and 3.3, allowing their direct use in practical applications. Some conclusion remarks are drawn in Section 3.4.

3.1 Gaussian Approximations of the PSF Models with an L^∞ Constraint

To derive Gaussian approximations of the PSFs given in (2.7), (2.8) and (2.9), we suppose in the following that the Gaussian functions are centered at the origin of the PSFs and are separable. We make these two assumptions not only because they simplify the calculus and are widely adopted as stated before, but also because centered separable Gaussians are the only Gaussian functions that preserve the intrinsic symmetries in the PSF models, as shown in Proposition 1 (proof in Appendix A.4).

Proposition 1 *Assuming a 3D Gaussian distribution*

$$g_\Sigma(\mathbf{x}) := (2\pi)^{-\frac{3}{2}}|\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where we denote $\mathbf{x} = (x, y, z)^T$*, the mean vector* $\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)^T$ *and the covariance matrix*

$\Sigma = [\sigma_{ij}]_{1\leq i,j\leq 3}$. *$g_\Sigma$ has circular symmetry about the z-axis and mirror symmetry about the xy-plane if and only if* $\mu = \mathbf{0}$ *and* Σ *is diagonal with* $\sigma_{11} = \sigma_{22}$.

Therefore, the 2D and 3D Gaussians g_σ are respectively given by:

$$g_{\sigma_\rho}(x,y) := A_1 \exp\left(-\frac{x^2+y^2}{2\sigma_\rho^2}\right) = A_1 \exp\left(-\frac{\rho^2}{2\sigma_\rho^2}\right) \tag{3.1}$$

$$g_{\sigma_\rho,\sigma_z}(x,y,z) := A_2 \exp\left(-\frac{x^2+y^2}{2\sigma_\rho^2} - \frac{z^2}{2\sigma_z^2}\right) = A_2 \exp\left(-\frac{\rho^2}{2\sigma_\rho^2} - \frac{z^2}{2\sigma_z^2}\right) \tag{3.2}$$

We want to find the best Gaussian parameter, i.e. $\sigma^* = \sigma_\rho^*$ in 2D and $\sigma^* = \{\sigma_\rho^*, \sigma_z^*\}$ in 3D, that minimizes the least squares (LSQ) criterion, i.e.

$$\sigma^* = \underset{\sigma>0}{\operatorname{argmin}} \|\mathrm{PSF} - g_\sigma\|_2^2 \tag{3.3}$$

In this section, we suppose that the Gaussian functions and the PSFs are normalized according to their L^∞ norm i.e. $\|g_\sigma\|_\infty = \|\mathrm{PSF}\|_\infty = 1$. We have thus $A_1 = A_2 = 1$. It can be noted that since the Gaussians, the WFFM PSFs, the LSCM PSFs and the DSCM PSFs (with d sufficiently large) are all maximized at the origin, this normalization imposes matching the peaks of g_σ and the PSF. In the following, we will denote by $\hat{\sigma}^*$ the proposed parameter and by σ^* the optimal parameter given by (3.3).

3.1.1 Gaussian approximations of 2D WFFM PSFs

We begin with the 2D paraxial WFFM PSF, which is known as "Airy disk" and can be directly derived from (1.27) and (2.7) by setting $z = 0$, i.e.

$$\mathrm{PSF}_{WFFM}(\rho) = \left[2\frac{J_1(k_{em}\mathrm{NA}\rho)}{k_{em}\mathrm{NA}\rho}\right]^2 \tag{3.4}$$

where J_1 is the first order Bessel function of the first kind and $k_{em} = \frac{2\pi}{\lambda_{em}}$ is the emission wavenumber. The solution to (3.3) is given by Proposition 2 (proof in Appendix A.5).

Proposition 2 *The unique solution to (3.3) under the condition* $\|g_\sigma\|_\infty = \|PSF\|_\infty = 1$ *is given by*

$$\sigma_\rho^* \approx 0.21\frac{\lambda_{em}}{NA} \tag{3.5}$$

This value (3.5) was first given in [38].

Unlike the paraxial case where the optimal parameter value can be found exactly, a direct minimization of (3.3) is difficult in the non-paraxial case since the non-paraxial PSF,

derived by inserting (1.26) into (2.7) with $z = 0$, has no closed form. However, we can note that the L^2 energy of the PSF ($\|\mathrm{PSF}_{WFFM}\|_2^2$) concentrates mostly in the main lobe, which is situated in a small neighborhood of the origin. Therefore, an approximation near the origin capturing the decay of the main lobe, e.g. $g_\sigma(\rho) \to \mathrm{PSF}_{WFFM}(\rho)$ as $\rho \to 0$, can be expected to give a good global approximation. This can be easily achieved by matching the Maclaurin series of the PSF ((A.8) in Appendix A.7) with that of a Gaussian function ((A.6) in Appendix A.7) with $z = 0$. The two expansions differ in their second and higher order terms. Thus, by imposing the equality on their second order terms, we obtain:

$$\hat{\sigma}_\rho^* = \frac{1}{nk_{em}} \left[\frac{4 - 7\cos^{\frac{3}{2}}\alpha + 3\cos^{\frac{7}{2}}\alpha}{7(1 - \cos^{\frac{3}{2}}\alpha)} \right]^{-\frac{1}{2}} \tag{3.6}$$

The positivity of the bracketed term in (3.6) can be easily verified, so $\hat{\sigma}_\rho^*$ is well defined. From a geometric point of view, this approach matches the principal curvatures of the graphs of the two functions at the origin [44], since the gradients and the second-order mixed derivatives of the PSF and of the Gaussian vanish at the origin.

The near-optimality of (3.6) under the original LSQ criterion (3.3) will be confirmed by the numerical results in Section 3.3. Yet, the consistency analysis below already sheds some light on why (3.6) can be considered as a "good" parameter.

3.1.1.1 Consistency of the Gaussian parameters

As $\alpha \to 0$ or $\mathrm{NA} \to 0$, i.e. as the system becomes paraxial, we have,

$$\lim_{\alpha \to 0+} \frac{(3.6)}{\sqrt{2}/(k_{em}\mathrm{NA})} = 1 \tag{3.7}$$

Therefore, (3.6) tends asymptotically to $\sqrt{2}/(k_{em}\mathrm{NA}) \approx 0.225\lambda_{em}/\mathrm{NA}$, a value only a few percent larger than (3.5). This implies that the non-paraxial parameter (3.6) gets close to the optimal solution in the paraxial limit.

3.1.2 Gaussian approximations of 2D LSCM PSFs

The Maclaurin expansion of the 2D paraxial LSCM PSF is given by (A.9) with $z = 0$. By applying the same series-matching method as in Section 3.1.1, the Gaussian parameter is found to be:

$$\hat{\sigma}_\rho^* = \sqrt{2} \left[\frac{c_1^2}{r^2} + \frac{4c_2 J_0(c_2) J_1(c_2) - 8J_1^2(c_2)}{r^2[J_0^2(c_2) + J_1^2(c_2) - 1]} \right]^{-\frac{1}{2}} \tag{3.8}$$

where $c_1 = k_{ex} r\mathrm{NA}$, $c_2 = k_{em} r\mathrm{NA}$ and $k_{ex} = \frac{2\pi}{\lambda_{ex}}$ is the excitation wavenumber.

In the non-paraxial case, due to the complexity of the non-paraxial PSF, the parameter derived by series matching has no closed-form expression, which is inconvenient for practical use. One solution to this is based on the observation that the non-paraxial LSCM PSF will be much simplified if the terms $|\mathbf{h}|$ in (2.8) are first approximated by Gaussian functions. This pre-approximation of the PSF model is reasonable, since the terms $|\mathbf{h}|$ share the same form as the 2D non-paraxial WFFM PSF, which can be very accurately approximated by the Gaussians with the parameters derived in the previous section (see Section 3.3 for the numerical results). By matching the series of this pre-approximated PSF, i.e. (A.10) with $z = 0$, and that of a Gaussian, we obtain:

$$\hat{\sigma}_\rho^* = \sqrt{2}\left[\frac{2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right] + r^2\sigma_{ex,\rho}^2}{\sigma_{ex,\rho}^2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]}\right]^{-\frac{1}{2}} \tag{3.9}$$

where $\sigma_{em,\rho}$ is given by (3.6) and $\sigma_{ex,\rho}$ is also expressed by (3.6) with k_{em} replaced by k_{ex}.

3.1.2.1 Consistency of the Gaussian parameters

First, we can see that as the system becomes paraxial, the non-paraxial parameter (3.9) tends asymptotically to the paraxial one (3.8):

$$\lim_{\alpha\to 0+}\frac{(3.9)}{(3.8)} = 1 \tag{3.10}$$

Second, we study the asymptotic behavior of (3.8) and (3.9) in the vanishing pinhole case ($r \to 0+$), i.e. as the system approaches ideal confocality. In this situation, the pinhole is modelled as a Dirac distribution and the LSCM PSF (2.8) simply reduces to the product of two WFFM PSFs. Maclaurin series matching is then applied for this ideal confocality case to derive the paraxial and non-paraxial Gaussian parameters, given by (3.11) and (3.12) respectively.

$$\hat{\sigma}_\rho^* = \frac{\sqrt{2}}{2\pi\mathrm{NA}}\frac{\lambda_{ex}\lambda_{em}}{(\lambda_{ex}^2+\lambda_{em}^2)^{1/2}} \tag{3.11}$$

$$\hat{\sigma}_\rho^* = \frac{\lambda_{ex}\lambda_{em}}{2\pi n(\lambda_{ex}^2+\lambda_{em}^2)^{\frac{1}{2}}}\left[\frac{7\cos^{\frac{3}{2}}\alpha - 3\cos^{\frac{7}{2}}\alpha - 4}{7(\cos^{\frac{3}{2}}\alpha - 1)}\right]^{-\frac{1}{2}} \tag{3.12}$$

The following results confirm the consistency of the parameters (3.8) and (3.9) with (3.11) and (3.12) as $r \to 0+$.

$$\lim_{r\to 0+}(3.8) = (3.11), \quad \lim_{r\to 0+}(3.9) = (3.12) \tag{3.13}$$

Interestingly, we can note that if the Stokes shift [27] is negligible, i.e. $\lambda_{ex} \approx \lambda_{em}$, both pairs of equations (3.5)(3.11) and (3.6)(3.12) imply that in the ideal confocality situation, the effective width of the LSCM PSF is approximately 1.4 times smaller than that of the WFFM PSF. Here, we see the well-known resolution gain factor in LSCM.

Finally, for a full-open pinhole ($r \to +\infty$), the equations below show that the Gaussian parameters (3.8) and (3.9) tend towards those of WFFM with λ_{em} replaced by λ_{ex}, which is consistent with the known fact that in this case an LSCM behaves as a WFFM.

$$\lim_{r\to+\infty}(3.8) = \frac{\sqrt{2}}{k_{ex}\mathrm{NA}} \approx 0.225\frac{\lambda_{ex}}{\mathrm{NA}}, \quad \lim_{r\to+\infty}(3.9) = \frac{1}{nk_{ex}}\left[\frac{7\cos^{\frac{3}{2}}\alpha - 3\cos^{\frac{7}{2}}\alpha - 4}{7(\cos^{\frac{3}{2}}\alpha - 1)}\right]^{-\frac{1}{2}} \tag{3.14}$$

3.1.3 Gaussian approximations of 2D DSCM PSFs

The DSCM case is more complex as the approximation quality varies with the adjacent pinhole distance d. We will assume that d is sufficiently large so that the PSF is almost circularly symmetric, or equivalently, that the contribution of individual excitation distributions from adjacent pinholes to the total PSF is negligible. This condition can be considered fulfilled if the support of the main lobe of the emission PSF contains only one main lobe of the excitation PSF. In terms of d, we should require:

$$d \geq d_0 := \frac{1}{2}\left(\frac{\lambda_{em}}{\lambda_{ex}} + D + 1\right) \ \mathrm{AU} \tag{3.15}$$

where "AU" stands for the "Airy Unit", i.e. $1\mathrm{AU} = 1.22 \cdot \lambda_{ex}/\mathrm{NA}$. Under this condition, the DSCM PSF approaches reasonably well the LSCM PSF, and the Gaussian parameters (3.8) and (3.9) can be applied respectively to the paraxial and non-paraxial DSCM PSFs (see also Section 3.3).

3.1.4 Gaussian approximations of 3D PSFs

The Maclaurin series of the 3D paraxial WFFM PSF, 3D non-paraxial WFFM PSF, 3D paraxial LSCM PSF and 3D pre-approximated non-paraxial LSCM PSF are given by (A.7), (A.8), (A.9) and (A.10) in Appendix A.7. The lateral and axial Gaussian parameters for these PSFs are found by the same series matching method as described in the 2D case. The same Gaussian parameters as LSCM are proposed for the DSCM PSFs under the condition (3.15).

Tab. 3.1, 3.2 and 3.3 summarize the proposed 2D and 3D Gaussian parameters. The consistency of the 3D parameters can be studied in the same manner as in the 2D case and the main results are listed below:

Paraxial system: As the NA becomes small, the 3D non-paraxial lateral and axial Gaussian parameters of WFFM, LSCM and DSCM tend asymptotically to the paraxial Gaussian parameters;

Ideal confocality: As the pinhole radius approaches zero, the 3D lateral and axial Gaussian parameters of LSCM and DSCM converge to the parameters derived in the vanishing pinhole situation;

Full-open pinhole: As the pinhole radius tends to infinity, the 3D Gaussian parameters of LSCM and DSCM approach those of WFFM with λ_{em} replaced by λ_{ex}, *except for* the non-paraxial axial parameter $\hat{\sigma}_z^*$. The latter exception implies that the 3D non-paraxial Gaussian approximations of LSCM and DSCM PSFs cannot be applied to large pinhole situations.

3.2 Gaussian Approximations of the PSF Models with an L^1 Constraint

In the applications where photometry should be preserved, it is important to consider the L^1 constraint $\|g_\sigma\|_1 = \|\text{PSF}\|_1 = 1$ in the optimization problem (3.3) instead of the L^∞ constraint. This constraint requires the PSF energy to be conserved by the Gaussian function used to approximate the PSF.

Note that this constrained optimization is infeasible for 3D WFFM PSFs, since they are not L^1 functions. In general, analytical solutions to this optimization problem are difficult to obtain except for the 2D paraxial WFFM case, where we can show the following result (proof in Appendix A.5).

Proposition 3 *The unique solutions to (3.3) under the condition* $\|g_\sigma\|_1 = \|PSF\|_1 = 1$ *is given by:*

$$\sigma_\rho^* \approx 0.22 \frac{\lambda_{em}}{NA} \tag{3.16}$$

The general result of the approximation of the 2D paraxial WFFM PSF with an L^p constraint ($1 \leq p \leq \infty$) is given by Proposition 4 (proof in Appendix A.6).

Proposition 4 *The solution to (3.3) for* $\|g_\sigma\|_p = \|PSF\|_p = 1$ *with* $1 \leq p \leq \infty$*, if it exists, satisfies necessarily the following equation (taking the limit for the case* $p = \infty$*):*

$$c^2\|h\|_p \left(\frac{p}{2\pi}\right)^{\frac{1}{p}} \sigma^{2-\frac{2}{p}} \left(1 - \frac{2}{p}\right) = 8\left[p^{-1}\left(e^{-c^2\sigma^2} I_0(c^2\sigma^2) - 1\right) + e^{-c^2\sigma^2} I_1(c^2\sigma^2)(p^{-1}+1)\right] \tag{3.17}$$

3.3 Numerical Evaluations of the Approximations

The approximation error is evaluated using the Relative Squared Error (RSE) defined by

$$\text{RSE} := \frac{\|\text{PSF} - g_{\hat{\sigma}^*}\|_2^2}{\|\text{PSF}\|_2^2} \tag{3.18}$$

where the Gaussian function and the PSF are both normalized according to their L^∞ norm or L^1 norm, depending on the approximation constraint used. Clearly, this criterion is essentially the same as the one defined in (3.3), since the squared L^2 norm of the PSF in (3.18) is just a normalizing constant. To compute the optimal Gaussian parameter σ^* of (3.3), a numerical LSQ fit is used. Then, the Parameter Relative Error (PRE), i.e. $|\hat{\sigma}^* - \sigma^*|/\sigma^*$, is evaluated.

3.3.1 Numerical results

In our simulations, λ_{ex} and λ_{em} are set to 488nm and 520nm respectively, which are two wavelengths frequently used in real experiments. The refractive index is set to $n = 1.515$, which is the typical value of immersion oils. In LSCM and DSCM, the pinhole diameter ranges from 0AU to 3AU, i.e. from a vanishing size to a large size. For DSCM, d is set to d_0. The NA varies from 0.2 to 0.7 in the paraxial cases, and from 0.8 to 1.4 in the non-paraxial cases. Given an NA value, the exact theoretical PSFs are computed using (2.7), (2.8) and (2.9), the Gaussian approximations are generated using the parameters shown in Tab. 3.1, 3.2 and 3.3, and then the approximation errors are evaluated. The results of the approximations with the L^∞ constraint are shown in Tab. 3.4, 3.6 and 3.7, where we present the minimal and maximal RSE values for the WFFM, LSCM and DSCM cases. The minimal and maximal PRE values of the lateral and axial parameters ($\hat{\sigma}_\rho^*$ and $\hat{\sigma}_z^*$) are also shown. The results of the approximation of the 2D paraxial WFFM PSF with the L^1 constraint are given in Tab. 3.5. Examples of the Gaussian approximations with the two constraints are presented in Fig. 3.1 and Fig. 3.3.

3.3.2 Discussion

The following conclusions can be drawn by examining Tab. 3.4, 3.6, 3.7 and 3.5.

2D approximations for WFFM, LSCM and DSCM The approximation accuracy for all 2D PSFs is very high, since we have RSE $< 2\%$ in WFFM (cf. Tab. 3.4 and 3.5), $\leq 2\%$ in LSCM (cf. Tab. 3.6) and $< 4\%$ in DSCM (cf. Tab. 3.7).

3D approximations for WFFM Tab. 3.4 shows that the approximations of the 3D WFFM PSFs are only average, as an RSE of about 17% is observed. However, this is not a defect of the Maclaurin series matching method, since the very low PRE values ($\simeq 2\%$) confirm the near-optimality of the proposed parameter $\hat{\sigma}^*$. Indeed, the numerical LSQ Gaussian fits are found to result in almost the same RSE (data not shown). It follows that an RSE of about 17% is the lower error bound, whatever the approximation approach used. This inaccuracy is actually due to the fact that the axial WFFM PSF decreases slowly (as $O(z^{-2})$). In contrast, the axial decreasing speed of the LSCM PSF with typical pinhole sizes is much higher (as $O(z^{-4})$), since the PSF is in the form of the product of excitation and emission PSFs both having $O(z^{-2})$ as decreasing rate. This can also be seen in Figure 3.1, where secondary lobes with significant amplitudes are present in Figure 3.1(d) (axial WFFM PSF), while they are almost invisible in Figure 3.1(e) (axial LSCM PSF). We also show the xz-planes of the PSFs, of their Gaussian approximations and of the approximation residues in Fig. 3.2. It can be seen from the residues that the WFFM secondary lobes decrease much more slowly than those of LSCM. Hence, Gaussian approximations perform much better for LSCM with typical pinhole sizes (see below).

3D approximations for LSCM and DSCM Tab. 3.6 and 3.7 show that, the 3D approximations are accurate for DSCM with typical pinhole sizes (RSE $< 7\%$, $D < 1$AU), and accurate for LSCM up to reasonably large pinhole sizes (RSE $< 9\%$, $D < 3$AU). In particular, the 3D approximations for LSCM with typical pinhole sizes are nearly perfect (RSE $< 1\%$, $D < 1$AU).

As pinholes become larger and larger, the PSFs tend asymptotically to WFFM PSFs. It follows from the previous discussion on 3D approximations for WFFM that in this case, any Gaussian approximation will become inaccurate. In the experiments, as can be seen in Tab. 3.6, the performance of our approximations for LSCM degrades (RSE $> 10\%$) as

$D \geq 3$AU. In the case of DSCM, the RSE values are larger than 10% as long as $D \geq 1$AU (cf. Tab. 3.7).

Therefore, we conclude that the regions where our 3D approximations perform accurately for LSCM and for DSCM are given by the sets of D satisfying $D < 3$AU and $D < 1$AU, respectively.

Accuracy of Gaussian parameters As can be seen from Tab. 3.4, 3.6 and 3.5, in the cases of WFFM, 2D LSCM, and 3D LSCM with $D < 3$AU, almost all the PRE values are within only a few percent, implying the near-optimality of the proposed Gaussian parameters. For 2D DSCM and for 3D DSCM with $D < 1$AU, the PRE values are generally larger than those in LSCM under the same condition but remain satisfactory (cf. Tab. 3.7).

Finally, we point out that as the optimal Gaussian parameters (3.5) and (3.16) derived for 2D paraxial WFFM PSF have undergone numerical approximations (see Appendix A.5), their PRE values deviate slightly from zero (cf. Tab. 3.4 and 3.5).

3.4 Conclusion

We have studied comprehensively the least squares Gaussian approximations of the diffraction-limited 2D/3D paraxial/non-paraxial PSFs of WFFM, LSCM and DSCM described using the Debye integrals. Optimal Gaussian parameters are derived for the 2D paraxial WFFM PSF, under both the L^∞ constraint imposing peak matching and the L^1 constraint imposing energy conservation. For the other PSFs, with the L^∞ constraint, near-optimal parameters in explicit forms are derived using Maclaurin series matching. We found that: (i) the 2D approximations are all very accurate; (ii) no accurate Gaussian approximation exists for 3D WFFM PSFs; (iii) with typical pinhole sizes, the 3D approximations are accurate for DSCM and nearly perfect for LSCM.

Tab. 3.1: Gaussian parameters for 2D PSFs

Microscope	Gaussian Parameter[1]
Paraxial WFFM (L^∞ constraint)	$\hat{\sigma}_\rho^* = 0.21\lambda_{em}/\mathrm{NA}$
Paraxial WFFM (L^1 constraint)	$\hat{\sigma}_\rho^* = 0.22\lambda_{em}/\mathrm{NA}$
Non-paraxial WFFM (L^∞ constraint)	$\hat{\sigma}_\rho^* = \frac{1}{nk_{em}}\left[\frac{4-7\cos^{\frac{3}{2}}\alpha+3\cos^{\frac{7}{2}}\alpha}{7(1-\cos^{\frac{3}{2}}\alpha)}\right]^{-\frac{1}{2}}$
Paraxial LSCM and DSCM ($d \geq d_0$, L^∞ constraint)	$\hat{\sigma}_\rho^* = \sqrt{2}\left[\frac{c_1^2}{r^2} + \frac{4c_2J_0(c_2)J_1(c_2)-8J_1^2(c_2)}{r^2[J_0^2(c_2)+J_1^2(c_2)-1]}\right]^{-\frac{1}{2}}$
Non-paraxial LSCM and DSCM ($d \geq d_0$, L^∞ constraint)	$\hat{\sigma}_\rho^* = \sqrt{2}\left[\frac{2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]+r^2\sigma_{ex,\rho}^2}{\sigma_{ex,\rho}^2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]}\right]^{-\frac{1}{2}}$

[1] $k_{ex} := \frac{2\pi}{\lambda_{ex}}$, $k_{em} := \frac{2\pi}{\lambda_{em}}$, $c_1 := k_{ex}r\mathrm{NA}$, $c_2 := k_{em}r\mathrm{NA}$, $\sigma_{em,\rho}$ is given by the expression of $\hat{\sigma}_\rho^*$ of the non-paraxial WFFM (L^∞ constraint) and $\sigma_{ex,\rho}$ is given by the same expression with k_{em} replaced by k_{ex}.

Tab. 3.2: Lateral Gaussian parameters for 3D PSFs (L^∞ constraint)

Microscope	Lateral Gaussian Parameter[1]
Paraxial WFFM	$\hat{\sigma}_\rho^* = \sqrt{2}/(k_{em}\mathrm{NA})$
Non-paraxial WFFM	$\hat{\sigma}_\rho^* = \frac{1}{nk_{em}}\left[\frac{4-7\cos^{\frac{3}{2}}\alpha+3\cos^{\frac{7}{2}}\alpha}{7(1-\cos^{\frac{3}{2}}\alpha)}\right]^{-\frac{1}{2}}$
Paraxial LSCM and DSCM ($d \geq d_0$)	$\hat{\sigma}_\rho^* = \sqrt{2}\left[\frac{c_1^2}{r^2} + \frac{4c_2J_0(c_2)J_1(c_2)-8J_1^2(c_2)}{r^2[J_0^2(c_2)+J_1^2(c_2)-1]}\right]^{-\frac{1}{2}}$
Non-paraxial LSCM and DSCM ($d \geq d_0$)	$\hat{\sigma}_\rho^* = \sqrt{2}\left[\frac{2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]+r^2\sigma_{ex,\rho}^2}{\sigma_{ex,\rho}^2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]}\right]^{-\frac{1}{2}}$

[1] $k_{ex} := \frac{2\pi}{\lambda_{ex}}$, $k_{em} := \frac{2\pi}{\lambda_{em}}$, $c_1 := k_{ex}r\mathrm{NA}$, $c_2 := k_{em}r\mathrm{NA}$, $\sigma_{em,\rho}$ is given by the expression of $\hat{\sigma}_\rho^*$ of the non-paraxial WFFM and $\sigma_{ex,\rho}$ is given by the same expression with k_{em} replaced by k_{ex}.

Tab. 3.3: Axial Gaussian parameters for 3D PSFs (L^∞ constraint)

Microscope	Axial Gaussian Parameter[1]
Paraxial WFFM	$\hat{\sigma}_z^* = 2\sqrt{6}\cdot n/(k_{em}\mathrm{NA}^2)$
Non-paraxial WFFM	$\hat{\sigma}_z^* = \frac{5\sqrt{7}(1-\cos^{\frac{3}{2}}\alpha)}{\sqrt{6}\cdot nk_{em}\left[4\cos^5\alpha-25\cos^{\frac{7}{2}}\alpha+42\cos^{\frac{5}{2}}\alpha-25\cos^{\frac{3}{2}}\alpha+4\right]^{\frac{1}{2}}}$
Paraxial LSCM and DSCM ($d \geq d_0$)	$\hat{\sigma}_z^* = 2\sqrt{6}\left[\frac{c_1^2\mathrm{NA}^2}{r^2n^2} - \frac{48c_2^2[J_0^2(c_2)+J_1^2(c_2)]-192J_1^2(c_2)}{n^2k_{em}^2r^4[J_0^2(c_2)+J_1^2(c_2)-1]}\right]^{-\frac{1}{2}}$
Non-paraxial LSCM and DSCM ($d \geq d_0$)	$\hat{\sigma}_z^* = \frac{\sigma_{ex,z}\sigma_{em,z}}{\left[\sigma_{ex,z}^2+\sigma_{em,z}^2\right]^{\frac{1}{2}}}$

[1] $k_{ex} := \frac{2\pi}{\lambda_{ex}}$, $k_{em} := \frac{2\pi}{\lambda_{em}}$, $c_1 := k_{ex}r\mathrm{NA}$, $c_2 := k_{em}r\mathrm{NA}$, $\sigma_{em,z}$ is given by the expression of $\hat{\sigma}_z^*$ of the non-paraxial WFFM and $\sigma_{ex,z}$ is given by the same expression with k_{em} replaced by k_{ex}.

Tab. 3.4: Approximation errors[§] on WFFM PSFs (L^∞ constraint)

RSE%	2D Parax.	2D Non-Parax.	3D Parax.		3D Non-Parax.	
	(0.4, 0.5)	(1.6, 1.7)	(16.0, 19.9)		(16.2, 17.4)	
PRE%	$\hat{\sigma}_\rho^*$	$\hat{\sigma}_\rho^*$	$\hat{\sigma}_\rho^*$	$\hat{\sigma}_z^*$	$\hat{\sigma}_\rho^*$	$\hat{\sigma}_z^*$
	(0.3, 1.5)	(7.0, 7.0)	(0.6, 2.0)	(2.1, 2.4)	(2.2, 2.6)	(0.9, 1.9)

[§] The top part of the table shows the RSE% and the bottom part shows the PRE%. In parentheses, the minimal and maximal errors are shown: (Min.Err.%, Max.Err.%). The NA varies from 0.2 to 0.7 in the paraxial cases and from 0.8 to 1.4 in the non-paraxial cases.

Tab. 3.5: Approximation errors[§] on the 2D paraxial WFFM PSF (L^1 constraint)

RSE%	PRE% ($\hat{\sigma}_\rho^*$)
(1.1, 1.3)	(0.0, 1.7)

[§] In parentheses, the minimal and maximal errors are shown: (Min.Err.%, Max.Err.%). The NA varies from 0.2 to 0.7.

Tab. 3.6: Approximation errors[§] on LSCM PSFs (L^∞ constraint)

RSE%	2D Parax.	2D Non-Parax.	3D Parax.		3D Non-Parax.	
D=0[†]	(0.1, 0.3)	(0.2, 0.3)	(0.3, 0.4)		(0.4, 0.4)	
D=0.25	(0.1, 0.2)	(0.3, 0.6)	(0.3, 0.5)		(0.5, 0.8)	
D=0.5	(0.1, 0.3)	(0.6, 0.6)	(0.4, 0.5)		(0.6, 0.7)	
D=1	(1.3, 2.0)	(1.0, 1.0)	(1.7, 2.2)		(1.5, 1.6)	
D=2	(1.1, 1.6)	(1.5, 1.6)	(5.3, 5.7)		(7.6, 8.5)	
D=3	(1.1, 1.6)	(1.5, 1.7)	(9.2, 9.6)		(11.8, 13.5)	
PRE%	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_z$	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_z$
D=0	(0.5, 3.1)	(2.8, 3.2)	(0.2, 2.4)	(2.5, 2.9)	(2.1, 2.5)	(2.2, 2.4)
D=0.25	(0.6, 3.0)	(3.7, 5.2)	(0.3, 2.2)	(2.4, 3.3)	(2.9, 4.3)	(1.9, 2.7)
D=0.5	(0.9, 3.1)	(5.2, 5.3)	(0.2, 2.0)	(1.7, 2.1)	(4.1, 4.2)	(0.9, 1.2)
D=1	(7.4, 9.0)	(5.4, 5.7)	(5.6, 7.3)	(0.0, 0.3)	(3.9, 4.1)	(7.6, 8.3)
D=2	(5.9, 7.5)	(6.9, 7.1)	(1.9, 3.5)	(3.7, 4.0)	(3.3, 3.5)	(20.6, 21.2)
D=3	(5.7, 7.2)	(6.8, 6.9)	(1.0, 2.6)	(2.6, 2.8)	(2.4, 2.8)	(24.0, 24.3)

[§] The top part of the table shows the RSE% and the bottom part shows the PRE%. In parentheses, the minimal and maximal errors are shown: (Min.Err.%, Max.Err.%). The NA varies from 0.2 to 0.7 in the paraxial cases and from 0.8 to 1.4 in the non-paraxial cases.

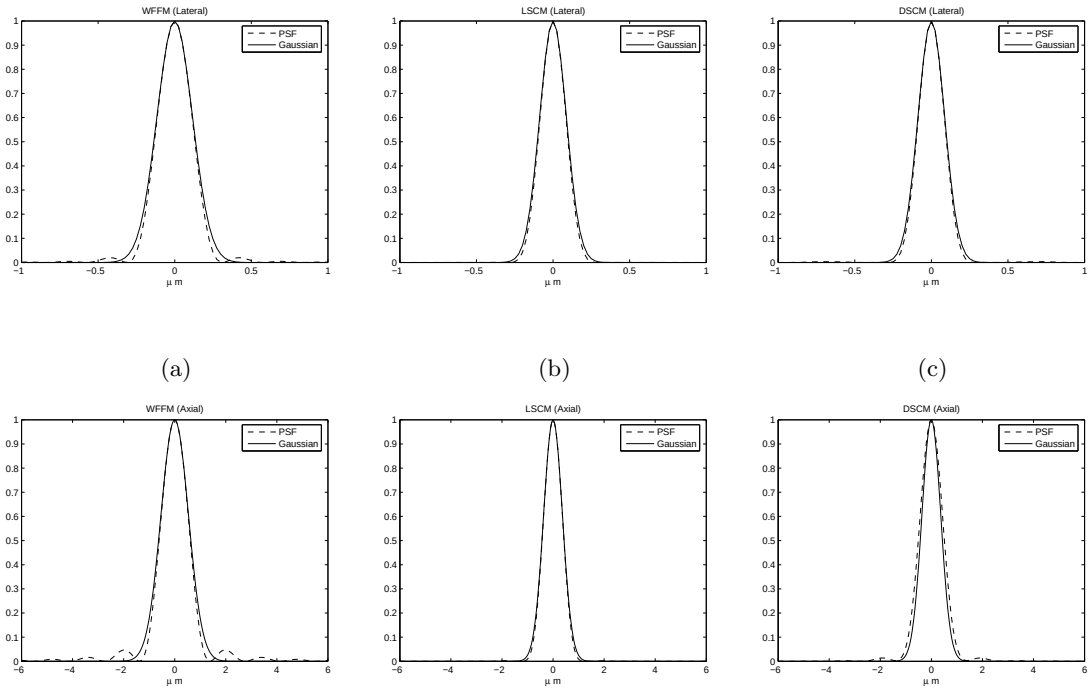
[†] D is the pinhole diameter and its unit is Airy Unit (1AU = $1.22 \cdot \lambda_{ex}/\mathrm{NA}$). D=0 corresponds to the vanishing pinhole situation.

Tab. 3.7: Approximation errors[§] on DSCM PSFs (L^∞ constraint)

RSE%	2D Parax.	2D Non-Parax.	3D Parax.		3D Non-Parax.	
D=0[†]	(3.3, 3.3)	(3.2, 3.4)	(6.1, 6.1)		(3.9, 5.5)	
D=0.25	(2.0, 2.1)	(2.2, 2.4)	(2.8, 2.8)		(2.3, 2.9)	
D=0.5	(0.3, 0.4)	(0.5, 0.7)	(2.3, 2.3)		(2.5, 3.3)	
D=1	(1.2, 1.2)	(2.1, 2.8)	(11.6, 11.8)		(16.2, 18.4)	
D=2	(3.8, 3.8)	(3.7, 4.0)	(16.4, 16.9)		(15.5, 16.7)	
D=3	(1.6, 1.7)	(1.6, 2.3)	(19.0, 21.9)		(23.2, 24.5)	
PRE%	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_z$	$\hat{\sigma}^*_\rho$	$\hat{\sigma}^*_z$
D=0	(12.6, 12.6)	(12.3, 12.7)	(14.4, 14.4)	(2.3, 2.3)	(13.5, 14.4)	(0.1, 3.4)
D=0.25	(9.7, 10.2)	(10.1, 10.8)	(10.5, 11.0)	(4.8, 4.9)	(10.5, 11.5)	(5.8, 8.2)
D=0.5	(3.1, 3.2)	(4.2, 5.2)	(3.4, 3.5)	(12.8, 13.0)	(4.1, 5.4)	(14.4, 18.9)
D=1	(5.7, 5.8)	(9.0, 10.4)	(5.6, 5.6)	(26.6, 26.6)	(8.8, 10.1)	(32.9, 35.9)
D=2	(12.4, 12.4)	(11.6, 12.0)	(9.1, 9.1)	(7.4, 7.5)	(8.9, 9.3)	(18.7, 21.0)
D=3	(7.1, 7.1)	(6.8, 7.7)	(2.1, 2.1)	(3.4, 3.4)	(1.9, 3.2)	(20.1, 23.4)

[§] The top part of the table shows the RSE% and the bottom part shows the PRE%. In parentheses, the minimal and maximal errors are shown: (Min.Err.%, Max.Err.%). The NA varies from 0.2 to 0.7 in the paraxial cases and from 0.8 to 1.4 in the non-paraxial cases. d is set to d_0.

[†] D is the pinhole diameter and its unit is Airy Unit (1AU $= 1.22 \cdot \lambda_{ex}/\mathrm{NA}$). D=0 corresponds to the vanishing pinhole situation.



(a) (b) (c)

(d) (e) (f)

Fig. 3.1: Examples of the Gaussian approximations of WFFM, LSCM and DSCM PSFs with an L^∞ constraint. Non-paraxial cases, $\lambda_{ex} = 488$nm, $\lambda_{em} = 520$nm, $n = 1.515$, NA $= 1.0$ and the pinhole diameter $D = 0.5$AU in LSCM and DSCM. (a) Lateral WFFM PSF; (b) Lateral LSCM PSF; (c) Lateral DSCM PSF ($d = d_0$); (d) Axial WFFM PSF; (e) Axial LSCM PSF; (f) Axial DSCM PSF ($d = d_0$).

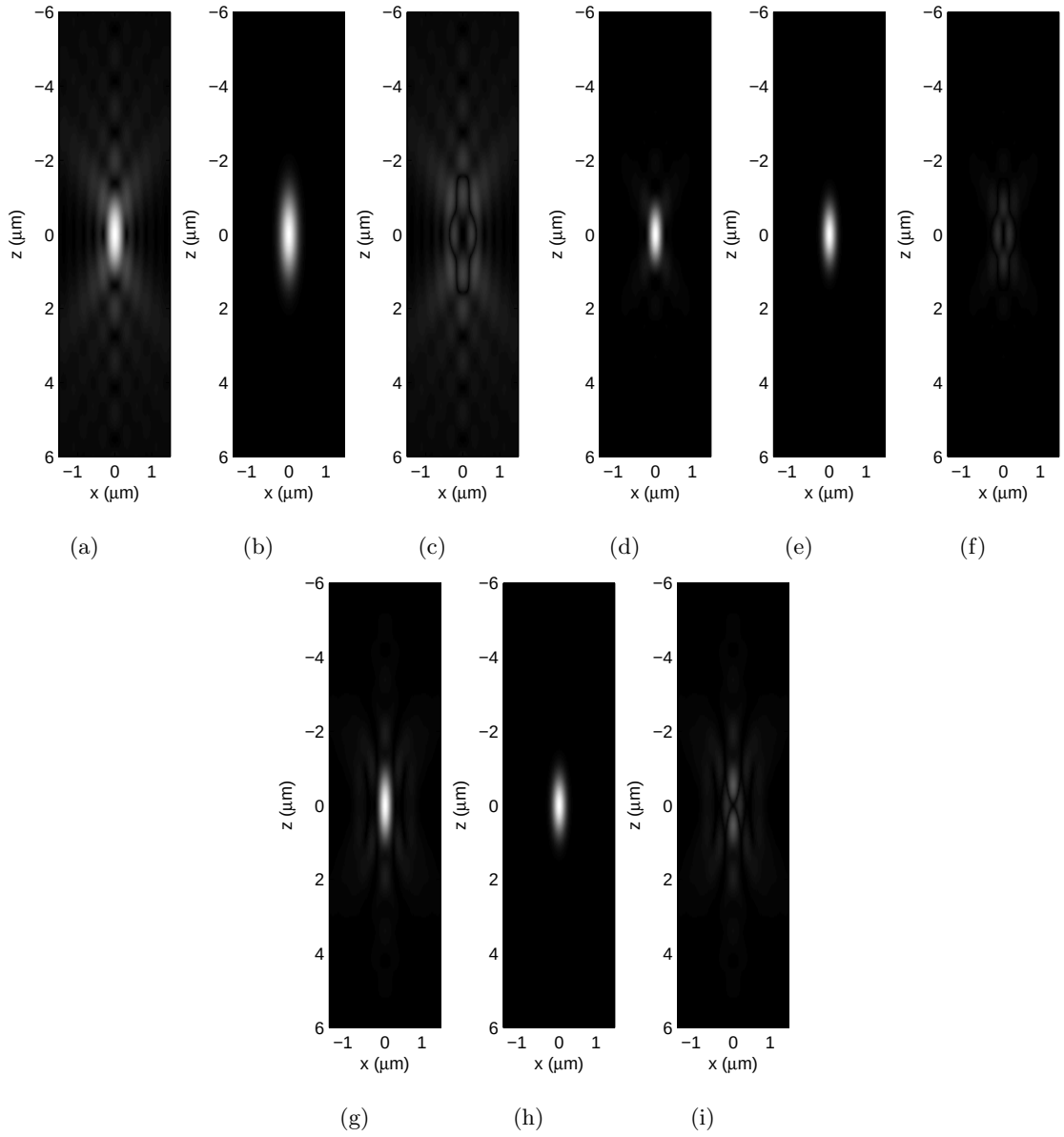
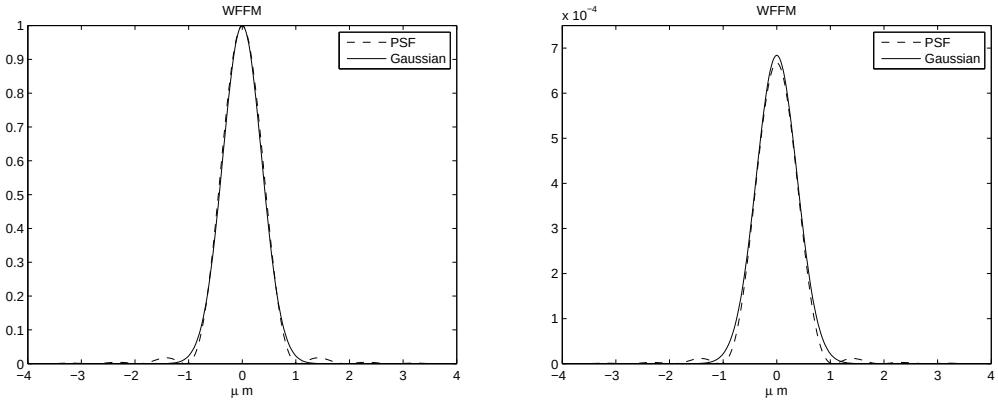


Fig. 3.2: xz-planes of nonparaxial microscope PSFs (NA = 1.0), of their Gaussian approximations and of the approximation residues. (a) WFFM PSF; (b) Gaussian approximation for (a); (c) Residue for (b); (d) LSCM PSF ($D = 0.5$AU); (e) Gaussian approximation for (d); (f) Residue for (e); (g) DSCM PSF ($D = 0.5\text{AU}, d = d_0$); (h) Gaussian approximation for (g); (i) Residue for (h).



(a) (b)

Fig. 3.3: Examples of the Gaussian approximations of 2D Paraxial WFFM PSF. (a) approximation with an L^∞ constraint; (b) approximation with an L^1 constraint. $\lambda_{em} = 520$nm, $n = 1.515$ and NA $= 0.3$.

Part II

Fluorescence Image Restoration

Chapter 4

Noise Sources in Fluorescence Microscope

The data acquired by a fluorescence microscope are contaminated by various noise sources in the microscope imaging system. As discussed in Chapter 2, principal instruments used for fluorescence detection are CCDs and PMTs. This chapter will study the noise sources contributing to image degradation in these two types of detectors, and propose for each microscope its noise model. The CCD noise is studied in Section 4.1, and the noise process in a PMT is discussed in Section 4.2. We conclude by summarizing our noise models for different microscopes in Section 4.3.

4.1 CCD Noise Sources

A typical CCD camera functions as follows. The incident fluorescence photons hit the light-sensing photodiode elements of the CCD, resulting in the liberation of electrons. Each absorbed photon will create on average a mobile electron and a corresponding positively charged electron hole. The resulting charge in each pixel is linearly proportional to the number of incident photons. During the exposure time, the charge of each pixel will be accumulated and stored in a potential well. Then, the charge will be amplified and converted to a proportional voltage by the output amplifier. Finally, the voltage will be quantified by an analog-to-digital converter (ADC) and the corresponding value will be stored in the image file in the memory. Several noises are involved in the above detection process.

Fluorescence photon noise The fluorescence photon noise, due to the stochastic nature of the photon emission, is inherent in all optical signals. Supposing the average photon flux to be λ, the observed photon number N_p follows a Poisson distribution with intensity parameter λ, i.e., $N_p \sim \mathcal{P}(\lambda)$. If we dispose of a high photon flux, N_p will be asymptotically normally distributed with both the mean and the variance equal to λ.

Dark noise The kinetic vibration of silicon atoms in the CCD substrate will liberate electrons or holes even when no incident fluorescence photon is present. The resulting charge will contribute to the final signal and is termed dark noise. Secondary sources of the dark noise are cosmic rays or external high-energy radiation from nearby sources such as an indoor illumination. The dark noise N_d follows also a Poisson distribution, i.e., $N_d \sim \mathcal{P}(\lambda_d)$, where λ_d represents the average dark flux. In practice, the CCDs are usually cooled (cooled CCD) to reduce the dark noise.

Readout noise The readout noise primarily originates from the imperfectness of the output amplifier during the process of converting charge into a voltage signal. This noise is usually described by a normal distribution, i.e., $N_r \sim \mathcal{N}(\mu, \sigma^2)$. Sometimes, the readout noise may be frequency-dependent.

Quantification noise Suppose that the output voltage of the amplifier is within $V \in [-V_M/2, V_M/2]$. Given Q bits for the quantification and V as the input voltage, the output of the ADC is given by $\lfloor V/q + 1/2 \rfloor$, where $q = V_M/2^Q$ is the quantification step. The quantification noise can thereby be written as $N_q = V - q\lfloor V/q + 1/2 \rfloor$, and $N_q \in [-q/2, q/2]$. As N_q is bounded, its mean μ_{N_q} and standard deviation σ_{N_q} are also and it is easy to show that

$$|\mu_q| \leq q/2 = 2^{-(Q+1)} V_M, \quad \sigma_{N_q} \leq q = 2^{-Q} V_M$$

That is, μ_{N_q} and σ_{N_q} decrease exponentially to 0 as Q augments. Thereby, the quantification noise becomes rapidly negligible as the quantification bits increase.

4.1.1 Noise in wide-field microscopes

By the above analysis, if the quantification noise is ignored, the final detected signal S can be written as

$$S = \alpha(N_p + N_d) + N_r$$

where α is the overall gain of the camera. As N_p and N_d are independent Poisson variables, $N_p + N_d$ is also Poisson distributed. Thus, S can be further simplified as

$$S = \alpha N_s + N_r, \quad N_s \sim \mathcal{P}(\lambda + \lambda_d), \quad N_r \sim \mathcal{N}(\mu, \sigma^2) \tag{4.1}$$

We name (4.1) as a mixed-Poisson-Gaussian (MPG) model. In WFFM, λ is generally high. This is because the camera collects not only the photons from the focused plane but those from out-of-focus planes as well. Hence, N_s becomes asymptotically normally ($\mathcal{AN}$) distributed. Thereby,

$$S = \alpha(\lambda + \lambda_d) + \mu + B, \quad B \sim \mathcal{AN}(0, \alpha^2(\lambda + \lambda_d) + \sigma^2)$$

We have therefore a non-stationary Gaussian noise. For a CCD having a dominating readout noise ($\sigma \gg \alpha$), the noise part B can be further approximated by a white noise, i.e.,

$$S \approx \alpha(\lambda + \lambda_d) + \mu + B, \quad B \sim \mathcal{N}(0, \sigma^2)$$

4.1.2 Noise in disc-scanning confocal microscopes

In DSCMs, the noise can no longer be approximated to be Gaussian. This is because the detection pinholes reject the photons from out-focus planes, so that we have only a very limited number of detected photons and the Gaussian approximation for N_s is no longer valid. In addition, in contrast to the WFFM case high gain CCDs are usually employed in DSCM systems such as EM-CCDs. Consequently, our noise is signal-dependent, and the resulting data has a MPG nature (4.1).

4.2 PMT Noises

In LSCM systems, the detection is carried out by PMTs which usually have high gains. A photocathode is a critical element of a PMT. When the incidence photons hit the photocathode, part of the excitation energy will be converted and produce a flux of photoelectrons. The number of photoelectrons will be amplified through a chain of dynodes, resulting in a current pulse. A PMT can normally function in two modes, i.e., analog mode and photon-counting mode. When the emission rate is high, the light pulses overlap and produce a continuous waveform. This current is then amplified and measured by an ADC. However, if the emission rate is very low, direct photon counting is possible, since there are very few pileups of photon events.

4.2.1 Noise in laser-scanning confocal microscopes

Similar to the CCD case, for a PMT operating in the analog mode, the four noise sources discussed in Section 4.1 are present. We point out that the dark noise in a PMT is originated from the thermal emission of electrons from the photocathode and the dynodes, and the radiation from nearby secondary sources. Thus, ignoring the quantification noise, the final signal model is given by the MPG one (4.1). For the same reason as in DSCM, an approximation to the MPG model by a Gaussian one will be inaccurate.

For a PMT operating in the photon-counting mode, the signal will be Poisson-distributed (no readout noise), i.e.,

$$S = N_p + N_d \sim \mathcal{P}(\lambda + \lambda_d) \tag{4.2}$$

4.3 Conclusion

We studied in this chapter the different noise sources that degrade the image acquisition in fluorescence microscopes. The final signal models for WFFM, LSCM and DSCM are summarized as follows.

WFFM The signal is corrupted by a Gaussian noise:

$$S = \alpha(\lambda + \lambda_d) + \mu + B, \quad B \sim \mathcal{AN}(0, \alpha^2(\lambda + \lambda_d) + \sigma^2) \tag{4.3}$$

With a dominating readout noise, we have the approximation

$$S \approx \alpha(\lambda + \lambda_d) + \mu + B, \quad B \sim \mathcal{N}(0, \sigma^2) \tag{4.4}$$

LSCM We distinguish the two modes of a PMT.

- For a PMT operating in the analog mode, the signal model is MPG:

$$S = \alpha N_s + N_r, \quad N_s \sim \mathcal{P}(\lambda + \lambda_d), \quad N_r \sim \mathcal{N}(\mu, \sigma^2) \tag{4.5}$$

- For a PMT operating in the photon-counting mode, the model is Poissonian:

$$S \sim \mathcal{P}(\lambda + \lambda_d) \tag{4.6}$$

DSCM The signal model is MPG:

$$S = \alpha N_s + N_r, \quad N_s \sim \mathcal{P}(\lambda + \lambda_d), \quad N_r \sim \mathcal{N}(\mu, \sigma^2) \tag{4.7}$$

The signal restoration in the presence of a Poisson noise or a MPG noise will be studied in Chapter 5 and Chapter 6, respectively.

Chapter 5

Poisson Noise Removal

Denoising Gaussian additive noise present in a signal has been very well studied in the literature. Vast successful approaches have been proposed such as the optimal filters [45][46][47][48][49], multiscale denoisers [50][51][52][53][54], PDE-based smoothing [55][56][57], and Bayesian estimators [58][59][60][61]. These methods are well applicable for WFFM images, but probably not so for LSCM data obtained in the photon-counting mode (see Chapter 4). In contrast to denoising Gaussian additive noise, Poisson data restoration has been much less studied.

Let us first formulate the problem of Poisson denoising. We observe a discrete dataset of counts $\mathbf{X} = (X_i)_{i\in\mathbb{Z}^q}$ where X_i is a Poisson random variable of intensity λ_i, i.e., $X_i \sim \mathcal{P}(\lambda_i)$. Here we suppose that X_i's are mutually independent. The denoising aims at estimating the underlying intensity profile $\Lambda = (\lambda_i)_{i\in\mathbb{Z}^q}$ from the observation $\mathbf{X}$.

The present chapter contributes to this subject by proposing two novel Poisson denoising methods following a review of the state of the art (Section 5.1). Our first approach is based on biorthogonal Haar-domain hypothesis tests (Section 5.2). This approach is particularly suitable for estimating smooth intensities for large datasets in real time. Our second method makes use of well designed multi-scale variance stabilizing transforms (VSTs). It can be combined with most multi-scale transforms to adapt different source morphologies in the restoration (Section 5.3).

We point out that in our formulation of the Poisson denoising problem, we have supposed a zero dark intensity, i.e., $\lambda_d = 0$ (see Chapter 4). The proposed denoisers can be easily adapted to the case of nonzero dark intensity as to be discussed in Section 5.4.

5.1 Existing Methods for Denoising Poisson Data

We summarize the existing intensity estimators into four categories, i.e., VST-based methods (Section 5.1.2), filters in the wavelet domain (Section 5.1.3), approaches using hypothesis tests (Section 5.1.4), and Bayesian and penalized maximum-likelihood (ML) estimators (Section 5.1.5). Many of these estimators are based on wavelet shrinkage, and we first provide a brief review of the wavelet background (Section 5.1.1). More details can be found, for example, in [62], [63], and [64].

5.1.1 Wavelet background

The wavelet transform represents a one-dimensional (1D) real-valued continuous-time signal $f(t), t \in \mathbb{R}$, in terms of shifts and dilations of a lowpass scaling function $\phi(t)$ and bandpass wavelet $\psi(t)$. Let us note $\phi_{j,n}(t) := 2^{-j/2}\phi(2^{-j}t - n)$, and $\psi_{j,n}(t) := 2^{-j/2}\psi(2^{-j}t - n)$. For special choices of these functions, the shifts and dilations $\{\phi_{J,n}, \psi_{j,m} | \, j \leq J, (n,m) \in \mathbb{Z}^2\}$ form an orthonormal basis of $L^2(\mathbb{R})$. Thereby, we have the signal representation

$$f(t) = \sum_n a_J[n]\phi_{J,n} + \sum_{j=-\infty}^{J} \sum_n d_j[n]\psi_{j,n} \tag{5.1}$$

$$a_J[n] = \langle f, \phi_{J,n} \rangle, \quad d_j[n] = \langle f, \psi_{j,n} \rangle \tag{5.2}$$

Here, $(a_j[n])_n$ and $(d_j[n])_n$ are respectively approximation coefficients (or scaling coefficients) and detail coefficients (or wavelet coefficients) at scale j; $\langle \cdot, \cdot \rangle$ denotes inner product. The above representation reveals the multiresolution feature of the wavelet analysis. The first sum in (5.1) is an approximation of f at scale (resolution) J, and the second double sum consists of refinements at finer and finer scales $j \leq J$. Unlike the infinitely supported Fourier basis, compactly supported wavelets with arbitrary number of vanishing moments[1] can be constructed [62]. Thereby, wavelets provide an analysis that adapts to the local signal variations.

The approximation and detail coefficients can be computed iteratively and efficiently. Using the fact that the scaling function and the wavelet are related by [64]

$$\phi(x/2) = \sqrt{2}\sum_n h[n]\phi(x-n), \quad \psi(x/2) = \sqrt{2}\sum_n g[n]\phi(x-n)$$

[1] A wavelet having p vanishing moments is orthogonal to any polynomial of degree $p-1$.

where $h[n]$ and $g[n]$ are discrete-time lowpass and bandpass filters, respectively. The approximation and detail coefficients can be recursed as

$$a_{j+1}[n] = \bar{h} \star a_j[2n] = \sum_k h[k-2n]a_j[k], \quad d_{j+1}[n] = \bar{g} \star a_j[2n] = \sum_k g[k-2n]a_j[k] \quad (5.3)$$

where $\bar{h}[n] := h[-n]$, and $\star$ denotes convolution. Let $\check{x}[n] = x[p]$ if $n = 2p$ and 0 otherwise, which is an up-sampling scheme. The reconstruction of the approximation coefficients can be obtained from those of the coarser resolutions

$$a_j[n] = \sum_k \check{a}_{j+1}[k]\tilde{h}[n-k] + \sum_k \check{d}_{j+1}[k]\tilde{g}[n-k] \quad (5.4)$$

where $\tilde{h} = h$ and $\tilde{g} = g$ for orthonormal wavelets [64]. (5.3) and (5.4) can be summarized into a discrete-time filter bank comprising lowpass and bandpass filters, decimators and upsamplers (Fig. 5.1). For a discrete signal $f[n]$, the recursions (5.3) and (5.4) define

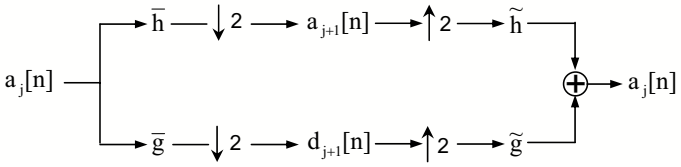


Fig. 5.1: Wavelet transform implemented by an iterative filter bank.

a discrete wavelet decomposition and inversion, respectively. The transform has a linear complexity $O(N)$ where N is the signal length. For two-dimensional (2D) image $f[n, m]$, the discrete wavelet transform is implemented by a 2D filter bank. If a separable wavelet basis is used, the filters h and g will also be separable. Then, the transform is obtained by alternating the application of the h and g filters on rows and columns of the image. The transform for higher dimensional signals can be likewise achieved.

It is important in many applications such as pattern recognition to have a translation-invariant representation of the signal. When a pattern is translated, its numerical descriptors should be translated but not modified. However, it is easy to verify that the wavelet representation (5.2) is not translation-invariant. The translation invariance can be regained by computing the approximation and detail coefficients over the integer grid

$$a_j[n] = \langle f(t), \phi_{j,n} \rangle, \quad \phi_{j,n} = 2^{-j/2}\phi(2^{-j}(t-n)) \quad (5.5)$$

$$d_j[n] = \langle f(t), \psi_{j,n} \rangle, \quad \psi_{j,n} = 2^{-j/2}\psi(2^{-j}(t-n)) \quad (5.6)$$

Practical computation of a_j and d_j can also be achieved by a recursion scheme, or "à trous" algorithm [65][66].

$$a_{j+1}[l] = \bar{h}^{\uparrow j} \star a_j[l] = \sum_k h[k] a_j[l + 2^j k], \quad d_{j+1}[l] = \bar{g}^{\uparrow j} \star a_j[l] = \sum_k g[k] a_j[l + 2^j k] \quad (5.7)$$

where $h^{\uparrow j}[l] = h[l]$ if $l/2^j \in \mathbb{Z}$ and 0 otherwise. The reconstruction is given by: $a_j[l] = \frac{1}{2}\left[(\tilde{h}^{\uparrow j} \star a_{j+1})[l] + (\tilde{g}^{\uparrow j} \star d_{j+1})[l]\right]$. We summarize the à trous algorithm in Fig. 5.2. Comparing

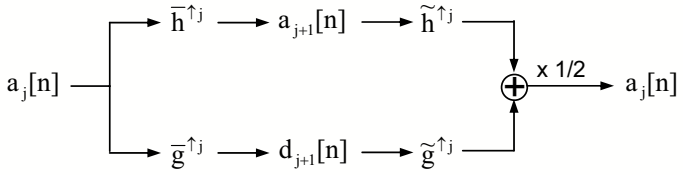


Fig. 5.2: Undecimated wavelet transform implemented by an iterative filter bank.

to Fig. 5.1, the decimators and the upsamplers are absent. Thus, this schema is also termed as undecimated wavelet transform (UWT). This transform has a complexity $O(N \log(N))$ which is only slightly larger than the decimated version.

The wavelet transform of real-world signals tends to be very sparse, with a few large scaling and wavelet coefficients dominating the representation, while most of the wavelet coefficients are almost zero. Formally speaking, if we approximate a signal by keeping the M vectors in a wavelet basis having the largest amplitudes of their coefficients, then the ℓ^2-norm of the approximation error can be written as a function decaying with M, i.e., $O(M^{-\alpha})$ where $\alpha > 0$. The larger is α, the higher is the decaying rate, and hence the sparser is the signal representation in the wavelet domain. It is shown that for signals of bounded variations wavelets offer the highest decay rate that cannot be improved by either optimal spline approximation or any approximation calculated in an orthonormal basis [64][67]. Bounded variation functions include those that are piecewise regular with a finite number of point discontinuities. This property makes the wavelet transform a very efficient representation for many practical signals. However, we should point out that data of a dimension higher than 1 can have more complex singularities other than the point-like discontinuities (such as line-like and curvilinear boundaries in 2D). For such data, wavelet system is no longer optimal, and other more efficient multiscale transforms have been developed (see Sections 5.1.5 and 5.3).

This sparsity property is also advantageous for signal recovery applications. Indeed, the wavelet transform will tend to concentrate the energy of the desired signal into a small number of large-amplitude (high signal-to-noise ratio (SNR)) coefficients. The other coefficients will have small amplitudes (low SNR) that are mainly due to noise. Thus, a simple coefficient-wise thresholding that removes insignificant coefficients and retains the significant ones will lead to noise removal. A general thresholding configuration can be written as

$$\tilde{d}_j[n] = h \cdot d_j[n]$$

where $0 \leq h \leq 1$. A zero-valued factor h completely removes the contribution of the basis function $\psi_{j,n}$; setting $h = 1$ leaves it intact. Letting $0 < h < 1$ shrinks the contribution of $\psi_{j,n}$ accordingly. Widely used thresholding functions include hard-thresholding

$$h = \begin{cases} 1 & |d_j[n]| \geq T \\ 0 & |d_j[n]| < T \end{cases}$$

and soft thresholding

$$h = \begin{cases} 1 - \frac{T \cdot \mathrm{sgn}(d_j[n])}{d_j[n]} & |d_j[n]| \geq T \\ 0 & |d_j[n]| < T \end{cases}$$

where T is an user-specified threshold value which is generally proportional to the noise level. In the Gaussian white noise case, with an appropriate choice of the threshold level, the soft thresholding procedure is near-minimax optimal over the bounded variation space (more generally a wide interval of the Besov scale) [68].

Denoising using the standard decimated wavelet transform sometimes exhibits visual artifacts due to the lack of translation invariance. Translation-invariant denoising can be achieved by the procedure of cycle-spinning [51]. The basic idea is to "average out" the translation dependence. For a range of cyclic shifts, one shifts the data, denoises the shifted data, and then unshifts the denoised data. Doing this for each of a range of shifts, and averaging the results so obtained, produces a reconstruction with far weaker artifacts. Cycle-spinning over the range of all cyclic shifts leads to exact translation-invariant estimation, which is equivalent to thresholding the UWT coefficients.

5.1.2 VST-based denoisers

Given Poisson data $\mathbf{X}$, each sample $X_i \sim \mathcal{P}(\lambda_i)$ has a variance $\mathrm{Var}[X_i] = \lambda_i$. Thus, the variance of $\mathbf{X}$ is signal-dependent. The aim of a VST T is to stabilize the data such that

each coefficient of $T(\mathbf{X})$ has an (asymptotically) constant variance, say 1, irrespective of the value of λ_i. In addition, for the VSTs to be presented below, $T(\mathbf{X})$ is asymptotically normally distributed. Thus, the VST-transformed data are approximately stationary and Gaussian. Once we are brought to the Gaussian denoising problem, any standard method such as wavelet shrinkage [69, 50][64] can be applied to obtain the denoised coefficients $\widehat{T(\mathbf{X})}$. Finally, the intensity is estimated by inverting the VST, i.e., $\hat{\lambda} = T^{-1}(\widehat{T(\mathbf{X})})$.

Anscombe transform [70][71] is a widely used VST which has a simple square-root form

$$T(X) := 2\sqrt{X + 3/8} \tag{5.8}$$

We can show that $T(X)$ is asymptotically normal as the intensity increases.

$$T(x) - 2\sqrt{\lambda} \underset{\lambda \to +\infty}{\xrightarrow{\mathcal{D}}} \mathcal{N}(0, 1) \tag{5.9}$$

It can be shown that Anscombe transform requires a high underlying intensity to stabilize the data correctly (typically for $\lambda \gtrapprox 10$, see also Section 5.3.1.3).

[72] proposed another VST based on the Fisz transform in the Haar wavelet domain (see Section 5.1.1 for a general wavelet review). It consists of the following steps.

1. Apply the Haar wavelet transform of $\mathbf{X}$ and modify the detail coefficients by

$$\tilde{d}_j[k] := \begin{cases} 0 & \text{if } a_j[k] = 0 \\ d_j[k]/\sqrt{a_j[k]} & \text{otherwise} \end{cases} \tag{5.10}$$

2. Apply the inverse Haar transform to the modified coefficients $(a_J, d_J, d_{J-1}, \ldots, d_1)$ to produce $\tilde{\mathbf{X}}$.

Here, $a_j[k]$ and $d_j[k]$ are respectively the Haar approximation and detail coefficients at scale j and location k. For Haar transform, $a_j[k]$ and $d_j[k]$ can be respectively written as a sum and a difference of two independent Poisson variables U and V, i.e., $a_j[k] = U + V$ and $d_j[k] = U - V$. Asymptotic Gaussianity and variance stabilization for the form $(U-V)/(U+V)^p$ where U and V are nonnegative, independent random variables, and $p > 0$ were studied by Fisz [73]. Thus (5.10) can be seen as a multi-scale Fisz transform ($p = 1/2$) applied in the Haar domain, and the entire transform defined by the above procedure $\tilde{\mathbf{X}} := T(\mathbf{X})$ is termed as a Haar-Fisz transform [72]. [72] shows that the transformed data $\tilde{\mathbf{X}}$ are asymptotically uncorrelated and normally distributed. This VST adapts to the data with lower photon rates than for Anscombe (typically $\lambda \gtrapprox 1$, see Section 5.3.1.3).

[74] extended the Haar-Fisz transform for any wavelet basis. A wavelet coefficient $d_j[k]$ can be written as a linear combination of the input data $d_j[k] = \sum_{l \in \mathcal{L}_{j,k}} g_{j,k}[l] X_l$. Here, $g_{j,k}$ is the band-pass filter at scale j and location k, and $\mathcal{L}_{j,k}$ is the index set where $g_{j,k}[l] \neq 0$ for $l \in \mathcal{L}_{j,k}$. We further denote $N_j[k] := \sum_{l \in \mathcal{L}_{j,k}} X_l$. Then, the wavelet coefficients will be normalized to be

$$\tilde{d}_j[k] := \begin{cases} d_j[k]/\sqrt{N_j[k]} & \text{if } N_j[k] \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

[74] shows that $\mathrm{Var}\left[\tilde{d}_j[k] | N_{j,k} \neq 0\right] = \|g_{j,k}\|_2^2 / L_{j,k}$ where $L_{j,k} = |\mathcal{L}_{j,k}|$. This stabilization approach is called conditional variance stabilization (CVS). If Haar basis is used, the CVS becomes the multi-scale Fisz transform. We should emphasis a delicate difference between the Haar-Fisz VST and the CVS. In the former method, the forward transform involves the inversion of the Haar transform. In other words, the Gaussinized data are obtained from a complete Haar decomposition and reconstruction. In contrast, once CVS produces the stabilized wavelet coefficients, they are immediately thresholded based on the normal statistics. These thresholded coefficients are then re-multiplied by the normalization factor $\sqrt{N_j[k]}$ before reconstructing the final intensity estimate. In summary, the stabilization and the coefficient estimation are separated in Haar-Fisz approach, whereas they are coupled together in the CVS framework.

5.1.3 Wavelet-domain filters

Nowak and Baraniuk [75], and Antoniadis and Sapatinas [76] proposed a wavelet domain shrinkage operator, which can be deemed as an approximative oracle attenuation under a wavelet basis. We write $d_j[k] = \sum_i g_{j,k}[i] X_i$. Let

$$\widehat{\sigma_{j,k}^2} = \sum_i g_{j,k}[i]^2 X_i$$

be an unbiased estimate of the noise power at the wavelet coefficient $d_j[k]$, and

$$\widehat{d_j[k]^2} = d_j[k]^2 - \widehat{\sigma_{j,k}^2}$$

be an unbiased estimate of the signal power at $d_j[k]$. Then, the shrinkage filter has the form

$$h = \max\left(\frac{\widehat{d_j[k]^2}}{\widehat{d_j[k]^2} + \widehat{\sigma_{j,k}^2}}, 0 \right) \tag{5.12}$$

This filter weights each noisy coefficient $d_j[k]$ by a factor that approaches 1 for high SNR and to 0 for low SNR. Hence, this filter can be interpreted as a data-adaptive wavelet-domain Wiener filter. As the total counts increase, this filter is asymptotically optimal in the mean-square error sense.

5.1.4 Hypothesis tests

The wavelet thresholding can be also implemented by applying a binary hypothesis test on each wavelet coefficient $d_j[k]$ [77]:

$$H_0 : d_j[k] = 0 \text{ vs. } H_1 : d_j[k] \neq 0$$

The coefficients of the null hypothesis are insignificant coefficients and those of the alternative hypothesis are significant. Note that since any wavelet has a zero mean, if $d_j[k]$ comes from a signal having a constant intensity within the wavelet support, then $d_j[k]$ would be zero if no noise were present. Thus, $d_j[k] \in H_0$ in this case.

Individual hypothesis tests are carried out in a coefficient-by-coefficient manner. First, the user pre-specifies a false positive rate (FPR) in the wavelet domain, say α. Then the p-value of each coefficient p_i is calculated under the null hypothesis H_0. Finally, all the coefficients with $p_i > \alpha$ will be zeroed.

If we desire to control global statistical error rates, multiple hypothesis tests should be used. For example, the Bonferroni over-conservative correction controls the probability of erroneously rejecting even one of the true null hypothesis, i.e., Family-Wise Error Rate (FWER). To upper bound FWER by α, the same individual-testing procedure is performed as above but with FPR set to α/N, where N is the number of wavelet coefficients. Alternatively, one can carry out the Benjamini and Hochberg procedure [78] to control the false discovery rate (FDR), i.e., the average fraction of false detections over the total number of detections

$$\text{FDR} := \mathbb{E}\left[\frac{|\text{FP}|}{|\text{FP}| + |\text{TP}|}\right] \tag{5.13}$$

where $|\text{FP}|$ and $|\text{TP}|$ are the number of false positives and that of true positives, respectively. The control of FDR has the following advantages over that of FWER: 1) it usually has a greater detection power; 2) it can easily handle correlated data [79]. The latter point allows the FDR control in non-orthogonal wavelet domains. Minimaxity of FDR has also been studied in various cases (see [80][81] for details).

For any hypothesis testing procedure, we need to know the distribution of the wavelet coefficients $d_j[k]$ under the null hypothesis to calculate the p-values. [82] shows the probability density function (*pdf*) of $d_j[k]$ to be

$$\begin{aligned} p(d_j[k] = w; \lambda_s) &= \sum_{n \geq 0} \star^{(n)} H_\psi(w) \frac{(\lambda_s)^n}{n!} e^{-\lambda_s} \\ \star^{(n)} H_\psi &= \begin{cases} \delta & n = 0 \\ \underbrace{H_\psi \star \cdots \star H_\psi}_{n \text{ times}} & n > 0 \end{cases} \end{aligned} \tag{5.14}$$

where H_ψ is the normalized histogram of ψ, and δ is the Dirac disrtibution. We suppose a constant intensity within the support of $\psi_{j,k}$, and λ_s is the total intensity in that support. For a general wavelet, the *pdf* has no closed form, and is therefore computationally complex in practice. To obtain distributions of manageable forms, simple wavelets are preferred, such as Haar. To the best of our knowledge, Haar is the only wavelet yielding a closed-form *pdf*, which is given by [83] ($n \geq 0$),

$$p(d_j[k] = X_1 - X_2 = n) = e^{-2\lambda} I_n(2\lambda), \quad X_{1,2} \sim \mathcal{P}(\lambda) \tag{5.15}$$

where $I_n(z)$ is the first kind modified Bessel function of order n. For negative n, the probability can be obtained by symmetry. The tail probability (i.e. p-value) is given by [84],

$$\Pr(d \geq n; \lambda) = \Pr\left(\chi^2_{(2n)}(2\lambda) < 2\lambda\right), \quad n \geq 1 \tag{5.16}$$

where $\chi^2_{(f)}(\Delta)$ is the non-central chi-square distribution with f degrees of freedom and Δ as non-centrality parameter. The individual hypothesis tests in the Haar domain can be equivalently implemented by thresholding operators. The details can be found in Section 5.2.

Kolaczyk [85] also attempted to generalize the hypothesis tests for wavelets other than Haar. The implementing thresholds are "corrected" versions of the usual Gaussian-based thresholds for the Poisson case. However, the asymptotic approximation he adopts may not allow reasonable solutions in low-count situations.

5.1.5 Bayesian and penalized ML estimators

Poisson intensity estimation has also been formulated in a Bayesian framework. Special attention has been focused on the multiscale analysis provided by the Haar transform,

$$\begin{aligned} a_0[k] &= X_k, \quad k = 0, 1, \ldots, 2^J - 1 \\ a_j[k] &= a_{j-1}[2k] + a_{j-1}[2k+1], \quad k = 0, 1, \ldots, 2^{J-j} - 1, j = 1, 2, \ldots, J \end{aligned}$$

The parameter $a_j[k]$'s are the Haar approximation coefficients. Let us define $\lambda_j[k]$ as the approximation coefficients of the underlying intensity Λ by applying the same transform. The relationship between a parent ($a_j[k]$) and a child ($a_{j-1}[2k]$) is of fundamental interest in multiscale analysis. This relationship is expressed by the conditional likelihood $p(a_{j-1}[2k]|a_j[k], \Lambda)$ which happens to follow a simple binomial distribution

$$p(a_{j-1}[2k]|a_j[k], \Lambda) = \text{Binomial}(a_{j-1}[2k]|\, a_j[k], \theta_j[k])$$

where $\theta_j[k] := \lambda_{j-1}[2k]/\lambda_j[k]$ which can be deemed as splitting factors that govern the inter-scale intensity assignment. This leads to the following factorization of the likelihood function

$$p(\mathbf{X}|\Lambda) = p(a_J[0]|\lambda_J) \prod_{j=1}^{J} \prod_{k=0}^{2^{J-j}-1} p(a_{j-1}[2k]|a_j[k], \theta_j[k])$$

where $\lambda_J = \sum_i \lambda_i$ is the total intensity. Now, Bayesian method consists of specifying a prior distribution on $\theta_j[k]$ and then derive the posterior distribution from the likelihood and the prior.

Kolaczyk [86] proposed the prior on the multiscale parameters $\theta_j[k]$ to be a mixture of a point mass at 1/2 and a symmetric beta distribution, i.e.,

$$\begin{aligned} \theta_j[k] &\sim \gamma_j[k]\frac{1}{2} + (1 - \gamma_j[k])B_j[k], \\ \gamma_j[k] &\sim \text{Bernoullio}(p_j), \\ B_j[k] &\sim \text{Beta}(A_j, A_j) \end{aligned}$$

where $0 \leq p_j \leq 1$ and $A_j > 0$ are hyperparameters. Then the final posterior mean of the intensity can be expressed in a recursive manner across the scales.

[87] moves beyond [86] by introducing a more general beta mixture for $\theta_j[k]$.

$$p(\theta_j[k]) = \sum_{i=1}^{M} p_i \frac{(\theta_j[k](1 - \theta_j[k]))^{s_i - 1}}{2B(s_i, s_i)}$$

where $B(\cdot,\cdot)$ is the Euler beta function, $s_i \geq 1$ are hyperparameters, $(p_i)_i$ are mixture probabilities and $\sum_{i=1}^{M} p_i = 1$. The fact that the beta family is conjugate to the likelihood function (binomial) leads to a posterior mean with a closed-form expression.

[88] considered the indirectly observed Poisson data, i.e., $\Lambda = P\mu$ where P is a known matrix of weight and μ is the intensity to estimate. The same Haar-based multiscale framework is used and an EM algorithm is introduced to derive the maximum a posterior (MAP) intensity estimate. Recently, [74] proposed to condition the total counts in the wavelet support when computing a given coefficient. This concept allows to generalize the Haar-based Bayesian framework to arbitrary wavelet. The low-intensity case apart, Bayesian approaches generally outperform the direct wavelet filtering [89].

Poisson denoising has also been formulated as a penalized maximum likelihood (ML) estimation problem [90][91][92][93] within wavelet, wedgelet and platelet dictionaries. Unlike wavelet-based estimators where the image is recursively decomposed into dyadic square partitions, the wedgelet partition [94] is based on a recursive dyadic square partition of the image in which the final nodes are allowed to terminate with a wedge instead of a square. The wedge split is defined by a line connecting two points on two sides of the square. Then, each cell of the partition is approximated by a constant. Platelet generalizes the wedgelet transform in that it fits each cell with a planar surface. That is, a platelet is a function of 3 parameters [95]

$$f_S(x,y) = (A_S x + B_S y + C_S) I_S(x,y)$$

where A_S, B_S, $C_S \in \mathbb{R}$, S is a dyadic square or wedge at a terminal node of an recursive partition. Clearly, platelets are localized atoms at various locations, scales and orientations that can produce accurate piecewise linear approximations to images consisting of smooth regions separated by smooth boundaries.

The penalized maximum-likelihood estimation with platelets is formulated as follows [93].

$$\hat{\Lambda} = \arg\min_{\Lambda \in \Gamma_P} -\log p(\mathbf{X}|\Lambda) + 2 \cdot \text{penalty}(\Lambda) \tag{5.17}$$

$$\text{penalty}(\Lambda) := \gamma \cdot \log n \cdot |\{\text{PlateletCoefficients}(\Lambda)\}| \tag{5.18}$$

where the penalty is proportional to the number of platelet coefficients associated with the estimate Λ. n is the total number of counts, and γ is a tunning parameter controlling the balance between the likelihood term and the penalty. Γ_P is the family of solutions consisting

of intensities corresponding to wedgelet partitions with platelet models fit to the end nodes.

This penalized MLE platelet framework produces near-minimax optimal intensity estimates for images composed of smooth surfaces (Hölder-α, $\alpha \in (1, 2]$) separated by smooth boundaries (Hölder-β, $\beta \in (1, 2]$) [92]. The platelet approximation for the intensities in this class outperforms those using Fourier basis, wavelets, or wedgelets. Indeed, Fourier does not perform well due to the singularities; wavelets can handle point-like singularities but are seriously challenged by boundaries in 2D; wedgelets can handle the boundaries but provide a piecewise constant approximations which perform poorly for nonconstant smooth intensity regions.

Ridgelet [52] and curvelet [96] systems are near optimal in representing line-like and twice-differentiable curvilinear singularities, respectively. To our knowledge, no Poisson denoising method has been proposed for the ridgelet and curvelet transforms.

5.2 Denoising by Hypothesis Tests in the Biorthogonal Haar Domain

The Haar-based methods reviewed in Section 5.1.4 apply hypothesis tests on Haar coefficients to control a user-specified FPR. When working with very large datasets or real-time applications, the decimated Haar transform is generally required to meet limited-memory or computation-time constraints. Unfortunately, for regular underlying intensities, decimation yields discontinuous estimates with strong "staircase" artifacts, thus significantly degrading the denoising performance.

In this section, we propose to combine the hypothesis testing framework with the decimated biorthogonal Haar (Bi-Haar) transform. The Bi-Haar filter bank is normalized such that the p-values of Bi-Haar coefficients (p_{BH}) approximate those of Haar (p_H) for high-intensity settings or large scales; for low-intensity settings and small scales, we show that p_{BH} are essentially upper-bounded by p_H. Thus, we may apply the Haar-based hypothesis tests to Bi-Haar coefficients to control a prefixed FPR. By doing so, we benefit from the regular Bi-Haar filter bank to gain a smooth estimate. A Fisher-approximation-based (FAB) threshold implementing the hypothesis tests is also established. We find that this approach even exhibits a performance comparable to the more time/space-consuming translation-invariant Haar (TI Haar or undecimated Haar) denoising in some of our experiments.

This section is organized as follows. The Bi-Haar domain tests are presented in Section 5.2.1. Section 5.2.2 details some thresholding operators implementing the tests. The final denoising algorithm is summarized in Section 5.2.3, and the numerical results are shown in Section 5.2.4.

5.2.1 Bi-Haar domain testing

Haar wavelet provides us with a manageable distribution under H_0. But due to the lack of continuity of Haar filters, its estimate can be highly irregular with strong "staircase" artifacts when decimation is involved.

To solve this dilemma between distribution manageability and reconstruction regularity, we propose to use the Bi-Haar wavelet. The implementation filter bank is given by [97]:

$$\begin{aligned} h &= 2^{-c}[1,1], & g &= 2^{-c}r[\tfrac{1}{8},\tfrac{1}{8},-1,1,-\tfrac{1}{8},-\tfrac{1}{8}]; \\ \tilde{h} &= 2^{c-1}r[-\tfrac{1}{8},\tfrac{1}{8},1,1,\tfrac{1}{8},-\tfrac{1}{8}], & \tilde{g} &= 2^{c-1}[1,-1] \end{aligned}$$

where c and $r = (1+2^{-5})^{-1/2}$ are normalizing factors, (h, g) and $(\tilde{h}, \tilde{g})$ are respectively the analysis and synthesis filter banks. Note that our Bi-Haar filter bank has an unusual normalization. The motivation behind this is to ensure that the Bi-Haar coefficients will have the same variance as the Haar ones at each scale. Let us also point out that to correct for the introduction of the factor r, the Bi-Haar coefficients must be multiplied by r^{-1} at each stage of the recursive reconstruction. For comparison, the Haar filter bank is ($h = 2^{-c}[1,1]$, $g = 2^{-c}[-1,1]$, $\tilde{h} = 2^{c-1}[1,1]$, $\tilde{g} = 2^{c-1}[1,-1]$). It follows that the synthesis Haar scaling function is discontinuous while that of Bi-Haar is almost Lipschitz [98][99]. Hence, the Bi-Haar reconstruction will be smoother.

At scale $j \geq 1$, let us define $\lambda_j = 2^j\lambda$ where λ is the underlying constant intensity. Then, a Haar coefficient can be written as $d_j^h = 2^{-cj}(X_1 - X_2)$ where $X_1, X_2 \sim \mathcal{P}(\lambda_j/2)$ are independent. We note $p_H := \Pr(d_j^h \geq 2^{-cj}k_0|H_0)$ to be the p-value of a Haar coefficient where $k_0 = 1, 2, \cdots$. Accordingly, a Bi-Haar coefficient can be written as $d_j^{bh} = 2^{-cj}r(X_3 - X_4 + \frac{1}{8}(X_1 - X_2))$, where $X_1, X_2 \sim \mathcal{P}(\lambda_j)$ and $X_3, X_4 \sim \mathcal{P}(\lambda_j/2)$ are all independent. We note $p_{BH} := \Pr(d_j^{bh} \geq 2^{-cj}k_0|H_0)$ to be the p-value of a Bi-Haar coefficient at the same critical threshold as for p_H. These definitions can be extended to higher dimensions ($q > 1$) straightforwardly.

For high-intensity settings or for large scales, d_j^h and d_j^{bh} will be asymptotically

normal with the same asymptotic variances $\sigma_h^2 = \sigma_{bh}^2 = 2^{qj(1-2c)}\lambda$ due to the normalized filter banks. Thereby, they will have asymptotically equivalent tail probabilities, i.e., $p_{BH} \approx p_H$.

For low intensity settings ($\lambda \ll 1$) and small scales, the following proposition (proof in Appendix A.8) shows for 1D signals that p_{BH} is essentially upper-bounded by p_H under H_0. The bounds for multidimensional data ($q > 1$) are also studied in Appendix A.8.

Proposition 5 *We have the following upper-bound for 1D signals*

$$p_{BH} \leq p_H + A(\lambda_j)(1 - 2p_H) \tag{5.19}$$

where

$$A(\lambda_j) = \frac{1}{2}\left[1 - e^{-2\lambda_j}\left(I_0(2\lambda_j) + 2\sum_{m=1}^{8} I_m(2\lambda_j)\right)\right]$$

As $\lambda \to 0+$, $A(\lambda_j) = \frac{2^{9j-7}}{2835}\lambda^9 + o(\lambda^9)$.

This theoretical bound is clearly confirmed by the numerical simulations shown in Table 5.1. Here we show the results for $\lambda_j \in [10^{-1}, 10^2]$ and different critical thresholds k_0 at the tails of the distributions. We indeed observe that p_{BH} is always strictly smaller than p_H.

Tab. 5.1: p_H and p_{BH}

$\lambda_j = 10^{-1}$		$\lambda_j = 10^0$		$\lambda_j = 10^1$		$\lambda_j = 10^2$	
$k_0 = 2$	$(1.15 \times 10^{-3},$ $1.17 \times 10^{-4})$	$k_0 = 4$	$(1.12 \times 10^{-3},$ $4.57 \times 10^{-4})$	$k_0 = 9$	$(3.97 \times 10^{-3},$ $2.48 \times 10^{-3})$	$k_0 = 20$	$(2.56 \times 10^{-2},$ $2.28 \times 10^{-2})$
$k_0 = 3$	$(1.91 \times 10^{-5},$ $1.87 \times 10^{-6})$	$k_0 = 5$	$(1.09 \times 10^{-4},$ $4.34 \times 10^{-5})$	$k_0 = 12$	$(2.12 \times 10^{-4},$ $1.26 \times 10^{-4})$	$k_0 = 30$	$(1.62 \times 10^{-3},$ $1.39 \times 10^{-3})$
$k_0 = 4$	$(2.38 \times 10^{-7},$ $2.28 \times 10^{-8})$	$k_0 = 6$	$(8.90 \times 10^{-6},$ $3.49 \times 10^{-6})$	$k_0 = 15$	$(6.60 \times 10^{-6},$ $3.78 \times 10^{-6})$	$k_0 = 40$	$(4.22 \times 10^{-5},$ $3.52 \times 10^{-5})$

Every parenthesis shows (p_H, p_{BH}) for 1D signals, where we always observe that $p_{BH} < p_H$.

5.2.2 Thresholds for individual tests

For individual tests controlling FPR, the HTs can be implemented by thresholding operators. In other words, one can find $\tilde{t}_j$ such that $\Pr(|d_j^{bh}| \geq \tilde{t}_j | H_0) \leq \alpha$ where α represents the controlled FPR. Now consider the Haar case and suppose that we have derived the Haar threshold t_j under the controlled FPR. Then, by setting $\tilde{t}_j := 2^{-cjq}\lceil 2^{cjq} t_j \rceil$ the results in Section 5.2.1 allow us to conclude that the FPR for a Bi-Haar test will always be upper-bounded by α. We point out that to simplify the presentation, t_j and $\tilde{t}_j$ are supposed to

be scale-dependent only, but scale *and* location-dependent thresholds can be derived using the same procedure presented below.

Section 5.2.2.1 reviews a Haar-domain threshold introduced by Kolaczyk [100], and then a more sensitive threshold is derived in Section 5.2.2.2.

5.2.2.1 CLTB threshold [101, 85, 100, 102]

The Haar coefficient can be written as $d_j^h = 2^{-cjq}(X_1 - X_2)$ where $X_1, X_2 \sim \mathcal{P}(\lambda_j/2)$ are independent. It follows from (5.16) that:

$$\Pr(d_j^h \geq t_j | H_0) = \Pr\left(\chi^2_{(2m_j)}(\lambda_j) < \lambda_j\right) \approx \Pr(\gamma\chi^2_{(f)} < \lambda_j) \tag{5.20}$$

$$\approx \Pr\left(Z > \frac{f - \lambda_j/\gamma}{\sqrt{2f}}\right) \tag{5.21}$$

where $m_j = 2^{cjq}t_j$, $\gamma = (2m_j + 2\lambda_j)/(2m_j + \lambda_j)$, $f = (2m_j + \lambda_j)^2/(2m_j + 2\lambda_j)$, $\chi^2_{(v)}$ is a central chi-square variable and $Z \sim \mathcal{N}(0,1)$. Here, two stages of approximation are used: 1) the non-central chi-square distribution is first approximated by a central one (5.20) [103]; 2) the central chi-square variable is then approximated by a normal one (5.21) using the central limit theorem (CLT). t_j is thus called the CLT-based (CLTB) threshold. Consequently, it remains to solve the equation (5.21) $= \alpha/2$, and the solution is given by:

$$t_j = 2^{-cjq-1}\left(z^2_{\alpha/2} + \sqrt{z^4_{\alpha/2} + 4 \cdot \lambda_j z^2_{\alpha/2}}\right) \tag{5.22}$$

where $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, and Φ is the standard normal *cdf*. Universal threshold can also be obtained by setting $z_{\alpha/2} = \sqrt{2 \ln N_j}$ in (5.22) where N_j is the total number of coefficients in one band at scale j.

5.2.2.2 FAB threshold

An improvement of CLTB threshold can be achieved by replacing (5.21) with an approximation of faster convergence, e.g., the following one proposed by Fisher [104]:

$$\sqrt{2\chi^2_{(f)}} \to \mathcal{N}(\sqrt{2f - 1}, 1), \quad f \to \infty \tag{5.23}$$

Therefore, (5.21) changes to:

$$\Pr\left(\gamma\chi^2_{(f)} < \lambda_j\right) \approx \Pr\left(Z > \sqrt{2f - 1} - \sqrt{\frac{2\lambda_j}{\gamma}}\right) \tag{5.24}$$

Let us denote:

$$G(m_j) := \sqrt{2f-1} - \sqrt{\frac{2\lambda_j}{\gamma}} = \sqrt{\frac{(2m_j+\lambda_j)^2}{m_j+\lambda_j} - 1} - \sqrt{\frac{\lambda_j(2m_j+\lambda_j)}{m_j+\lambda_j}} \tag{5.25}$$

It remains to solve $G(m_j) = z_{\alpha/2}$, which leads to a quartic equation in m_j:

$$\begin{aligned} 16m_j^4 + \left[16\lambda_j - 8(z_{\alpha/2}^2+1)\right] m_j^3 + \left[(z_{\alpha/2}^2+1)^2 - (20z_{\alpha/2}^2+12)\lambda_j + 4\lambda_j^2\right] m_j^2 \\ + \left[2(z_{\alpha/2}^2+1)^2\lambda_j - 16z_{\alpha/2}^2\lambda_j^2 - 4\lambda_j^2\right] m_j + (z_{\alpha/2}^2+1)^2\lambda_j^2 - 4z_{\alpha/2}^2\lambda_j^3 = 0 \end{aligned} \tag{5.26}$$

The final Fisher-approximation-based (FAB) threshold t_j is obtained from m_j^*, the solution of (5.26). Owing to the following results (proof in Appendix A.9), we do not need to write out the explicit expression of m_j^*, which could be rather complex:

Proposition 6 *The feasible condition for m_j is given by (5.27), and the feasible solution m_j^* exists and is unique.*

$$m_j \geq \frac{1}{8}\left[z_{\alpha/2}^2 - 2\lambda_j + 1 + \left(z_{\alpha/2}^4 + (12\lambda_j+2)z_{\alpha/2}^2 + 4\lambda_j^2 + 12\lambda_j + 1\right)^{1/2}\right] \tag{5.27}$$

Proposition 6 implies that we can use any numerical quartic-equation solver, e.g. Hacke's method [105], to find the four solutions of (5.26). One and only one of the solutions will satisfy (5.27), which is m_j^*. The universal threshold can also be derived in the same way as in the CLTB case.

Fig. 5.3 compare the values scaled by 2^j of the exact threshold, the CLTB threshold and the FAB threshold as functions of λ_j where we set $c = q = 1$. The scaling makes the threshold values independent of the scale j. But this does not bias the comparison as all the thresholds are scaled by the same factor. We can clearly see that for any intensities, the FAB threshold provides a much more accurate approximation to the exact one than for CLTB.

5.2.3 Summary of the denoising algorithm

Note that the thresholds t_j (more generally the p-value of d_j^h) depend on the background rate at scale j (i.e. λ_j). Without any prior knowledge, it can be estimated by the values of the approximation coefficients at scale $j+1$ (i.e. a_{j+1}). Here, the wavelet denoising should be carried out in a coarse-to-fine manner, outlined in Algorithm 1.

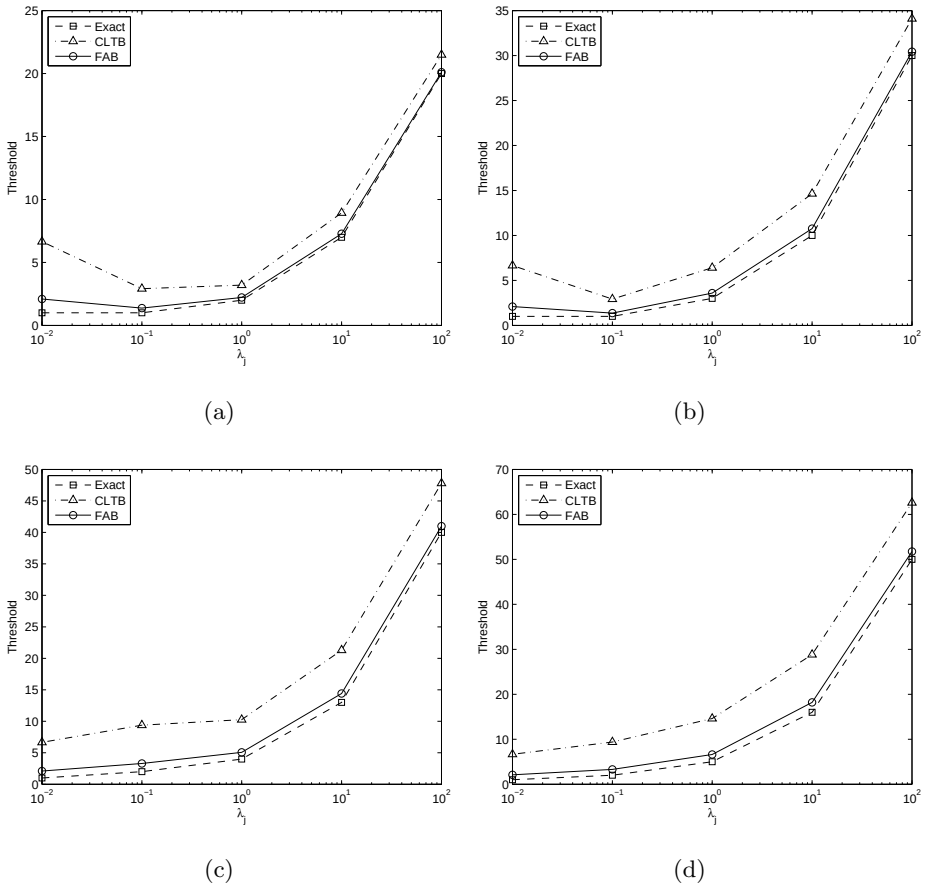


Fig. 5.3: Comparison of the exact threshold, the CLTB and FAB thresholds as functions of λ_j. The thresholds are scaled by 2^j. $c = q = 1$. The exact threshold implements a (a) 2σ thresholding level; (b) 3σ thresholding level; (c) 4σ thresholding level; (d) 5σ thresholding level.

Algorithm 1 Poisson noise removal by hypothesis tests in the Bi-Haar domain

1: Bi-Haar transform of $\mathbf{v}$ up to $j = J$ to obtain a_J and d_j^{bh} $(1 \le j \le J)$
2: **for** $j = J$ down to 1 **do**
3: $\quad \hat{\lambda}_j = 2^{jq}\lambda$ if λ is known; otherwise $\hat{\lambda}_j = \max(2^{cjq}a_j, 0)$
4: $\quad$ Testing d_j^{bh} by applying thresholds $\tilde{t}_j$ for a prefixed FPR $= \alpha$
5: $\quad$ Reconstruct a_{j-1} by inverse Bi-Haar transform
6: **end for**
7: Positivity projection: $\hat{\Lambda} = \max(a_0, 0)$

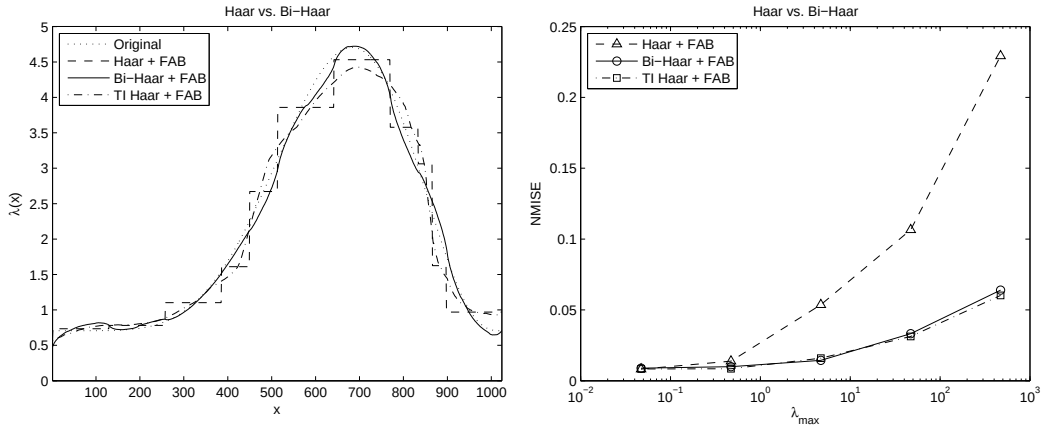
5.2.4 Results

5.2.4.1 Haar vs. Bi-Haar denoising for regular intensities

To compare Haar and Bi-Haar denoising for regular intensities, we generate noisy signals from the "Smooth" function [89] (see Fig. 5.4(a)) and measure the Normalized Mean Integrated Squared Error (NMISE) per bin from the denoised signals. The NMISE is defined as: NMISE $:= \mathbb{E}[(\sum_{i=1}^{N}(\hat{\lambda}_i - \lambda_i)^2/\lambda_i)/N]$, where $(\hat{\lambda}_i)_i$ is the intensity estimate. Note that the denominator λ_i plays the role of variance stabilization in the error measure.

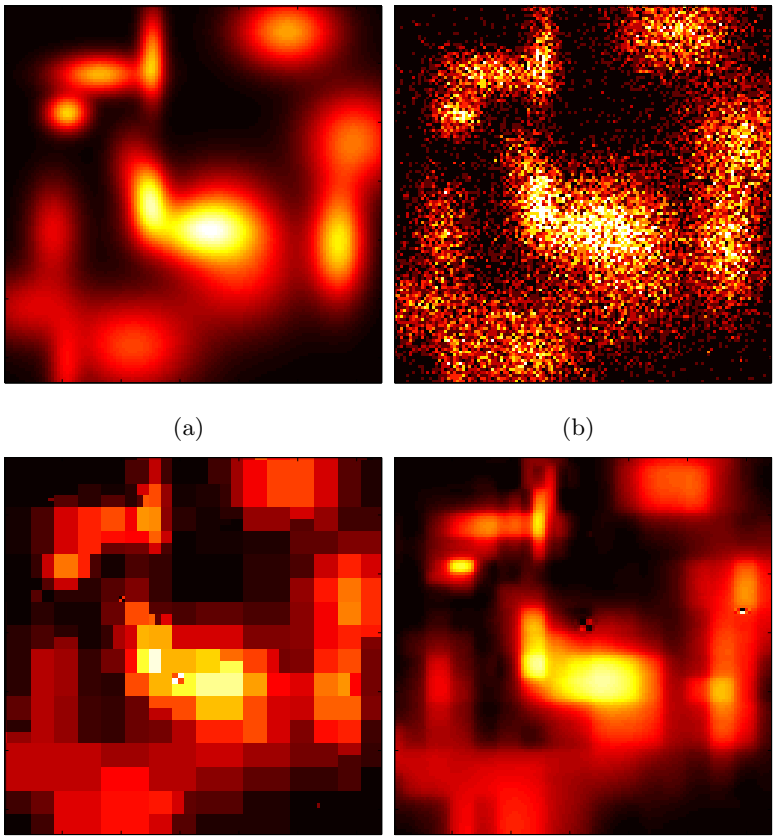
Fig. 5.4(a) shows the denoising examples given by Haar, Bi-Haar and TI Haar estimations, where FAB thresholds are applied. The original intensity function is scaled to cover a wide range of intensities, and Fig. 5.4(b) compares the NMISEs (measured from 100 replications) of the three estimators as functions of the underlying peak intensity.

It can be seen that the Bi-Haar estimate is much more regular than the Haar one, and is even almost as good as TI Haar at every intensity level under the NMISE criterion. This surprising performance is gained with the same complexity as in the Haar denoising, i.e., $O(N)$ only, as opposed to $O(N \log N)$ in the TI Haar case.



(a) (b)

Fig. 5.4: Denoising the "Smooth" function (length = 1024). Estimates from Haar, Bi-Haar and TI Haar (undecimated) are compared. $\alpha = 10^{-3}$ and $J = 7$. (a) denoising results; (b) NMISEs.



(a) (b)

(c) (d)

Fig. 5.5: Denoising a mixture of 15 simulated Gaussian sources. Image size 128×128. $\lambda \in [0.05, 7.11]$. FAB threshold, FPR $= 10^{-3}$ and $J = 4$. (a) original image; (b) Poisson counts; (c) Haar + FAB (NMISE $= 0.13$); (d) Bi-Haar + FAB (NMISE $= 0.069$).

5.2.4.2 Restoration of simulated images

We test our method on a 2D image where a mixture of 15 Gaussian sources of different variances are simulated (Fig. 5.5)(a). The peak intensity in the image is $\lambda_{max} = 7.06$. After adding a constant background ($\lambda_B = 0.05$) and a Poisson noise, we obtain the image of counts in Fig. 5.5(b). The restored images using Haar and Bi-Haar wavelets are respectively shown in Fig. 5.5(c) and Fig. 5.5(d). Here, the FAB thresholding is employed. The smoothness gained by the Bi-Haar wavelet can be clearly seen. Moreover, Bi-Haar estimation is more accurate (NMISE $= 6.9 \times 10^{-2}$) than that of Haar (NMISE $= 13.0 \times 10^{-2}$) where the NMISEs are measured from 100 replications.

5.3 Denoising by Multi-scale Variance Stabilizing Transform

The proposed Bi-Haar based estimation is fast, and produces regular estimates. However, for applications not requiring real-time processing, undecimated transforms are always preferred than the decimated one. In that case, the gain offered by using Bi-Haar rather than Haar wavelet is marginal. Another limitation of this approach is that it is designed for Bi-Haar wavelet only. For line-like or curvilinear sources, this wavelet filter is clearly unadapted. For q-dimensional ($q \geq 2$) isotropic sources, the biorthogonal wavelet transform (more generally for any standard separable wavelet transform) does not provide an isotropic analysis. Thereby, this method leads to a suboptimal estimation for those sources.

In this section, we propose another Poisson intensity estimator which is more flexible in adapting different source types and is also more powerful. It is based on a VST designed to stabilize the variance of a filtered discrete Poisson process, yielding a near Gaussian process. This new transform, which can be deemed as an extension of the Anscombe transform to filtered data, is simple, fast and efficient in (very) low-count situations. The rationale behind the benefits of stabilizing a filtered version of the original process is as follows. It is well known that the performance of the Anscombe VST deteriorates as the intensity becomes low [97] (typically for $\lambda < 10$), i.e., as the SNR decreases. Hence, one can alleviate this limitation and enhance the performance of the VST if the SNR is increased before stabilization. This can be achieved by pre-filtering the original process provided that the filter acts as an "averaging" kernel, and more generally, a low-pass filter. A detailed asymptotic analysis will support these claims.

By recognizing that a large family of multiscale transforms are computed from filtering equations (e.g. wavelets), the proposed VST can be seamlessly combined with their filter banks, leading to multiscale VSTs (MS-VSTs). Toward the goal of Poisson denoising, we are allowed to choose or design the most adaptive transform for the sources to be restored based on their morphology. Indeed, owing to recent advances in modern harmonic analysis, different multiscale transforms were shown to be very effective in sparsely representing different kinds of information. For example, to represent efficiently isotropic singularities and regular structures, a qualified candidate is the wavelet transform [64][97]. The ridgelet transform [52] has been shown to be very effective in representing global lines in an image. The curvelet system [96][53] is highly suitable for representing smooth ($\mathcal{C}^2$) images away from $\mathcal{C}^2$ contours. These transforms are also computationally attractive particularly

in large-scale applications. We will show that our new VST can be easily combined with these different multiscale geometrical decompositions, yielding asymptotically normally distributed coefficients with known variances. A classical hypothesis testing framework is then adopted to detect the significant coefficients, and a sparsity-driven iterative scheme is proposed to reconstruct the final estimate. We show that the MS-VST approach provides a very effective denoiser capable of recovering important structures of various (isotropic, line-like and curvilinear) shapes in (very) low-count images.

This section is organized as follows. In Section 5.3.1, a detailed analysis is provided to characterize the VST. Section 5.3.2 outlines the denoising setting using MS-VST with wavelets. Section 5.3.2.2 and 5.3.2.3 show how the VST can be combined with the isotropic undecimated wavelet transform and the standard separable undecimated wavelet transform, respectively. Denoising by MS-VST combined with ridgelets and curvelets are respectively presented in Section 5.3.3 and 5.3.4.

5.3.1 VST of a filtered Poisson process

Given a Poisson process $\mathbf{X} := (X_i)_i$ where X_i's are independent and $X_i \sim \mathcal{P}(\lambda_i)$, $Y_j := \sum_i h[i]X_{j-i}$ is the filtered process obtained by convolving $(X_i)_i$ with a discrete filter h. We will use Y to denote any one of Y_j. Let us define $\tau_k := \sum_i (h[i])^k$ for $k = 1, 2, \cdots$. In addition, we adopt a local homogeneity assumption that $\lambda_{j-i} = \lambda$ for all i within the support of h.

5.3.1.1 VST-heuristics

It can be seen that the variance of Y ($\mathrm{Var}\,[Y]$) is proportional to the intensity λ. To stabilize $\mathrm{Var}\,[Y]$, we seek a transformation $Z := T(Y)$ such that $\mathrm{Var}\,[Z]$ is (asymptotically) constant, say 1, irrespective of the value of λ.

Heuristically, the Taylor expansion gives us $T(Y) \approx T(\mu_Y) + T'(\mu_Y)(Y - \mu_Y)$, where $\mu_Y := \mathbb{E}\,[Y] = \lambda\tau_1$. We then have $\mathrm{Var}\,[Z] \approx T'(\mu_Y)^2 \cdot \mathrm{Var}\,[Y] = T'(\mu_Y)^2 \cdot \lambda\tau_2$. Hence, by setting $\mathrm{Var}\,[Z] = 1$, we obtain a differential equation $T'(\mu_Y) = {\mu_Y}^{-1/2}\sqrt{\tau_1/\tau_2}$, of which the solution is given by $T(Y) = 2\sqrt{\tau_1/\tau_2}\sqrt{Y}$. This implies that the square-root transform could serve as a VST. It is possible to use higher order Taylor expansions to find VST of different forms, but solving the associated differential equations is found difficult since they are highly non-linear.

5.3.1.2 VST-rigor

We define the square-root transform T as follows:

$$T(Y) := b \cdot \operatorname{sgn}(Y + c)|Y + c|^{1/2} \tag{5.28}$$

where b is a normalizing factor. Lemma 2 (proof in Appendix A.10) confirms our heuristics that T is indeed a VST for a filtered Poisson process (with a nonzero-mean filter) in that $T(Y)$ is asymptotically normally distributed with a stabilized variance as λ becomes large.

Lemma 2 (Square root as VST) *If $\tau_1 \neq 0$, $\|h\|_2, \|h\|_3 < \infty$, then we have:*

$$sgn(Y + c)\sqrt{|Y + c|} - sgn(\tau_1)\sqrt{|\tau_1|\lambda} \xrightarrow[\lambda \to +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{\tau_2}{4|\tau_1|}\right) \tag{5.29}$$

where $sgn(\cdot)$ is the sign function.

This result holds true for any $c \in \mathbb{R}$, of which the value controls the convergence rate in (5.29). The next section provides an analysis of the asymptotic rate and determines the optimal value of c.

5.3.1.3 Optimal parameter of the VST

To simplify the asymptotic analysis, we assume a non-negative filter h and a positive constant c (a non-positive h with a negative c can also be considered). Thus, our VST is simplified to $Z := T(Y) = b\sqrt{Y + c}$. We can now derive the asymptotic expansions of $\mathbb{E}[Z]$ and $\operatorname{Var}[Z]$ as stated in Proposition 7 (proof in Appendix A.11). Note that the last point in the proposition results directly from Lemma 2.

Proposition 7 (Optimal parameter of the VST)

(i) *Define $Z := b\sqrt{Y + c}$. Then we have:*

$$\mathbb{E}[Z] = b\sqrt{\lambda\tau_1} + b\frac{4c\tau_1 - \tau_2}{8\tau_1^{3/2}}\lambda^{-1/2} + O_{\lambda\to+\infty}(\lambda^{-1}) \tag{5.30}$$

$$\begin{aligned}\operatorname{Var}[Z] = {} & b^2\frac{\tau_2}{4\tau_1} + b^2\left(\frac{7\tau_2^2}{32\tau_1^3} - \frac{2\tau_2 c + \tau_3}{8\tau_1^2}\right)\lambda^{-1} + b^2\left(\frac{5\tau_4 + 16c^2\tau_2 + 16c\tau_3}{64\tau_1^3}\right. \\ & \left. - \frac{17\tau_2\tau_3 + 21c\tau_2^2}{32\tau_1^4} + \frac{75\tau_2^3}{128\tau_1^5}\right)\lambda^{-2} + O_{\lambda\to+\infty}(\lambda^{-5/2})\end{aligned} \tag{5.31}$$

(ii) *For the VST to be second order accurate and Z to have asymptotic unit variance, b and c must satisfy:*

$$c = \frac{7\tau_2}{8\tau_1} - \frac{\tau_3}{2\tau_2} \quad , \quad b = b_1 := 2\sqrt{\frac{\tau_1}{\tau_2}} \tag{5.32}$$

(iii) *For b and c as above, $Z - b_1\sqrt{\tau_1\lambda} \xrightarrow[\lambda\to+\infty]{\mathcal{D}} \mathcal{N}(0,1)$.*

Proposition 7 tells us that for the chosen value of c, the first order term in the expansion (5.31) disappears, and the variance is almost constant up to a second order residue. As a normalizing factor, the value of b does not influence the convergence rate. Note that if there is no filtering ($h = \delta$), the value of c in (5.32) equals 3/8, i.e., the value of the Anscombe VST.

Now fix c to the value given in (5.32). Once the asymptotic expectation is normalized to $\sqrt{\lambda}$, the coefficient of the higher-order term $O(\lambda^{-1/2})$ in (5.30) is given by (5.33). Similarly, the asymptotic variance being normalized to 1, the coefficient of the term $O(\lambda^{-2})$ in (5.31) is shown in (5.34).

$$C_{\mathbb{E}} = \frac{5\tau_2^2 - 4\tau_1\tau_3}{16\tau_1^2\tau_2} \tag{5.33}$$

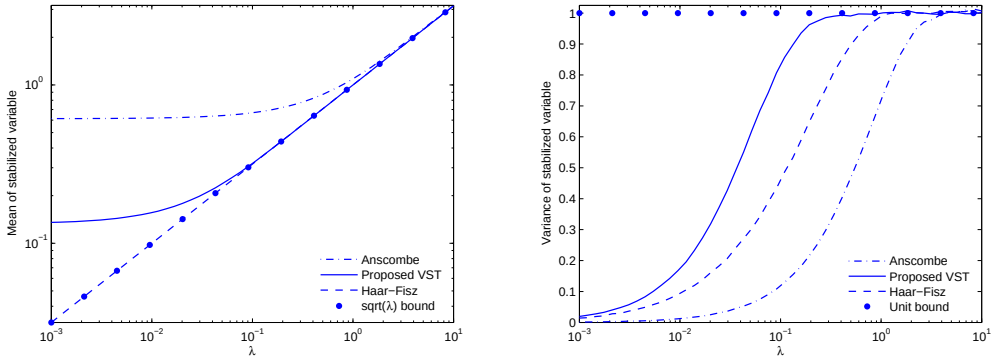
$$C_{\mathrm{Var}} = \frac{5\tau_1^2\tau_2\tau_4 + 13\tau_2^4 - 4\tau_1^2\tau_3^2 - 13\tau_1\tau_2^2\tau_3}{16\tau_1^4\tau_2^2} \tag{5.34}$$

These higher-order coefficients (5.33) and (5.34) can be used to evaluate the stabilization efficiency for a given filter. The ideal filters will be those minimizing (5.33) and (5.34). Tab. 5.2 shows the values of $C_{\mathbb{E}}$ and C_{Var} for different filters, where $h = \delta$ corresponds to the Anscombe VST (no filtering). Note that the values for the Anscombe VST are 10 or even 100 times larger than for the other cases, indicating the benefits of filtering prior to the stabilization. This is also confirmed by the simulations depicted in Fig. 5.6, where the estimates of $\mathbb{E}[Z]$ (resp. $\mathrm{Var}[Z]$) obtained from $2 \cdot 10^5$ replications are plotted as a function of the intensity λ for Anscombe [70] (dashed-dotted), Haar-Fisz [72] (dashed) and our VST (solid). The asymptotic bounds (dotted), i.e., $\sqrt{\lambda}$ for the expectation and 1 for the variance, are also shown. It can be seen that for increasing intensity, $\mathbb{E}[Z]$ and $\mathrm{Var}[Z]$ stick to the theoretical bounds at different rates depending on the VST used. Quantitatively, Poisson variables transformed using the Anscombe VST can be reasonably considered to be unbiased and stabilized for $\lambda \gtrapprox 10$, using Haar-Fisz for $\lambda \gtrapprox 1$, and using our VST (after low-pass filtering with the chosen h) for $\lambda \gtrapprox 0.1$.

Tab. 5.2: $C_{\mathbb{E}}$ and C_{Var} of different filters

Filter[1] h	$C_{\mathbb{E}}$	C_{Var}
δ (Anscombe)	6.25×10^{-2}	6.25×10^{-2}
2D Average $= h_A \otimes h_A$	6.94×10^{-3}	7.72×10^{-4}
2D B_3-Spline $= h_{B_3} \otimes h_{B_3}$	-4.94×10^{-4}	-3.45×10^{-4}

[1] $h_A = [1\ 1\ 1]/3$; $h_{B_3} = [1\ 4\ 6\ 4\ 1]/16$; $\otimes$ denotes the tensor product.



(a) (b)

Fig. 5.6: Behavior of (a) $\mathbb{E}[Z]$ and (b) $\mathrm{Var}[Z]$ as a function of the underlying intensity, for the Anscombe VST, 2D Haar-Fisz VST, and the proposed VST with a low-pass filter $h =$ 2D B_3-Spline filter.

5.3.2 Denoising by MS-VST+wavelets

5.3.2.1 General settings

In this section, the proposed VST will be incorporated within the multiscale framework offered by the (non-necessarily separable) UWT, giving rise to the MS-VST. The undecimated transform is used because it provides translation-invariant denoising. Below, we first discuss the 1D denoising case, and then the multidimensional extension will be straightforward (see Section 5.3.2.2 and 5.3.2.3).

The VST can be combined with the UWT in the following way: since $(\bar{h}^{\uparrow j})_j$ are low-pass filters (so have nonzero means), we can first stabilize the approximation coefficients $(a_j)_j$ using the VST, and then compute in the standard way the detail coefficients from

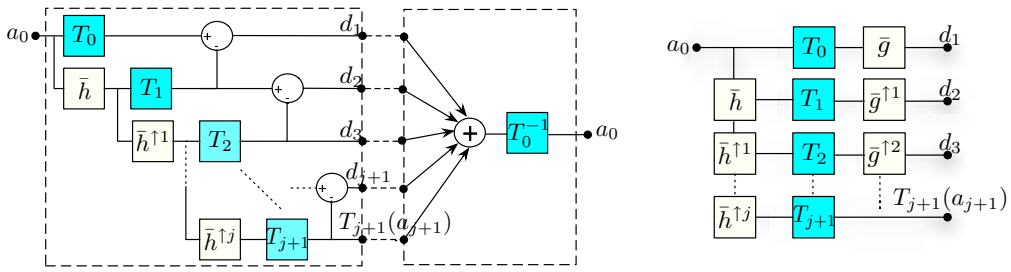
the stabilized a_j's. Note that the VST is now scale-dependent (hence MS-VST). By doing so, the asymptotic stabilized Gaussianity of the a_j's will be transferred to the d_j's, as will be shown later. Thus, the distribution of the d_j's being known (Gaussian), we can detect the significant coefficients by classical hypothesis tests. With the knowledge of the detected coefficients, the final estimate can be reconstructed. In summary, UWT denoising with the MS-VST involves the following three main steps:

1. **Transformation (Sections 5.3.2.2 and 5.3.2.3)**: Compute the UWT in conjunction with the MS-VST;

2. **Detection (Section 5.3.2.4)**: Detect significant detail coefficients by hypothesis tests;

3. **Estimation (Section 5.3.2.5)**: Reconstruct the final estimate iteratively using the knowledge of the detected coefficients.

The last step needs some explanation. The signal reconstruction requires inverting the MS-VST-combined UWT after the detection step. However, the nonlinearity of the MS-VST makes a direct inversion impossible in the general case. Even for the isotropic UWT (IUWT) which uses special filter banks yielding an invertible MS-VST, the direct inverse will be seen to be suboptimal. Hence, we propose to reformulate the reconstruction as a convex optimization problem and solve it by an iterative steepest descent algorithm (Section 5.3.2.5).



(a) (b)

Fig. 5.7: Diagrams of the MS-VST in 1D for (a) the IUWT and (b) the standard UWT.

5.3.2.2 MS-VST+isotropic UWT

The IUWT [106] uses the filter bank $(h, g = \delta - h, \tilde{h} = \delta, \tilde{g} = \delta)$ where h is typically a symmetric low-pass filter such as the B_3-Spline filter. The particular structure of the analysis filters (h, g) leads to the iterative decomposition scheme shown in the left part of (5.35). The reconstruction is trivial, i.e., $a_0 = a_J + \sum_{j=1}^{J} d_j$. This algorithm is widely used in biomedical imaging [107] and astronomy [97] to detect isotropic objects.

As stated in Section 5.3.2.1, we apply the VST on the a_j's resulting in the stabilization procedure shown in the right part of (5.35):

$$\text{IUWT} \left\{ \begin{array}{rcl} a_j & = & \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j & = & a_{j-1} - a_j \end{array} \right. \Longrightarrow \begin{array}{c} \text{MS-VST} \\ + \\ \text{IUWT} \end{array} \left\{ \begin{array}{rcl} a_j & = & \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j & = & T_{j-1}(a_{j-1}) - T_j(a_j) \end{array} \right. \tag{5.35}$$

Note that the filtering step on a_{j-1} can be rewritten as a filtering on $a_0 := \mathbf{X}$, i.e., $a_j = h^{(j)} \star a_0$, where $h^{(j)} = \bar{h}^{\uparrow j-1} \star \cdots \star \bar{h}^{\uparrow 1} \star \bar{h}$ for $j \geq 1$ and $h^{(0)} = \delta$. T_j is the VST operator at scale j (see Lemma 2):

$$T_j(a_j) = b^{(j)} \operatorname{sgn}(a_j + c^{(j)}) \sqrt{|a_j + c^{(j)}|} \tag{5.36}$$

Let us define $\tau_k^{(j)} := \sum_i \left(h^{(j)}[i]\right)^k$. Then according to (5.32), the constant $c^{(j)}$ associated to $h^{(j)}$ should be set to

$$c^{(j)} := \frac{7\tau_2^{(j)}}{8\tau_1^{(j)}} - \frac{\tau_3^{(j)}}{2\tau_2^{(j)}} \tag{5.37}$$

This stabilization procedure is directly invertible as we have:

$$a_0 = T_0^{-1} \left[T_J(a_J) + \sum_{j=1}^{J} d_j \right] \tag{5.38}$$

The decomposition scheme and the inversion of MSVST+IUWT are also illustrated in Fig. 5.7(a).

Asymptotic Distribution of the Detail Coefficients

Theorem 3 (Asymptotic distribution of d_j) *Setting $b^{(j)} := sgn(\tau_1^{(j)})/\sqrt{|\tau_1^{(j)}|}$, if λ is constant within the support of the filter $h^{(j)}[k - \cdot]$, then we have:*

$$d_j[k] \xrightarrow[\lambda \to +\infty]{\mathcal{D}} \mathcal{N}\left(0, \quad \frac{\tau_2^{(j-1)}}{4{\tau_1^{(j-1)}}^2} + \frac{\tau_2^{(j)}}{4{\tau_1^{(j)}}^2} - \frac{\langle h^{(j-1)}, h^{(j)} \rangle}{2\tau_1^{(j-1)}\tau_1^{(j)}}\right) \tag{5.39}$$

The proof of this theorem is given in Appendix A.12. This is a very useful result showing that the detail coefficients issued from locally homogeneous parts of the signal (null hypothesis H_0, see Section 5.3.2.4) follow asymptotically a centered normal distribution with an intensity-independent variance which only relies on the filter h and the current scale. Hence, the stabilized variance (also the constants $b^{(j)}$, $c^{(j)}$, $\tau_k^{(j)}$) can all be pre-computed once for any given h.

Extension to the Multi-dimensional Case The filter bank in qD ($q > 1$) becomes $(h_{q\mathrm{D}}, g_{q\mathrm{D}} = \delta - h_{q\mathrm{D}}, \tilde{h}_{q\mathrm{D}} = \delta, \tilde{g}_{q\mathrm{D}} = \delta)$ where $h_{q\mathrm{D}} = \otimes_{i=1}^{q} h$. Note that $g_{q\mathrm{D}}$ is in general nonseparable. The MS-VST decomposition scheme remains the same as (5.35), and the asymptotic result above holds true. The complexity for pre-computing $b^{(j)}$, $c^{(j)}$, $\tau_k^{(j)}$ and the stabilized variance in (5.39) remains the same as in the 1D case.

5.3.2.3 MS-VST+standard UWT

In this section, we show how the MS-VST can be used to stabilize the wavelet coefficients of a standard separable UWT. In the same vein as (5.35), we apply the VST on the approximation coefficients $(a_j)_j$, leading to the following scheme (see also the block-diagram of Fig. 5.7(b)):

$$\mathrm{UWT}\left\{\begin{array}{rcl} a_j &=& \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &=& \bar{g}^{\uparrow j-1} \star a_{j-1} \end{array}\right. \quad \Longrightarrow \quad \begin{array}{c}\text{MS-VST} \\ + \\ \text{UWT}\end{array} \quad \left\{\begin{array}{rcl} a_j &=& \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &=& \bar{g}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \end{array}\right. \tag{5.40}$$

where $T_j(a_j) = b^{(j)} \operatorname{sgn}(a_j + c^{(j)}) \sqrt{|a_j + c^{(j)}|}$, and $c^{(j)}$ is defined as in (5.37).

Asymptotic Distribution of the Detail Coefficients

Theorem 4 (Asymptotic distribution of d_j) *Setting $b^{(j)} := 2\sqrt{|\tau_1^{(j)}|/\tau_2^{(j)}}$, if λ is constant within the support of the filter $(\bar{g}^{\uparrow j-1} \star h^{(j-1)})[k - \cdot]$, then $d_j[k] \xrightarrow[\lambda \to +\infty]{\mathcal{D}} \mathcal{N}(0, \sigma_j^2)$, where*

$$\sigma_j^2 = \frac{1}{\tau_2^{(j-1)}} \sum_{m,n} \bar{g}^{\uparrow j-1}[m] \bar{g}^{\uparrow j-1}[n] \sum_k h^{(j-1)}[k] h^{(j-1)}[k + m - n] \tag{5.41}$$

Parallel to Theorem 3, Theorem 4 (proof in Appendix A.12) shows the asymptotic normality of the detail coefficients issued from locally homogeneous parts of the signal (null hypothesis H_0, see Section 5.3.2.4). Here, the values of $b^{(j)}$, $c^{(j)}$, $\tau_k^{(j)}$ and σ_j can all be pre-computed once (h, g) has been chosen.

Extension to the Multi-dimensional Case The scheme (5.40) can be extended straightforward to higher dimensional cases, and the asymptotic result above holds true. For example, in the 2D case, the UWT is given by the left part of (5.42) and the version combined with the MS-VST is given on the right:

$$\text{UWT}\begin{cases} a_j &= \bar{h}^{\uparrow j-1}\bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j^1 &= \bar{g}^{\uparrow j-1}\bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j^2 &= \bar{h}^{\uparrow j-1}\bar{g}^{\uparrow j-1} \star a_{j-1} \\ d_j^3 &= \bar{g}^{\uparrow j-1}\bar{g}^{\uparrow j-1} \star a_{j-1} \end{cases} \Longrightarrow \begin{array}{c}\text{MS-VST}\\ +\\ \text{UWT}\end{array} \begin{cases} a_j &= \bar{h}^{\uparrow j-1}\bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j^1 &= \bar{g}^{\uparrow j-1}\bar{h}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \\ d_j^2 &= \bar{h}^{\uparrow j-1}\bar{g}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \\ d_j^3 &= \bar{g}^{\uparrow j-1}\bar{g}^{\uparrow j-1} \star T_{j-1}(a_{j-1}) \end{cases} \tag{5.42}$$

where $hg \star a$ is the convolution of a by the separable filter hg, i.e., convolution first along the rows by h and then along the columns by g. The complexity for pre-computing the constants $b^{(j)}$, $c^{(j)}$, $\tau_k^{(j)}$ and σ_j remains the same as in the 1D case.

5.3.2.4 Detection by wavelet-domain hypothesis testing

Our wavelet-domain detection is formulated by hypothesis tests, i.e., testing the null hypothesis $H_0 : d_j[k]$ is insignificant against the alternative $H_1 : d_j[k]$ is significant. Note that wavelet coefficients computed from locally homogeneous parts of the signal are insignificant. Indeed, if there were no noise, these coefficients issued from the classical UWT scheme would be zero-valued, as any wavelet has a zero mean. Thanks to Theorems 3 and 4, the distribution of the stabilized $d_j[k]$ under the null hypothesis H_0 is now known (Gaussian). Thus, according to the statistical error rate intended to be controlled, individual tests and multiple tests are all applicable.

5.3.2.5 Iterative reconstruction

Following the detection step, we have to invert the MS-VST schemes to reconstruct the estimate. For the standard UWT case, direct reconstruction procedure is unavailable since the convolution (by $\bar{g}^{\uparrow j-1}$) operator and the VST operator T_{j-1} do not commute in (5.40). For the IUWT case, the estimate can be reconstructed by [Eq.(5.38)]. However, this direct MS-VST inversion followed by a positivity projection[2] could entail a loss of important structures in the estimate. Here, we propose to reformulate the reconstruction as a convex optimization problem described below, and solve it iteratively. This procedure will be

[2]Positivity projection because Poisson intensity is always nonnegative.

shown to better preserve the significant structures in the data than the direct inverse. In the following, we will concentrate on the 1D case for clarity.

We suppose that the underlying intensity function Λ is sparsely represented in the wavelet domain. We define the multiresolution support [108] $\mathcal{M}$, which is determined by the set of detected significant coefficients at each scale j and location k, i.e.,

$$\mathcal{M} := \{(j,k) \mid \text{if } d_j[k] \text{ is significant}\} \tag{5.43}$$

The estimation is then formulated as a constrained sparsity-promoting minimization problem in terms of the wavelet coefficients $\mathbf{d}$. A component of $\mathbf{d}$ can be indexed by the usual scale-location index (j,k) (i.e. $d_j[k]$). The indices can also be renumbered so that $\mathbf{d}$ is mapped to a vector in $\mathbb{R}^L$. In this case, a component of $\mathbf{d}$ is indexed in a 1D way, i.e., $d[i]$. Hereafter, both notations will be used. Our optimization problem is given by

$$\begin{aligned} &\min_{\mathbf{d}\in\mathcal{C}} J(\mathbf{d}), \quad J(\mathbf{d}) := \|\mathbf{d}\|_1 \\ &\text{where } \mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2,\ \mathcal{S}_1 := \{\mathbf{d} | d_j[k] = (\mathcal{W}\mathbf{X})_j[k], (j,k) \in \mathcal{M}\},\ \mathcal{S}_2 := \{\mathbf{d} | \mathcal{R}\mathbf{d} \geq 0\} \end{aligned} \tag{5.44}$$

where $\mathcal{W}$ represents the wavelet transform operator, and $\mathcal{R}$ its (weak-generalized) left inverse (synthesis operator). It can be seen that we seek the sparsest solution by minimizing the ℓ^1 objective [109][110] within the feasible set $\mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2$. The set $\mathcal{S}_1$ requires that the significant elements of $\mathbf{d}$ preserve those of the data $\mathbf{X}$; the set $\mathcal{S}_2$ ensures a positive intensity estimate.

(5.44) is a convex optimization problem which be cast as a Linear Programming (LP) and solved using interior-point methods. However, the computational complexity of the LP solver increases dramatically with the size of the problem. Classical projected (sub-)gradient method is also difficult to apply here since the projector on the feasible set is unknown. Below we propose a much faster alternative based on the hybrid steepest descent (HSD) [111]. The HSD approach allows minimizing convex functionals over the intersection of fixed point sets of nonexpansive mappings. It is much faster than LP, and in our problem, the nonexpansive mappings do have closed forms. We first establish the following theorem (proof in Appendix A.13).

Theorem 5 *Let* $\mathbf{d} \in \mathbb{R}^L$. *Define the following optimization problem* ($\epsilon \geq 0$)*:*

$$\begin{aligned} &\min_{\mathbf{d}\in\mathcal{C}_B} J_\epsilon(\mathbf{d}), \quad J_\epsilon(\mathbf{d}) := \textstyle\sum_{i=1}^{L} \sqrt{d[i]^2 + \epsilon} \\ &\text{where } \mathcal{C}_B := \mathcal{S}_1 \cap \mathcal{S}_2 \cap \mathcal{S}_3,\ \mathcal{S}_3 := \{\mathbf{d} |\ \|\mathbf{d}\|_2 \leq B, B \geq \|\mathcal{W}\mathbf{X}\|_1\} \end{aligned} \tag{5.45}$$

Define the HSD iteration scheme [111] ($k \geq 0$):

$$\mathbf{d}_\epsilon^{(k+1)} := T_{\mathcal{C}_B}\mathbf{d}_\epsilon^{(k)} - \beta_{k+1}\nabla J_\epsilon\left(T_{\mathcal{C}_B}\mathbf{d}_\epsilon^{(k)}\right) \tag{5.46}$$

where ∇J_ϵ is the gradient of J_ϵ, and $T_{\mathcal{C}_B} := P_{\mathcal{S}_3} \circ P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2}$,

$$P_{\mathcal{S}_3}\mathbf{d} := \frac{\mathbf{d}}{\|\mathbf{d}\|_2}\cdot\min(\|\mathbf{d}\|_2, B); \quad (P_{\mathcal{S}_1}\mathbf{d})_j[k] := \begin{cases} (\mathcal{W}\mathbf{X})_j[k] & (j,k) \text{ in } \mathcal{M} \\ \mathbf{d}_j[k] & \text{otherwise} \end{cases}; \quad Q_{\mathcal{S}_2}\mathbf{d} := \mathcal{W}P_+\mathcal{R}\mathbf{d} \tag{5.47}$$

where P_+ represents the projection on the nonnegative orthant, and $P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_3}$ are the projectors onto their respective constraint sets. The step sequence satisfies:

$$\lim_{k\to\infty}\beta_k = 0, \quad \sum_{k\geq 1}\beta_k = +\infty \quad \text{and} \quad \sum_{k\geq 1}|\beta_k - \beta_{k+1}| < +\infty \tag{5.48}$$

Suppose that in (ii)-(v) below $\mathcal{W}$ represents a tight frame decomposition and $\mathcal{R}$ its pseudo-inverse operator. Then we have:

(i) *The solution set of (5.44) is the same as that of (5.45) with $\epsilon = 0$;*

(ii) *$T_{\mathcal{C}_B}$ is nonexpansive, and its fix point set is $Fix(T_{\mathcal{C}_B}) = \mathcal{C}_B \neq \emptyset$;*

(iii) *$\forall\epsilon > 0$, with any $\mathbf{d}_\epsilon^{(0)} \in \mathbb{R}^N$, $\mathbf{d}_\epsilon^{(k)} \underset{k\to+\infty}{\longrightarrow} \mathbf{d}_\epsilon^*$, where $\mathbf{d}_\epsilon^*$ is the unique solution to (5.45);*

(iv) *As $\epsilon \to 0^+$, the sequence $(\mathbf{d}_\epsilon^*)_{\epsilon>0}$ is bounded. Therefore, it has at least one limit point;*

(v) *As $\epsilon \to 0^+$, every limit point of the sequence $(\mathbf{d}_\epsilon^*)_{\epsilon>0}$ is a solution to (5.44).*

Theorem 5 implies that in practice instead of directly solving (5.44), one can solve its smoothed version (5.45) by applying (5.46) with a small ϵ. In practice, $T_{\mathcal{C}_B}$ can be simplified to $T_{\mathcal{C}_B} = T_{\mathcal{C}} := P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2}$. Indeed, for real problems the exact value of B is not important, and it can be considered to be sufficiently large so that the constraint $\mathcal{S}_3$ is always satisfied. We also point out that although Theorem 5 assumes a tight frame decomposition and pseudo-inverse reconstruction, in our experiments, it has been observed that the iterations (5.46) applied equally to general frame decompositions and inverses, and performed very well even with $\epsilon = 0$ (see results in Section 5.3.2.6). For $\epsilon = 0$, (5.49) rewrites:

$$\mathbf{d}^{(k+1)} := T_{\mathcal{C}}\mathbf{d}^{(k)} - \beta_{k+1}\nabla J\left(T_{\mathcal{C}}\mathbf{d}^{(k)}\right) \quad \text{where} \quad \nabla J(\mathbf{d}) = \text{sgn}(\mathbf{d}) \tag{5.49}$$

where $\nabla J(\mathbf{d})[i] = \text{sgn}(d[i])$ is the limiting gradient[3] of J_ϵ as $\epsilon \to 0+$. (5.49) is implemented in practice as a soft thresholding with a threshold β_{k+1} (noted as $\text{ST}_{\beta_{k+1}}$). Now the MS-VST denoising using the IUWT and the standard UWT is presented in Algorithm 2 and 3 respectively. In Algorithm 2, step 1 – 6 obtain a first estimate of Λ by directly

Algorithm 2 MS-VST + IUWT

Require: $a_0 := \mathbf{X}$; a low-pass filter h,

Detection

1: **for** $j = 1$ to J **do**
2: Compute a_j and d_j using (5.35).
3: Hypothesis testing d_j assuming the normal statistics (Theorem 3), get the estimate $\hat{d}_j$, and update $\mathcal{M}$.
4: **end for**

Estimation

5: Estimate $\mathbb{E}[T_0(a_0)]$ by: $\widehat{T_0 a_0} = \sum_{j=1}^{J} \hat{d}_j + T_J(a_J)$
6: Estimate $\mathbb{E}[a_0]$ by: $\hat{a}_0 = \text{Var}[T_0(a_0)] + \widehat{T_0 a_0}^2 - c^{(0)}$
7: Initialize $\mathbf{d}^{(0)} = \mathcal{W} P_+ \hat{a}_0$
8: **for** $k = 1$ to $N_{\max}$ **do**
9: $\tilde{\mathbf{d}} := P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2} \mathbf{d}^{(k-1)}$
10: $\hat{\mathbf{d}} := \mathbf{d}^{(k)} := \text{ST}_{\beta_k}[\tilde{\mathbf{d}}]$.
11: **end for**
12: Get the estimate $\hat{\Lambda} = P_+ \mathcal{R} \hat{\mathbf{d}}$.

Algorithm 3 MS-VST + Standard UWT

Require: $a_0 := \mathbf{X}$; a wavelet filter bank $(h, g, \tilde{h}, \tilde{g})$,

Detection

1: **for** $j = 1$ to J **do**
2: Compute a_j and d_j using (5.40).
3: Hypothesis testing d_j assuming the normal statistics (Theorem 4) and update $\mathcal{M}$.
4: **end for**

Estimation

5: Initialize $\mathbf{d}_j^{(0)}[k] = (\mathcal{W}\mathbf{X})_j[k]$, if $(j,k) \in \mathcal{M}$; 0 otherwise.
6: **for** $k = 1$ to $N_{\max}$ **do**
7: $\tilde{\mathbf{d}} := P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2} \mathbf{d}^{(k-1)}$
8: $\hat{\mathbf{d}} := \mathbf{d}^{(k)} := \text{ST}_{\beta_k}[\tilde{\mathbf{d}}]$.
9: **end for**
10: Get the estimate $\hat{\Lambda} = P_+ \mathcal{R} \hat{\mathbf{d}}$.

inverting MS-VST+IUWT after zeroing the insignificant wavelet coefficients. The direct

[3]Clearly, $\nabla J(\mathbf{d})$ is also a sub-gradient of J. The sub-gradient of J is given by $\partial J(\mathbf{d})[i] = \text{sgn}(d[i])$ if $d[i] \neq 0$ and $\partial J(\mathbf{d})[i] \in [-1, 1]$ otherwise.

inverse serves as the initialization of the iterations. In step 6, the term $\mathrm{Var}\left[T_0(a_0)\right]$ corrects the bias due to squaring an estimate. Indeed, if $Z = \sqrt{a_0 + c^{(0)}}$, then $\lambda = \mathbb{E}\left[a_0\right] = \mathbb{E}\left[Z^2\right] - c^{(0)} = \mathrm{Var}\left[Z\right] + \mathbb{E}\left[Z\right]^2 - c^{(0)}$. We can also see that every iteration of (5.49) involves a projection onto $\mathcal{S}_1$ that restores all the significant coefficients. This actually results in a better preservation of the important structures than the direct inverse (see also the results in Section 5.3.2.6).

In Algorithm 3 the initialization is given by the detected significant wavelet coefficients (step 5). In both algorithms, $N_{\max}$ is the maximal number of iterations. A possible choice of the sequence $(\beta_k)_k$ is a linearly decreasing one: $\beta_k = \frac{N_{\max}-k}{N_{\max}-1}, k = 1, 2, \cdots, N_{\max}$. It can be noted that for $(\beta_k)_k$ chosen as above, the conditions in (5.48) are all satisfied as $N_{\max} \to \infty$. The computational cost of the whole denoising is dominated by the iterative estimation step. This step involves an analysis and a synthesis at each iteration and thus has a complexity of $O(2N_{\max}V)$, where $V = O(N \log N)$ is the complexity of UWT and N is the number of data samples.

5.3.2.6 Applications

Simulated confocal image restoration We have simulated an image containing disk-like isotropic sources on a constant background (see Fig. 5.8(a)) where the pixel size is $100nm \times 100nm$. From the leftmost column to the rightmost one, source radii increase from $50nm$ to $350nm$. This image has been convolved with a Gaussian PSF with a standard deviation of $103nm$. This function approximates the PSF of a nonparaxial LSCM ($\lambda_{ex} = 488nm$, $\lambda_{ex} = 520nm$, NA = 1.0, $D = 1$AU, see Chapter 3). The source amplitudes range from 0.08 to 4.99, and the background level is 0.03. This spot grid can be deemed as a model for cellular vesicles of different sizes and intensities. A realization of the photon-count image is shown in Fig. 5.8(b). We present the restoration results given by Anscombe [71] (Fig. 5.8(c)), Haar-Fisz [72] (Fig. 5.8(d)), CVS [74] (Fig. 5.8(e)), Haar+CLTB thresholding [100] (Fig. 5.8(f)), Platelet estimation [112, 92, 93](Fig. 5.8(g)), and the MS-VST denoiser using iterative (Fig. 5.8(h)) and direct (Fig. 5.8(i)) reconstructions. IUWT has been used to produce the results in Fig. 5.8(c)(d)(e)(i)(j); standard Haar UWT is used in Fig. 5.8(f); cycle spinning with a total of 25 shifts is employed in Fig. 5.8(d)(g) to attenuate the block artifacts. The controlled FPR in all the wavelet-based methods is set to 5×10^{-3}; for the platelet approach, the penalizing factor γ is set to $1/3$.

As revealed by Fig. 5.8, all the estimators perform comparatively well at high intensity levels (right part of the images). For low-intensity sources, Haar-Fisz, CVS, Platelets and the MS-VST are the most sensitive approaches. We can see that the IUWT-based methods preserve better the isotropic source shapes than the other methods. Some residual noise can be seen in the estimate of CVS.

We also quantify the performances in terms of the NMISE per bin based on 5 replications. The MS-VST denoiser results in the second lowest error, which is slightly larger than that of the platelet estimate. The platelet estimator offers an efficient piecewise linear approximation to the image. However, on the isolated smooth spots, it tends to alter the isotropic shapes and produces some artifacts. The regularity in the result could be improved by averaging a larger number of cyclic shifts, but leading to a very time-consuming procedure (a computation-time benchmark is shown for a real example in Section 5.3.4.3).

Finally, we can also observe that the iterative reconstruction Fig. 5.8(i) is more effective in restoring low-flux sources (see the upper part of the image) than the direct inverse Fig. 5.8(j). This phenomenon is clearly expected.

Other application: astronomical image restoration Our method can be equally applied to the image restoration problems in astronomical domains. Fig. 5.9 compares the restoration methods on a galaxy image. The FDR control is employed in Anscombe, Haar-Fisz, CVS, TI Haar hypothesis tests [100], and the MS-VST methods. Among all the results, Haar-Fisz, CVS, Platelets and the MS-VST estimates detect more faint sources. It is found that Haar-Fisz, Haar hypothesis tests, Platelets and the MS-VST with iterative construction generate comparable low NMISE values, among which the iterative MS-VST leads to the smallest one.

5.3.3 Denoising by MS-VST+ridgelets

5.3.3.1 The ridgelet transform

The ridgelet transform [52] has been shown to be very effective for representing global lines in an image. Ridgelet analysis may be constructed as a wavelet analysis in the Radon domain. Recall that the 2D Radon transform of an object f is the collection of line integrals indexed by $(\theta, t) \in [0, 2\pi) \times \mathbb{R}$ given by

$$Rf(\theta, t) = \int_{\mathbb{R}^2} f(x_1, x_2)\delta(x_1 \cos\theta + x_2 \sin\theta - t)\, dx_1 dx_2 \tag{5.50}$$

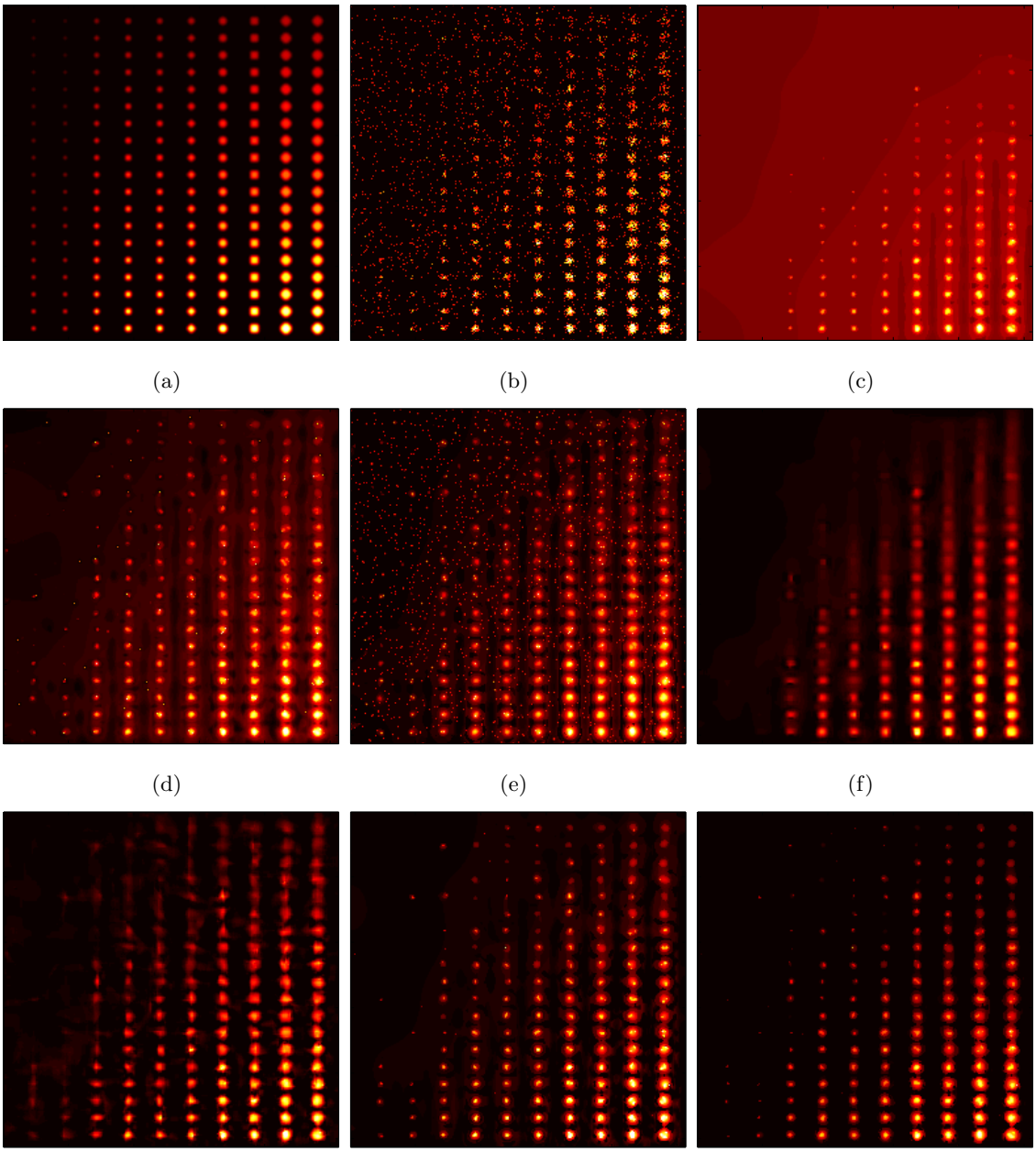


Fig. 5.8: Denoising an image of simulated spots of different radii (image size: 256×256). (a) simulated sources (amplitudes $\in [0.08, 4.99]$; background $= 0.03$); (b) observed counts; (c) Anscombe-denoised image (IUWT, $J = 5$, FPR $= 5 \times 10^{-3}$, NMISE $= 2.34$); (d) Haar-Fisz-denoised image (IUWT, $J = 5$, FPR $= 5 \times 10^{-3}$, 25 cyclic shifts where 5 for each of the axes, NMISE $= 0.33$); (e) CVS-denoised image (IUWT, $J = 5$, FPR $= 5 \times 10^{-3}$, NMISE $= 0.81$); (f) image denoised by HT-based Haar thresholding (Haar UWT, $J = 5$, FPR $= 5 \times 10^{-3}$, NMISE $= 0.10$); (g) platelet-denoised image ($\gamma = 1/3$, 25 random cyclic shifts, NMISE $= 0.059$); (h) MS-VST-denoised image (IUWT, $J = 5$, FPR $= 5 \times 10^{-3}$, $N_{max} = 20$ iterations, NMISE $= 0.069$); (i) MS-VST-denoised image (IUWT, $J = 5$, FPR $= 5 \times 10^{-3}$, direct inverse, NMISE $= 0.073$).

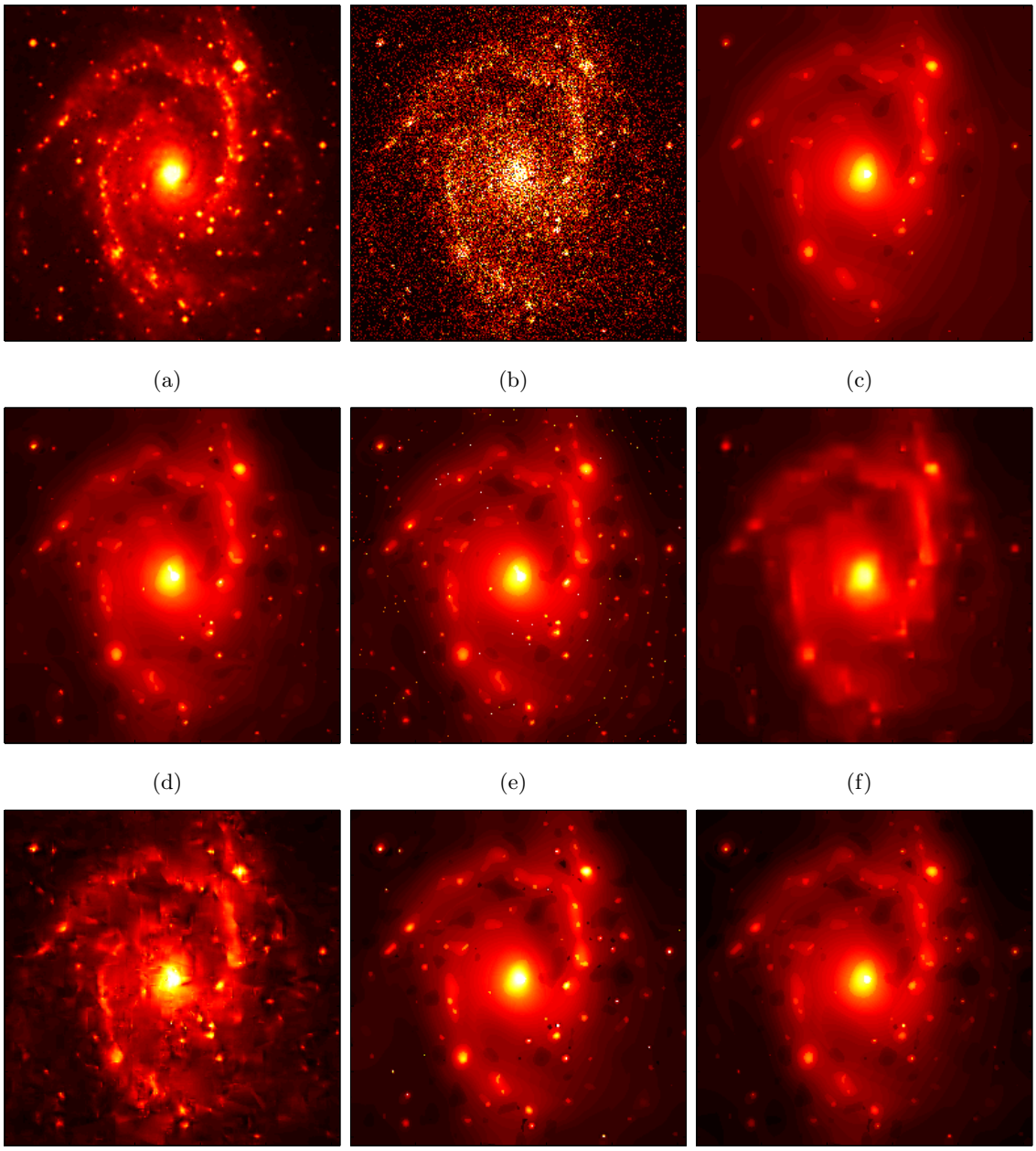


Fig. 5.9: Denoising a galaxy image (image size: 256×256). (a) galaxy image (intensity $\in [0, 5]$); (b) observed counts; (c) Anscombe-denoised image (IUWT, B_3-spline filter bank, $J = 5$, FDR $= 0.1$, NMISE $= 0.15$); (d) Haar-Fisz-denoised image (IUWT, B_3-spline filter bank, $J = 5$, FDR $= 0.1$, 25 cyclic shifts where 5 for each of the axes, NMISE $= 0.04$); (e) CVS-denoised image (IUWT, B_3-spline filter bank, $J = 5$, FDR $= 0.1$, NMISE $= 0.074$); (f) denoised image by Haar HTs (Haar UWT, $J = 5$, FDR $= 0.1$, NMISE $= 0.036$); (g) Platelet-denoised image ($\gamma = 1/3$, 25 random cyclic shifts, NMISE $= 0.038$) (h) MS-VST-denoised image (IUWT, B_3-spline filter bank, $J = 5$, FDR $= 0.1$, $N_{max} = 20$ iterations, NMISE $= 0.035$); (i) MS-VST-denoised image (IUWT, B_3-spline filter bank, $J = 5$, FDR $= 0.1$, direct inverse, NMISE $= 0.051$).

where δ is the Dirac distribution. Then the ridgelet transform is precisely the application of a 1D wavelet transform to the slices of the Radon transform where the angular variable θ is constant and t is varying. For each scale $s > 0$, position $t \in \mathbb{R}$ and angle $\theta \in [0, 2\pi)$, the 2D ridgelet function $\psi_{s,t,\theta}$ is defined from a 1D wavelet function ψ as:

$$\psi_{s,t,\theta}(x_1, x_2) = s^{-1/2} \cdot \psi((x_1 \cos\theta + x_2 \sin\theta - t)/s) \tag{5.51}$$

A ridgelet is constant along the lines $x_1 \cos\theta + x_2 \sin\theta = const$. Transverse to a ridge is a wavelet.

Thus, the basic strategy for calculating the continuous ridgelet transform is first to compute the Radon transform $Rf(t, \theta)$ and second, to apply a 1D wavelet transform to the slices $Rf(\cdot, \theta)$. Different digital ridgelet transforms can be derived depending on the choice of both the Radon algorithm and the wavelet decomposition [113]. For example, the Slant Stack Radon (SSR) transform [114][115] is a good candidate, which has the advantage of being geometrically accurate, and is used in our experiments. The inverse SSR has however the drawback to be iterative. If computation time is an issue, the recto-polar Radon transform is a good alternative. More details on the implementation of these Radon transforms can be found in [53][114][115][113].

5.3.3.2 MS-VST with ridgelets

As a Radon coefficient is obtained from an integration of the pixel values along a line, the noise in the Radon domain follows also a Poisson distribution. Thus, we can apply the 1D MS-VST wavelet detection described in Section 5.3.2 to the slices of the Radon transform. Let $\mathcal{M} := \{(\theta, j, k)\}$ denote the ridgelet multi-resolution support, where (θ, j, k) indicates that the stabilized ridgelet coefficient at projection angle θ, scale j and location k is significant. $\mathcal{M}$ being available, we can formulate a constrained ℓ^1-minimization problem in the same way as in the wavelet case (Section 5.3.2.5), which is then solved by HSD iterations. Hence, the Ridgelet Poisson denoising algorithm consists of the following three steps:

Algorithm 4 MS-VST + Ridgelets

1: Apply the Radon transform.
2: For each Radon slice, apply the 1D MS-VST+UWT detection and update $\mathcal{M}$.
3: Apply the HSD iterations to the ridgelet coefficients before getting the final estimate.

5.3.3.3 Experimental results

We have simulated an image with smooth ridges shown in Fig. 5.10(a). The peak intensities of the vertical ridges vary progressively from 0.1 to 0.5; the inclined ridge has a maximum intensity of 0.3; the background level is 0.05. A Poisson-count image is shown in Fig. 5.10(b). The biorthogonal 7/9 filter bank [64] is used in the Anscombe (Fig. 5.10(c)), Haar-Fisz (Fig. 5.10(d)), CVS (Fig. 5.10(e)), and MS-VST+UWT (Fig. 5.10(g)) approaches. Denoising using Haar HTs is given by Fig. 5.10(f). The estimates by Platelets and by MS-VST+Ridgelets are shown in Fig. 5.10(h) and Fig. 5.10(i), respectively. Due to the very low-count setting, the Anscombe estimate is highly biased. Among all the wavelet-based methods, MS-VST+UWT leads to the smallest error, but is outperformed by the Platelet and the MS-VST-based ridgelet estimates. The two latter methods result in the lowest NMISE values among all the competitors. Clearly, this is because wavelets are less adapted to line-like sources. It can also be seen that the shape of the ridges is better preserved by the ridgelet-based estimate.

5.3.4 Denoising by MS-VST+curvelets

5.3.4.1 The first generation curvelet transform

The ridgelet transform is efficient for finding only the lines of the size of the image. To detect line segments, a partitioning need to be introduced. The image is first decomposed into smoothly overlapping blocks of side-length B pixels, and the ridgelet transform is applied independently on each block. This is called the local ridgelet transform. The curvelet transform [116][117] opens the possibility to analyze an image with different block sizes, but with a single transform. The idea is to first decompose the image into a set of wavelet bands using the IUWT, and to analyze each band with a local ridgelet transform. The block size is changed at every other scale. The coarsest resolution of the image (a_J) is not processed. This transform has been shown to provide optimal approximation rate for piecewise C^2 images away from C^2 contours, and is very effective in detecting anisotropic structures of different lengths. More details can be found in [116][53].

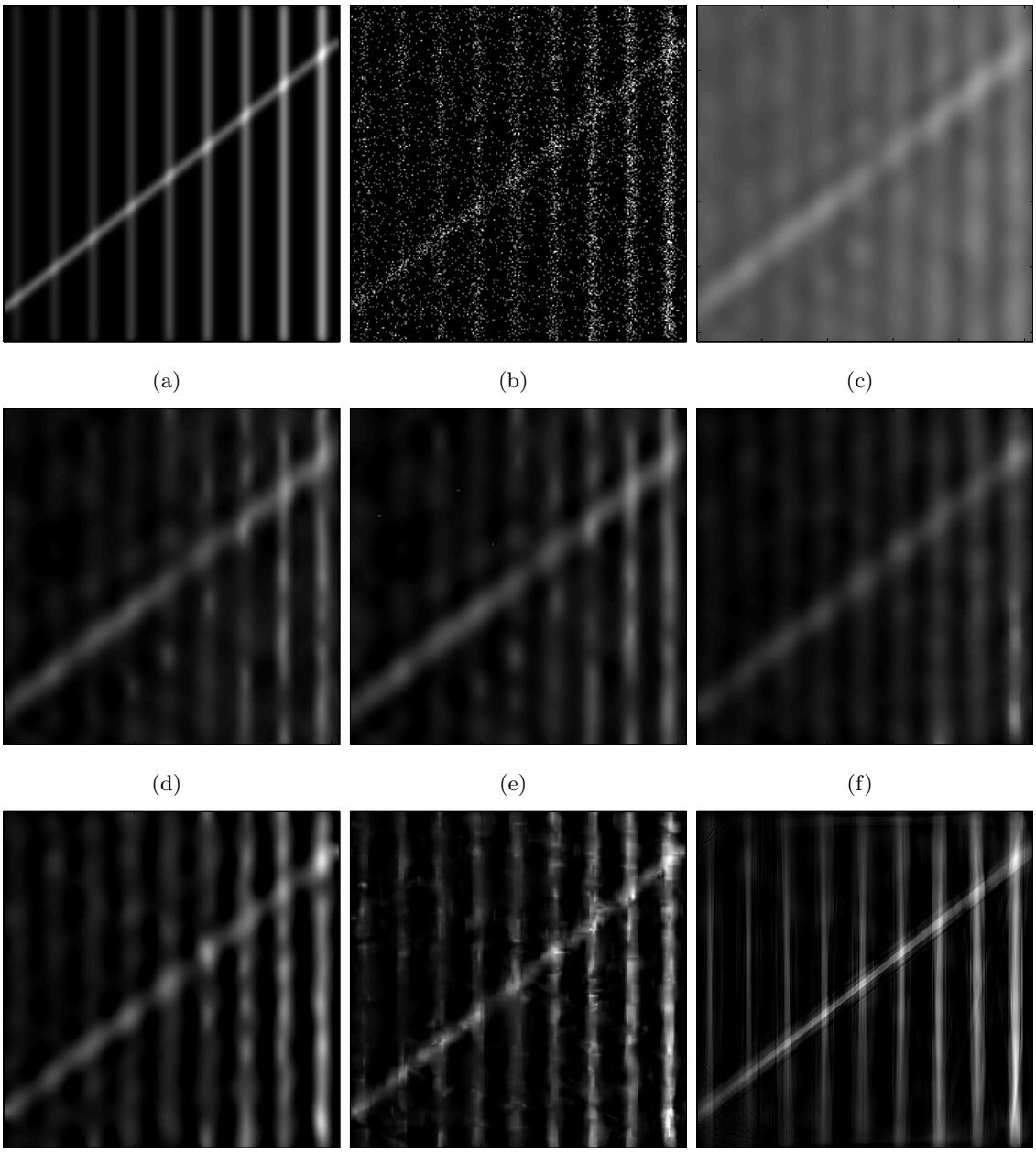


Fig. 5.10: Poisson denoising of smooth ridges (image size: 256×256). (a) intensity image (the peak intensities of the 9 vertical ridges vary progressively from 0.1 to 0.5; the inclined ridge has a maximum intensity of 0.3; background = 0.05); (b) Poisson noisy image; (c) Anscombe-denoised image (UWT, 7/9 filter bank, $J = 4$, FDR = 10^{-7}, NMISE = 0.83); (d) Haar-Fisz-denoised image (UWT, 7/9 filter bank, $J = 4$, FDR = 10^{-7}, 25 cyclic shifts where 5 for each of the axes, NMISE = 0.035); (e) CVS-denoised image (UWT, 7/9 filter bank, $J = 4$, FDR = 10^{-7}, NMISE = 0.034); (f) image denoised by Haar+FDR ($J = 4$, FDR = 10^{-7}, NMISE = 0.044); (g) image denoised by MS-VST+UWT (7/9 filter bank, $J = 4$, FDR = 10^{-7}, $N_{max} = 10$ iterations, NMISE = 0.023); (h) Platelet-denoised image ($\gamma = 1/3$, 25 random cyclic shifts, NMISE = 0.017); (i) MS-VST+Ridgelets ($J = 4$, FDR = 10^{-7}, $N_{max} = 10$ iterations, NMISE = 0.017).

5.3.4.2 MS-VST with curvelets

As the first step of the algorithm is an IUWT, we can stabilize each resolution level in the same way as described in Section 5.3.2.2. We then apply the local ridgelet transform on each stabilized wavelet band. Significant Gaussianized curvelet coefficients will be detected by HTs from which the curvelet multiresolution support $\mathcal{M}$ is derived. Finally, the same to the wavelet and ridgelet case, we solve a constrained ℓ^1-minimization problem on the curvelet coefficients by HSD iterations before reconstructing the estimate. We now present a sketch of the Poisson curvelet denoising algorithm:

Algorithm 5 MS-VST + Curvelets

1: Apply the MS-VST+IUWT with J scales to get the stabilized wavelet subbands $(d_j)_j$.
2: set $B_1 = B_{\min}$
3: **for** $j = 1$ to J **do**
4: Partition the subband d_j with blocks of side-length B_j and apply the digital ridgelet transform to each block to obtain the stabilized curvelet coefficients.
5: Test the stabilized curvelet coefficients to obtain $\mathcal{M}$.
6: **if** j modulo $2 = 1$ **then**
7: $B_{j+1} = 2B_j$
8: **else**
9: $B_{j+1} = B_j$
10: **end if**
11: **end for**
12: Apply the HSD iterations to the curvelet coefficients before getting the final estimate.

It is not as straightforward as with the wavelet and ridgelet transforms to derive the asymptotic noise variance in the stabilized curvelet domain. In our experiments, we derived them using simulated data with Poisson noise only. After having checked that the standard deviation in the curvelet bands becomes stabilized as the intensity level λ increases (which means that the stabilization is working properly), we stored this standard deviation $\sigma_{j_1,j_2,l}$ for each wavelet scale j_1, each ridgelet scale j_2, and each direction angle l. Then, once the stabilized curvelet transform is applied to our data, these values of $(\sigma_{j_1,j_2,l})_{j_1,j_2,l}$ serve in the hypothesis testing framework described in Section 5.3.2.4 to test the significance of each stabilized curvelet coefficient at each scale (j_1, j_2) and direction angle l.

5.3.4.3 Experimental results

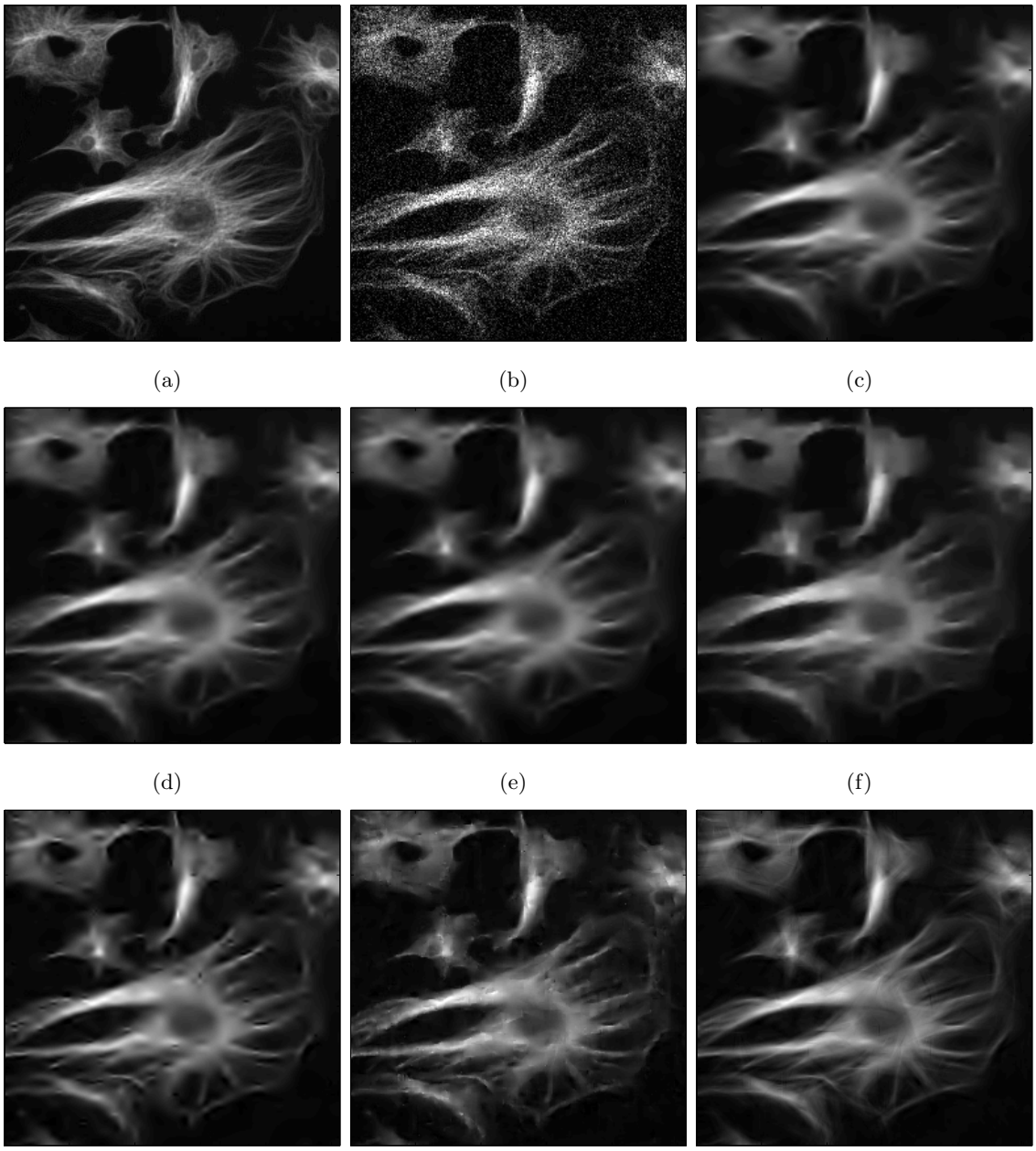
Fig. 5.11 compares the methods on a fluorescent tubulin filaments stained with Bodipy FL goat anti-mouse IgG[4]. The FDR control is employed in Anscombe (Fig. 5.11(c)), Haar-Fisz (Fig. 5.11(d)), CVS (Fig. 5.11(e)), Haar hypothesis tests (Fig. 5.11(f)), MS-VST+UWT (Fig. 5.11(g)), and MS-VST+Curvelet (Fig. 5.11(i)). MS-VST+UWT outperforms all the wavelet-based methods; among all the compared approaches, MS-VST+Curvelet leads to the best result both quantitatively and visually. For this example, we also evaluated the computation time of the tested methods on a 1.1GHz PC, giving: Anscombe (C++ codes, 4 sec), Haar-Fisz (C++ codes, 90 sec), CVS (Matlab codes, 3 sec), Haar HTs (C++ codes, 8 sec), MS-VST+UWT (C++ codes, 18 sec), Platelets (Matlab MEX codes, 2404 sec), MS-VST+Curvelet (Matlab codes, 1287 sec). This time benchmark shows that our MS-VST+UWT provides a fast solution among the wavelet-based estimators; MS-VST+Curvelet is more computationally intensive but is about one time faster than platelet denoising in our example.

5.4 Conclusion

In this chapter, we have introduced two novel Poisson denoisers. The first one is based on the hypothesis tests in the biorthogonal Haar domain, which is suitable for estimating regular intensities within real-time applications. We also proposed a more accurate threshold implementing individual hypothesis tests in the Haar/Bi-Haar domain. The second Poisson intensity estimator is based on a filter-adapted variance stabilization method and we have shown that it can be easily combined with various multiscale transforms such as the undecimated wavelet (isotropic and standard), the ridgelet and the curvelet transforms. This MS-VST approach enjoys the following advantages:

- It is efficient and sensitive in detecting faint features at a very low-count rate;
- We have the choice to integrate the VST with the multiscale transform we believe to be the most suitable for restoring a given kind of morphological features (isotropic, line-like, curvilinear, etc);

[4]The image is available on the ImageJ website `http://rsb.info.nih.gov/ij`



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 5.11: Poisson denoising of fluorescent tubulins (image size: 256×256). (a) intensity image (intensity $\in [0.53, 16.93]$); (b) Poisson noisy image; (c) Anscombe-denoised image (UWT, 7/9 filter bank, $J = 4$, FDR $= 0.1$, NMISE $= 0.095$); (d) Haar-Fisz-denoised image (UWT, 7/9 filter bank, $J = 4$, FDR $= 0.1$, 25 cyclic shifts where 5 for each of the axes, NMISE $= 0.096$); (e) CVS-denoised image (UWT, 7/9 filter bank, $J = 4$, FDR $= 0.1$, NMISE $= 0.10$); (f) denoised image by Haar+FDR (Haar UWT, $J = 4$, FDR $= 0.1$, NMISE $= 0.10$; (g) denoised image by MS-VST+UWT (UWT, 7/9 filter bank, $J = 4$, $N_{\max} = 5$ iterations, FDR $= 0.1$, NMISE $= 0.090$); (h) platelet-denoised image ($\gamma = 1/3$, 25 random cyclic shifts, NMISE $= 0.079$); (i) denoised image by MS-VST+Curvelets ($J = 4$, $N_{\max} = 5$ iterations, FDR $= 0.1$, NMISE $= 0.078$).

- The computation time is similar to that of a Gaussian denoising, which makes our denoising method capable of processing large data sets.

Comparison to competing methods in the literature show that the MS-VST is very competitive offering performance as good as state-of-the-art approaches, with low computational burden.

In all our presentation, a zero dark intensity λ_d has been assumed. In reality, λ_d is always positive. It can be estimated by calibrating the microscope camera. However, this value may vary from one experiment to another due to the temperature change. Thus, it may be more accurate to directly estimate in the data at hand by taking the average in a object-free zone. Then, in Algorithm 1, the estimate of λ_j should be modified to $\hat{\lambda}_j = \max(2^{cjq}a_j, 2^{jq}\lambda_d)$, and the positivity projection should be changed to $\hat{\Lambda} = \max(a_0, \lambda_d)$. In the MS-VST estimations (5.44), it suffices to replace the positivity constraint of the estimate $\mathcal{R}\mathbf{d} \geq 0$ by $\mathcal{R}\mathbf{d} \geq \lambda_d$.

Chapter 6

Mixed-Poisson-Gaussian Noise Removal

For a LSCM working under the analog mode, or a DSCM, the fluorescence images are corrupted by photon-counting and camera readout noises. Statistically, the data is modeled by a MPG process. For data restoration and object detection applications, it would be rather complicated to directly deal with such processes since the distribution of a MPG sample has an infinite Gaussian mixture structure. Many denoisers proposed in the literature are based on simplified assumptions such as a pure Gaussian or a pure Poisson process (see e.g. [118][112][119][120][39][121] and the references therein), which model only partially the noise statistics. To fully take into account the MPG process, a widely used approach is to apply the generalized Anscombe VST (GAT) [97] to preprocess the MPG data such that they become stabilized and Gaussianized. Standard restoration and detection algorithms derived from Gaussianity assumptions can then be applied on the transformed data. After processing the data in the transform domain, the VST is inverted if the data reconstruction is needed. GAT preprocessing has been successfully applied to fluorescence sequence restoration [122] when high photon flux are present. For low-count settings where the performace of GAT degrades, [122] proposed to use several Radon projections to increase the photon flux within a patch-based denoising algorithm [49][123].

In this chapter, we will extend the concept in Section 5.3 and propose a VST which is capable of stabilizing and Gaussianizing a filtered MPG process (Section 6.1). This transform can be seen as a generalization of GAT for filtered data. We then combine this

VST with the IUWT leading to a multiscale VST (MS-VST, Section 6.2) where the wavelet coefficients are stabilized and Gaussianized. We demonstrate the usefulness of the MS-VST for noise removal (Section 6.2) and for spot detection (Section 6.3) in confocal microscopy. In the first case, owing to the Gaussianized wavelet bands, significant wavelet coefficients are detected using classical multiple hypothesis tests under a false discovery rate (FDR) control. A sparsity-driven iterative scheme is applied to properly reconstruct the final estimate. In the second case, the Gaussianization provided by the VST allows us to directly apply a multiscale correlation based spot detector for Gaussian data [107]. Experiments show that 1) compared with the simplified assumption of a Gaussian or a Poisson noise, the MPG model is more realistic and results in a higher denoising performance and less false positives in the detection; 2) for MPG data with low Poisson intensities the MS-VST-based denoising and detection outperform those using GAT.

6.1 VST for a Filtered MPG Process

A fluorescence image is modeled by a MPG process $\mathbf{X} := (X_i)_{i\in\mathbb{Z}^q}$ where

$$X_i = \alpha U_i + V_i, \quad U_i \sim \mathcal{P}(\lambda_i), \quad V_i \sim \mathcal{N}(\mu, \sigma^2) \tag{6.1}$$

where $\alpha > 0$ is the overall gain of the microscope imaging system, U_i is a Poisson variable modeling the photon counting, V_i is a normal variable representing the camera readout noise, and all $(U_i)_i$ and $(V_i)_i$ are assumed mutually independent. Similar to the Poisson denoising case, it suffices to consider a zero dark intensity. Given a discrete filter h, we define a filtered MPG process as:

$$Y_i := \sum_{j\in\mathbb{Z}^q} h[j] X_{i-j} \tag{6.2}$$

We will use the index-free notations X and Y to denote any one of X_i and Y_i respectively. We recall that τ_k denotes the quantity $\sum_i (h[i])^k$ for $k = 1, 2, \cdots$.

We adopt a local stationarity assumption that, i.e., $\lambda_i = \lambda$ within the support of h. This will serve to derive the distribution under the null hypothesis in our detection step. The variance of Y, i.e., $\mathrm{Var}\,[Y] = (\alpha^2\lambda + \sigma^2)\tau_2$, is an affine function of the Poisson intensity λ. We seek a VST $T(Y)$ such that $\mathrm{Var}\,[T(Y)]$ is (asymptotically) constant, irrespective of the value of λ. Parallel to Lemma 2, Lemma 3 (proof in Appendix A.14) shows that

the square-root transform [Eq.(6.4)] can serve as a VST for stabilizing and Gaussianizing a filtered MPG process with a nonzero mean filter.

Lemma 3 (square root as VST) *If $\tau_1 \neq 0$, and τ_1, $\|h\|_2$, and $\|h\|_3$ are all finite, then we have*

$$T(Y) - b \cdot sgn(\tau_1)\sqrt{|\tau_1|\alpha\lambda} \xrightarrow[\lambda \to +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{\alpha b^2 \tau_2}{4|\tau_1|}\right) \tag{6.3}$$

where

$$T(Y) := b \cdot sgn(Y + c)|Y + c|^{1/2}, \quad b \neq 0, \ c \in \mathbb{R} \tag{6.4}$$

In (6.3), b is only a normalizing factor and (6.3) holds true for any real value of c. As the Poisson case, the convergence rate in (6.3) varies with the value of c, and we will determine its asymptotic optimal value.

6.1.1 Optimal parameter of the VST

Without loss of generality, we suppose that $\tau_1 > 0$. Then, $\Pr(Y + c > 0)$ can be made arbitrarily close to 1 as $\lambda \to +\infty$. Thus, in the asymptotic regime our VST will essentially take the form $T(Y) := bT_0(Y) := b\sqrt{Y + c}$. Expanding $T_0(Y)$ by Taylor series about the point $Y = \mathbb{E}[Y]$ up to the 4th order term, and by applying the expectation, one can show the asymptotic expectation and variance of $T(Y)$:

$$\mathbb{E}[b_1 T_0] \approx \sqrt{\lambda} + \underbrace{\frac{4\tau_1(\tau_1\mu + c) - \tau_2\alpha}{8\tau_1^2\alpha}}_{C_{\mathbb{E}}(c)} \lambda^{-1/2} \tag{6.5}$$

$$\mathrm{Var}\,[b_2 T_0] \approx 1 + \underbrace{\frac{8\tau_1^2\tau_2(\sigma^2 - \alpha\mu) - 4\tau_1\alpha(2\tau_2 c + \tau_3\alpha) + 7\tau_2^2\alpha^2}{8\alpha^2\tau_1^2\tau_2}}_{C_{\mathrm{Var}}(c)} \lambda^{-1} \tag{6.6}$$

where $b_1 = 1/\sqrt{\tau_1\alpha}$ and $b_2 = 2\sqrt{\tau_1}/\sqrt{\alpha\tau_2}$. These settings normalize respectively the asymptotic expectation and variance to $\sqrt{\lambda}$ and 1, both values being independent of the filter h. The value of c is chosen so that the variance residual term $O(\lambda^{-1})$ is canceled in (6.6), i.e.,

$$c^* := -\mu\tau_1 + \frac{\tau_1\sigma^2}{\alpha} + \frac{7\alpha\tau_2}{8\tau_1} - \frac{\alpha\tau_3}{2\tau_2} \tag{6.7}$$

[124] also proposed to compute the value by minimizing $C_{\mathbb{E}}^2 + |C_{Var}|$ (η set to 1/2 in [124]), which is a bias-variance trade-off criterion. However, experiments show that these different settings impact little the outputs.

This VST reduces to GAT if $h = \delta$ (Dirac filter, i.e., no filtering). Moreover, we can find appropriate filters such that the resulting VSTs are more efficient than GAT. As an example, Fig. 6.1 compares the mean $\mathbb{E}[b_1 T_0]$ and the variance $\mathrm{Var}\,[b_2 T_0]$ for GAT and for our VST with $h = \text{2D } B_3$-spline filter, both as functions of the underlying intensity. The plots are obtained from a Monte-Carlo simulation with 2×10^5 replications. Stabilizations with different values of the gain are tested. Other parameters are $\mu = 10$ and $\sigma = 1$. It is clear that as the intensity increases, the mean and the variance resulted from our VST stick to the asymptotic bounds much faster than those given by GAT, particularly as α is large.

In practice, we need to estimate μ, σ, and α if they are unknown a priori. As the MPG distribution has an infinite mixture structure, the classical expectation-maximization (EM) estimator has no closed forms. A faster and simpler approach is to match the first four cumulants of a MPG process with the k-statistics [125] of the observations in a uniform image background region. This follows from the fact that the k-statistics are the minimum variance unbiased estimators for cumulants. Computational details are shown in Appendix B. Moment matching can also be applied [126]. We should point out that these approaches are not robust against outliers. Robust estimation of MPG parameters is one of the future perspectives to improve our method.

6.2 Fluorescence Image Denoising Using MS-VST

We can now combine the VST with a large family of multiscale transforms implemented by iterative filters, leading to MS-VSTs. Isotropic structures are often presented in biological fluorescence images due to micrometric subcellular sources (marked genes, vesicles, etc.). In this chapter we will concentrate on these sources and combine the proposed VST with the IUWT, which adapts very well the isotropic features in images. For line-like and curvilinear sources, ridgelet and curvelet transforms can be used. The combination of the VST with these transforms follows the same way as in the pure Poisson noise case (see Chapter 5).

The MPG noise removal by MS-VST involves the same three steps as in the Poisson case. First, we compute the VST-combined IUWT. The wavelet bands will be shown to be stabilized and Gaussianized with known variances (Section 6.2.1). Second, we detect significant coefficients by applying classical hypothesis tests (Section 6.2.2). Finally, the image estimate is reconstructed by inverting the VST-combined IUWT which is done iteratively

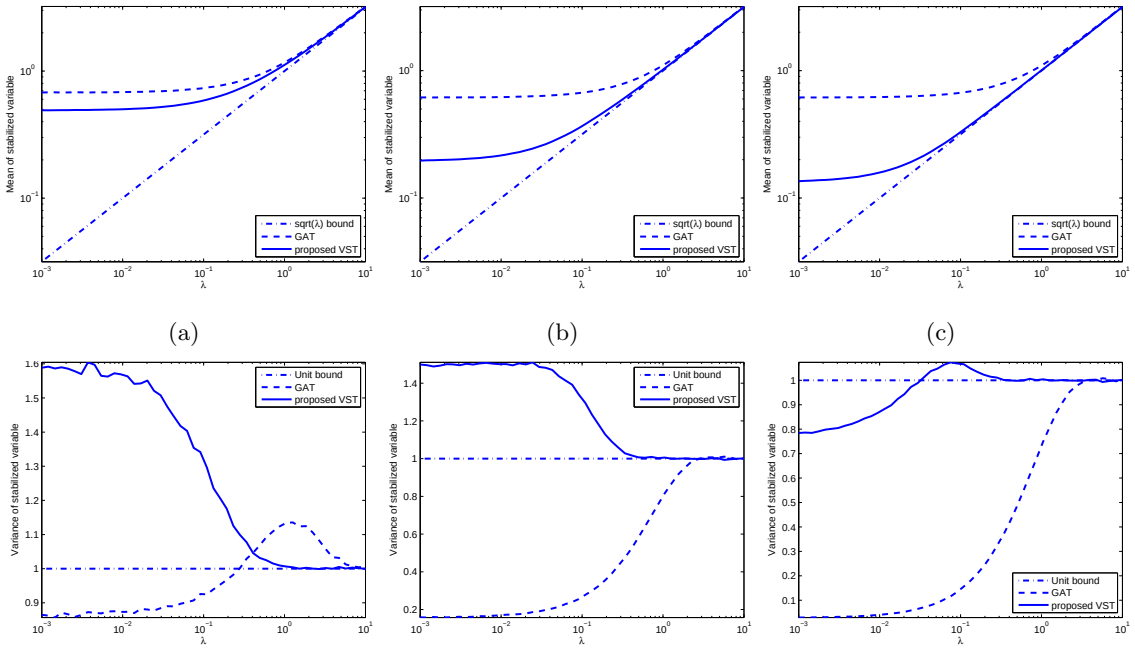


Fig. 6.1: Stabilized mean and variance as functions of intensity λ, for GAT and for the proposed VST with h = 2D B_3-spline filter ($= h_{B_3} \otimes h_{B_3}$, $h_{B_3} = [1, 4, 6, 4, 1]/16$, where $\otimes$ represents the tensor product). Other parameters are: $\mu = 10$ and $\sigma = 1$. (a) $\mathbb{E}[b_1 T_0]$ ($\alpha = 2$); (b) $\mathbb{E}[b_1 T_0]$ ($\alpha = 5$); (c) $\mathbb{E}[b_1 T_0]$ ($\alpha = 10$); (d) $\mathrm{Var}\,[b_2 T_0]$ ($\alpha = 2$); (e) $\mathrm{Var}\,[b_2 T_0]$ ($\alpha = 5$); (f) $\mathrm{Var}\,[b_2 T_0]$ ($\alpha = 10$).

(Section 6.2.3). For the clarity of presentation, we assume 1D data, but the results hold indeed in any dimensional case.

6.2.1 IUWT-based MS-VST

The VST is coupled with IUWT in the usual way:

$$\left\{\begin{array}{lcl} a_j &=& \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &=& a_{j-1} - a_j \end{array}\right. \Rightarrow \left\{\begin{array}{lcl} a_j &=& \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &=& T_{j-1}(a_{j-1}) - T_j(a_j) \end{array}\right. \tag{6.8}$$

where $a_0 := \mathbf{X}$ is the original MPG data in the form of (6.1). T_j is the VST operator at scale j (cf. Eq.(6.4)):

$$T_j(a_j) = b^{(j)} \operatorname{sgn}(a_j + c^{(j)})|a_j + c^{(j)}|^{1/2}$$

The constants $b^{(j)}$ and $c^{(j)}$ are associated to the filter $h^{(j)}$, and $c^{(j)}$ is set to c^*. Parallel to Theorem 3, Theorem 6 (proof in Appendix A.15) shows the asymptotic stabilized Gaussianity of the wavelet coefficients resulted from the MS-VST scheme.

Theorem 6 (Stabilized Gaussianity of d_j) *Setting $b^{(j)} := sgn(\tau_1^{(j)})/\sqrt{\alpha|\tau_1^{(j)}|}$, if λ is constant within the support of the filter $h^{(j)}[k-\cdot]$, we have that $d_j \xrightarrow[\lambda\to+\infty]{\mathcal{D}} \mathcal{N}(0,\sigma_j^2)$, where*

$$\sigma_j^2 := \frac{\tau_2^{(j-1)}}{4[\tau_1^{(j-1)}]^2} + \frac{\tau_2^{(j)}}{4[\tau_1^{(j)}]^2} - \frac{\langle h^{(j-1)}, h^{(j)}\rangle}{2\tau_1^{(j-1)}\tau_1^{(j)}} \tag{6.9}$$

where $\tau_k^{(j)} := \sum_i (h^{(j)}[i])^k$. This result shows that the asymptotic variance of d_j depends only on the wavelet filter bank and the current scale, and thus can be pre-computed once h is chosen.

6.2.2 Detection of significant coefficients by FDR

Wavelet denoising is achieved by testing binary hypothesis: $\forall\, d_j[l],\ H_0 : d_j[l] = 0$ vs. $H_1 : d_j[l] \neq 0$. The distribution of d_j under the null hypothesis H_0 is provided by Theorem 6. We apply the multiple hypothesis testing controlling the FDR, i.e., Benjamini-Hochberg procedure. We use $\hat{d}_j$ to denote the thresholded wavelet coefficients.

6.2.3 Sparsity-driven iterative reconstruction

We formulate the following ℓ^1 minimization problem.

$$\begin{aligned} &\min_{\mathbf{d}\in\mathcal{C}} J(\mathbf{d}) := \|\mathbf{d}\|_1 \text{ where } \mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2 \\ &\mathcal{S}_1 := \{\mathbf{d} | d_j[k] = (\mathcal{W}\mathbf{X})_j[k], (j,k) \in \mathcal{M}\} \text{ and } \mathcal{S}_2 := \{\mathbf{d} | \mathcal{R}\mathbf{d}[l] \geq \mu, \forall\, l\} \end{aligned} \tag{6.10}$$

Only the constraint of $\mathcal{S}_2$ is changed compared with the Poisson case. Here, $\mathcal{S}_2$ assures a model-consistent estimate since $\mathbb{E}[X_i] = \alpha\lambda_i + \mu \geq \mu$. Then, this problem is solved by HSD iterations [111].

$$\mathbf{d}^{(k)} := T_{\mathcal{C}}\mathbf{d}^{(k-1)} - \beta_k \operatorname{sgn}\left(T_{\mathcal{C}}\mathbf{d}^{(k-1)}\right) \tag{6.11}$$

$\mathbf{d}^{(0)}$ can be initialized by the wavelet coefficients of the direct MS-VST inverse. The operator $T_{\mathcal{C}}$ is defined as $T_{\mathcal{C}} := P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2}$, where

$$(P_{\mathcal{S}_1}\mathbf{d})_j[k] := \begin{cases} (\mathcal{W}\mathbf{X})_j[k] & (j,k) \in \mathcal{M} \\ \mathbf{d}_j[k] & \text{otherwise} \end{cases} \quad ; \quad Q_{\mathcal{S}_2}\mathbf{d} := \mathcal{W}P_\mu\mathcal{R}\mathbf{d}$$

Here, $P_\mu(\mathbf{x})[l] := \max(\mathbf{x}[l], \mu)$. The final image estimate is given by $\hat{a}_0 = P_\mu\mathcal{R}\mathbf{d}^{(N_{max})}$ where N_{max} is the maximum number of iteration. Algorithm 6 summarizes the whole MPG-denoising steps.

Algorithm 6 MPG noise removal by MS-VST

Require: $a_0 := \mathbf{X}$; a low-pass filter h; wavelet-domain FDR bound γ_w

Detection

1: **for** $j = 1$ to J **do**
2: Compute a_j and d_j using the MS-VST [Eq.(6.8)].
3: Apply the Benjamini-Hochberg tests on d_j to get $\hat{d}_j$ and $\mathcal{M}$.
4: **end for**

Estimation

5: Direct MS-VST inverse: $\hat{a}_0^* := T_0^{-1}[\sum_{j=1}^{J}\hat{d}_j + T_J(a_J)]$
6: Initialize $\mathbf{d}^{(0)} = \mathcal{W}P_\mu\hat{a}_0^*$
7: **for** $k = 1$ to N_{max} **do**
8: $\mathbf{d}^{(k)} := \mathrm{ST}_{\beta_k}[P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2}\mathbf{d}^{(k-1)}]$.
9: **end for**
10: Get the estimate $\hat{a}_0 = P_\mu\mathcal{R}\mathbf{d}^{(N_{max})}$.

6.2.4 Results

In all our experiments, the low-pass filter h in the IUWT is set to the 2D B_3-spline filter.

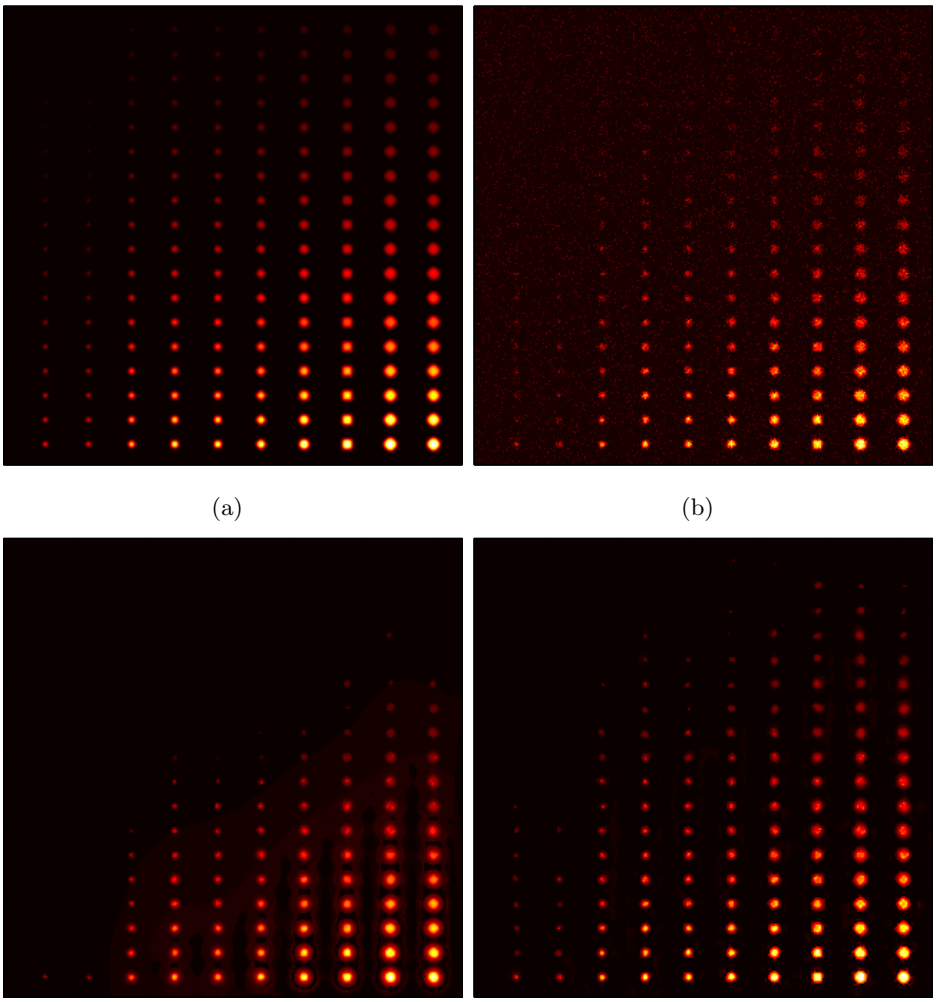
6.2.4.1 MS-VST vs. GAT

We first test our denoising approach on a simulated 18×10 isotropic-source grid (pixel size = 100 nm) shown in Fig. 6.2. From the leftmost to the rightmost column, the source radii increase from $50nm$ to $350nm$. The image is then convolved with a 2D Gaussian function with a standard deviation $\sigma_g = 103nm$, which simulates a typical LSCM PSF. Fig. 6.2(a) shows the sources with amplitudes $\lambda_A \in [0.05, 30]$. After adding a MPG noise ($\alpha = 20, \mu = 10, \sigma = 1$), we obtain Fig. 6.2(b). Fig. 6.2(c) and (d) respectively show the denoising examples using GAT and MS-VST. The controlled FDR is set to 10^{-2}. We can see that MS-VST is more sensitive than GAT since more faint sources are restored.

The performance of the two methods is also quantified using the NMISE measure, i.e., NMISE $:= \mathbb{E}[(\sum_{i=1}^{N}(\widehat{\mathbb{E}(\mathbf{X})}_i - \alpha\lambda_i - \mu)^2/(\alpha^2\lambda_i + \sigma^2))/N]$. Based on 10 replications, the NMISE for GAT and for MS-VST approaches are evaluated to be 0.19 and 0.05 respectively. This shows that the MS-VST denoising is more accurate.

6.2.4.2 Denoising comparison with different noise assumptions

Second, our method is applied on real fluorescence confocal images where the necessity of the MPG noise model is illustrated. Fig. 6.3(a) shows one confocal slice of a drosophila melanogaster ovary. The part of nurse cells consist of many nucleus and GFP (Green Fluorescent Protein)-marked Staufen genes. These genes have isotropic intensity profiles. This image is then denoised by different denoisers whose essential difference is in the noise assumption. Fig. 6.3(b) results from the IUWT denoising assuming a Gaussian noise. The noise variance is pre-estimated using the median absolute deviation (MAD) estimator in the domain of Daubechies wavelet with 4 vanishing moments. Fig. 6.3(c) is given by the MS-VST denoiser derived for Poisson noise (Chapter 5). Fig. 6.3(d) is yielded by our proposed approach which is adapted for MPG noise. The MPG parameters are estimated by cumulant matching. The local zooms of these denoising results are presented in Fig. 6.3(e)-(h). The same FDR bound (FDR $= 10^{-2}$) is set for wavelet coefficient detection in all the denoisers. It is clear that only our method assuming a MPG noise outputs an almost noise-free image, while residual noises persist in the other results. It is also interesting to notice that the

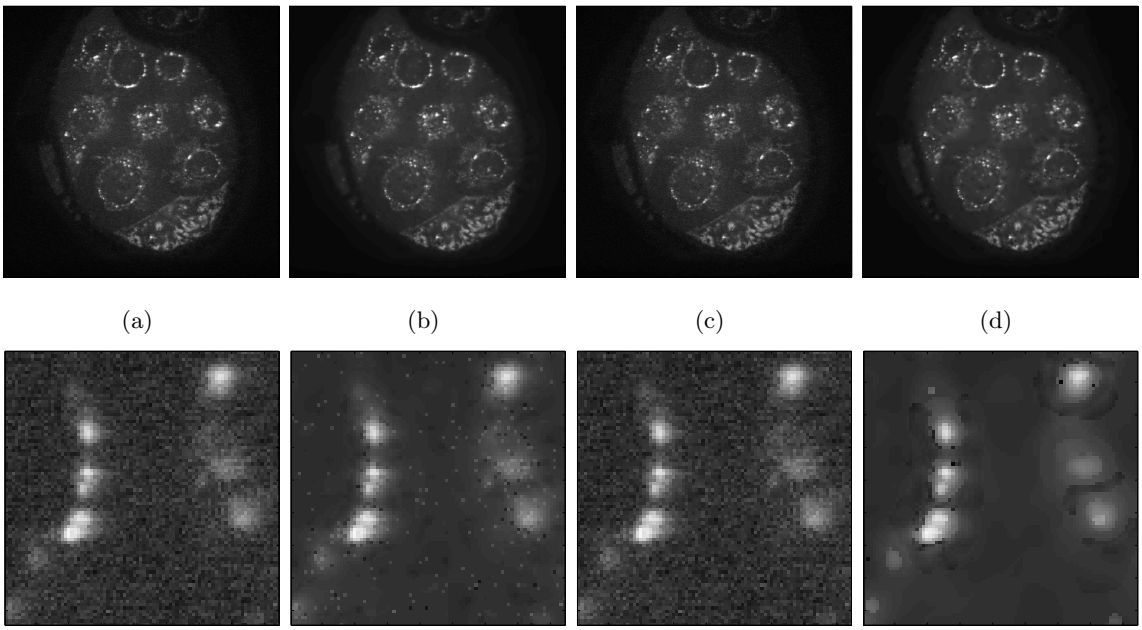


(a) (b)

(c) (d)

Fig. 6.2: Simulated source denoising using IUWT. h = 2D B_3-spline filter, $J = 5$, and FDR = 0.01. (a) simulated sources (amplitudes $\lambda_A \in [0.05, 30]$; background = 0.05); (b) MPG noisy image ($\alpha = 20, \mu = 10, \sigma = 1$); (c) GAT denoised image, NMISE = 0.19; (d) MS-VST denoised image ($N_{max} = 10$ iterations), NMISE = 0.05.

Gaussian noise model leads to an estimate with less residual noise than the Poisson model.



(a) (b) (c) (d)

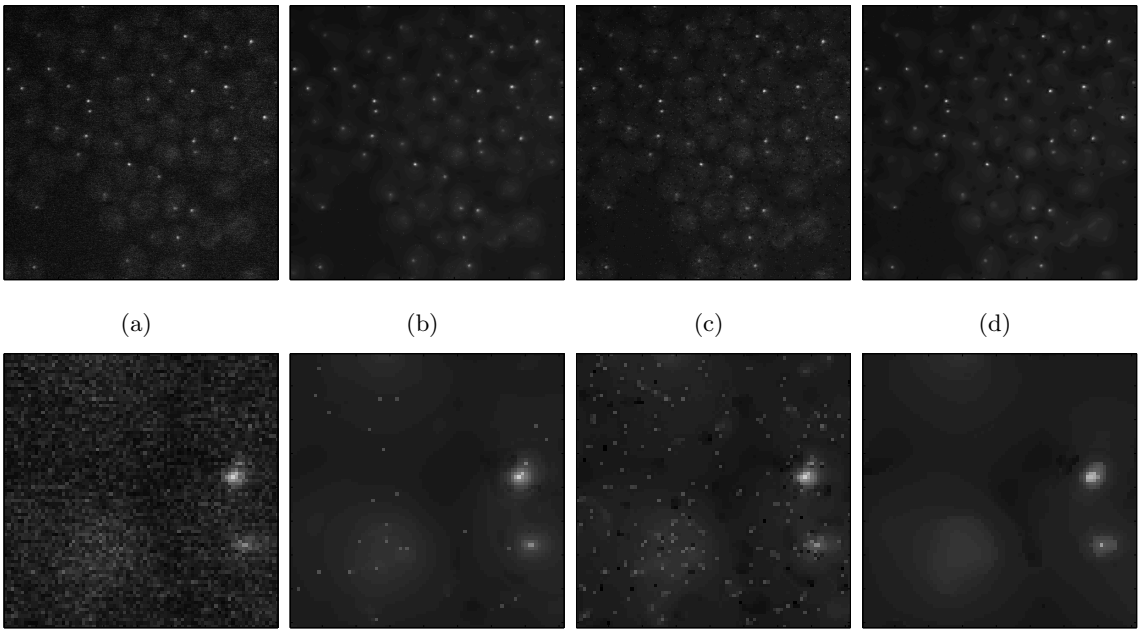
(e) (f) (g) (h)

Fig. 6.3: IUWT denoising of a drosophila melanogaster ovary image (estimates $\hat{\alpha} = 178.6$, $\hat{\mu} = 981.0$, $\hat{\sigma} = 33.8$). $h =$ 2D B_3-Spline filter, $J = 5$, and FDR $= 10^{-2}$. (a) observed image; (b) denoised image assuming a Gaussian noise; (c) denoised image assuming a Poisson noise (Chapter 5, $N_{max} = 10$ iterations); (d) image denoised by the proposed method assuming a MPG noise ($N_{max} = 10$ iterations); (e) local zoom of (a); (f) local zoom of (b); (g) local zoom of (c); (h) local zoom of (d).

We repeat this comparison on an image of saccharomyces cerevisiae cells acquired by a DSCM (Fig. 6.4). In each cell, a subtelomere is marked which exhibits a spot structure. The same denoising performances can be observed as for Fig. 6.3.

6.3 Detection of Fluorescent Spots Using MS-VST

Fluorescent bright spots are formed by marked nanometric cellular or subcellular components. Information such as their number and their locations can be important for the study of the spatial distribution and the dynamics of these objects [127][128][129].



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 6.4: IUWT denoising of an image of saccharomyces cerevisiae cells (estimates $\hat{\alpha} = 93.4$, $\hat{\mu} = 1016.5$, $\hat{\sigma} = 26.6$). $h =$ 2D B_3-Spline filter, $J = 5$, and FDR $= 10^{-2}$. (a) observed image; (b) denoised image assuming a Gaussian noise; (c) denoised image assuming a Poisson noise (Chapter 5, $N_{max} = 10$ iterations); (d) image denoised by the proposed method assuming a MPG noise ($N_{max} = 10$ iterations); (e) local zoom of (a); (f) local zoom of (b); (g) local zoom of (c); (h) local zoom of (d).

Difficulties as a low SNR in microscopy images and complex nonuniform backgrounds underly spot detections. Various spot extractors have been proposed in the literature to suppress the influence of noise and background. For example, in the context of detecting telomeres [127] deconvolves the image by assuming a Poisson noise followed by a top-hat filter to eliminate the background shading. Then the spots are extracted by thresholds. [130] proposed to partition the image using quad-tree into regions of small intensity variation, and apply a Fuzzy c-means clustering procedure in each subimage to find spots (peroxisomes). An aspect-ratio criterion is also introduced to separate close spots. If the number of spots is small, least-square fit using a mixture of Gaussians is also widely used [36][131][38]. The underlying hypothesis are that the objects are small compared with the PSF, the background can be described by a simple parametric model (e.g., uniform, linear, etc.), and the photon flux is high such that the Poisson noise becomes asymptotically normal. In the video-microscopy context, [122] proposed to detect vesicles by either applying a model-selection based threshold or a *a contrario* decision in the image where the background has been subtracted. The background is estimated by fitting a linear temporal model into the denoised video sequence [49][123][122]. In [107], significant positive isotropic wavelet coefficients across several selected scales are correlated (i.e., multiplied). This is based on the observation that the local maxima of the wavelet coefficients tend to propagate across several scales if they are due to the spots while they are not if due to noise. The significant wavelet coefficients are detected by wavelet thresholds for Gaussian noise. The scales are selected according to the size of the interested spots. Regular backgrounds are easily eliminated owing to the vanishing moments of the wavelet. The correlated image is thresholded and binarized, the connected components are determined as putative bright spots, and their centroids are computed as spot locations.

Here, thanks to the MS-VST, we can generalize the IUWT-based detector [107] to the MPG case in a straightforward way. We first derive the multiresolution support $\mathcal{M}$ by making hypothesis tests in the stabilized wavelet domain, and then multiply the wavelet bands with positive significant coefficients of the original image. The correlated wavelet band is then binarized by a threshold $T \geq 0$. This procedure is summarized in Algorithm 7 whose output d_p is a binary image. In all our experiments T is set to 0, i.e., all positive parts in the correlated band are retained as detections.

With this approach, the false detection rate in a spot-free noisy image can be controlled:

Algorithm 7 Spot detection by MS-VST

Require: $a_0 := \mathbf{X}$; a low-pass filter h; wavelet-domain FDR; indices of bands to be multiplied I_s; binarizing threshold $T \geq 0$

1: **for** $j = 1$ to J **do**
2: Compute a_j and d_j using the MS-VST [Eq.(6.8)].
3: Apply the Benjamini-Hochberg tests on d_j to get $\mathcal{M}$.
4: **end for**
5: **for** $j = 1$ to J **do**
6: Compute a_j and d_j using the classical IUWT [Eq.(6.8)].
7: $d_j[l] = 0$ if $d_j[l] < 0$ or $(j, l) \notin \mathcal{M}$.
8: **end for**
9: $d_p := \Pi_{j \in I_s} d_j$
10: $\forall\, l, d_p[l] = 1$ if $d_p[l] > T$, and otherwise $d_p[l] = 0$.

Proposition 8 *The probability of erroneously detecting spots in an object-free homogeneous MPG noisy image ($\lambda_i \equiv \lambda$) is upper bounded by the controlled FDR in the wavelet domain.*

The proof of this proposition is very simple: as all detections in pure noise are false discoveries, the FDR [Eq.(5.13)] reduces to the global probability of false detection, from which this proposition follows.

6.3.1 Results

6.3.1.1 MS-VST vs. GAT

Although Proposition 8 assures a false positive rate control in a pure noisy image, it is more interesting to quantify the detection sensibility and specificity when both signal and noise are present. For this purpose, 100 detection replications were made on the simulated spot grid Fig. 6.5(a) (the same as Fig. 6.2(a)) with different noise realizations. We set $I_s = \{3, 4\}$, and we compare the detection performances resulted from the use of GAT and of the MS-VST. A connected component is considered as a good detection if its centroid falls into the support of one of the spots. Multiple good detections in the same spot support will be counted as one good detection only. Now, Fig. 6.5(d) shows the true positive rates (TPRs) as functions of FDR $\in [10^{-9}, 10^{-1}]$ for the two methods. The maximal measured FDR of the detections are very low in both methods: 0% and 0.38% respectively for GAT

and for the MS-VST. This shows their good specificity against noise. However, we can see that the detection using the MS-VST always leads to a sensibility of about 20% higher than GAT. Typical detection examples by GAT and the MS-VST are shown in Fig. 6.5(b) and Fig. 6.5(c) respectively. The circles indicate the true spot locations, and the crosses indicate the centroids of the detections. Visually, the MS-VST detects more spots owing to its higher sensibility to faint objects.

6.3.1.2 Detection comparison with different noise assumptions

We compare the spot extractions using different noise assumptions in the image of saccharomyces cerevisiae cells where the subtelomeres are objects of interest (Fig. 6.6(a)). The detection settings are $I_s = \{3, 4\}$ and $\text{FDR} = 10^{-2}$. Fig. 6.6(b), Fig. 6.6(c) and Fig. 6.6(d) are results yielded by the detection using a Gaussian noise assumption, a Poisson noise assumption, and a MPG noise assumption. Different noise modeling leads to different multiresolution supports, and thus to different detections. The biological specimen preparation ensures that each cell in the image contains at most one spot. Therefore, multiple detections in one cell are due to false positives. The cells can be seen in Fig. 6.6(a) which are bulb-shaped objects holding the subtelomeres. The results show that all the three detectors correctly extract almost all the visible spots, but the detectors based on Gaussian or Poisson noise assumptions result in more false positives.

Furthermore, cellular structures in the image background are not detected. Indeed, as wavelets are band-pass filters, background information is mostly encoded in the approximation band. Since the spot detection are realized in the wavelet domain, regular backgrounds are successfully suppressed.

6.4 Conclusion

The noise in LSCM and DSCM images is modeled by a MPG process. We have designed a VST to stabilize and Gaussianize a filtered MPG process. This VST is then combined with the IUWT yielding the MS-VST. We have shown the effectiveness of the MS-VST approach in removing MPG noise and in extracting spots from nonuniform backgrounds. Our experiments show that 1) compared with standard denoising methods adopting the simplified assumption of a Gaussian or a Poisson noise, the MPG model is more realistic and results in a higher denoising performance and less false positives in the detection; 2)

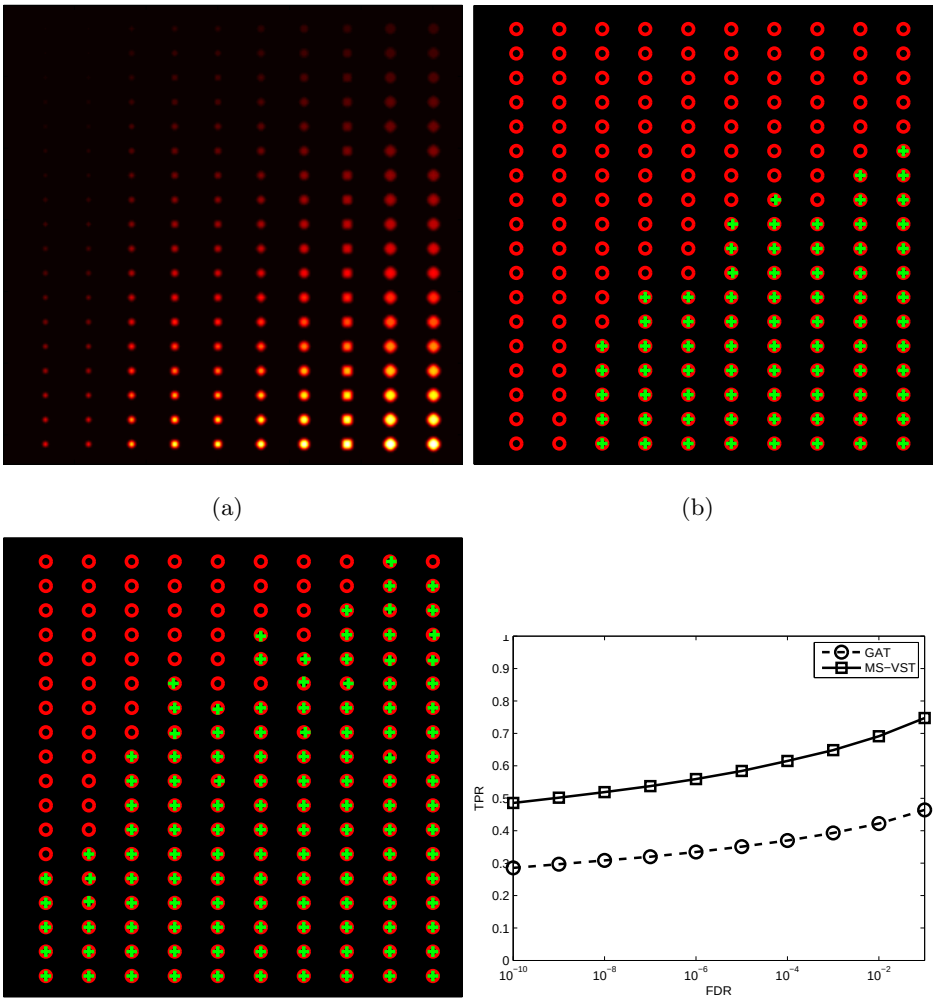
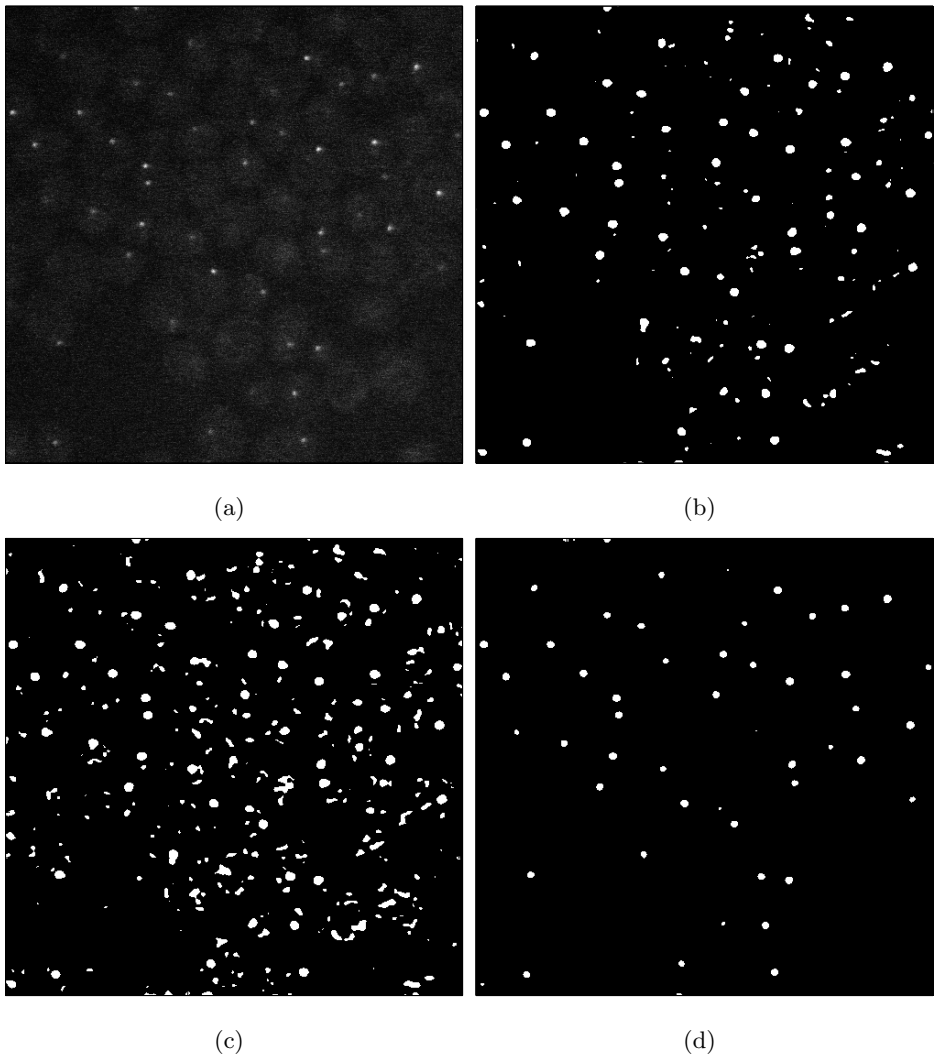


Fig. 6.5: Detection of simulated spots. MPG parameters: $\alpha = 20, \mu = 10, \sigma = 1$; detection settings: $h =$ 2D B_3-spline filter, and $I_s = \{3, 4\}$. (a) simulated sources (amplitudes $\lambda_A \in [0.05, 30]$; background $= 0.05$); (b) example of the GAT-based detection (FDR$= 10^{-2}$; 75 good detections among 180, 0 false detection); (c) example of the MS-VST-based detection (FDR$= 10^{-2}$; 125 good detections among 180, 0 false detection); (d) TPRs as functions of FDR$\in [10^{-9}, 10^{-1}]$.



(a) (b)

(c) (d)

Fig. 6.6: Detection of subtelomeres of saccharomyces cerevisiae cells. h = 2D B_3-Spline filter, FDR = 10^{-2}, and $I_s = \{3, 4\}$. (a) observed image; (b) extracted spots with a Gaussian noise assumption; (c) extracted spots with a Poisson noise assumption (Chapter 5); (d) extracted spots with a MPG noise assumption.

for MPG data with low Poisson intensities the MS-VST-based denoising and detection outperform those using the GAT.

Part III

Super-resolution Object Detection

Chapter 7

Super-resolution Spot Detection

Classical 2D WFFM resolution limit is predicted by the Rayleigh criterion. Suppose that we have two point sources generating two Airy patterns. If the first zero of one Airy image superposes the peak of the other, then the distance between these two Airy patterns is defined as the Rayleigh distance, which is given by $d_R := 0.61\lambda_{em}/\mathrm{NA}$. Rayleigh states that any two point sources separated with a lateral distance smaller than d_R are "unresolvable". Clearly, the limiting distance d_R is formulated in a deterministic way. It does not take the statistical nature of fluorescence imaging into account. For example, the influence of SNR on the source-detection performance is not reflected.

In this chapter, we will first briefly review in Section 7.1 recent studies on the limiting resolution in detection-theoretic, estimation-theoretic and information-theoretic points of view. Our main contribution to this subject is that we propose to use the VST to study the resolution with Poisson or MPG observations in these frameworks. This approach significantly simplifies the analysis by Gaussianizing the problem, and leads to asymptotically consistent results with closed-form expressions. The localization accuracy limits are also studied. The main conclusion is that the resolution is not limited, and Rayleigh's limit can indeed be broken if the SNR in the image is sufficiently high. Then, in Section 7.2 we extends an existing super-resolution approach which combines parametric model fitting and model-order selection for practically determining an unknown number of spots in the image.

7.1 Resolution Limit

Below, to simplify the presentation, only the lateral resolution will be considered in the detection and estimation-theoretic frameworks.

7.1.1 Detection-theoretic resolution limit

The resolution limit has been studied in a statistical decision framework [132][133][134][135][136]. To resolve two point sources, we test a binary hypothesis, i.e., H_0 : one source is present vs. H_1 : two sources are present. The useful information we want to know is the relationship between the smallest detectable source separation under a given FPR, a given SNR, and a given minimum true positive rate (TPR).

7.1.1.1 Gaussian observations

Below, we follow the presentation of [135][136]. To clarify the idea, let us consider an ideal situation where we observe an image $\mathbf{y} = (y(\mathbf{x}_n))_n$ of two symmetrically placed point sources along the x-axis with a lateral separation d, i.e.,

$$y(\mathbf{x}_n) = f(\mathbf{x}_n; d) + w(\mathbf{x}_n),\ n = 1, \ldots, N \tag{7.1}$$

$$f(\mathbf{x}_n; d) = A \cdot h(x_n - d/2, y_n, z_n) + A \cdot h(x_n + d/2, y_n, z_n) \tag{7.2}$$

where $\mathbf{x}_n := [x_n, y_n, z_n]^T$ is a sample point, N is the number of pixels, $w(\mathbf{x}_n) \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian noise process, A is the amplitude of the point sources, h is the PSF, and d is the source separation which is assumed to be the only unknown parameter. Note that if the pixelation is considered, the function h is the effective PSF, i.e., the convolution of the optical PSF and the pixel indicating function $1_{\mathbf{C}}$ which takes value 1 within a pixel support $\mathbf{C}$ and value zero elsewhere. The binary hypothesis test can now be equivalently formulated as $H_0 : d = 0$ vs. $H_1 : d > 0$, which can be again rewritten as

$$\begin{cases} H_0 : & y(\mathbf{x}_n) = f_0(\mathbf{x}_n) + w(\mathbf{x}_n), \forall\, n \\ H_1 : & y(\mathbf{x}_n) = f(\mathbf{x}_n; d) + w(\mathbf{x}_n), \forall\, n \end{cases} \tag{7.3}$$

where $f_0(\mathbf{x}_n) = 2Ah(\mathbf{x}_n)$. This is a composite hypothesis testing problem which is typically solved by the general likelihood ratio test (GLRT) procedure. That is, we first carry out the ML estimation of d, then use this estimated value to form the standard Neyman-Pearson test. This detection problem is highly nonlinear and a second order Maclaurin

approximation for f around $d = 0$ is employed to linearize the problem, i.e.,

$$f(\mathbf{x}_n; d) \approx 2Ah(\mathbf{x}_n) + \frac{A}{4} h_{11}(\mathbf{x}_n) d^2$$

where $h_{11}(\mathbf{x}_n) = \frac{\partial^2 h}{\partial x^2}(\mathbf{x}_n)$. As we will see, by doing so, a closed-form expression can be derived for the minimum separation. This approximation is reasonable since we are interested in small d only. Note that $2Ah(\mathbf{x}_n)$ is known to the detector and is independent of d. Thus, the test can be further simplified as

$$\begin{cases} H_0: & z(\mathbf{x}_n) = w(\mathbf{x}_n), \forall\, n \\ H_1: & z(\mathbf{x}_n) = \frac{A}{4} h_{11}(\mathbf{x}_n) d^2 + w(\mathbf{x}_n), \forall\, n \end{cases} \tag{7.4}$$

where $z(\mathbf{x}_n) = y(\mathbf{x}_n) - 2Ah(\mathbf{x}_n)$. It can be easily verified that the GLRT statistic is given by $T(\mathbf{z}) = (\mathbf{h}_{11}^T \mathbf{h}_{11})^{-1/2} \mathbf{h}_{11}^T \mathbf{z}$, where $\mathbf{h}_{11} = [h_{11}(\mathbf{x}_1), \ldots, h_{11}(\mathbf{x}_N)]^T$ and $\mathbf{z} = [z(\mathbf{x}_1), \ldots, z(\mathbf{x}_N)]^T$. It can be shown that this test is uniformly most powerful (UMP) in the sense that it produces the highest detection probability for all values of d under a given FPR. Note that $T(\mathbf{z})$ is normally distributed under H_0 and H_1, and it follows that

$$d \geq \frac{2}{\sqrt{A/\sigma}(\mathbf{h}_{11}^T \mathbf{h}_{11})^{1/4}} \left[\Phi^{-1}(1 - p_f) - \Phi^{-1}(1 - p_d)\right]^{1/2} \tag{7.5}$$

where Φ is the standard normal *cdf*, p_f and p_d are respective the FPR and the minimum TPR to attain. If we define the SNR to be the ratio SNR $:= A/\sigma$, then $d \propto \text{SNR}^{-1/2}$. A more comprehensive analysis of the minimum detectable separation for asymmetric sources, under-Nyquist sampling (aliased image), and PSF mismatch situations can be found in [135][136].

7.1.1.2 Poisson and MPG observations

To our knowledge, the limiting resolution has not been studied in the detection-theoretic framework with Poisson and MPG observations. Here, we propose to apply the VST (6.4) on the data, which significantly facilitates the analysis by bringing the problem to the Gaussian case, and leads to closed-form expressions. Clearly, this approach is not valid for (very) low-count images where the Poisson intensity level falls out of the effective domain of the VST.

We specify the MPG image model

$$\begin{aligned} y(\mathbf{x}_n) &= \alpha F(\mathbf{x}_n; d) + w(\mathbf{x}_n), \quad F(\mathbf{x}_n; d) \sim \mathcal{P}(f(\mathbf{x}_n; d)), w(\mathbf{x}_n) \sim \mathcal{N}(\mu, \sigma^2) \\ f(\mathbf{x}_n; d) &= A \cdot h(x_n - d/2, y_n, z_n) + A \cdot h(x_n + d/2, y_n, z_n) \end{aligned} \tag{7.6}$$

(7.6) reduces to a Poisson model if $\alpha = 1$ and $\mu = \sigma = 0$. Direct application of GLRT will not result in a closed-form expression for the test statistic due to the MPG distribution. However, by applying the VST we can treat the observations as if they were issued from a Gaussian process, i.e.,

$$T(y(\mathbf{x}_n)) = 2\sqrt{f(\mathbf{x}_n; d)} + w_T(\mathbf{x}_n) \tag{7.7}$$

where $T(x) = 2/\sqrt{\alpha}\,\mathrm{sgn}(x + c^*)|x + c^*|^{1/2}$, and $w_T(\mathbf{x}_n) \sim \mathcal{N}(0, 1)$. No information is lost since T is invertible.

Now using the second order series approximation on the transformed signal $2\sqrt{f(\mathbf{x}_n; d)}$, our test becomes

$$\begin{cases} H_0: & z(\mathbf{x}_n) = w_T(\mathbf{x}_n), \forall\, n \\ H_1: & z(\mathbf{x}_n) = \frac{\sqrt{2A}}{8\sqrt{h(\mathbf{x}_n)}} h_{11}(\mathbf{x}_n) d^2 + w_T(\mathbf{x}_n), \forall\, n \end{cases} \tag{7.8}$$

Similar calculus as in the Gaussian case shows that

$$d \geq \frac{2 \cdot 2^{1/4}}{A^{1/4} \left[\sum_{n=1}^{N} h_{11}(\mathbf{x}_n)^2 / h(\mathbf{x}_n)\right]^{1/4}} \left[\Phi^{-1}(1 - p_f) - \Phi^{-1}(1 - p_d)\right]^{1/2} \tag{7.9}$$

According to the MPG statistics, we can define $\mathrm{SNR} := \alpha A / \sqrt{\alpha^2 A + \sigma^2}$. (7.9) shows that $d \propto A^{-1/4}$, or $d \propto \mathrm{SNR}^{-1/2}$.

As we have mentioned, (7.9) is valid for high enough Poisson intensities only. In particular, if the readout noise level increases, higher intensity is required to correctly stabilize the data. As an example, for the setting $\alpha = 10$ and $\mu = 1$, Fig. 7.1 shows the minimum intensity required by the VST as a function of the readout noise standard deviation. The curve is obtained from Monte-Carlo simulations. The MPG data are considered well stabilized as long as the mean and the variance of the VST-transformed data are within the $\pm 2.5\%$ range of the theoretic asymptotic values.

7.1.1.3 Some extensions

Multivariate case Up to now, we have supposed that the only unknown parameter is the source separation. A more realistic signal model is given by

$$f(\mathbf{x}_n; \theta) = A_1 \cdot h(x_n + d_1, y_n + d_2, z_n + d_3) + A_2 \cdot h(x_n + d_4, y_n + d_5, z_n + d_6)$$

where $\theta = [A_1, A_2, d_1, \ldots, d_6]$ are unknown parameters. A second order Maclaurin approximation on f around $(d_1, \ldots, d_6) = \mathbf{0}$ will lead to a linear model detection problem, and the

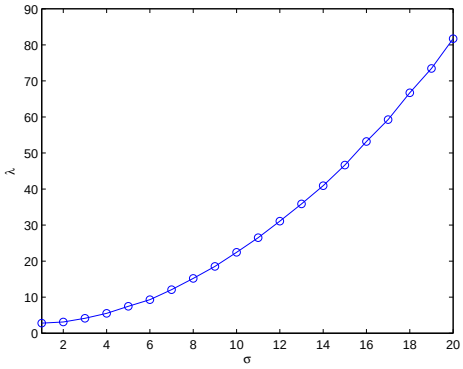


Fig. 7.1: Minimum Poisson intensity (λ) required to correctly stabilize the MPG data as a function of the readout noise standard deviation (σ). Other MPG parameters are $\alpha = 10$ and $\mu = 1$.

relation between θ and (SNR, p_f, p_d) can be derived using the same GRLT framework (see [135][136] for details).

7.1.2 Estimation-theoretic resolution limit

The resolution limit has also been extensively studied using the statistical estimation theory [133][137][138][139][140][141]. The fundamental idea is to construct a parametric model of the image in which the distance between the point sources is a parameter to estimate. Then, the Cramér-Rao lower bound (CRLB) is applied, which provides a fundamental limit on the attainable precision for estimating the distance by any unbiased estimator. Although other MSE lower bounds are also applicable leading to tighter bounds (see Appendix C for a brief review of the classical MSE bounds), CRLB has the advantage of having a simple form. Consequently, a source separation is considered resolvable if it is not smaller than the limiting standard deviation given by CRLB.

7.1.2.1 Gaussian observations

Consider the same image model as in (7.1),

$$y(\mathbf{x}_n) = f(\mathbf{x}_n; d) + w(\mathbf{x}_n) \tag{7.10}$$

$$f(\mathbf{x}_n; d) = A \cdot h(x_n - d/2, y_n, z_n) + A \cdot h(x_n + d/2, y_n, z_n) \tag{7.11}$$

Here, the distance d is the only unknown parameter to estimate.

Let T_{xy} and T_z represent the size of a voxel on the x and y-axis, and that on z-axis respectively. Accordingly, we denote the sampling frequency by $\mathbf{f}_s = [f_{xy}, f_z]$. We assume $\mathbf{f}_s$ to be sufficiently high (e.g., at least Nyquist for band-limited signals), such that the Fisher information (C.2) can be approximately calculated by an integral[1]

$$\mathcal{I}(d) = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(\frac{\partial f}{\partial d}(\mathbf{x}_n; d)\right)^2 \approx \frac{f_{xy}^2 f_z}{\sigma^2}\int_{\mathbb{R}^3}\left(\frac{\partial f}{\partial d}(\mathbf{x}; d)\right)^2 d\mathbf{x} \tag{7.12}$$

For h to be a Gaussian with unit amplitude $h(x, y, z) := \exp[-(x^2+y^2)/(2\sigma_{xy}^2)]\exp[-z^2/(2\sigma_z^2)]$,

$$(7.12) = \frac{\pi^{3/2}A^2 f_{xy}^2 f_z \sigma_z}{8\sigma^2}\left[2 + e^{-\frac{d^2}{4\sigma_{xy}^2}}\left(\frac{d^2}{\sigma_{xy}^2} - 2\right)\right]$$

Here, h is normalized according to its L^∞ norm. Other normalization may also be used leading to slightly different bounds, but the asymptotic behavior of the bounds as a function of the SNR remains the same. Now, the standard deviation of d (σ_d) is bounded by CRLB,

$$\sigma_d \geq \frac{1}{\sqrt{\mathcal{I}(d)}} = \frac{4\sqrt{3}\sigma_{xy}\sigma}{3A\sqrt{f_{xy}^2 f_z \pi^{3/2}\sigma_z}}d^{-1} + \frac{5\sqrt{3}\sigma}{36A\sigma_{xy}\sqrt{f_{xy}^2 f_z \pi^{3/2}\sigma_z}}d + o(d^2) \tag{7.13}$$

Here, we expand the CRLB in its Maclaurin series. For small d, we can ignore the term $o(d^2)$ above, and the smallest resolvable distance d_0 can be explicitely written out

$$d_0 = \frac{12\sigma_{xy}}{\left[36\sqrt{3}\sigma_{xy}\frac{A}{\sigma}\sqrt{f_{xy}^2 f_z \pi^{3/2}\sigma_z} - 15\right]^{1/2}} \tag{7.14}$$

(7.14) shows that $d_0 \propto \mathrm{SNR}^{-1/2}$.

7.1.2.2 Poisson and MPG observations

A comprehensive treatment of CRLB with Poisson and MPG observations can be found in [140], where the resolution is considered under the influence of dark intensity, pixelation, and unsynchronized photobleaching. Our following treatment is less sophisticated but is sufficient in providing instructive and consistent results. We will make use of the VST (6.4) to bring the problem to the Gaussian case.

The same MPG image model as (7.6) will be used

$$y(\mathbf{x}_n) = \alpha F(\mathbf{x}_n; d) + w(\mathbf{x}_n) \tag{7.15}$$

[1]The Fisher information can also be approximately calculated in the Fourier domain using Plancherel formula [135], which yields the same result.

Since y has an infinite-mixture distribution, exact calculation of Fisher information will result in complex expressions. However, by applying the VST the problem is Gaussianized, i.e.,

$$T(y(\mathbf{x}_n)) = 2\sqrt{f(\mathbf{x}_n; d)} + w_T(\mathbf{x}_n)$$

The Fisher information can now be easily calculated by the continuous integral approximation

$$\mathcal{I}(d) = \sum_{n=1}^{N} \frac{\left(\frac{\partial f}{\partial d}(\mathbf{x}_n; d)\right)^2}{f(\mathbf{x}_n; d)} \approx f_{xy}^2 f_z \int_{\mathbb{R}^3} \frac{\left(\frac{\partial f}{\partial d}(\mathbf{x}; d)\right)^2}{f(\mathbf{x}; d)} d\mathbf{x}$$

For a unit-amplitude Gaussian PSF, CRLB yields that

$$\sigma_d \geq \frac{1}{\sqrt{\mathcal{I}(d)}} = \frac{2^{1/4}\sigma_{xy}}{\sqrt{f_{xy}^2 f_z \pi^{3/2} \sigma_z A}} d^{-1} + \frac{2^{1/4}}{4\sigma_{xy}\sqrt{f_{xy}^2 f_z \pi^{3/2} \sigma_z A}} d + o(d^2)$$

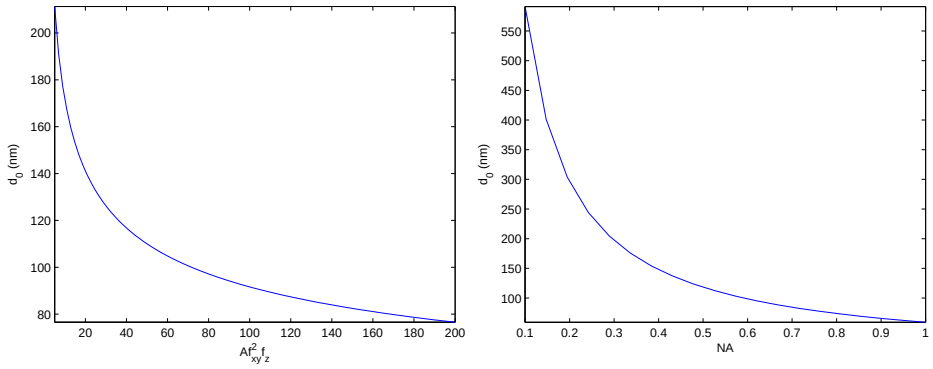
Ignoring $o(d^2)$, the minimal resolvable distance is given by

$$d_0 = \frac{2\sqrt{2}\sigma_{xy}}{\left[4\sqrt{2A}\sigma_{xy}\sqrt{f_{xy}^2 f_z \pi^{3/2}\sigma_z\sqrt{2}} - 2\right]^{1/2}} \tag{7.16}$$

(7.16) shows that $d_0 \propto A^{-1/4}$, or $d_0 \propto \mathrm{SNR}^{-1/2}$.

Remark We should not deduce from (7.16) that the minimum resolvable distance decreases if we increase the sampling rate $\mathbf{f}_s$. Indeed, unlike the Gaussian observation case which represents a "photon-unlimited" imaging, most biological fluorescence acquisition is photon limited as is modeled by a Poisson or a MPG process. Thus, the peak amplitude A of one source in (7.11) should be understood as the peak value at the sampling rate $\mathbf{f}_s$. That is, A depends on $\mathbf{f}_s$. Actually, for a fixed field of view and a fixed camera exposure time, the increasing of $\mathbf{f}_s$ means a decreasing of the pixel size on the camera. Due to the photon-limited imaging, the average number of photons arrived at every pixel will decrease accordingly and so does the value of A. For sufficiently high rates $\mathbf{f}_s$, $Af_{xy}^2 f_z$ approximates a constant. The same attention should also be called for the detection-theoretic limits derived from the MPG observations, since $\mathbf{f}_s$ is implicitly involved in the finite sum in the expression of the minimum separation (7.9).

Fig. 7.2(a) shows the minimum resolvable distance as a function of $Af_{xy}^2 f_z$ according to (7.16). An LSCM (NA = 0.5, $\lambda_{ex} = 488$nm, $\lambda_{em} = 520$nm, $D = 1$AU) is supposed, and σ_{xy} and σ_z are derived from the values given in Chapter 3. Clearly, higher SNR implies



(a) (b)

Fig. 7.2: An LSCM ($\lambda_{ex} = 488$nm, $\lambda_{em} = 520$nm, $D = 1$AU) is supposed, and σ_{xy} and σ_z are derived from the values given in Chapter 3. (a) d_0 as a function of $Af_{xy}^2 f_z$ according to (7.16) (NA = 0.5); (b) d_0 as a function of NA where d_0's are computed with the PSFs being L^1-normalized and $Af_{xy}^2 f_z = 50$.

smaller resolvable distance d_0. We have also studied the influence of the value of NA on d_0 (Fig. 7.2(b)). Here, d_0's are computed with the PSFs h being L^1-normalized ($\|h\|_1 = 1$) such that the images of the PSFs have equivalent average total photon counts. We can see that a higher resolution can be achieved by using an objective of a higher NA.

7.1.2.3 Some extensions

Localization precision limit The same framework can be applied to study the localization precision limit for a single particle [142][133][143][144][145][141]. Take the scalar estimation case, the observed image model is given by

$$y(\mathbf{x}_n) = A \cdot h(x_n - d, y_n, z_n) + w(\mathbf{x}_n) \tag{7.17}$$

where $w(\mathbf{x}_n) \sim \mathcal{N}(0, \sigma^2)$, and d is the lateral location parameter to estimate. For h having a unit-amplitude Gaussian profile, the Fisher information is given by

$$\mathcal{I}(d) \;=\; \frac{A^2}{\sigma^2} \sum_{n=1}^{N} \left(\frac{\partial h}{\partial x}(x_n - d, y_n, z_n) \right)^2 \approx \frac{A^2 f_{xy}^2 f_z \sigma_z \pi^{3/2}}{2\sigma^2}$$

Thus, the precision limit is given by

$$\sigma_d \geq \frac{1}{\sqrt{\mathcal{I}(d)}} = \frac{\sqrt{2}}{f_{xy}\sqrt{f_z \sigma_z}\pi^{3/4}} \frac{\sigma}{A} \tag{7.18}$$

This shows that $\sigma_d \propto \mathrm{SNR}^{-1}$.

If the image is of a MPG nature, i.e., $y(\mathbf{x}_n) = \alpha F(\mathbf{x}_n) + w(\mathbf{x}_n)$ where $F(\mathbf{x}_n) \sim \mathcal{P}(A \cdot h(x_n - d, y_n, z_n))$ and $w(\mathbf{x}_n) \sim \mathcal{N}(\mu, \sigma^2)$, by applying the VST we have that

$$T(y(\mathbf{x}_n)) = 2\sqrt{A \cdot h(x_n - d, y_n, z_n)} + w_T(\mathbf{x}_n)$$

Similar calculus as in the Gaussian case shows that

$$\begin{aligned} \mathcal{I}(d) &\approx 2\sqrt{2}Af_{xy}^2 f_z \sigma_z \pi^{3/2} \\ \sigma_d &\geq \frac{1}{\sqrt{\mathcal{I}(d)}} = \frac{1}{\sqrt{A}\left[2\sqrt{2}f_{xy}^2 f_z \sigma_z \pi^{3/2}\right]^{1/2}} \end{aligned} \tag{7.19}$$

Again, we find that $\sigma_d \propto \mathrm{SNR}^{-1}$, or in other words σ_d is inversely proportional to the square root of the total underlying Poisson intensity [144]. Note that (7.19) is different from the value derived in [144] just by a constant factor, due to the different normalizations on the source profile h.

Multivariate estimation A more realistic signal model is given by

$$\begin{aligned} f(\mathbf{x}_n; \theta) &:= A_1 h(x_n + d_1, y_n + d_2, z_n + d_3) + A_2 h(x_n + d_4, y_n + d_5, z_n + d_6) \\ \theta &:= [A_1, A_2, d_1, \ldots, d_6] \end{aligned} \tag{7.20}$$

Here, the parameter θ to estimate is a vector. The CRLB follows from inverting the Fisher information matrix associated with θ. The case of nonidentical source profiles can also be studied similarly. Details can be found, e.g., in [144][140][141].

7.1.3 Information-theoretic resolution limit

Kosarev [146] has proposed to study the resolution limit using Shannon's channel capacity theory (see Appendix D for a brief introduction). Here, using the same information-theoretic framework we will summarize and extend [146] to study the limiting resolution for a finite number and a upper-bounded but unknown number of sources with Gaussian, Poisson and MPG observations, and the localization accuracy limit. The information-theoretic approach exhibits an advantage over the detection-theoretic or estimation-theoretic methods in the aspect that it can tackle more easily $K > 2$ (or even an unknown number of) sources. The disadvantage of this approach is that it leads to looser bounds. In the presentation below, no assumption will be made on the sampling rate.

7.1.3.1 Gaussian observations

We observe an image y of N pixels,

$$y(\mathbf{x}_n) = f(\mathbf{x}_n) + w(\mathbf{x}_n), \quad 1 \leq n \leq N \tag{7.21}$$

where f is the noise-free image, and $w(\mathbf{x}_n) \sim \mathcal{N}(0, \sigma^2)$ is the noise process.

A finite number of sources We assume that f is generated from the convolution of the PSF h with K point sources of equal amplitudes A, i.e.,

$$f(\mathbf{x}) = A \sum_{i=1}^{K} h(\mathbf{x} - \mathbf{p}_i) \tag{7.22}$$

where $\mathbf{p}_i$ is the position vector of the i-th source. Let us suppose h to be a Gaussian PSF of unit amplitude.

(7.21) can be deemed as one use of N parallel Gaussian channels. These channels have a capacity, i.e., the upper bound on the maximum information bits per use that can be reliably transmitted. This is predicted by Shannon's capacity theorem (Theorem 11). We will calculate the capacity of our channels which will bound the maximum information in our image, from which we can derive the resolution limit. Before applying the capacity theorem we should find the energy constraint. Supposing for a moment that the image support is periodized so that every pixel plays a symmetric role. Assuming further that $(\mathbf{p}_i)_i$ are independent and uniformly distributed in an observed image volume V much larger than the PSF size, then we have that

$$\begin{aligned}
\mathbb{E}\left[\sum_{n=1}^{N} f(\mathbf{x}_n)^2\right] &= A^2 \sum_{n=1}^{N} \sum_{i,j=1}^{K} \mathbb{E}\left[h(\mathbf{x}_n - \mathbf{p}_i) h(\mathbf{x}_n - \mathbf{p}_j)\right] \\
&\approx A^2 N \left(\sum_{i=j} \frac{\sigma_{xy}^2 \sigma_z \pi^{3/2}}{V} + \sum_{i \neq j} \left[\frac{2\sqrt{2}\pi^{3/2} \sigma_{xy}^2 \sigma_z}{V} \right]^2 \right) \\
&= \frac{N A^2 K \sigma_{xy}^2 \sigma_z \pi^{3/2}}{V^2} \left(V + 8\pi^{3/2} (K-1) \sigma_{xy}^2 \sigma_z \right) =: P_K
\end{aligned}$$

Dropping off the periodization, it follows immediately that

$$\mathbb{E}\left[\sum_{n=1}^{N} f(\mathbf{x}_n)^2\right] \leq P_K$$

Therefore, by Theorem 11 the total information bits transmitted by the whole image with N pixels are bounded by

$$\begin{aligned} C &\leq \frac{1}{2}\sum_{n=1}^{N}\log_2\left[1+\frac{P_K/N}{\sigma^2}\right] \\ &= \log_2\left[1+\frac{A^2K\sigma_{xy}^2\sigma_z\pi^{3/2}}{\sigma^2V^2}\left(V+8\pi^{3/2}(K-1)\sigma_{xy}^2\sigma_z\right)\right]^{N/2} \end{aligned} \tag{7.23}$$

Now, we have at most 2^C distinguishable messages. Thus, if we partition the image volume by cubes of sidelength Δ, we will have a 3D grid of $M = V/\Delta^3 = NT_{xy}^2T_z/\Delta^3$ cubes. Thereby, we have a total of C_M^K different configurations (ignoring orders) of these K sources at the resolution level Δ. These configurations form the entire message set. Thus, the minimal resolution level for K sources are given by the solution of Δ of the equation $2^C = C_M^K$.

If $K \ll M$, we have $C_M^K \approx M^K/K!$. Thus,

$$\Delta \approx \frac{V^{\frac{1}{3}}}{(K!)^{\frac{1}{3K}}\left[1+\frac{A^2K\sigma_{xy}^2\sigma_z\pi^{3/2}}{\sigma^2V^2}\left(V+8\pi^{3/2}(K-1)\sigma_{xy}^2\sigma_z\right)\right]^{\frac{Vf_{xy}^2f_z}{6K}}} \tag{7.24}$$

Now, if V is large, (7.24) can be again approximated by

$$\Delta \approx \frac{V^{\frac{1}{3}}}{(K!)^{\frac{1}{3K}}}\exp\left[-\frac{A^2\sigma_{xy}^2\sigma_z\pi^{3/2}}{6\sigma^2T_{xy}^2T_z}\right] \tag{7.25}$$

This shows that $\Delta \propto \exp(-\mathrm{SNR}^2)$.

An upper-bounded but unknown number of sources If no prior information of the exact source number is available, but if M is large enough such that it can be deemed as an upper bound of the source number, the message set includes a total of $\sum_{K=0}^{M}C_M^K = 2^M$ different configurations. Supposing the source number K follows a distribution with mean μ_K and variance σ_K^2, the total energy constraint becomes

$$\begin{aligned} \mathbb{E}\left[\sum_{n=1}^{N}f(\mathbf{x}_n)^2\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{n=1}^{N}f(\mathbf{x}_n)^2\middle| K\right]\right] \approx \mathbb{E}[P_K] \\ &= \frac{NA^2\sigma_{xy}^2\sigma_z\pi^{3/2}}{V^2}\left(\mu_KV+8\pi^{3/2}\sigma_{xy}^2\sigma_z(\sigma_K^2+\mu_K^2-\mu_K)\right) \end{aligned}$$

We derive the resolution level for this situation,

$$\Delta = \frac{V^{\frac{1}{3}}}{\left[\log_2\left(1+\frac{A^2\sigma_{xy}^2\sigma_z\pi^{3/2}}{\sigma^2V^2}\left[\mu_K V + 8\pi^{3/2}\sigma_{xy}^2\sigma_z(\sigma_K^2+\mu_K^2-\mu_K)\right]\right)^{N/2}\right]^{\frac{1}{3}}} \tag{7.26}$$

For a large V, we have

$$\Delta \approx \frac{V^{\frac{1}{3}}}{\left(\frac{A}{\sigma}\right)^{\frac{2}{3}}\left[\frac{1}{2}\sigma_{xy}^2\sigma_z\pi^{3/2}\mu_K f_{xy}^2 f_z \log_2 e\right]^{\frac{1}{3}}} \tag{7.27}$$

This shows that $\Delta \propto \mathrm{SNR}^{-2/3}$.

7.1.3.2 Poisson and MPG observations

Our image model after applying the VST (6.4) is given by

$$T(y(\mathbf{x}_n)) = 2\sqrt{f(\mathbf{x}_n)} + w(\mathbf{x}_n), \quad 1 \le n \le N \tag{7.28}$$

where $w(\mathbf{x}_n) \sim \mathcal{N}(0,1)$.

A finite number of sources If K sources are present in the image with a large volume V, the energy constraint is given by

$$\mathbb{E}\left[\sum_{n=1}^{N} 4f(\mathbf{x}_n)\right] \approx \frac{ANK8\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z}{V} \tag{7.29}$$

The total capacity of these parallel channels is given by

$$C = \log_2\left(1+\frac{AK8\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z}{V}\right)^{\frac{N}{2}} \tag{7.30}$$

For $K \ll M$, the resolution level is thus

$$\Delta \approx \frac{V^{\frac{1}{3}}}{(K!)^{\frac{1}{3K}}\left(1+\frac{AK8\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z}{V}\right)^{\frac{Vf_{xy}^2 f_z}{6K}}} \tag{7.31}$$

For a large V, the resolution level is approximated by

$$\Delta \approx \frac{V^{\frac{1}{3}}}{(K!)^{\frac{1}{3K}}}\exp\left(-\frac{4}{3}A\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z f_{xy}^2 f_z\right) \tag{7.32}$$

This shows that $\Delta \propto \exp(-\mathrm{SNR}^2)$.

An upper-bounded but unknown number of sources Again, we assume M to be an upper-bound of the source number. We have the energy constraint

$$\mathbb{E}\left[\sum_{n=1}^{N} 4f(\mathbf{x}_n)\right] \approx \frac{AN\mu_K 8\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z}{V} \tag{7.33}$$

The capacity is given by

$$C = \log_2\left[1 + \frac{A\mu_K 8\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z}{V}\right]^{Vf_{xy}^2 f_z/2} \tag{7.34}$$

The resolution level is given by

$$\Delta = \frac{V^{\frac{1}{3}}}{\left(\log_2\left[1 + \frac{A\mu_K 8\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z}{V}\right]^{Vf_{xy}^2 f_z/2}\right)^{\frac{1}{3}}} \tag{7.35}$$

which can be approximated for a large V by

$$\Delta \approx \frac{V^{\frac{1}{3}}}{A^{\frac{1}{3}}\left(4\sqrt{2}\pi^{3/2}\mu_K\sigma_{xy}^2\sigma_z f_{xy}^2 f_z \log_2 e\right)^{\frac{1}{3}}} \tag{7.36}$$

This shows again that $\Delta \propto \mathrm{SNR}^{-2/3}$.

7.1.3.3 Some extensions

Localization accuracy limit We set $K = 1$ in (7.24) and (7.25) to derive the localization accuracy limits.

$$\Delta \approx \frac{V^{\frac{1}{3}}}{\left[1 + \frac{A^2\sigma_{xy}^2\sigma_z\pi^{3/2}}{\sigma^2 V}\right]^{\frac{Vf_{xy}^2 f_z}{6}}} \tag{7.37}$$

$$(\text{large } V)\ \Delta \approx V^{\frac{1}{3}}\exp\left[-\frac{A^2\sigma_{xy}^2\sigma_z\pi^{3/2}f_{xy}^2 f_z}{6\sigma^2}\right] \tag{7.38}$$

We set $K = 1$ in (7.31) and (7.32) to derive the localization accuracy limit for MPG data.

$$\Delta \approx \frac{V^{\frac{1}{3}}}{\left(1 + \frac{8\sqrt{2}\pi^{3/2}A\sigma_{xy}^2\sigma_z}{V}\right)^{\frac{Vf_{xy}^2 f_z}{6}}} \tag{7.39}$$

$$(\text{large } V)\ \Delta \approx V^{\frac{1}{3}}\exp\left(-\frac{4}{3}A\sqrt{2}\pi^{3/2}\sigma_{xy}^2\sigma_z f_{xy}^2 f_z\right) \tag{7.40}$$

7.1.3.4 Looseness of the bounds

It can be seen that all the information-theoretic bounds derived above are looser than those derived in the detection/estimation frameworks. The first reason is that Shannon's capacity theorem considers only the energy constraint, and the PSF shape is not fully taken into account. Indeed, at a first look, the bounds (7.27) and (7.36) could be counter-intuitive because they decrease as the mean number of sources increases. However, this is plausible in the Shannon's sense that as the total energy constraint $\mathbb{E}[P_K]$ is an increasing function of μ_K, SNR rises for larger μ_K. Thereby, more information bits can be transmitted in the Gaussian channels. There is a second reason for the looseness of the bounds. In effect, only some special coding/decoding schemes for the message set can achieve the channel capacity, such as random coding and jointly typical decoding (see [147]). In reality, the coding for our messages of the source locations is implicitly determined by the microscope imaging system, which could be suboptimal such that the real information rate is much lower than Shannon's capacity.

7.2 Super-resolution Spot Detection

Section 7.1 showed that it is theoretically possible to localize highly accurately an individual point source and to identify close sources with a separation smaller than Rayleigh's distance, provided that we have a sufficient SNR.

Concerning practical algorithm for localizing a single particle, [148] extracted the object of interest by thresholding after linear filtering of the image. The location is estimated by the center-of-mass of the object. Recently, parametric method such as least-squares PSF fittings, and in general, ML estimations are investigated [149][36][131][144][141]. Indeed, as a subresolutional particle has a microscope PSF profile, the image can be modeled by a theoretical PSF. For Poisson observations, the ML estimation is asymptotically optimal in the sense that as the observed photon number increases (or SNR increases), the estimator attains the CRLB, i.e., the ultimate localization precision limit. ML estimation is also proposed to localize a single particle with one or a few defocus observation planes and a WFFM PSF model considering spherical aberrations [150].

By modeling the image as a mixture of PSFs, the ML approach allows to localize a known number of particles separated even below the Rayleigh's distance (super-resolution)

[37][140]. In applications of quantum dots, the blinking statistics can be analyzed by an independent component analysis [151] to separate nearby dots.

If the source number K is unknown, a model selection procedure has to be carried out to determine this number. [152] proposed to impose prior distributions on the source separations. Then, MAP estimation is carried out for every possible number of sources and the most probable values are retained. However, in real biological applications what the true priors for the separations should be is not clear. In [38], a super-resolution detection of submicrometric fluorescent tags has been proposed for Gaussian data. These tags were modeled as point-like sources, and the PSF as a Gaussian, yielding an image model consisting of a mixture of Gaussians. In a first step, spots are pre-detected based on a "spottiness" score equal to the product of the local average and the Gaussian curvature of the image intensity, providing a score image that is then thresholded using an empirical threshold. In a second step, a parametric-model fitting routine refines the positions of spots, and iteratively inserts new Gaussians into the model. The acceptance of new spots is controlled by hypothesis tests, i.e., $H_0 : K = M$ vs. $H_1 : K \geq M + 1$. The test statistic is the ratio of the residues of the least-square fits with M and $M + 1$ sources, which is assumed to have a F-distribution under H_0. The tests are carried out from a small M to larger ones only, and the tests terminate as long as the fit no longer improves significantly.

Here, we propose a source localization framework where the source number K is unknown and the data can be of a MPG nature. This approach can be seen as an extension of [38]. A pre-detection will be carried out followed by an iterative parametric-model fitting with model-order selections in order to refine the number of spots and their locations. We will mainly discuss the MPG case where the generalized least-squares fit and the VST-based least-squares fit will be both tested. We will also evaluate the performance of a variety of model selection methods for this super-resolution problem.

7.2.1 Pre-detection

We will describe the pre-detection scheme, which aims to provide a good initial guess of the number, the positions, and the amplitudes of spots in the image. This initial guess will then be improved upon by the model-fitting approach described in section 7.2.2.

Our MPG image model is given by

$$y(\mathbf{x}_n) = \alpha F(\mathbf{x}_n) + w(\mathbf{x}_n) \tag{7.41}$$

$$F(\mathbf{x}_n) \sim \mathcal{P}(f(\mathbf{x}_n)), \quad f(\mathbf{x}_n) = b + \sum_{i=1}^{K} A_i h(\mathbf{x}_n - \mathbf{p}_i) \tag{7.42}$$

$$w(\mathbf{x}_n) \sim \mathcal{N}(\mu, \sigma^2) \tag{7.43}$$

where $b \geq 0$ represents a dark intensity background, A_i and $\mathbf{p}_i$ respectively describe the amplitude and the position vector of i-th spot, K is the total number of spots, and h is the unit-amplitude source profile.

The pre-detection is based on the MS-VST+IUWT scheme (chapter 6). A slight modification of the detection method proposed in section 6.3 will allow us to estimate the spot amplitudes in addition to their locations. Our pre-detection consists of the following steps. First, in an object-free background region, we estimate (α, b, μ, σ) by cumulant matching. Second, we carry out MS-VST+IUWT detection, and once the multi-resolution support is available, we reconstruct the image from the significant wavelet coefficients with the coarsest approximation band set to zero. This eliminates most regular background as wavelets are band-pass filters. Third, in the reconstructed image the pixels with an intensity below a threshold T will be set to zero. In our experiments, the positive part is retained (i.e. $T = 0$) since we are interested in bright spots. Finally, all local maxima are extracted as putative spots with their amplitudes. Parallel to Proposition 8, wavelet FDR control will lead to detection error probability control in an object-free homogeneous MPG noisy image. Let us suppose that M_0 spots are pre-detected.

7.2.2 Model fitting and model-order selection

The spot parameters computed at this stage are a first guess, which must be improved upon for several reasons. The pre-detection scheme described above could produce false detection due to noise. If two or more spots are very close to each other or overlap, the pre-detection may produce only one hit. Moreover, the localization precision of the pre-detection is limited by the size of the pixels. We therefore seek to improve the pre-detection using a parametric-model estimation procedure. This second step aims to: (i) resolve multiple spots that produced a single predetection hit, possibly with super-resolution; (ii) remove residual false detections; (iii) achieve subpixelic localization.

7.2.2.1 Parametric object model

Each spot profile h is described by the convolution of the underlying spot intensity distribution with the microscope PSF. The Gaussian approximative PSF will be used with standard deviations (σ_{xy}, σ_z) calculated in chapter 3. We will also suppose the underlying source intensity distribution to be a Gaussian with a standard deviation σ_0. σ_0 represents the size of the source and we assume that typical range of the source size is available from prior biophysical knowledges. This modeling results in a closed-form expression for h, i.e.,

$$h(x, y, z) = \exp\left[-\frac{x^2 + y^2}{2(\sigma_0^2 + \sigma_{xy}^2)} - \frac{z^2}{2(\sigma_0^2 + \sigma_z^2)}\right]$$

(7.42) is now a mixture of Gaussians and the parameter set for the signal f is given by $\Theta_K = \{(\mathbf{p}_i)_{1\leq i\leq K}, (A_i)_{1\leq i\leq K}, \sigma_0, b\}$ where the subscript K indicates the model order. To specify that f depends on Θ_K, we will write $f(\mathbf{x}; \Theta_K)$.

7.2.2.2 Iterative model fitting

Given a value of K, θ can be determined by the ML estimation. The exact ML estimator has no closed forms due to the MPG distribution. However, by applying the VST on y, we get an asymptotic normal model,

$$T(y(\mathbf{x})) = 2\sqrt{f(\mathbf{x}; \Theta_K)} + w_T(\mathbf{x}) \tag{7.44}$$

where $w_T(\mathbf{x}) \sim \mathcal{N}(0, 1)$. Clearly, the ML estimation for this transformed image model is simply the least-squares fitting, i.e.,

$$\hat{\theta}_K = \arg\min_{\Theta_K} \sum_{n=1}^{N} \left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n; \Theta_K)}\right]^2 \tag{7.45}$$

Constraints on the parameters can also be imposed based on prior knowledges such as the positivity, i.e., $A_i, \sigma_0, b \geq 0$. In our experiments, we also test the generalized least-squares fitting on the MPG observations [131] where the variance at each sample is considered in the fit, i.e.,

$$\hat{\theta}_K = \arg\min_{\Theta_K} \sum_{n=1}^{N} \frac{(y(\mathbf{x}_n) - \alpha f(\mathbf{x}_n; \Theta_K) - \mu)^2}{\alpha^2 f(\mathbf{x}_n; \Theta_K) + \sigma^2} \tag{7.46}$$

It remains to determine the value of K by a model-order selection approach. Three categories of order selection methods will be evaluated, i.e., hypothesis tests, information

criteria, and MDL criteria. In the discussion below, we will concentrate on the model selection with the VST-based fit (7.45), since it has a log-likelihood form which is theoretically required by many model-selection methods.

Hypothesis tests At the onset, we have a list of spots L_s which initially contains the pre-detection ones. The model fitting is initialized with the parameter of pre-detection Θ_{M_0}. The initial fit (7.45) produces a refined estimation of the spot parameters allowing subpixelic localization.

At this stage, however, close and overlapping spots remain to be resolved. We therefore follow up a bottom-up addition of spots [38]. To do this, we sequentially try to add spots to the model using hypothesis tests to decide whether to accept spot addition or not. After adding a spot to the current model Θ_M, we fit the new model Θ_{M+1} to the image, and then ask whether the new fit is significantly better than the previous fit. This is done by testing $H_0 : K = M$ against the alternative $H_1 : K > M$. The GLRT statistic $-2\ln R(y)$ [153] is constructed, i.e.,

$$\begin{aligned} -2\ln R(y) &= -2\ln \frac{\sup_{\Theta_M} \text{Likelihood}(\Theta_M|T(y))}{\sup_{\Theta_{M+1}} \text{Likelihood}(\Theta_{M+1}|T(y))} \\ &= -2\ln \frac{\exp\left(-\frac{1}{2}\sum_{n=1}^{N}\left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n;\hat{\theta}_M)}\right]^2\right)}{\exp\left(-\frac{1}{2}\sum_{n=1}^{N}\left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n;\hat{\theta}_{M+1})}\right]^2\right)} \end{aligned} \tag{7.47}$$

Under H_0 we test $-2\ln R(y)$ using a Chi-square distribution with a degree of freedom $|\Theta_{M+1}| - |\Theta_M| = 4$, and a significance level α. The spot addition is accepted only if the test is significant. Then, this new spot will be inserted into L_s, and the new test becomes $H_0 : K = M+1$ vs. $H_1 : K > M+1$. If H_0 is maintained, this spot addition is refused since the smaller model Θ_M provides an equally good fit to the data as Θ_{M+1}, and another spot in L_s is examined. This procedure is repeated until all the spots in L_s have been examined.

The bottom-up procedure allows to distinguish close and overlapping spots, but we should also remove superfluous spots due to false positives of the pre-detection. For this reason, we carry out a top-down procedure for spot removal, where the hypothesis test described above is applied in reverse. That is, one spot in L_s will be removed from the model as long as the quality of fit is not significantly degraded. The output of the algorithm is the spot parameters from the final fit.

Besides GLRT, [38] has proposed the F-test. The test statistic is the ratio of two Chi-square residues,

$$F := \frac{\sum_{n=1}^{N}\left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n;\hat{\theta}_M)}\right]^2 / r_M}{\sum_{n=1}^{N}\left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n;\hat{\theta}_{M+1})}\right]^2 / r_{M+1}} \tag{7.48}$$

Under H_0, F is tested using a Fisher distribution of r_M and r_{M+1} degrees of freedom. Here $r_M = N - p_M$, and $p_M = 4M + 2$ is the number of free parameters of the model Θ_M.

It can be seen that the probability of erroneously detecting objects in an object-free noisy image is still upper-bounded by the wavelet FDR value using this bottom-up and top-down hypothesis testing framework. If in addition the distribution of the pre-detection number is known for a pure noisy image, say, $\alpha_p = \Pr(M_0 = p|H_0)$, the detection error rate can be easily shown to be

$$\Pr(\hat{K} \neq 0|H_0) = \sum_{b=1}^{\infty}\sum_{p=1}^{b} C_{2b-p}^{b}[1-(1-\alpha)^b]\alpha^{b-p}(1-\alpha)^b\alpha_p$$

where $\hat{K}$ is the number of finally detected spots.

Finally, we should point out that in GLRT, as $(y(\mathbf{x}_k))_k$ are not identically distributed, the Chi-square distribution for H_0 is only an approximation. Similarly, in F-test, as the numerator and the denominator in (7.48) are not independent statistics, the Fisher distribution is also an approximation under the null hypothesis.

Information criteria In addition to HTs, information criteria (see [154][155] and the references therein) are also widely used for order selection. They include, e.g., Akaike information criterion (AIC), Corrected AIC rule (AICc) and Bayesian information criterion (BIC).

These criteria can be derived by maximizing the relative Kullback-Leibler information under different hypothesis (see [154] for details). They can be written in a general penalized-likelihood form

$$\begin{aligned}\hat{K} &= \arg\min_K \mathrm{IC}_K, \quad \mathrm{IC}_K = -2\ln\left[\text{Likelihood}(\hat{\theta}_K|T(y))\right] + \eta\cdot p_K \\ &= \arg\min_K \sum_{n=1}^{N}\left[T(y(\mathbf{x}_n)) - 2\sqrt{f(\mathbf{x}_n;\hat{\theta}_K)}\right]^2 + \eta(4K+2)\end{aligned} \tag{7.49}$$

where $\eta = 2$ for AIC, $\eta = 2N/(N-K-1)$ for AICc, and $\eta = \ln N$ for BIC. These criteria can be combined with our iterative bottom-up/top-down schemes in a similar way as for the HTs. For example, in the bottom-up procedure, IC_K is evaluated for each spot addition, and the new spot is accepted only if it induces a smaller value of IC_K.

An interesting relation between the information criteria and the GLRT statistic is noted in [156], i.e.,

$$\mathrm{IC}_M - \mathrm{IC}_{M+1} = -2\ln R(y) - 4\eta$$

This shows that the information criteria correspond to particular choices of the p-value in GLRT.

MDL criteria Like information criteria, the Minimum Description Length (MDL) principle for model-order selection can also be written in a penalized-likelihood form [157][158]

$$\hat{K} \;=\; \arg\min_K -2\ln\Big[\mathrm{Likelihood}(\hat{\theta}_K|T(y))\Big] + p_K \ln\frac{N}{2\pi} \tag{7.50}$$

Although (7.50) shares a similar form with (7.49), they are derived using fundamentally different assumptions. Roughly speaking, for information criteria, we assume that there exists a true distribution from which the observations are sampled, and the goal of model selection is to find the model closest to the truth. For the MDL principle, however, whether a true distribution even exits is considered inherently unanswerable. The goal of model selection is to identify the model that allows to describe the data with fewest symbols, i.e., the model that achieves the greatest data compression [158].

The penalization term in (7.50) is derived by assuming the existence of an upper-bound of the model order. If this hypothesis is dropped, the MDL should be corrected to [157]

$$\hat{K} \;=\; \arg\min_K -2\ln\Big[\mathrm{Likelihood}(\hat{\theta}_K|T(y))\Big] + p_K \ln\frac{N}{2\pi} + 2\ln(K+1) \tag{7.51}$$

The term $2\ln(K+1)$ represents the additional symbols necessary to encode the model index K which is not bounded [159]. This criterion will be denoted as MDLi.

7.2.3 Detection within Gaussian data

Although not discussed above, the proposed framework can be easily applied to detect spots in Gaussian noisy data. The image model is given by $y(\mathbf{x}_n) = f(\mathbf{x}_n;\Theta_K) + w(\mathbf{x}_n)$ where f is

shown in (7.42) and $w(\mathbf{x}_n) \sim \mathcal{N}(0, \sigma^2)$. Regarding the pre-detection, the method proposed in [38] can be adopted. We can also resort to IUWT-based detection as proposed for the MPG case but without using the MS-VST. Parametric model fitting for Gaussian data boils down to the least-squares fitting. The model-order selections completely resemble the way described for the MPG case with the use of VST, since VST just brings us to a Gaussian image model.

7.2.4 Experiments

To quantitatively evaluate the performance of the algorithm, we run series of experiments on simulated MPG data. The images were generated using the model of (7.42) and the following parameters: voxel sizes $T_x = T_y = 0.1\mu m$, $T_z = 0.3\mu m$, image field $5.1\mu m \times 5.1\mu m \times 6.3\mu m$, the standard deviation of the underlying source profile $\sigma_0 = 0.02\mu m$, and Gaussian PSF standard deviations $\sigma_{xy} = 0.103$ and $\sigma_z = 0.368$ which approximate a non-paraxial LSCM with NA = 1, an oil immersion objective, a pinhole of diameter 1AU, an excitation wavelength $\lambda_{ex} = 488nm$, and an emission wavelength $\lambda_{em} = 520nm$. MPG noise with the following parameters are generated: gain $\alpha = 20$, dark intensity $b = 0.01$, readout mean $\mu = 5$ and standard deviation $\sigma = 2$. Source amplitude A is varied to produce images at various SNR, defined to be

$$\text{SNR} := \frac{\alpha A}{\sqrt{\alpha^2(A+b)+\sigma^2}}$$

7.2.4.1 Localizing an individual spot

We first test our algorithm on images of a single spot, without the complications due to overlapping fluorescence from several spots. Example images of SNR $\in [5, 15]$ are shown in Fig. 7.3. We have tested both the VST-based least-squares fit and the generalized least-squares fit; all the model-order selection methods discussed in section 7.2.2.2 are implemented. For each combination of SNR, model-fitting approach and model-order selection criterion, we have run 50 replications of the detection under different noise realizations. Based on these trials, the following quantities are evaluated: the correct detection rate (CDR), i.e., the percentage that the algorithm detect exactly one spot in the image; RMSE_x, RMSE_y and RMSE_z, i.e., the root mean squared errors of the x-coordinate, y-coordinate and z-coordinate of the spot center; PRE of A, i.e., the relative error on the amplitude estimation. The RMSEs and PREs are evaluated only for those correctly detected spots.

Numeric results are shown in Tab. 7.1, 7.2 and 7.3. CDR equal or higher than 80% are emphasized with bold face.

It can be seen that only the order-selection by F-test leads to high CDRs for all the SNRs. Moreover, all the localization errors are in the range of several nanometers for the lateral coordinates and tens of nanometers for the axial one. The localization becomes more accurate as SNR increases. The accuracy of the amplitude estimation is also high, with errors of a few percent only.

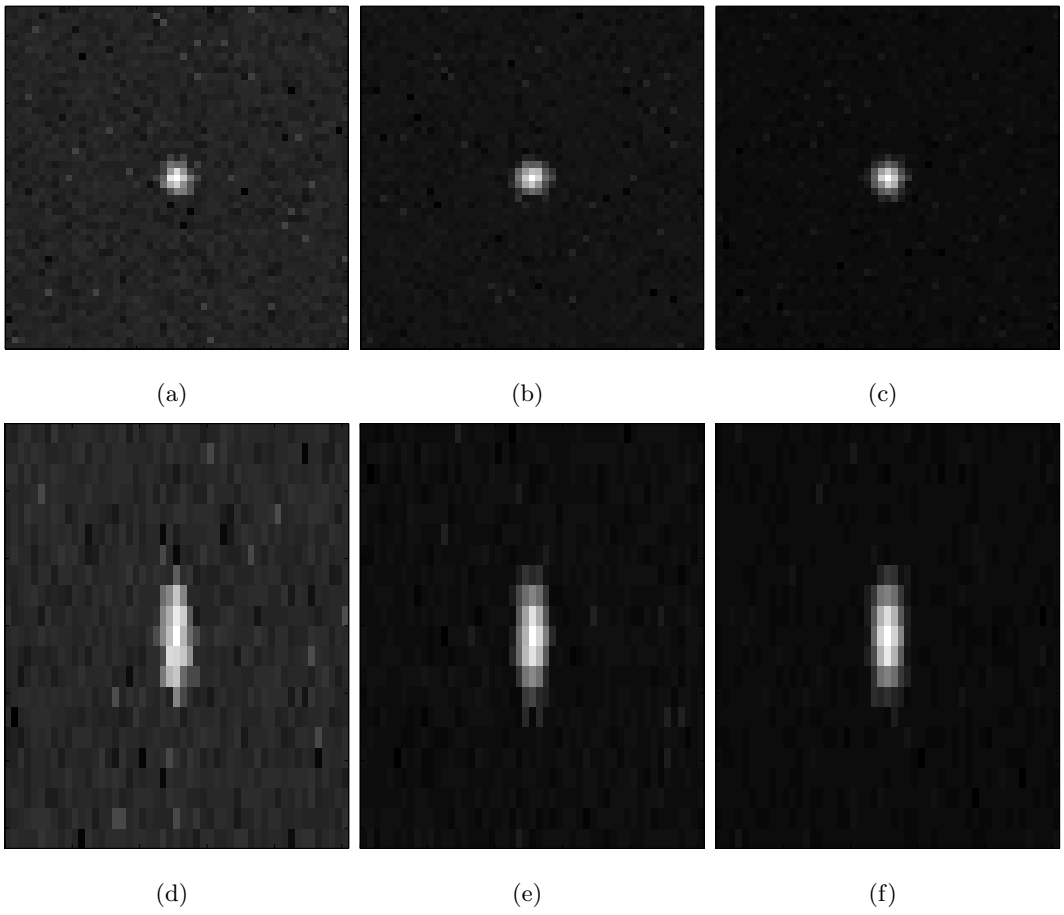


Fig. 7.3: Simulated individual spot at different SNRs. xy-planes passing through the spot center for (a) SNR = 5; (b) SNR = 10; (c) SNR = 15; yz-planes passing through the spot center for (d) SNR = 5; (e) SNR = 10; (f) SNR = 15. The image intensities are shown in a square-root scale.

Tab. 7.1: Localization performance (SNR = 5)

Order Sel.	CDR%	RMSE_x (nm)	RMSE_y (nm)	RMSE_z (nm)	PRE (A)%
GLRT*	(0, 6)	(-, 3.43)	(-, 3.02)	(-, 9.77)	(-, 9.33)
GLRT	(2, 4)	(4.28, 3.99)	(0.21, 4.81)	(2.5, 23.95)	(15.22, 5.45)
F-test*	(**100**, **100**)	(4.62, 5.27)	(4.77, 4.97)	(17.29, 19.31)	(4.86, 7.39)
F-test	(**100**, **100**)	(4.62, 5.27)	(4.77, 4.97)	(17.29, 19.31)	(4.86, 7.39)
AIC	(2, 2)	(8.28, 4.97)	(4.55, 2.16)	(12.63, 9.09)	(1.29, 8.89)
AICc	(0, 2)	(-, 4.97)	(-, 2.16)	(-, 9.09)	(-, 8.89)
BIC	(36, 76)	(4.48, 5.12)	(4.26, 5.30)	(15.7, 19.66)	(4.63, 6.92)
MDL	(30, 66)	(4.36, 5.72)	(4.88, 5.02)	(14.49, 19.82)	(5.50, 6.81)
MDLi	(20, 60)	(5.04, 5.31)	(4.89, 5.45)	(18.03, 18.83)	(3.75, 7.33)

In every parenthesis, the first value is from the VST-based least-squares fit, and the second is given by the generalized least-squares fit. In the pre-detection step, $J = 3$ is used in IUWT and the wavelet-domain FDR is set to 10^{-6}. The p-value for GLRT* and F-test* is $\alpha = 0.05$; that for GLRT and F-test is set to $\alpha = 0.1$. 50 replications are carried out for each order-selection method and each model fitting method. CDR equal or higher than 80% are emphasized with bold face.

Tab. 7.2: Localization performance (SNR = 10)

Order Sel.	CDR%	RMSE_x (nm)	RMSE_y (nm)	RMSE_z (nm)	PRE (A)%
GLRT*	(36, 8)	(1.68, 1.71)	(1.88, 1.29)	(9.47, 15.31)	(5.19, 5.16)
GLRT	(22, 10)	(1.53, 1.53)	(1.88, 1.22)	(10.00, 13.72)	(4.11, 6.19)
F-test*	(**100**, **100**)	(2.23, 2.12)	(2.16, 2.40)	(9.07, 10.07)	(6.67, 5.47)
F-test	(**100**, **100**)	(2.23, 2.12)	(2.16, 2.40)	(9.07, 10.07)	(6.67, 5.47)
AIC	(28, 6)	(1.57, 1.90)	(2.06, 3.23)	(9.74, 16.17)	(4.63, 2.87)
AICc	(22, 8)	(1.82, 2.31)	(2.21, 2.83)	(9.09, 14.28)	(4.37, 4.26)
BIC	(58, 78)	(2.29, 2.26)	(2.20, 2.58)	(8.35, 9.47)	(6.72, 4.82)
MDL	(56, 68)	(2.35, 2.23)	(2.09, 2.57)	(8.44, 9.88)	(7.31, 4.64)
MDLi	(64, 66)	(1.84, 2.36)	(2.10, 2.44)	(9.01, 9.22)	(6.83, 5.30)

In every parenthesis, the first value is from the VST-based least-squares fit, and the second is given by the generalized least-squares fit. In the pre-detection step, $J = 3$ is used in IUWT and the wavelet-domain FDR is set to 10^{-6}. The p-value for GLRT* and F-test* is $\alpha = 0.05$; that for GLRT and F-test is set to $\alpha = 0.1$. 50 replications are carried out for each order-selection method and each model fitting method. CDR equal or higher than 80% are emphasized with bold face.

Tab. 7.3: Localization performance (SNR = 15)

Order Sel.	CDR%	RMSE$_x$ (nm)	RMSE$_y$ (nm)	RMSE$_z$ (nm)	PRE (A)%
GLRT*	(42, 24)	(1.21, 1.64)	(1.97, 1.48)	(4.67, 4.84)	(4.33, 6.98)
GLRT	(34, 20)	(1.28, 1.53)	(2.11, 1.72)	(4.83, 5.34)	(4.16, 6.52)
F-test*	(**100**, **100**)	(1.65, 1.53)	(1.77, 1.82)	(4.66, 5.74)	(4.50, 5.54)
F-test	(**100**, **100**)	(1.65, 1.53)	(1.77, 1.82)	(4.66, 5.74)	(4.50, 5.54)
AIC	(40, 14)	(1.22, 1.81)	(2.09, 1.88)	(4.94, 5.70)	(4.13, 7.34)
AICc	(36, 20)	(1.28, 1.73)	(1.98, 1.58)	(4.13, 6.20)	(3.83, 7.36)
BIC	(64, **82**)	(1.50, 1.57)	(1.83, 1.70)	(4.96, 5.26)	(3.81, 4.92)
MDL	(54, 76)	(1.26, 1.56)	(1.91, 1.75)	(4.67, 5.41)	(3.62, 4.82)
MDLi	(58, 74)	(1.33, 1.56)	(1.84, 1.64)	(5.08, 5.44)	(3.76, 4.95)

In every parenthesis, the first value is from the VST-based least-squares fit, and the second is given by the generalized least-squares fit. In the pre-detection step, $J = 3$ is used in IUWT and the wavelet-domain FDR is set to 10^{-6}. The p-value for GLRT* and F-test* is $\alpha = 0.05$; that for GLRT and F-test is set to $\alpha = 0.1$. 50 replications are carried out for each order-selection method and each model fitting method. CDR equal or higher than 80% are emphasized with bold face.

7.2.4.2 Resolving close spots

In order to test the ability of the algorithm to separate spots whose fluorescence overlaps, we now simulate images of two spots with various distances d_{xy} along x-axis, and SNR. By the property that the dip between the two superimposed Airy functions at the Rayleigh distance amounts to 73.5% of the maximum intensity, we derive the corresponding Rayleigh distance for Gaussian approximative PSF to be $d_R \approx 2.8\sigma_1$ where $\sigma_1 = [\sigma_0^2 + \sigma_{xy}^2]^{1/2}$. We have tested the algorithm for $d_{xy} = d_R$, $d_{xy} = d_R/2$ and $d_{xy} = d_R/3$. The numeric results are shown in Tab. 7.4, where the rate of correctly resolving the two spots is provided for each combination of SNR, fitting method and model-order selection method. We say that the two spots are correctly detected if the algorithm outputs exactly two spots, each center of which falls inside the neighborhood of one of the true spot centers. The neighborhood of a spot center is defined as an ellipsoidal region, which is centered at the spot position and whose principal lateral and axial radii are $1.96\sigma_1$ and $1.96\sigma_2$, respectively. Note that the interval $[-1.96\sigma, 1.96\sigma]$ corresponds to the 95% confidence interval for a normal distribution.

Several conclusions can be drawn by analyzing Tab. 7.4. First, good order-selection

methods for super-resolution are F-test, BIC, MDL, and MDLi. F-test achieves a near perfect detection rate for the sources at the Rayleigh distance. A resolution factor of 2 is reliably gained, i.e. for $d_{xy} = d_R/2$, by F-test ($\alpha = 0.05$), the generalized least-squares fit and SNR ≥ 10, or by F-test ($\alpha = 0.1$), the VST-based fit and SNR ≥ 15. Second, VST-based fit requires in general higher SNRs than for the generalized least-squres fit to resolve the spots. This is expected as the performance of the VST degrades for low Poisson intensities, i.e., small SNRs, such that the asymptotic model (7.44) is no longer valid. Third, F-test loses rapidly its sensitivity as SNR lowers or d_{xy} decreases. For example, the tests totally fail at $d_{xy} = d_R/3$. However, BIC, MDL and MDLi may remain effective (detection rate $\approx 80\%$ – 90%) when the resolvability of F-test drops to zero.

An example of $d_{xy} = d_R/2$ is given in Fig. 7.4 at SNR = 15, where the source centers are plotted in green.

We have also evaluated the axial resolvability of the approach by simulating spots with various axial distances d_z and SNR. The corresponding Rayleigh distance is given by $d_R = 2.8\sigma_2$ where $\sigma_2 = [\sigma_0^2 + \sigma_z^2]^{1/2}$. The results are shown in Tab. 7.5. Similar conclusions as for the lateral resolution case can be drawn. Note that BIC, MDL and MDLi perform less well than for the lateral cases.

Tab. 7.4: Lateral resolution performance

SNR	GLRT*	GLRT	F-test*	F-test	AIC	AICc	BIC	MDL	MDLi
$d_{xy} = d_R$									
5	(32,16)	(32,16)	(**98**,**100**)	(**100**,**100**)	(28,16)	(22,4)	(46,76)	(64,60)	(62,66)
10	(50,12)	(66,16)	(**100**,**100**)	(**100**,**100**)	(52,12)	(50,12)	(**90**,**80**)	(**90**,68)	(**90**,68)
15	(68,24)	(58,10)	(**100**,**100**)	(**100**,**100**)	(64,16)	(60,16)	(**82**,**92**)	(**82**,**94**)	(**82**,**90**)
$d_{xy} = d_R/2$									
5	(48,46)	(48,30)	(0,0)	(0,0)	(46,40)	(38,44)	(54,**90**)	(62,**88**)	(66,**84**)
10	(54,42)	(74,28)	(0,**100**)	(0,**100**)	(48,34)	(50,40)	(**86**,**90**)	(**84**,**90**)	(**86**,**92**)
15	(**84**,48)	(78,50)	(54,**100**)	(**98**,**100**)	(66,52)	(68,54)	(**88**,**86**)	(**88**,**88**)	(**88**,**88**)
$d_{xy} = d_R/3$									
5	(14,32)	(16,26)	(0,0)	(0,0)	(12,22)	(14,30)	(6,42)	(6,52)	(6,42)
10	(74,54)	(**80**,44)	(0,0)	(0,0)	(66,58)	(62,52)	(44,**84**)	(56,**82**)	(56,**88**)
15	(64,44)	(62,38)	(0,0)	(0,0)	(50,42)	(60,36)	(**84**,**80**)	(**84**,**82**)	(**84**,**82**)

The percentages of correctly resolving the two spots are shown. In every parenthesis, the first value is from the VST-based least-squares fit, and the second is given by the generalized least-squares fit. In the pre-detection step, $J = 3$ is used in IUWT and the wavelet-domain FDR is set to 10^{-6}. The p-value for GLRT* and F-test* is $\alpha = 0.05$; that for GLRT and F-test is set to $\alpha = 0.1$. 50 replications are carried out for each order-selection method, each model fitting method, and each SNR. Percentages equal or higher than 80% are emphasized with bold face.

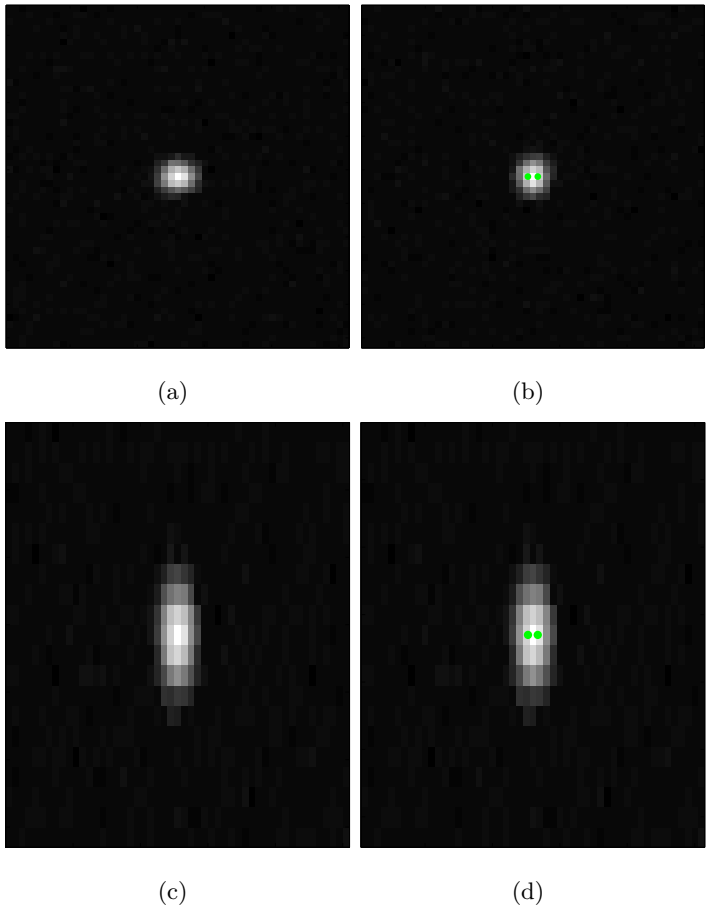


Fig. 7.4: Two simulated close spots separated with a lateral distance $d_{xy} = d_R/2$ at SNR = 15. The two sources can be reliably resolved by F-test ($\alpha = 0.05$) with the generalized least-squares fit, or by F-test ($\alpha = 0.1$) with the VST-based fit. (a) the xy-plane passing through the spot center; (b) the xy-plane as for (a) with spot positions indicated in green; (c) the xz-plane passing through the spot center; (d) the xz-plane as for (c) with spot positions indicated in green. The image intensities are shown in a square-root scale.

Tab. 7.5: Axial resolution performance

SNR	GLRT*	GLRT	F-test*	F-test	AIC	AICc	BIC	MDL	MDLi
$d_z = d_R$									
5	(30,12)	(16,10)	(**100**,**100**)	(**100**,**100**)	(20,10)	(16,4)	(46,68)	(38,68)	(36,60)
10	(34,24)	(22,18)	(**100**,**100**)	(**100**,**100**)	(26,16)	(18,16)	(50,**80**)	(44,76)	(48,74)
15	(28,16)	(16,10)	(**100**,**100**)	(**100**,**100**)	(18,12)	(22,10)	(32,54)	(34,54)	(42,56)
$d_z = d_R/2$									
5	(24,34)	(10,24)	(0,4)	(0,44)	(16,26)	(20,30)	(60,**90**)	(54,**92**)	(40,**80**)
10	(34,38)	(26,36)	(2,**100**)	(54,**100**)	(34,26)	(26,28)	(54,**82**)	(52,**80**)	(54,76)
15	(54,34)	(52,32)	(**100**,**100**)	(**100**,**100**)	(52,30)	(42,32)	(64,76)	(68,72)	(74,72)
$d_z = d_R/3$									
5	(38,50)	(28,42)	(0,0)	(0,0)	(34,54)	(40,40)	(36,**80**)	(50,78)	(50,74)
10	(72,54)	(62,52)	(0,0)	(0,14)	(78,54)	(**80**,52)	(**82**,**80**)	(76,**82**)	(**84**,78)
15	(78,**82**)	(**82**,74)	(0,**82**)	(0,**96**)	(**82**,78)	(**84**,72)	(**88**,**92**)	(**88**,**88**)	(**88**,**90**)

The percentages of correctly resolving the two spots are shown. In every parenthesis, the first value is from the VST-based least-squares fit, and the second is given by the generalized least-squares fit. In the pre-detection step, $J = 3$ is used in IUWT and the wavelet-domain FDR is set to 10^{-6}. The p-value for GLRT* and F-test* is $\alpha = 0.05$; that for GLRT and F-test is set to $\alpha = 0.1$. 50 replications are carried out for each order-selection method, each model fitting method and each SNR. Percentages equal or higher than 80% are emphasized with bold face.

7.2.5 Fast super-resolution detection

In practice, the bottom-up and top-down detections may lead to a high computational complexity. One possible way to accelerate the algorithm is to consider projective models. That is, we integrate the data $(y(\mathbf{x}_k))_k$ along parallel lines or planes to obtain lower dimensional data (2D or 1D). Note that the projected data follow also a MPG distribution since we sum independent MPG variables by projections. In addition, the signal model will remain to be Gaussian mixtures as projections of a Gaussian result in low-dimensinal Gaussian functions. Thus, we can fit the projective model into a few projections. Note that this projection-model approach resembles that proposed in [160], where projective templates for 3D local curvilinear structures are designed and incorporated in an active contour framework to reconstruct DNA filaments from the stereo-micrographs of Cryo-electron microscopy. In

[160], only two projective views are available due to biophysical constraints. However in our case as the original data are available, any number of projections can be produced in any lower dimension. This approach accelerates significantly the detection, but at a cost of sensibility and the resolvability. The reason is that, as most energy of a LSCM or a DSCM PSF concentrates in a local support, the SNR of the projected data can be much lower than that of the original data especially when the projections are applied in a large image.

Another possibility is to fit the model using only a few defocused planes as proposed by [150]. However, this requires a more complex PSF model, e.g., a diffraction model including aberration terms, such that the diffraction patterns change rapidly as a function of defocus. A highly complex PSF model may have no closed forms, which could even slow down the fitting. Another limitation is that if the defocus is large, the SNR on the planes can be so small that we may even lose the detection of an isolated source.

7.3 Conclusion

In this chapter, we have reviewed and studied the limiting resolution of point sources using detection-theoretic, estimation-theoretic and information-theoretic methods. We proposed to use the VST in the theoretical study such that asymptotically the resulting limits have closed-forms for Poisson or MPG observations. It is then immediately clear that the Rayleigh limit can indeed be broken as long as we have a sufficiently high SNR. Practically, we extended an existing super-resolution detection method which unites iterative parametric model fitting and model-order selection. Experiments show that the generalized least-squares fit and the VST-based fit combined with F-test as order selection criterion effectively localize individual spot and close spots with overlapping fluorescence below the Rayleigh's limit.

Chapter 8

Super-resolution Rod Detection

The super-resolution detection approach for spots can be well extended to detect other objects with highly constrained shapes. Here, we will describe the detection of rod-like objects. The application background is as follows. In the study of the bacteria *Shigella flexneri* [161], we want to understand the cellular invasion process. Toward this purpose we have to track these bacteria in microscope videos, and at the first place, we need to detect and localize these objects in each frame. These bacteria are rod-shaped and we will propose below an analytical model for the fluorescence distribution of a rod. Clearly, the same detection-theoretic, estimation-theoretic, and information-theoretic frameworks as in the case of point sources can be carried out to show that highly accurate localization and resolution below the Rayleigh limit are possible. For example, the detection-theoretic bounds remain the same as (7.5) and (7.9) with h being a rod profile; for another instance, [162] derived the CRLB for localizing a line, i.e., a rod with its length tending to infinity.

Parallel to the spot case, super-resolution rod detection consists of a pre-detection and an iterative parametric-model fitting with model-order selections. Experiments show that the method allows to accurately localize the objects and to distinguish overlapping segments with super-resolution. Examples of extracting fluorescent Shigella bacteria by this approach are also shown.

8.1 Parametric Rod Model

The fluorescence distribution of a single rod is modeled by the convolution of the PSF with a finite segment of a length $L = 2l$. The segment is assumed to have a Gaussian cross-

section profile of which the standard deviation is denoted as σ_0. σ_0, which is proportional to the diameter of the cross section, represents thus the thickness of the rod. We assume that typical ranges of the rod length and cross-section diameter are available from prior biophysical knowledges. Then, thanks to the Gaussian approximative PSF the analytical rod model has a closed form and is given by

$$
\begin{aligned}
R(x,y,z) &= C\int_{-1}^{1} l\exp\left[-\frac{(x-tl\sin(\phi)\cos(\theta))^2+(y-tl\sin(\phi)\sin(\theta))^2}{2\sigma_1^2}-\frac{(z-tl\cos(\phi))^2}{2\sigma_2^2}\right]dt \\
&= \frac{A}{2}\exp\left[-\frac{\sigma_2^2(\sin(\theta)\sin(\phi)x-\cos(\theta)\sin(\phi)y)^2}{2\sigma_1^2(\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2)}\right. \\
&\left.-\frac{x^2\cos(\phi)^2+y^2\cos(\phi)^2+z^2\sin(\phi)^2-\cos(\theta)\sin(2\phi)xz-\sin(\theta)\sin(2\phi)yz}{2(\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2)}\right] \\
&\cdot\left(\operatorname{erf}\left[\frac{\sqrt{2}(l\sigma_2^2\sin(\phi)^2+l\cos(\phi)^2\sigma_1^2+\cos(\theta)\sin(\phi)\sigma_2^2x+\sin(\theta)\sin(\phi)\sigma_2^2y+\cos(\phi)\sigma_1^2z)}{2\sigma_1\sigma_2\sqrt{\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2}}\right]\right. \\
&\left.+\operatorname{erf}\left[\frac{\sqrt{2}(l\sigma_2^2\sin(\phi)^2+l\cos(\phi)^2\sigma_1^2-\cos(\theta)\sin(\phi)\sigma_2^2x-\sin(\theta)\sin(\phi)\sigma_2^2y-\cos(\phi)\sigma_1^2z)}{2\sigma_1\sigma_2\sqrt{\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2}}\right]\right) \\
&\cdot\left(\operatorname{erf}\left[\frac{\sqrt{2}l(\sigma_2^2\sin(\phi)^2+\sigma_1^2\cos(\phi)^2)}{2\sigma_1\sigma_2\sqrt{\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2}}\right]\right)^{-1}
\end{aligned}
\tag{8.1}
$$

where A the amplitude, C a normalizing constant so that R has a peak value of A, ϕ and θ are respectively the angle formed by the rod and z-axis and that formed by the rod and x-axis (see Fig. 8.1), $\sigma_1 = [\sigma_{xy}^2+\sigma_0^2]^{1/2}$, $\sigma_1 = [\sigma_z^2+\sigma_0^2]^{1/2}$, and $\operatorname{erf}(x) := 2\Phi(\sqrt{2}x)-1$ is the error function.

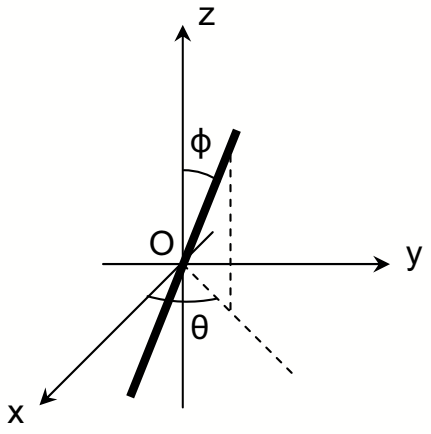


Fig. 8.1: The coordinate system for a rod.

8.1.1 Asymptotic models

It can be verified that as $l \to 0$, the rod model tends to the spot model (Gaussian) as expected, i.e.,

$$R(x,y,z) \overset{l\to 0}{\longrightarrow} A\exp\left[-\frac{x^2+y^2}{2\sigma_1^2}-\frac{z^2}{2\sigma_2^2}\right] \tag{8.2}$$

As $l \to \infty$, the error functions in (8.1) tend to 1, and thus

$$\begin{aligned} R(x,y,z) \overset{l\to\infty}{\longrightarrow} A\exp\Bigg[&-\frac{\sigma_2^2(\sin(\theta)\sin(\phi)x-\cos(\theta)\sin(\phi)y)^2}{2\sigma_1^2(\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2)} \\ &-\frac{x^2\cos(\phi)^2+y^2\cos(\phi)^2+z^2\sin(\phi)^2-\cos(\theta)\sin(2\phi)xz-\sin(\theta)\sin(2\phi)yz}{2(\cos(\phi)^2\sigma_1^2+\sigma_2^2\sin(\phi)^2)}\Bigg] \end{aligned} \tag{8.3}$$

This is the fluorescence distribution for an infinite-length line with a Gaussian cross section.

8.2 Rod Pre-detection

For Gaussian data, we follow the idea of spot pre-detection proposed in [38]. We first perform a matched filtering by convolution of the image y with the Gaussian filter g approximating the microscope PSF to increase the SNR. Then, a pixelic score is constructed for each local maximum in the pre-filtered image as the product of the local average intensity I_M with the mean curvature I_K, i.e.,

$$S(y) := I_K(y)\cdot I_M(y) = (\triangle \star g \star y)\cdot(A \star g \star y) \tag{8.4}$$

where $\triangle$ is a convolution kernel approximating the Laplacian operator, and A is the 3×3 constant averaging matrix. We point out that the Gaussian curvature κ has been used for computing the score of spots [38], but this is less adapted for rods, because a point on a segment has both a large and a small principal curvatures, yielding lower values of the product for κ. Both for this reason and for its linearity, we prefer the mean curvature I_K. This score was found to be highly discriminating for rods embedded in very noisy images.

For an image of Gaussian white noise, I_K and I_M are two correlated normal variables thank to the linearity of the convolution operators. We found that the score S is asymptotically normally distributed as stated in Proposition 9.

Proposition 9 *If $X = (X_i)_{i\in\mathbb{Z}^d}$ are i.i.d. normal variables where $X_i \sim \mathcal{N}(\mu, \sigma^2)$, and H is normalized $\|H\|_1 = 1$, then we have that $S(X) \sim \mathcal{AN}(\mu_S, \sigma_S^2)$, as $\frac{\mu}{\sigma\|H\|_2} \to \infty$, where*

$$\begin{aligned} \mu_S &= \sigma^2\langle H, L\rangle \\ \sigma_S^2 &= \sigma^2\mu^2\|L\|_2^2 + \sigma^4(\langle H, L\rangle^2 + \|H\|_2^2\|L\|_2^2) \end{aligned}$$

Here, $L = \triangle \star g$ is the Laplacian of Gaussian filter, and $H = A \star g$. Proposition 9 can be proved using the asymptotic-normality result of the product of two normal variables [163][164]. The distribution of S being known, we can threshold the score S with a control on the false positive rate.

In order to evaluate μ_S and σ_S, we require the image background μ and the Gaussian noise variance σ^2. Here, we estimate $\hat{\mu}$ from the median of y and $\hat{\sigma}$ from the median absolute deviation estimation [64] of the Daubechies D8 wavelet coefficients at the finest scale. To each of M_0 detected pixels $(x_{i,0}, y_{i,0}, z_{i,0})_{i=1..M_0}$, we associate the orientation angles $\phi_{i,0}$ and $\theta_{i,0}$ computed from the eigenvector of the Hessian corresponding to the lowest eigenvalue, an amplitude $A_{i,0} = I_M(x_{i,0}, y_{i,0}, z_{i,0}) - \hat{\mu}$, and an initial assumed length $L_{i,0} = L_0$.

For MPG data, we apply the same MS-VST+IUWT detection as in the spot case. The image is reconstructed from the significant wavelet coefficients with the coarsest approximation band set to zero. We retain the positive part and extract all local maxima. For each local maximum, we compute the same initial rod parameters as for the Gaussian case.

8.3 Rod Mixture Model Fitting

For Gaussian data, the image model is given by $y(\mathbf{x}_n) = f(\mathbf{x}_n; \Theta_K) + w(\mathbf{x}_n)$ where $w(\mathbf{x}_n) \sim \mathcal{N}(0, \sigma^2)$. For MPG data, the image model is provided by (7.42) with $f(\mathbf{x}_n; \Theta_K)$ being the underlying Poisson intensity. In both cases, $f(\mathbf{x}_n; \Theta_K)$ is modeled by a mixture of rods, i.e.,

$$f(\mathbf{x}_n; \Theta_K) = b + \sum_{i=1}^{K} A_i R_i(\mathbf{x}_n - \mathbf{p}_i) \tag{8.5}$$

where R_i is given by (8.1), $(\mathbf{p}_i)_i$ are the centroids of rods, and the parameter set is given by $\Theta_K := \{(\mathbf{p}_i)_i, (A_i)_i, (\phi_i)_i, (\theta_i)_i, (l_i)_i, \sigma_0, b\}$ for $i = 1, \ldots, K$. Now, the same bottom-up and top-down scheme as in the spot case can be carried out in order to localize an unknown number of rods with super-resolution.

8.4 Experiments

We will evaluate the performance of the algorithm by series of experiments on Gaussian data. For our simulations, the images are generated from the model (8.5) with the following parameters: $\sigma_0 = 0.08\mu m$, $\phi = \theta = \pi/4$, $L = 2\mu m$, readout noise mean $\mu = 5$ and standard deviation $\sigma = 2$. The image size, pixel size, the standard deviations of Gaussian PSF are the same as described in Section 7.2.4. Source amplitude A is varied to produce images at various SNR, defined to be SNR $= A/\sigma$.

8.4.1 Localizing an individual rod

We have tested our method for detecting a single rod. Example images of SNR $\in [5, 15]$ are shown in Fig. 8.2. For each SNR, we have run 20 replications of the detection with least-squares fitting and F-test as model selection method. Numerical results are shown in Tab. 8.1 where we present the correct detection rate, the RMSEs on the coordinates of the rod center, ϕ, θ, and the PREs on the length L and the amplitude A.

As in the case of spot detection, F-test with least-squares fit reliably detects exactly one rod underlying the robustness of the algorithm to noise. It can also be seen that the localization accuracy is within tens of nanometers only for even the lowest SNR. The orientation is estimated with an error around 1 degree. The accuracy of rod length and intensity estimation is also very high, with errors of a few percent only.

Tab. 8.1: Localization performance (Rod)

Order Sel.	CDR%	RMSE_x (nm)	RMSE_y (nm)	RMSE_z (nm)	RMSE_ϕ (deg.)	RMSE_θ (deg.)	PRE (L)%	PRE (A)%
SNR = 5								
F-test	**100**	9.77	9.37	17.75	1.25	0.73	2.73	2.61
SNR = 10								
F-test	**100**	6.36	4.75	8.80	0.42	0.31	0.94	1.08
SNR = 15								
F-test	**100**	3.20	3.27	5.42	0.36	0.25	0.77	0.76

In the pre-detection step, the global false positive rate is set to 10^{-6} using Bonferroni correction. The p-value of F-test is $\alpha = 0.05$ ($\alpha = 0.1$ is also tested and leads to the same results). 20 replications are carried out for each SNR. CDR equal or higher than 80% are emphasized with bold face.

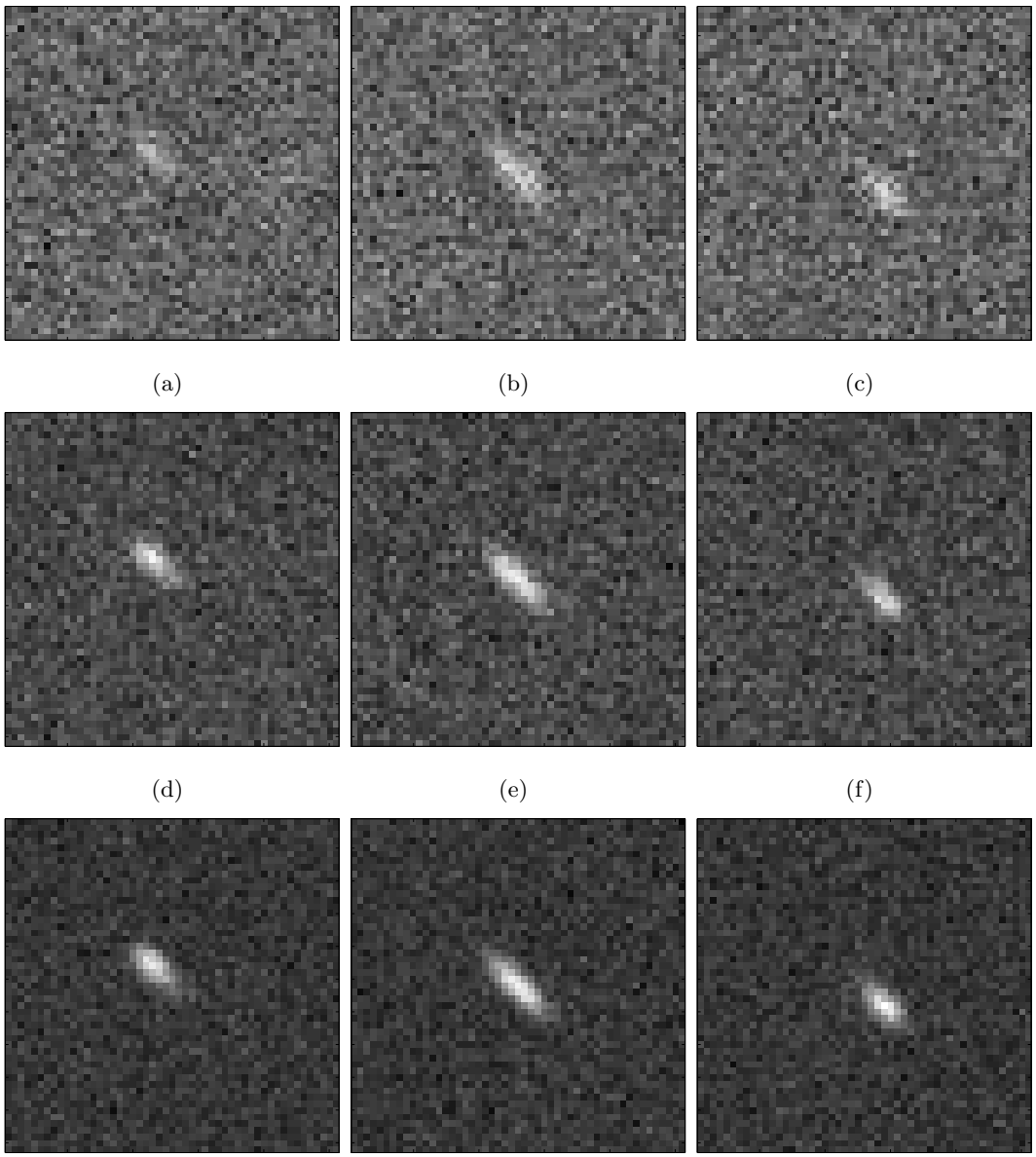


Fig. 8.2: Simulated individual rod at different SNRs. (a) slice $z = 2.4\mu m$, SNR = 5; (b) slice $z = 3\mu m$, SNR = 5; (c) slice $z = 3.6\mu m$, SNR = 5; (d) slice $z = 2.4\mu m$, SNR = 10; (e) slice $z = 2.4\mu m$, SNR = 10; (f) slice $z = 2.4\mu m$, SNR = 10; (g) slice $z = 2.4\mu m$, SNR = 15; (h) slice $z = 2.4\mu m$, SNR = 15; (i) slice $z = 2.4\mu m$, SNR = 15.

8.4.2 Resolving close rods

We test the detection approach in the images where two close rods are simulated at various SNRs. To derive the corresponding Rayleigh limit for the rod case, we adopt the same principle described in Section 7.2.4.2. In our case, the lateral Rayleigh distance is given by $d_R \approx 0.49\mu m$, and the axial one equals $d_R \approx 1.12\mu m$. Our model-order selection method is set to F-test. The correct resolving rates are presented in Tab. 8.2 for $d_{xy} = d_R$, $d_{xy} = d_R/2$ and $d_{xy} = d_R/3$. The numerical results for axial resolution performance are shown in Tab. 8.3. It can be seen that as the SNR increases, a resolution beyond the Rayleigh limit can be gained.

Tab. 8.2: Lateral resolution performance (Rods)

	$d_{xy} = d_R$		$d_{xy} = d_R/2$		$d_{xy} = d_R/3$	
SNR	F-test*	F-test	F-test*	F-test	F-test*	F-test
5	**90**	**90**	0	0	0	0
10	**100**	**100**	0	0	0	0
15	**100**	**100**	40	**80**	0	0
20	**100**	**100**	**100**	**100**	0	0

The percentages of correctly resolving the two rods are shown. In the pre-detection step, the global false positive rate is set to 10^{-6} using Bonferroni correction. The p-value of F-test* is $\alpha = 0.05$; that for F-test is set to $\alpha = 0.1$. 10 replications are carried out for each order-selection method and each SNR. Percentages equal or higher than 80% are emphasized with bold face.

Tab. 8.3: Axial resolution performance (Rods)

	$d_z = d_R$		$d_z = d_R/2$		$d_z = d_R/3$	
SNR	F-test*	F-test	F-test*	F-test	F-test*	F-test
5	**80**	**80**	0	0	0	0
10	**90**	**90**	0	60	0	0
15	**100**	**90**	**100**	**100**	0	0
20	**100**	**100**	**100**	**100**	0	0

The percentages of correctly resolving the two rods are shown. In the pre-detection step, the global false positive rate is set to 10^{-6} using Bonferroni correction. The p-value of F-test* is $\alpha = 0.05$; that for F-test is set to $\alpha = 0.1$. 10 replications are carried out for each order-selection method and each SNR. Percentages equal or higher than 80% are emphasized with bold face.

8.4.3 Detecting Shigella flexneri

Finally, two examples of detecting Shigella bacilli [161] in real images are shown in Fig. 8.3 and Fig. 8.4. The images of bacilli are obtained by a spinning-disc confocal system (UltraView RS; Perkin-Elmer) with an Orca D-ER camera and a 63× Apochromat 1.4 objective. The pixel size is set to $0.205\mu m \times 0.205\mu m \times 0.2\mu m$. The acquisition of these images was ill-conditioned: 1) the system seems to be not well adjusted or have strong aberrations as its PSF size was found to be significantly larger than a theoretical DSCM PSF. For this reason, instead of using the Gaussian parameters shown in chapter 3, we have to empirically adjust their values to be $\sigma_{xy} = 0.2\mu m$ and $\sigma_z = 0.5\mu m$; 2) the noise distribution is found to be approximately symmetric, and thus we have adopted a Gaussian assumption. This could be due to a run-down camera that produces high dominant readout noises. Despite these imperfect conditions, our algorithm still succeeded in detecting the bacilli in the images. In each figure, we show the image model generated from the rod parameters of the detections. In Fig. 8.3, the three bacteria are well detected. The two close rods at the bottom of the image hardly distinguishable by eyes are also precisely identified. The existence of two rods at this place is confirmed by our later video frames where the two bacteria are eventually well separated. This demonstrates the super-resolution ability of our method.

Another example is given by Fig. 8.4 where the three bacilli form a junction in the 3D space. Our approach has resolved this complex spatial configuration by identifying the three bacilli.

8.5 Conclusion

In this chapter, a rod parametric model has been constructed. Then, we have extended the super-resolution spot detection framework to detect an unknown number of rod-like objects. The experiments show that the method allows to accurately localize the objects and to distinguish overlapping segments with super-resolution. We have also applied this approach to extract Shigella bacilli, where two examples are provided demonstrating the effectiveness of the method in resolving close bacteria and complexly configured bacteria.

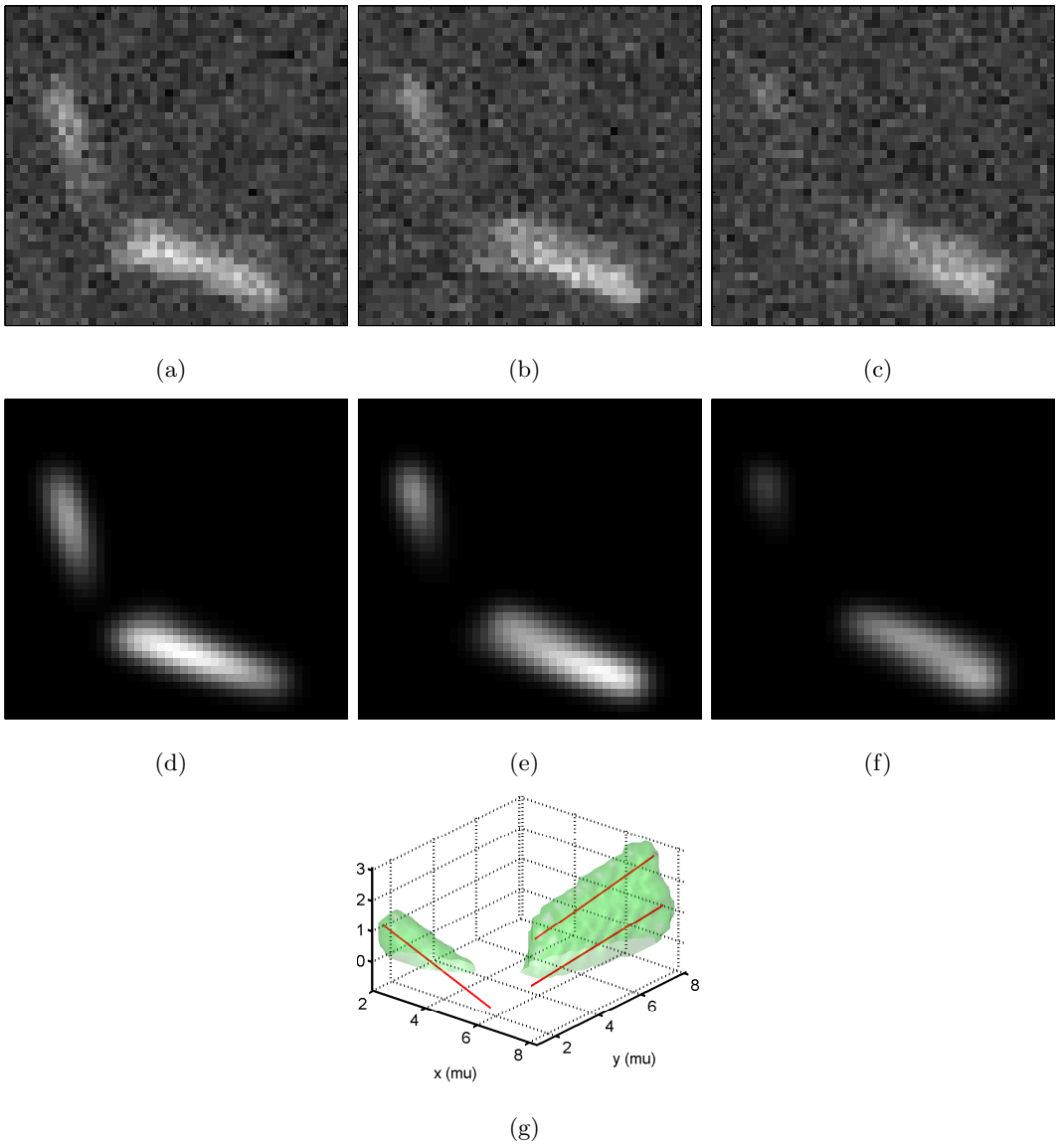


Fig. 8.3: Detection of Shigella bacilli. Three bacilli exist in the image. Two of them at the bottom of the image are very close that are hardly distinguishable by eyes. The algorithm succeeded in identifying the three rods. In the pre-detection step, the global false positive rate is set to 10^{-6} using Bonferroni correction. F-test as model-selection method is used with $\alpha = 0.05$. Observed slice at (a) $z = 0.2\mu m$; (b) $z = 0.8\mu m$; (c) $z = 1.4\mu m$; Model generated from the rod parameters of the detections at (d) $z = 0.2\mu m$; (e) $z = 0.8\mu m$; (f) $z = 1.4\mu m$; (g) 3D reconstruction of the detected rods.

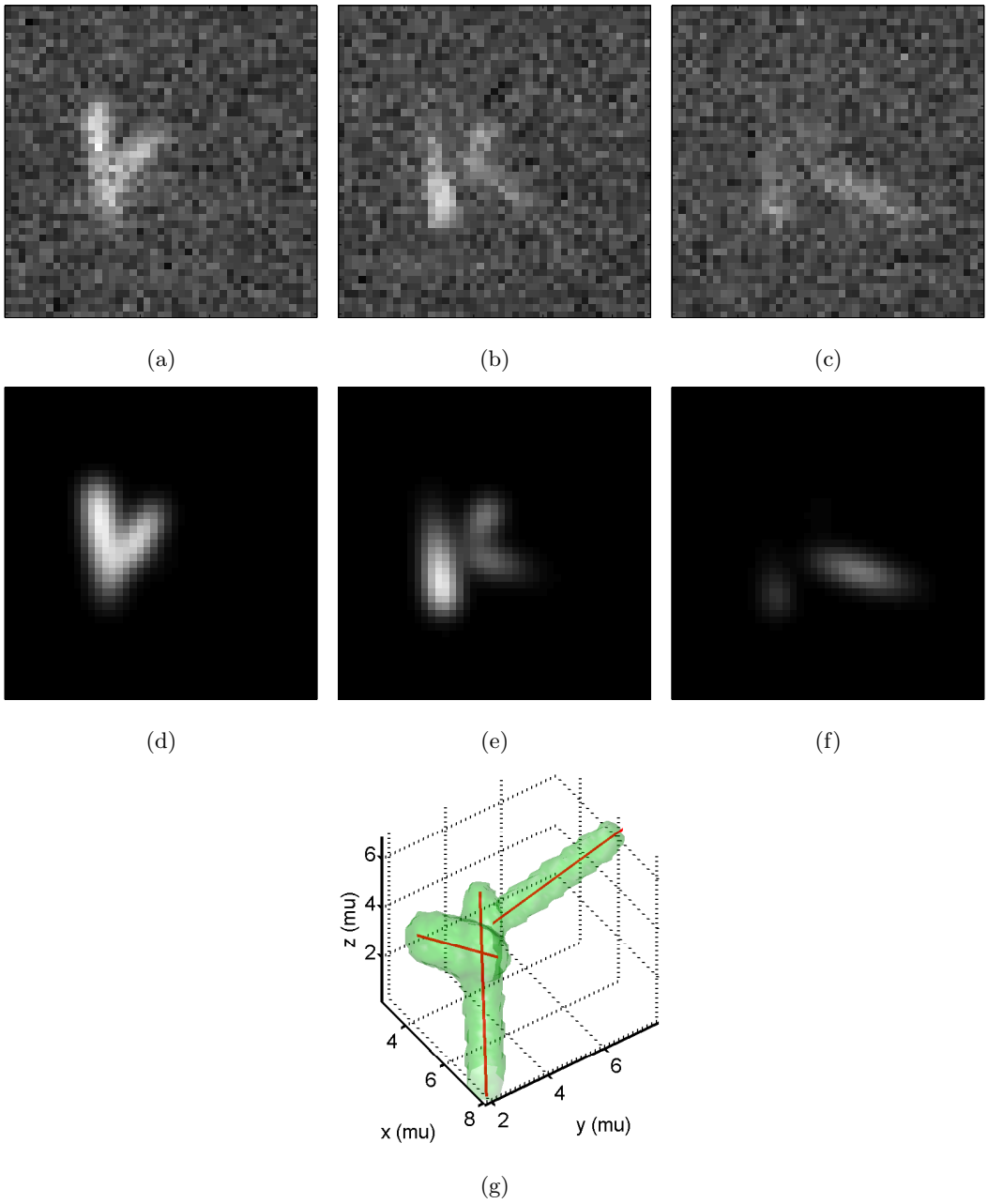


Fig. 8.4: Detection of Shigella bacilli. Three bacilli are in the image forming a junction in the 3D space. The algorithm succeeded in resolving the complex spatial configuration by identifying the three rods. In the pre-detection step, the global false positive rate is set to 10^{-6} using Bonferroni correction. F-test as model-selection method is used with $\alpha = 0.05$. Observed slice at (a) $z = 2.8\mu m$; (b) $z = 3.8\mu m$; (c) $z = 4.8\mu m$; Model generated from the rod parameters of the detections at (d) $z = 2.8\mu m$; (e) $z = 3.8\mu m$; (f) $z = 4.8\mu m$; (g) 3D reconstruction of the detected rods.

Conclusion and Perspectives

The first part of the thesis has discussed the modeling of the optical systems of fluorescence microscopes. We have comprehensively studied the least squares Gaussian approximations of the diffraction-limited 2D/3D paraxial/non-paraxial PSFs of WFFM, LSCM and DSCM described using the Debye diffraction integrals. Optimal Gaussian parameters are derived for the 2D paraxial WFFM PSF, under both the L^∞ and L^1 normalizations. For the other PSFs, with the L^∞ normalization, near-optimal parameters in explicit forms are derived using Maclaurin series matching. We found that the 2D approximations are all very accurate; no accurate Gaussian approximation exists for 3D WFFM PSFs; with typical pinhole sizes, the 3D approximations are accurate for DSCM and nearly perfect for LSCM. These Gaussian approximative PSF models allow fast computation and greatly simplify the modeling of biological objects under these microscopes.

This work can be extended along several lines. First, in order to treat more accurately the non-paraxial diffraction, the approximations can be studied using a vectorial Debye diffraction model instead of the current scalar one. The only detail is that, as the PSF is no longer circularly symmetric about the z-axis in general except for a circularly polarized laser [11], the standard deviations on x-axis and y-axis of the Gaussian should be different. Second, we could also consider the approximation in the presence of aberrations, e.g., spherical aberration. This is, however, not trivial as the PSF looses the mirror symmetry about the xy-plane. Thereby, approximations with functions other than Gaussians may be more appropriate.

The second problem we addressed is the fluorescence image denoising. The images produced by LSCM and DSCM are of either Poisson or MPG statistics. Concerning the Poisson case, we have proposed two approaches for removing the noise. One method relies on biorthogonal Haar-domain hypothesis tests, which is suitable for fast estimating smooth

intensities from large datasets. Our second method makes use of a well designed VST allowing to Gaussianize and stabilize a filtered Poisson process. This VST is then combined with wavelets, ridgelets and curvelets yielding MS-VSTs. We have shown the MS-VST approach to be very effective in recovering important structures of various (isotropic, line-like and curvilinear) shapes within (very) low-count images. This MS-VST method has also been extended to remove MPG noise and to extract fluorescent spots embedded in MPG noisy data.

Concerning the future work, first, the curvelet denoising could be improved if the VST is applied after the Radon transform in the local ridgelet transform step, rather than on the wavelet coefficients as proposed here. This is however not trivial and requires further investigations. Second, new multiscale transforms have been recently proposed such as the fast curvelet transform [165] and the wave atom transform [166], and it would also be very interesting to investigate how our MS-VST could be linked to them. Third, here we have considered the denoising with a single multiscale transform only. If the data contains features with different morphologies, it could be better to introduce several multiscale transforms in the denoising algorithm. This could be done in a very similar way as in the Gaussian noise case [54]. Finally, it is worth investigating how this MS-VST framework can be applied in deconvolution problems.

The third part of our work has focused on super-resolution object detection. We have reviewed and extended the results of resolution limits for point-like sources under detection-theoretic, estimation-theoretic and information-theoretic points of view. In particular, we propose to apply the VST to study the limiting resolution with Poisson or MPG data, leading to asymptotically consistent results with closed-form expressions. We have seen that it is theoretically possible to localize highly precisely an individual point source and to identify close sources with a separation smaller than Rayleigh's distance, provided that we have a sufficient SNR. In practice, a general framework has also been proposed based on an existing super-resolution approach to localize an unknown number of point-like or rod-like sources. This method allows not only to localize sources having complex spatial configurations, but also to detect close objects with super-resolution.

Currently, this proposed approach presents two limitations. First, the image model is considered as a mixture of the sources of interest plus a uniform background. In practice, the background can be highly complex which may contain other biological structures such as membranes, cell kernels, actin filaments, etc. For Gaussian observations with piecewise

regular backgrounds, we may consider the model fitting in the wavelet domain (or some band-pass transformed domain). The wavelet vanishing moments automatically cancel out most regular background such that we are brought to the uniform background case. For MPG observations, a VST preprocessing may be necessary to obtain additive Gaussian noise before applying the wavelet fit. The wavelet (or some band-pass filter) should be well designed such that we can obtain close-form expressions for the source profiles in the transform domain. For highly structured background, the detection could be achieved by combining the source models with the source separation framework [167]. Second, the iterative model fitting and model selection is a time consuming procedure. Projective model fitting and fitting with a few defocused planes may accelerate the algorithm but both present some limitations. Hence, fast super-resolution detection requires more investigations in our future work.

Appendix

Appendix A

Proofs of the Results

A.1 Proof of Lemma 1

proof: On one hand, as $U, U' \in \mathcal{C}^2(W)$, the divergence theorem tells us

$$\begin{aligned}\int_{\mathcal{W}} U\triangle U' - U'\triangle U\, dW &= \int_{\mathcal{W}} \nabla\cdot(U\nabla U' - U'\nabla U)\, dW \\ &= -\int_{\mathcal{S}} U\frac{\partial U'}{\partial \mathbf{n}} - U'\frac{\partial U}{\partial \mathbf{n}}\, dS \end{aligned} \tag{A.1}$$

On the other hand, as U and U' satisfy (1.3), we have

$$\int_{\mathcal{W}} U\triangle U' - U'\triangle U\, dW = 0 \tag{A.2}$$

This shows (1.4). □

A.2 Proof of Theorem 1

proof: The auxiliary function U' has a singularity at $s = 0$. Thus, we will surround P by a small sphere S' of radius ϵ (see Fig. A.1) and we have by Lemma 1 that

$$\int_{\mathcal{S}\cup\mathcal{S}'} U\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial \mathbf{n}}\, dS = 0 \tag{A.3}$$

Hence,

$$\begin{aligned}\int_{\mathcal{S}} U\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial \mathbf{n}}\, dS &= -\int_{\mathcal{S}'} U\frac{e^{iks}}{s}\left(ik - \frac{1}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial \mathbf{n}}\, dS' \\ &= -\int_{\Omega}\left[U\frac{e^{ik\epsilon}}{\epsilon}\left(ik - \frac{1}{\epsilon}\right) - \frac{e^{ik\epsilon}}{\epsilon}\frac{\partial U}{\partial \mathbf{n}}\right]\epsilon^2\, d\Omega \\ &= 4\pi U(P), \quad \epsilon \to 0\end{aligned}$$

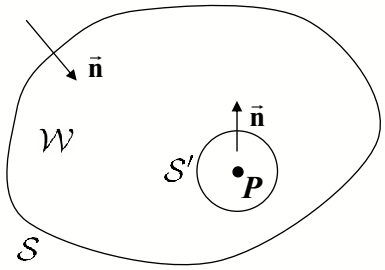


Fig. A.1: Proof of the integral theorem of Helmholtz and Kirchhoff

where $d\Omega$ is an element of the solid angle. Then, (1.5) follows. □

A.3 Proof of Theorem 2

proof: By Theorem 1, we have

$$U(P) = \frac{1}{4\pi}\int_{\mathcal{A}\cup\mathcal{B}\cup\mathcal{C}} U\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\frac{\partial U}{\partial \mathbf{n}}\, dS$$

The integral on $\mathcal{B}$ equals zero by the hypothesis. Let us verify that the integral on $\mathcal{C}$ vanishes as $R \to +\infty$. As $\frac{e^{iks}}{s} = \frac{e^{ikR}}{R}$ on $\mathcal{C}$, we have that

$$\frac{1}{4\pi}\int_{\mathcal{C}} U\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{ikR}}{R}\right) - \frac{e^{ikR}}{R}\frac{\partial U}{\partial \mathbf{n}}\, dS = -\frac{1}{4\pi}\frac{e^{ikR}}{R}\int_{\mathcal{C}} U\left(ik - \frac{1}{R}\right) + \frac{\partial U}{\partial \mathbf{n}}\, dS$$
$$= -\frac{1}{4\pi}e^{ikR}\left(\int_{\Omega} R\left[ikU + \frac{\partial U}{\partial \mathbf{n}}\right]\, d\Omega - \int_{\Omega} U\, d\Omega\right) \overset{R\to+\infty}{\longrightarrow} 0$$

where Ω is the solid angle subtended by $\mathcal{C}$ at P. The limit zero is obtained by the hypothesis on $\mathcal{C}$ and the fact that U is vanishing as R tends to infinity.

Now the original integral involves only the surface $\mathcal{A}$. We can verify that

$$\frac{\partial}{\partial \mathbf{n}}\left(\frac{e^{iks}}{s}\right) = \frac{e^{iks}}{s}\left(ik - \frac{1}{s}\right)\cos(\mathbf{n}, \mathbf{s})$$

and

$$\frac{\partial U}{\partial \mathbf{n}} = \frac{e^{-ikf}}{f}\left(ik + \frac{1}{f}\right)A + \frac{e^{-ikf}}{f}\frac{\partial A}{\partial \mathbf{n}} = \frac{e^{-ikf}}{f}\left(ik + \frac{1}{f}\right)A$$

Now, we can rewrite the original integral as

$$U(P) = \frac{1}{4\pi}\int_{\mathcal{A}} \frac{e^{ik(s-f)}}{sf}\left[A\left(ik - \frac{1}{s}\right)\cos(\mathbf{n}, \mathbf{s}) - A\left(ik + \frac{1}{f}\right)\right]\, dS$$

As $f, s \gg \lambda$, we have that $ik - \frac{1}{s} \approx ik$, and $ik - \frac{1}{f} \approx ik$. Then, (1.16) follows. □

A.4 Proof of Proposition 1

proof: The sufficiency is obvious and we prove the necessity as follows. Let us write

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_{xy}, \mu_z)^T, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{13}^T & \sigma_{33} \end{bmatrix}$$

where $\boldsymbol{\mu}_{xy} = (\mu_x, \mu_y)^T$, and Σ_{11} and Σ_{13} are matrices of sizes 2×2 and 2×1 respectively. Then the conditional distribution of z given $\mathbf{c} := (x, y)^T = (x_0, y_0)^T$ is an 1D Gaussian g_1 with the mean and the variance respectively as

$$\mu_{z|xy} = \mu_z + \Sigma_{13}^T \Sigma_{11}^{-1} (\mathbf{c} - \boldsymbol{\mu}_{xy}), \quad \Sigma_{z|xy} = \sigma_{33} - \Sigma_{13}^T \Sigma_{11}^{-1} \Sigma_{13}$$

Note that $\Sigma_{z|xy}$ is given by the Schur complement of Σ with respect to Σ_{11}. Since g_Σ has mirror symmetry about the xy-plane, g_1 is centered at $z = 0$. This is true for any $\mathbf{c}$, which shows that $\mu_z = 0$ and $\Sigma_{13} = \mathbf{0}$. Similarly, we can write out the conditional distribution of (x, y) given $z = z_0$, which is a 2D Gaussian. Then, using the fact that Σ is positive definite and the circular symmetry of g_Σ, it can be easily shown that $\boldsymbol{\mu}_{xy} = \mathbf{0}$ and that Σ_{11} is diagonal with $\sigma_{11} = \sigma_{22}$. □

A.5 Proofs of Proposition 2 and Proposition 3

We recall the expressions of the Airy function h and the Gaussian function g_σ:

$$h(\rho) = \left(2 \frac{J_1(c\rho)}{c\rho} \right)^2 \quad \text{and} \quad g_\sigma(\rho) = \exp\left(-\frac{\rho^2}{2\sigma^2} \right)$$

where $c = k_{em}\text{NA}$. We can rewrite the objective function to minimize by including the constraint of L^p normalization for $1 \leq p \leq \infty$,

$$E_p(\sigma) := \left\| \frac{h}{\|h\|_p} - \frac{g_\sigma}{\|g_\sigma\|_p} \right\|_2^2$$

In the following, I_n denotes the n-th order modified Bessel functions of the first kind.

proof: We will only prove the case $p = 1$, as the proof for the case $p = \infty$ is completely similar.

As $\|h\|_1 = \frac{4\pi}{c^2}$ (cf. Ref. [168]), we have (the calculus details are in the proof of Proposition 4):

$$\frac{\partial E_1}{\partial \sigma}(\sigma) = \frac{3e^{c^2\sigma^2} - 4I_0(c^2\sigma^2) - 8I_1(c^2\sigma^2)}{2\pi\sigma^3 e^{c^2\sigma^2}} \tag{A.4}$$

We define the numerator of (A.4) as $f(u) := 3e^u - 4I_0(u) - 8I_1(u)$ where $u = c^2\sigma^2$. Then we have, on one hand, $f(u) < 0$ in a right neighborhood of the origin; on the other hand, $f(u) \to +\infty$ as $u \to +\infty$. Therefore, there must exist at least one u^* such that $f(u^*) = 0$.

We will prove the uniqueness of u^*. The series expansion of $f''(u)$ is given by:

$$\begin{aligned} f''(u) &= 1 + \sum_{k=1}^{\infty}(k+1)(k+2)c_k \cdot u^k \\ c_k &= \begin{cases} \frac{3}{(2n)!} - \frac{4}{2^{2n}(n!)^2} & k = 2n-2 \\ \frac{3}{(2n+1)!} - \frac{4}{2^{2n}n!(n+1)!} & k = 2n-1 \end{cases} \qquad n = 1, 2, \ldots \end{aligned}$$

It can be shown from induction that $c_k \geq 0$, so $f''(u) > 0$ for $u \geq 0$. This strict convexity of f on $[0, +\infty)$ together with the fact that $f(0) = -1 < 0$ ensures the uniqueness of u^*.

We have thus $f(u) < 0$ on $(0, u^*)$ and $f(u) > 0$ on $(u^*, +\infty)$. Numerically, $u^* \approx 1.9116$. The proof is completed by noting the positivity of the denominator of equation (A.4). □

A.6 Proof of Proposition 4

proof: The case $p = \infty$ is proven in Ref. [169]. For $1 \leq p < \infty$, note that $\|g_\sigma\|_p = (2\pi)^{\frac{1}{p}}\sigma^{\frac{2}{p}}p^{-\frac{1}{p}}$, and $\|h\|_p$ is a constant independent of σ. Then, we differentiate the objective function with respect to σ, and for $\sigma > 0$ the differentiate operator and the integral can be shown to be interchangeable:

$$\begin{aligned} \frac{d}{d\sigma}\left\|\frac{h}{\|h\|_p} - \frac{g_\sigma}{\|g_\sigma\|_p}\right\|_2^2 &= 2\pi\int_0^\infty \rho\frac{d}{d\sigma}\left(\frac{h(\rho)}{\|h\|_p} - \frac{g_\sigma(\rho)}{\|g_\sigma\|_p}\right)^2 d\rho \\ &= 4\pi\left(\frac{p}{2\pi}\right)^{\frac{1}{p}}(T_0 + T_1 + T_2) \end{aligned} \tag{A.5}$$

where

$$\begin{aligned} T_0 &= \left(\frac{p}{2\pi}\right)^{\frac{1}{p}}\sigma^{-\frac{4}{p}-1}\int_0^\infty \exp\left(-\frac{\rho^2}{\sigma^2}\right)\left[\sigma^{-2}\rho^2 - \frac{2}{p}\right]\rho\, d\rho = \left(\frac{p}{2\pi}\right)^{\frac{1}{p}}\frac{p-2}{2p}\sigma^{1-\frac{4}{p}} \\ T_1 &= \frac{8\sigma^{-\frac{2}{p}-1}}{pc^2\|h\|_p}\int_0^\infty J_1^2(c\rho)\exp\left(-\frac{\rho^2}{2\sigma^2}\right)\rho^{-1}\, d\rho = \frac{4\sigma^{-\frac{2}{p}+1}}{p\|h\|_p}\cdot\frac{1-\exp(-c^2\sigma^2)[I_0(c^2\sigma^2)+I_1(c^2\sigma^2)]}{c^2\sigma^2} \\ T_2 &= -\frac{4\sigma^{-\frac{2}{p}-3}}{\|h\|_p c^2}\int_0^\infty J_1^2(c\rho)\exp\left(-\frac{\rho^2}{2\sigma^2}\right)\rho\, d\rho = -\frac{4\sigma^{-\frac{2}{p}-1}}{\|h\|_p c^2}\exp(-c^2\sigma^2)I_1(c^2\sigma^2) \end{aligned}$$

The results of the integrals T_1 and T_2 are taken from Ref. [168]. Finally, the equation (3.17) is obtained by setting (A.5) to zero. □

A.7 List of Maclaurin Expansions of the PSFs

To simplify the expressions, we define the intensity distributions given by the paraxial and non-paraxial Debye integrals i.e. h_p and h_{np}, and the Gaussian function g_{σ_ρ,σ_z} as follows.

$$\begin{aligned} h_p(x,y,z;\lambda) &:= \left| \int_0^1 J_0(kt\sqrt{x^2+y^2}\sin\alpha) e^{-\frac{i}{2}kzt^2\sin^2\alpha} t\, dt \right|^2 \\ h_{np}(x,y,z;\lambda) &:= \left| \int_0^\alpha \sqrt{\cos\theta} J_0(k\sqrt{x^2+y^2}\sin\theta) e^{ikz\cos\theta}\sin\theta\, d\theta \right|^2 \\ g_{\sigma_\rho,\sigma_z}(x,y,z) &:= \exp\left(-\frac{x^2+y^2}{2\sigma_\rho^2} - \frac{z^2}{2\sigma_z^2}\right) \end{aligned}$$

where the wavenumber in the object space $k = n\frac{2\pi}{\lambda}$. The excitation and emission wavenumbers are denoted as $k_{ex} = \frac{2\pi}{\lambda_{ex}}$ and $k_{em} = \frac{2\pi}{\lambda_{em}}$, respectively. We further denote $\mathbf{x} := (x,y,z)^T$.

Below shows the second-order Maclaurin expansion of the 3D separable Gaussian function:

$$g_{\sigma_\rho,\sigma_z}(x,y,z) = 1 - \frac{1}{2\sigma_\rho^2}(x^2+y^2) - \frac{1}{2\sigma_z^2}z^2 + o(\|\mathbf{x}\|^2) \tag{A.6}$$

The expansions of the 3D paraxial and non-paraxial WFFM PSFs are given by:

$$4h_p(x,y,z;\lambda_{em}) = 1 - \frac{k_{em}^2\mathrm{NA}^2}{4}(x^2+y^2) - \frac{k_{em}^2\mathrm{NA}^2\sin^2\alpha}{48}z^2 + o(\|\mathbf{x}\|^2) \tag{A.7}$$

$$\begin{aligned} &\frac{9}{4(1-\cos^{\frac{3}{2}}\alpha)^2}h_{np}(x,y,z;\lambda_{em}) = 1 - \frac{n^2k_{em}^2(4-7\cos^{\frac{3}{2}}\alpha+3\cos^{\frac{7}{2}}\alpha)}{14(1-\cos^{\frac{3}{2}}\alpha)}(x^2+y^2) \\ &-\frac{3n^2k_{em}^2\left(4+4\cos^5\alpha-25\cos^{\frac{7}{2}}\alpha+42\cos^{\frac{5}{2}}\alpha-25\cos^{\frac{3}{2}}\alpha\right)}{175(1-\cos^{\frac{3}{2}}\alpha)^2}z^2 + o(\|\mathbf{x}\|^2) \end{aligned} \tag{A.8}$$

By denoting $c_1 := k_{ex}r\mathrm{NA}$ and $c_2 := k_{em}r\mathrm{NA}$, the expansion of the 3D paraxial LSCM PSF is given by:

$$\begin{aligned} &\frac{4k_{em}^2\mathrm{NA}^2}{\pi[1-J_0^2(c_2)-J_1^2(c_2)]}h_p(x,y,z;\lambda_{ex})\int_{t_1^2+t_2^2\le r^2} h_p(x-t_1,y-t_2,z;\lambda_{em})\,dt_1dt_2 \\ &= 1 - \frac{1}{4}\left[\frac{c_1^2}{r^2} + \frac{4c_2J_0(c_2)J_1(c_2) - 8J_1^2(c_2)}{r^2[J_0^2(c_2)+J_1^2(c_2)-1]}\right](x^2+y^2) \\ &-\frac{1}{48}\left[\frac{c_1^2\mathrm{NA}^2}{r^2n^2} - \frac{48c_2^2[J_0^2(c_2)+J_1^2(c_2)] - 192J_1^2(c_2)}{n^2k_{em}^2r^4[J_0^2(c_2)+J_1^2(c_2)-1]}\right]z^2 + o(\|\mathbf{x}\|^2) \end{aligned} \tag{A.9}$$

If the excitation PSF and the emission PSF are modeled by Gaussian functions $g_{\sigma_{ex,\rho},\sigma_{ex,z}}$

and $g_{\sigma_{em,\rho},\sigma_{em,z}}$, respectively, the expansion of the 3D LSCM PSF is given by:

$$\begin{aligned}
&\frac{\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)}{2\pi\sigma_{em,\rho}^2\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]} g_{\sigma_{ex,\rho},\sigma_{ex,z}}(x,y,z)\int_{t_1^2+t_2^2\le r^2} g_{\sigma_{em,\rho},\sigma_{em,z}}(x-t_1,y-t_2,z)\,dt_1dt_2\\
&= 1-\frac{1}{4}\frac{2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]+r^2\sigma_{ex,\rho}^2}{\sigma_{ex,\rho}^2\sigma_{em,\rho}^4\left[\exp\left(\frac{r^2}{2\sigma_{em,\rho}^2}\right)-1\right]}(x^2+y^2)\\
&-\frac{1}{2}\frac{\sigma_{ex,z}^2+\sigma_{em,z}^2}{\sigma_{ex,z}^2\sigma_{em,z}^2}z^2+o(\|\mathbf{x}\|^2) \qquad \text{(A.10)}
\end{aligned}$$

A.8 Proof of Proposition 5

proof: We note that

$$p_H = \Pr(d_j^h \ge 2^{-cj}k_0|H_0) = \sum_{k\ge k_0} e^{-\lambda_j} I_k(\lambda_j) = \sum_{k\le -k_0} e^{-\lambda_j} I_{|k|}(\lambda_j)$$

where $k_0 \ge 1$. The p-value of d_{BH} is given by

$$\begin{aligned}
p_{BH} &= \Pr(X_1 - X_2 + 8(X_3 - X_4) \ge \lceil 8k_0/r\rceil | H_0)\\
&= \sum_{k\in\mathbb{Z}} \Pr(X_3 - X_4 = k|H_0) \sum_{n=\lceil 8k_0/r\rceil}^{\infty} \Pr(X_1 - X_2 = n - 8k|H_0)
\end{aligned}$$

where $X_{1,2} \sim \mathcal{P}(\lambda_j)$, $X_{3,4} \sim \mathcal{P}(\lambda_j/2)$, and $(X_i)_i$ are independent. Now we have,

$$\begin{aligned}
p_{BH} &= \sum_{k\ge k_0} \Pr(X_3 - X_4 = k|H_0)\cdot\\
&\quad \sum_{n=\lceil 8k_0/r\rceil}^{\infty} \Pr(X_1 - X_2 = n - 8k|H_0) + \Pr(X_1 - X_2 = n + 8k|H_0)\\
&+ \sum_{|k|<k_0} \Pr(X_3 - X_4 = k|H_0) \sum_{n=\lceil 8k_0/r\rceil}^{\infty} \Pr(X_1 - X_2 = n - 8k|H_0) \qquad \text{(A.11)}\\
&\le p_H + \sum_{|k|<k_0} e^{-\lambda_j} I_{|k|}(\lambda_j) \underbrace{\sum_{n=\lceil 8k_0/r\rceil}^{\infty} e^{-2\lambda_j} I_{|n-8k|}(2\lambda_j)}_{T}
\end{aligned}$$

To bound T, we use the identity [170] $e^x = I_0(x) + 2\sum_{n=1}^{\infty} I_n(x)$. As $r < 1$, we have

$$T \le e^{-2\lambda_j}\sum_{n\ge 9} I_n(2\lambda_j) = \frac{1}{2}\left[1 - e^{-2\lambda_j}\left(I_0(2\lambda_j) + 2\sum_{m=1}^{8} I_m(2\lambda_j)\right)\right] =: A(\lambda_j)$$

Thus, $p_{BH} \le p_H + A(\lambda_j)(1 - 2p_H)$. As $\lambda \to 0+$, we have that $A(\lambda_j) = \frac{2^{9j-7}}{2835}\lambda^9 + o(\lambda^9)$. □

The same arguments can be carried out to bound p_{BH} in multi-dimensional cases. As an example, let us consider 2D data. A 2D wavelet transform will produce bands of horizontal, vertical and diagonal detail coefficients, i.e., $d_{j;H}$, $d_{j;V}$, and $d_{j;D}$. For horizontal and vertical coefficients, we have that $p_{BH} \leq p_H + A(\lambda_j)(1-2p_H)$, where $\lambda_j := 4^j\lambda$. For diagonal coefficients, it can be shown that $p_{BH} \leq p_H + B(\lambda_j)(1-2p_H)$, where

$$B(\lambda_j) := \frac{1}{2}\left[1 - \sum_{n=-64}^{64}\sum_{k=-\infty}^{\infty} e^{-8\lambda_j} I_{|k|}(4\lambda_j) I_{|n-8k|}(4\lambda_j)\right] \tag{A.12}$$

To see the behavior of $B(\lambda_j)$ as the intensity becomes small, we note $B(\lambda_j) = B_K(\lambda_j) + \epsilon_K$. Here, B_K is given by (A.12) with k ranging from $-K$ to K, and ϵ_K is the residual which can be made arbitrary small as K increases. Then, we have for all $K \geq 8$ that $B_K(\lambda_j) = \frac{8}{567}\lambda_j^9 + o(\lambda_j^9)$. Clearly, this procedure can be continued for higher dimensional cases ($q > 2$).

A.9 Proof of Proposition 6

proof: The facts that $G(m_j) = z_{\alpha/2}$, $z_{\alpha/2} > 0$, $2f - 1 \geq 0$, $m_j > 0$ and $\lambda_j \geq 0$ show (5.27).

Next, when the equality in (5.27) holds, we have:

$$G(m_j) = \sqrt{z_{\alpha/2}^2 + \frac{\lambda_j}{2m_j}(z_{\alpha/2}^2 + 1)} - \sqrt{\frac{\lambda_j}{2m_j}(z_{\alpha/2}^2 + 1)} \leq z_{\alpha/2}$$

The existence and uniqueness of the feasible solution follow from the fact that G is a strictly increasing function under (5.27), and that $G(m_j) \to +\infty$ as $m_j \to +\infty$. □

A.10 Proof of Lemma 2

We first prove the following lemma.

Lemma 4 (Filtered MPG process with a high intensity) *Suppose a filtered MPG process $Y := \sum_i h[i]X_i$, where $X_i := U_i + V_i$, $U_i \sim \mathcal{P}(\lambda)$, $V_i \sim \mathcal{N}(\mu, \sigma^2)$ and all U_i and V_i are mutually independent. Assuming $c \in \mathbb{R}$, $\tau_1 < \infty$, $\tau_2 < +\infty$ and $\|h\|_3 < +\infty$, we have:*

$$\sqrt{\tau_2\lambda}\left(\frac{Y + c}{\tau_2\lambda} - \frac{\tau_1}{\tau_2}\right) \xrightarrow[\lambda\to+\infty]{\mathcal{D}} \mathcal{N}(0,1) \tag{A.13}$$

proof: The characteristic function of the LHS of (A.13) is:

$$\begin{aligned}
& \mathbb{E}\left[\exp\left(jt\frac{\sum_i h[i](U_i+V_i)+c-\tau_1\lambda}{\sqrt{\tau_2\lambda}}\right)\right] \\
= \;& \exp\left[jt\left(\frac{c}{\sqrt{\tau_2\lambda}}-\sqrt{\lambda}\frac{\tau_1}{\sqrt{\tau_2}}\right)\right]\cdot\Pi_i\exp\left[\lambda\left(\exp\left[j\frac{h[i]t}{\sqrt{\tau_2\lambda}}\right]-1\right)\right]\exp\left[j\mu\frac{h[i]t}{\sqrt{\tau_2\lambda}}-\frac{1}{2}\sigma^2\frac{h[i]^2t^2}{\tau_2\lambda}\right] \\
= \;& \exp\left[\frac{jt(c+\mu\tau_1)}{\sqrt{\tau_2\lambda}}-\frac{t^2\sigma^2}{2\lambda}-\frac{t^2}{2}\right]\cdot\exp\left[\lambda\sum_i\sum_{k\geq 3}\frac{j^k}{k!}\frac{h[i]^kt^k}{(\tau_2\lambda)^{k/2}}\right]\overset{\lambda\to+\infty}{\longrightarrow}e^{-\frac{1}{2}t^2}\quad(\forall t\in\mathbb{R})
\end{aligned}$$

Indeed, for any $t\neq 0$, by $\|h\|_3<+\infty$, we have:

$$\begin{aligned}
\lambda\left|\sum_i\sum_{k\geq 3}\frac{j^k}{k!}\frac{h[i]^kt^k}{(\tau_2\lambda)^{k/2}}\right| &\leq \lambda\sum_i\sum_{k\geq 3}\frac{1}{k!}\frac{|h[i]t|^k}{(\tau_2\lambda)^{k/2}}=\lambda\sum_i\left(\exp\left[\frac{|h[i]t|}{\sqrt{\tau_2\lambda}}\right]-1-\frac{|h[i]t|}{\sqrt{\tau_2\lambda}}-\frac{|h[i]t|^2}{2\tau_2\lambda}\right) \\
&= \frac{|t|^3}{3!\tau_2^{3/2}\lambda^{1/2}}\sum_i|h[i]|^3\exp\left[\frac{|h[i]|}{\sqrt{\tau_2\lambda}}\xi\right]\overset{\lambda\to+\infty}{\longrightarrow}0
\end{aligned}$$

where $\xi\in(0,|t|)$. The case of $t=0$ is trivial. We conclude by applying Lévy's continuity theorem. □

Now, we prove Lemma 2.

proof: We define the function $f(x):=\mathrm{sgn}(x)\sqrt{|x|}$. Since $\tau_1\neq 0$, we have $f'(\tau_1/\tau_2)=\sqrt{\tau_2}/(2\sqrt{|\tau_1|})$. By the result of Lemma 4 and by setting $\mu=\sigma=0$, we obtain the desired result by applying the Delta-method with the function f:

$$\sqrt{\tau_2\lambda}\left[f\left(\frac{Y+c}{\tau_2\lambda}\right)-f\left(\frac{\tau_1}{\tau_2}\right)\right]=\mathrm{sgn}(Y+c)\sqrt{|Y+c|}-\mathrm{sgn}(\tau_1)\sqrt{|\tau_1|\lambda}\overset{\mathcal{D}}{\to}\mathcal{N}\left(0,\frac{\tau_2}{4|\tau_1|}\right)$$

□

A.11 Proof of Proposition 7

The proposition will be proved in four stages.

A.11.1 Proof – stage I

Expand $T(Y)$ in the neighborhood of $Y=\mu_Y$, we obtain

$$T(Y)=b\sqrt{Y+c}=b\sqrt{\mu_Y+c}+b\frac{1}{2}\frac{Y-\mu_Y}{\sqrt{\mu_Y+c}}-b\frac{(Y-\mu_Y)^2}{8(\mu_Y+c)^{3/2}}+\cdots+R_s \tag{A.14}$$

where the Lagrangian form of the remainder R_s is given by

$$R_s := b\frac{(-1)^{s-1}(2s-3)!!}{2^s s!}\frac{(Y-\mu_Y)^s}{(\xi+c)^{s-1/2}} \quad (s>1) \tag{A.15}$$

with ξ strictly between μ_Y and Y. The following lemma gives an asymptotic bound on the expectation of the remainder R_s. The proof is provided in the following sections A.11.2 and A.11.3.

Lemma 5 (Asymptotic bound of $\mathbb{E}[|R_s|]$ with a high intensity) *Consider $Y := \sum_i h[i]X_i$ a filtered Poisson process where h is a nonnegative FIR filter with $\tau_1 > 0$. If $s > 1$ and $c > 0$, then $\mathbb{E}[|R_s|] = O_{\lambda\to+\infty}(\lambda^{-\frac{s-1}{2}})$.*

A.11.2 Proof – stage II

First, let us recall the Cramér-Chernoff inequality [171].

Lemma 6 ***(Cramér-Chernoff)*** *Let $(X_i)_{1\le i\le n}$ be i.i.d. real random variables. Consider the sum $S_n := \sum_{i=1}^n X_i$. Let $M(t) := \mathbb{E}\left[e^{tX_1}\right]$ be the moment generating function of X_1 and define $I_X(x) := \sup_{t\in\mathbb{R}}(tx - \log M(t))$ for $x \in \mathbb{R}$ (I_X is thus $[0,+\infty]$ valued). Then, we have,*

$$\Pr(S_n \le nx) \le e^{-nI_X(x)}, \quad x \le \mathbb{E}[X_1]$$

Remark If we define $F(t) := tx - \log M(t)$, we have $F(t) = (x - \mathbb{E}[X_1])t + o(t)$ for small t. Therefore, if $x \neq \mathbb{E}[X_1]$, $I_X(x)$ is strictly positive. It can also be shown by Cauchy-Schwarz inequality that $F(t)$ is concave and is strictly concave if X_i is not almost surely a constant.

Second, we have the following lemma,

Lemma 7 *Consider a filtered Poisson process $Y := \sum_{i=1}^n h[i]U_i$ where $U_i \sim \mathcal{P}(\lambda)$ are independent, and h is an FIR filter with $\tau_1 > 0$. Then, for all $c^* \in (0, \tau_1/\sqrt{\tau_2})$, there exists $\beta > 0$ depending only on h and c^* such that,*

$$\Pr\left(Y \le \lambda(\tau_1 - c^*\sqrt{\tau_2})\right) \le e^{-\lambda\beta}$$

proof: Rewrite Y as follows:

$$Y := \sum_{i=1}^n h[i]U_i = \sum_{i=1}^n h[i]\sum_{j=1}^{\lambda/a} W_{i,j} = \sum_{j=1}^{\lambda/a}\sum_{i=1}^n h[i]W_{i,j} = \sum_{j=1}^{\lambda/a} T_j, \quad T_j := \sum_{i=1}^n h[i]W_{i,j}$$

where $\exists\, a > 0$ such that $\lambda/a \in \mathbb{N}$ and $W_{i,j} \sim \mathcal{P}(a)$ are i.i.d. Poisson variables. It can be noted that $(T_j)_j$ are also i.i.d. variables.

We will apply Lemma 6 on Y. First let us calculate $I_T(x)$ as follows:

$$I_T(x) := \sup_{t\in\mathbb{R}}(tx - \log M_T(t)) = \sup_{t\in\mathbb{R}}\left(tx - \sum_{i=1}^{n} a\left(e^{h[i]t} - 1\right)\right) \tag{A.16}$$

where M_T is the moment generating function of T_1. We will evaluate $I_T(x)$ at $x_0 := a(\tau_1 - c^*\sqrt{\tau_2}) > 0$. Since T_1 is not almost surely a constant, $I_T(x_0)$ must be attained for a unique t_0. Thus, setting $x = x_0$, we take the derivative of the term in the sup of (A.16) and set it to zero, resulting in the equation necessarily satisfied by t_0:

$$\sum_{i=1}^{n} h[i]\left(1 - e^{h[i]t_0}\right) = c^*\sqrt{\tau_2} \tag{A.17}$$

$I_T(x_0)$ is given by:

$$I_T(x_0) = a\beta, \quad \beta = t_0(\tau_1 - c^*\sqrt{\tau_2}) - \sum_{i=1}^{n}\left(e^{h[i]t_0} - 1\right) \tag{A.18}$$

Both (A.17) and (A.18) show that t_0 and β depend only on h and c^*. We have in addition $I_T(x_0) > 0$, since $x_0 < \tau_1 a$. We can now apply Lemma 6, giving:

$$\Pr\left(Y \le x_0\lambda/a\right) = \Pr\left(Y \le \lambda(\tau_1 - c^*\sqrt{\tau_2})\right) \le e^{-I_T(x_0)\lambda/a} = e^{-\lambda\beta}$$

□

A.11.3 Proof – stage III

Now we are at the point to prove Lemma 5.

proof: It can be seen from (A.15) that R_s satisfies:

$$|R_s| \le B_s := \frac{|b|}{2s}\frac{|Y - \mu_Y|^s}{|\xi + c|^{s-\frac{1}{2}}} \tag{A.19}$$

Denote $\mu_Y := \lambda\tau_1$ and $\sigma_Y := \sqrt{\lambda\tau_2}$. We have,

$$\begin{aligned}
\mathbb{E}\left[B_s\right] &= \int \mathbf{1}\Big(y \ge \mu_Y - c^*\lambda^{\frac{1}{2}}\sigma_Y\Big)B_s\, dP_Y + \int \mathbf{1}\Big(0 \le y < \mu_Y - c^*\lambda^{\frac{1}{2}}\sigma_Y\Big)B_s\, dP_Y \\
&\le \frac{|b|}{2s}\frac{\mathbb{E}\left[|Y - \mu_Y|^s\right]}{(\mu_Y - c^*\lambda^{\frac{1}{2}}\sigma_Y + c)^{s-\frac{1}{2}}} + \frac{|b|}{2s}\frac{\mu_Y^s}{c^{s-\frac{1}{2}}}\Pr\left(0 \le Y < \lambda(\tau_1 - c^*\sqrt{\tau_2})\right) \\
&\le \frac{|b|}{2s}\frac{\mathbb{E}\left[|Y - \mu_Y|^s\right]}{(\lambda(\tau_1 - c^*\sqrt{\tau_2}) + c)^{s-\frac{1}{2}}} + \frac{|b|}{2s}\frac{\lambda^s\tau_1^s}{c^{s-\frac{1}{2}}}\cdot e^{-\lambda\beta}
\end{aligned} \tag{A.20}$$

where there exists $c^* \in (0, \tau_1/\sqrt{\tau_2})$ and the second term in (A.20) results from Lemma 7. Then,

$$(A.20) = \mathbb{E}\left[|Y - \mu_Y|^s\right] \cdot O_{\lambda \to +\infty}(\lambda^{-s+\frac{1}{2}})$$

We will conclude by showing that $\tilde{M}_s := \mathbb{E}\left[|Y - \mu_Y|^s\right] = O_{\lambda \to +\infty}(\lambda^{s/2})$. The moment M_n and the cumulant κ_n of the centered random variable $(Y - \mu_Y)$ are related by:

$$\begin{aligned} M_n &= \kappa_n + \sum_{p=2}^{n-2} C_{n-1}^p M_p \kappa_{n-p} \quad (n \geq 2) \\ &= \underbrace{\kappa_n}_{O(\lambda)} + \underbrace{\sum_{p=2}^{n-2} C_{n-1}^p \kappa_p \kappa_{n-p}}_{O(\lambda^2)} + \underbrace{\sum_{p=2}^{n-2} \sum_{q=2}^{p-2} C_{n-1}^p C_{p-1}^q \kappa_{n-p} \kappa_{p-q} \kappa_q}_{O(\lambda^3)} + \cdots \end{aligned} \tag{A.21}$$

It can be shown by induction that M_n is a polynomial of $\kappa_2, \cdots, \kappa_n$, which has a minimal order 1 and a maximal order $\lfloor n/2 \rfloor$. The p-th cumulant of $(Y - \mu_Y)$ is $\kappa_p = \lambda \tau_p$ for $p \geq 2$. Therefore $M_n = O_{\lambda \to +\infty}(\lambda^{\lfloor n/2 \rfloor})$. Consequently, $\tilde{M}_k$ satisfies:

$$\begin{aligned} \tilde{M}_{2k} &:= \mathbb{E}\left[|Y - \mu_Y|^{2k}\right] = M_{2k} = O_{\lambda \to +\infty}(\lambda^k) \\ \tilde{M}_{2k+1} &:= \mathbb{E}\left[|Y - \mu_Y|^{2k+1}\right] = \mathbb{E}\left[|Y - \mu_Y|^k |X - \mu_Y|^{k+1}\right] \leq M_{2k}^{1/2} M_{2k+2}^{1/2} = O_{\lambda \to +\infty}(\lambda^{\frac{2k+1}{2}}) \end{aligned}$$

This shows $\tilde{M}_s = O_{\lambda \to +\infty}(\lambda^{s/2})$. □

A.11.4 Proof – stage IV

We are at the final point to prove Proposition 7.

proof: (i) Using (A.14) and (A.15), we can derive the Taylor expansion of $\mathbb{E}[Z]$ about $\lambda = +\infty$ up to order $s = 3$. Applying Lemma 5, (5.30) follows. (5.31) can be proved similarly. (ii) The parameter c is chosen to cancel out the first-order term in (5.31); the parameter b does not influence the variance stabilization asymptotic rate and it can be set to b_1 to get an asymptotic unit variance. (iii) The last statement follows from Lemma 2. □

A.12 Proofs of Theorem 3 and Theorem 4

We will prove Theorem 3 below, and Theorem 4 can be proved in the same way. We first establish the following lemma.

Lemma 8 *Let* $\mathbf{F}_j := [a_{j-1} + c^{(j-1)}, a_j + c^{(j)}]^T$ *and* $\boldsymbol{\mu}_j := [\tau_1^{(j-1)}\lambda, \tau_1^{(j)}\lambda]^T$. *Suppose* $\tau_1^{(j-1)}, \tau_1^{(j)} < \infty$, $0 < \tau_2^{(j-1)}, \tau_2^{(j)} < +\infty$, *and* $\|h^{(j-1)}\|_3, \|h^{(j)}\|_3 < +\infty$. *Then we have:*

$$\frac{\mathbf{F}_j - \boldsymbol{\mu}_j}{\sqrt{\lambda}} \xrightarrow[\lambda\to+\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_j), \quad \Sigma_j = \begin{bmatrix} \tau_2^{(j-1)} & \langle h^{(j-1)}, h^{(j)} \rangle \\ \langle h^{(j-1)}, h^{(j)} \rangle & \tau_2^{(j)} \end{bmatrix} \tag{A.22}$$

proof: Define $\mathbf{Z} := (\mathbf{F}_j - \boldsymbol{\mu}_j)/\sqrt{\lambda}$ and $\mathbf{c} := [c^{(j-1)}, c^{(j)}]^T$. We calculate the characteristic function of Z as follows:

$$\begin{aligned}
\mathbb{E}\left[\exp(i\langle \mathbf{t}, \mathbf{Z}\rangle)\right] &= \exp\left(-\frac{i}{\sqrt{\lambda}}\langle \mathbf{t}, \boldsymbol{\mu}_j\rangle\right) \mathbb{E}\left[\exp\left(\frac{i}{\sqrt{\lambda}}\langle \mathbf{t}, \mathbf{F}_j\rangle\right)\right] \\
&= \exp\left(-\frac{i}{\sqrt{\lambda}}\langle \mathbf{t}, \boldsymbol{\mu}_j - \mathbf{c}\rangle\right) \Pi_k \mathbb{E}\left[\exp\left(\frac{i}{\sqrt{\lambda}}\left(t_1 h^{(j-1)}[k] + t_2 h^{(j)}[k]\right) X_k\right)\right] \\
&= \exp\left(-\frac{i}{\sqrt{\lambda}}\langle \mathbf{t}, \boldsymbol{\mu}_j - \mathbf{c}\rangle\right) \exp\left[\sum_k \lambda\left(\exp\left(i\frac{t_1 h^{(j-1)}[k] + t_2 h^{(j)}[k]}{\sqrt{\lambda}}\right) - 1\right)\right]
\end{aligned}$$

Similar the proof of Lemma 4, if $\|h^{(j-1)}\|_3, \|h^{(j)}\|_3 < +\infty$, then as $\lambda \to +\infty$,

$$\mathbb{E}\left[\exp(i\langle \mathbf{t}, \mathbf{Z}\rangle)\right] \to \exp\left[-\frac{1}{2}\sum_k (t_1 h^{(j-1)}[k] + t_2 h^{(j)}[k])^2\right] = \exp\left(-\frac{1}{2}\mathbf{t}^T \Sigma_j \mathbf{t}\right)$$

where Σ_j is a positive semi-definite matrix:

$$\Sigma_j = \begin{bmatrix} \sum_k h^{(j-1)}[k]^2 & \sum_k h^{(j-1)}[k]h^{(j)}[k] \\ \sum_k h^{(j-1)}[k]h^{(j)}[k] & \sum_k h^{(j)}[k]^2 \end{bmatrix} = \begin{bmatrix} \tau_2^{(j-1)} & \langle h^{(j-1)}, h^{(j)} \rangle \\ \langle h^{(j-1)}, h^{(j)} \rangle & \tau_2^{(j)} \end{bmatrix}$$

We conclude by applying Lévy's continuity theorem. □

We will now prove Theorem 3:

proof: Define $\boldsymbol{\mu}_0 := [\tau_1^{(j-1)}, \tau_1^{(j)}]^T$ and $g(x_1, x_2) := b^{(j-1)} \operatorname{sgn}(x_1)\sqrt{|x_1|} - b^{(j)} \operatorname{sgn}(x_2)\sqrt{|x_2|}$. Then the gradient of g is given by:

$$\nabla g(x_1, x_2) = \frac{1}{2}\left[b^{(j-1)}|x_1|^{-1/2}, \quad -b^{(j)}|x_2|^{-1/2}\right]^T, \quad x_1, x_2 \neq 0$$

From Lemma 8 we have,

$$\sqrt{\lambda}\left(\frac{\mathbf{F}_j}{\lambda} - \boldsymbol{\mu}_0\right) \xrightarrow[\lambda\to+\infty]{\mathcal{D}} \mathcal{N}(0, \Sigma_j)$$

where Σ_j is given by (A.22). Applying the multivariate Delta-method for the function g gives:

$$\begin{aligned}
&\sqrt{\lambda}\left[b^{(j-1)}\operatorname{sgn}(a_{j-1}+c^{(j-1)})\sqrt{\frac{|a_{j-1}+c^{(j-1)}|}{\lambda}}-b^{(j)}\operatorname{sgn}(a_j+c^{(j)})\sqrt{\frac{|a_j+c^{(j)}|}{\lambda}}-\right.\\
&\left.\left(b^{(j-1)}\operatorname{sgn}(\tau_1^{(j-1)})\sqrt{|\tau_1^{(j-1)}|}-b^{(j)}\operatorname{sgn}(\tau_1^{(j)})\sqrt{|\tau_1^{(j)}|}\right)\right]\\
&=b^{(j-1)}\operatorname{sgn}(a_{j-1}+c^{(j-1)})\sqrt{|a_{j-1}+c^{(j-1)}|}-b^{(j)}\operatorname{sgn}(a_j+c^{(j)})\sqrt{|a_j+c^{(j)}|}\\
&=d_j\xrightarrow[\lambda\to+\infty]{\mathcal{D}}\mathcal{N}(0,\sigma_0^2)
\end{aligned}$$

where,

$$\sigma_0^2=[\nabla g(\boldsymbol{\mu}_0)]^T\,\Sigma_j\,[\nabla g(\boldsymbol{\mu}_0)]=\frac{\tau_2^{(j-1)}}{4{\tau_1^{(j-1)}}^2}+\frac{\tau_2^{(j)}}{4{\tau_1^{(j)}}^2}-\frac{\langle h^{(j-1)},h^{(j)}\rangle}{2\tau_1^{(j-1)}\tau_1^{(j)}}$$

□

A.13 Proof of Theorem 5

We will first prove Lemma 9. Given a Hibert space $\mathcal{H}$ with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and induced norm $\|\cdot\|_{\mathcal{H}}$, we call a mapping $V:\mathcal{H}\to\mathcal{H}$ *nonexpansive* if for all $x,y\in\mathcal{H}$, $\|Vx-Vy\|_{\mathcal{H}}\leq\|x-y\|_{\mathcal{H}}$. Suppose that a mapping $V:\mathcal{H}\to\mathcal{H}$ is nonexpansive and $Fix(V)\neq\emptyset$. Then V is *attracting* (w.r.t. $Fix(V)$) if for every $x\notin Fix(V), y\in Fix(V)$, we have $\|Vx-y\|_{\mathcal{H}}<\|x-y\|_{\mathcal{H}}$. Properties of nonexpansive and attracting mappings can be found in [172]. We also define η-strongly monotone mappings as follows. For a given set $S\subset\mathcal{H}$, a mapping $V:\mathcal{H}\to\mathcal{H}$ is η-strongly monotone over S if there exists $\eta>0$ such that $\langle Vx-Vy,x-y\rangle_{\mathcal{H}}\geq\eta\|x-y\|_{\mathcal{H}}^2$ for all $x,y\in S$. Let us point out that in our case, $\mathcal{H}=\mathbb{R}^L$.

Lemma 9 *With the same notations as in Theorem 5, we have:*

(a) *$\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{C}_B$ are all closed convex nonempty sets;*

(b) *$P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_3}$ are attracting, and $Fix(P_{\mathcal{S}_1})=\mathcal{S}_1$ and $Fix(P_{\mathcal{S}_3})=\mathcal{S}_3$;*

(c) *$Fix(Q_{\mathcal{S}_2})=\mathcal{S}_2$, and if $\mathcal{W}$ represents a tight frame and $\mathcal{R}$ is the pseudo-inverse operator, then $Q_{\mathcal{S}_2}$ is nonexpansive;*

(d) *If V_1 is attracting, V_2 is nonexpansive, and $Fix(V_1)\cap Fix(V_2)\neq\emptyset$, then $V:=V_1\circ V_2$ is nonexpansive with $Fix(V)=Fix(V_1)\cap Fix(V_2)$.*

proof: (a) As $\|\mathbf{d}\|_{\mathcal{H}}^2 = \|\mathbf{d}\|_2^2 \leq \|\mathbf{d}\|_1^2$, we have that $\mathcal{W}\mathbf{x} \in \mathcal{C}_B$. $\mathcal{C}_B$ is thus nonempty. The closeness and convexity can also be easily verified. (b) results from the fact that $P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_3}$ are the projectors onto $\mathcal{S}_1$ and $\mathcal{S}_3$, respectively. To prove (c), we have by assumption on $\mathcal{W}$ and $\mathcal{R}$ that $\|\mathcal{W}\|\|\mathcal{R}\| = 1$ (see [64]). Then using the fact that P_+ is a projector (so nonexpansive), we have:

$$\|Q_{\mathcal{S}_2}x - Q_{\mathcal{S}_2}y\|_{\mathcal{H}} \leq \|\mathcal{W}\|\|P_+\mathcal{R}x - P_+\mathcal{R}y\|_{\mathcal{H}_0} \leq \|\mathcal{W}\|\|\mathcal{R}x - \mathcal{R}y\|_{\mathcal{H}_0} \leq \|\mathcal{W}\|\|\mathcal{R}\|\|x-y\|_{\mathcal{H}} \leq \|x-y\|_{\mathcal{H}}$$

where $\mathcal{H}_0$ is the initial Hilbert space (the space of $\mathbf{x}$). So $Q_{\mathcal{S}_2}$ is nonexpansive. $Fix(Q_{\mathcal{S}_2}) = \mathcal{S}_2$ can be easily verified. To prove (d), V can be easily verified to be nonexpansive. It is obvious that $Fix(V_1) \cap Fix(V_2) \subseteq Fix(V_1 \circ V_2)$. To prove the other inclusion, pick $x \in Fix(V_1 \circ V_2)$. It is sufficient to show that $x \in Fix(V_2)$. Suppose that $x \notin Fix(V_2)$, then necessarily $V_2x \notin Fix(V_1)$. Now pick any $y \in Fix(V_1) \cap Fix(V_2)$. Since V_1 is attracting, we have:

$$\|x - y\|_{\mathcal{H}} = \|V_1 \circ V_2 x - y\|_{\mathcal{H}} < \|V_2 x - y\|_{\mathcal{H}} = \|V_2 x - V_2 y\|_{\mathcal{H}} \leq \|x - y\|_{\mathcal{H}}$$

which is absurd. Thus $Fix(V_1) \cap Fix(V_2) = Fix(V_1 \circ V_2) = Fix(V)$. □

We prove now Theorem 5.

proof: (i) can be easily verified. (ii) is a direct result of the point (d) of Lemma 9. To prove (iii), we note that J_ϵ is convex and $\nabla J_\epsilon(\mathbf{d})[i] = d[i](d[i]^2 + \epsilon)^{-1/2}$. It can be verified that $\nabla J_\epsilon(\mathbf{d})$ is $\epsilon^{-1/2}$-Lipschitzian and $\epsilon(B^2+\epsilon)^{-3/2}$-strongly monotone over $T_{\mathcal{C}_B}(\mathbb{R}^L)$. Then (iii) results from the convergence theorem of HSD [111]. (iv) is obvious. To prove (v), we have for any convergent subsequence of d_ϵ^*, say $d_{\epsilon_j}^* \underset{\epsilon \to 0^+}{\longrightarrow} d_0^*$, that:

$$\forall \mathbf{d} \in \mathcal{C}_B, \quad J(\mathbf{d}_{\epsilon_j}^*) = \sum_{i=1}^{L} |\mathbf{d}_{\epsilon_j}^*[i]| \leq \sum_{i=1}^{L} \sqrt{\mathbf{d}_{\epsilon_j}^*[i]^2 + \epsilon_j} \leq \sum_{i=1}^{L} \sqrt{\mathbf{d}[i]^2 + \epsilon_j} \tag{A.23}$$

Then by taking the limit $\epsilon \to 0^+$ on both sides of (A.23), we have $\|\mathbf{d}_0^*\|_1 \leq \|\mathbf{d}\|_1$. $\mathbf{d}_0^* \in \mathcal{C}_B$ since $\mathcal{C}_B$ is closed. $\mathbf{d}_0^*$ is thus a solution to (5.45) with $\epsilon = 0$, and hence also a solution to (5.44) by (i). □

A.14 Proof of Lemma 3

Lemma 3 can be proved similarly as for Lemma 2.

proof: We have by Lemma 4 that

$$\sqrt{\tau_2\lambda}\left(\frac{Y_k+c}{\alpha\tau_2\lambda}-\frac{\tau_1}{\tau_2}\right)\underset{\lambda\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}}\mathcal{N}(0,1) \tag{A.24}$$

We obtain the desired result by applying the Delta-method with the function $f(x) := b\cdot\operatorname{sgn}(x)\sqrt{\alpha|x|}$ and (A.24). □

A.15 Proof of Theorem 6

The proof Theorem 6 is similar to that of Theorem 3.

proof: We define $F_j := [a_{j-1}+c^{(j-1)}, a_j+c^{(j)}]^T$ and $\mu_0 := \alpha[\tau_1^{(j-1)}, \tau_1^{(j)}]^T$. Lévy's continuity theorem shows that

$$\sqrt{\lambda}\left(\frac{F_j}{\lambda}-\mu_0\right)\underset{\lambda\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}}\mathcal{N}(0,\Sigma_j) \tag{A.25}$$

where

$$\Sigma_j := \alpha^2\begin{bmatrix}\tau_2^{(j-1)} & \langle h^{(j-1)}, h^{(j)}\rangle\\ \langle h^{(j-1)}, h^{(j)}\rangle & \tau_2^{(j)}\end{bmatrix}$$

Define the function $g(x_1,x_2) := b^{(j-1)}\operatorname{sgn}(x_1)\sqrt{|x_1|} - b^{(j)}\operatorname{sgn}(x_2)\sqrt{|x_2|}$, where $b^{(j)} := \operatorname{sgn}(\tau_1^{(j)})/\sqrt{\alpha|\tau_1^{(j)}|}$. Applying the Delta-method with g and (A.25), we obtain the desired asymptotic normality with variance $\sigma_j^2 = \nabla^T g(\mu_0)\Sigma_j\nabla g(\mu_0)$, which is given by (6.9). □

Appendix B

MPG Parameter Estimation by Cumulant Matching

Given n i.i.d. MPG variables $X_i = \alpha U_i + V_i$ where $U_i \sim \mathcal{P}(\lambda)$, $V_i \sim \mathcal{N}(\mu, \sigma^2)$ and U_i and V_i are independent. We want to estimate α, μ and σ^2. The first four cumulants of X are given by

$$\begin{aligned} \kappa_1 &= \alpha\lambda + \mu,\ \kappa_2 = \alpha^2\lambda + \sigma^2 \\ \kappa_3 &= \alpha^3\lambda,\ \kappa_4 = \alpha^4\lambda \end{aligned}$$

Suppose k_1, k_2, k_3, k_4 are k-statistics [125] for the four cumulants, then the MPG parameters can be estimated by

$$\begin{aligned} \hat{\alpha} &= \frac{k_4}{k_3} \\ \hat{\mu} &= k_1 - \frac{k_3}{\hat{\alpha}^2} = k_1 - \frac{k_3^3}{k_4^2} \\ \hat{\sigma^2} &= k_2 - \frac{k_3}{\hat{\alpha}} = k_2 - \frac{k_3^2}{k_4} \end{aligned}$$

Appendix C

MSE Lower Bounds

We briefly summarize the classical results on the MSE lower bounds based on the Cauchy-Schwarz inequality. More details can be found in [173][174].

Lemma 10 (Cauchy-Schwarz inequality) *Given any random vectors* $\mathbf{T}$ *and* $\mathbf{S}$ *and supposing* $\mathbb{E}[\mathbf{S}\mathbf{S}^T]$ *to be invertible, we have:*

$$\mathbb{E}[\mathbf{T}\mathbf{T}^T] \geq \mathbb{E}[\mathbf{T}\mathbf{S}^T]\mathbb{E}^{-1}[\mathbf{S}\mathbf{S}^T]\mathbb{E}[\mathbf{S}\mathbf{T}^T] \tag{C.1}$$

The equality holds iff $\mathbf{T} = \mathbb{E}[\mathbf{T}\mathbf{S}^T]\mathbb{E}^{-1}[\mathbf{S}\mathbf{S}^T]\mathbf{S}$ *a.s.*

proof: Let $\mathbf{Z} = \mathbf{T} - \mathbb{E}[\mathbf{T}\mathbf{S}^T]\mathbb{E}^{-1}[\mathbf{S}\mathbf{S}^T]\mathbf{S}$. We have that

$$0 \leq \mathbb{E}[\mathbf{Z}\mathbf{Z}^T] = \mathbb{E}[\mathbf{T}\mathbf{T}^T] - \mathbb{E}[\mathbf{T}\mathbf{S}^T]\mathbb{E}^{-1}[\mathbf{S}\mathbf{S}^T]\mathbb{E}[\mathbf{S}\mathbf{T}^T]$$

Note that if the equality holds, $\mathbb{E}[Z_i^2] = 0$ for all i. □

We point out that the RHS of (C.1) is invariant under the exchange of rows or columns in $\mathbf{S}$. Below, we will study three well known MSE lower bounds derived using Cauchy-Schwarz inequality, i.e., Cramér-Rao (CRLB), Bhattacharyya (BHLB), and Barankin lower bounds (BLB). Different bounds result from different choices of the score function $\mathbf{S}$.

C.1 Cramér-Rao Lower Bound

Suppose that we have a *pdf* $f(\mathbf{x};\theta)$ of a random vector $\mathbf{X} \in \mathbb{R}^n$, where $\theta \in \mathbb{R}^d$. The Fisher information matrix is a $d \times d$ matrix with element

$$\mathcal{I}_{i,j}(\theta) := \mathbb{E}_\theta \left[\frac{\partial}{\partial \theta_i} \ln f(\mathbf{X};\theta) \frac{\partial}{\partial \theta_j} \ln f(\mathbf{X};\theta) \right] \tag{C.2}$$

We define also the following regularity conditions for f:

1. There exists $\mathcal{D} = \mathcal{A} \cup \mathcal{B} \subseteq \mathbb{R}^n$ where $\mathcal{A} := \{\mathbf{x} | f(\mathbf{x};\theta) > 0\}$ and $\mathcal{B}$ has a zero Lebesgue measure such that $\mathcal{D}$ is independent of θ;

2. $\forall i, \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta_i} f(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial}{\partial \theta_i} \int_{\mathbb{R}^n} f(\mathbf{x};\theta)\, d\mathbf{x}$, and $\forall i,j$, $\int_{\mathbb{R}^n} T_i(\mathbf{x}) \frac{\partial}{\partial \theta_j} f(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial}{\partial \theta_j} \int_{\mathbb{R}^n} T_i(\mathbf{x}) f(\mathbf{x};\theta)\, d\mathbf{x}$, where $T(\mathbf{X})$ is an estimator of θ.

It is straightforward to verify that under the regularity conditions, $\mathcal{I}(\theta) = N\mathcal{I}_1(\theta)$ where $\mathcal{I}$ and $\mathcal{I}_1$ are the Fisher information matrices associated to N i.i.d. variables $(\mathbf{X}_i)_{i=1,\cdots,N}$ and $\mathbf{X}_1$, respectively. CRLB chooses the following score function $S(\mathbf{x},\theta) := \nabla_\theta \ln f(\mathbf{x};\theta)$.

Theorem 7 (CRLB) *Define $\psi_T(\theta) := \mathbb{E}_\theta[T(\mathbf{X})]$. Assuming the regularity conditions and that $\mathcal{I}$ is invertible, then:*

$$Cov_\theta[T(\mathbf{X})] \geq \nabla\psi_T(\theta)\mathcal{I}^{-1}(\theta)\nabla^T\psi_T(\theta) \tag{C.3}$$

The equality (C.3) holds iff $T = \psi_T(\theta) + \nabla\psi_T(\theta)\mathcal{I}^{-1}(\theta)S$ a.s.

proof: We can verify under the assumptions of the theorem that 1) $\mathbb{E}_\theta[S] = 0$; 2) $\forall\, \alpha \in \mathbb{R}^d, \mathbb{E}_\theta[(T-\alpha)S^T] = \mathbb{E}_\theta[TS^T] = \nabla\psi_T(\theta)$; 3) $\mathbb{E}_\theta[SS^T] = \mathcal{I}(\theta)$. Setting $\mathbf{T} := T - \mathbb{E}_\theta[T]$ and $\mathbf{S} := S$, (C.1) leads to (C.3). □

Let us define a collection of estimators $\mathcal{T}_\psi := \{T | \mathbb{E}_\theta[T] = \psi(\theta)\}$. This is the set of T's whose expectations depend only on a fixed function $\psi(\theta)$. Then (C.3) becomes a uniform lower bound for $\mathcal{T}_\psi$. Note that all unbiased estimators form a family $\mathcal{T}_\psi$ with $\psi(\theta) = \theta$.

Corollary 1 (CRLB for MSE)

$$\begin{aligned}
\mathbb{E}_\theta[(T(\mathbf{X})-\theta)(T(\mathbf{X})-\theta)^T] &\geq \underbrace{(\psi_T(\theta)-\theta)(\psi_T(\theta)-\theta)^T}_{bias} + \underbrace{\nabla\psi_T(\theta)\mathcal{I}^{-1}(\theta)\nabla^T\psi_T(\theta)}_{variance} \\
\mathbb{E}_\theta[\|T(\mathbf{X})-\theta\|^2] &\geq \underbrace{\|\psi_T(\theta)-\theta\|^2}_{bias} + \underbrace{tr\left(\nabla\psi_T(\theta)\mathcal{I}^{-1}(\theta)\nabla^T\psi_T(\theta)\right)}_{variance}
\end{aligned}$$

Corollary 2 (CRLB for unbiased estimators) *If $T(\mathbf{X})$ is an unbiased estimator, then:*

$$Cov_\theta[T(\mathbf{X})] \geq \mathcal{I}^{-1}(\theta) \tag{C.4}$$

C.2 Bhattacharyya Lower Bound

If f satisfies some more regularity conditions, tighter bounds than CRLB can be derived. BHLB chooses the following score function:

$$S^T := [\underbrace{f_1, f_2, \ldots, f_d}_{\partial\theta^1}, \underbrace{f_{11}, f_{12}, \ldots, f_{dd}}_{\partial\theta^2}, \ldots, \underbrace{f_{11\cdots1}, \ldots, f_{dd\cdots d}}_{\partial\theta^k}]/f$$

where $f_{i_1 i_2 \ldots i_m} := \partial^m f / \partial\theta_{i_1}\partial\theta_{i_2}\cdots\partial\theta_{i_m}$. Clearly, this choice of S reduces to that in CRLB if $k = 1$. Since more derivatives are involved in S, we need more stringent regularity conditions to derive BHLB:

1. the same condition on $\mathcal{D}$ as for CRLB;
2. $\int \frac{\partial^m}{\partial\theta_{i_1}\partial\theta_{i_1}\cdots\partial\theta_{i_m}} f(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial^m}{\partial\theta_{i_1}\partial\theta_{i_1}\cdots\partial\theta_{i_m}} \int f(\mathbf{x};\theta)\, d\mathbf{x}$, and $\int T(\mathbf{x}) \frac{\partial^m}{\partial\theta_{i_1}\partial\theta_{i_1}\cdots\partial\theta_{i_m}} f(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial^m}{\partial\theta_{i_1}\partial\theta_{i_1}\cdots\partial\theta_{i_m}} \int T(\mathbf{x}) f(\mathbf{x};\theta)\, d\mathbf{x}$, for $1 \le m \le k$.

T is an estimator of θ. We define $\tilde{\mathcal{I}}(\theta) := \mathbb{E}_\theta[SS^T]$. We can verify that 1) $\mathbb{E}_\theta[S] = 0$; 2) $\forall\, \alpha \in \mathbb{R}^d, \mathbb{E}_\theta[(T-\alpha)S^T] = \mathbb{E}_\theta[TS^T] = [\nabla\psi_T(\theta), \nabla^{(2)}\psi_T(\theta), \ldots, \nabla^{(k)}\psi_T(\theta)]$.

Theorem 8 (BHLB) *Under the regularity conditions, we have*

$$Cov_\theta(T) \ge [\nabla\psi_T(\theta), \nabla^{(2)}\psi_T(\theta), \ldots, \nabla^{(k)}\psi_T(\theta)]\tilde{\mathcal{I}}^{-1}(\theta)[\nabla\psi_T(\theta), \nabla^{(2)}\psi_T(\theta), \ldots, \nabla^{(k)}\psi_T(\theta)]^T$$

BHLB reduces to the CRLB for $k = 1$.

Corollary 3 (BHLB for unbiased estimators) *If T is an unbiased estimator for θ, then $Cov_\theta(T) \ge (\tilde{\mathcal{I}}^{-1}(\theta))_{1..d,1..d}$.*

C.3 Barankin Lower Bound

BLB chooses the following score function:

$$S(\mathbf{x},\theta) := \int_\Theta \frac{f(\mathbf{x};\tilde{\theta})}{f(\mathbf{x};\theta)} \mathbf{f}(\tilde{\theta};\theta)\, d\tilde{\theta}$$

where $\mathbf{f}(\tilde{\theta};\theta) = \mathbf{f}_{(1)}(\tilde{\theta};\theta) - \mathbf{f}_{(2)}(\tilde{\theta};\theta)$ and $\mathbf{f}_{(1)} \in \mathbb{R}^d$ and $\mathbf{f}_{(2)} \in \mathbb{R}^d$ are two arbitrary vector functions of density on $\tilde{\theta}$, i.e., $\int_\Theta \mathbf{f}_{(i)}(\tilde{\theta};\theta)\, d\tilde{\theta} = \mathbf{1}_{d\times 1}$, and $\Theta \subseteq \mathbb{R}^d$ is the parameter space.

Thus, we have $\mathbb{E}_\theta[S] = 0$. We can verify that

$$\begin{aligned}
\mathbb{E}_\theta[TS^T]_{i,j} &= \int_\Theta \mathbb{E}_{\tilde{\theta}}[T_i] f_j(\tilde{\theta};\theta)\, d\tilde{\theta} \\
\mathbb{E}_\theta[SS^T]_{i,j} &= \int_{\Theta\times\Theta} \mathbb{E}_\theta \left[\frac{f(\mathbf{X};\tilde{\theta}_1) f(\mathbf{X};\tilde{\theta}_2)}{f^2(\mathbf{X};\theta)} \right] f_i(\tilde{\theta}_1;\theta) f_j(\tilde{\theta}_2;\theta)\, d\tilde{\theta}_1 d\tilde{\theta}_2
\end{aligned}$$

Theorem 9 (BLB) *We have:*

$$Cov_\theta(T) \geq \sup_{\mathbf{f}_{(i)}} \mathbb{E}_\theta[TS^T] \mathbb{E}_\theta^{-1}[SS^T] \mathbb{E}_\theta[ST^T]$$

If T is an unbiased estimator of a scalar θ, then:

$$Var_\theta(T) \geq \sup_{\mathbf{f}_{(i)}} \frac{\left(\mathbb{E}_{(1)}[\tilde{\theta}] - \mathbb{E}_{(2)}[\tilde{\theta}] \right)^2}{\int_{\mathbb{R}^n} \frac{\left(\int_\Theta f(\mathbf{x};\tilde{\theta}) [f_{(1)}(\tilde{\theta};\theta) - f_{(2)}(\tilde{\theta};\theta)]\, d\tilde{\theta} \right)^2}{f(\mathbf{x};\theta)}\, d\mathbf{x}} \tag{C.5}$$

Note that since $\mathbf{f}_{(1)}$ and $\mathbf{f}_{(2)}$ are arbitrary, sup operator is used in BLB. We can also notice that BLB does not require any regularity conditions. Barankin [175] has shown that (C.5) is the greatest lower bound.

Appendix D

Capacity of Real Gaussian Channels

We briefly summarize the information-theoretic results on the capacity of parallel Gaussian channels. More details can be found in [147].

Definition 1 (Differential entropy) *The differential entropy of a random vector* $\mathbf{X} \in \mathbb{R}^n$ *is defined as*

$$h(\mathbf{X}) := -\mathbb{E}[\log_2(f(\mathbf{X}))] = -\int f(\mathbf{x}) \log_2(f(\mathbf{x}))\, d\mathbf{x} \tag{D.1}$$

where $f(\mathbf{x})$ *is the density of* $\mathbf{X}$ *if it exists.*

The unit of entropy is (information) bit. For a complex random vector $\mathbf{X} = \mathbf{X}_c + j\mathbf{X}_s \in \mathbb{C}^n$, its differential entropy is defined as $h(\mathbf{X}) := h(\mathbf{X}_c, \mathbf{X}_s)$. If $\mathbf{X}$ is independent of $\mathbf{Y}$, then $h(\mathbf{X}, \mathbf{Y}) = h(\mathbf{X}) + h(\mathbf{Y})$.

Definition 2 (Conditional differential entropy) *The conditional differential entropy of* $\mathbf{X} \in \mathbb{R}^n$ *given* $\mathbf{Y} \in \mathbb{R}^m$ *is defined as*

$$h(\mathbf{X}|\mathbf{Y}) := -\mathbb{E}[\log_2(f(\mathbf{X}|\mathbf{Y}))] = -\int f(\mathbf{x}, \mathbf{y}) \log_2(f(\mathbf{x}|\mathbf{y}))\, d\mathbf{x} d\mathbf{y} \tag{D.2}$$

where $f(\mathbf{x}|\mathbf{y})$ *is the conditional density of* $\mathbf{X}$ *given* $\mathbf{Y}$ *if it exists.*

Note that $h(\mathbf{X}|\mathbf{Y}) = h(\mathbf{X}, \mathbf{Y}) - h(\mathbf{Y})$.

Definition 3 (Mutual information) *The mutual information of* $\mathbf{X}$ *and* $\mathbf{Y}$ *is defined as* $I(\mathbf{X}; \mathbf{Y}) := h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) = h(\mathbf{X}) + h(\mathbf{Y}) - h(\mathbf{X}, \mathbf{Y})$.

It can be verified that $I(\mathbf{X};\mathbf{Y}) = D(f(\mathbf{x},\mathbf{y})\|f(\mathbf{x})f(\mathbf{y}))$, where $D(\cdot||\cdot)$ denotes the Kullback-Leibler distance.

A (M,n) code consists an index set $\mathcal{M} := \{1,2,\ldots,M\}$, an encoding function f and a decoding function g. $\mathcal{M}$ is used to index M distinct messages. The encoding function f generates for each index a code of length n. This code is transmitted in a channel and undergoes some distortion due to the imperfectness of the channel such as noise. The decoding function g associate an element in $\mathcal{M}$ to each received vector of length n. The sending index being i, the probability that $g \neq i$ is denoted as e_i. We define the maximal probability of error $e^{(n)}$ for a (M,n) code to be $\max_{i\in\mathcal{M}} e_i$. The rate of a (M,n) code is defined to be $R := (\log_2 M)/n$ bits per transmission. A rate R is said achievable if there exists a sequence of $(\lceil 2^{nR}\rceil, n)$ codes such that $e^{(n)} \to 0$ as $n \to \infty$. The capacity of a channel C is the supremum of all achievable rates.

Proposition 10 $\mathbf{X} \in \mathbb{C}^m$, $\mathbf{Y},\mathbf{Z} \in \mathbb{C}^n$, *and* $g : \mathbb{C}^m \mapsto \mathbb{C}^n$. *Write* $g(\mathbf{X}) := g_c(\mathbf{X}_c,\mathbf{X}_s) + jg_s(\mathbf{X}_c,\mathbf{X}_s)$ *and suppose that* g_c, g_s *are derivable. Then, for* $\mathbf{Z} := g(\mathbf{X}) + \mathbf{Y}$ *we have* $h(\mathbf{Z}|\mathbf{X}) = h(\mathbf{Y}|\mathbf{X})$.

proof: By the assumptions we can show that $f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_c,\mathbf{x}_s,\mathbf{z}_c,\mathbf{z}_s) = f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}_c,\mathbf{x}_s,\mathbf{z}_c - g_c(\mathbf{x}_c,\mathbf{x}_s),\mathbf{z}_s - g_s(\mathbf{x}_c,\mathbf{x}_s))$, from which we derive that $f_{\mathbf{Z}|\mathbf{X}}(\mathbf{z}_c,\mathbf{z}_s) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{z}_c - g_c(\mathbf{X}_c,\mathbf{X}_s),\mathbf{z}_s - g_s(\mathbf{X}_c,\mathbf{X}_s))$. Thus,

$$\begin{aligned} h(\mathbf{Z}|\mathbf{X}) &:= -\mathbb{E}[\log_2(f_{\mathbf{Z}|\mathbf{X}}(\mathbf{Z}_c,\mathbf{Z}_s))] \\ &= -\mathbb{E}[\log_2(f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Z}_c - g_c(\mathbf{X}_c,\mathbf{X}_s),\mathbf{Z}_s - g_s(\mathbf{X}_c,\mathbf{X}_s)))] \\ &= -\mathbb{E}[\log_2(f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_c,\mathbf{Y}_s))] =: h(\mathbf{Y}|\mathbf{X}) \end{aligned}$$

□

It can be verified that for a normal vector $\mathbf{X} \sim \mathcal{N}(\mu, K)$ where $\mu \in \mathbb{R}^n$ and K the covariance matrix, the entropy is given by $h(\mathbf{X}) = \frac{1}{2}\log_2((2\pi e)^n|K|)$.

Theorem 10 (Entropy bound - real case) *Let* $\mathbf{X} \in \mathbb{R}^n$ *be a continuous random vector of a zero mean and a covariance* K. *Then,*

$$h(\mathbf{X}) \leq \frac{1}{2}\log_2[(2\pi e)^n|K|]$$

with equality iff $\mathbf{X} \sim \mathcal{N}(0,K)$.

proof: Let g be any density satisfying $\int x_i x_j g(\mathbf{x})\, d\mathbf{x} = K_{ij}$ for all i, j. Let $f(\mathbf{x})$ be the normal density. Then, by noting $\log_2(f(\mathbf{x}))$ to be a quadratique form, we have that:

$$0 \leq D(g||f) = \int g(\mathbf{x}) \log_2(g(\mathbf{x})/f(\mathbf{x}))\, d\mathbf{x} = -h(g) - \int g(\mathbf{x}) \log_2(f(\mathbf{x}))\, d\mathbf{x}$$
$$= -h(g) - \int f(\mathbf{x}) \log_2(f(\mathbf{x}))\, d\mathbf{x} = -h(g) + h(f)$$

□

Theorem 11 (Capacity of parallel real Gaussian channels) *Suppose that we have k independent real Gaussian channels, $Y[i] = X[i] + W[i]$ for $i = 1, 2, \ldots, k$ where $W[i] \sim \mathcal{N}(0, \sigma_i^2)$. Under the total energy constraint $\mathbb{E}[\sum_{i=1}^k X[i]^2] \leq P$, the total capacity is given by*

$$C = \frac{1}{2} \sum_{i=1}^{k} \log_2 \left(1 + \frac{\epsilon[i]}{\sigma_i^2}\right) \tag{D.3}$$

where $\epsilon[i] = (\beta - \sigma_i^2)_+$, and β is determined by $\sum_{i=1}^k \epsilon[i] = P$.

proof: It can be shown that C is given by the supremum of the mutual information under the energy constraint [147]

$$C := \sup_{\substack{f_{\mathbf{X}} \\ \mathbb{E}[\sum_{i=1}^k X[i]^2] \leq P}} I((X[i])_i; (Y[i])_i) = \sup_{\substack{\epsilon \geq 0, f_{\mathbf{X}} \\ \sum_{i=1}^k \epsilon[i] \leq P \\ \mathbb{E}[X[i]^2] \leq \epsilon[i]}} I((X[i])_i; (Y[i])_i) \tag{D.4}$$

where by Proposition 10,

$$I((X[i])_i; (Y[i])_i) = h((Y[i])_i) - h((W[i])_i) \leq \sum_{i=1}^{k} h(Y[i]) - h(W[i])$$

with equality iff $(Y[i])_i$ are independent. By $Var[Y_i] = \mathbb{E}[X_i^2] - \mu_i^2 + \sigma_i^2$, by Theorem 10, and using the constraint in (D.4), we have that $h(Y_i) \leq \frac{1}{2} \log_2[2\pi e|\epsilon[i] + \sigma[i]^2|]$ with equality iff $X[i]$ is centered, $\mathbb{E}[X[i]^2] = \epsilon[i]$, and that $Y[i]$ is Gaussian (i.e. $X[i]$ is Gaussian). Now,

$$(D.4) = \sup_{\substack{\epsilon \geq 0 \\ \sum_{i=1}^k \epsilon[i] \leq P}} \frac{1}{2} \sum_{i=1}^{k} \log_2 \left(1 + \frac{\epsilon[i]}{\sigma_i^2}\right) \tag{D.5}$$

Then, by inverting the sign of the objective function, we can verify that (D.5) becomes a convex minimization problem. It can be solved by Lagrange multiplier and we can verify that $\epsilon[i] = (\beta - \sigma_i^2)_+$, and $\sum_{i=1}^k \epsilon[i] = P$ satisfy the KKT condition. □

Publications and Softwares

Journal Publications

- B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," *Applied Optics*, vol. 46, No. 10, pp. 1819-1829, 2007.

- B. Zhang, J. M. Fadili, J.-L. Starck, and S. W. Digel, "Fast Poisson Noise Removal by Biorthogonal Haar Domain Hypothesis Testing," *Statistical Methodology*, Elsevier, 2007, submitted.

- B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, Ridgelets and Curvelets for Poisson Noise Removal," *IEEE Transactions on Image Processing*, 2007, submitted.

Conference Publications

- B. Zhang, J. M. Fadili, J.-L. Starck, and J.-C. Olivo-Marin, "Multiscale variance-stabilizing transform for mixed-Poisson-Gaussian processes and its applications in bioimaging," *IEEE International Conference on Image Processing*, 2007.

- P. Pankajakshan, B. Zhang, L. Blanc-Féraud, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, "Parametric blind deconvolution for confocal laser scanning microscopy," *IEEE Engineering in Medicine and Biology Society International Conference*, 2007.

- B. Zhang, J. M. Fadili, and J.-L. Starck, "Multi-scale variance stabilizing transform for multi-dimensional Poisson count image denoising," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 81-84, 2006.

- B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "A study of Gaussian approximations of fluorescence microscopy PSF models," *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XIII, Proceedings of SPIE*, vol. 6090, pp. 104-114, 2006.

- B. Zhang, J. Enninga, J.-C. Olivo-Marin, and C. Zimmer, "Automated super-resolution detection of fluorescent rods in 2D," *IEEE International Symposium on Biomedical Imaging*, pp. 1296-1299, 2006.

Developed Softwares

- **PSFMODEL** It allows to calculate the 2D/3D paraxial/non-paraxial PSFs of WFFM, LSCM and DSCM using Debye's scalar diffraction integrals. It computes also the Gaussian approximation parameters for these PSFs. The software is implemented in Java. Reference: Chapters 2 and 3.
- **MWIR** The software allows to estimate the underlying Poisson intensity by hypothesis tests in the Haar or Bi-Haar domain. It is implemented in C++. Reference: Section 5.2.
- **MSVST** The software allows to estimate the underlying Poisson intensity by the MS-VST method. IUWT+MS-VST and UWT+MS-VST are implemented in C++, and Ridgelet+MS-VST is written in Matlab. Reference: Section 5.3.
- **MSVSTMPG** It allows to restore a MPG image using IUWT+MS-VST. It is coded in Matlab. Reference: Section 6.2.
- **SPOTDETECT** It allows to detect bright spots in a MPG image by using IUWT+MS-VST with scale correlation. It is implemented in Matlab. Reference: Section 6.3.
- **SUPERSPOT** It allows to detect spots with super-resolution in a Gaussian or MPG image. Least-square fitting is available for Gaussian data. For MPG data, VST-based least-square fitting and the generalized least-square fitting are both implemented. Model selections using hypothesis tests, information criteria, and MDL criteria are available. It is coded in Matlab. Reference: Chapter 7.
- **SUPERROD** It allows to detect rods with super-resolution in a Gaussian or MPG image. The same fitting and model-selection options as for SUPERSPOT are available. It is coded in Matlab. A preliminary version is written in Java. Reference: Chapter 8.

References

[1] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge University Press, 2002.

[2] J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. McGraw-Hill, 1996.

[3] T. P. Kurzweg, J. A. Martinez, S. P. Levitan, and M. T. Shomsky, "Modeling optical MEM systems," *J. Model. Simul. Microsyst.*, vol. 2, no. 1, pp. 21–34, 2001.

[4] M. Gu, *Advanced Optical Imaging Theory*. Springer, 2000.

[5] L. Tao and C. Nicholson, "The three-dimensional point spread functions of a microscope objective in image and object space," *J. Microscopy*, vol. 178, no. 3, pp. 267–271, 1995.

[6] R. Juškaitis, "Measuring the Real PSF of High Numerical Aperture Microscope Objective Lenses," in *Handbook of Biological Confocal Microscopy*, 3rd ed., J. B. Pawley, Ed. Springer, 2006, ch. 11.

[7] S. F. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," *J. Opt. Soc. Am. A*, vol. 8, no. 10, pp. 1601–1613, 1991.

[8] H. T. M. van der Voort and K. C. Strasters, "Restoration of confocal images for quantitative image analysis," *J. Microscopy*, vol. 178, no. 2, pp. 165–181, 1995.

[9] C. J. R. Sheppard, "Effects of aberrating layers and tube length on confocal imaging properties," *Optik*, vol. 87, no. 1, pp. 34–38, 1991.

[10] B. Richards and E. Wolf, "Electromagnetic diffraction in optical systems. pt. ii – Structure of the image field in an aplanatic system," in *Proc. Roy. Soc. (A)*, vol. 253, 1959, pp. 358–379.

[11] H. T. M. van der Voort and G. J. Brakenhoff, "3-D image formation in high-aperture fluorescence confocal microscopy: a numerical analysis," *J. Microscopy*, vol. 158, no. 1, pp. 43–54, 1990.

[12] P. Török, S. J. Hewlett, and P. Varga, "On the series expansion of high-aperture, vectorial diffraction integrals," *J. Mod. Opt.*, vol. 44, no. 3, pp. 493–503, 1997.

[13] P. Török and P. Varga, "Electromagnetic diffraction of light focused through a stratified medium," *Appl. Opt.*, vol. 36, pp. 2305–2312, 1997.

[14] P. Török, "Propagation of electromagnetic dipole waves through dielectric interfaces," *Opt. Lett.*, vol. 25, pp. 1463–1465, 2000.

[15] O. Haeberle, M. Ammar, H. Furukawa, K. Tenjimbayashi, and P. Török, "Point spread function of optical microscopes imaging through stratified media," *Optics Express*, vol. 11, no. 22, pp. 2964–2969, 2003.

[16] A. S. van de Nes, L. Billy, S. F. Pereira, and J. J. M. Braat, "Calculation of the vectorial field distribution in a stratified focal region of a high numerical aperture imaging system," *Optics Express*, vol. 12, no. 7, pp. 1281–1293, 2004.

[17] S. W. Hell, G. Reiner, C. Cremer, and E. H. K. Stelzer, "Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index," *J. Microsc.*, vol. 169, pp. 391–405, 1993.

[18] A. Egner and S. W. Hell, "Equivalence of the Huygens-Fresnel and Debye approach for the calculation of high aperture point-spread functions in the presence of refractive index mismatch," *J. Microsc.*, vol. 193, no. 3, pp. 244–249, 1999.

[19] S. Hell, E. Lehtonen, and E. Stelzer, *Confocal fluorescence microscopy: Wave optics considerations and applications to cell biology.* Verlag Chemie, 1992, pp. 145–160.

[20] S. Kimura and C. Munakata, "Calculation of three-dimensional optical transfer function for a confocal scanning fluorescent microscope," *J. Opt. Soc. Am. A*, vol. 6, no. 7, pp. 1015–1019, 1989.

[21] ——, "Dependence of 3-D optical transfer functions on the pinhole radius in a fluorescent confocal optical microscope," *Appl. Opt.*, vol. 29, no. 20, pp. 3007–3011, 1990.

[22] S. F. Gibson, "Modeling the 3-D imaging properties of the fluorescence light microscope," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, Pa., 1990.

[23] C. J. R. Sheppard and C. J. Cogswell, *Confocal Microscopy.* Academic Press, 1990, ch. Three-dimensional imaging in confocal microscopy, pp. 143–169.

[24] M. Gu and C. J. R. Sheppard, "Confocal fluorescent microscopy with a finite-sized circular detector," *J. Opt. Soc. Am. A*, vol. 9, no. 1, pp. 151–153, 1992.

[25] M. Martínez-Corral, "Point spread function engineering in confocal scanning microscopy," in *Wave-Optical Systems Engineering II.* Proc. SPIE, 2003, pp. 112–122.

[26] J.-A. Conchello and J. W. Lichtman, "Theoretical analysis of a rotating-disk partially confocal scanning microscope," *Appl. Opt.*, vol. 33, no. 4, pp. 585–596, 1994.

[27] D. R. Sandison, R. M. Williams, K. S. Wells, J. Strickler, and W. W. Webb, "Quantitative Fluorescence Confocal Laser Scanning Microscopy (CLSM)," in *Handbook of Biological Confocal Microscopy*, 2nd ed., J. B. Pawley, Ed. Plenum Press, 1995, ch. 3.

[28] E. Gratton and M. J. vandeVen, "Laser Sources for Confocal Microscopy," in *Handbook of Biological Confocal Microscopy*, 2nd ed., J. B. Pawley, Ed. Plenum Press, 1995, ch. 5.

[29] M. Petráň, M. Hadravský, J. Benes, R. Kucera, and A. Boyde, "The tandem scanning reflected light microscope. Part 1 – The principle and its design," in *Proc. Roy. Micro. Soc.*, vol. 20. Blackwell Scientific Publications Ltd, 1985, pp. 125–129.

[30] G. Q. Xiao and G. S. Kino, "A real-time confocal scanning optical microscope," in *Scanning Imaging Technology, Proc. SPIE*, vol. 809, 1987, pp. 107–113.

[31] G. Q. Xiao, T. R. Corle, and G. S. Kino, "Real-time confocal scanning optical microscope," *Appl. Phys. Lett.*, vol. 53, pp. 716–718, 1988.

[32] T. Tanaami, S. Otsuki, N. Tomosada, Y. Kosugi, M. Shimizu, and H. Ishida, "High-speed 1-frame/ms scanning confocal microscope with a microlens and Nipkow disks," *Applied Optics*, vol. 41, no. 22, pp. 4704–4708, 2002.

[33] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii, *Electrodynamics of continuous media.* Oxford: Pergamon Press, 1984.

[34] C. M. Anderson, G. N. Georgiou, I. E. G. Morrison, G. V. W. Stevenson, and R. J. Cherry, "Tracking of cell surface receptors by fluorescence digital imaging microscopy using a charge-coupled device camera," *J. Cell Sci.*, vol. 101, pp. 415–425, 1992.

[35] G. J. Schütz, H. Schindler, and T. Schmidt, "Single-molecule microscopy on model membranes reveals anomalous diffusion," *Biophys. J.*, vol. 73, pp. 1073–1080, 1997.

[36] M. K. Cheezum, W. F. Walker, and W. H. Guilford, "Quantitative Comparison of Algorithms for Tracking Single Fluorescent Particles," *Biophys. J.*, vol. 81, no. 4, pp. 2378–2388, 2001.

[37] A. Santos and I. T. Young, "Model-based resolution: applying the theory in quantitative microscopy," *Appl. Opt.*, vol. 39, no. 17, pp. 2948–2958, 2000.

[38] D. Thomann, D. R. Rines, P. K. Sorger, and G. Danuser, "Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement," *J. Microscopy*, vol. 208, no. 1, pp. 49–64, 2002.

[39] F. Rooms, W. Philips, and D. S. Lidke, "Simultaneous degradation estimation and restoration of confocal images and performance evaluation by colocalization analysis," *J. Microscopy*, vol. 218, no. 1, pp. 22–36, 2005.

[40] J. C. G. Blonk, A. Don, H. van Aalst, and J. J. Birmingham, "Fluorescence photobleaching recovery in the confocal scanning light microscope," *J. Microscopy*, vol. 169, no. 3, pp. 363–374, 1993.

[41] K. Braeckmans, L. Peeters, N. N. Sanders, S. C. D. Smedt, and J. Demeester, "Three-Dimensional Fluorescence Recovery after Photobleaching with the Confocal Scanning Laser Microscope," *Biophys. J.*, vol. 85, pp. 2240–2252, 2003.

[42] G. J. Streekstra and J. van Pelt, "Analysis of tubular structures in three-dimensional confocal images," *Network: Comput. Neural Syst.*, vol. 13, pp. 381–395, 2002.

[43] J.-A. Conchello, "Superresolution and convergence properties of the expectation-maximization algorithm for maximum-likelihood deconvolution of incoherent images," *J. Opt. Soc. Am. A*, vol. 15, no. 10, pp. 2609–2619, 1998.

[44] L. J. van Vliet, "Grey-Scale Measurements in Multi-Dimensional Digitized Images," Ph.D. dissertation, Delft University, The Netherlands, 1993.

[45] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* Wiley, 1949.

[46] J. W. Wood and V. K. Ingle, "Kalman filtering in two dimensions," *IEEE Trans. Inform. Theory*, 1977.

[47] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. of 6th ICCV*, 1998, pp. 839–846.

[48] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, p. 490530, 2005.

[49] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Processing*, vol. 15, no. 10, pp. 2866–2878, 2006.

[50] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[51] D. L. Donoho and R. R. Coifman, "Translation-Invariant De-Noising," in *Wavelets and Statistics*, ser. Lecture Notes in Statistics, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, vol. 103.

[52] E. Candès and D. Donoho, "Ridgelets: the key to high dimensional intermittency?" *Philosophical Transactions of the Royal Society of London A*, vol. 357, pp. 2495–2509, 1999.

[53] J.-L. Starck, E. Candès, and D. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 131–141, 2002.

[54] J. L. Starck, "Very high quality image restoration by combining wavelets and curvelets," *Proc. SPIE*, vol. 4478, pp. 9–19, 2001.

[55] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[56] P. Perona and J. Malik, "Scale-space and edge detection using anistropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

[57] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*. Springer, 2001.

[58] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, pp. 721–741, 1984.

[59] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. of ICIP*, vol. 1, 1996, pp. 379–382.

[60] H. Chipman, E. Kolaczyk, and R. McCulloch, "Adaptive bayesian wavelet shrinkage," *J. Am. Statist. Ass.*, vol. 92, pp. 1413–1421, 1997.

[61] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy. Statist. Soc. ser. B*, vol. 60, no. 4, pp. 725–749, 1998.

[62] I. Daubechies, *Ten Lectures on Wavelets*. SIAM, 1992.

[63] Y. Meyer, *Wavelets and Operators*. Cambridge University Press, 1992.

[64] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1998.

[65] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform," in *Wavelets: Time-Frequency Methods and Phase-Space*. Springer-Verlag, 1989, pp. 286–297.

[66] M. J. Shensa, "Discrete wavelet transforms: Wedding the à trous and Mallat algorithms," *IEEE Trans. Signal Processing*, vol. 40, pp. 2464–2482, 1992.

[67] R. A. DeVore, "Nonlinear approximation," *Acta Numerica*, pp. 51–150, 1998.

[68] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. American Statist. Assoc.*, vol. 90, pp. 1200–1224, 1995.

[69] D. L. Donoho, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.

[70] F. J. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, vol. 35, pp. 246–254, 1948.

[71] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data," *Proc. Symp. Applied Mathematics: Different Perspectives on Wavelets*, vol. 47, pp. 173–205, 1993.

[72] P. Fryźlewicz and G. P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *J. Comp. Graph. Stat.*, vol. 13, pp. 621–638, 2004.

[73] M. Fisz, "The limiting distribution of a function of two independent random variables and its statistical application," *Colloquium Mathematicum*, vol. 3, pp. 138–146, 1955.

[74] M. Jansen, "Multiscale Poisson data smoothing," *J. Roy. Statist. Soc. ser. B*, vol. 68, no. 1, pp. 27–48, 2006.

[75] R. D. Nowak and R. G. Baraniuk, "Wavelet-Domain Filtering for Photon Imaging Systems," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 666–678, 1999.

[76] A. Antoniadis and T. Sapatinas, "Wavelet shrinkage for natural exponential families with quadratic variance functions," *Biometrika*, vol. 88, pp. 805–820, 2001.

[77] A. Antoniadis, J. Bigot, and T. Sapatinas, "Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study," *Journal of Statistical Software*, vol. 6, no. 6, 2001.

[78] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. Roy. Statist. Soc. ser. B*, vol. 57, no. 1, pp. 289–300, 1995.

[79] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[80] F. Abramovich, Y. Benjamini, D. Donoho, and I. Johnstone, "Adapting to Unknown Sparsity by controlling the False Discovery Rate," *Annals of Statistics*, vol. 34, no. 2, pp. 584–653, 2006.

[81] D. L. Donoho and J. Jin, "Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data," *Annals of Statistics*, vol. 34, no. 6, pp. 2980–3018, 2006.

[82] A. Bijaoui and G. Jammal, "On the distribution of the wavelet coefficient for a Poisson noise," *Signal Processing*, vol. 81, pp. 1789–1800, 2001.

[83] J. G. Skellman, "The frequency distribution of the difference between two Poisson variates belonging to different populations," *J. Roy. Statist. Soc. ser. A*, vol. 109, p. 296, 1946.

[84] N. L. Johnson, "On an extension of the connexion between Poisson and χ^2-distributions," *Biometrika*, vol. 46, pp. 352–363, 1959.

[85] E. D. Kolaczyk, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statist. Sinica*, vol. 9, pp. 119–135, 1999.

[86] ——, "Bayesian multiscale models for Poisson processes," *J. Amer. Statist. Ass.*, vol. 94, no. 447, pp. 920–933, 1999.

[87] K. E. Timmermann and R. D. Nowak, "Multiscale Modeling and Estimation of Poisson Processes with Application to Photon-Limited Imaging," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 846–862, 1999.

[88] R. D. Nowak and E. D. Kolaczyk, "A statistical multiscale framework for Poisson inverse problems," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1811–1825, 2000.

[89] P. Besbeas, I. D. Feis, and T. Sapatinas, "A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts," *Internat. Statist. Rev.*, vol. 72, no. 2, pp. 209–237, 2004.

[90] S. Sardy, A. Antoniadis, and P. Tseng, "Automatic smoothing with wavelets for a wide class of distributions," *J. Comput. Graph. Stat.*, vol. 13, no. 2, pp. 399–421, 2004.

[91] R. M. Willett and R. D. Nowak, "Fast, Near-Optimal, Multiresolution Estimation of Poisson Signals and Images," in *EUSIPCO*, San Diego, CA, 2004.

[92] ——, "Multiscale Poisson Intensity and Density Estimation," *IEEE Trans. Inform. Theory*, 2005, submitted.

[93] R. Willett, "Multiscale Analysis of Photon-Limited Astronomical Images," in *SCMA IV*, 2006.

[94] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Ann. Statist.*, vol. 27, p. 859897, 1999.

[95] R. M. Willett and R. D. Nowak, "Multiscale likelihood analysis and image reconstruction," in *SPIE Wavelets X*, vol. 5207, San Diego, CA, Aug. 2003, pp. 97–111.

[96] E. J. Candès and D. L. Donoho, "Curvelets – A Surprisingly Effective Nonadaptive Representation For Objects with Edges," in *Curve and Surface Fitting: Saint-Malo 1999*, A. Cohen, C. Rabut, and L. Schumaker, Eds. Nashville, TN: Vanderbilt University Press, 1999.

[97] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis*. Cambridge University Press, 1998.

[98] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1053–1060, 1995.

[99] O. Rioul, "Simple regularity criteria for subdivision schemes," *SIAM Journal on Mathematical Analysis*, vol. 23, no. 6, pp. 1544–1576, 1992.

[100] E. D. Kolaczyk, "Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics," *The Astrophysical Journal*, vol. 534, pp. 490–505, 2000.

[101] ——, "Nonparametric Estimation of Gamma-Ray Burst Intensities Using Haar Wavelets," *The Astrophysical Journal*, vol. 483, pp. 340–349, 1997.

[102] C. Charles and J. P. Rasson, "Wavelet denoising of Poisson-distributed data and applications," *Computational Statistics and Data Analysis*, vol. 43, no. 2, pp. 139–148, 2003.

[103] P. B. Patnaik, "The non-central χ^2- and F-distributions and their applications," *Biometrika*, vol. 36, pp. 202–232, 1949.

[104] R. A. Fisher, *Contributions to mathematical statistics*. New York: Wiley, 1950.

[105] J. E. Hacke, "Solving the quartic," *Amer. Math. Monthly*, vol. 48, pp. 327–328, 1941.

[106] J.-L. Starck, M. Fadili, and F. Murtagh, "The Undecimated Wavelet Decomposition and its Reconstruction," *IEEE Transactions on Image Processing*, 2006, accepted.

[107] J. C. Olivo-Marin, "Extraction of spots in biological images using multiscale products," *Pattern Recognition*, vol. 35, no. 9, pp. 1989–1996, 2002.

[108] J. L. Starck, A. Bijaoui, and F. Murtagh, "Multiresolution Support Applied to Image Filtering and Deconvolution," *CVIP: Graphical Models and Image Processing*, vol. 57, no. 5, pp. 420–431, Sept. 1995.

[109] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *PNAS*, vol. 100, no. 5, pp. 2197–2202, 2003.

[110] D. L. Donoho, "For Most Large Underdetermined Systems of Linear Equations, the minimal ℓ^1-norm solution is also the sparsest solution," Department of Statistics of Stanford Univ., Tech. Rep., 2004.

[111] I. Yamada, "The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings," in *Inherently Parallel Algorithm in Feasibility and Optimization and their Applications*. Elsevier, 2001, pp. 473–504.

[112] R. M. Willett and R. D. Nowak, "Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 332–350, 2003.

[113] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," *Advances in Imaging and Electron Physics*, vol. 132, 2004.

[114] A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli, and J. Waldén, "Fast Slant Stack: A notion of Radon transform for data in a cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible," *SIAM J. Sci. Comput.*, 2001, to appear.

[115] D. Donoho and A. Flesia, "Digital Ridgelet Transform Based on True Ridge Functions," in *Beyond Wavelets*, J. Schmeidler and G. Welland, Eds. Academic Press, 2002.

[116] D. L. Donoho and M. R. Duncan, "Digital curvelet transform: strategy, implementation and experiments," in *Proc. Aerosense 2000, Wavelet Applications VII*, H. H. Szu, M. Vetterli, W. Campbell, and J. R. Buss, Eds., vol. 4056. SPIE, 2000, pp. 12–29.

[117] J.-L. Starck, E. Candes, and D. Donoho, "Astronomical image representation by the curvelet tansform," *Astronomy and Astrophysics*, vol. 398, pp. 785–800, 2003.

[118] J. B. de Monvel, S. L. Calvez, and M. Ulfendahl, "Image restoration for confocal microscopy: Improving the limits of deconvolution, with application to the visualization of the mammalian hearing organ," *Biophys. J.*, vol. 80, pp. 2455–2470, 2001.

[119] C. Kervrann and A. Trubuil, "An adaptive window approach for poisson noise reduction and structure preserving in confocal microscopy," in *International Symposium on Biomedical Imaging (ISBI'04)*, vol. 1, 2004, pp. 788–791.

[120] J. Boulanger, C. Kervrann, and P. Bouthemy, "An adaptive statistical method for denoising 4D fluorescence image sequences with preservation of spatio-temporal discontinuities," in *International Conference on Image Processing (ICIP'05)*, vol. 2, 2005, pp. 145–148.

[121] G. P. Bernad, L. Blanc-Feraud, and J. Zerubia, "A restoration method for confocal microscopy using complex wavelet transform," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, vol. 2, 2005, pp. 621–624.

[122] J. Boulanger, "Estimation non-paramétrique et contributions à l'analyse de séquences d'images; Modélisation, simulation et estimation du trafic intra-cellulaire dans des séquences de vidéo-microscopie," Univ. Rennes I, Tech. Rep., 2007, ph.D. thesis.

[123] J. Boulanger, C. Kervrann, and P. Bouthemy, "Space-time adaption for patch based image sequence restoration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 1096–1102, 2007.

[124] B. Zhang, J. M. Fadili, J.-L. Starck, and J.-C. Olivo-Marin, "Multiscale variance-stabilizing transform for mixed-poisson-gausian processes and its applications in bioimaging," in *ICIP*, 2007, accepted.

[125] C. Rose and M. D. Smith, *Mathematical Statistics with Mathematica*. Springer-Verlag, 2002, ch. 7.2C: k-Statistics: Unbiased Estimators of Cumulants, pp. 256–259.

[126] P. M. Heffernan, "Unbiased estimation of central moments by using U-statistics," *J. Roy. Statist. Soc. B*, vol. 59, no. 4, pp. 861–863, 1997.

[127] B. J. Vermolen, Y. Garini, S. Mai, V. Mougey, T. Fest, T. C.-Y. Chuang, A. Y.-C. Chuang, L. Wark, and I. T. Young, "Characterizing the three-dimensional organization of telomeres," *Cytometry*, vol. 67A, no. 2, pp. 144–150, 2005.

[128] N. Arhel, A. Genovesio, K.-A. Kim, S. Miko, E. Perret, J.-C. Olivo-Marin, S. Shorte, and P. Charneau, "Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes," *Nature Methods*, vol. 3, pp. 817–824, 2006.

[129] E. Meijering, I. Smal, and G. Danuser, "Tracking in Molecular Bioimaging," *IEEE Signal Processing Mag.*, vol. 23, no. 3, pp. 46–53, 2006.

[130] T. D. Pham, D. I. Crane, T. H. Tran, and T. H. Nguyen, "Extraction of fluorescent cell puncta by adaptive fuzzy segmentation," *Bioinformatics*, vol. 20, no. 14, pp. 2189–2196, 2004.

[131] R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescence probes," *Biophys. J.*, vol. 82, pp. 2775–2783, 2002.

[132] J. L. Harris, "Resolving power and decision theory," *J. Opt. Soc. Am.*, vol. 54, pp. 606–611, 1964.

[133] C. W. Helstrom, "Detection and resolution of incoherent objects by a background-limited optical system," *J. Opt. Soc. Am.*, vol. 59, p. 164175, 1969.

[134] A. J. den Dekker, "Model-based optical resolution," *IEEE Trans. Instrum. Meas.*, vol. 46, pp. 798–802, 1997.

[135] M. Shahram and P. Milanfar, "Imaging below the diffraction limit: a statistical analysis," *IEEE Trans. Image Processing*, vol. 13, no. 5, pp. 677–689, 2004.

[136] ——, "Statistical and information-theoretic analysis of resolution in imaging," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3411–3437, 2006.

[137] C. W. Helstrom, "Resolvability of objects from the standpoint of statistical parameter estimation," *J. Opt. Soc. Am.*, vol. 60, p. 659666, 1970.

[138] E. Bettens, D. V. Dyck, A. J. den Dekker, J. Sijbers, and A. van den Bos, "Model-based two-object resolution from observations having counting statistics," *Ultramicroscopy*, vol. 77, p. 3748, 1999.

[139] S. T. Smith, "Statistical resolution limits and the complexified Cramér-Rao bound," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1597–1609, 2005.

[140] S. Ram, E. S. Ward, and R. J. Ober, "Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy," *PNAS*, vol. 103, no. 12, pp. 4457–4462, 2006.

[141] ——, "A stochastic analysis of performance limits for optical microscopes," *Multidimensional Systems and Signal Processing*, vol. 17, no. 1, pp. 27–57, 2006.

[142] E. J. Farrell, "Information content of photoelectric star images," *J. Opt. Soc. Am.*, vol. 56, pp. 578–587, 1966.

[143] W. T. Cathey, B. R. Frieden, W. T. Rhodes, and C. K. Rushforth, "Image gathering and processing for enhanced resolution," *J. Opt. Soc. Am. A*, vol. 1, pp. 241–250, 1984.

[144] R. J. Ober, S. Ram, and E. S. Ward, "Localization accuracy in single-molecule microscopy," *Biophys. J.*, vol. 86, pp. 1185–1200, 2004.

[145] S. Ram, E. S. Ward, and R. J. Ober, "How accurately can a single molecule be localized in three dimensions using a fluorescence microscope?" in *Proc. of SPIE*, vol. 5699, 2005.

[146] E. L. Kosarev, "Shannon's superresolution limit for signal recovery," *Inverse Problems*, vol. 6, pp. 55–76, 1990.

[147] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[148] R. N. Ghosh and W. W. Webb., "Automated detection and tracking of individual and clustered cell surface low density lipoprotein receptor molecules," *Biophys. J.*, vol. 66, p. 13011318, 1994.

[149] U. Kubitscheck, O. Kuckmann, T. Kues, and R. Peters, "Imaging and tracking of single GFP molecules in solution," *Biophys. J.*, vol. 78, pp. 2170–2179, 2000.

[150] F. Aguet, D. V. D. Ville, and M. Unser, "A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles," *Optics Express*, vol. 13, no. 26, pp. 10 503–10 522, 2005.

[151] K. A. Lidke, B. Rieger, T. M. Jovin, and R. Heintzmann, "Superresolution by localization of quantum dots using blinking statistics," *Optics Express*, vol. 13, no. 18, pp. 7052–7062, 2005.

[152] A. Mohammad-Djafari, "A Bayesian estimation method for detection, localization and estimation of superposed sources in remote sensing," in *Proc. of SPIE*, vol. 3163, 1997, pp. 73–82.

[153] G. Casella and R. L. Berger, *Statistical Inference*. Duxbury, 2001.

[154] P. Stoica and Y. Selén, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Mag.*, pp. 36–47, Jul. 2004.

[155] E. Lebarbier and T. Mary-Huard, "Le critère BIC: fondements théoriques et interprétation," INRIA, Tech. Rep. 5315, 2004.

[156] P. Stoica, Y. Selén, and J. Li, "On information criteria and the generalized likelihood ratio test of model order selection," *IEEE Signal Processing Lett.*, vol. 11, no. 10, pp. 794–797, 2004.

[157] P. Grünwald, *Advances in Minimum Description Length: Theory and Applications.* MIT Press, 2005.

[158] J. I. Myung, D. J. Navarro, and M. A. Pitt, "Model selection by normalized maximum likelihood," *J. Math. Psych.*, vol. 50, no. 2, pp. 167–179, 2005.

[159] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Stat.*, vol. 11, pp. 416–431, 1983.

[160] M. Jacob, T. Blu, C. Vaillant, J. H. Maddocks, and M. Unser, "3-D shape estimation of DNA molecules from stereo Cryo-electron micro-graphs using a projection-steerable snake," *IEEE Trans. Image Processing*, vol. 15, no. 1, pp. 214–227, 2006.

[161] J. Enninga, J. Mounier, P. Sansonetti, and G. T. V. Nhieu, "Studying the secretion of type III effectors into host cells in real time," *Nature Methods*, vol. 2, no. 12, pp. 959–965, 2005.

[162] K. Rohr and S. Wörz, "High-precision localization and quantification of 3D tubular structures," in *International Symposium on Biomedical Imaging (ISBI)*, 2006, pp. 1160–1163.

[163] C. C. Craig, "On the frequency function of xy," *Ann. Math. Statist.*, vol. 7, pp. 1–15, 1936.

[164] L. A. Aroian, "The probability function of the product of two normally distributed variables," *Ann. Math. Statist.*, vol. 18, pp. 265–271, 1947.

[165] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, 2006.

[166] L. Demanet and L. Ying, "Wave Atoms and sparsity of oscillatory patterns," *Appl. Comput. Harmon. Anal.*, 2006, submitted.

[167] J. Bobin, J.-L. Starck, J. M. Fadili, Y. Moudden, and D. L. Donoho, "Morphological component analysis: An adaptive thresholding strategy," *IEEE Trans. Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.

[168] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products.* Academic Press, 1967.

[169] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "A study of Gaussian approximations of fluorescence microscopy PSF models," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XIII of Proc. SPIE, in press*, vol. 6090, 2006.

[170] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions.* Dover, 1970.

[171] S. Varadhan, *Large Deviations and Applications.* Philadelphia: SIAM, 1984.

[172] H. H. Bauschke and J. M. Borwein, "On Projection Algorithms for Solving Convex Feasibility Problems," *SIAM Review*, vol. 38, no. 3, pp. 367–426, 1996.

[173] V. Piera and N. Javier, "Sample covariance based parameter estimation for digital communications," Univ. Politècnica de Catalunya, Tech. Rep., 2005, chap. 2.

[174] H. L. V. Trees, *Detection, estimation, and modulation theory.* Wiley, 2001.

[175] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, vol. 20, no. 4, pp. 477–501, 1949.