# Learning and optimization for 3D super-resolution in fluorescence microscopy

Vasiliki Stergiopoulou

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

# THÈSE DE DOCTORAT

# Apprentissage et optimisation pour la super-résolution 3D en microscopie de fluorescence


## Vasiliki STERGIOPOULOU

Laboratoire d'Informatique, de Signaux et Systèmes de Sophia Antipolis (I3S)
et Centre Inria d'Université Côte d'Azur


**Présentée en vue de l'obtention du grade de docteur en** Automatique, Traitement du Signal et des Images **d'**Université Côte d'Azur

**Dirigée par :** Laure BLANC-FÉRAUD, DR CNRS, Université Côte d'Azur, France
**Co-encadrée par :** Luca CALATRONI, CR CNRS, Université Côte d'Azur, France
Sébastien SCHAUB, IR CNRS, Sorbonne Université, France

**Soutenue le :** 30/01/23

**Devant le jury, composé de :**
Vicente ZARZOSO, Full Professor, Université Côte d'Azur, France
Arrate MUÑOZ BARRUTIA, Full Professor, Universidad Carlos III de Madrid, Spain
Sébastien BOURGUIGNON, Assistant Professor (HDR), École Centrale de Nantes, France
Daniel SAGE, Research Engineer, École Polytechnique Fédérale de Lausanne, Switzerland
Luca ZANNI, Full Professor, University of Modena and Reggio Emilia, Italy
Luca CALATRONI, CR CNRS, Université Côte d'Azur, France
Sébastien SCHAUB, IR CNRS, Sorbonne Université, France

UNIVERSITÉ CÔTE D'AZUR | INTERDISCIPLINARY INSTITUTE FOR ARTIFICIAL INTELLIGENCE 3IA CÔTE D'AZUR | FRANCE 2030

# APPRENTISSAGE ET OPTIMISATION POUR LA SUPER-RÉSOLUTION 3D EN MICROSCOPIE DE FLUORESCENCE

---

***Learning and optimization for 3D super-resolution in fluorescence microscopy***

**Vasiliki STERGIOPOULOU**

## Jury :

**Président du jury**

Vicente ZARZOSO, Full Professor, Université Côte d'Azur, France

**Rapporteurs**

Arrate MUÑOZ BARRUTIA, Full Professor, Universidad Carlos III de Madrid, Spain
Sébastien BOURGUIGNON, Assistant Professor (HDR), École Centrale de Nantes, France

**Examinateurs**

Daniel SAGE, Research Engineer, École Polytechnique Fédérale de Lausanne, Switzerland
Luca ZANNI, Full Professor, University of Modena and Reggio Emilia, Italy

**Directeur de thèse**

Laure BLANC-FÉRAUD, DR CNRS, Université Côte d'Azur, France

**Co-encadrants de thèse**

Luca CALATRONI, CR CNRS, Université Côte d'Azur, France
Sébastien SCHAUB, IR CNRS, Sorbonne Université, France

---

Université Côte d'Azur

Vasiliki STERGIOPOULOU

***Apprentissage et optimisation pour la super-résolution 3D en microscopie de fluorescence***

xxiii+129 p.

Ce document a été préparé avec LaTeX2e et la classe `these-ISSS` version v. 2.10.

*Impression : output.tex – 27/4/2023 – 14:55*

*Révision pour la classe : these-ISSS.cls,v 2.10 2020/06/24 14:16:37 mpelleau*

# Apprentissage et optimisation pour la super-résolution 3D en microscopie de fluorescence


## Résumé

L'objectif de cette thèse est de développer de nouvelles méthodes de reconstruction algorithmique pour la *super-résolution* en microscopie de fluorescence. La résolution spatiale des images obtenues par les microscopes à fluorescence standards est physiquement limitée à cause de la diffraction de la lumière visible, ce qui signifie qu'il est difficile d'étudier des entités de taille inférieure à la barrière de diffraction qui est d'environ $200$ nm dans le plan latéral et $500$ nm dans la direction axiale. Plusieurs techniques de super-résolution peuvent dépasser ces limites en employant des protocoles complexes. En comparaison, les techniques de super-résolution proposées dans ce manuscrit ne nécessitent pas de fluorophores spéciaux, de prototypes complexes, de fixation de l'échantillon ni de longs temps d'acquisition, ce qui permet l'imagerie de cellules vivantes à l'aide d'un *équipement standard* avec peu de dommages à l'échantillon observé.

Notre première contribution est la méthode COL0RME (*COvariance-based* $\ell_0$ *super-Resolution Microscopy with intensity Estimation*) qui reconstruit dans un cadre variationnel une image super-résolue latéralement étant donné une courte pile temporelle d'images. Dans cette méthode, nous exploitons l'indépendance entre émetteurs distincts ainsi que la distribution parcimonieuse des molécules fluorescentes via l'utilisation d'un terme de régularisation approprié défini sur la matrice de covariance des émetteurs. La fonction de coût à minimiser est non-différentiable et, généralement, non-convexe. Pour sa minimisation numérique, nous utilisons donc des schémas itératifs basés sur le gradient proximal.

Nous présentons également une approche de *super-résolution 3D* pour améliorer la résolution latérale et axiale d'une fine couche adjacente à la lamelle de verre dans les applications d'imagerie par microscopie à réflexion interne totale (Total Internal Reflection Fluorescence (TIRF) en anglais). Nous combinons la méthode COL0RME qui effectue la super-résolution dans le plan latéral, avec une procédure de reconstruction 3D dans la direction axiale en utilisant la microscopie TIRF multi-angle (MA-TIRF).

Nous avons ensuite proposé deux approches hybrides non supervisées qui combinent la modélisation physique des acquisitions temporelles en microscopie de fluorescence avec des *approches génératives* pour le problème de la déconvolution des images. Pour la première approche, nous utilisons des Generative Adversarial Networks (GANs) en combinaison avec une modélisation non-linéaire des fluctuations de fluorescence, et pour la seconde, des Variational Auto-Encoders (VAEs) ont été employés.

Enfin, nous étendons la méthode COL0RME avec l'intérêt d'obtenir une reconstruction plus adaptée aux géométries des échantillons (par exemple, les filaments). À cette fin, nous remplaçons l'opérateur proximal du terme de régularisation employé dans COL0RME, par un *débruiteur d'image* (c'est-à-dire un réseau de neurones pré-entraîné), suivant le cadre de reconstruction Plug-and-Play (PnP).



**Mots-clés :** microscopie de fluorescence, super-résolution, optimisation parcimonieuse, régularisation non convexe, problèmes inverses régularisés, réseaux neuronaux

# Learning and optimization for 3D super-resolution in fluorescence microscopy


## Abstract

The objective of this thesis is to develop new algorithmic reconstruction methods for *super-resolution* in fluorescence microscopy. The spatial resolution of images obtained by standard fluorescence microscopes is physically limited due to the diffraction of visible light, which means that it is difficult to study entities of size smaller than the diffraction barrier that is around $200$ nm in the lateral plane and $500$ nm in the axial direction. Several super-resolution techniques can overcome those limits by employing complex protocols. In comparison to them, the super-resolution techniques that are proposed in this manuscript do not require special fluorophores, complex prototypes, sample fixation nor long acquisition times, allowing live cell imaging by means of *standard equipment* with little damage to the sample under observation.

Our first contribution is the method COL0RME (*COvariance-based $\ell_0$ super-Resolution Microscopy with intensity Estimation*) which reconstructs in a variational framework a laterally super-resolved image given a short temporal stack of frames. In this method, we exploit the independence between distinct emitters as well as the sparse distribution of the fluorescent molecules via the use of an appropriate regularization term defined on the emitters' covariance matrix. The cost function to minimize is non-differentiable and, generally, non-convex. For its numerical minimization, we thus use proximal gradient-based iterative schemes.

We, further, present a *3D super-resolution* approach to improve both lateral and axial resolution of a thin layer adjacent to the coverslip in Total Internal Reflection Fluorescence (TIRF) microscopy imaging applications. To do so, we combine the method COL0RME that performs super-resolution in the lateral plane, with a 3D reconstruction procedure in the axial direction using Multi-Angle TIRF (MA-TIRF).

We then proposed two unsupervised hybrid approaches that combine the physical modeling of fluorescence microscopy time-lapse acquisitions with *generative approaches* for the problem of image deconvolution. For the first approach, we used Generative Adversarial Networks (GANs) in combination with a non-linear modeling of fluorescent fluctuations, while for the second, Variational Auto-Encoders (VAEs) were employed.

Finally, we extend the method COL0RME with the interest of obtaining a reconstruction more adapted to the sample geometries (e.g. filaments). To this end, we replace the proximity operator of the regularization term employed in COL0RME, with an *image denoiser* (i.e. a pretrained network), following the Plug-and-Play (PnP) reconstruction framework.



**Keywords:** fluorescence microscopy, super-resolution, sparse optimization, non-convex regularization, regularized inverse problems, neural networks

# Acknowledgments

I would like to express my deepest gratitude to the members of my PhD defense jury, Prof. Arrate Muñoz Barrutia, Dr. Sébastien Bourguignon, Dr. Daniel Sage, Prof. Vicente Zarzoso, and Prof. Luca Zanni, for their invaluable time, effort, and insightful comments in reviewing my thesis. Their constructive feedback and suggestions have significantly contributed to the improvement of my work.

I am immensely grateful to my thesis director, Dr. Laure Blanc-Féraud, for her persistent support, guidance, and encouragement throughout my doctoral journey. Her door was always open, and her kind smile provided me with a safe refuge during challenging times. I am also indebted to my thesis co-supervisor, Dr. Luca Calatroni, for his expert advice, mathematical help, insightful discussions, but also gastronomic recommendations and amazing dinners we shared together. Thank you, Luca, for your invaluable support and guidance. I would like to extend my sincere appreciation to my second thesis co-supervisor, Dr. Sébastien Schaub, for introducing me to the fascinating world of microscopy and patiently explaining complex concepts.

I am grateful to Dr. Emmanuel Soubies, whose passion for imaging has inspired me, and to Prof. Carola-Bibiane Schönlieb, Dr. Subhadip Mukherjee, and the CIA group for hosting me during my exchange program and making me feel at home. Collaborating with Dr. Henrique Goulart, Mayeul Cachia, and Sai Muttavarapu has been a great pleasure, and I thank them for their professionalism, creativity, and humor.

I also have amazing memories of my friends at the I3S lab, who have made my experience unforgettable. Special thanks to Arne, Bastien, Cedric, Cyprien, Diana, Eva, Gabriele, Marta, Melpo, Mayeul, Rudan, Sai, Somia, Xavier, and many more at Morpeheme and Mediacoding, for their companionship, coffee breaks, gossip, and fun moments. I would like also to thank Andre, Andrew, Amelie, Arnab, Dalia, Fernando, Giulia, Ninad, Remy, and Thomas for making me feel welcome and part of their group, even though we were on different floors of the building. I am also grateful to the administration, and I owe a special debt of gratitude to Nadia, the best team assistant, and friend, for her constant support and kindness.

During my three years in Nice, I met some amazing people who became like family to me. Although I can't name them all here, I want to thank the ones who had a big impact on me. Marta, you were the best company I could have asked for during our time in Cambridge, summer/winter schools, and I'm so grateful to have you as a lifelong friend. Ninad and Diana, thanks for all the crazy moments we shared in and out of the lab. Your friendship made my PhD journey so much more fun. Melpo and Dimitris, thank you for hosting me, sharing your delicious Greek meals, and for the great time we had together. Somia, your support, coffee/fruit/plank breaks, and inspiring chats while walking by the sea meant a lot to me. Eva, we started our doctoral journey almost at the same time, and I appreciate all the support you gave me during the tough times, like making fresh orange juice when I was feeling sick or tired. Hazal, I loved having you in Nice for a long summer, and the trips we took in the French Riviera will always be unforgettable. Lastly, to my music band mates Louis, Stephane, Thibaud, Santiago, Gabriele, and Giulia, thanks for making music with me and creating special memories.

I would also like to express my gratitude to the individuals who have supported me throughout my academic journey. To my dear friends from university and high school, including Eleutheria,

Leonida, Rita, Elisavet, Lampi, Nikoleta, and Foteini, your persistent encouragement has meant the world to me. I also want to extend my thanks to my loving family, including my cousins, uncles, aunts, and grandma, who have always been there for me.

Finally, I am grateful to my mum, dad, brother, and boyfriend, Christos, who have been my pillars of strength, love, and inspiration, even from afar. I could not have achieved this without them.

# Notations

The complex numbers are denoted by $\mathbb{C}$, the real numbers by $\mathbb{R}$, the integers by $\mathbb{I}$ and the natural numbers by $\mathbb{N}$. Positive real numbers are denoted as $\mathbb{R}_+$, while vectors of real numbers of dimension $N$ as $\mathbb{R}^N$. The notation $|\Omega|$ is used to notate the cardinality of the set $\Omega$.

Scalars are denoted by small or big italic letters (e.g. $\alpha \in \mathbb{R}$ ), as well as, the continuously-defined functions (e.g., $F \in L_2(\mathbb{R})$). Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively (e.g., $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{A} \in \mathbb{R}^{N\times M}$). The Hermitian transpose of a complex matrix $\mathbf{B} \in \mathbb{C}^{N\times M}$ is denoted by $\mathbf{B}^*$, while for a real matrix $\mathbf{A} \in \mathbb{R}^{N\times M}$ is simply the transpose and is denoted by $\mathbf{A}^\intercal$. The $n$-th element of a vector will be denoted as $x_n$ or $(\mathbf{x})_n$, while the element in the $n$-th row and $m$-th column of a matrix will be denoted as $A_{nm}$ or $\mathbf{A}(n, m)$. The $n$-th column of a matrix will be denoted as $\mathbf{A}_n$, while the restriction of $\mathbf{A}$ to the columns indexed by the elements of $\Omega$, as $\mathbf{A}_\Omega = ([\mathbf{A}]_{\Omega_1}, \cdots, [\mathbf{A}]_{\Omega_{|\Omega|}}) \in \mathbb{R}^{N\times|\Omega|}$.

For a vector $\mathbf{x} \in \mathbb{R}^N$, $\|\mathbf{x}\|_p \in \mathbb{R}$ is the p-norm, defined as $\|\mathbf{x}\|_p^p = \sum_i |x_i|^p$. With $p = 0$, we refer to the $\ell_0$-norm defined as $\|\mathbf{x}\|_0 = |\{i : x_i \neq 0\}|$. If there is no index (e.g., $\|\mathbf{x}\|$), it is the $\ell_2$-norm.

We denote by $\mathbf{1}_N \in \mathbf{R}^N$ and $\mathbf{1}_{N\times M} \in \mathbf{R}^{N\times M}$ a vector of ones of size $N$ and a matrix of ones of size $N \times M$, respectively. Similar for $\mathbf{0}_N \in \mathbf{R}^N$ and $\mathbf{0}_{N\times M} \in \mathbf{R}^{N\times M}$, that is a vector and a matrix full of zeros. With $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ is the identity square matrix of size $N \times N$.

The column-wise vectorization of a matrix $\mathbf{A} \in \mathbb{R}^{N\times M}$ will be denoted as $\mathbf{v}_\mathbf{A}$ or $\text{vec}(\mathbf{A})$ or with the same lower-case letter $\mathbf{a} \in \mathbb{R}^{NM}$. Finally, the main diagonal of a square matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ will be denoted by $\text{diag}(\mathbf{D}) \in \mathbb{R}^N$.

# Acronyms

**ADMM** Alternate Directions Method of Multipliers

**CEL0** Continuous Exact relaxation of the $\ell_0$-norm

**COL0RME** COvariance-based $\ell_0$ super-Resolution fluorescence Microscopy with intensity Estimation

**DL** Deep Learning

**DRS** Douglas-Rachford Splitting

**FISTA** Fast - ISTA

**FluoGAN** Fluorescent image deconvolution microscopy via GAN learning

**FWHM** Full-Width at Half-Maximum

**GAN** Generative Adversarial Network

**GD** Gradient Descent

**IRL1** Iteratively Reweighted $\ell_1$

**ISTA** Iterative Shrinkage-Thresholding Algorithm

**JI** Jaccard Index

**KL** Kullback-Leibler

**LSPARCOM** Learned - SPARCOM

**MA-TIRF** Multi Angle - TIRF

**NA** Numerical Aperture

**PGD** Proximal Gradient Descent

**PnP** Plug-and-Play

**PSF** Point Spread Function

**PSNR** Peak Signal-to-Noise Ratio

**SMLM** Single Molecule Localization Microscopy

**SNR** Signal-to-Noise Ratio

**SOFI** Super-resolution Optical Fluctuation Imaging

**SPARCOM** SPARsity-based super-resolution COrrelation Microscopy

**SRRF** Super-Resolution Radial Fluctuations

**TIRF** Total Internal Reflection Fluorescence

**TV** Total Variation

**VAE** Variational Auto-Encoder

# List of Figures

# List of Algorithms

# Contents

## Model-aware Learning Approaches

# CHAPTER 1

# Introduction: Seeing through light

*Fluorescence microscopy is a popular imaging technique that allows the study of living cells and cellular organelles. However, the resolution of images obtained by fluorescence microscopes is physically limited due to the diffraction of visible light. To overcome this limitation, super-resolution fluorescence microscopy techniques have been proposed in the literature.*

*In this chapter, we provide an overview of the limitations of fluorescence microscopy and ways to overcome them. Specifically, in Section 1.1 we give more details on the critical limit of optical resolution and we present the most common microscopy modalities, in Section 1.2 we review existing super-resolution approaches, in Section 1.3 we introduce basic notions on inverse problems, in Section 1.4 we formulate mathematically the problem we are challenged to solve and, finally, in Section 1.5 we present a brief summary of our main research contributions.*

## 1.1 A journey in time: From Abbe to Super-Resolution

Fluorescence microscopy was developed in the early 20th century and is a very useful tool that allows the observation of cells and tissues with high precision. It allows the acquisition of spatial and temporal information about objects that are either intrinsically fluorescent (auto-fluorescence) or coupled to extrinsic fluorescent molecules in samples so small that cannot be seen by the naked eye [Herman, 1998, Sanderson et al., 2014].

In practice, fluorescent molecules absorb some specific light wavelengths (or colors) and emit some others with longer wavelengths. The main function of fluorescence microscopes is, therefore, to deliver excitation energy to the fluorescent species in the sample and to separate the much weaker emitted fluorescence light from the brighter excitation light so that it reaches the detector and, finally, a high contrast image is generated. The light separation is usually achieved by optical filters and the key to successful imaging is their selection according to the indicators used (i.e. fluorescent dyes).

The main advantages of fluorescence microscopy are the very high contrast of a highly specific molecule and the fact that it allows the user to select the structure of interest that he/she wants to visualize. Another paradigm tells us that the fluorescence intensity is considered proportional to the amount of the dye. As the excitation and the emission spectra are in general characteristics of the structure and composition of the molecule, fluorescence microscopy further allows the analysis of complex mixtures of molecular species [Herman, 1998].

### 1.1.1 Diffraction limit

Despite the significant improvements in the quality of fluorescence microscopes during the last century through various measures such as aberration correction, theoretical and technical innovations, and many more, there is still an ultimate limit in the spatial resolution that is imposed by the diffraction of visible light. In consequence, the highest achievable resolution that can be obtained in fluorescence microscopy is governed by some fundamental physical laws that cannot be overcome by physical means. These diffraction barriers restrict the ability of the optical instrument to distinguish between two objects which are too close to each other.

The response of the microscope to an extremely small object, such as a single fluorescent protein molecule is described by the Point Spread Function (PSF). For a perfect optical system, the PSF in the focal plane would be a diffraction pattern that is called Airy disk. Based on that and soon after the physicist Ernst Abbe advanced the diffraction-limited resolution theory in 1873 [Abbe, 1873], Lord Rayleigh established a standard formula to characterize the spatial resolution of an optical device [Lord Rayleigh, 1879]. According to Rayleigh, the resolution limit is equal to the minimum resolvable distance of two point sources, and two point sources are considered just resolved when the maximum of one diffraction pattern coincides with the first minimum of the other, as shown in Figure 1.1. This results to a resolution:

$$d_{x,y} = 0.61\frac{\lambda}{\mathrm{NA}} \tag{1.1}$$

where $\lambda$ is the emission wavelength and NA is the numerical aperture of the objective. The Numerical Aperture (NA) is given by the formula $\mathrm{NA} = n \sin\theta$ and depends on the refractive index ($n$) of the objective immersion medium and the half-angle ($\theta$) of the cone of light collected by the lens [Herman, 1998]. The lateral resolution $d_{x,y}$ is also identical to the Full-Width at Half-Maximum (FWHM) of the microscope's PSF.

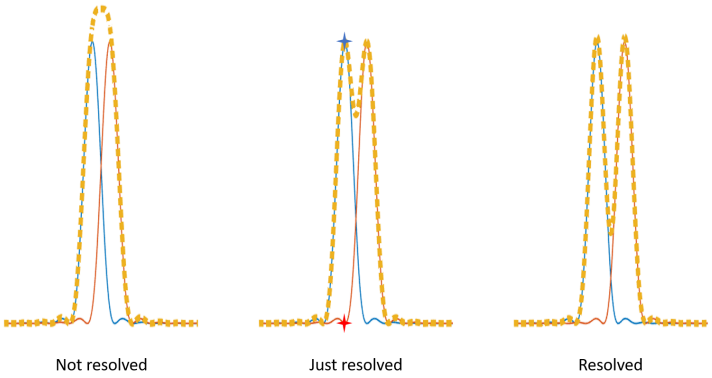



Figure 1.1 – The Rayleigh criterion: Two points are considered as just resolved when the maximum of one diffraction pattern coincides with the first minimum of the other.

According to Abbe's and Rayleigh's theory, the images obtained by fluorescence microscopes consist of many overlapping diffraction-limited spots with different intensities. The only way to improve the lateral resolution is to minimize the size of the diffraction-limited spots either by increasing the NA of the objective lens or by decreasing the wavelength of the emitted light. However, even under ideal conditions, when imaging with visible light, the lateral resolution cannot drop under the level of 200 nm.

A similar resolution limit exists also for the axial direction or the direction of the optical axis and is given by the formula:

$$d_z = \frac{2\lambda}{\mathrm{NA}^2} \tag{1.2}$$

Under ideal conditions, the axial resolution achieved by a classical fluorescence microscope used to image a sample cannot be less than 500 nm, a higher limit than the lateral plane, as the PSF forms an elliptical pattern in the axial direction. In Figure 1.2, we report a visualization of a realistic 3D widefield microscope PSF, generated by the Java software package "PSF Generator" [Griffa et al., 2010, Kirshner et al., 2013] and the Born and Wolf diffraction model [Born and Wolf, 1999].

### 1.1.2 An overview of fluorescence microscopy modalities

Different microscopy techniques can be used to enhance the visualization and contrast of an image depending on the application. Each method has advantages and disadvantages, but all use the same fluorescence mechanism to observe a biological process. The most well-known fluorescence microscopy techniques are widefield, confocal, and Total Internal Reflection Fluorescence (TIRF) microscopy. A brief overview of these three modalities is given below. Note that TIRF microscopy is a bit more detailed, as we mainly use it in our real data experiments.

**Widefield microscopy** Widefield microscopy is the most common fluorescence microscopy modality. It took its name because the whole (wide) field of view is illuminated. Observation and excitation of fluorescence are done from only one side of the sample and the entire sample is exposed at the same time to light, making it the simplest and fastest fluorescence modality.

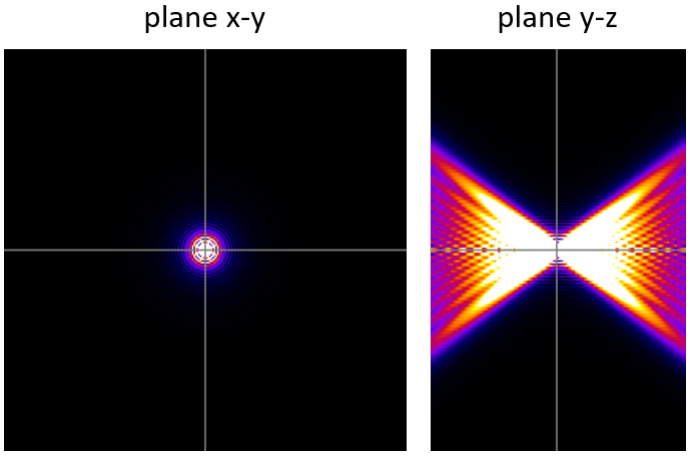



Figure 1.2 – An illustration of a realistic 3D PSF of a widefield microscope using the Born and Wolf diffraction model.

However, because fluorescence signals from all focal planes are detected, contrast is poor in thick samples, while thinner samples, such as adherent cells, are preferred.

**Confocal microscopy** In comparison to widefield microscopy for which the entire sample is illuminated at the same time, in confocal microscopy a laser light focused into a focal point (a predefined spot at a specific depth in the sample) is used to illuminate the sample little by little. Only fluorescence signals from the lighted spot can access the light detector because a pinhole blocks out-of-focus signals from entering the optical channel. The basis of modern confocal microscopes is the "double focusing system" patented by Minsky in 1961 [Minsky, 1961].

Confocal microscopy has several advantages over conventional widefield microscopy. The most important is the ability to make thin optical slices, eliminating unwanted light above and below the focal plane, providing better contrast, and improving spatial resolution. However, the acquisition procedure is much more time-consuming due to the precise scanning required. Therefore, spinning disk microscopy, a variant of confocal microscopy that scans the sample using a disk that contains more than one pinhole, was developed to reduce image acquisition time [Gräf et al., 2005].

**Total Internal Reflection Fluorescence (TIRF) microscopy** Another important fluorescence microscopy modality is the TIRF Microscopy [Axelrod, 1981, Axelrod, 2008]. It is an elegant modality that provides the excitation of fluorophores in an extremely thin axial region due to the low penetration depth of the evanescent field produced by total internal reflection. The main advantage of a TIRF microscope is that it offers a fine signal-to-noise ratio and improved spatial resolution as the out-of-focus fluorescence is almost eliminated.

TIRF microscopy exploits the properties of the evanescent field, an electromagnetic field, appearing in a small region adjacent to the interface between two media with different refractive indices. When the light propagates from a medium with a high refractive index to a medium with a lower refractive index, there is a critical angle ($\alpha_c$) beyond which the total amount of light is reflected. For angles $\alpha > \alpha_c$ the light beam is totally reflected, so the electromagnetic field

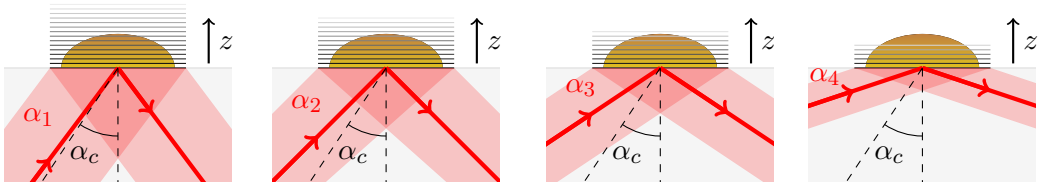



Figure 1.3 – Incident angles of the illumination in TIRF microscopy [source: [Soubies, 2016]]

(evanescent wave) penetrates the medium with an intensity $I(z, \alpha)$ that exponentially decays in the axial direction $z$ and is given by:

$$I(z, \alpha) = I_0(\alpha) e^{-zp(\alpha)}, \quad p(\alpha) = \frac{4\pi n_i}{\lambda_{\text{exc}}} \sqrt{\sin^2(\alpha) - \sin^2(a_c)} \tag{1.3}$$

where $I_0(\alpha)$ is the intensity at the interface, $p(\alpha)$ the inverse of the penetration depth, $n_i$ the refractive index of the incident medium, $\lambda_{\text{exc}}$ the excitation wavelength and $\alpha_c = \arcsin(n_t/n_i)$, with $n_t$ being the refractive index of the transmitted medium. See [Axelrod, 1981, Axelrod, 2008, Martin-Fernandez et al., 2013] for more details.

According to (1.3), the depth of the evanescent field decreases as the angle of incidence moves away from the critical angle. This can also be seen visually in Figure 1.3 if we imagine that the evanescent field is represented by lines parallel to the interface of the two media. Therefore, the optimal angle to choose depends on the proximity of the sample to the interface at which total internal reflection occurs and on the thickness of the required optical sectioning [Martin-Fernandez et al., 2013].

There are many other fluorescence microscopy imaging techniques in the literature. Among them, we recall for instance, 2-photon microscopy [Denk et al., 1990] which simultaneously excites two photons instead of one but at a longer wavelength than the emitted light, and light-sheet microscopy (see [Olarte et al., 2018] for a review) which creates a thin sheet of excitation light by splitting the excitation and detection of fluorescence into two separate light paths.

For all of the fluorescence modalities above, it is possible to image also a sample in three dimensions (3D). A special reference is given to the TIRF modality, where someone could use the Multi Angle - TIRF (MA-TIRF) microscope, which is simply a TIRF microscope with an additional module changing the angle of the illumination light beam. It provides access to 3D information through optical sectioning. More details on the MA-TIRF microscope are given in Chapter 3.

### 1.1.3 Super-resolution: The only way forward

The common disadvantage of all the fluorescence microscopy modalities described in the previous subsection is that they produce images with low spatial resolution due to light diffraction. This diffraction barrier has created a new challenge as many entities of biological interest are smaller than such a barrier (see Figure 1.4). A variety of methods called *super-resolution* techniques, can break the diffraction limit of light and will be discussed below.

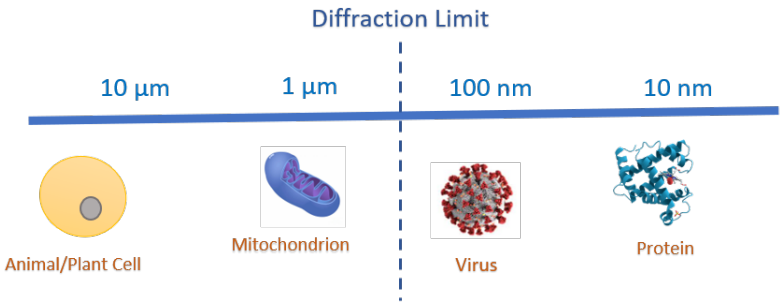



Figure 1.4 – Microscopy scale: There are important biological entities below the diffraction limit.

## 1.2 Super-Resolution fluorescence microscopy

Super-resolution fluorescence microscopy is a set of techniques capable of providing images with a higher resolution than the one imposed by the diffraction barrier. There are many super-resolution techniques in the literature and all of them have advantages and disadvantages depending on the type of sample to be imaged, the purposes served (e.g., not to damage the sample), or the microscopy equipment available (e.g., microscope and fluorophores) and many more. For the purposes of this thesis, we choose to classify into these two categories: illumination- and single-molecule-based super-resolution methods, which require a specific configuration, and fluctuation-based super-resolution methods, which can be applied in standard setups.

### 1.2.1 Illumination- and Single-Molecule-based Super-Resolution methods

It was not until the late 20th century that a microscopic modality overcoming the diffraction limit and significantly improving spatial resolution emerged. Such modality is known as Stimulated Emission Depletion (STED) microscopy [Hell and Wichmann, 1994]. In STED microscopy, the size of the PSF is reduced due to a depletion light beam which induces stimulated emission from molecules in a doughnut-shape around the center of the PSF and thus switches them off. In other words, the STED microscope minimizes the area of illumination at the focal point and therefore improves its resolution. Although the spatial resolution achieved is promising (e.g. lateral resolution of less than 50 nm can be reached), STED techniques have a relatively slow acquisition process and require a very expensive commercial setup as well as special fluorophores with specific properties (e.g., fluorophores that can return to the steady state with a high-powered laser).

Soon after STED, Structured Illumination Microscopy (SIM) methods that use patterned illumination to excite the sample, appeared in the literature [Gustafsson, 2000]. Differently from STED, images here can be recovered with high temporal resolution via high-speed acquisitions that cause comparatively little damage to the sample, but at the cost of a relatively low spatial resolution and, more importantly, the requirement of a specific illumination setup. Random illumination microscopy (RIM) [Mudry et al., 2012] has been proposed as a more robust imaging technique than SIM, but the maximum super-resolution gain of RIM is the same as that of SIM [Idier et al., 2018].

A large and powerful family of imaging techniques achieving nanometric resolution are the ones often known as Single Molecule Localization Microscopy (SMLM) techniques. Among them, methods such as Photo-Activated Localization Microscopy (PALM) [Betzig et al., 2006]

and Stochastic Optical Reconstruction Microscopy (STORM) [Rust et al., 2006] have been designed to create a super-resolved image (achieving up to 20 nm of resolution) by activating and precisely localizing only a few molecules in each of thousands of acquired frames. The localization is done by some software used to detect and accurately localize point sources, see, e.g. [Sage et al., 2015, Sage et al., 2019] for a review. For their use, these methods need specific photoactivatable, photoswitchable, and binding-activated fluorophores, among others [Li and Vaughan, 2018], as well as a large number (typically thousands) of sparse acquired frames leading to a poor temporal resolution and large exposure times which can significantly damage the sample.

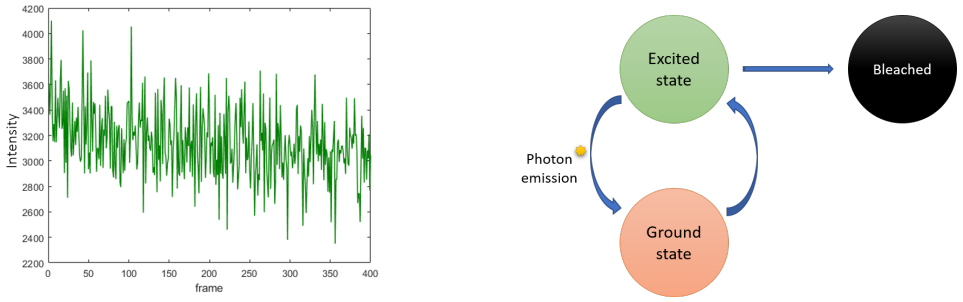
Many 3D super-resolution techniques have also appeared in the literature. Often they are extensions of the techniques above, that overcome the diffraction limitations. The widely known ones are 3D-STED [Harke et al., 2008] which can achieve very good levels of spatial resolution but at the expense of fast photobleaching (that is the change of the structure of the fluorophore so that it can no longer fluoresce) and the requirement of special fluorophores, 3D-SIM [Gustafsson et al., 2008] that has short acquisition times (improved temporal resolution), but limited spatial resolution and 3D-STORM [Huang et al., 2008] that can achieve also very good levels of spatial resolution but at the expense of long acquisition times. The most resolutive method in 3D, to our knowledge, is iPALM [Shtengel et al., 2009], reaching up to 10 nm of axial resolution, which combines the STORM approach and interferometry at the price of a highly complex prototype.

### 1.2.2 Fluctuation-based Super-Resolution methods in standard setups

During the last decade, new approaches taking advantage of the independent stochastic temporal fluctuations/blinking of conventional fluorescent emitters appeared in the literature. No specific material is needed here, neither for the illumination setup nor for the fluorescent dyes. Due to standard acquisition settings, temporal resolution properties are drastically improved. The idea is to acquire a stack of images at a high temporal rate, typically $10 - 100$ images/s, by means of common microscopes (such as widefield, confocal, or TIRF) using standard fluorophores, and then to exploit their independent fluctuations to compute a single super-resolved image.

To further illustrate what we mean by fluctuations, Figure 1.5a shows the temporal behavior of a single pixel containing several fluctuating/blinking fluorescent molecules, while Figure 1.5b shows the states in which a fluorescent molecule passes through: it passes randomly (with a certain probability) from the excited to the ground state by emitting a photon and returns to the excited state to start the circle again until it bleaches, i.e. it is no longer fluorescent.

Several methods exploiting fluorophore fluctuations have been proposed in recent years. First of all, the Super-resolution Optical Fluctuation Imaging (SOFI) [Dertinger et al., 2009] is a powerful technique where a second and/or higher order statistical analysis is performed, which leads to a reduction of the size of the PSF and thus to an improvement of the resolution (of a factor almost proportional to the order of the cumulants used, see [Dertinger et al., 2010]). However, due to an amplification of heterogeneities in molecular luminosity and blinking statistics, the use of higher-order cumulants is limited and so is the resolution. An extension of SOFI that combines several cumulant orders and achieves better resolution levels than SOFI is the method balanced-SOFI (bSOFI) [Geissbuehler et al., 2012]. However, the spatial resolution still cannot reach the same levels of PALM/STORM (SMLM methods referred to in the previous subsection).




(a) (b)

Figure 1.5 – (a) The temporal profile of a pixel in terms of fluorescence intensity (measured in arbitrary units) captured at a rate of 40 frames per second (fps), (b) The states of a fluorescent molecule.

Another method taking as input a temporal stack of diffraction-limited frames is Super-Resolution Radial Fluctuations (SRRF) [Gustafsson et al., 2016] microscopy. In this method, super-resolution is achieved by calculating, for each frame, a degree of local symmetry based on the intrinsic symmetry of the microscope PSF. SRRF calculates the degree of local gradient convergence (called 'radiality' by the authors) on a sub-pixel basis over the entire frame. For a sub-pixel that is very close to a fluorescent molecule, the result is a high degree of local gradient convergence and thus a high radiality value. In contrast, for a sub-pixel displaced from a true fluorescent molecule, the neighboring gradients have low convergence because they either point in the direction of the true molecule or are randomly aligned if there is no molecule nearby (only noise). To generate a single super-resolution image from the resulting stack of radiality images, temporal analysis is done by either estimating a time average or performing, as in SOFI, higher-order statistical analysis. SRRF has a wide applicability since it is very easy to manipulate: the author has accompanied the method with an easy-to-use plug-in for the freeware Fiji [1] [Laine et al., 2019]. However, SRRF often creates reconstruction artifacts which may limit its use in view of accurate analysis (see 'Numerical Results' sections in the following chapters).

Other methods which belong to the same category and are worth mentioning are: the method 3B [Cox et al., 2011], which uses Bayesian analysis and takes advantage of the blinking and bleaching events of standard fluorescent molecules, the method Entropy-based Super-resolution Imaging (ESI) [Yahiatene et al., 2015] that computes entropy pixel-wise values, weighted with higher order statistics and the method Spatial COvariance REconstructive (SCORE) [Deng et al., 2014] that analyzes intensity statistics, similarly to SOFI, but further reduces noise and computational cost by computing only a few components that have a significant contribution to the intensity variances of the pixels.

In addition, the approach SPARsity-based super-resolution COrrelation Microscopy (SPARCOM) [Solomon et al., 2018, Solomon et al., 2019] exploits, as SOFI, both the lack of correlation between distinct emitters as well as the sparse distribution of the fluorescent molecules via the use of an $\ell_1$ regularization defined on the emitters' covariance matrix. Along the same lines, a Deep Learning (DL) method exploiting algorithmic unrolling, called Learned - SPARCOM (LSPARCOM) [Dardikman-Yoffe and Eldar, 2020], has recently been introduced (see

1. https://github.com/HenriquesLab/NanoJ-Core

[Monga et al., 2021] for a review on algorithm unrolling methods). Differently from plain SPARCOM, the advantage of LSPARCOM is that neither previous knowledge of the PSF nor any heuristic choice of the regularization parameter for tuning the sparsity level is required. As far as the reconstruction quality is concerned, both SPARCOM and LSPARCOM create some artifacts under challenging imaging conditions, for example when the noise and/or background level are relatively high (see numerical results in Section 2.4).

A super-resolution method exploiting the temporal fluctuation of conventional fluorescent dyes that is worth mentioning as it is based on an off-the-grid variational approach is the one by Laville *et al.* [Laville et al., 2022]. The great advantage of the gridless approaches (see [Laville et al., 2021] for a review) is that there are no limitations imposed by the size of the discrete grid considered. However, when the method is applied to biological images, especially those containing filament structures, it produces a point-like reconstruction.

Finally, without using higher-order statistics, a constrained tensor modeling approach that estimates a map of local molecule densities and their overall intensities, as well as a matrix-based formulation that promotes structure sparsity via an $\ell_0$ type regularizer, are available in [de Morais Goulart et al., 2019]. These approaches can achieve good temporal resolution levels, but the spatial resolution is limited.

## 1.3 Mathematical Formulation as an Inverse Problem

Each imaging system can be represented approximately by a model; this is the case, also, for fluorescence microscopes. When interested in quantities that are not directly accessible from the observations but through the model of observations, we have to solve an inverse problem. In this section, we first present the general framework of inverse problems in imaging as well as a broad overview of possible ways to solve them, and then we formulate the fluorescence microscopy imaging model.

### 1.3.1 Introduction to Inverse Problems

When a physical quantity of interest is not directly accessible from measurements, it is common to proceed by observing other quantities that are connected with it through some physical laws. The notion of an *inverse problem* corresponds to the idea of inverting these physical laws to gain indirect access to the quantity of interest [Idier, 2008].

A very large and interesting subfield of inverse problems is the one of imaging, having applications from microscopy to medicine and astronomy. The calculation of an image of X-ray attenuations from a sinogram in X-ray computed tomography (CT) or a "clean" image from noisy, blurred, and possibly undersampled data in fluorescence microscopy imaging applications are some examples of inverse problems.

In many realistic settings, the image of interest, therefore, has to be estimated from data, by exploiting the model describing the physical laws, called the forward model. In general, we seek to find a solution $x$ that belongs to the set $\mathcal{X}$, i.e., $x \in \mathcal{X}$, where $\mathcal{X}$ contains all possible values that $x$ can take, of the following inverse problem:

$$\text{find } x \quad \text{s.t.} \quad \mathbf{y} = \mathcal{H}(x) \in \mathbb{R}^M. \tag{1.4}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the discrete vector of measurements and $\mathcal{H} : \mathcal{X} \to \mathbb{R}^M$ is the forward operator that maps $x$ to the measurements $\mathbf{y}$. For example, in X-ray CT, $x$ represents the desired image

of X-ray attenuations, $\mathbf{y}$ is the measurement data (i.e. the resulting sinogram) and $\mathcal{H}$ is a Radon transform operator. In fluorescence microscopy, $x$ would be the high-resolution blur- and noise-free image, $\mathbf{y}$ the corrupted data acquired by the CCD camera of a fluorescence microscope, and $\mathcal{H}$ an operator related to the PSF of the microscope.

The acquisition process, however, is impossible to be perfectly simulated or characterized since there are many factors (noise, aging of the system, etc.) that cannot be precisely determined. Therefore, in most methods the operator $\mathcal{H}$ of the imaging system is replaced by a deterministic operator $H$ and a random noise term $\mathbf{n} \in \mathbb{R}^M$, as follows:

$$\mathcal{H}(x) \approx H(x) + \mathbf{n}. \tag{1.5}$$

Finally, for simplification, let us assume that the reconstruction belongs to a discrete domain $\mathbf{x} \in \mathbb{R}^L$ and the operator $H$ is linear so that we can write the forward model as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1.6}$$

where $\mathbf{H} \in \mathbb{R}^{M \times L}$ is the matrix representing the linear operator $H$.

### 1.3.2 Ill-posed Inverse Problems

Often, the inverse problems we meet in practice have a nasty tendency to be "naturally unstable", meaning that if there are errors, even tiny, on the data, "naive" inversion methods turn out to not be robust [Idier, 2008].

The French mathematician Jacques Hadamard, in the early 20th century, characterized these unstable problems as *ill-posed* [Hadamard, 1902]. According to Hadamard, a problem is considered to be well-posed if its solution (a) exists, (b) is unique, and (c) is continuous under infinitesimal changes of the data, in the sense that if a solution changes considerably for infinitesimal changes of the data, it is not considered a solution in a physical sense [Bertero and Boccacci, 1998]. The problem is thus ill-posed if it violates any of these three conditions [Hadamard, 1902]. In problems arising from physics, even if the first two conditions can be verified, the stability condition is often violated. Error amplifications have to be taken into consideration to prevent unstable results.

We recall that the discrete inverse problem, we have formulated in the previous section, has now the following form:

$$\text{find } \mathbf{x} \in \mathbb{R}^L \quad \text{s.t.} \quad \mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \in \mathbb{R}^M. \tag{1.7}$$

Finding a solution to this problem is a challenging task, especially in imaging applications where one often aims to reconstruct an image with a larger number of pixels ($L$) from an image with a smaller number of pixels ($M$), where $L > M$. This leads to a matrix $\mathbf{H}$ with no inverse and to a problem with infinitely-many solutions, violating the condition of uniqueness in Hadamard's sense. Even when the matrix $\mathbf{H}$ is non-singular, so it has a matrix inverse and the problem has a unique solution, the solution can be very sensitive to noise, violating the third condition in Hadamard's sense. The following example illustrates this easily.

In case the distribution of the noise is known and the matrix $\mathbf{H}$ is non-singular (an assumption to avoid violating the condition of uniqueness), we could find $\mathbf{x}$ solving a maximum-likelihood (ML) estimation problem as follows:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) = \arg\min_{\mathbf{x}} -\ln p(\mathbf{y}|\mathbf{x}) \tag{1.8}$$

where $p(\mathbf{y}|\mathbf{x})$ is the likelihood of observing $\mathbf{y}$ given $\mathbf{x}$. If we further assume that the noise $\mathbf{n}$ is a vector of independent and identically distributed (i.i.d.) white Gaussian random variables with zero mean and independent from data $\mathbf{y}$, the solution is given by the minimization of the following least square term, also called data fidelity term:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2. \tag{1.9}$$

In case the noise is not Gaussian but rather another type, e.g. Poisson, a different term is minimized.

If $\mathbf{H}^*\mathbf{H}$ is invertible, as the function minimized is quadratic, we will find the unique solution $\hat{\mathbf{x}} = (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{y}$, with $\mathbf{H}^*$ the Hermitian transpose of $\mathbf{H}$. Using the expression of the forward model, $\mathbf{y} = \mathbf{H}\mathbf{x}^{\text{exact}} + \mathbf{n}$, where $\mathbf{x}^{\text{exact}}$ is the exact or true solution to the inverse problem, we will find that: $\hat{\mathbf{x}} = \mathbf{x}^{\text{exact}} + (\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*\mathbf{n}$. As can be seen, the estimated solution contains the exact solution but also an additional term that corrupts the solution with noise. Especially when the matrix $\mathbf{H}$ is ill-conditioned or equivalently has singular values close to 0, something that is true for many forward operators used to represent physical models, the inversion of $\mathbf{H}^*\mathbf{H}$ will result in a noise explosion, corrupting totally the solution with noise.

### 1.3.3 How to overcome ill-posedness

As we saw in the previous section a discrete problem such as (1.7) is in general ill-posed. Either it has infinitely-many solutions, in case we would like to find a better resolution than the one of our measurements ($L > M$), or either it has a unique but unacceptable solution due to a complete corruption of the real solution by a small error in the measurements (when $L <= M$).

In the first scenario, we are trying to find a way to select one among an infinite number of solutions, while in the second we are looking for a physically accepted solution that is not a solution to the problem but only an approximate solution in the sense that it does not reproduce the data exactly but only within the experimental errors [Bertero and Boccacci, 1998]. The question now is how to choose good solutions.

#### 1.3.3.1 Regularization Methods

One way to select a good solution among many is to include some *prior* knowledge or information on the solution to the optimization problem. By adding this information we reduce the set of solutions and can discriminate between good and bad solutions, in the sense of solutions that have a physical meaning and others that do not. These methods which take into account prior knowledge of the solution are called *regularization methods*. The first theory of regularization methods was formulated by A.N. Tikhonov, a mathematician, and geophysicist, in 1963 [Tikhonov, 1963]. Therefore many articles and books have been written on regularization theory (see for example [Engl et al., 2000]).

Possible *prior* information we may have could be that the intensity of the solution or the solution itself is not too big. Another one could be that the solution we want to retrieve has smooth intensities and therefore its derivatives should be small enough. Furthermore, especially in biology, it is very common to search only for non-negative solutions, which is also prior information.

This *a priori* information is typically encoded into a function $\mathcal{R} : \mathbb{R}^L \to \mathbb{R}$ and can be taken into account as a constraint or in a penalized form. The minimization problem expressed as a

constrained optimization problem has the two following forms depending on where we would like to use the constraint, in the data fidelity term or the prior:

$$\arg\min_{\mathbf{x}} \mathcal{R}(\mathbf{x}) \quad \text{s. t.} \quad \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \leq M\sigma^2, \quad \sigma > 0, \tag{1.10}$$

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \quad \text{s. t.} \quad \mathcal{R}(\mathbf{x}) \leq \kappa, \quad \kappa > 0. \tag{1.11}$$

In the penalized formulation of the optimization problem, we minimize the following objective function containing both the data fidelity term and the prior term:

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\mathcal{R}(\mathbf{x}). \tag{1.12}$$

The parameters $\sigma^2 > 0$ and $\kappa > 0$ used in the constrained formulations are easy to interpret. For example, if we would like to constrain the data term and if we assume that the data are corrupted with white Gaussian noise, that justifies also the appearance of the least square term, we can choose $\sigma^2$ to be the constant noise variance. On the other side the parameter $\lambda > 0$, often referred to as the regularization parameter, is a bit more difficult to interpret. The role of this parameter is to control the relative strength of the data-fidelity term and the regularization term, in the minimization.

The choice of either the penalized or the constrained formulation is a personal choice. Practically, these two problems are equivalent, while theoretically, we have equivalence only in convex cases. In the following, we will choose the penalized form because we have mostly "soft" constraints and the minimization of the penalized form is in general simpler and faster.

#### 1.3.3.2 Tikhonov Regularization

The Tikhonov regularization, also called ridge regression, is a possible way to regularize inverse problems. The regularization function used in this case has the form $\|\mathbf{\Gamma}\mathbf{x}\|_2^2$, where $\mathbf{\Gamma} \in \mathbb{R}^{\tilde{L}\times L}$ is a general matrix, with $\tilde{L} \in \mathbb{R}$. It penalizes large values in $\mathbf{x}$ in case $\mathbf{\Gamma} = \mathbf{I}_L$ or, depending on the matrix $\mathbf{\Gamma}$, another unwanted property. The problem (1.12) takes the form:

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{\Gamma}\mathbf{x}\|_2^2 \tag{1.13}$$

The solution to this problem is:

$$\hat{\mathbf{x}} = (\mathbf{H}^*\mathbf{H} + 2\lambda\mathbf{\Gamma}^*\mathbf{\Gamma})^{-1}\mathbf{H}^*\mathbf{y} \tag{1.14}$$

where the symbol $^*$ refers to the Hermitian transpose.

With this formulation if we ask $\mathbf{\Gamma}$ to be full rank then we will be able to have a unique solution in the case we want to recover an image with more pixels than the measurements ($L > M$). Moreover, the noise amplification will be restricted because with the right selection of $\mathbf{\Gamma}$, the matrix we want to invert will not be ill-conditioned, as it will have all singular values away from 0.

However, sometimes, even if the inverse exists, it is difficult from a computational point of view to compute it, due to high dimensions. For this reason, an iterative gradient schema of the following format can be used:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \tau\nabla\mathcal{F}(\mathbf{x}^k), \quad k \geq 0, \tau > 0 \tag{1.15}$$

where $\mathcal{F}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$ and $\nabla\mathcal{F}(\mathbf{x}^k)$, the gradient of $\mathcal{F}$ with respect to (w.r.t.) $\mathbf{x}^k$. It is given by:

$$\nabla\mathcal{F}(\mathbf{x}^k) = (\mathbf{H}^*\mathbf{H} + 2\lambda\boldsymbol{\Gamma}^*\boldsymbol{\Gamma})\mathbf{x}^k - \mathbf{H}^*\mathbf{y}. \tag{1.16}$$

This iterative schema is called Gradient Descent (GD) with a fixed step size $\tau$ and is used to minimize differentiable functions $\mathcal{F}$, like the one in (1.13). The GD is guaranteed to converge when $\tau \leq \frac{2}{L}$, where $L > 0$ is the Lipschitz constant of $\nabla\mathcal{F}$. In addition to GD, there are accelerated versions such as the Accelerated Gradient Descent introduced by Nesterov in 1983 [Nesterov, 1983], and more advanced methods like the Conjugate Gradient method introduced by Hestenes in 1952 [Hestenes and Stiefel, 1952].

#### 1.3.3.3 Sparse Representation

During the last few decades, it has been observed that many real-world images, can be approximated by a sparse representation on a certain basis. By sparse we mean that only a few non-zero coefficients are required to represent a high-dimensional image. The goal is therefore to penalize images with many non-zero coefficients. This could be done by replacing the quadratic regularization term $\|\boldsymbol{\Gamma}\mathbf{x}\|_2^2$ with another one promoting sparsity.

The most suitable term that penalizes non-sparse entries is the $\ell_0$ pseudo-norm that counts the number of the non-zero elements and is defined as $\|\mathbf{z}\|_0 = |\{i : z_i \neq 0\}|$. It is only a pseudo-norm because it is invariant by multiplication, i.e., $\lambda \neq 0$, $\lambda\|\mathbf{x}\|_0 = \|\mathbf{x}\|_0$. By abuse of terminology, we will refer to it as the $\ell_0$-norm. Therefore the minimization problem will have the following form:

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\boldsymbol{\Gamma}\mathbf{x}\|_0. \tag{1.17}$$

However problem (1.17) is non-convex and non-continuous due to the combinatorial nature of the $\ell_0$-norm and thus very challenging to solve. To make the problem tractable the $\ell_0$- norm is replaced by another function.

**Convex $\ell_1$ relaxation** The first choice is the $\ell_1$-norm that is the simplest sparsity-inducing function. It is defined as $\|\mathbf{z}\|_1 = \sum_i |z_i|$. Besides convexity, the key difference between using the $\ell_0$ and the $\ell_1$-norm is that the $\ell_0$ provides a correct interpretation of sparsity by counting only the number of the non-zero coefficients, while the $\ell_1$ depends also on the magnitude of the coefficients. However, its use as a sparsity-promoting regularizer is nowadays well-established (see, e.g., [Candès et al., 2007]). Using now the $\ell_1$-norm the minimization problem takes the form:

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\boldsymbol{\Gamma}\mathbf{x}\|_1. \tag{1.18}$$

If we select the operator $\boldsymbol{\Gamma}$ to be equal to the identity $\mathbf{I}_L$, the minimization problem (1.18) is typically known under the name Least Absolute Shrinkage and Selection Operator (LASSO) [Tibshirani, 1996] and has the following form:

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1. \tag{1.19}$$

To solve the LASSO problem (1.19), a popular way is using a sub-category of the proximal forward-backward splitting algorithms, the Iterative Shrinkage-Thresholding Algorithm (ISTA), see e.g., [Chambolle et al., 1998, Daubechies et al., 2003, Bect et al., 2004]. It is important to

specify that such algorithms can also be used in the more general problem (1.18). Each iteration of ISTA involves a gradient update step followed by a shrinkage/soft-thresholding step, i.e.:

$$\mathbf{x}^{k+1} = \mathcal{T}_{\lambda\tau}\left(\mathbf{x}^k - \tau\mathbf{H}^*\left(\mathbf{H}\mathbf{x}^k - \mathbf{y}\right)\right), \quad \kappa \geq 0, \tau > 0 \tag{1.20}$$

where $\mathcal{T}_\alpha : \mathbb{R}^L \to \mathbb{R}^L$ is the soft-thresholding operator defined as:

$$(\mathcal{T}_\alpha(\mathbf{z}))_i = (|x_i| - \alpha)_+ \text{sgn}(x_i), \quad \mathbf{z} \in \mathbb{R}^L, \tag{1.21}$$

with $x_i$, $i \in \{1, \ldots, L\}$, being the $i$-th element of the vector $\mathbf{x} \in \mathbb{R}^L$ and $(\cdot)_+$ the positive part.

The soft-thresholding or shrinkage operator defined in (1.21) is simply the proximal mapping of the $\ell_1$-norm and can be derived easily in 1D, using the definition of the proximal operator (see the Definition 1.1).

**Definition 1.1.** The proximal operator, $\text{prox}_f : \mathbb{R}^L \to \mathbb{R}^L$ of a convex, proper, lower semi-continuous function $f : \mathbb{R}^L \to \mathbb{R}$ is given by:

$$\text{prox}_f(\mathbf{z}) = \arg\min_{\mathbf{u}\in\mathbb{R}^L} f(\mathbf{u}) + \frac{1}{2}\|\mathbf{z} - \mathbf{u}\|_2^2, \quad \mathbf{z} \in \mathbb{R}^L. \tag{1.22}$$

There are also faster algorithms than ISTA solving (1.18). An algorithm that keeps the simplicity of ISTA but has an improved convergence rate is the Fast - ISTA (FISTA) [Beck and Teboulle, 2009].

However, sometimes it may be desirable not to enforce sparsity directly on the image, as for LASSO, but rather to the matrix-vector product of the image with a specific family of filters that we would like to mostly vanish. After noticing that real-world images have often sharp edges and piece-wise constant regions, a classical idea is enforcing sparsity on the spatial derivatives of the image. To do so, it is sufficient to replace $\mathbf{\Gamma}$ in (1.18) with a discrete gradient operator that computes finite differences with respect to all image dimensions. The penalty introduced is the so-called Total-Variation (TV) penalty, first used by Rudin, Osher and Fatemi in 1992 [Rudin et al., 1992] for the task of denoising (when $\mathbf{H} = \mathbf{I}_L$ and $M = L$ in (1.18)).

Problems with $\mathbf{\Gamma}$ different from the identity $\mathbf{I}_L$ are more difficult to be solved than the LASSO problem. More than forward-backward splitting gradient methods, saddle point methods such as Primal-Dual methods [Chambolle and Pock, 2011], Alternate Directions Method of Multipliers (ADMM) [Boyd et al., 2011] or Douglas-Rachford Splitting (DRS) method [Douglas and Rachford, 1956] could be used.

**Non-convex exact relaxation** Another way to circumvent the difficulty imposed by the $\ell_0$-norm in (1.17) without penalizing the magnitude of $\mathbf{x}$ (a bad side-effect that the $\ell_1$ penalty has), consists in using other, still non-convex, but continuous function. One example is the Continuous Exact relaxation of the $\ell_0$-norm (CEL0) proposed by Soubies *et al.* in [Soubies et al., 2015] that is continuous, non-convex, and preserves the global minima of the original $\ell_2 - \ell_0$ problem while removing some local ones. The minimization problem (1.17), with $\mathbf{\Gamma} = \mathbf{I}_L$, reformulates as follows:

$$\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \Phi_{\text{CEL0}}(\mathbf{x}; \lambda) \tag{1.23}$$

where the CEL0 penalty has an explicit expression given by:

$$\Phi_{\text{CEL0}}(\mathbf{x};\lambda) = \sum_{i=1}^{L^2} \lambda - \frac{\|\mathbf{H}_i\|^2}{2}\left(|x_i| - \frac{\sqrt{2\lambda}}{\|\mathbf{H}_i\|}\right)^2 \mathbb{1}_{\{|x_i| \leq \frac{\sqrt{2\lambda}}{\|\mathbf{H}_i\|}\}}, \tag{1.24}$$

with $\mathbf{H}_i$ being the $i$-th column of the matrix $\mathbf{H}$.

To minimize (1.23), the authors in [Gazagnes et al., 2017] propose to use the Iteratively Reweighted $\ell_1$ (IRL1) algorithm [Ochs et al., 2015] with FISTA [Beck and Teboulle, 2009] as an inner solver.

## 1.4 Inverse Problems in Fluorescence Microscopy

As in this thesis we deal with inverse problems in fluorescence microscopy, in this section, we discuss the model describing acquisitions by a fluorescence microscope. Given the model, the acquisitions from the fluorescence microscope, an estimate of the PSF, and using some of the optimization tools described in Section 1.3, we aim to reconstruct images with a spatial resolution beyond the diffraction limit to resolve small details.

We start with the continuous framework, to better present the fluorescence microscopy imaging model. Let us assume that in the sample there are $K > 1$ single, independently blinking/fluctuating, emitters. Each emitter is located at a position $\mathbf{r}_k \in \mathbb{R}^N$, where $N$ the space dimensions, and has an intensity over time $t = 1, \ldots, T$, collected in the vector $\mathbf{s}_k \in \mathbb{R}^T$. Then, the true image intensity at position $\mathbf{r}$ and time $t$ is given by:

$$x(\mathbf{r}, t) = \sum_{k=1}^{K} \delta(\mathbf{r} - \mathbf{r}_k)\mathbf{s}_k(t), \tag{1.25}$$

where $\delta(\cdot)$ is the Dirac's delta function.

If the microscope has a PSF $v(\mathbf{r})$ that is spatial- and time-invariant, then the diffraction-limited image $y(\mathbf{r}, t)$ at position $\mathbf{r}$ and time $t$ will be given by the convolution of the PSF of the system with $x(\mathbf{r}, t)$:

$$y(\mathbf{r}, t) = \sum_{k=1}^{K} v(\mathbf{r} - \mathbf{r}_k)\mathbf{s}_k(t). \tag{1.26}$$

We now discretize the model (1.26) using a finer grid for the locations of the emitters we want to retrieve, so as to localize them with better accuracy. In the discretized model we include also a noise component. The discretized model for all $t = 1, \ldots, T$ with $T > 0$, reads:

$$\mathbf{y_t} = M_q\left(U\left(\mathbf{x_t}\right)\right) + \mathbf{n_t}. \tag{1.27}$$

This model links the acquisition $\mathbf{y_t} \in \mathbb{R}^M$ with the non-directly accessible but only through the model true image $\mathbf{x_t} \in \mathbb{R}^L$ that belongs in a $q^N$-times finer grid, as soon as $L = q^N M$. In the model there is, also, the composition of the convolution operator $U : \mathbb{R}^L \to \mathbb{R}^L$ associated with the PSF and the downsampling operator $M_q : \mathbb{R}^L \to \mathbb{R}^M$ that simply returns the sum of every $q^N$ non-overlapping consecutive pixel blocks. Finally, there is the noise component $\mathbf{n_t}$, which describes the presence of signal-independent read noise and model errors. It is often modeled as a vector of independent and identically distributed (i.i.d.) Gaussian random variables with a zero mean and a constant variance.

Moreover, in real scenarios, the background, which is modeled here by the vector $\mathbf{b} \in \mathbb{R}^M$ and gathers contributions from out-of-focus (and ambient) fluorescent molecules, is present along with another type of noise, the signal-dependent photon noise. A more accurate fluorescence microscopy model thus is:

$$\mathbf{y_t} = \mathcal{P}\left(M_q\left(U\left(\mathbf{x_t}\right)\right) + \mathbf{b_t}\right) + \mathbf{n_t}, \tag{1.28}$$

where, for $\mathbf{w} \in \mathbb{R}^M$, $\mathcal{P}(\mathbf{w})$ represents the realization of a multivariate Poisson variable of parameter $\mathbf{w}$.

## 1.5 Structure of the thesis

In this thesis, we are proposing reconstruction algorithms for 2D (lateral plane) and 3D (lateral plane and axial direction) live-cell fluorescence microscopy imaging applications. Differently to well-known super-resolution techniques, e.g. STED [Hell and Wichmann, 1994], SMLM (see [Sage et al., 2015] for a review), etc., we use standard fluorophores, not complex setups and we have fast acquisition procedures. In short, we are able to reconstruct images with fine spatial and temporal resolution without damaging the sample and without the need for special equipment.

We divide the thesis into two main parts. In the first part, we present model-based approaches: for 2D super-resolution (in Chapter 2) and for 3D super-resolution in TIRF imaging applications (in Chapter 3). In the second part, we present hybrid approaches, combining data-driven models, based on deep learning, with the fluorescence microscopy physical-model knowledge we have. More precisely, the second part includes fluorescence image reconstruction using model-aware generative approaches (in Chapter 4) and Plug-and-Play (PnP) priors (in Chapter 5).

To begin, in Chapter 2 we present the method COvariance-based $\ell_0$ super-Resolution fluorescence Microscopy with intensity Estimation (COL0RME) for super-resolution in the lateral plane. The method has two steps, both of them based on a variational framework. The first step is called *support estimation* and aims at precisely localizing the fluorescent emitters in a finer grid than the one of the acquisitions. In this step, we solve a sparse optimization problem in the covariance domain. By working with temporal covariance matrices we have the opportunity to exploit the temporal statistical independence of the intensities of distinct fluorescent emitters. The second step is called *intensity estimation* and as the name suggests, real intensity values of the previously accurately localized emitters are estimated. An optimization problem is now solved in the image domain. We evaluate the performance of the method on both simulated and real datasets and we observe significant improvements in resolution, competitive and sometimes better than other state-of-the-art approaches.

In Chapter 3 we present a method for 3D image super-resolution when using a MA-TIRF microscope. With the MA-TIRF microscope we are able to acquire images under different angles of illumination and using a dedicated reconstruction algorithm, we have access to depth information of a thin layer adjacent to the microscope coverslip. A pre-existing 3D MA-TIRF reconstruction algorithm reaches very good levels of axial resolution but has limited spatial resolution. With the method we propose, we maintain good levels of axial resolution but we improve also the lateral one. We validate the proposed method on simulated data and on challenging real MA-TIRF acquisitions, showing significant resolution improvements.

In Chapter 4 we are proposing two unsupervised hybrid approaches that combine the physical modeling of fluorescence microscopy time-lapse acquisitions with generative approaches to deal with the problem of image deconvolution. In the first approach, we minimize a distance between the distribution of the observed data (i.e. the temporal stack of images acquired by a standard fluorescence microscope) and the one of samples generated by a model-aware simulator. Following Generative Adversarial Network (GAN) framework, the distance between these two distributions is computed using adversarial training of two competing architectures: a physics-inspired *generator* simulating the fluctuation behavior of the observed images combined with blur and undersampling, and a standard convolutional *discriminator*. The proposed method has been applied to both simulated and real data, while it reaches very good levels of lateral resolution. In the second approach, we introduce the fluorescence microscopy physical model into the decoder of a standard Variational Auto-Encoder (VAE). In such a way, we are able to predict a whole distribution of noise- and blur-free images, from which we are then able to sample. The method has so far only been tested on simulated data showing deconvolution results.

Finally, in Chapter 5 we present an extension of the support estimation step of COL0RME (the method presented in Chapter 2) for a more accurate and less dotted fluorescent emitter localization. Being inspired by the Plug-and-Play (PnP) priors, a framework to solve ill-posed inverse problems through the integration of physical and learned models, we impose in the loss function that we minimize through an iterative scheme, a learned regularizer in the form of an image denoiser. The regularizer is parameterized by a deep neural network trained on simulated data characterized by the desired structures of interest (filament-like structures in our case). Some first deconvolution results of the method on simulated and real data are available in Chapter 5.

# PART

# Model-based Approaches

# CHAPTER 2

# COL0RME: Localization and quantification of fluorescent signal

*In this chapter, we present COL0RME, a method for super-resolution fluorescence microscopy that takes as an input a temporal stack of diffraction-limited images and gives as an output a super-resolved image. In Section 2.1 we introduce the model of the acquisition process as well as its reformulation in the covariance domain. The proposed approach is divided into two steps which we present in Section 2.2 and 2.3, respectively. In Section 2.4 we report the results of COL0RME, as well as of other approaches, when applied to simulated and real data, and finally, in Section 2.5 we discuss the performance of COL0RME allowing some possible improvements which will be discussed in the following chapters.*

---

**Related Publications:**

[Stergiopoulou et al., 2021]
V. Stergiopoulou, J. H. de Morais Goulart, S. Schaub, L. Calatroni and L. Blanc-Féraud, "COL0RME: Covariance-Based l0 Super-Resolution Microscopy with Intensity Estimation," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 349-352, doi: 10.1109/ISBI48211.2021.9433976.

[Stergiopoulou et al., 2022a]
V. Stergiopoulou, L. Calatroni, J. H. de Morais Goulart, S. Schaub, and L. Blanc-Féraud, "COL0RME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation," Biological Imaging, vol. 2, p. e1, 2022, doi: 10.1017/S2633903X22000010.

## 2.1 Mathematical Modeling and Problem Formulation

In Section 1.4, the general imaging model for fluorescence microscopy was introduced. In this section, recalling (1.28), we limit ourselves to the case of 2-dimensional (2D) images. In order to benefit from the independent stochastic fluctuations of distinct emitters appearing at different frames, we further require the acquisition of more than one frame ($T > 1$). Moreover, we introduce the reformulated model in the covariance domain which helps us to exploit suitable independence assumptions between emitters.

### 2.1.1 Model formulation in the image domain

For natural scalars $T, M \in \mathbb{N}$ and $t \in \{1, 2, \dots, T\}$, let $\mathbf{Y}_t \in \mathbb{R}^{M\times M}$ denote the blurred, noisy and down-sampled image frame acquired at time $t$. We look for a high-resolution image $\mathbf{X} \in \mathbb{R}^{L\times L}$ being defined as $\mathbf{X} := \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t$. It belongs on a $q$-times finer grid, with $q \in \mathbb{N}$ and $L = qM$. The image formation model describing the acquisition process at each $t$ can be written as:

$$\mathbf{Y}_t = \mathcal{P}\left(M_q\left(U\left(\mathbf{X}_t\right)\right) + \mathbf{B}_t\right) + \mathbf{E}_t, \qquad \forall t = 1, 2, \dots, T, \tag{2.1}$$

where,

— $U : \mathbb{R}^{L\times L} \to \mathbb{R}^{L\times L}$ is the convolution operator defined by the PSF of the optical imaging system. The PSF can be approximated by a Gaussian kernel $\mathbf{u} \in \mathbb{R}^{K\times K}$, where $K \times K$ denotes the support of $\mathbf{u}$, with standard deviation equal to $\mathrm{FWHM}/2\sqrt{2\ln 2}$ (see formula (1.1) for the theoretical value of the FWHM of the PSF). The convolution operator simply performs a 2D convolution of $\mathbf{x}_t$ with the 2D Gaussian kernel.

— $M_q : \mathbb{R}^{L\times L} \to \mathbb{R}^{M\times M}$ is a down-sampling operator used in (2.1) since the reconstructed and the measurement data are in two different grids. We refer to these grids as *fine grid* (with size $L \times L$) and *coarse grid* (with size $M \times M$), respectively. The factor $q$ connecting the two grids, with $L = qM$, is called the *super-resolution factor*. In this chapter, we typically set $q = 4$. The down-sampling operator maps an image from the fine to the coarse grid, by summing every $q$ consecutive pixels in both dimensions. Figure 2.1 illustrates the under-sampling: a pixel in the coarse grid (colored red) is computed after summing the $q \times q$ pixels of the fine grid (colored blue).

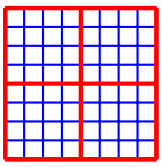


Figure 2.1 – An example of the fine grid (blue) and the coarse grid (red) with $q = 4$.

— $\mathbf{B}_t \in \mathbb{R}^{M\times M}$ is a matrix used to model background contributions, i.e. the image containing out-of-focus fluorescent molecules. Motivated by experimental observations showing that the blinking/fluctuating behavior of such molecules is not visible after convolution with wide defocused PSFs, we assume that the background is temporally constant $\mathbf{B}$ (i.e., does not depend on $t$), while it smoothly varies in space.

— $\mathbf{E}_t \in \mathbb{R}^{M\times M}$ is a matrix describing the presence of signal-independent readout noise. Such noise is generated by the electronics of the camera during the readout process and more

precisely by the conversion of charge carriers (i.e. photons) into a quantified voltage signal and the subsequent analog-to-digital (A/D) conversion, amplification, and processing. The readout noise is added uniformly to every image pixel and therefore it can be modeled as a matrix of independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance $\sigma^2 \in \mathbb{R}^+$.

— $\mathcal{P}(\mathbf{W})$, with $\mathbf{W} = M_q\left(U\left(\mathbf{X}_t\right)\right) + \mathbf{B}_t \in \mathbb{R}_+^{M\times M}$ models the presence of signal-dependent photon noise. The photon noise is a basic form of uncertainty associated with the detection of photons by a photoelectric sensor (e.g. CCD camera), inherent to the particle nature of light and the independence of photon detections. The intensity detected by the camera can thus be modeled by a Poisson distribution parameterized by the exact photon count $\mathbf{W}$. Mathematically, $\mathcal{P}(\mathbf{W})$ thus represents the random matrix drawn from a multidimensional Poisson distribution with parameter $\mathbf{W}$.

We assume that there is no displacement of the specimen during the imaging period, which is a reasonable assumption whenever short acquisition times are considered.

Taking into account that the background is temporally constant, we can write the following:

$$\mathbf{Y}_t = \mathcal{P}\left(M_q\left(U\left(\mathbf{X}_t\right)\right) + \mathbf{B}\right) + \mathbf{E}_t = \mathcal{P}\left(M_q\left(U\left(\mathbf{X}_t\right)\right)\right) + \mathcal{P}\left(\mathbf{B}\right) + \mathbf{E}_t, \quad \forall t = 1, 2, \ldots, T, \tag{2.2}$$

where the second equality holds due to the independence between $M_q\left(U\left(\mathbf{X}_t\right)\right)$ and $\mathbf{B}$. To further simplify the model we assume that $\mathbf{B}$ has sufficiently large entries, a reasonable assumption in real acquisitions, so that $\mathcal{P}(\mathbf{B})$ can be approximated as $\mathcal{P}(\mathbf{B}) \approx \hat{\mathbf{B}}$ with each element of $\hat{\mathbf{B}}$ following a normal distribution or $\hat{B}_{ij} \sim \mathcal{N}(B_{ij}, B_{ij})$ , where $(i, j) \in \{1, \ldots, M\}^2$. We thus consider:

$$\mathbf{Y}_t = \mathcal{P}\left(M_q\left(U\left(\mathbf{X}_t\right)\right)\right) + \hat{\mathbf{B}} + \mathbf{E}_t, \qquad \forall t = 1, 2, \ldots, T. \tag{2.3}$$

We now further approximate the variance of each element of $\hat{\mathbf{B}}$ with a constant $b \in \mathbb{R}_+$ to be interpreted as the average of $\mathbf{B}$, and introduce a random term $\mathbf{N}_t$ following a multivariate normal distribution, with the mean of every entry being 0 and the variance $s$. The variance $s$ is given by the following formula $s = \sigma^2 + b$ which takes into account both the underlying electronic noise and the noise bias induced by $\mathbf{B}$. With simple manipulations, we find:

$$\hat{\mathbf{B}} + \mathbf{E}_t = \mathbf{B} + \mathbf{N}_t, \tag{2.4}$$

where the independence between $\hat{\mathbf{B}}$ and $\mathbf{E}_t$ has been exploited.

By further neglecting the Poisson noise dependence in the term $\mathcal{P}\left(M_q\left(U\left(\mathbf{X}_t\right)\right)\right)$ we can thus retrieve the simplified model:

$$\mathbf{Y}_t = M_q\left(U\left(\mathbf{X}_t\right)\right) + \mathbf{B} + \mathbf{N_t}, \qquad \forall t = 1, 2, \ldots, T, \tag{2.5}$$

In vectorized form, model (2.5) reads:

$$\mathbf{y}_t = \mathbf{\Psi}\mathbf{x}_t + \mathbf{b} + \mathbf{n}_t, \tag{2.6}$$

where $\mathbf{\Psi} \in \mathbb{R}^{M^2\times L^2}$ is the (known) matrix representing the composition $M_q \circ U$, while $\mathbf{y}_t \in \mathbb{R}^{M^2}$, $\mathbf{x}_t \in \mathbb{R}^{L^2}$, $\mathbf{b} \in \mathbb{R}^{M^2}$ and $\mathbf{n}_t \in \mathbb{R}^{M^2}$ are the column-wise vectorizations of $\mathbf{Y}_t$, $\mathbf{X}_t$, $\mathbf{B}$ and $\mathbf{N}_t$ in (2.5), respectively.

### 2.1.2 Model formulation in the covariance domain

In order to exploit the uncorrelation over time of distinct fluorophore emissions, we reformulate the model in the covariance domain. This idea was previously exploited by the SOFI approach [Dertinger et al., 2009] and was shown to reduce the FWHM of the PSF, therefore enhancing the resolution. In particular, the use of second-order statistics for a Gaussian PSF corresponds to a reduction factor of the FWHM of $\sqrt{2}$. This is illustrated in Remark 2.1.1. When $n$-order statistics are used, the reduction factor will be equal to $\sqrt{n}$ (see [Dertinger et al., 2009] for more details). In subsequent work, by efficiently re-weighting the Fourier transform of the PSF, the authors result in a reduction factor close to $n$ [Dertinger et al., 2010].

*Remark 2.1.1 –* We start from model (1.26) defined in the continuous setting describing an image distorted only due to diffraction to prove that a reduction of the FWHM of a factor equal to $\sqrt{2}$ occurs whenever second-order statistics and a Gaussian PSF are used.

We recall that the 2D image at position $\mathbf{r} = [r_x, r_y] \in \mathbb{R}^2$ and time $t > 0$ is given by:

$$y(\mathbf{r}, t) = \sum_{k=1}^{N} v(\mathbf{r} - \mathbf{r}_k) s_k(t), \tag{2.7}$$

where $v$ is the PSF, $\mathbf{r}_k$ the position of the emitter $k$ and $s_k(t)$ its intensity over time.

Following [Dertinger et al., 2009], the second-order auto-correlation function $\tilde{R}_y(\mathbf{r}, \varrho)$ of $y(\mathbf{r}, t)$ with $t = 1, \ldots, T$, for a time-lag $\varrho \in \mathbb{N}$, is given by:

$$\begin{aligned}
\tilde{R}_y(\mathbf{r}, \varrho) &= \mathbb{E}_y\{(y(\mathbf{r}, t+\varrho) - \mathbb{E}_y\{y(\mathbf{r}, t+\varrho)\}) \cdot (y(\mathbf{r}, t) - \mathbb{E}_y\{y(\mathbf{r}, t)\})\} \\
&= \sum_{j,k} v(\mathbf{r} - \mathbf{r}_j) v(\mathbf{r} - \mathbf{r}_k)\, \mathbb{E}_s\{(s_j(t+\varrho) - \mathbb{E}_s\{s_j(t+\varrho)\}) \cdot (s_k(t) - \mathbb{E}_s\{s_k(t)\})\} \\
&= \sum_{k} v^2(\mathbf{r} - \mathbf{r}_k)\, \mathbb{E}_s\{(s_k(t+\varrho) - \mathbb{E}_s\{s_k(t+\varrho)\}) \cdot (s_k(t) - \mathbb{E}_s\{s_k(t)\})\},
\end{aligned} \tag{2.8}$$

where $\mathbb{E}$ denotes the expected value and $\cdot$ denotes element-wise multiplication. In the previous equation, it is assumed that the different emitters are uncorrelated in time and therefore all the cross-correlation terms, $\mathbb{E}_s\{(s_j(t+\varrho) - \mathbb{E}_s\{s_j(t+\varrho)\}) \cdot (s_k(t) - \mathbb{E}_s\{s_k(t)\})\}$ for $k \neq j$, vanish.

The second-order auto-correlation function $\tilde{R}_y(\mathbf{r}, \varrho)$ is given therefore from the convolution of the emitters auto-correlation function with a distribution that is the square of the original PSF $v$. When the PSF is approximated by a Gaussian function, we thus observe a width reduction of a factor $\sqrt{2}$ in both dimensions, as:

$$v^2(\mathbf{r}) = \left(\exp\left(-\frac{r_x^2 + r_y^2}{2\eta^2}\right)\right)^2 = \exp\left(-\frac{r_x^2 + r_y^2}{2(\eta/\sqrt{2})^2}\right). \tag{2.9}$$

The standard deviation $\eta$ of the original PSF is thus reduced by a factor $\sqrt{2}$, as it is for its FWHM, because the relationship between the two is linear.

The main disadvantage of SOFI is the non-linear dependence of the resulted image contrast on emitter brightness. Not only the PSF $u(\mathbf{r})$ is raised to the power of $n$, for cumulants of order $n$, but also the molecular brightness. This results to a SOFI image with very high dynamic range, limiting the use of higher cumulant orders for improving the resolution.

To overcome this problem, the method SPARCOM [Solomon et al., 2018, Solomon et al., 2019] restricts the computation of only order-2 statistics and tries to gain resolution from other sources. First of all, for a time-lag $\varrho$, the whole temporal covariance matrix $\mathbf{R}_{\mathbf{y}}(\varrho)$ is considered. In comparison, auto-correlation or cross-correlation SOFI [Dertinger et al., 2009, Dertinger et al., 2010], consider only a part of this matrix (e.g. the main diagonal for auto-correlation SOFI), without exploiting all the available information. Moreover, SPARCOM exploits the sparse distribution of the fluorescent emitters. This will be discussed, with more details, in the next section as it does not directly affect the observation model but only gives prior knowledge of the solution we are looking for.

Similarly to SPARCOM, in our model formulation we will use the whole temporal covariance matrix and set the time-lag $\varrho$ equal to 0. With the previous notation we are referring to the matrix $\mathbf{R}_{\mathbf{y}}(0)$, but we will use $\mathbf{R}_{\mathbf{y}}$ for short. To formulate the model, we consider the discrete frames $\{\mathbf{y}_t\}_{t=1}^{T}$ as $T$ realizations of a random variable $\mathbf{y}$ with covariance matrix defined by:

$$\mathbf{R}_{\mathbf{y}} = \mathbb{E}_{\mathbf{y}}\{(\mathbf{y} - \mathbb{E}_{\mathbf{y}}\{\mathbf{y}\})(\mathbf{y} - \mathbb{E}_{\mathbf{y}}\{\mathbf{y}\})^{\mathsf{T}}\}, \tag{2.10}$$

where $\mathbb{E}_{\mathbf{y}}\{\cdot\}$ denotes the expected value computed w.r.t. to the unknown law of $\mathbf{y}$. We estimate $\mathbf{R}_{\mathbf{y}}$ by computing the empirical covariance matrix, i.e.:

$$\mathbf{R}_{\mathbf{y}} \approx \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{y}_t - \overline{\mathbf{y}})(\mathbf{y}_t - \overline{\mathbf{y}})^{\mathsf{T}},$$

where $\overline{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t$ denotes the empirical temporal mean. From (2.6) and (2.10), we thus deduce the relation:

$$\mathbf{R}_{\mathbf{y}} = \mathbf{\Psi}\mathbf{R}_{\mathbf{x}}\mathbf{\Psi}^{\mathsf{T}} + \mathbf{R}_{\mathbf{n}}, \tag{2.11}$$

where $\mathbf{R}_{\mathbf{x}} \in \mathbb{R}^{L^2\times L^2}$ and $\mathbf{R}_{\mathbf{n}} \in \mathbb{R}^{M^2\times M^2}$ are the covariance matrices, with $\varrho = 0$, of $\{\mathbf{x}_t\}_{t=1}^{T}$ and $\{\mathbf{n}_t\}_{t=1}^{T}$, respectively. As the background $\mathbf{b}$ is stationary by assumption, its covariance matrix is zero.

Recalling now that the emitters are uncorrelated, we deduce that $\mathbf{R}_{\mathbf{x}}$ is diagonal, as all the cross-correlations are equal to 0. We thus set $\mathbf{r}_{\mathbf{x}} := \mathrm{diag}(\mathbf{R}_{\mathbf{x}}) \in \mathbb{R}^{L^2}$. Furthermore, by the i.i.d. assumption on $\mathbf{n}_t$, we have that $\mathbf{R}_{\mathbf{n}} = s\mathbf{I}_{\mathbf{M^2}}$, where $s \in \mathbb{R}_+$ and $\mathbf{I}_{\mathbf{M^2}}$ is the identity matrix in $\mathbb{R}^{M^2\times M^2}$.

Note that the model in equation (2.11) is different from the one used in SPARCOM and presented in [Solomon et al., 2018, Solomon et al., 2019], as noise contributions are also taken into account by including in the model the diagonal covariance matrix $\mathbf{R}_{\mathbf{n}}$. Finally, the vectorized form of the model in the covariance domain can thus be written as:

$$\mathbf{r}_{\mathbf{y}} = (\mathbf{\Psi} \odot \mathbf{\Psi})\mathbf{r}_{\mathbf{x}} + s\mathbf{v}_{\mathbf{I}}, \tag{2.12}$$

where $\odot$ denotes the Khatri–Rao (column-wise Kronecker) product, $\mathbf{r}_{\mathbf{y}} \in \mathbb{R}^{M^4}$ is the column-wise vectorization of $\mathbf{R}_{\mathbf{y}}$ and $\mathbf{v}_{\mathbf{I}} = \mathrm{vec}(\mathbf{I}_{\mathbf{M^2}})$.

### 2.1.3 COL0RME: A two-step approach

We now introduce the method COL0RME, the abbreviation for *COvariance-based $\ell_0$ super-Resolution Microscopy with intensity Estimation*. The inspiration for such an acronym comes from the fact that we do not only find where the fluorescent molecules are, but we also calculate, for

each molecule position, the intensity values (a quantity proportional to the number of photons emitted by the molecules), so we "color" them.

To be more precise, COL0RME solves two minimization problems with the overall goal to find a super-resolved 2D image $\mathbf{x}$, belonging to a finer grid than the acquisitions grid. We are firstly interested in precisely localizing the fluorescent molecules by finding the positions of the pixels in the fine grid that have at least one fluorescent molecule. This is equivalent to finding the support $\Omega$ of the image $\mathbf{x}$, defined as the positions of the non-zero elements, i.e. $\Omega := \{i : x_i \neq 0\} \subset \{1, \dots, L^2\}$. The support of the final image $\mathbf{x}$ is exactly the same as the support of the temporal auto-covariance matrix of the emitters $\mathbf{r_x}$, after the reasonable assumption that there is no displacement of the specimen during the imaging period whenever short time acquisitions are considered. Therefore we can write: $\Omega := \{i : (\mathbf{r_x})_i \neq 0\}$ and seek it in the first minimization problem. As discussed before, looking at second-order statistics, we expect an improvement of a factor equal to $\sqrt{2}$. As a second step, a second minimization problem is solved, using the original fluorescence microscopy model (2.6) and first-order statistics (mean over time). The goal of this step is to find the real intensity values only for the pixels belonging to the already estimated support $\Omega$. Mathematically, we can write $\mathbf{x} \in \mathbb{R}^{|\Omega|}$, while with $|\Omega|$ we denote the cardinality of the set $\Omega$.

For all $t$ and given $\mathbf{\Psi}$, $\mathbf{y}_t$ and therefore $\mathbf{r_y}$, the two problems can thus be formulated as:

$$\text{find} \quad \Omega := \{i : (\mathbf{r_x})_i \neq 0\}, s \in \mathbb{R} \quad \text{s.t.} \quad \mathbf{r_x}, s \quad \text{solves (2.12)}, \tag{COL0RME-1}$$

$$\text{find} \quad \mathbf{x} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \in \mathbb{R}^{|\Omega|} \text{ and } \mathbf{b} \in \mathbb{R}^{M^2} \quad \text{s.t.} \quad \mathbf{x}, \mathbf{b} \quad \text{solves time-averaged (2.6)}. \tag{COL0RME-2}$$

We solve the two problems sequentially because, in the second problem, we use the solution of the first one. We refer to these two problems as support and intensity estimation steps. In Figure 2.2, we report an overview of both steps by visualizing the inputs/outputs of each, as well as the interaction between them. On the right side of the figure, we can see the two main outputs of COL0RME: the support $\Omega$ containing the locations of the fine-grid pixels with at least one fluorescent molecule, and the intensity $\mathbf{x} \in \mathbb{R}^{L^2}$ whose non-null values are estimated only on $\Omega$. More details for these steps are given in Section 2.2 and 2.3, respectively.

## 2.2 Support Estimation

In this section, we aim at solving the problem COL0RME-1. Similarly to SPARCOM [Solomon et al., 2019], our approach makes use of the fact that the vector $\mathbf{r_x}$ is sparse. In our work, we further include the estimation of the variance $s = \sigma^2 + b > 0$ for dealing with more challenging scenarios including noise.

### 2.2.1 Formulation of the optimization problem

In order to compare different *a-priori* regularity constraints on the solution, we make use of different regularization terms, whose importance is controlled by a regularization hyperparameter

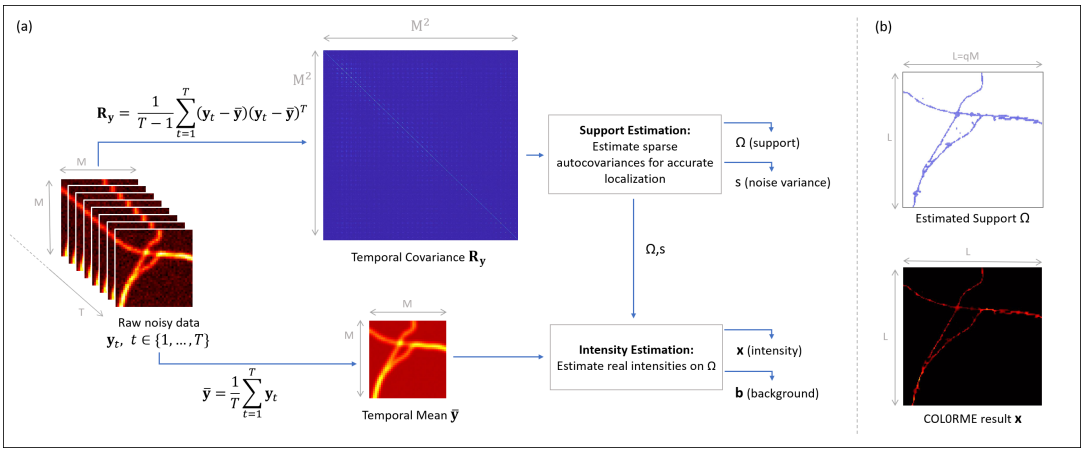



Figure 2.2 – COL0RME pipeline. (a) An overview of the two steps, Support Estimation and Intensity Estimation, (b) The two main outputs of COL0RME: the binary image identifying the support $\Omega$ and the intensity image $\mathbf{x}$.

$\lambda > 0$ to be chosen appropriately. By further introducing a non-negativity constraint for both variables $\mathbf{r_x}$ and $s$, we thus aim to solve:

$$\underset{\mathbf{r_x}\geq 0,\ s\geq 0}{\arg\min}\ \mathcal{F}(\mathbf{r_x}, s) + \mathcal{R}(\mathbf{r_x}; \lambda), \tag{2.13}$$

where the data fidelity term is defined from (2.12) as:

$$\mathcal{F}(\mathbf{r_x}, s) = \frac{1}{2}\|\mathbf{r_y} - (\mathbf{\Psi} \odot \mathbf{\Psi})\mathbf{r_x} - s\mathbf{v_I}\|_2^2, \tag{2.14}$$

and $\mathcal{R}(\cdot; \lambda)$ is a sparsity-promoting penalty. As discussed in Section 1.3.3.3, ideally, one would like to make use of the $\ell_0$-norm to enforce sparsity. However solving the resulting non-continuous, non-convex and combinatorial minimization problem is a NP-hard optimization problem. We bypass this problem by using the continuous exact relaxation of the $\ell_0$-norm (CEL0) proposed by Soubies *et al.* in [Soubies et al., 2015] in a tight way of promoting sparsity without modifying the problem too much. The CEL0 regularization is continuous, non-convex and non-differentiable. It preserves the global minima of the ideal $\ell_2 - \ell_0$ problem one would like ideally to solve and it may remove some local ones. The CEL0 penalty is defined as:

$$\mathcal{R}(\mathbf{r_x}; \lambda) = \Phi_{\text{CEL0}}(\mathbf{r_x}; \lambda) = \sum_{i=1}^{L^2} \lambda - \frac{\|\mathbf{A}_i\|^2}{2}\left(|(\mathbf{r_x})_i| - \frac{\sqrt{2\lambda}}{\|\mathbf{A}_i\|}\right)^2 \mathbb{1}_{\{|(\mathbf{r_x})_i| \leq \frac{\sqrt{2\lambda}}{\|\mathbf{A}_i\|}\}}, \tag{2.15}$$

where $\mathbf{A}_i = (\mathbf{\Psi} \odot \mathbf{\Psi})_i$ denotes the $i$-th column of the matrix $\mathbf{A} := \mathbf{\Psi} \odot \mathbf{\Psi}$.

A different, convex way of favoring sparsity consists in taking as regularizer the $\ell_1$-norm:

$$\mathcal{R}(\mathbf{r_x}; \lambda) = \lambda\|\mathbf{r_x}\|_1. \tag{2.16}$$

This way of promoting sparsity was previously used in SPARCOM [Solomon et al., 2019].

Finally, in order to model situations where extended piece-wise constant structures are considered, we consider a different regularization term favoring gradient-sparsity by means of the Total

Variation (TV) regularization defined in a discrete setting as follows:

$$\mathcal{R}(\mathbf{r_x};\lambda) = \lambda TV(\mathbf{r_x}) = \lambda \sum_{i=1}^{L^2} \left( |(\mathbf{r_x})_i - (\mathbf{r_x})_{n_{i,1}}|^2 + |(\mathbf{r_x})_i - (\mathbf{r_x})_{n_{i,2}}|^2 \right)^{\frac{1}{2}}, \tag{2.17}$$

where $(n_{i,1}, n_{i,2}) \in \{1, \ldots, L^2\}^2$ indicate the locations of the horizontal and vertical nearest neighbor pixels of pixel $i$, as shown in Figure 2.3. For the computation of the TV penalty, Neumann boundary conditions have been used.

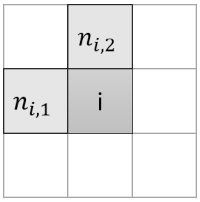



Figure 2.3 – The one-sided nearest horizontal and vertical neighbors of the pixel $i$ used to compute the gradient discretization in (2.17).

To solve (2.13) we use the Alternate Minimization algorithm between $s$ and $\mathbf{r_x}$ [Attouch et al., 2010]. The pseudo-code is reported in Algorithm 1. Note that, at each $k \geq 1$, the update for the variable $s$ is efficiently computed through the following explicit expression:

$$s^{k+1} = \frac{1}{M^2} \mathbf{v_I}^\intercal (\mathbf{r_y} - (\mathbf{\Psi} \odot \mathbf{\Psi}) {\mathbf{r_x}}^k).$$

The result is then projected onto the positive subspace to ensure that the update for the variable $s$ is non-negative.

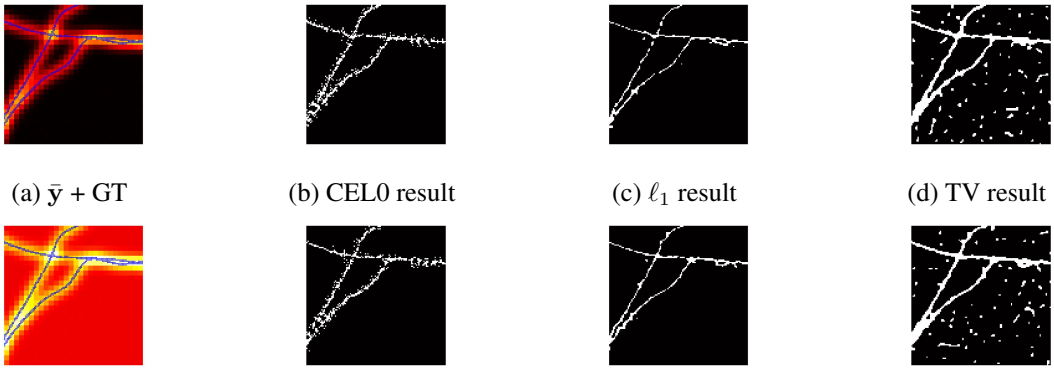
Concerning the update of $\mathbf{r_x}$, different algorithms are used depending on the choice of the regularization term. For the CEL0 penalty (2.15) we used the Iteratively Reweighted $\ell_1$ algorithm (IRL1) [Ochs et al., 2015], following Gazagnes et al. [Gazagnes et al., 2017] with Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [Beck and Teboulle, 2009] as inner solver. If the $\ell_1$ norm (2.16) is chosen, FISTA is used. Finally, when the TV penalty (2.17) is employed, the Primal-Dual Splitting Method in [Condat, 2013] was considered.

**Algorithm 1** COL0RME, Step I: Support Estimation

**Require:** $\mathbf{r_y} \in \mathbb{R}^{M^4}, {\mathbf{r_x}}^0 \in \mathbb{R}^{L^2}, \lambda > 0$
  **repeat**
    $s^{k+1} = \underset{s \in \mathbb{R}_+}{\arg\min}\, \mathcal{F}({\mathbf{r_x}}^k, s)$
    ${\mathbf{r_x}}^{k+1} = \underset{\mathbf{r_x} \in \mathbb{R}_+^{L^2}}{\arg\min}\, \mathcal{F}(\mathbf{r_x}, s^{k+1}) + \mathcal{R}(\mathbf{r_x};\lambda)$
  **until** convergence
  **return** $\Omega_{\mathbf{x}}, s$

Following the description provided by Attouch *et al.* in [Attouch et al., 2010], the convergence of Algorithm 1 can be guaranteed if an additional quadratic term is introduced in the objective function of the second minimization sub-problem, as the first one is exactly solved at each iteration. Nonetheless, empirical convergence was observed also without such additional terms.

To evaluate the performance of the first step of COL0RME using the different regularization penalties described above, we created two noisy simulated datasets, with low background (LB) and high background (HB) and used them to test the first step of COL0RME. More details on the two datasets are available in the following Section 2.4.1. The results obtained by using the three different regularizers, namely CEL0, $\ell_1$-norm and TV penalty, are reported in Figure 2.4. On the left part of the figure, we show the superimposed diffraction-limited image (temporal mean of the stack) with 4-times zoom on ground truth (GT) support (with blue). In this example, we chose the regularization parameter $\lambda$ heuristically. More details on the selection of such a parameter are given in Section 2.2.3.



(a) $\bar{\mathbf{y}}$ + GT (b) CEL0 result (c) $\ell_1$ result (d) TV result

(e) $\bar{\mathbf{y}}$ + GT (f) CEL0 result (g) $\ell_1$ result (h) TV result

Figure 2.4 – (a) GT and diffraction limited image for the noisy simulated LB dataset with stack size of T=500 frames, (b-d) Results with different regularization penalties for the LB dataset, (e) GT and diffraction limited image for the noisy simulated HB dataset with T=500 frames (f-h) Results with different regularization penalties for the HB dataset.

We observe that the reconstruction obtained by the TV regularizer does not provide precise localization results. To give an example, the separation of the two filaments on the top-right corner is not visible and while the junction of the other two filaments on the bottom-left should appear further down, we see that those filaments are erroneously glued together. We thus will not consider TV further in our tests and rather focus on the comparison between CEL0 and $\ell_1$ only.

The Jaccard indices of both results obtained when using CEL0 and $\ell_1$, have been computed. The Jaccard Index (JI), is a quantity in the range $[0, 1]$ computed as the ratio between correct detections (CD) and the sum of correct detections, false negatives (FN), and false positives (FP), that is JI $:= CD/(CD + FN + FP)$, up to a tolerance $\delta > 0$, measured in nm (see [Sage et al., 2015]). A correct detection occurs when one pixel at most $\delta$ nm away from a ground truth pixel is added to the support. In order to match the pixels from the estimated support to the ones from the ground truth, we employ the standard Gale–Shapley algorithm [Gale and Shapley, 1962]. Once the matching has been performed, we can simply count the number of ground truth pixels that have not been detected (false negative) and the number of pixels in the estimated support which have not been matched to any ground truth pixel (false positive). A small example of the way to count CD, FP, and FN is reported in Figure 2.5.

Figure 2.6 reports the average JI computed from 20 different noise realizations, as well as an error bar (vertical lines) representing the standard deviation, for several stack sizes. A tolerance

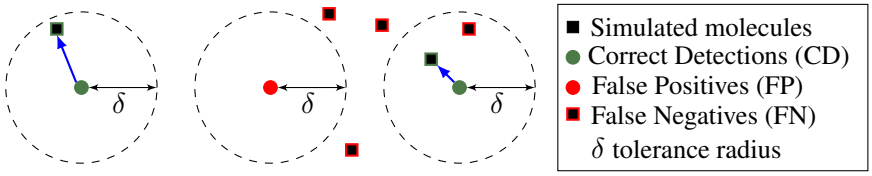
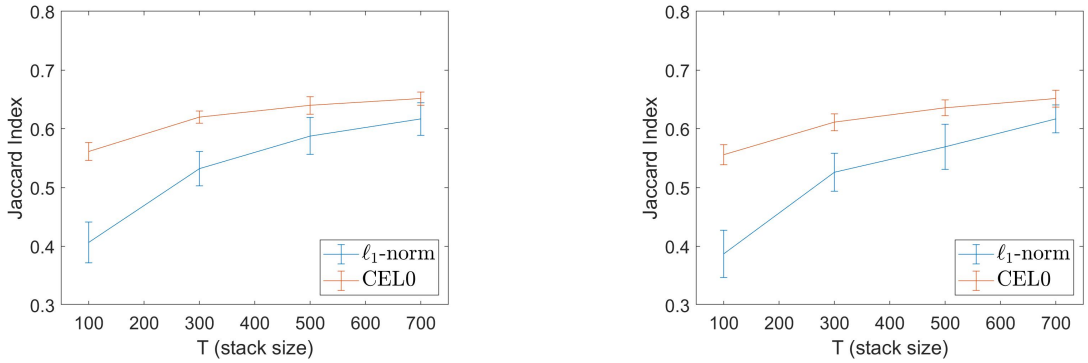



Figure 2.5 – Jaccard Index metric for accessing localizations

precision $\delta = 40$ nm is set so that we allow the CD, that needed to be counted for the computation of the JI, to be found not only in the same pixel but also in any of the 8-neighboring pixels, since the pixel size 25 nm. According to the figure, a slightly better JI is obtained when CEL0 is used. In addition, for both regularizers, an increase in the number of frames leads to better JI, hence better localization.

As the reader may notice, such quantitative assessment could look inconsistent with the visual results reported in Figure 2.4. Indeed, by definition, the JI tends to assume higher values whenever more CD are found even in presence of more FP (as it happens for the CEL0 reconstruction), while it gets more penalized when FN occur, as they affect the computation "twice", reducing the numerator and increasing the denominator.




(a) LB dataset

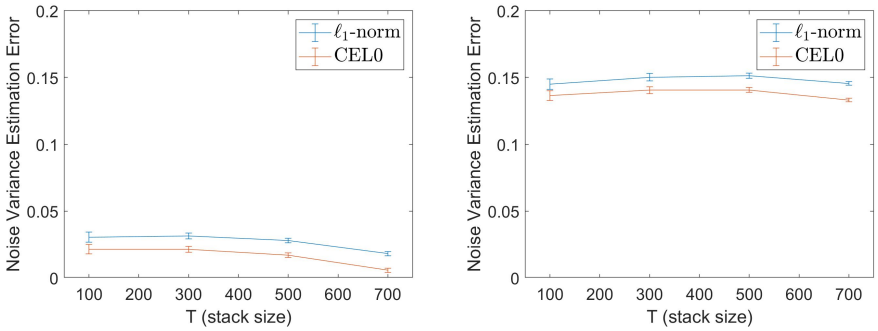
(b) HB dataset

Figure 2.6 – Jaccard Index values with tolerance $\delta = 40nm$ for the LB and HB dataset, for different stack sizes and regularization terms. The mean and the standard deviation (error bars) for 20 different noise realization are presented.

### 2.2.2 Noise variance estimation

Along with the estimations of the emitter covariance matrix, the estimation of the noise variance in the joint model (2.13) allows for much more precise results even in challenging acquisition conditions featuring noise. In Figure 2.7 we show the relative error, defined as: $\texttt{error} = |s - \sigma^2_{\text{GT}}|/|\sigma^2_{\text{GT}}|$, between the computed noise variance $s = \sigma^2 + b$ and the true variance of the electronic noise $\sigma^2_{\text{GT}}$ used to produce simulated low-background (LB) and high-background (HB) data.

The relative error is higher in the case of the HB dataset, which is expected, as in our noise variance estimation $s$ there is a bias coming from the background $b$. In the case of the LB dataset,




(a) LB dataset (b) HB dataset

Figure 2.7 – The relative error committed in noise variance estimation, for the LB and HB dataset and for different stack sizes and regularization penalty choices. The mean and the standard deviation (error bars) of 20 different noise realization are presented.

as the background is low, the bias is sufficiently small so that it is barely visible in the error graph. In our experiments, Gaussian noise with a corresponding SNR of approximately 16 dB is being used, while the value of $\sigma^2_{\text{GT}}$ is on average equal to $7.11 \times 10^5$ for the LB dataset and $7.13 \times 10^5$ for the HB dataset. Note that, in general, the estimation of the noise variance $s$ obtained by COL0RME is quite precise.

### 2.2.3 Hyper-parameter Selection

The selection of the regularization parameter value $\lambda$ in (2.13) is critical, as it determines the sparsity level of the support of the emitters. For its estimation, we start computing reference values $\lambda_{max}$, defined as the smallest regularization parameters for which the identically zero solution is found. It is indeed possible to compute such a $\lambda_{max}$ from the data and the model operators for both regularization terms CEL0 and $\ell_1$ (see [Soubies, 2016] and [Koulouri et al., 2021]). Once such values are known what we actually need to compute is a fraction $\gamma \in (0, 1)$ of $\lambda_{max}$ corresponding to the choice $\lambda = \gamma\lambda_{max}$. For the CEL0 regularizer the expression for $\lambda_{max}$ (see Proposition 10.9 in [Soubies, 2016]) is:

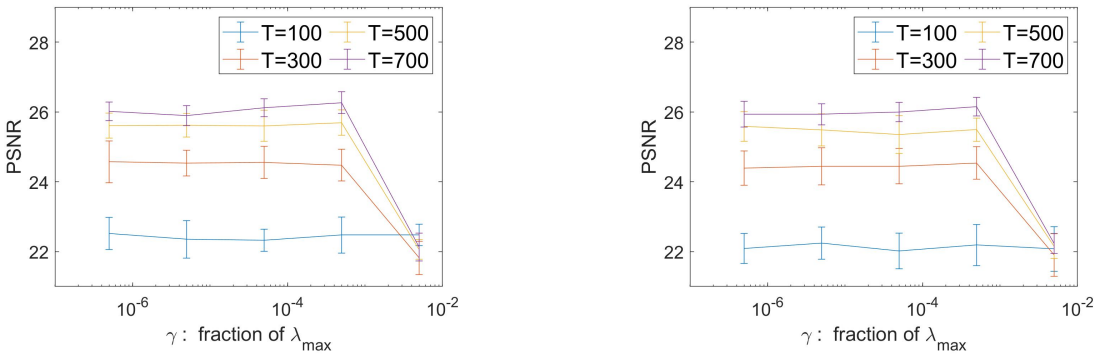
$$\lambda^{CEL0}_{max} := \max_{1\leq i\leq L^2} \frac{\langle \mathbf{A}_i, \mathbf{r_y}\rangle^2}{2\|\mathbf{A}_i\|^2}, \tag{2.18}$$

Regarding the $\ell_1$-norm regularization penalty, $\lambda_{max}$ is given by:

$$\lambda^{\ell_1}_{max} := \|\mathbf{A}^\intercal \mathbf{r_y}\|_\infty = \max_{1\leq i\leq L^2} \langle \mathbf{A}_i, \mathbf{r_y}\rangle. \tag{2.19}$$

As far as the $\ell_1$-norm is used as the regularization term in (2.13), we report in Figure 2.8 a graph showing how the Peak Signal-to-Noise Ratio (PSNR) value of the final estimated intensity image (i.e. after the application of the second step of COL0RME discussed in following Section 2.3) varies for the two datasets considered depending on $\lambda$. It can be observed that for a large range of values $\lambda$, the final PSNR remains almost the same. Although this may look a bit surprising at a first sight, we remark that the robustness of the result is due, essentially, to the second step of the algorithm where false localizations related to an underestimation of $\lambda$ can be corrected through the intensity estimation step. Note, however, that in the case of an overestimation of $\lambda$, points

contained in the original support are definitively lost so no benefit is obtained from the intensity estimation step, hence the overall PSNR decreases.




(a) LB dataset

(b) HB dataset

Figure 2.8 – The PSNR values of the final COL0RME image obtained using the $\ell_1$-norm regularizer for support estimation, at various $\gamma$ values, evaluated on both the LB and HB datasets. The mean and the standard deviation (error bars) of 20 different noise realizations are presented.

When the CEL0 penalty is used for support estimation, a heuristic parameter selection strategy can be used to improve the localization results and avoid fine parameter tuning. More specifically, the non-convexity of the model can be 'leveraged' by considering an algorithmic restarting approach to improve the support reconstruction quality. An initial large value of $\lambda$ can be fixed, typically $\lambda = \gamma\lambda_{max}^{CEL0}$ with $\gamma \approx 5 \times 10^{-4}$, so as to achieve a very sparse reconstruction. Then, the support estimation Algorithm 1 can be run and iteratively repeated with a new initialization $\mathbf{r_x}^0$ (that is, restarted) several times. While keeping $\lambda$ fixed along this procedure, a wise choice of the initialization $\mathbf{r_x}^0$ depending, but not being equal, to the previous output can be used to enrich the support.

There are many ways to choose the new initialization since both deterministic and stochastic strategies can be used. We chose a deterministic way based on the following idea: for every pixel belonging to the solution of the previous restarting, we find its closest neighbor. Then, we define the middle point between the two and we include it in the initialization of the current restarting. A small example is given in the Figure 2.9. The yellow points belong to the support estimation of the previous restarting. Starting from them we define the red points, used for the initialization of the current restarting.

Non-convexity is here exploited by changing, for a fixed $\lambda$, the initialization at each algorithmic restart, so that new local minimizers (corresponding to possible support points) are computed. The final support image can thus be computed as the superposition of the different solutions computed at each restarting. In such a way, a good result for a not-finely-tuned value of $\lambda$ is computed. Clearly the price to pay for such procedure is an increased computational cost since the optimization algorithm has to be run several times.

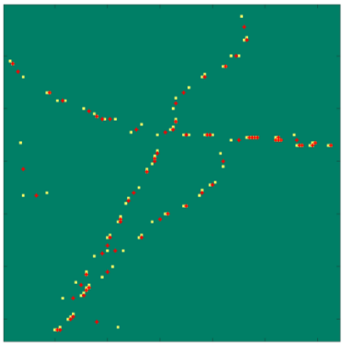


Figure 2.9 – The yellow pixels belong to the support estimated in the previous restarting, while the red pixels belong to the initialization that is used in the current restarting.

## 2.3 Intensity Estimation

From the first step of COL0RME described in Section 2.2, we obtain a sparse estimation of $\mathbf{r_x} \in \mathbb{R}^{L^2}$. Its support $\Omega := \{i : (\mathbf{r_x})_i \neq 0\}$ is thus defined. We are now interested in solving the problem (COL0RME-2) to retrieve the intensity information of the signal $\mathbf{x}$ in correspondence with the estimated support $\Omega$.

### 2.3.1 Formulation of the optimization problem for intensity estimation

We propose an intensity estimation procedure for $\mathbf{x}$ restricted only to the pixels of interest. Under this constraint, it is thus reasonable to consider a regularization term favoring smooth intensities on $\Omega$, in agreement to what is typically found in real images. In order to take into account the modeling of blurry and out-of-focus fluorescent molecules, we further include in our model (2.6) a regularization term for smooth background estimation. We can thus consider the following joint minimization problem:

$$\underset{\mathbf{x}\in\mathbb{R}_+^{|\Omega|},\, \mathbf{b}\in\mathbb{R}_+^{M^2}}{\arg\min} \frac{1}{2}\|\mathbf{\Psi_\Omega x} - (\overline{\mathbf{y}} - \mathbf{b})\|_2^2 + \frac{\mu}{2}\|\nabla_\Omega \mathbf{x}\|_2^2 + \frac{\beta}{2}\|\nabla \mathbf{b}\|_2^2, \tag{2.20}$$

where the data term models the presence of Gaussian noise, $\overline{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{y_t}$ is the temporal empirical average of the acquired data, $\mu, \beta > 0$ are regularization parameters and $\mathbf{\Psi_\Omega} \in \mathbb{R}^{M^2 \times |\Omega|}$ is a matrix whose $i$-th column is extracted from $\mathbf{\Psi}$ for all indexes $i \in \Omega$. The regularization term on $\mathbf{x}$ is the squared norm of the discrete gradient restricted to $\Omega$, i.e.:

$$\|\nabla_\Omega \mathbf{x}\|_2^2 := \sum_{i\in\Omega} \sum_{j\in\mathcal{N}(i)\cap\Omega} (x_i - x_j)^2,$$

where $\mathcal{N}(i)$ denotes the $3\times 3$-pixel neighborhood of $i \in \Omega$. Note that, according to this definition, $\nabla_\Omega \mathbf{x}$ denotes a (redundant) isotropic discretization of $\nabla \mathbf{x}$ evaluated for each pixel in the support $\Omega$. It coincides with the standard definition of $\nabla \mathbf{x}$ restricted to points in the support $\Omega$.

The non-negativity constraints on $\mathbf{x}$ and $\mathbf{b}$ as well as the one restricting the estimation of $\mathbf{x}$ on $\Omega$ can be relaxed by using suitable smooth penalty terms, so that, finally, the following

unconstrained optimization problem is addressed:

$$\underset{\mathbf{x}\in\mathbb{R}^{L^2},\,\mathbf{b}\in\mathbb{R}^{M^2}}{\arg\min}\ \frac{1}{2}\|\mathbf{\Psi x}-(\overline{\mathbf{y}}-\mathbf{b})\|_2^2+\frac{\mu}{2}\|\nabla\mathbf{x}\|_2^2+\frac{\beta}{2}\|\nabla\mathbf{b}\|_2^2+\frac{\alpha}{2}\left(\|\mathbf{I_\Omega x}\|_2^2+\sum_{i=1}^{L^2}[\phi(x_i)]^2+\sum_{i=1}^{M^2}[\phi(b_i)]^2\right), \tag{2.21}$$

where the parameter $\alpha \gg 1$ can be chosen arbitrarily high to enforce the constraints. $\mathbf{I_\Omega} \in \mathbb{R}^{L^2\times L^2}$ is a diagonal matrix acting as characteristic function of $\Omega$, i.e. defined as:

$$\mathbf{I_\Omega}(i,i) = \begin{cases} 0 & \text{if } i \in \Omega, \\ 1 & \text{if } i \notin \Omega \end{cases}, \qquad \forall i \in \{1,...,L^2\}. \tag{2.22}$$

The function $\phi : \mathbb{R} \to \mathbb{R}$ is used to penalize negative entries, being defined as:

$$\phi(z) := \begin{cases} 0 & \text{if } z \geq 0, \\ z & \text{if } z < 0 \end{cases}, \qquad \forall z \in \mathbb{R}. \tag{2.23}$$

By considering the unconstrained optimization problem in (2.21) instead of the original constrained one in (2.20), we facilitate the development of an automatic parameter selection strategy, as we describe in detail in Section 2.3.2.

To solve problem (2.21) we use the Alternate Minimization algorithm, see Algorithm 2. In the following Sections 2.3.1.1 and 2.3.1.2, we provide more details on the solution of the two minimization sub-problems.

**Algorithm 2** COL0RME, Step II: Intensity Estimation

**Require:** $\overline{\mathbf{y}} \in \mathbb{R}^{M^2}, \mathbf{x}^0 \in \mathbb{R}^{L^2}, \mathbf{b}^0 \in \mathbb{R}^{M^2}, \mu, \beta > 0, \alpha \gg 1$

**repeat**

$\mathbf{x}^{k+1} = \underset{\mathbf{x}\in\mathbb{R}^{L^2}}{\arg\min}\ \frac{1}{2}\|\mathbf{\Psi x} - (\overline{\mathbf{y}} - \mathbf{b^k})\|_2^2 + \frac{\mu}{2}\|\nabla\mathbf{x}\|_2^2 + \frac{\alpha}{2}\left(\|\mathbf{I_\Omega x}\|_2^2 + \sum_{i=1}^{L^2}[\phi(x_i)]^2\right)$

$\mathbf{b}^{k+1} = \underset{\mathbf{b}\in\mathbb{R}^{M^2}}{\arg\min}\ \frac{1}{2}\|\mathbf{b} - (\overline{\mathbf{y}} - \mathbf{\Psi x}^{k+1})\|_2^2 + \frac{\beta}{2}\|\nabla\mathbf{b}\|_2^2 + \frac{\alpha}{2}\sum_{i=1}^{M^2}[\phi(b_i)]^2$

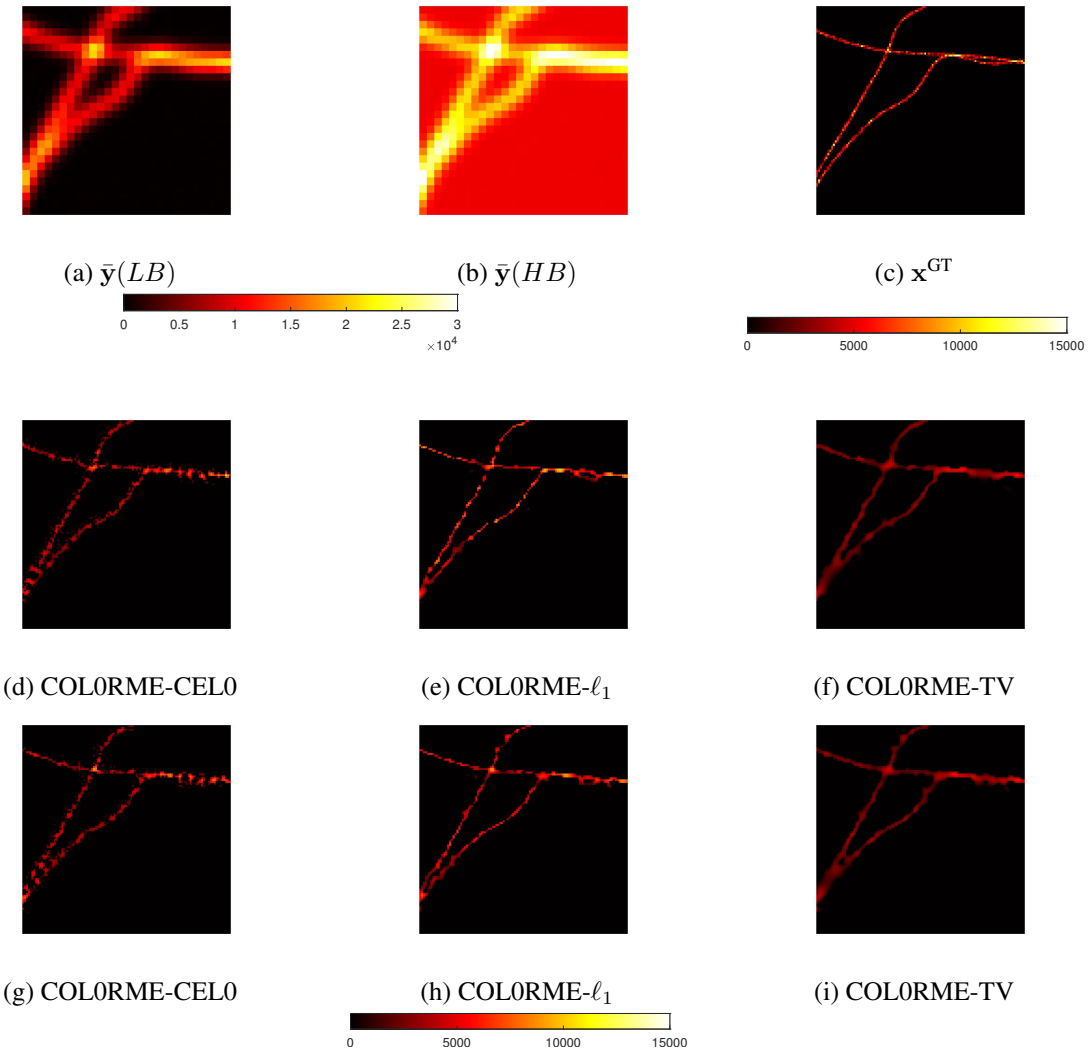
**until** convergence

**return** $\mathbf{x}, \mathbf{b}$

Intensity estimation results can be found in Figure 2.10. Thus, (2.21) is used for intensity/background estimation on the supports $\Omega_\mathcal{R}$ estimated from the first step of COL0RME using $\mathcal{R} = \text{CEL0}$, $\mathcal{R} = \ell_1$ and $\mathcal{R} = \text{TV}$. We refer to these solutions as COL0RME-CEL0, COL0RME-$\ell_1$ and COL0RME-TV, respectively. The result of TV, even after the use of the second step does not allow for the observation of significant details (e.g. the separation of the two filaments on the bottom left corner).

A quantitative assessment for the other two regularization penalty choices, is available in Figure 2.11. More precisely we compute the Peak-Signal-to-Noise-Ratio (PSNR), given the following formula:

$$\text{PSNR}_{\text{dB}}(\mathbf{x}^{\text{GT}}, \mathbf{x}) = 10\log_{10}\left(\frac{\text{MAX}^2_{\mathbf{x}^{\text{GT}}}}{\text{MSE}}\right), \qquad \text{MSE} = \frac{1}{L^2}\sum_{i=1}^{L^2}\left((\mathbf{x}^{\text{GT}})_i - (\mathbf{x})_i\right)^2, \tag{2.24}$$

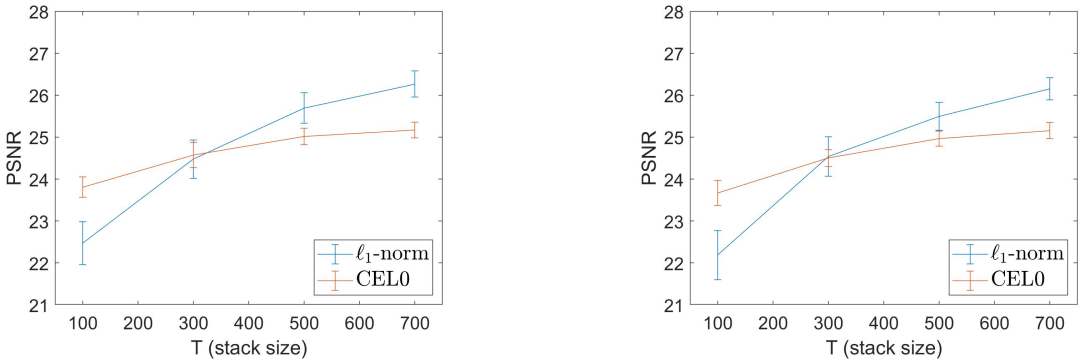



(a) $\bar{\mathbf{y}}(LB)$ (b) $\bar{\mathbf{y}}(HB)$ (c) $\mathbf{x}^{\mathrm{GT}}$

(d) COL0RME-CEL0 (e) COL0RME-$\ell_1$ (f) COL0RME-TV

(g) COL0RME-CEL0 (h) COL0RME-$\ell_1$ (i) COL0RME-TV

Figure 2.10 – (a-b) Diffraction limited image $\bar{\mathbf{y}}$ with $T = 500$ frames for the LB dataset and for the HB dataset (4x zoom), (c) GT intensity image, (d-f) Reconstructions for the noisy simulated dataset with LB, (g-i) Reconstruction for the noisy simulated dataset with HB. For all COL0RME intensity estimations, the same colorbar, presented at the bottom of the figure, has been used.




(a) LB dataset (b) HB dataset

Figure 2.11 – COL0RME PSNR values for two different datasets (LB and HB dataset), stack sizes, and regularization penalty choices. The mean and the standard deviation (error bars) of 20 different noise realizations are presented.

where $\mathbf{x}^{\text{GT}} \in \mathbb{R}^{L^2}$ is the ground truth (GT) intensity image used as a reference image, $\mathbf{x} \in \mathbb{R}^{L^2}$ the image we want to evaluate using the PSNR metric and $\text{MAX}_{\mathbf{x}^{\text{GT}}}$ the maximum value of the image $\mathbf{x}^{\text{GT}}$. The higher the PSNR, the better the quality of the reconstructed image.

As shown in Figures 2.10 and 2.11, when only a few frames are considered (eg. $T = 100$ frames, high temporal resolution), the method performs better when the CEL0 penalty is used for the support estimation. However, when longer temporal sequences are available (e.g. $T = 500$ or $T = 700$ frames) the method performs better when the $\ell_1$-norm is used instead. In addition to this, for both penalizations, PSNR improves as the number of temporal frames increases, as expected.

Background estimation results are reported in Figure 2.12 where (2.21) is used for intensity/background estimation on the supports $\Omega_{\mathcal{R}}$, with $\mathcal{R} = \text{CEL0}$ and $\mathcal{R} = \ell_1$. Note the different scales between the diffraction limited and background images for a better visualization of the results. In the figure we further report the constant background generated by the SOFI Simulation Tool [Girsault et al., 2016], the software used to generate our simulated data (more details in Section 2.4.1). Although the results look different due to the considered space-variant regularization on $\mathbf{b}$, the variations are very small. The estimated background is smooth, as expected, with higher values estimated near the simulated filaments and values closer to the true background found away from them.

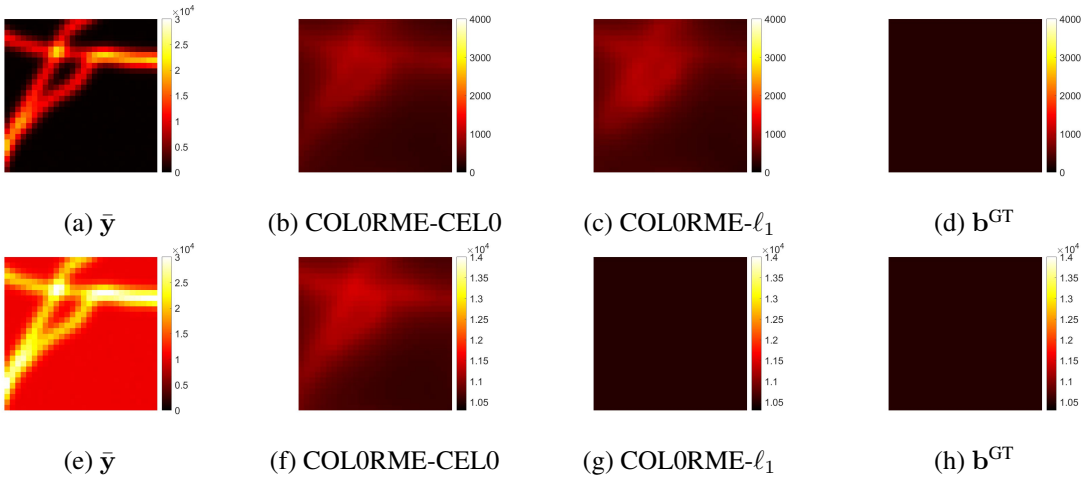
#### 2.3.1.1 Update of x

In order to find at each $k \geq 1$ the optimal solution $\mathbf{x}^{k+1} \in \mathbb{R}^{L^2}$ of the first sub-problem, we need to solve a minimization problem of the form:

$$\mathbf{x}^{k+1} = \underset{\mathbf{x} \in \mathbb{R}^{L^2}}{\arg\min}\ g(\mathbf{x}; \mathbf{b}^k) + h(\mathbf{x}), \tag{2.25}$$

where, for $\mathbf{b}^k \in \mathbb{R}^{M^2}$ being fixed at each iteration, $g(\cdot; \mathbf{b}^k) : \mathbb{R}^{M^2} \to \mathbb{R}_+$ is a proper and convex function with Lipschitz gradient, defined by:

$$g(\mathbf{x}; \mathbf{b}^k) := \frac{1}{2}\|\mathbf{\Psi}\mathbf{x} - (\overline{\mathbf{y}} - \mathbf{b}^k)\|_2^2 + \frac{\mu}{2}\|\nabla \mathbf{x}\|_2^2, \tag{2.26}$$



(a) $\bar{\mathbf{y}}$ (b) COL0RME-CEL0 (c) COL0RME-$\ell_1$ (d) $\mathbf{b}^{\text{GT}}$

(e) $\bar{\mathbf{y}}$ (f) COL0RME-CEL0 (g) COL0RME-$\ell_1$ (h) $\mathbf{b}^{\text{GT}}$

Figure 2.12 – (a-d) LB dataset results: Diffraction limited image $\bar{\mathbf{y}}$ with T=500 (4x zoom), Background estimation result on estimated support using CEL0 and $\ell_1$ regularization, GT background image. (e-h) HB dataset results: Diffraction limited image $\bar{\mathbf{y}}$ with T=500 (4x zoom), Background estimation result on estimated support using CEL0 and $\ell_1$ regularization, GT background image.

and where the function $h : \mathbb{R}^{L^2} \to \mathbb{R}$ encodes the penalty terms:

$$h(\mathbf{x}) = \frac{\alpha}{2}\left(\|\mathbf{I}_{\boldsymbol{\Omega}}\mathbf{x}\|_2^2 + \sum_{i=1}^{L^2}[\phi(x_i)]^2\right). \tag{2.27}$$

Solution of (2.25) can be obtained iteratively, using, for instance, Proximal Gradient Descent (PGD) algorithm, whose iteration is defined by:

$$\mathbf{x}^{n+1} = \text{prox}_{h,\tau}(\mathbf{x}^n - \tau\nabla g(\mathbf{x}^n)), \quad n = 1, 2, .., \tag{2.28}$$

where $\nabla g(\cdot)$ denotes the gradient of $g$, $\tau \in (0, \frac{1}{L_g}]$ is the algorithmic step-size chosen within a range depending on the Lipschitz constant of $\nabla g$, here denoted by $L_g$, to guarantee convergence. The proximal update in (2.28) can be computed explicitly using the computations reported in Appendix A. For each $\mathbf{w} \in \mathbb{R}^{L^2}$ there holds element-wise:

$$\left(\text{prox}_{h,\tau}(\mathbf{w})\right)_i = \text{prox}_{h,\tau}(w_i) = \begin{cases} \frac{w_i}{1+\alpha\tau\mathbf{I}_{\boldsymbol{\Omega}}(i,i)} & \text{if } w_i \geq 0, \\ \frac{w_i}{1+\alpha\tau(\mathbf{I}_{\boldsymbol{\Omega}}(i,i)+1)} & \text{if } w_i < 0. \end{cases} \tag{2.29}$$

*Remark 2.3.1* – As the reader may have noticed, we consider the proximal gradient descent algorithm (2.28) for solving (2.25), even though both functions $g$ and $h$ in (2.26) and (2.27) respectively, are smooth and convex, hence, in principle, (accelerated) gradient descent algorithms could be used. Note, however, that the presence of the large penalty parameter $\alpha \gg 1$ would significantly slow down convergence speed in such case as the step size $\tau$ would be constrained to the smaller range $(0, \frac{1}{L_g+\alpha}]$. By considering the penalty contributions in terms of their proximal operators, this limitation doesn't affect the range of $\tau$, and convergence is still guaranteed [Combettes and Wajs, 2005] in a computationally fast way through the update (2.29).

#### 2.3.1.2 Update of b

As far as the estimation of the background is concerned, the minimization problem we aim to solve at each $k \geq 1$ takes the form:

$$\mathbf{b}^{k+1} = \underset{\mathbf{b}\in\mathbb{R}^{M^2}}{\arg\min}\ r(\mathbf{b}; \mathbf{x}^{k+1}) + q(\mathbf{b}), \tag{2.30}$$

where:

$$r(\mathbf{b}; \mathbf{x}^{k+1}) := \frac{1}{2}\|\mathbf{b} - (\overline{\mathbf{y}} - \mathbf{\Psi}\mathbf{x}^{k+1})\|_2^2 + \frac{\beta}{2}\|\nabla\mathbf{b}\|_2^2, \qquad q(\mathbf{b}) := \frac{\alpha}{2}\sum_{i=1}^{M^2}[\phi(b_i)]^2.$$

Note that $r(\cdot; \mathbf{x}^{k+1}) : \mathbb{R}^{M^2} \to \mathbb{R}_+$ is a convex function with $L_r$-Lipschitz gradient and $q : \mathbb{R}^{M^2} \to \mathbb{R}_+$ encodes (large, depending on $\alpha \gg 1$) penalty contributions. Recalling Remark 2.3.1, we thus use again the proximal gradient descent algorithm for solving (2.30). The desired solution can thus be found by iterating:

$$\mathbf{b}^{n+1} = \text{prox}_{q,\delta}(\mathbf{b}^n - \delta\nabla r(\mathbf{b}^n)), \quad n = 1, 2, .., \tag{2.31}$$

where $\delta \in (0, \frac{1}{L_r}]$ is the step-size parameter. The proximal operator $\text{prox}_{q,\delta}(\cdot)$, has an explicit expression and it is defined element-wise for $i = 1, \ldots, M^2$ by:

$$\Big(\text{prox}_{q,\delta}(\mathbf{d})\Big)_i = \text{prox}_{q,\delta}(d_i) = \begin{cases} d_i & \text{if } d_i \geq 0, \\ \frac{d_i}{1+\alpha\delta} & \text{if } d_i < 0. \end{cases} \tag{2.32}$$

### 2.3.2 Hyper-parameter selection

In this section, we provide some details on the estimation of the parameter $\mu$ in (2.20), which is crucial for an accurate intensity estimation. The other two regularization parameters $\beta$ and $\alpha$ do not need fine tuning: they are both chosen sufficiently large, with $\beta$ allowing for smooth background and with high $\alpha$ to enforce the required constraints (positivity for both intensity and background and restriction to the predefined support only for the intensity estimation).

The problem we look at in this second step is:

$$\text{find} \quad \mathbf{x} \in \mathbb{R}^{L^2} \quad \text{s.t.} \quad \overline{\mathbf{y}} = \mathbf{\Psi}\mathbf{x} + \mathbf{b} + \overline{\mathbf{n}}, \tag{2.33}$$

where the quantities denoted by a bar sign correspond to the temporal averages of the vectorized model in (2.6), so that, e.g., $\overline{\mathbf{n}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{n_t}$. The temporal realizations $\mathbf{n_t}$ of the random vector $\mathbf{n}$ follow a normal distribution with zero mean and covariance matrix $s\mathbf{I}_{M^2}$, where $s$ has been estimated in the first step of the algorithm, see Section 2.2.2. Consequently, the vector $\overline{\mathbf{n}}$ follows also a normal distribution with zero mean and covariance matrix equal to $\frac{s}{T}\mathbf{I}_{M^2}$. As both $s$ and $T$ are thus known at this point, we can use the discrepancy principle, a well-known *a-posteriori* parameter-choice strategy (see, e.g., [Hansen, 2010, Gfrerer, 1987]), to efficiently estimate the hyper-parameter $\mu$ appearing in (2.21).

To detail how the procedure is applied to our problem, we write $\mathbf{x}_\mu$ in the following to highlight the dependence of $\mathbf{x}$ on $\mu$. According to the discrepancy principle, an optimal regularization parameter $\mu$ is chosen so that the residual norm of the regularized solution satisfies:

$$\|\overline{\mathbf{y}} - \Psi\hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}\|_2^2 = \nu_{DP}^2\|\overline{\mathbf{n}}\|_2^2, \tag{2.34}$$

where $\hat{\mathbf{x}}_\mu \in \mathbb{R}^{L^2}$ and $\hat{\mathbf{b}} \in \mathbb{R}^{M^2}$ are the solutions of (2.20). By definition, the expected value of $\|\overline{\mathbf{n}}\|_2^2$ is:

$$\mathbb{E}\{\|\overline{\mathbf{n}}\|_2^2\} = M^2 \frac{s}{T}, \tag{2.35}$$

which can be used as an approximation of $\|\overline{\mathbf{n}}\|_2^2$ for $M^2$ big enough. The scalar value $\nu_{DP} \approx 1$ is a 'safety factor' that plays an important role in the case when a good estimate of $\|\overline{\mathbf{n}}\|_2$ is not available. In such situations a value $\nu_{DP}$ closer to 2 is used. As detailed in Section 2.2.2, the estimation of $s$ is rather precise in this case, hence we fix $\nu_{DP} = 1$ in the following.

We now define the function $f(\mu) : \mathbb{R}_+ \to \mathbb{R}$ as:

$$f(\mu) = \frac{1}{2}\|\overline{\mathbf{y}} - \boldsymbol{\Psi}\hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}\|_2^2 - \frac{\nu_{DP}^2}{2}\|\overline{\mathbf{n}}\|_2^2. \tag{2.36}$$

We want to find the value $\hat{\mu}$ such that $f(\hat{\mu}) = 0$. This can be done iteratively, using, e.g., Newton's method whose iterations read:

$$\mu_{n+1} = \mu_n - \frac{f(\mu_n)}{f'(\mu_n)}, \quad n = 1, 2, ... \tag{2.37}$$

In order to be able to compute easily the values $f(\mu)$ and $f'(\mu)$, the values $\hat{\mathbf{x}}_\mu \in \mathbb{R}^{L^2}$, $\hat{\mathbf{b}} \in \mathbb{R}^{M^2}$ and $\hat{\mathbf{x}}'_\mu = \frac{\partial}{\partial\mu}\hat{\mathbf{x}}_\mu \in \mathbb{R}^{L^2}$ need to be computed, as it can be easily noticed by writing the expression of $f'(\mu)$ which reads:

$$f'(\mu) = \frac{\partial}{\partial\mu}\{\frac{1}{2}\|\overline{\mathbf{y}} - \boldsymbol{\Psi}\hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}\|_2^2\} = (\hat{\mathbf{x}}'_\mu)^\intercal \boldsymbol{\Psi}^\intercal(\overline{\mathbf{y}} - \boldsymbol{\Psi}\hat{\mathbf{x}}_\mu - \hat{\mathbf{b}}). \tag{2.38}$$

The values $\hat{\mathbf{x}}_\mu$ and $\hat{\mathbf{b}}$ can be found by solving the minimization problem (2.20). As far as $\hat{\mathbf{x}}'_\mu$ is concerned, we report in Appendix B the steps necessary for its computation. We note that in order to compute such quantity, the use of a relaxed support/non-negativity constraint defined by the smooth quadratic terms discussed above is fundamental as it allows a closed-form expression. One can show in fact that $\hat{\mathbf{x}}'_\mu$ is the solution of the following minimization problem:

$$\hat{\mathbf{x}}'_\mu = \underset{\mathbf{x}\in\mathbb{R}^{L^2}}{\arg\min} \frac{1}{2}\|\boldsymbol{\Psi}\mathbf{x}\|_2^2 + \frac{\mu}{2}\|\nabla\mathbf{x} + \mathbf{c}\|_2^2 + \frac{\alpha}{2}\left(\|\mathbf{I}_{\boldsymbol{\Omega}}\mathbf{x}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu}\mathbf{x}\|_2^2\right), \tag{2.39}$$

where $\mathbf{c}$ is a known quantity defined by $\mathbf{c} = \frac{1}{\mu}\nabla\hat{\mathbf{x}}_\mu$, and the diagonal matrix $\mathbf{I}_{\hat{\mathbf{x}}_\mu} \in \mathbb{R}^{L^2\times L^2}$ identifies the support of $\hat{\mathbf{x}}_\mu$, i.e.:

$$I_{\hat{x}_\mu}(i,i) = \begin{cases} 0 & \text{if } (\hat{\mathbf{x}}_\mu)_i \geq 0, \\ 1 & \text{if } (\hat{\mathbf{x}}_\mu)_i < 0. \end{cases}$$

We can thus find $\hat{\mathbf{x}}'_\mu$ by iterating:

$$\mathbf{x'}_\mu^{n+1} = \text{prox}_{\overline{h},\tau}(\mathbf{x'}_\mu^n - \tau\nabla\overline{g}(\mathbf{x'}_\mu^n)), \quad n = 1, 2, .., \tag{2.40}$$

where

$$\overline{g}(\mathbf{x}) := \frac{1}{2}\|\boldsymbol{\Psi}\mathbf{x}\|_2^2 + \frac{\mu}{2}\|\nabla\mathbf{x} + \mathbf{c}\|_2^2, \qquad \overline{h}(\mathbf{x}) := \frac{\alpha}{2}\left(\|\mathbf{I}_{\boldsymbol{\Omega}}\mathbf{x}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu}\mathbf{x}\|_2^2\right). \tag{2.41}$$

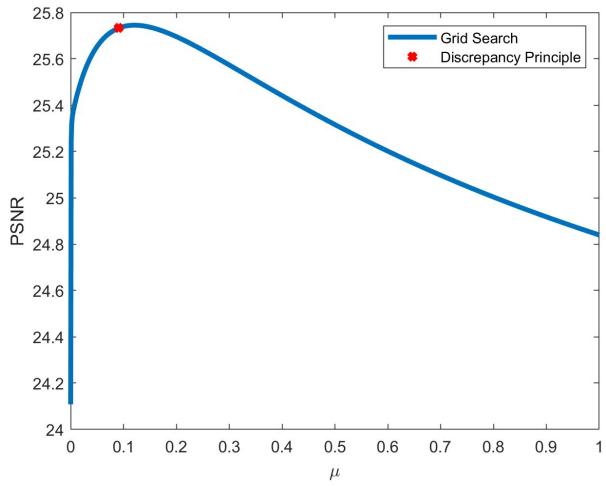



Figure 2.13 – The solid blue line shows the PSNR values computed by solving (2.21) for several values of $\mu$ within a specific range.

For $\mathbf{z} \in \mathbb{R}^{L^2}$, the proximal operator $\text{prox}_{\overline{h},\tau}(\mathbf{z})$ can be obtained following the computations in Appendix A:

$$(\text{prox}_{\overline{h},\tau}(\mathbf{z}))_i = \text{prox}_{\overline{h},\tau}(z_i) = \frac{z_i}{1 + \alpha\tau\left(\mathbf{I}_{\boldsymbol{\Omega}}(i,i) + \mathbf{I}_{\hat{\mathbf{x}}_\mu}(i,i)\right)}, \tag{2.42}$$

while

$$\nabla\overline{g}(\mathbf{x}') = (\boldsymbol{\Psi}^\intercal\boldsymbol{\Psi} + \mu\nabla^\intercal\nabla)\mathbf{x}' + \nabla^\intercal\nabla\hat{\mathbf{x}}_\mu, \tag{2.43}$$

and the step-size $\tau \in (0, \frac{1}{L_{\overline{g}}}]$, with $L_{\overline{g}} = \|\boldsymbol{\Psi}^\intercal\boldsymbol{\Psi} + \mu\nabla^\intercal\nabla\|_2$ the Lipschitz constant of $\nabla\overline{g}$. A pseudo-code explaining the procedure is reported in Algorithm 3.

**Algorithm 3** Discrepancy Principle

**Require:** $\overline{\mathbf{y}} \in \mathbb{R}^{M^2}, \mathbf{x}^0 \in \mathbb{R}^{L^2}, \mathbf{b}^0 \in \mathbb{R}^{M^2}, \mu_0, \beta > 0, \alpha \gg 1$
  **repeat**
    Find $\hat{\mathbf{x}}_{\mu_n}, \hat{\mathbf{b}}$ using Algorithm 2
    Find $\hat{\mathbf{x}}'_{\mu_n}$ solving (2.39)
    Compute $f(\mu_n), f'(\mu_n)$ from (2.36) and (2.38)
    $\mu_{n+1} \leftarrow \mu_n - \frac{f(\mu_n)}{f'(\mu_n)}$
  **until** convergence
  **return** $\hat{\mu}$

Finally, in Figure 2.13, a numerical example showing the good estimation of the parameter $\hat{\mu}$ is given. For this example, we used the HB dataset with $T = 500$ frames (Figure 2.15c) and the $\ell_1$-norm regularization penalty. The red cross shows the PSNR value obtained in correspondence with the parameter $\hat{\mu}$ computed by applying the Discrepancy Principle. We note that such a value is very close to one maximizing the PSNR metric.

## 2.4 Numerical Results

In this section, we compare the method COL0RME with state-of-the-art methods that exploit the temporal fluctuations/blinking of fluorophores. We apply them to both simulated and real data. More precisely, we compare: COL0RME-CEL0 (using the CEL0 regularization in the support estimation), COL0RME-$\ell_1$ (using the $\ell_1$-norm regularization in the support estimation), SRRF [Gustafsson et al., 2016], SPARCOM [Solomon et al., 2019] and LSPARCOM [Dardikman-Yoffe and Eldar, 2020].

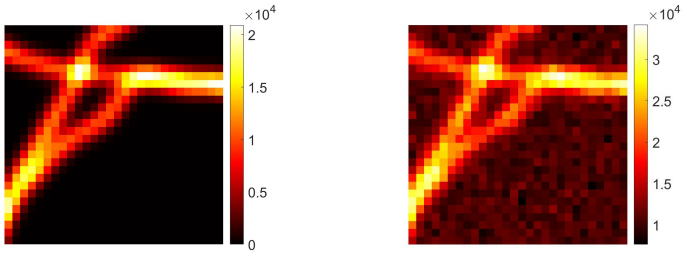
### 2.4.1 Simulated Data

To evaluate the method COL0RME we choose images of tubular structures that simulate standard microscope acquisitions with standard fluorescent dyes. In particular, the spatial pattern (see Figure 2.15a) is taken from the MT0 microtubules training dataset uploaded for the SMLM Challenge of 2016 [1]. The temporal fluctuations are obtained by using the SOFI Simulation Tool [Girsault et al., 2016]. This simulation software, implemented in MATLAB, generates realistic temporal stacks of images with fluctuating intensities, similar to the ones obtained from real microscopes, as it makes use of parameters specific to some microscope setup and some of the sample's main properties. However, differently from the fluctuating microscopic data presented in Section 2.4.2, the blinking generated by the SOFI Simulation Tool has a more distinctive "on-off" behavior.

For the experiments presented in this section, we generate a video of a total number of 700 frames following model (2.1). However, we evaluate the methods using the first $T \in \{100, 300, 500, 700\}$ frames, so as to examine further the trade-off between temporal and spatial resolution. The frame rate is fixed at 100 frames per second (fps) and the pixel size is 100 nm. Regarding the optical parameters, we set the numerical aperture equal to 1.4 and the emission wavelength to 525 nm, so that the FWHM of the PSF is equal to 228.75 nm. The fluorophore parameters are set as follows: 20 ms for on-state average lifetime, 40 ms for off-state average lifetime, and 20 s for average time until bleaching. The emitter density is equal to 10.7 emitters/pixel/frame, while 500 photons are emitted, on average, by a single fluorescent molecule in every frame.

We create two datasets with the main difference between them being the background level. More precisely we create a low-Background (LB) dataset, where the background is equal to 50 photons/pixel/frame, and a high-Background (HB) dataset, where the background is equal to 2500 photons/pixel/frame. For both datasets, we proceed as follows: initially, Poisson noise is added to simulate the photon noise (see (2.1)); subsequently, the number of photons recorded by each camera pixel is converted into an electric charge in accordance with the quantum efficiency and gain of the camera that has been set to 0.7 and 6 respectively (thus resulting in an overall gain of 4.2); finally, Gaussian noise is added. In order to give a visual inspection of the background and noise, in Figure 2.14, one frame of the HB dataset is presented before and after the background/noise addition. As we want, also, to provide a quantitative assessment, we measure the quality of the reconstruction of the final sequence of $T$ frames ($\{\mathbf{y}_t\}_{t=1}^{T}$) using the Signal-to-Noise Ratio (SNR),

1. http://bigwww.epfl.ch/smlm/datasets/index.html




(a) (b)

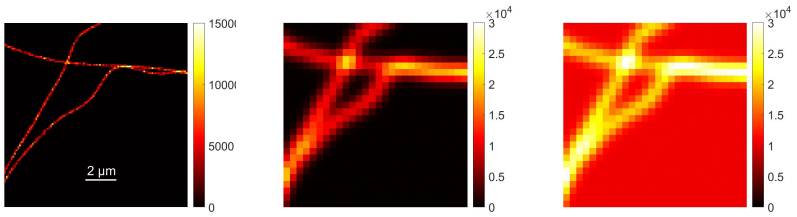
Figure 2.14 – One frame of the HB dataset, before (left) and after (right) the addition of background and the simulated noise degradation process. (a) A convoluted and down-sampled image $\mathbf{\Psi x}_t^{\text{GT}}$ obtained from a ground truth frame $\mathbf{x}_t^{\text{GT}}$, (b) A frame of the final noisy sequence: $\mathbf{y}_t$. Note the different colormaps to better capture the presence of noise and background.

given by the following formula:

$$\text{SNR}_{\text{dB}}(\mathbf{y}^*, \mathbf{y}) = 10 \log_{10} \left( \frac{\frac{1}{TM^2} \sum_{i=1}^{TM^2} (\mathbf{y}^*)_i^2}{\frac{1}{TM^2} \sum_{i=1}^{TM^2} \left((\mathbf{y}^*)_i - (\mathbf{y})_i\right)^2} \right), \tag{2.44}$$

where $\mathbf{y}^* \in \mathbb{R}^{TM^2}$ is the reference image and $\mathbf{y} \in \mathbb{R}^{TM^2}$ the image we want to evaluate, both of them in a vectorized form. As a reference, we choose the convoluted and down-sampled ground truth frames (see one frame of the reference sequence in Figure 2.14a). The SNR values for a sequence of $T = 500$ frames for the LB and HB datasets are $15.57$dB and $-6.07$dB, respectively. A negative value is found for the HB dataset due to the very high background used in this case.

The average image of the stack of each dataset as well as the ground truth intensity image are reported in Figure 2.15. In the LB dataset, due to the high signal values, the background is not visible. Further, as the observed microscopic images and the reconstructed ones belong to different grids their intensity values are not comparable, hence we can not use the same colorbar to represent them. The intensity of one pixel in the coarse grid is indeed the summation of the intensity values of $q \times q$ pixels in the fine grid, where $q$ is the super-resolution factor and in this case $q = 4$. For this reason, we use two different colorbars.




(a) GT image (b) $\bar{\mathbf{y}}$ (LB) (c) $\bar{\mathbf{y}}$ (HB)

Figure 2.15 – The GT intensity image and the diffraction limited images $\bar{\mathbf{y}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{y_t}$ for the two datasets with a 4x zoom, for a sequence of T=500 frames

Regarding the use of COL0RME-CEL0 and COL0RME-$\ell_1$, a regularization parameter equal to $\lambda = 5 \times 10^{-4} \times \lambda_{max}^{CEL0}$ and $\lambda = 5 \times 10^{-4} \times \lambda_{max}^{\ell_1}$, respectively, is used in the support estimation. The hyper-parameters $\alpha$ and $\beta$ are set as follows: $\alpha = 10^6$, $\beta = 20$. For the method COL0RME-CEL0 the algorithmic restarting approach is used for a better support estimation. It stops when there are no additional pixels added to the estimated support or if a maximum number of 10 restarts is reached. Such a number was empirically determined by preliminary simulations. For the method SRRF we are using the NanoJ SRRF plugin for Fiji [2] [Laine et al., 2019]. Concerning the method SPARCOM, we make use of the MATLAB code available online [3]. As regularization penalty we choose the $\ell_1$-norm with a regularization parameter equal to $10^{-10}$ and we avoid the post-processing step (the convolution with a small Gaussian function) for most precise localization. Finally, we test the method LSPARCOM, using the code that is available online [4] and the tubulin (TU) training set that is provided.

In Figure 2.16 we compare the reconstructions of the methods COL0RME-CEL0, COL0RME-$\ell_1$, SRRF, SPARCOM and LSPARCOM for the LB dataset and in Figure 2.17 for the HB dataset, for a sequence of T = 500 frames. Results for different stack sizes are available in Figure 2.18. Quantitative metrics like the Jaccard Index (JI) for the localization precision and the Peak-Signal-to-Noise-ration (PSNR) for the evaluation of the estimated intensities are only available for the methods COL0RME-CEL0 and COL0RME-$\ell_1$ (see Figures 2.6, 2.11). For the rest of the methods, the JI values are very low (i.e. the localization performance is bad) due to background and noise artifacts in the reconstructions that lead to the appearance of many false positives. The PSNR metric cannot be computed for the reconstructions obtained by SRRF, SPARCOM, and LSPARCOM, since these methods do not reconstruct intensity levels.

In both the LB and HB dataset, and for T= 500 frames, the best reconstruction, visually, is the one obtained by COL0RME-$\ell_1$, as it is able to achieve a more clear separation of the filaments in the critical regions (yellow and green zoom boxes). The method COL0RME-CEL0 achieves also a good result, even though the separation of the filaments, which are magnified in the green box, is not as evident. The same happens also when the method SPARCOM is used. Finally, the reconstruction obtained by SRRF and LSPARCOM, is slightly misleading.

### 2.4.2 Real Data

To show the effectiveness of our method for handling real-world data, we apply COL0RME to an image sequence acquired from a Total Internal Reflection Fluorescence (TIRF) microscope. To be more precise, the data we are using have been obtained from a Multi-Angle TIRF (MA-TIRF) microscope, with a fixed angle above the critical one. More details about this imaging modality are available in Section 1.1.2.

A sequence of $T = 500$ frames has been acquired, with an acquisition time equal to 25 s. Tubulins in endothelial cells are stained with the Alexa Fluor 488. The variance of fluctuations over time for a pixel of a region of interest is measured lying in the range $5 \times 10^5 - 7 \times 10^5$. The image $\bar{\mathbf{y}}$ is shown in Figure 2.19, together with one frame $\mathbf{y}_t$ extracted from the entire stack. The FWHM of the PSF has been measured experimentally and is equal to 292.03 nm, while the CCD camera has a pixel of size 106 nm.

2. https://github.com/HenriquesLab/NanoJ-SRRF
3. https://github.com/KrakenLeaf/SPARCOM
4. https://github.com/gilidar/LSPARCOM

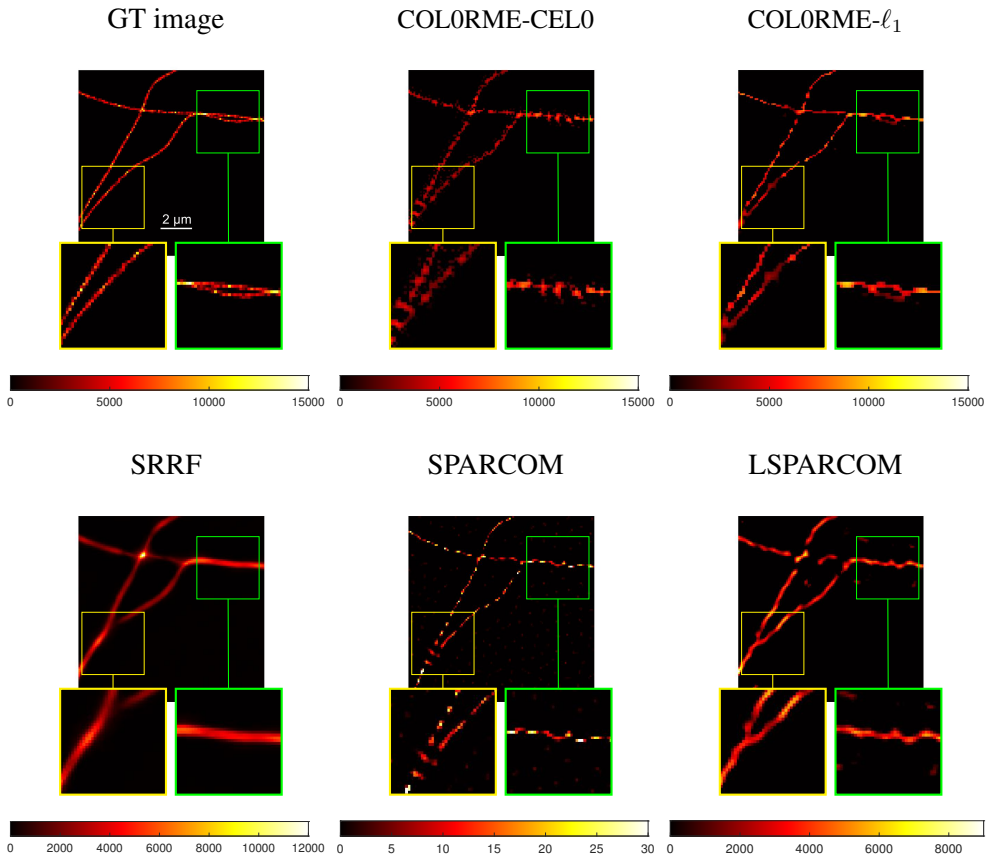



Figure 2.16 – Results for the LB dataset with $T = 500$. Note that the methods SRRF, SPARCOM, and LSPARCOM do not estimate real intensity values. Among the compared methods, only COL0RME is capable of estimating intensity values, while the other methods estimate the mean of a radiality image sequence (SRRF) and normalized autocovariances (SPARCOM, LSPARCOM)

The results of COL0RME-CEL0 and COL0RME-$\ell_1$ can be found in Figure 2.19. Experiments using different stack sizes have been done showing that the more frames we use (up to a point that we do not have many molecules bleached), the more continuous filaments we find. However, by acquiring only $T = 500$ frames we have a good balance between temporal and spatial resolution. For this reason we present here only results using a stack of 500 frames.

For the method COL0RME-CEL0 the regularization parameter $\lambda$ is equal to $\lambda = 5 \times 10^{-4} \times \lambda_{max}^{CEL0}$ and the algorithmic restarting approach has been used. Regarding the method COL0RME-$\ell_1$, we denote $\lambda = 5 \times 10^{-6} \times \lambda_{max}^{\ell_1}$, a relatively small value so as to be sure that we will include all the pixels that contain fluorescent molecules. Even if we underestimate $\lambda$ and find more false positives in the support estimation, after the second step of the algorithm, the final reconstruction is corrected, as explained in Section 2.2.3. The hyper-parameters $\alpha$ and $\beta$ are equal to: $\alpha = 10^6$, $\beta = 20$. Using either regularizer the spatial resolution is enhanced, as can be observed from the yellow zoom boxes. However, the reconstruction obtained by both COL0RME-CEL0 and COL0RME-$\ell_1$ suffers from intrinsic biases (i.e. punctuated reconstructions) due to the limitations arising from the experimental difficulties to get a staining sufficiently homogeneous for

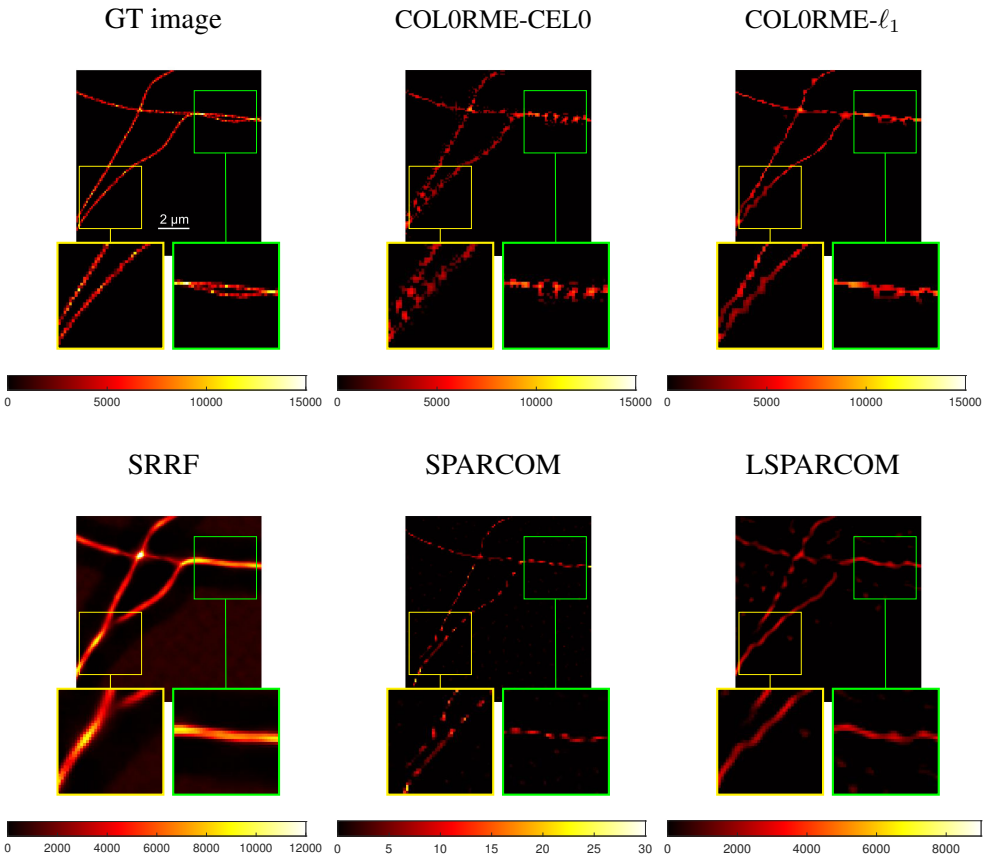



Figure 2.17 – Results for the HB dataset with $T = 500$.

this imaging resolution but also due to the sparsity-promoting regularization terms we have used. Furthermore, there are a few filaments that do not seem to be well reconstructed, especially using the COL0RME-CEL0 method, e.g. the one inside the green box.

Finally, a comparison of the methods COL0RME-CEL0 and COL0RME-$\ell_1$ with the other state-of-the-art methods is available in Figure 2.20. The parameters used for the methods SRRF, SPARCOM, and LSPARCOM, are explained in Section 2.4.1. Here, we further use a post-processing step (i.e. a convolution with a small Gaussian function) as suggested by the authors of SPARCOM. The methods COL0RME-CEL0 and COL0RME-$\ell_1$ seem to have the most precise localization, by reconstructing thin filaments, as shown in the cross-section plotted in Figure 2.20, though a bit punctuated. The most visually appealing result is the one of SRRF, where the filaments have a more continuous structure, however from the cross-section, we can see that the resolution is not much improved compared to other methods. SPARCOM and LSPARCOM do not perform very well in this real image sequence due to, mainly, background artifacts.

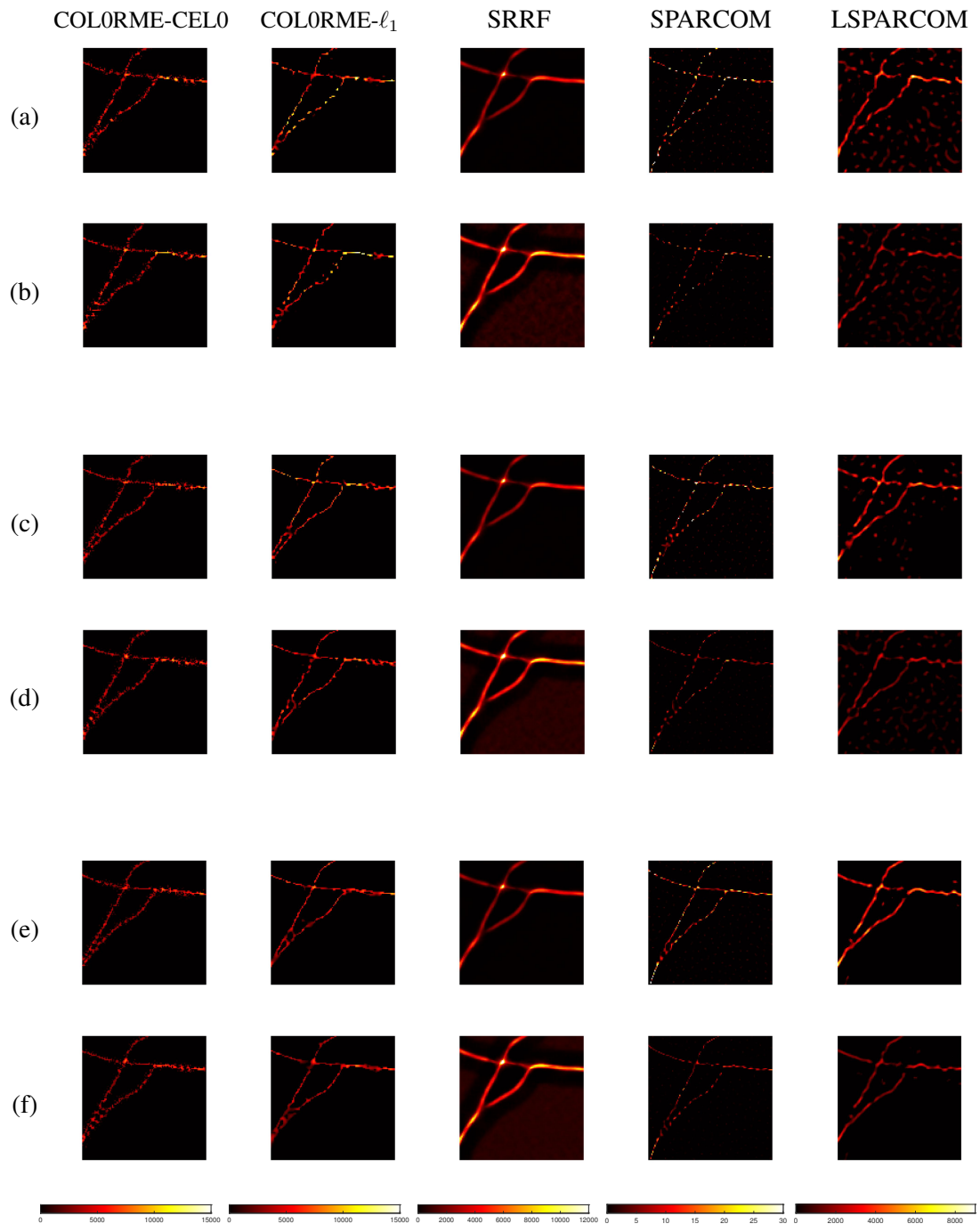



Figure 2.18 – Result for the LB and HB dataset (upper and lower row respectively) with stack sizes (a-b) $T = 100$ frames, (c-d) $T = 300$ frames and $T = 700$ frames.

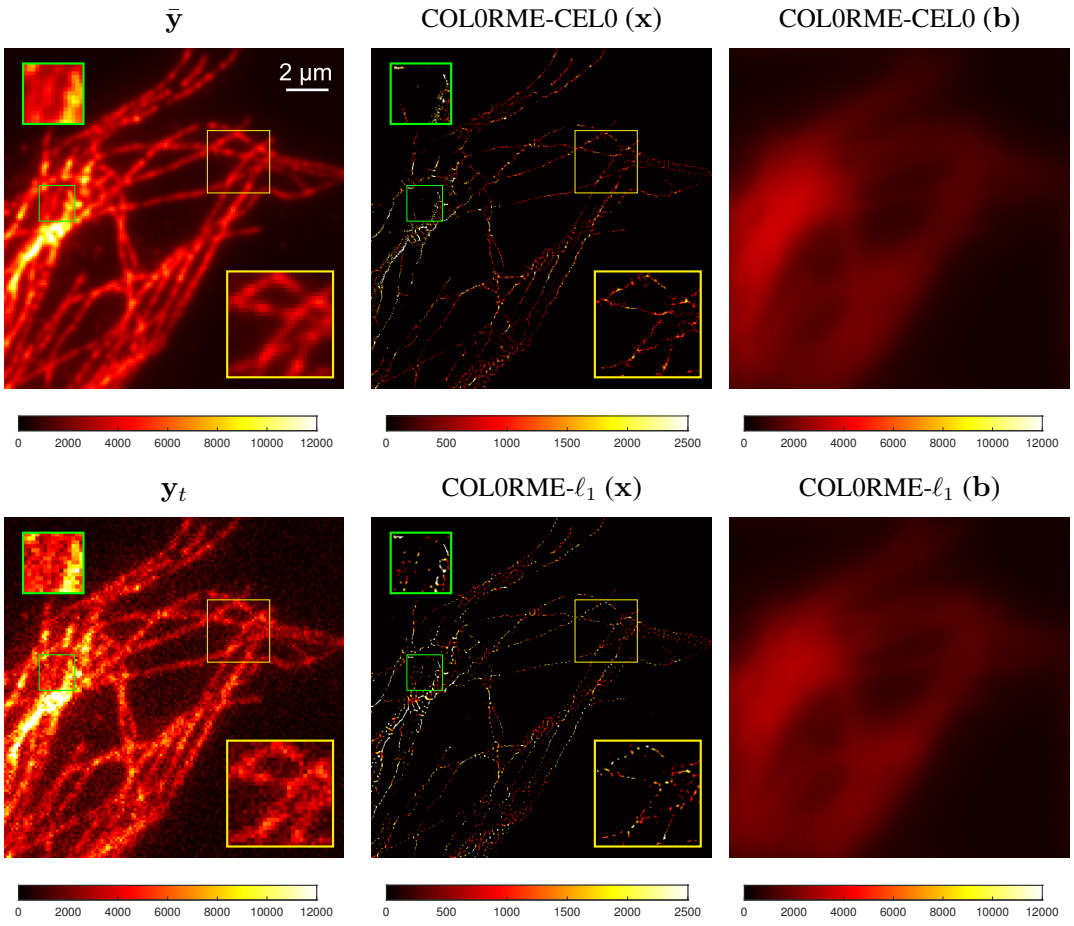



Figure 2.19 – Real TIRF data, $T = 500$ frames. Diffraction limited image $\bar{\mathbf{y}}$ (4x zoom), a frame $\mathbf{y}_t$ from the stack (4x zoom), the intensity and background estimation of the methods COL0RME-CEL0 and COL0RME-$\ell_1$.

## 2.5 Discussion

In this chapter, we proposed and discussed the model and the performance of COL0RME, a method for super-resolution microscopy imaging based on the sparse analysis of the stochastic fluctuations of molecules' intensities. Similarly to other methods exploiting temporal fluctuations, COL0RME relaxes the requirements for special equipment (microscope and fluorescent dyes) and allows for live-cell imaging, due to the good temporal resolution and the low-power lasers employed.

In comparison with competing methods, COL0RME achieves higher spatial resolution than other methods exploiting fluctuations while having a sufficient temporal resolution (acquisition time of $10 - 25$ s for estimating a super-resolved image). COL0RME is based on two different steps: in the former, accurate molecule localization and noise estimation are achieved by solving non-smooth convex/non-convex optimization problems in the covariance domain; in the latter, intensity information is retrieved in correspondence with the estimated support only. Our numerical

results show that COL0RME outperforms competing approaches in terms of localization precision.

To the best of our knowledge, COL0RME is the only super-resolution method exploiting temporal fluctuations that is capable of retrieving intensity-type information, both for the signal and for a spatially-varying background, which are of fundamental interest in biological data analysis. For both steps, automatic parameter selection strategies are detailed. We remark that the strategy of intensity estimation could be applied to the other competing super-resolution methods in the literature to improve performance. Several results obtained on both simulated and real data are discussed, showing the superior performance of COL0RME in comparison with analogous methods such as SPARCOM, LSPARCOM and SRRF.

A drawback of COL0RME is that, due to the choice of the $\ell_0$- or $\ell_1$-type regularizations, punctuated reconstructions are found. This could be corrected by choosing an appropriate regularizer, tailored to favor fine structures, similar to those observed in our simulated or real data. It is a challenging problem that we are aiming to address in Chapter 5.

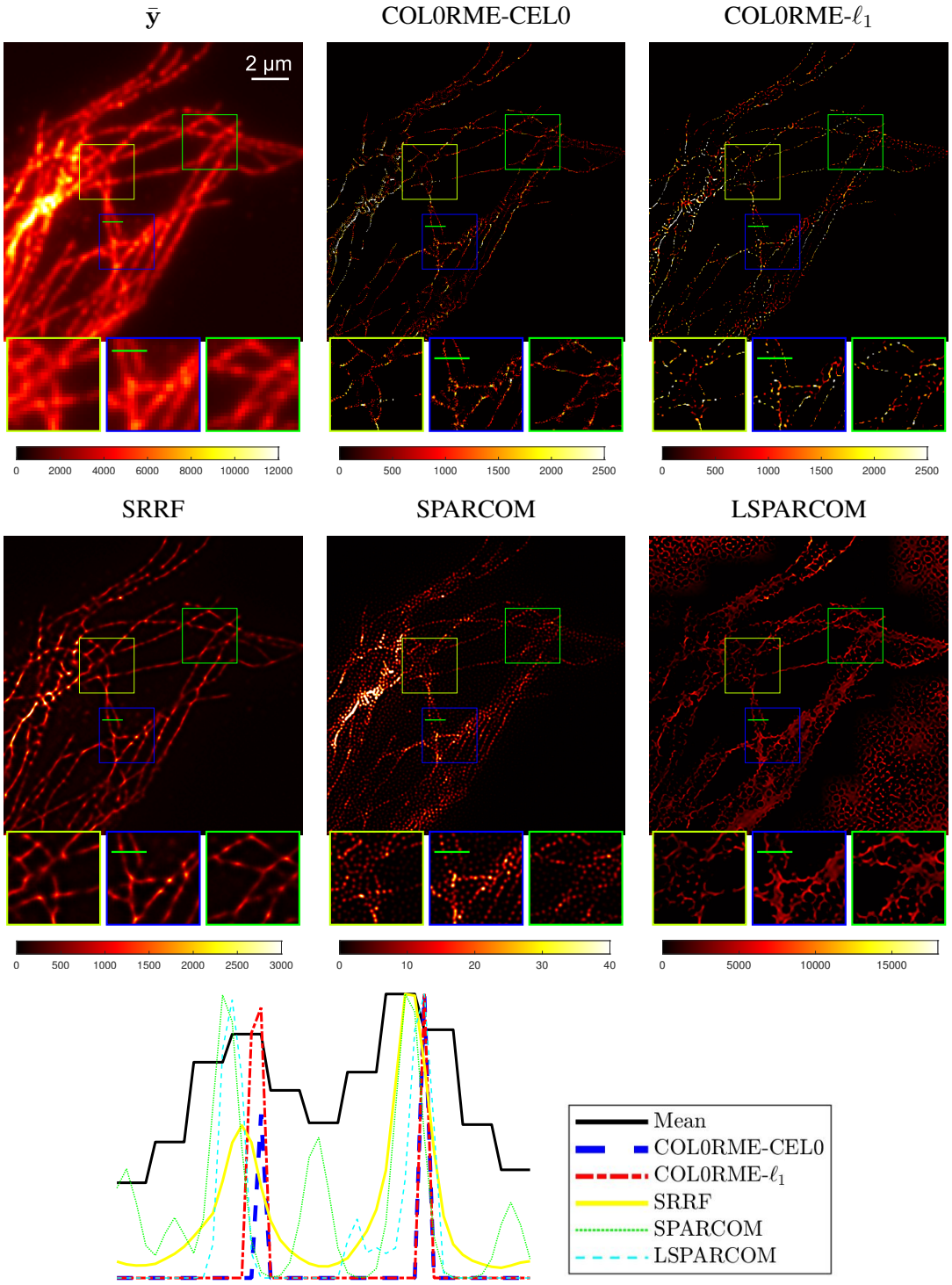



Figure 2.20 – Real TIRF data, $T = 500$ frames. Diffraction limited image $\bar{\mathbf{y}}$ (4x zoom), Comparisons between the method that exploit the temporal fluctuations, Normalized cross-section along the green line presented in the diffraction-limited and reconstructed images, but also in the blue zoom-boxes. Description of colorbars: real intensity values for $\bar{\mathbf{y}}$ and COL0RME in two different grids, mean of the radiality image sequence for SRRF, normalized autocovariances for SPARCOM and LSPARCOM.

# CHAPTER 3

# 3D Super-Resolution by MA-TIRF microscopy

*In this chapter, we present 3D MA-TIRF COL0RME, a 3D super-resolution approach that improves both the lateral and the axial spatial resolution of images of a thin layer adjacent to the coverslip in Total Internal Reflection Fluorescence (TIRF) imaging applications.*
*Specifically, in Section 3.1 we introduce the principles of multi-angle total internal reflection fluorescence (MA-TIRF) microscopy, in Section 3.2 we formulate the mathematical model through which the 3D super-resolved image is recovered, in Section 3.3 and Section 3.4 we detail the method 3D MA-TIRF COL0RME and the results obtained, respectively. Finally, in Section 3.5 we discuss the results and the improvements with respect to pre-existing approaches.*

---

**Related Publication:**

[Stergiopoulou et al., 2022b]
V. Stergiopoulou, L. Calatroni, S. Schaub and L. Blanc-Féraud, "3D Image Super-Resolution by Fluorophore Fluctuations and MA-TIRF Microscopy Reconstruction (3D-COL0RME)," 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1-4, doi: 10.1109/ISBI52829.2022.9761572.

## 3.1 From TIRF to Multi-Angle TIRF Microscopy

In this section, we first describe the principles of Total Internal Reflection Fluorescence (TIRF) microscopy by providing more details with respect to what is discussed in Chapter 1 (see Section 1.1.2). Next, we provide some basic details on Multi-Angle TIRF (MA-TIRF) microscopy which, in addition to TIRF, provides depth information and we see how to extract three-dimensional (3D) information from MA-TIRF acquisitions.

### 3.1.1 An overview of TIRF microscopy

The first TIRF microscope was developed by Daniel Axelrod in the early 1980s [Axelrod, 1981] and has been used ever since in a wide range of applications, for example, in the study of molecules or organelles in living culture cells near the contact regions with the glass coverslip. Unlike many other fluorescence microscopy techniques (see for example STED [Hell and Wichmann, 1994], SMLM [Sage et al., 2015], etc.), in TIRF microscopy there is no need for special fluorophores with specific properties since a broad spectrum of common fluorophores can be used for excitation.

TIRF microscopes allow for a way to selectively excite fluorophores only in a thin region (of a few hundred nm) adjacent to the coverslip, without exciting other fluorophores further away. The selective excitation of fluorophores offers many advantages, including a very low background, little out-of-focus fluorescence, low photobleaching, and low photodamage, as only fluorophores in the thin region are exposed to light. It is an interesting technique allowing us to observe, for example, the activity at the cell membrane, the exchange between the cell and its environment, etc.

In order to understand how the TIRF microscope works in detail we start by describing the optical phenomena of total internal reflection. When light propagates from an incident medium having a high refractive index $n_i$ to a transmitted medium with a lower refractive index $n_t$, it is possible to show the existence of a critical angle $\alpha_c$, beyond which light is totally reflected (see, for example, Figure 3.1: light with angle $\alpha_1 > \alpha_c$ is totally reflected). The critical angle $\alpha_c$ can be computed as a consequence of Snell's law via the formula:

$$\alpha_c = \sin^{-1} \frac{n_t}{n_i}. \tag{3.1}$$

For angles $\alpha > \alpha_c$, while the light beam is totally reflected back to the incident medium, there is a small penetration of the reflected light across the interface between the two media which then propagates parallel to the surface, creating an electromagnetic field in the second (transmitted) medium. Such electromagnetic field is also called *evanescent wave*, it is adjacent to the interface $z = 0$ (see Figure 3.1) and penetrates into the medium with an intensity $I(z, \alpha)$, given by:

$$I(z, \alpha) = I_0(\alpha) e^{-zp(\alpha)}, \tag{3.2}$$

where

$$p(\alpha) = \frac{4\pi n_i}{\lambda_{exc}} (\sin^2(\alpha) - \sin^2(\alpha_c))^{\frac{1}{2}} \tag{3.3}$$

is the inverse of the penetration depth and $\lambda_{exc} > 0$ is the excitation wavelength. For each $\alpha$, $I_0(\alpha)$ is the intensity of the evanescent field at the interface $z = 0$ (see [Axelrod, 1981, Axelrod, 2008] for more details).

As we can see from (3.2), the intensity of the evanescent wave decays exponentially in the $z$ direction (axial direction or direction of the optical axis). Furthermore, the decay is much faster when the penetration depth is small (or equivalently the inverse of the penetration depth $p(\alpha)$ given by (3.3) is large). The penetration depth varies with the angle of incidence of the illumination beam $\alpha$ and, in fact, the further $\alpha$ is from $\alpha_c$, the smaller it is. Figure 3.1 illustrates this. The lines parallel to the interface denote the evanescent field whose intensity fades as $z$ increases while fading much faster once $\alpha$ moves away from $\alpha_c$.

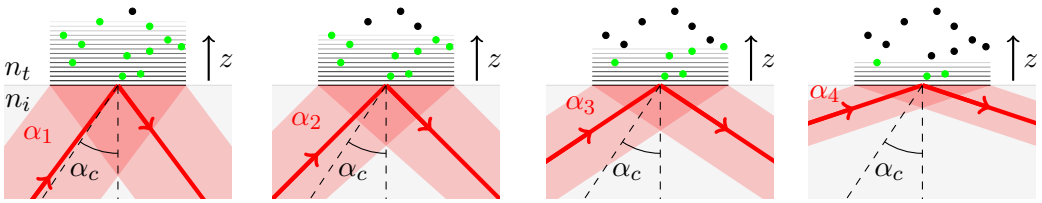



Figure 3.1 – Excited (green) and non-excited (black) fluorescent molecules for different incident angles of the illumination in TIRF microscopy.

By changing the angle and thus the penetration depth, we are able to tune the size of the illumination area in which the fluorophores will be excited. This happens because the evanescent field is capable of exciting fluorophores only in a limited region where its intensity is high enough. To better illustrate this, we assume that the small discs shown in Figure 3.1 are the fluorescent molecules, having a green color when excited by the evanescent field and a black color when not. Depending on the specific application, it is natural to select a different angle and, therefore, to excite fluorophores in a smaller or bigger area.

Another way in which selective excitation could be useful is to provide information about the depth of individual fluorescent emitters. By acquiring more than one image using different angles of illumination, optical sectioning along the $z$ axis can be achieved. In this way, it is possible to extract three-dimensional information about the biological structures observed, with the possibility of pushing the limits of (axial) resolution encountered in conventional microscopy.

### 3.1.2 Introduction to MA-TIRF microscopy

The MA-TIRF microscope is a TIRF microscope with an additional module that allows the quick change of the angle of incidence of the illumination beam. MA-TIRF microscopy benefits from all the good properties of TIRF microscopy (e.g. use of conventional fluorophores, low photobleaching/photodamage, ability to image live cells), but also, it provides depth information, by means of a collection of many TIRF images taken at different angles of incidence (always below the critical angle).

Given a stack of MA-TIRF acquisitions, a 3D super-resolved image can be estimated by means of suitable reconstruction methods. Some important works that have reached very good levels of axial resolution are the ones by [Boulanger et al., 2014, Dos Santos et al., 2014, Soubies et al., 2019b]. For example, while the axial resolution is about 500 nm for a confocal microscope and about 250 nm for a light sheet microscope, reconstruction from MA-TIRF acquisitions can provide images with an axial resolution of a few tens of nanometers in a thin layer of 500-800 nm adjacent to the coverslip, using the methods mentioned above.

In order to give a broad idea of how to reconstruct the 3D image $f : X \times \mathbb{R} \to \mathbb{R}_+$, with $X \subset \mathbb{R}^2$ being the image domain, given a set of images $\{\mathbf{y}_{\alpha_n} \in \mathbb{R}^{M^2}\}_{n=1}^{N_n}$ obtained with different incident angles of the illumination beam $\{\alpha_n : \alpha_n > \alpha_c\}_{n=1}^{N_n}$, we provide the simplified model:

$$(\mathbf{y}_{\alpha_n})_i = \int_{\mathbf{r}\in X_i} \int_{z=0}^{+\infty} I(z, \alpha_n) f(\mathbf{r}, z) dz d\mathbf{r}, \quad \forall i \in 1, \ldots, M^2, \tag{3.4}$$

where the vector $\mathbf{r} \in X_i$ denotes the lateral variables in $X_i \subset X$, the region occupied by the pixel $i$, $\forall i \in \{1, \ldots, M^2\}$, and $z \in \mathbb{R}_+$ denotes the axial variable. The intensity factor $I(z, \alpha_n)$ of the evanescent field associated with the angle $\alpha_n$ is given by (1.3). The 3D volume $f$ will be found from model (3.4) after solving the inverse problem which is ill-posed.

Note that this is a simplified model in which we have not included the effect of PSF, background, and noise, as we only use it to explain how 2D images of different incident angles can lead us to the reconstruction of a 3D image. The more detailed model can be found in the following Section 3.2.

## 3.2 Mathematical Modeling and Problem formulation

Images acquired by fluorescence microscopes, as already discussed in the previous chapters of this thesis, are subject to different types of distortion. The principal causes of distortion of the measured images are the blurring effect due to light diffraction, the out-of-focus fluorescence, as well as, the different types of noise. The mathematical model describing the acquisition process of image data $\mathbf{y}_{\alpha_n} \in \mathbb{R}^{M^2}$ obtained by means of a MA-TIRF microscope with incident angle of the illumination beam $\alpha_n > \alpha_c$, is given by:

$$(\mathbf{y}_{\alpha_n})_i = \mathcal{P}\Big( \int_{\mathbf{r}\in X_i} \int_{\mathbf{r}'\in X} \int_{z=0}^{+\infty} I(z, \alpha_n) v(\mathbf{r} - \mathbf{r}', z) f(\mathbf{r}', z)\, dz\, d\mathbf{r}'\, d\mathbf{r} + (\mathbf{b}_{\alpha_n})_i \Big) + (\mathbf{e}_{\alpha_n})_i,$$
$$\forall i \in \{1, \ldots, M^2\}, \tag{3.5}$$

where the vector $\mathbf{r}' \in X$ denotes the lateral variables in the image domain $X \subset \mathbb{R}^2$. In this model, we include also the 3D PSF $v : X \times \mathbb{R} \to \mathbb{R}$ of the optical system which is convoluted with the 3D volume $f$ we aim to reconstruct. By $\mathbf{b}_{\alpha_n} \in \mathbb{R}^{M^2}$ we denote the background term (collecting contributions from the out-of-focus and the ambient fluorophores) for each angle $\alpha_n$ and by $\mathbf{e}_{\alpha_n} \in \mathbb{R}^{M^2}$ a vector of independent and identically distributed (i.i.d) Gaussian entries of zero mean and constant variance $\sigma^2 \in \mathbb{R}_+$ modeling the presence of electronic noise. Finally, $\mathcal{P}(w)$ represents the realization of a Poisson variable $w \in \mathbb{R}$ and here models the presence of signal-dependent photon noise.

Following Section 2.1.1, we simplify model (3.5) by neglecting the Poisson noise dependence and introduce a new random term $\mathbf{n}_{\alpha_n} \in \mathbb{R}^{M^2}$ following a normal distribution with zero mean and constant variance $s^{\alpha_n} \in \mathbb{R}_+$ for all pixels. The variance $s^{\alpha_n}$ includes the variance of the electronic noise $\sigma^2$ but also a Poisson noise bias induced by $\mathbf{b}_{\alpha_n}$ (see Section 2.1.1 for more details). The

simplification of model (3.5) thus reads:

$$(\mathbf{y}_{\alpha_n})_i = \int_{\mathbf{r}\in X_i} \int_{\mathbf{r}'\in X} \int_{z=0}^{+\infty} I(z,\alpha_n) v(\mathbf{r}-\mathbf{r}',z) f(\mathbf{r}',z)\, dz\, d\mathbf{r}'\, d\mathbf{r} + (\mathbf{b}_{\alpha_n})_i + (\mathbf{n}_{\alpha_n})_i, \quad \forall i \in \{1,\dots,M^2\}. \tag{3.6}$$

Following [Soubies et al., 2019b], we can consider the PSF to be constant along the axial direction. This is a reasonable approximation since the excitation layer is very small due to the fast decay of the evanescent field, hence in this thin layer, the PSF exhibits only very little changes. We can thus set $\tilde{v}(\cdot) := v(\cdot, 0)$ to denote the 2D PSF $\tilde{v} : X \to \mathbb{R}$ which does not depend on $z$ and write:

$$(\mathbf{y}_{\alpha_n})_i = \int_{\mathbf{r}\in X_i} \int_{\mathbf{r}'\in X} \tilde{v}(\mathbf{r}-\mathbf{r}') \int_{z=0}^{+\infty} I(z,\alpha_n) f(\mathbf{r}',z)\, dz\, d\mathbf{r}'\, d\mathbf{r} + (\mathbf{b}_{\alpha_n})_i + (\mathbf{n}_{\alpha_n})_i, \quad \forall i \in \{1,\dots,M^2\}. \tag{3.7}$$

By discretizing (3.7), we find:

$$\mathbf{y}_{\alpha_n} = \mathbf{M_q U W}_{\alpha_n} \mathbf{f} + \mathbf{b}_{\alpha_n} + \mathbf{n}_{\alpha_n}, \tag{3.8}$$

where $\mathbf{f} \in \mathbb{R}^{L^2 \times N_z}$ is the discretized 3D volume to reconstruct, with $L = qM$ and $q \in \mathbb{N}$. $W_{\alpha_n} : \mathbb{R}^{L^2\times N_z} \to \mathbb{R}^{L^2}$ is an operator representing the weighted summation of the $N_z$ slices of $\mathbf{f}$, with weights related to the angle $\alpha_n$ and equal to the intensity factors $\{I(z,\alpha_n)\}_{z=1}^{N_z}$ of the evanescent field. The operator $M_q : \mathbb{R}^{L^2} \to \mathbb{R}^{M^2}$ is a down-sampling operator summing every $q$ consecutive pixels in both dimensions and $U : \mathbb{R}^{L^2} \to \mathbb{R}^{L^2}$ is a convolution operator defined by the 2D PSF of the optical imaging system. The operators $M_q$ and $U$ were described in detail in Section 2.1.1.

Revisiting equation (3.8) and setting $\mathbf{x}_{\alpha_n} := W_{\alpha_n}\mathbf{f} \in \mathbb{R}^{L^2}$ to be the laterally super-resolved image for the incident angle $\alpha_n$, it is straightforward to decouple the problem into two sub-problems. For the first sub-problem, given the laterally super-resolved images $\{\mathbf{x}_{\alpha_n} \in \mathbb{R}^{L^2}\}_{n=1}^{N_n}$, the task is to retrieve the elevation in the $z$-direction in each pixel $\mathbf{f} \in \mathbb{R}^{L^2\times N_z}$, that is:

$$\text{find} \quad \mathbf{f} \in \mathbb{R}^{L^2\times N_z} \quad \text{s.t.} \quad \mathbf{x}_{\alpha_n} = \mathbf{W}_{\alpha_n}\mathbf{f}, \tag{3.9}$$

where $\mathbf{W}_{\alpha_n} \in \mathbb{R}^{L^2 N_z \times L^2}$ is the matrix associated to the discrete TIRF operator $W_{\alpha_n} : \mathbb{R}^{L^2\times N_z} \to \mathbb{R}^{L^2}$.

The goal of the second sub-problem is to find the set of laterally super-resolved images $\{\mathbf{x}_{\alpha_n}\}_{n=1}^{N_n}$. In order to benefit from the independent stochastic fluctuations of distinct emitters appearing at different frames, we acquire more than one frame ($T > 1$) at each angle $\{\alpha_n\}_{n=1}^{N_n}$, similarly to what we have done in Chapter 2. By acquiring a sequence of $T$ images $\{\mathbf{y}_{\alpha_n,t} \in \mathbb{R}^{M^2}\}_{t=1}^{T}$, for each angle $\alpha_n$, and by exploiting the independence of the random fluctuations of individual fluorescent molecules, we thus aim to find $\mathbf{x}_{\alpha_n} \in \mathbb{R}^{L^2}$, the laterally super-resolved image for each angle $\{\alpha_n\}_{n=1}^{N_n}$. Mathematically, the problem reads:

$$\text{find} \quad \mathbf{x}_{\alpha_n} := \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{\alpha_n,t} \quad \text{s.t.} \quad \mathbf{y}_{\alpha_n,t} = \mathbf{\Psi}\mathbf{x}_{\alpha_n,t} + \mathbf{b}_{\alpha_n} + \mathbf{n}_{\alpha_n,t}, \quad \forall t = \{1,\dots.T\}, \tag{3.10}$$

where $\mathbf{\Psi} \in \mathbb{R}^{M^2\times L^2}$ is a matrix performing similarly to the composition of the operators $M_q$ and $U$, while the background $\mathbf{b}_{\alpha_n}$ is considered temporally constant.

## 3.3 3D MA-TIRF COL0RME

In this section, we describe how the two ill-posed inverse problems (3.9) and (3.10) can be solved in a sequential way by means of appropriate sparse regularisation models. Note that by solving (3.10), a super-resolved image $\hat{\mathbf{x}}_{\alpha_n}$ can be estimated for each incident angle $\alpha_n$ of the illumination beam. Those images serve as input data for problem (3.9) where the objective consists in finding the desired 3D image $\hat{\mathbf{f}}$ with improved spatial and axial resolution.

For solving (3.10) we use COL0RME, the approach presented in Chapter 2. For solving (3.9) we use the 3D MA-TIRF reconstruction algorithm proposed in previous works [Soubies et al., 2019b, Soubies et al., 2019a]. The overall method is called 3D MA-TIRF COL0RME.

### 3.3.1 Support and intensity estimation via COL0RME

We briefly recall that the main idea of COL0RME consists in exploiting the temporal and spatial independence of the fluorescent emitters by providing a sparse approximation of their second-order statistics. A formulation of problem (3.10) in the covariance domain is the following (see Section 2.1.2 for more details):

$$\text{find} \quad \mathbf{r}_{\mathbf{x}}^{\alpha_n}, s^{\alpha_n} \quad \text{s.t.} \quad \mathbf{r}_{\mathbf{y}}^{\alpha_n} = (\mathbf{\Psi} \odot \mathbf{\Psi})\mathbf{r}_{\mathbf{x}}^{\alpha_n} + s^{\alpha_n}\mathbf{v}_{\mathbf{I}}, \quad \forall n \in \{1, \dots, N_n\}, \tag{3.11}$$

where, for all angles $\alpha_n$, $\mathbf{r}_{\mathbf{y}}^{\alpha_n} \in \mathbb{R}^{M^4}$ is the vectorized form of the covariance matrix of the raw data $\{\mathbf{y}_{\alpha_n,t}\}_{t=1}^{T}$, $\mathbf{r}_{\mathbf{x}}^{\alpha_n} \in \mathbb{R}^{L^2}$ is the vector of the auto-covariances of the high-resolution images $\{\mathbf{x}_{\alpha_n,t}\}_{t=1}^{T}$, $s^{\alpha_n} \in \mathbb{R}_+$ is the (unknown) noise variance, $\odot$ denotes the Khatri-Rao product and $\mathbf{v}_{\mathbf{I}} \in \mathbb{R}^{M^4}$ is the vectorized form of the identity matrix $\mathbf{I}_{\mathbf{M^2}}$.

For each angle, by finding $\mathbf{r}_{\mathbf{x}}^{\alpha_n}$, we have access to the support of the laterally super-resolved image $\hat{\mathbf{x}}_{\alpha_n}$ as the two images $\mathbf{r}_{\mathbf{x}}^{\alpha_n}$ and $\hat{\mathbf{x}}_{\alpha_n}$ share the same support. However, due to the MA-TIRF setup, the support of the entire sample can be found by solving (3.11) only in correspondence with the angle ($\tilde{\alpha}$) closest to the critical one, as it will contain information of molecules located in the whole depth of investigation. This corresponds to a significant computational gain as (3.11) needs thus to be solved only once, for the angle $\tilde{\alpha}$. By denoting with $\tilde{\mathbf{r}}_{\mathbf{y}}$ the covariance matrix associated to $\{\mathbf{y}_{\tilde{\alpha},t}\}_{t=1}^{T}$, the following problem can be thus considered instead of solving (3.11):

$$\text{find} \quad \tilde{\mathbf{r}}_{\mathbf{x}}, \tilde{s} \quad \text{s.t.} \quad \tilde{\mathbf{r}}_{\mathbf{y}} = (\mathbf{\Psi} \odot \mathbf{\Psi})\tilde{\mathbf{r}}_{\mathbf{x}} + \tilde{s}\mathbf{v}_{\mathbf{I}}, \tag{3.12}$$

or in terms of a minimization problem:

$$(\hat{\mathbf{r}}_{\mathbf{x}}, \hat{s}) \in \underset{\tilde{\mathbf{r}}_{\mathbf{x}} \in \mathbb{R}_+^{L^2},\ \tilde{s} \in \mathbb{R}_+}{\arg\min} \left( \frac{1}{2}\|\tilde{\mathbf{r}}_{\mathbf{y}} - (\mathbf{\Psi} \odot \mathbf{\Psi})\tilde{\mathbf{r}}_{\mathbf{x}} - \tilde{s}\mathbf{v}_{\mathbf{I}}\|_2^2 + \mathcal{R}(\tilde{\mathbf{r}}_{\mathbf{x}}; \lambda) \right), \tag{3.13}$$

where $\lambda > 0$ is a regularization parameter, $\mathcal{R}(\cdot; \lambda)$ is a sparsity-promoting penalty, that could be either the $\ell_1$-norm or the continuous exact relaxations of the $\ell_0$ (CEL0) [Soubies et al., 2015] penalty. Numerically, problem (3.13) can be solved as discussed in Section 2.2.

Intensity can now be estimated using the second step of COL0RME as discussed in Section 2.3. In our scenario, for each angle $\alpha_n$, $n \in \{1, ..., N_n\}$, the mean intensity image $\hat{\mathbf{x}}_{\alpha_n}$ restricted to the estimated support $\Omega$ and the smoothly varying background $\hat{\mathbf{b}}_{\alpha_n}$ can be estimated from the

empirical temporal mean of the acquired stack $\{\overline{\mathbf{y}}_{\alpha_n}\}_{n=1}^{N_n}$ by solving:

$$(\hat{\mathbf{x}}_{\alpha_n}, \hat{\mathbf{b}}_{\alpha_n}) \in \underset{\mathbf{x}_{\alpha_n}\in\mathbb{R}_+^{|\Omega|},\, \mathbf{b}_{\alpha_n}\in\mathbb{R}_+^{M^2}}{\arg\min} \left(\frac{1}{2}\|\mathbf{\Psi\Omega x}_{\alpha_n} - (\overline{\mathbf{y}}_{\alpha_n} - \mathbf{b}_{\alpha_n})\|_2^2 + \frac{\mu}{2}\|\nabla_\Omega \mathbf{x}_{\alpha_n}\|_2^2 + \frac{\beta}{2}\|\nabla \mathbf{b}_{\alpha_n}\|_2^2\right), \quad \forall n \in \{1, \ldots, N_n\}, \quad (3.14)$$

where the parameter $\mu$ is automatically estimated via discrepancy principle (see Section 2.3.2) and $\beta > 0$ does not require very fine tuning. Problem (3.14) can be solved efficiently via (proximal) gradient-type algorithms (see Section 2.3 for more details).

### 3.3.2 MA-TIRF reconstruction

Having all the estimated COL0RME images $\{\hat{\mathbf{x}}_{\alpha_n}\}_{n=1}^{N_n}$ at hand, we can use them to solve the problem (3.9). To estimate the 3D super-resolved image $\hat{\mathbf{f}} \in \mathbb{R}^{L^2\times N_z}$, we follow [Soubies et al., 2019b, Soubies et al., 2019a] and look for solutions of

$$\hat{\mathbf{f}} \in \underset{\mathbf{f}\in\mathbb{R}_+^{L^2\times N_z}}{\arg\min} \left(\sum_{n=1}^{N_n} \frac{1}{2}\|\mathbf{W}_{\alpha_n}\mathbf{f} - \hat{\mathbf{x}}_{\alpha_n}\|_2^2 + \kappa\mathcal{R}(\mathbf{f})\right), \quad (3.15)$$

where $\{\hat{\mathbf{x}}_{\alpha_n}\}_{n=1}^{N_n}$ are the super-resolved COL0RME images, $\mathbf{W}_{\alpha_n} : \mathbb{R}^{L^2 N_z \times L^2}$ is the discrete TIRF matrix related to the angle $\alpha_n$, $\kappa > 0$ is the regularization parameter and $\mathcal{R}(\cdot)$ is the regularization penalty.

Different types of regularization function $\mathcal{R}$ can be used. Two good choices of regularizers according to [Soubies et al., 2019b, Soubies et al., 2019a] are the Hessian Shatten-norm of order one proposed by [Lefkimmiatis et al., 2013] or the Total-Variation (TV) regularization [Rudin et al., 1992], depending on whether Hessian- or gradient-sparsity is desired, respectively. In our work, we use the TV penalty, which performs well in our experiments and benefits from faster computations since it computes first-order finite differences along the three dimensions and not second-order, which is the case if the Hessian operator is applied to an image.

To solve (3.15) we use the code provided online [1] linked to the articles [Soubies et al., 2019b, Soubies et al., 2019a]. To be more precise, an extension of the well-known Alternate Directions Method of Multipliers (ADMM) [Boyd et al., 2011] is used, since in (3.15) the sum of more than two functions is minimized. The third function is the indicator function $i_{\geq 0}(\mathbf{f})$ defined as $i_{\geq 0}(\mathbf{f}) = \{0 \text{ if } \mathbf{f} \in \mathbb{R}_+^{L^2\times N_z}, +\infty \text{ otherwise}\}$, which appears because of the restriction that $\mathbf{f}$ must be positive.

## 3.4 Results

We apply 3D MA-TIRF COL0RME on both simulated and real data acquired by a MA-TIRF microscope. The simulated data allow us to evaluate the performance of the method since we have access to the ground truth 3D images which we can compare with the reconstructed ones. A more careful treatment was needed for the processing of real data.

1. https://github.com/esoubies/MA-TIRF_Reconstruction

### 3.4.1 Simulated data

We start by applying 3D MA-TIRF COL0RME on simulated tubulin structures. To simulate the data, we first set the 3D spatial pattern (see Figure 3.4a) using the SMLM 2016 MT0 microtubules dataset that is available online [2]. Afterwards, we simulate the temporal fluctuations using the SOFI simulation tool [Girsault et al., 2016]. The fluctuations' parameters are chosen as: 20 ms for an on-state average lifetime, 40 ms for an off-state average lifetime, 35 s for an average time until bleaching (so that little bleaching, around 18% is practically observed) and frame rate of 100 frames per second (fps).

Model (3.5) is used for the generation of the simulated data. The physics behind the MA-TIRF microscope is now exploited so that, depending on the angle of incidence of the illuminating beam, a different part of the sample is illuminated and with varying intensity of the evanescent field at the various $z$ levels. For five different incident angles $\{\alpha_n\}_{n=1}^5$ of the illumination beam, with $\alpha_c < \alpha_1 < \cdots < \alpha_5$, a stack of 500 frames is simulated. The PSF used has an FWHM of approximately 229nm while the pixel size is chosen to be equal to 100nm. The images have a size of $64 \times 64$ pixels. The spatially varying background was added to the acquisitions, signal-dependent Poisson noise, as well as additive Gaussian noise of signal-to-noise ratio (SNR) equal to 14.75 dB.

In Figure 3.2, a single frame of the acquired stack as well as its temporal mean (diffraction-limited image) is shown for each incident angle of the illumination beam in the first and second line, respectively. In the third line, we report the laterally super-resolved images $\{\hat{\mathbf{x}}_{\alpha_n}\}_{n=1}^5$ computed by solving COL0RME models (3.13)-(3.14) at each angle with a super-resolution factor of 4. The $\ell_1$-norm penalty was used for the support estimation. Please observe the different colormap ranges between the diffraction-limited and the reconstructed images. This is happening because the two sets of images are belonging to two different grids (coarse and fine grid respectively) and the intensity of one pixel in the coarse grid is the summation of the intensities of $q \times q$ pixels in the fine grid.

In Figure 3.3, we verify that the global intensity of the super-resolved COL0RME images $\hat{\mathbf{x}}_{\alpha_n}$ decreases exponentially with respect to the inverse of the penetration depth $p(\alpha_n)$ when $n$ increasing. The exponential relationship between these two quantities comes from (3.2). The same behavior is observed, as expected, for the global intensity of the diffraction-limited images $\{\bar{\mathbf{y}}_{\alpha_n}\}_{n=1}^5$. The small bias observed between the two intensity curves is expected due to the absence of background in $\{\hat{\mathbf{x}}_{\alpha_n}\}_{n=1}^5$. This figure confirms that we can use $\{\hat{\mathbf{x}}_{\alpha_n}\}_{n=1}^5$ in problem (3.15), aiming to recover the 3D super-resolved image $\hat{\mathbf{f}}$.

In Figure 3.4a we show the 3D ground-truth (GT) image and in Figure 3.4b the 3D reconstruction $\hat{\mathbf{f}}$ obtained by solving (3.15) using $\{\hat{\mathbf{x}}_{\alpha_n}\}_{n=1}^5$ as input. Note that different colors here denote different depth levels. For comparison, we further plot in Figure 3.4c the result of the MA-TIRF approach (without deconvolution) previously considered in [Soubies et al., 2019b, Soubies et al., 2019a] that estimates an axially super-resolved image. We can clearly observe that the latter, compared to the 3D MA-TIRF COL0RME reconstruction shows poor spatial resolution, as expected.

2. http://bigwww.epfl.ch/smlm/datasets/index.html

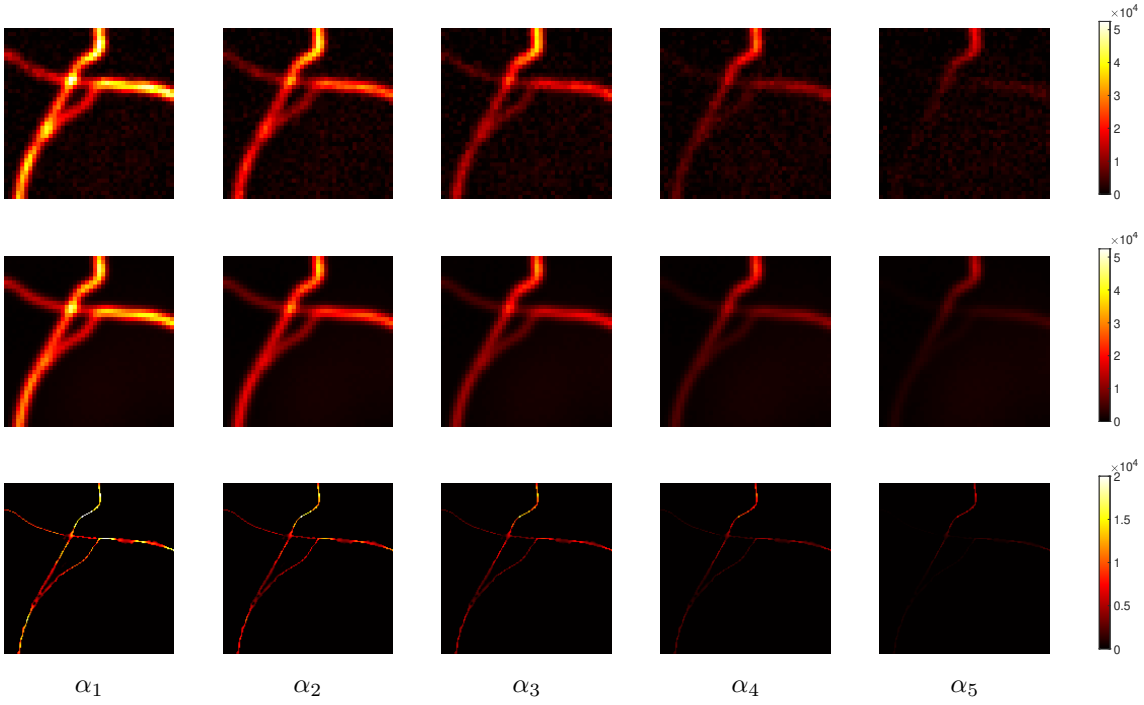



Figure 3.2 – For each $\{\alpha_n\}$, $n = 1, \ldots, 5$: (first row) One frame of the acquired stack $\mathbf{y}_{\alpha_n,t}$, (second row) the temporal mean $\bar{\mathbf{y}}_{\alpha_n}$, (third row) the laterally super-resolved COL0RME images $\hat{\mathbf{x}}_{\alpha_n}$.

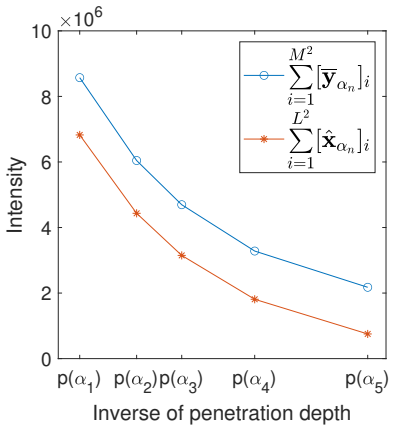
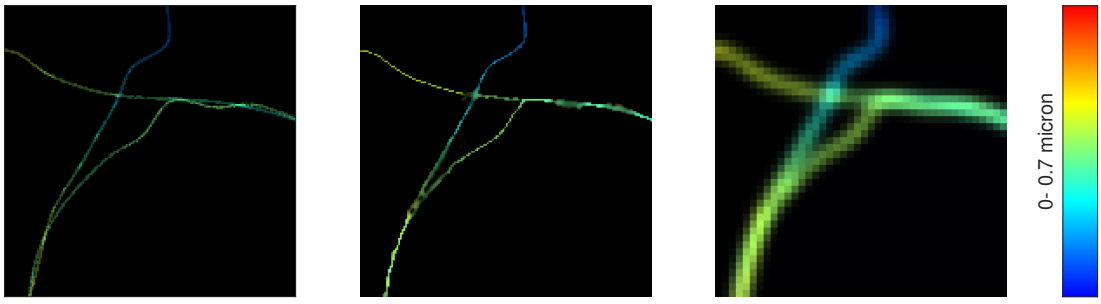



Figure 3.3 – Exponential decay of the global intensity of the diffraction-limited and the super-resolved COL0RME images with respect to the inverse of the penetration depth $\{p(\alpha_n)\}_{n=1}^{5}$, see (3.2).

### 3.4.2 Real MA-TIRF data

We now apply 3D MA-TIRF COL0RME on a dataset of images of tubulins that are acquired by a real MA-TIRF microscope. The tubulins imaged are found in bovine aortic endothelial (BAE)




(a) (b) (c)

Figure 3.4 – (a) Ground truth 3D tubulin image, (b) Super-resolved 3D MA-TIRF COL0RME reconstruction (c) MA-TIRF reconstruction result. Colour quantifies sample depth.

cells and are stained with the Alexa Fluor 488. The same data but with a fixed angle close to the critical one was used in the experiments of Chapter 2.

The custom MA-TIRF microscope that was used for the real data acquisition is developed at the Institute de Biologie Valrose (iBV) in Nice and is one of 3 MA-TIRF prototypes in France. It benefits from the rapid change in the incident angle of the excitation light thanks to dedicated electronics, i.e. fast galvanometric mirrors, which allows it to record MA-TIRF data for 3D reconstruction very fast, making it compatible with live cell imaging. More information on the microscope setup is available in [Soubies et al., 2019b].

A sequence of $T = 500$ frames is acquired for each of five different angles $\alpha_n, q = 0, \dots, 4$ of the illumination beam. Since the first angle is smaller than the critical angle or similarly $\alpha_0 < \alpha_c$, no propagation of the evanescent wave is observed for it. However, such an angle is used only for a more precise support estimation. Namely, as a final support we consider the superposition of the supports estimated for angles $\alpha_0$ and $\alpha_1$. Some additional details on the acquisitions are: the total acquisition time of the whole dataset is about 2 min, the pixel size of the CCD camera used is 106 nm, the images have a size of $144 \times 144$ pixel, and the FWHM of the PSF was measured experimentally and is equal to 292.03 nm.

In the first row of Figure 3.5 the diffraction-limited images $\{\bar{\mathbf{y}}_{\alpha_n}\}_{n=1}^4$ are shown, while in the second row, we report the laterally super-resolved COL0RME images $\{\hat{\mathbf{x}}_{\alpha_n}\}_{n=1}^4$ using a super-resolution factor $q = 4$. Figure 3.6a reports the 3D MA-TIRF COL0RME image in a 2D representation using color to quantify the depth of the sample, while Figure 3.6b reports the 3D representation of the same image in gray scale. Finally, Figure 3.7 shows the 3D reconstructions computed by both the 3D MA-TIRF COL0RME method in the upper left part and the standard MA-TIRF approach with only background removal [Soubies et al., 2019a] in the lower right part. Both methods achieve very good levels of axial resolution, but 3D MA-TIRF COL0RME also achieves very good levels of lateral resolution and is able to separate nearby filaments that would otherwise not be visible.

## 3.5 Discussion

In this chapter, we presented a 3D super-resolution approach to improve the spatial resolution of a MA-TIRF microscope over the three dimensions. Differently from state-of-the-art ap-

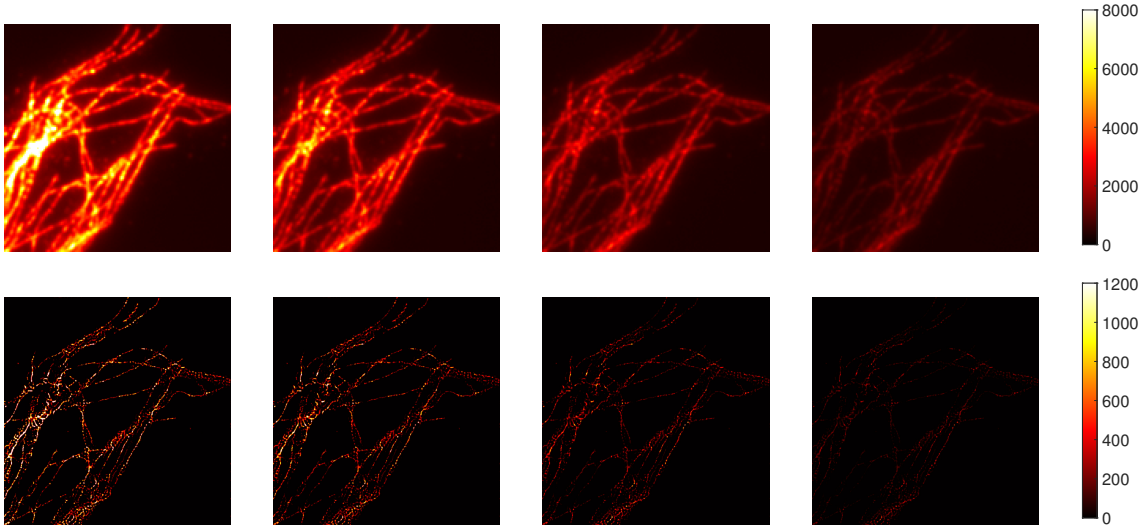
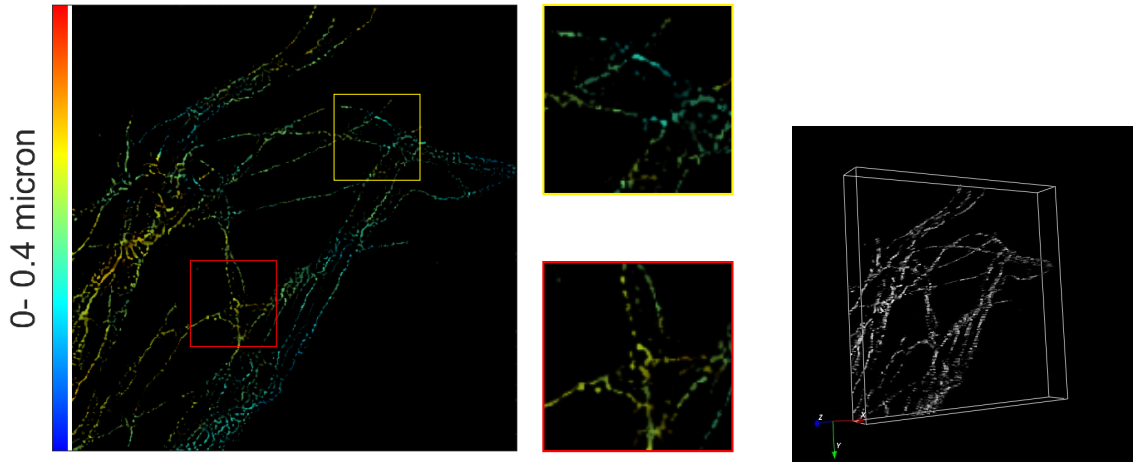



Figure 3.5 – For each $\alpha_n, n = 1, \ldots, 4$ of the illumination beam: (first row) The temporal mean of the stack $\bar{\mathbf{y}}_{\alpha_n}$, (second row) the 2D COL0RME results ($\hat{\mathbf{x}}_{\alpha_n}$).

proaches, the approach called 3D MA-TIRF COL0RME does not require the use of special fluorophores, while it is ideal for live cell imaging.

In short, 3D MA-TIRF COL0RME improves both lateral and axial resolution by combining sparsity-based modeling for accurate molecule localization and intensity estimation in the lateral plane with a 3D reconstruction procedure in the axial direction using MA-TIRF. To improve the lateral resolution, the COL0RME approach detailed in Chapter 2 is used, while to improve the resolution in the axial direction, a MA-TIRF reconstruction algorithm performing depth estimation based on a variational formulation of the inverse problem is employed.

The proposed methods have been validated on simulated MA-TIRF blinking-type data and on challenging real MA-TIRF acquisitions, showing significant resolution improvements. Comparisons with the pre-existing MA-TIRF reconstruction [Soubies et al., 2019b, Soubies et al., 2019a] show significant improvements in spatial resolution allowing accurate localization of molecules in all three dimensions. In addition, real intensity values can be estimated, a quantity of fundamental interest in the analysis of biological images.




(a) (b)

Figure 3.6 – Super-resolved 3D MA-TIRF COL0RME reconstruction. (a) 2D representation of the 3D image (color quantifies sample depth), (b) Gray-scale 3D representation.

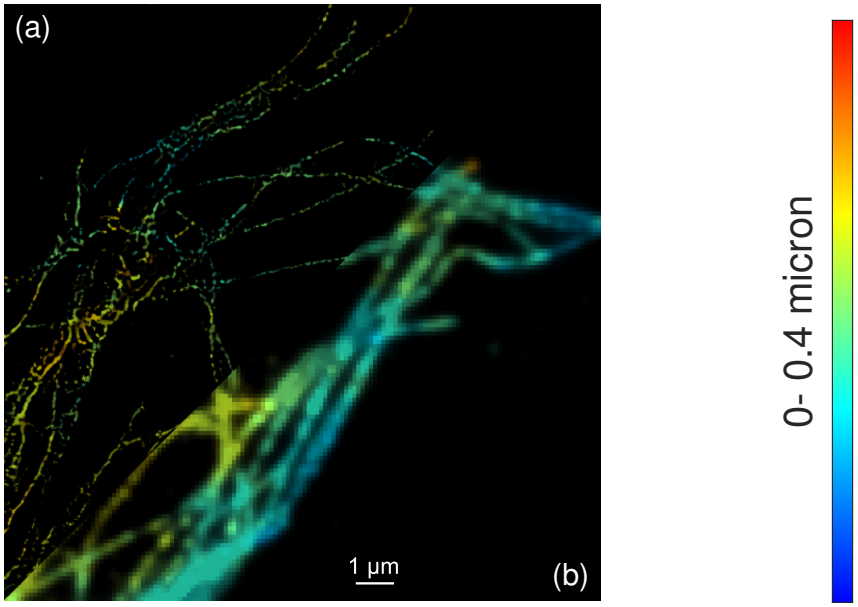



Figure 3.7 – (a) Super-resolved 3D MA-TIRF COL0RME image, (b) Standard MA-TIRF reconstruction. Color quantifies sample depth.

# PART

# Model-aware Learning Approaches

# CHAPTER 4

# Fluorescence image deconvolution by model-aware generative approaches

*In this chapter we present two model-aware generative approaches, which we named FluoGAN and DivBlurring, for fluorescence image deconvolution. To begin with, in Section 4.1 we introduce the main ideas of GANs and VAEs, while in Section 4.2 we present state-of-the-art model-aware generative methods for addressing inverse problems in imaging. Then, FluoGAN and DivBlurring are presented in detail in Sections 4.3 and 4.4, respectively. Finally, in Section 4.5 a discussion for the proposed hybrid approaches is given.*

---

**Related Publication for method FluoGAN:**

[Cachia et al., 2023]
M. Cachia, V. Stergiopoulou, L. Calatroni, S. Schaub, L. Blanc-Féraud. "Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN)," Inverse Problems, 39(5):054006, 2022, doi: 10.1088/1361-6420/acc889

The work presented in the second part of this chapter (method DivBlurring) was carried out in collaboration with Sai Muttavarapu as part of his Master 2 thesis project under my supervision.

## 4.1 Introduction to Generative Approaches

During the last few years, Deep Learning (DL) approaches have been used a lot to solve inverse problems in biological imaging, providing good results and very fast inference time compared to traditional regularized least square methods. Among them, model-aware generative methods have attracted attention due to their unsupervised nature, i.e. they do not require pairs of corrupted/ground-truth images, and due to their non-standard ability to learn the distribution of the given data that is proven useful for solving inverse problems. Popular generative approaches are the Generative Adversarial Networks (GANs) and the Variational Auto-Encoders (VAEs). For a review of generative approaches see [Bond-Taylor et al., 2022, Goodfellow et al., 2016].

### 4.1.1 Generative Adversarial Networks (GANs)

Generative Adversarial Networks or GANs, first introduced by Goodfellow *et al.* [Goodfellow et al., 2014], are based on a competitive interplay between two neural networks. The first network, the generator $G_{\psi}$ with parameters $\psi$, produces samples $\mathbf{y}^{\text{sim}} = G_{\psi}(\mathbf{z})$ from an input noise variable $\mathbf{z}$ uniformly distributed in $[0, 1]$, as shown in Figure 4.1. The second network, the discriminator $D_{\varphi}$ with parameters $\varphi$, tries to distinguish between samples drawn from the training data and samples drawn from the generator. To do so, it computes a probability value $D_{\varphi}(\mathbf{y})$, a scalar within the interval $[0, 1]$, that indicates the probability of $\mathbf{y}$ being a real sample (i.e. $\mathbf{y}$ belongs to the training set) rather than a fake sample (i.e. it is an output of the generator). Therefore, $D_{\varphi}(\mathbf{y}) \approx 1$ denoting that $\mathbf{y}$ is drawn from the distribution $p^{real}$ of the training data with high probability, while $D_{\varphi}(\mathbf{y}) \approx 0$ meaning that $\mathbf{y}$ is likely to be a sample created by the generator.

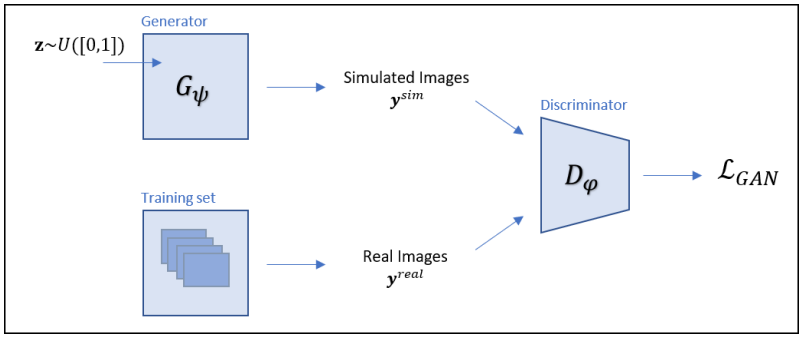



Figure 4.1 – Illustration of a GANs architecture.

The two networks compete against each other during the training process. The generator wants to trick the discriminator, while the discriminator tries not to be fooled. The simplest way to formulate the learning process in GANs is by means of a zero-sum game. The discriminator wants to make sure that its decision over real data $\mathbf{y}^{\text{real}} \sim p^{real}$ is correct by maximizing $\mathbb{E}_{\mathbf{y}^{\text{real}}}\{\log D_{\varphi}(\mathbf{y}^{\text{real}})\}$, while it outputs a probability close to zero when exposed to fake data $\mathbf{y}^{\text{sim}} \sim p^{sim}$, so maximizing $\mathbb{E}_{\mathbf{y}^{\text{sim}}}\{\log(1 - D_{\varphi}(\mathbf{y}^{\text{sim}}))\}$, with $\mathbf{y}^{\text{sim}} = G_{\psi}(\mathbf{z})$. On the other side, the generator is trained to increase the chances of $D_{\varphi}$ making a mistake and producing a fake sample with high probability, thus minimizing $\mathbb{E}_{\mathbf{y}^{\text{sim}}}\{\log(1 - D_{\varphi}(\mathbf{y}^{\text{sim}}))\}$.

We have, therefore, the following minimax game:

$$\boldsymbol{\psi}^* = \arg\min_{\boldsymbol{\psi}} \max_{\boldsymbol{\varphi}} \mathcal{L}_{\text{GAN}}(\boldsymbol{\psi}, \boldsymbol{\varphi}), \tag{4.1}$$

where the loss function $\mathcal{L}_{\text{GAN}}(\boldsymbol{\psi}, \boldsymbol{\varphi})$ is defined by:

$$\mathcal{L}_{\text{GAN}}(\boldsymbol{\psi}, \boldsymbol{\varphi}) = \mathbb{E}_{\mathbf{y}^{\text{real}}}\{\log D_{\boldsymbol{\varphi}}(\mathbf{y}^{\text{real}})\} + \mathbb{E}_{\mathbf{y}^{\text{sim}}}\{\log(1 - D_{\boldsymbol{\varphi}}(\mathbf{y}^{\text{sim}}))\} \tag{4.2}$$

Such an adversarial learning scheme progressively drives the generative model $G_{\boldsymbol{\psi}}$ to capture the unknown statistical distribution of the training data. We can see the whole problem as a minimization of a suitable distance between $p^{sim}$, the distribution of the simulated data, and $p^{real}$, the distribution of the real data. When the Jensen-Shannon divergence is used as such distance, the original GAN in the form (4.2), is obtained. As recalled in Section 4.3.2.1, however, other choices are possible such the Wasserstein distance [Arjovsky et al., 2017] which improves training stability.

### 4.1.2 Variational Auto-Encoders (VAEs)

Variational Auto-Encoders or VAEs were first introduced by Kingma & Welling [Kingma and Welling, 2014] and similarly to GANs, they fall into the class of generative models. They are able to learn unknown distributions of training data $\mathbf{y}$, using a latent variable $\mathbf{z}$ following a simple distribution $p(\mathbf{z})$.

VAEs consist of two collaborative networks: the decoder network $G_{\boldsymbol{\theta}}(\mathbf{z})$, acting like a GAN generator, that takes a point $\mathbf{z}^k$ from a lower-dimensional latent space and generates an image $\mathbf{y}^k$ following a likelihood distribution $p_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{z})$; the encoder $F_{\boldsymbol{\phi}}(\mathbf{y})$ that takes the real image $\mathbf{y}$ and encodes it to parameters $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ used to describe the distribution $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})$. For distribution $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})$, defining the latent space, is convenient to approximate it with a normal distribution, $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\mathbf{I})$. An overview of the architecture of VAEs is shown in Figure 4.2.

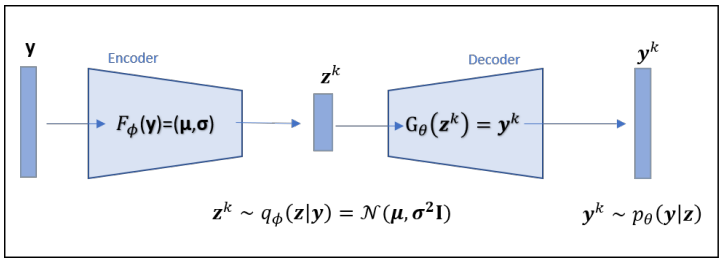



Figure 4.2 – The architecture of VAEs. The encoder network $F_{\boldsymbol{\phi}}$ outputs representation $\mathbf{z}$ of real data $\mathbf{y}$. The decoder network $G_{\boldsymbol{\theta}}$ learns to reconstruct the data $\mathbf{y}$ given a representation $\mathbf{z}$.

The generative model $p_{\boldsymbol{\theta}}(\mathbf{y})$ in VAEs is given by:

$$\log p_{\boldsymbol{\theta}}(\mathbf{y}) = \log\left(\int p_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{z})p(\mathbf{z})d\mathbf{z}\right). \tag{4.3}$$

We observe that in principle the decoder alone (together with a suitable prior $p(\mathbf{z})$) is good enough to describe the generative model in (4.3) since it is independent of the encoder parameters $\boldsymbol{\phi}$. However, as the integral in (4.3) is intractable, an analytic solution or efficient estimator

cannot be found. To compute a solution, the encoder model $q_{\phi}(\mathbf{z}|\mathbf{y})$ has to be introduced. In [Kingma and Welling, 2014] it is shown that instead of maximizing the marginal likelihood $p_{\boldsymbol{\theta}}(\mathbf{y})$, one could maximize the evidence lower bound (ELBO) $\mathcal{L}^{\text{ELBO}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$ bounding the likelihood of the data from below. Since in optimization, it is more convenient to minimize functions, the authors of [Kingma and Welling, 2014] define $\mathcal{L}^{\text{VAE}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) = -\mathcal{L}^{\text{ELBO}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$, given by:

$$\mathcal{L}^{\text{VAE}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) = \mathcal{L}^{\text{R}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) + \mathcal{L}^{\text{KL}}_{\boldsymbol{\phi}}(\mathbf{y}) \tag{4.4}$$

where $\mathcal{L}^{\text{R}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$ is the reconstruction loss defined by:

$$\mathcal{L}^{\text{R}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) = -\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})}\{\log p_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{z})\} \tag{4.5}$$

and $\mathcal{L}^{\text{KL}}_{\boldsymbol{\phi}}(\mathbf{y})$ is the Kullback-Leibler (KL) divergence loss, defined by:

$$\mathcal{L}^{\text{KL}}_{\boldsymbol{\phi}}(\mathbf{y}) = \text{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})||p(\mathbf{z})). \tag{4.6}$$

More details on the derivation of loss (4.4) are given in [Kingma and Welling, 2014]. Therefore in order to make sure that (4.3) suits the distribution of given training data $\mathbf{y}$, both parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ have to be optimized.

While $\mathcal{L}^{\text{KL}}_{\boldsymbol{\phi}}(\mathbf{y})$ can be computed analytically, the expected value defining $\mathcal{L}^{\text{R}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$ can be empirically approximated by drawing samples $\mathbf{z}^k$, with $k \geq 1$, from the distribution $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})$. However, since sampling operation cannot be differentiated, the gradient computations in the back-propagation are not possible anymore. Following the reparametrization trick proposed by [Kingma and Welling, 2014] and considering Gaussian assumption on the latent variable $\mathbf{z} \sim q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\mathbf{I})$, a valid reparameterization of $\mathbf{z}$ is $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \cdot \boldsymbol{\sigma}$, where $\boldsymbol{\epsilon}$ is an auxiliary noise variable $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\cdot$ denotes element-wise multiplication. Such reparameterization allows for efficient gradient computations.

Finally, if together with the Gaussian assumption on the latent variable, we further assume that the prior $p(\mathbf{z})$ is Gaussian we can obtain a closed form for the KL-loss $\mathcal{L}^{\text{KL}}_{\boldsymbol{\phi}}(\mathbf{y})$ that is given by:

$$\mathcal{L}^{\text{KL}}_{\boldsymbol{\phi}}(\mathbf{y}) \simeq -\frac{1}{2}\sum_{j=1}^{J}(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2) \tag{4.7}$$

where $j \in \{1, \ldots, J\}$ represents dimension of the latent vector $\mathbf{z}$ and $\sigma_j$, $\mu_j$ are the j$^{\text{th}}$ elements of vectors $\boldsymbol{\sigma}$, $\boldsymbol{\mu}$, respectively.

## 4.2 Hybrid generative methods for inverse problems

Several adaptations of the well-known GANs and VAEs can in fact lead to the solution of inverse problems in biological imaging in a fully unsupervised way since truth images are not often available in biological imaging problems. Relying on the knowledge of the underlying physical model, such adaptations are further able to reconstruct a physically meaningful solution and avoid estimating a solution that does not serve the physical model.

Recently, a hybrid approach called CryoGAN [Gupta et al., 2021], was designed to combine a physically-inspired simulator playing the role of a GAN generator for the problem of volume reconstruction from noisy projections in Cryo-Electron Microscopy (Cryo-EM). In CryoGAN the

volume of interest becomes the learnable parameter of the generator. It is estimated upon an adversarial training where, given noisy projections $\{\mathbf{y}_i\}_i$, an estimate of the desired volume at iteration $l \geq 0$ is passed through the simulator to obtain simulated noisy projections $\{\mathbf{y}_l^{\text{sim}}\}_l$ being 'close enough' to the given data in a distributional sense. The approach is proved to be very effective, allowing for an efficient volume estimation using a simple linear physical model which allows for deriving reconstruction guarantees. A similar approach is used in [Zehni and Zhao, 2021] for multi-segment reconstruction.

Another interesting approach proposed by Prakash *et al.* [Prakash et al., 2021], consists of a denoising approach for microscopy images based on fully-convolutional VAEs. The method is called Diversity Denoising (DivNoising): it is unsupervised as it explicitly includes the modeling of the noise into the decoder of a traditional VAE and overcomes the problem of having to choose a single solution by estimating a whole distribution of denoised images. In DivNoising, the authors assume that noisy images $\mathbf{y}$ are created from a clean signal $\mathbf{x}$ via a known noise model (NM), i.e., $\mathbf{y} \sim p_{\text{NM}}(\mathbf{y}|\mathbf{x})$, and try to estimate the posterior $p(\mathbf{x}|\mathbf{y}) \propto p_{\text{NM}}(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ where $p(\mathbf{x})$ is the prior. The DivNoising architecture is available in Figure 4.3. Differently from traditional VAE, $\mathbf{y}$, the input image to the encoder and the image used to produce the latent vector $\mathbf{z}$, is an image corrupted by noise. DivNoising is capable of generating not only one but several diverse samples from the approximated posterior, $\mathbf{z}^k \sim q_\phi(\mathbf{z}|\mathbf{y})$, giving rise to several estimations of denoised images $\mathbf{x}^k$. In order to compute a single final image, the authors propose to either approximate the minimum mean square error (MMSE) estimate by averaging the images $\mathbf{x}^k$ or find the maximum a posteriori (MAP) estimate.

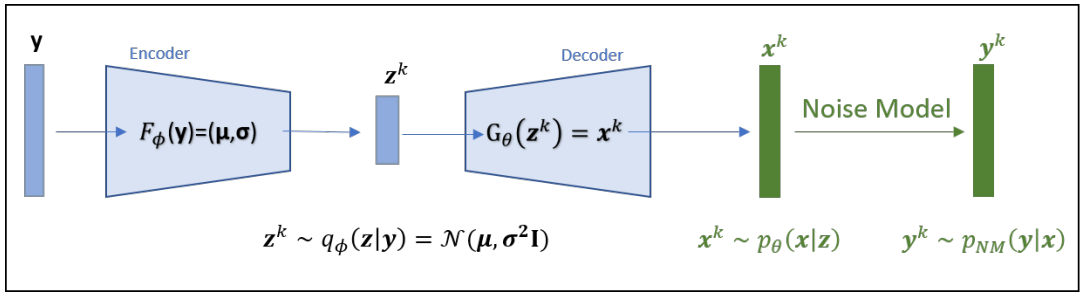



Figure 4.3 – The architecture of DivNoising VAE: explicit incorporation of the imaging noise model into the decoder.

Although we have only described in detail two approaches, there are other works in which generative models have been applied to inverse problems, see for example [Bora et al., 2017, Shah and Hegde, 2018, Bora et al., 2018, Goncharova et al., 2020, Duff et al., 2021, Goh et al., 2022].

## 4.3 FluoGAN: Fluorescence image deconvolution microscopy via GANs

We start our discussion by proposing a hybrid strategy called Fluorescent image deconvolution microscopy via GAN learning (FluoGAN). We keep the protocol of acquiring a stack of images with fluctuating molecules which we then use within an optimization procedure to find one single high-resolution image of the sample of interest (see Section 2.1.1). Here, we define a different

data fidelity term which is now given by a distance between two distributions, one given by the observed time-lapse acquisitions considered as samples of the observed (real) distribution and the other generated by a physical model of stochastic fluctuations. This distance and optimization procedure is realized in the framework of a GAN architecture where the generator network is replaced by the generative physical model (simulator).

In Section 4.3.1 we detail the physical model of fluorescence image acquisition used as a simulator, in Section 4.3.2 we revise the GAN formulation from an optimization perspective, and in Section 4.3.3 we describe in detail the modeling and optimization aspects of the proposed FluoGAN approach. Numerical results are reported in Section 4.3.4. The proposed method is validated first on realistic simulated data containing spatial patterns for assessing the resolution achieved, then on real data of a commercial calibrated slide and, finally, of *Ostreopsis cf Ovata* alga acquired by standard epifluorescence microscopes.

### 4.3.1 A non-linear model for stochastic fluctuations

We present in this section the simplified modeling used in the following as a reference simulator mimicking the physical processes creating blur and noise – including fluorophore fluctuations – in the observed image data.

We recall from (2.1) that the model in a vectorized form describing the acquisitions $\{\mathbf{y_t} \in \mathbb{R}^{M^2}\}_{t=1}^T$ by standard fluorescent microscopes of true images $\{\mathbf{x_t} \in \mathbb{R}^{L^2}\}_{t=1}^T$ belonging to a $q$-times finer grid, with $L = qM$ and $T > 1$, is given by:

$$\mathbf{y_t} = \mathcal{P}\left(\mathbf{\Psi x_t} + \mathbf{b_t}\right) + \mathbf{e_t}, \quad \forall t \in \{1, \dots, T\}, \tag{4.8}$$

where $\mathbf{\Psi} \in \mathbb{R}^{M^2 \times L^2}$ is a (known) matrix describing the composition of the operators $M_q : \mathbb{R}^{L^2} \to \mathbb{R}^{M^2}$ and $U : \mathbb{R}^{L^2} \to \mathbb{R}^{L^2}$, a down-sampling operator summing every $q$ consecutive pixels in both dimensions and a convolution operator defined by the PSF of the system, respectively. We assume $\mathbf{b}_t = \mathbf{b} \in \mathbb{R}^{M^2}$ i.e. the background image is temporally constant (see Section 2.1.1 for more details) and $\mathbf{e}_t \in \mathbb{R}^{M^2}$, the matrix describing the electronic noise, that is modeled as a realization of a multidimensional random vector with Gaussian distribution of zero mean and covariance matrix $\sigma^2 \mathbf{I}_{M^2}$, i.e. $\mathbf{e}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{M^2})$. Finally, $\mathcal{P}(\cdot)$ denotes the presence of Poisson noise.

The true images $\{\mathbf{x}_t\}_{t=1}^T$ capture the fluctuations of individual fluorophores. There are different ways to describe the fluctuations of standard emitters (e.g. the temporal fluctuations generated by the SOFI simulation tool [Girsault et al., 2016] have an "on-off" behavior). Here, the emission of a single fluorophore over time is described by a Poisson distribution, since we believe that such modeling is the closest to reality. Therefore we assume that the fluctuations in time of an image $\mathbf{x}$ can be modeled by $\mathcal{P}(\mathbf{x})$, the notation $\mathcal{P}(\mathbf{x})$ standing for the random vector drawn from a multidimensional Poisson distribution with vector parameter $\mathbf{x}$. These Poisson variables are independent of each other (in time and space). The images$\{\mathbf{x}_t\}_{t=1}^T$ are therefore considered as realizations of the Poisson random variable.

Similarly we denote $\mathbf{y}$ a random variable with realizations $\{\mathbf{y}_t\}_{t=1}^T$ and $\mathbf{e}$ a Gaussian random variable with realizations $\{\mathbf{e}_t\}_{t=1}^T$. By further incorporating a scale parameter $\alpha > 0$ describing the gain of the sensor, i.e. the amplification factor applied to the image by the image sensor, which can be easily estimated in practice and incorporated into the model, we can define the following forward observation model:

$$\mathbf{y} = \alpha\, \mathcal{P}\left(\mathbf{\Psi}\mathcal{P}(\mathbf{x}) + \mathbf{b}\right) + \mathbf{e}. \tag{4.9}$$

Due to the presence of the double Poisson distributions, model (4.9) is difficult to handle in computations. To alleviate such limitation, we make use of the following approximation model:

$$\mathbf{y} = \alpha\, \mathcal{P}\left(\mathbf{\Psi x} + \mathbf{b}\right) + \mathbf{e}, \tag{4.10}$$

which we claim to provide an acceptable approximation in the estimation process. Fluctuations of fluorophores are in fact passed through the PSF and undersampling operations, which both are processes smoothing out the fluctuations themselves. The Poisson noise model then adds further Poisson fluctuations. We thus use (4.10) in the estimation process.

The setup is composed of *generated images* which are realizations $\{\mathbf{y}_t^{\text{sim}}\}_{t=1}^T$ of the multivariate random variable $\mathbf{Y}^{\text{sim}}$ following a distribution $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ given by (4.10) with quantities $\mathbf{x}$ and $\mathbf{b}$ that will play the role of learnable parameters to be improved within the training so as to generate samples closer and closer to the measured data. In mathematical term we have:

$$\mathbf{Y}^{\text{sim}}(\mathbf{x}, \mathbf{b}) \sim p^{\text{sim}}(\mathbf{x}, \mathbf{b}) := \alpha\, \mathcal{P}\left(\mathbf{\Psi x} + \mathbf{b}\right) + \mathbf{e}. \tag{4.11}$$

### 4.3.2 The inverse problem formulation

Interpreting the given temporal sequence of $T > 1$ noisy, blurred and undersampled images $\{\mathbf{y}_t^{\text{real}}\}_{t=1}^T$ as the acquired $T$ realizations of an unknown distribution $p^{\text{real}}$, the data-driven inverse problem formulation of the problem thus reads:

$$\text{given } \left\{\mathbf{y}_t^{\text{real}}\right\}_{t=1}^T \quad \text{find } (\mathbf{x}, \mathbf{b}) \in \mathbb{R}^{L^2} \times \mathbb{R}^{M^2} \quad \text{s.t.} \quad p^{\text{sim}}(\mathbf{x}, \mathbf{b}) \sim p^{\text{real}} \tag{4.12}$$

where $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ is defined in (4.11). In other words, given the measurements $\{\mathbf{y}_t^{\text{real}}\}_{t=1}^T$, we aim at estimating learnable parameters $(\mathbf{x}, \mathbf{b})$ such that the distribution $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ is as close as possible (in a sense which has to be specified) to the unknown underlying distribution $p^{\text{real}}$.

We can thus aim at solving (4.12) by formulating the problem from an optimization perspective as the minimization of a suitable distance between $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ and $p^{\text{real}}$ over $\mathbf{x}$ and $\mathbf{b}$. Due to ill-posedness, we may further incorporate additional and physically-consistent assumptions on the desired signal $\mathbf{x}$ and the background $\mathbf{b}$, which stabilize the optimization and eliminate potential solutions without any physical meaning. We thus consider the following optimization problem:

$$\min_{\mathbf{x} \in \mathbb{R}_+^{L^2},\, \mathbf{b} \in \mathbb{R}_+^{M^2}} d(p^{\text{sim}}(\mathbf{x}, \mathbf{b}), p^{\text{real}}) + \mathcal{R}_1(\mathbf{x}) + \mathcal{R}_2(\mathbf{b}), \tag{4.13}$$

where the relevant distance $d$ comparing $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ and $p^{\text{real}}$ and the assumptions on $\mathbf{x}$ and $\mathbf{b}$ with the corresponding choice of $\mathcal{R}_1$ and $\mathcal{R}_2$ have to be specified, along with a tailored optimization algorithm.

#### 4.3.2.1 Comparing distributions

**The $\ell_2$-distance** The easiest way to compare $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ with $p^{\text{real}}$ is to compute the difference between their mean. In our modeling, we suppose that both distributions are, up to some zero-mean Gaussian noise modeling, Poisson-like distributions and, as such, characterized by their multidimensional parameter being both their expected value and variance. For the simulated image samples such parameter is thus given by $\mathbf{\Psi x} + \mathbf{b}$ and thus depend on the desired parameters $\mathbf{x}$ and

$\mathbf{b}$, but for the real ones this is of course unknown. Given a batch of images, an unbiased estimate of such value is given nonetheless by the sample average $\bar{\mathbf{y}}^{\text{real}} := \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t^{\text{real}}$. Defining similarly $\bar{\mathbf{y}}^{\text{sim}}(\mathbf{x},\mathbf{b}) := \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t^{\text{sim}}(\mathbf{x},\mathbf{b})$, we can choose:

$$d(p^{\text{sim}}(\mathbf{x},\mathbf{b}),p^{\text{real}}) = \|\mathbb{E}[\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b})] - \mathbb{E}[\mathbf{Y}^{\text{real}}]\|_2^2 \simeq \|\bar{\mathbf{y}}^{\text{sim}}(\mathbf{x},\mathbf{b}) - \bar{\mathbf{y}}^{\text{real}}\|_2^2, \tag{4.14}$$

where $\mathbf{Y}^{\text{real}}$ is the multidimensional random variable associated with the observed samples. Among the many other distances that could be here considered, the advantage of this choice is that it can be computed explicitly, so its minimization is straightforward. Of course, it has the major limitation of comparing only two mean images instead of the full datasets. As a consequence, it does not exploit all the diversity contained in the data.

A better choice improving the amount of information carried out by (4.14) consists in comparing distributions using the mean $\ell_2$ distance between individual samples drawn uniformly from a batch $B \subset \{1,\dots,T\}$ and the mean of the observed data, i.e.:

$$d(p^{\text{sim}}(\mathbf{x},\mathbf{b}),p^{\text{real}}) = \mathbb{E}\left[\|\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b}) - \mathbb{E}[\mathbf{Y}^{\text{real}}]\|^2\right] \simeq \frac{1}{|B|}\sum_{t\in B}\|\mathbf{y}_t^{\text{sim}}(\mathbf{x},\mathbf{b}) - \bar{\mathbf{y}}^{\text{real}}\|_2^2 \tag{4.15}$$

where $\{y_{\text{sim}}^t(x,b)\}_{t\in B}$ being a set of realisations of the random variable $\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b})$.

**From KL divergence to Wasserstein distance** The use of the Kullback-Leibler (KL) divergence to compare two probability distributions is rather classical. For two absolutely continuous probability distributions with respect to a measure $\mu$ on $\Omega$ and with densities $p_1$ and $p_2$, such quantity is defined by:

$$d(p_1,p_2) = \int_\Omega \log\left(\frac{p_1(w)}{p_2(w)}\right)p_2(w)d\mu(w).$$

In their vanilla form, GANs [Goodfellow et al., 2014] rely on the asymptotic minimization of the weighted sum of such quantity (called Jensen–Shannon Divergence) to minimize the distance between $p^{\text{sim}}$ and $p^{\text{real}}$ (see (4.2)). However, it is often not possible to compute this divergence explicitly in cases when the two distributions have disjoint support and/or when either of the two densities is unknown, as it happens in our case for $p^{\text{real}}$. When it is possible, moreover, such quantity turns out to be often equal to infinity [Arjovsky et al., 2017].

Another choice that allows overcoming such limitations is inspired by the field of optimal transport with the so-called 1-Wasserstein ($W_1$) functional also called as *earth-mover distance*, see, e.g., [Ambrosio et al., 2005] for a survey. Its functional form reads:

$$d(p_1,p_2) = \inf_{\gamma\in\Pi(p_1,p_2)}\mathbb{E}_{(\mathbf{u},\mathbf{v})\sim\gamma}\left[\|\mathbf{u}-\mathbf{v}\|\right], \tag{4.16}$$

where $\Pi(p_1,p_2)$ stands for the set of all joint distributions $\gamma(\mathbf{u},\mathbf{v})$ with marginals equal to $p_1$ and $p_2$, respectively. Intuitively, the $W_1$ distance can be understood as the cost corresponding to an optimal transport plan a worker should do to transform one earth heap into another with a different position or shape. Contrarily to the KL divergence, such distance is well defined also for distributions with disjoint support as it depends on the Euclidean distance between supports. The $W_1$ distance is therefore particularly suited for comparing distributions, as it was shown in [Arjovsky et al., 2017]. Note, however, that is in general not possible to compute this distance explicitly although it is sometimes possible to define relaxed optimization problems corresponding to the minimization of $W_1$ [Gulrajani et al., 2017]. This point is discussed in Section 4.3.3 below.

#### 4.3.2.2 Regularizations

Recalling (4.13), it is now beneficial to provide some rationale on how to choose the regularization terms $\mathcal{R}_1$ and $\mathcal{R}_2$ favoring *a-priori* regularity on the desired super-resolved image $\mathbf{x}$ and on the background image $\mathbf{b}$, respectively.

As discussed in Section 1.3.3.3, we can for instance enforce sparsity by using the convex $\ell_1$-norm. Therefore, we set $\mathcal{R}_1(\mathbf{x}) = \lambda_1 \|\mathbf{x}\|_1$ and deal with the non-smoothness associated with such choice by means of proximal-gradient algorithms, (see [Parikh and Boyd, 2014] for a review). As far as the background $\mathbf{b}$ is concerned, we can choose a gradient-smoothing term, i.e. $\mathcal{R}_2(\mathbf{b}) = \frac{\lambda_2}{2}\|\nabla \mathbf{b}\|_2^2$.

Note that both signal sparsity and background smoothness are indeed two reasonable assumptions also made in the method COL0RME presented in Chapter 2.

### 4.3.3 FluoGAN

We aim to define a generative adversarial learning approach to combine distribution fit between $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ and $p^{\text{real}}$, with regularization. Recalling (4.13), we observe that this problem is closely related to a GAN-type problem since it involves the minimization of a distance between two distributions. However, two significant differences have to be highlighted:

1. Since we have formulated in (4.11) the physical model of fluorescence fluctuations, our knowledge of the unknown random variable $\mathbf{Y}^{\text{sim}}(\mathbf{x}, \mathbf{b})$ depending on the two quantities of interest $\mathbf{x}$ and $\mathbf{b}$ does not start from scratch. Instead of a generator network and inspired by previous works [Gupta et al., 2021, Zehni and Zhao, 2021], it thus appears natural to consider (4.11) as a simulator of the direct problem. The input learnable parameters $(\mathbf{x}, \mathbf{b}) \in \mathbb{R}_+^{L^2} \times \mathbb{R}_+^{M^2}$ of this "network" are no longer parameters of a neural network, but the quantities of interest we want to retrieve.
2. A GAN is usually trained to get, at convergence, a satisfying generator that is used to create data distributed like the training ones. In our approach, the goal is different since we look into a way of producing fake blurred, noisy, and undersampled data by means of the given simulator, so that when $p^{\text{sim}}(\mathbf{x}, \mathbf{b})$ matches $p^{\text{real}}$, the corresponding $\mathbf{x}$ and $\mathbf{b}$ are the desired solutions of the inverse problem (4.12).

We call FluoGAN (Fluorescence image deconvolution microscopy via generative adversarial learning) the proposed method, given the particular applicative context considered here. It is therefore inspired by GANs but it has the goal of solving an ill-posed inverse problem given a training set of measured images and by exploiting the underlying physical knowledge. The FluoGAN architecture is reported in Figure 4.4 alongside with the regular GAN architecture for comparison. In both cases, the generator $G_{\psi}$ (or, equivalently, the simulator) is trained in an adversarial manner with the discriminator $D_{\varphi}$. The proposed method can be seen as a specialized GAN architecture where the additional knowledge available on the direct problem allows us to replace $G_{\psi}$ with a more interpretable physics-based function. FluoGAN compares the empirical distribution of the real images with the output distribution of the samples generated by $G_{\psi}$. Then, it computes gradients of suitable loss functions minimizing such distance and updates accordingly the discriminator parameters and simulator inputs.

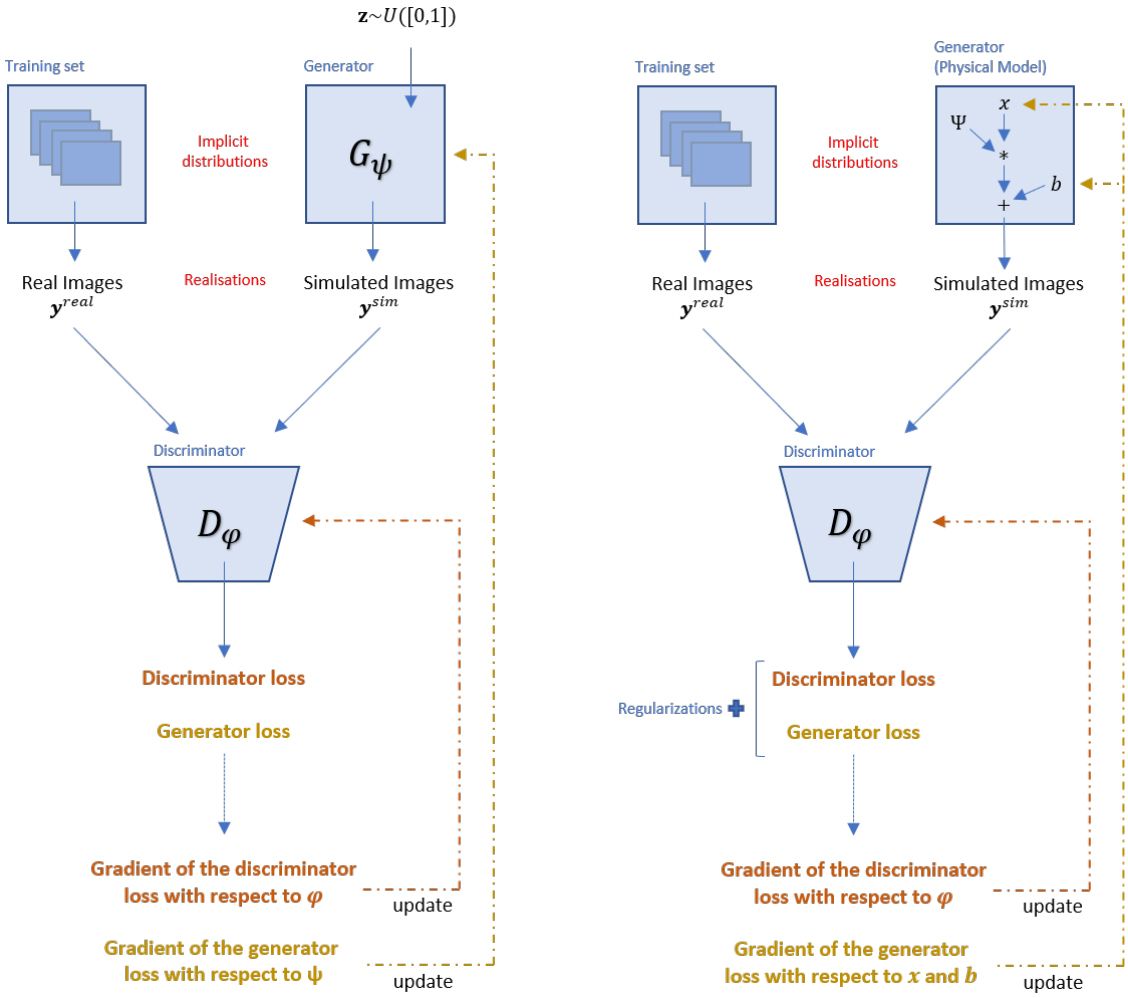



Figure 4.4 – Comparison between the architecture of a regular GAN (left) and FluoGAN (right).

#### 4.3.3.1 Formulation as an optimization problem

Inspired by the GAN modeling, we formulate the adversarial training process as a competing procedure defined in terms of two functions to be optimized alternatively and corresponding to the optimization process of the generator and the discriminator, respectively. We note that the minimax game between the generator and the discriminator described in Section 4.1.1, is done by alternately minimizing the loss function defined for each of them.

To maintain the same notation as for GANs, we will denote in the following by $G_{\mathbf{x},\mathbf{b}}$ the physically-inspired generator (i.e. the simulator) providing for a given input $(\mathbf{x}, \mathbf{b})$ an output $\mathbf{y}^{\text{sim}}$ being the realization of the random variable $\mathbf{Y}^{\text{sim}}(\mathbf{x}, \mathbf{b})$ in (4.11). We consider the problem:

$$\min_{\mathbf{x},\mathbf{b}} \mathbb{E}\Big[\mathcal{L}_{G_{\mathbf{x},\mathbf{b}}}(\mathbf{x}, \mathbf{b}, \varphi)\Big], \tag{4.17}$$

$$\min_{\varphi} \mathbb{E}\Big[\mathcal{L}_{D_\varphi}(\mathbf{x}, \mathbf{b}, \varphi)\Big]. \tag{4.18}$$

In such formulation:

— $\mathcal{L}_{G_{\mathbf{x},\mathbf{b}}}$ is a loss term enforcing equality in distribution between $p^{\text{real}}$ and $p^{\text{sim}}(\mathbf{x},\mathbf{b})$ in some sense. Recalling (4.15) we can choose for instance:

$$\mathcal{L}_{G_{\mathbf{x},\mathbf{b}}}(\mathbf{x},\mathbf{b},\boldsymbol{\varphi}) := \|\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b}) - \mathbb{E}[\mathbf{Y}^{\text{real}}]\|_2^2 - D_{\boldsymbol{\varphi}}(\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b})). \tag{4.19}$$

— $\mathcal{L}_{D_{\boldsymbol{\varphi}}}$ is a loss term which should enforce the discriminator to be maximized (i.e. $D_{\boldsymbol{\varphi}} \approx 1$) on real images, while it should be minimized ($D_{\boldsymbol{\varphi}} \approx 0$), on simulated images. A natural choice would thus be:

$$\mathcal{L}_{D_{\boldsymbol{\varphi}}}(\mathbf{x},\mathbf{b},\boldsymbol{\varphi}) := D_{\boldsymbol{\varphi}}(\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b})) - D_{\boldsymbol{\varphi}}(\mathbf{Y}^{\text{real}})$$

To avoid the well-known convergence instabilities of such adversarial training, it is rather classical to further introduce a regularisation term called *gradient penalty*, to promote gradient updates with a norm close to 1, see [Gulrajani et al., 2017]. This is in fact a penalized formulation of the constraint on the discriminator to be 1-Lipschitz which was observed to correspond to the minimization of the 1-Wasserstein distance (4.16) by duality arguments [Arjovsky et al., 2017]. Note, that the gradient penalty term should be applied on the domain where the discriminator is applied. To enforce that, it is therefore applied on images $\mathbf{Y}^{\text{mix}}$ uniformly chosen at random between simulated images and real images as $\mathbf{Y}^{\text{mix}}(\mathbf{x},\mathbf{b}) := \eta\mathbf{Y}^{\text{real}} + (1-\eta)\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b})$ with $\eta \sim \mathcal{U}([0,1])$, where $\mathcal{U}([0,1])$ stands for the uniform distribution on $[0,1]$.

Introducing now parameters $\gamma, \lambda_1, \lambda_2 \geq 0$, $\lambda_D > 0$ and $\delta \in \{0,1\}$ (we make this choice to assess the effect of the presence of the discriminator term in the loss) and incorporating in (4.17) the physical-inspired regularization terms discussed in Section 4.3.2.2 we consider the empirical risk formulation of (4.17)-(4.18) which reads:

$$\min_{\mathbf{x}\in\mathbb{R}_+^{L^2},\, \mathbf{b}\in\mathbb{R}_+^{M^2}} \mathbb{E}\left[\frac{\gamma}{2}\|\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b}) - \mathbb{E}[\mathbf{Y}^{\text{real}}]\|^2 - \delta D_{\boldsymbol{\varphi}}(\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b}))\right] + \lambda_1\|\mathbf{x}\|_1 + \frac{\lambda_2}{2}||\nabla\mathbf{b}||^2 \tag{4.20}$$

$$\min_{\boldsymbol{\varphi}\in\Phi} \mathbb{E}\left[D_{\boldsymbol{\varphi}}(\mathbf{Y}^{\text{sim}}(\mathbf{x},\mathbf{b})) - D_{\boldsymbol{\varphi}}(\mathbf{Y}^{\text{real}}) + \lambda_D\left(\|\nabla_{\mathbf{y}} D_{\boldsymbol{\varphi}}(\mathbf{Y}^{\text{mix}}(\mathbf{x},\mathbf{b}))\| - 1\right)^2\right], \tag{4.21}$$

where during optimization an empirical estimation on batches $B \subset \{1,\dots,T\}$ drawn uniformly at random can be performed for approximating the expected values for both $\mathbf{Y}^{\text{sim}}$ and $\mathbf{Y}^{\text{real}}$. Recalling (4.19), we observe that in (4.20) we are enforcing distributional equality using (4.15) and a GAN discriminator with parameters to be trained via (4.21), but, furthermore, we are enforcing physically-motivated regularisation. Practically, and similarly as for GANs, problem (4.20) is minimised at each epoch for few (say, $k_{\text{G}}$) iterations, then, similarly, (4.21) is minimised for $k_{\text{D}}$ iterations by freezing the quantities $\mathbf{x}^{k_{\text{G}}}, \mathbf{b}^{k_{\text{G}}}$ previously computed. This alternate procedure continues till convergence. In order to resort to gradient-based solvers for the problem (4.20), some details should be given concerning the computation of gradients w.r.t. the variables $\mathbf{x}$ and $\mathbf{b}$. In particular we notice that the computation of gradients w.r.t. $\mathbf{x}$ and $\mathbf{b}$ requires the computation of the quantities $\nabla_{\mathbf{x}}\, \mathbf{y}_t^{\text{sim}}(\mathbf{x},\mathbf{b})$ and $\nabla_{\mathbf{b}}\, \mathbf{y}_t^{\text{sim}}(\mathbf{x},\mathbf{b})$ which both rely on the non-linear Poisson model given by (4.11). The computations for such gradients are reported in Appendix C.

#### 4.3.3.2 Discriminator architecture

Inspired by [Gupta et al., 2021], we choose as a discriminator a standard convolutional neural network $\mathbb{R}^{L^2} \to [0,1]$ with 3 main layers and 2 fully connected layers, as shown in Figure

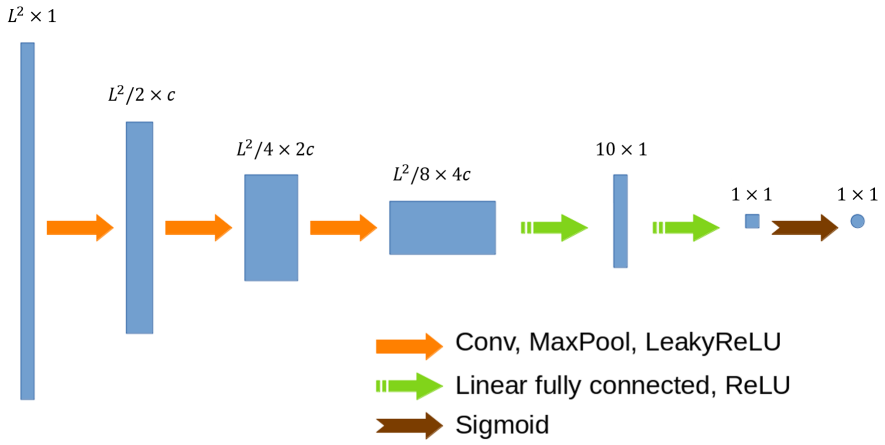



Figure 4.5 – Discriminator architecture

4.5. Each convolutional layer is followed by a max pooling layer and a non-linear ReLU (rectified linear unit) activation function. The number of channels increases and is doubled after each convolutional layer. The last layer is a sigmoid function that returns an output value within the range $[0, 1]$. This setup is a common choice for this kind of network since it gives more capacity to the network while reducing the resolution of the intermediate layers. However, the number of convolutional layers is limited by the input size and the size of the kernel. Indeed it is not possible to apply a convolutional kernel on an image smaller than the kernel. For this reason, there are only 3 convolutional layers here. In practice and for the size of images considered, we have seen that the chosen architecture gives enough capacity to the network. The discriminator is trained to yield $0$ given a simulated image as input and $1$ in case the input is a real image. However, as for standard GANs, it is used passively to improve the quality of simulated images.

#### 4.3.3.3 Optimization algorithms

The adversarial minimization problems (4.20) and (4.21) can be solved by means of tailored optimization algorithms suited to the particular regularization of the loss considered. To train the generator, we use the FISTA accelerated proximal algorithm [Beck and Teboulle, 2009] which can deal effectively with the non-smoothness of the $\ell_1$-norm. The choice of the learning rates has critical importance in order to achieve convergence. If such a parameter is too large then the algorithm may completely diverge, while if it is too small the convergence is too slow. Such parameter depends also on the proportion of noise and signal in the data. There is indeed a trade-off between the sparsity constraint parameter and the learning rates. For our experiments, we choose the learning rates manually, making sure to obtain a converging algorithm. We tested also backtracking strategies to automatize this choice. In practice, however, the constant values chosen here after empirical tuning seem to be general enough to process all the different kinds of data we present in our results without any change.

For the discriminator loss in (4.18), we do not have the same differentiability problem, and the choice of the corresponding learning rate is not data-constrained.

| $B$ | 32 |
|---|---|
| $\tau_\varphi$ | $10^{-6}$ |
| $\tau_{\mathbf{x}}$ | 1 |
| $\tau_{\mathbf{b}}$ | 1 |

Table 4.1 – Optimisation parameters.

To perform such optimization, we thus used the stochastic ADAM algorithm, which is a common choice to train a neural network of the form described in Section 4.3.3.2. In Table 4.1 we report the algorithmic parameters (batch size $B$ for ADAM, learning rate $\tau_\varphi$ for the optimization of (4.21) and initial learning rates $\tau_{\mathbf{x}}$ and $\tau_{\mathbf{b}}$ for the optimization in (4.20)) used in the experiments below. Note that due to the use of backtracking, a possible overestimation of both $\tau_{\mathbf{x}}$ and $\tau_{\mathbf{b}}$ can be corrected along the iterations.

### 4.3.4 Numerical results

FluoGAN is applied to three different types of data in order to evaluate its performance in comparison with state-of-the-art approaches, namely COL0RME (presented in Chapter 2) and SRRF [Gustafsson et al., 2016]. We first use synthetic data (see Section 4.3.4.1) by simulating fluctuating, blurred and noisy data acquired using conventional fluctuating/blinking fluorophores. Next, we apply FluoGAN to images of a phantom sample with known patterns acquired by a real fluorescence microscope. The results are presented in Section 4.3.4.2. Finally, we applied FluoGAN on challenging real images of the unicellular algae *Ostreopsis cf. Ovata*, acquired by an epifluorescent microscope (see Section 4.3.4.3).

#### 4.3.4.1 A simulated *ARGO-CR*-type validation dataset

We started applying FluoGAN to synthetic images for reconstruction quality evaluation. The synthetic images have been generated by us considering physically-plausible parameters of the microscope configuration as well as properties of the sample (e.g. fluctuation behavior, out-of-focus fluorescence, etc.). For data generation, we used model (4.9). The spatial pattern used is shown in Figure 4.6. Its structure resembles the pattern of a calibrated slide ARGO-CR (Argolight, Pessac, France) used in practice to assess quantitatively reconstruction capacities, see section 4.3.4.2. To be more precise, the pattern used consists of 14 sets of parallel lines of 100 nm width. The separation distance $d$ (center-to-center distance) between the two middle lines of each set is gradually increasing at a rate of 30 nm. Thanks to this setup, we can easily see the resolution level reached by each of the approaches considered.

Using the Poisson model (see (4.9)) to simulate the temporal fluctuations, we make a video of 500 frames at a frame rate of 100 frames per second (fps). The pixel size is equal to 100 nm while the full width at half maximum (FWHM) of the PSF is 324 nm. A spatially smoothly varying, temporally constant, background is added to simulate the presence of the out-of-focus fluorescent molecules and an average bleaching time (i.e. time during which the fluorophore stays emissive) of 20 s has been used. Reconstructions of the simulated diffraction-limited data are available for the FluoGAN, COL0RME, and SRRF methods in Figure 4.7, using a super-resolution factor $q = 6$ for all three approaches. The method COL0RME allows us to select between different regularization penalties in the support estimation step of the method (see Chapter 2 for more details) and for the purpose of this experiment the $\ell_1$-norm penalty was chosen due to the better and most time-efficient reconstruction. Regarding the method SRRF, for the temporal analysis of the radiality image sequence we are using the auto-correlation analysis provided in SRRF's Fiji plugin [Laine et al., 2019] that gives the best results.

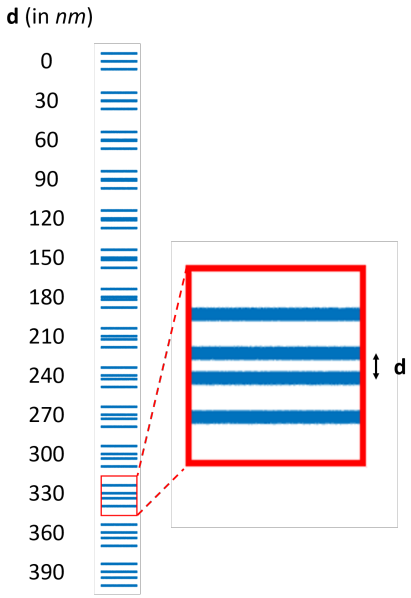



Figure 4.6 – Spatial structure of simulated/real calibrated sample (ARGO-CR slide, Argolight)

From the reconstruction shown in Figure 4.7, it is clear that FluoGAN achieves better results than both COL0RME and SRRF. FluoGAN is indeed able to reconstruct two parallel lines with a separation distance of 120 nm between their centers or only 20 nm between their closest edges (given that the width of the lines is 100 nm). COL0RME is more resolutive than SRRF, achieving a resolution of $80 - 110$ nm computed as the ability to reconstruct two lines with the closest edge-to-edge distance of such value, while SRRF reaches a resolution of around 200 nm measured in the same way. Finally, FluoGAN and COL0RME more than SRRF estimate real intensity values, while SRRF preserves better the "continuous" structure of the sample.

#### 4.3.4.2 Real *ARGO-CR* dataset

In the second experiment, we consider images acquired by an ARGO-CR calibrated slide using an epifluorescence microscope. In more detail, images were acquired on an AxioObserver Z1 (Zeiss, Germany) with 63x/1.15W Korr LD C-Apochromat objective, ORCA Flash 4.0 camera (Hamamatsu, Japan), 540-570nm LED excitation, and 581-619nm emission filter. The spatial pattern is similar to the one used in our simulated data, with the distance between the two middle filaments of each set increasing by 30nm and varying from 0 to 390nm (see Figure 4.6 for a graphical representation). Here we can estimate the resolution reached by the methods under real conditions. Only FluoGAN and SRRF are applied to these data, as due to a slight sample shift during acquisition, these data are not adaptable for COL0RME, which is based on the estimation of temporal covariance matrices.

The pixel size of the sCMOS camera used is equal to 103 nm, while the FWHM of the PSF is estimated to be 270 nm. Only 500 images were acquired with a frame rate of 10 fps, for a total acquisition time of less than 1 minute. For the reconstruction, a super-resolution factor of $q = 6$ was used for both methods. For FluoGAN, the 500 acquired images constituted the training set and 5000 iterations were performed in order to obtain the reconstruction presented in Figure 4.8.

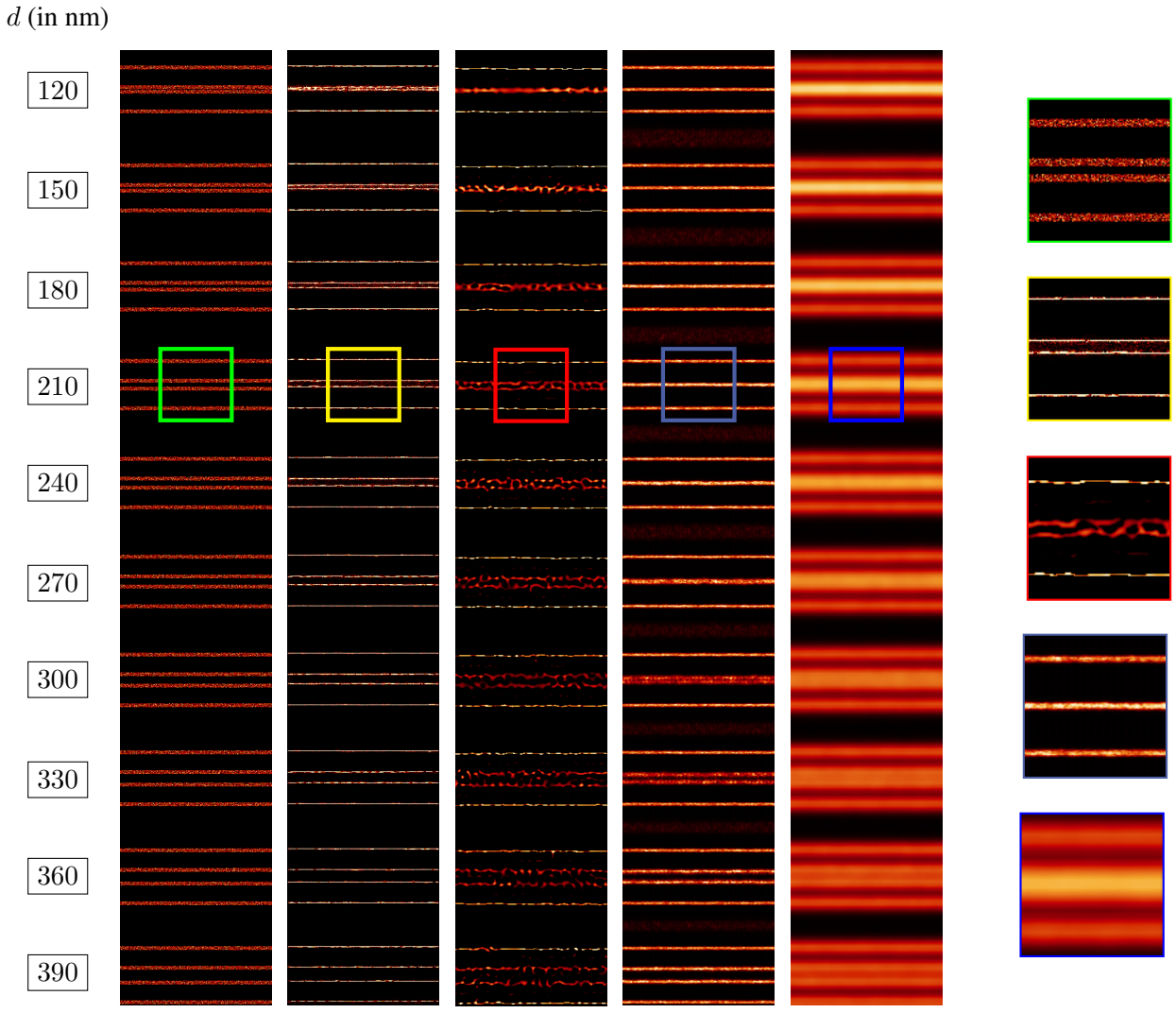



Figure 4.7 – Numerical results on a simulated dataset with zooms (on the right side of the figure), with enhanced contrast for better visualization. The separation distance $d$, used to measure the resolution, is given for each set of lines. The FWHM of the PSF in this experiment is equal to 324nm.

Compared to SRRF, FluoGAN achieves significantly better resolution levels. However, we have noticed that the structures reconstructed by FluoGAN are a bit thinner than the real ones.

We then tested the performance of FluoGAN with ($\delta = 1$) and without ($\delta = 0$) the presence of the discriminator in problem (4.20) to assess how important its role is in the overall estimation process. The two results are available in Figure 4.9. We observed that even if the $\ell_2$ distance (4.15) allows us to recover the main pattern, the explicit presence of the discriminator significantly improves the reconstruction. The result with no discriminator can nonetheless be used as an educated guess for the initialization of FluoGAN. In Figure 4.10 we provide a schema assessing the reconstruction precision and the resolution achieved in both cases. We note that without a discriminator the distance between filaments reconstructed by FluoGAN is systematically under-estimated. The discriminator plays here the double role of both allowing the separation of filaments with 120 nm resolution (separates two parallel lines with a distance of 120 nm between their centers or only 20 nm between their closest edges) and improving the reconstruction precision for all filaments. In Table 4.2 the computational costs required to run FluoGAN on image stacks of different sizes with/without the discriminator are reported.

| input images size | FluoGAN ($\delta = 1$) | FluoGAN ($\delta = 0$) |
|---|---|---|
| $50 \times 50$ | 10 min | 3 min |
| $50 \times 266$ | 60 min | 20 min |

Table 4.2 – Computation time for 5000 iterations of FluoGAN on a small GPU Nvidia GeForce GTX 950M from 2015 with 2GB of dedicated memory.

In Figure 4.11 we report the convergence graphs along the iterations. Unlike other optimization problems, it is difficult to visualize convergence since the simulator and discriminator have adverse goals and the difficulty is to maintain both in competition preventing one from overtaking the other.

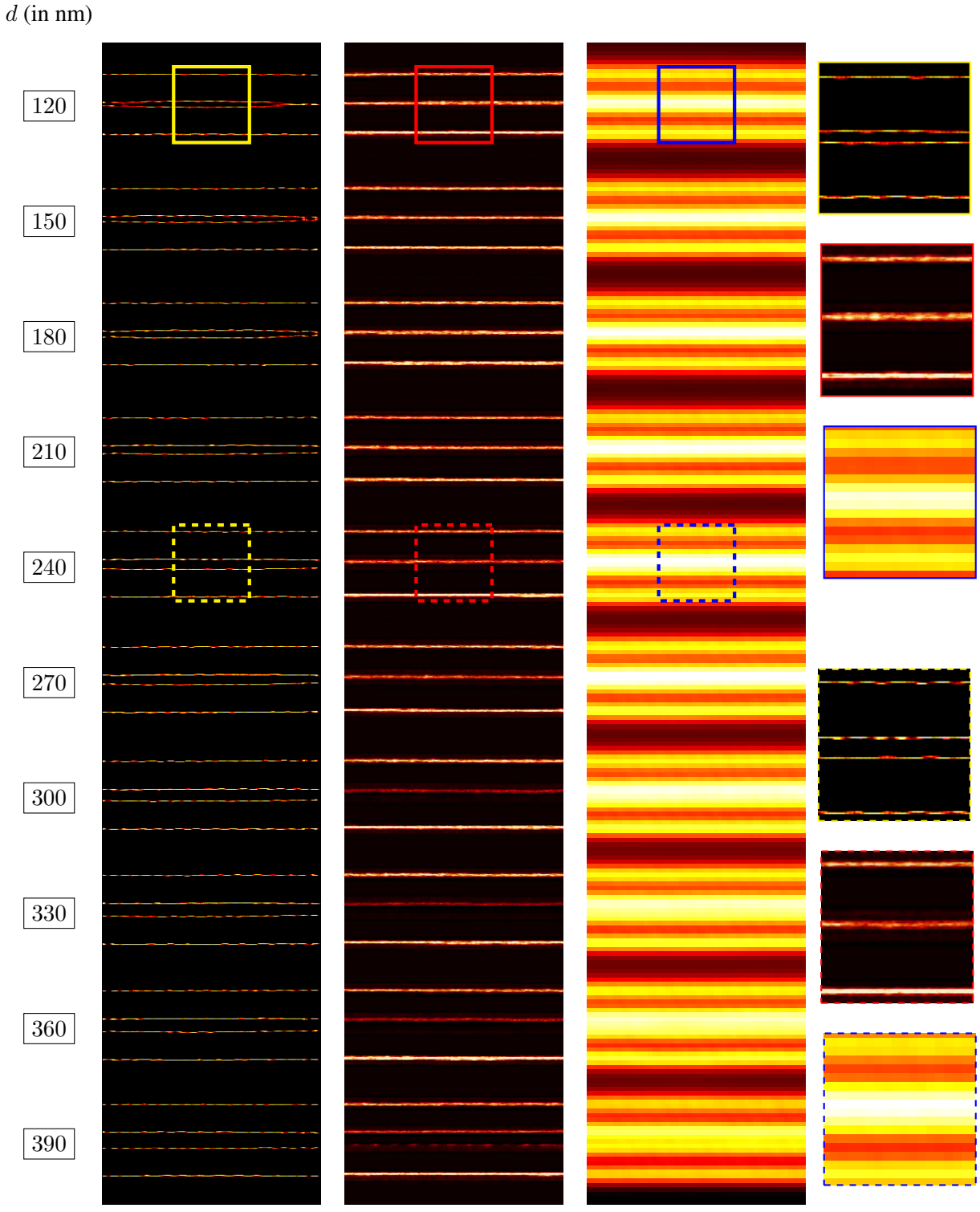
#### 4.3.4.3 Real data: Ostreopsis images

To test FluoGAN on challenging real biological samples, we considered a dataset of the unicellular alga *Ostreopsis cf. Ovata* (see Figure 4.12a). Such a dataset has several difficulties.

First, it shows a 20µm thick sample that can exhibit a strong out-of-focus signal which is clearly the case here for the microtubules staining as they form a cortical structure all around the alga. Another difficulty is the reduced transparency and non-negligible turbidity, which induces PSF distortion. Due to those complications, confocal microscopes cannot acquire through the whole thickness of the alga. To avoid this, we consider acquisitions obtained by an epifluorescence microscope. The flexibility and the data-adaptivity of FluoGAN allow dealing with such data, whereas standard approaches fail.

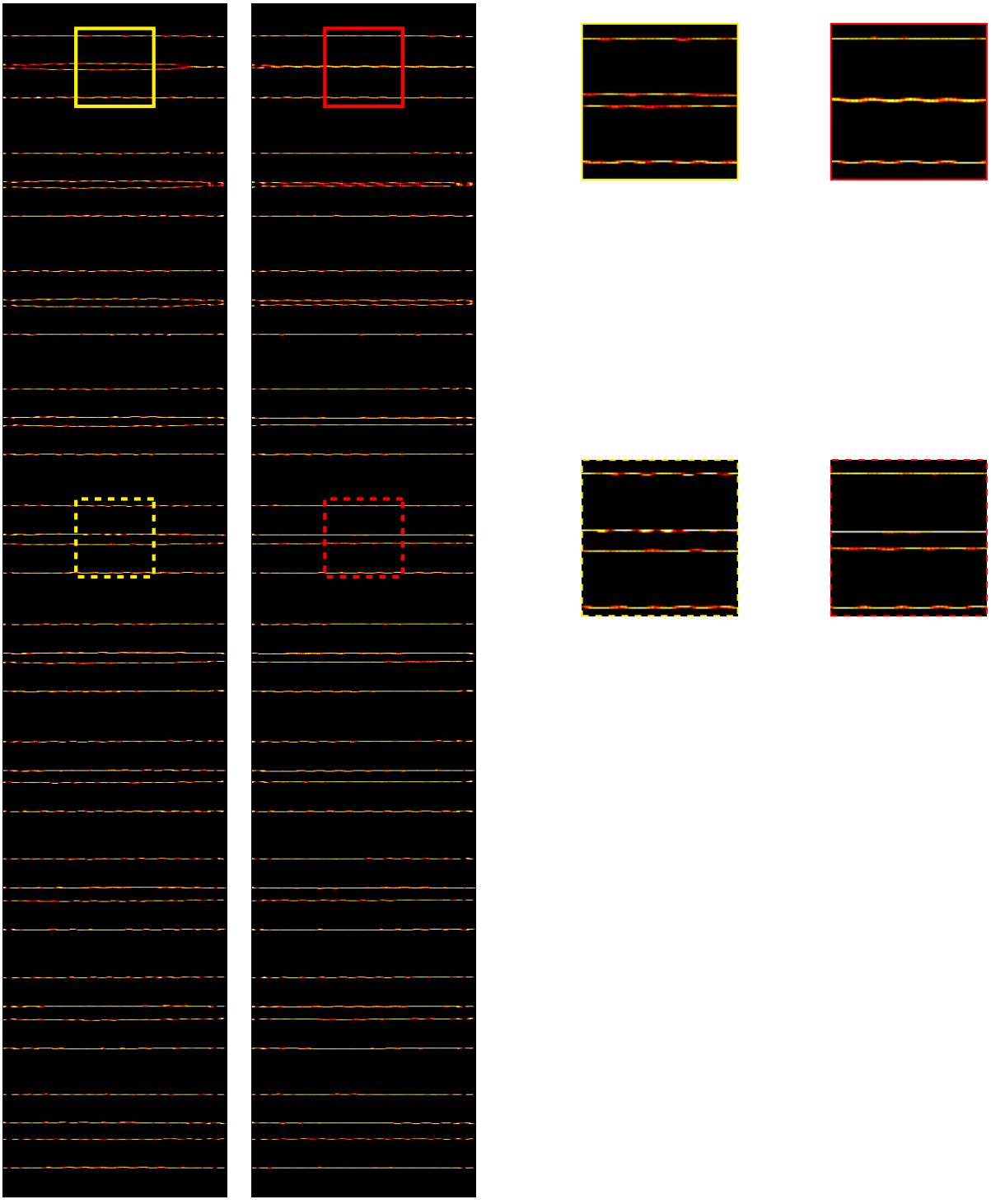
For the experimental design, microtubules were stained with TRITC dye, but also with Hoechst for DNA (for details see [Velasquez, 2021]). We focused on the tip of a microtubule bundle at the ventral pole (see Figure 4.12b). Microtubules play an important role in cell division and they are an element of explanation of the proliferation of these algae, and therefore, important to be studied. Epifluorescence images have been acquired with the same protocol as for ARGO-SR sample (see details in Section 4.3.4.2).

The reconstruction performed by FluoGAN is shown in Figure 4.12c, while the parameters used are reported in Table 4.3. We compare the results obtained by FluoGAN with the ones ob-



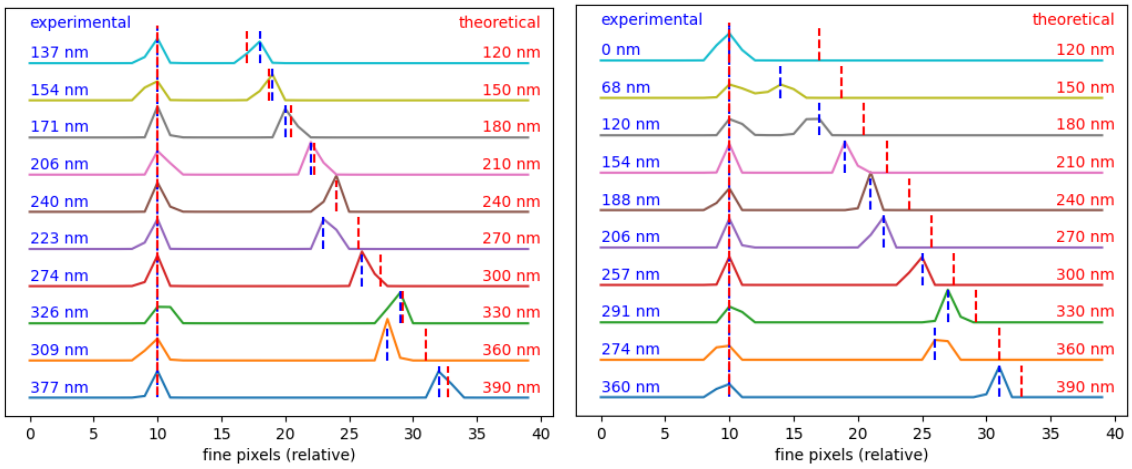

(a) FluoGAN (b) SRRF (c) $\bar{y}$

Figure 4.8 – Numerical results for *ARGO-CR* data with zooms. The distance $d$ is given for each set of lines. The FWHM of the PSF is estimated to be equal to 270nm.



(a) FluoGAN $\delta = 1$ (b) FluoGAN $\delta = 0$

Figure 4.9 – Comparison of FluoGAN results with (left, $\delta = 1$) and without (right $\delta = 0$) discriminator. Since the discriminator relies on convolutional layers, it improves first the center of the image. Due to the boundary effects related to the width of convolution kernels, boundary regions are not compared.




(a) FluoGAN, $\delta = 1$ (b) FluoGAN, $\delta = 0$

Figure 4.10 – Distances between filaments in Figure 4.8 reconstructed by FluoGAN, with (left) and without (right) discriminator. Each colored 1D profile corresponds to the $y$-averaged intensity of the reconstructed image plotted along the $x$-axis (excluding boundaries). Taking the left-most peak as a reference, the red dotted lines denote the theoretical location of the filament to reconstruct, while the blue dotted lines are placed in correspondence with the maximum value of the filament reconstructed by FluoGAN.

| | | |
|---|---|---|
| c10cm | Sensor's pixel size | 103nm |
| | FWHM of the PSF (estimated) | 325nm |
| | Undersampling rate | 6 |
| | Iteration number | 1000 |
| | Number of images in training set | 500 |

Table 4.3 – Parameters used for *ostreopsis* image reconstruction

tained by SRRF. Regarding SRRF and differently from the two previous experiments, gradient weighting is performed, as suggested by the authors, to deal with the low signal-to-noise ratio (SNR) regime. The reconstruction obtained by SRRF is reported in Figure 4.12d. Compared to the solution obtained by SRRF, the FluoGAN result shows fewer background artifacts and promotes better continuity of filaments and intensity reconstruction. Furthermore, the reconstruction obtained by FluoGAN reduces the appearance of ghost filaments.

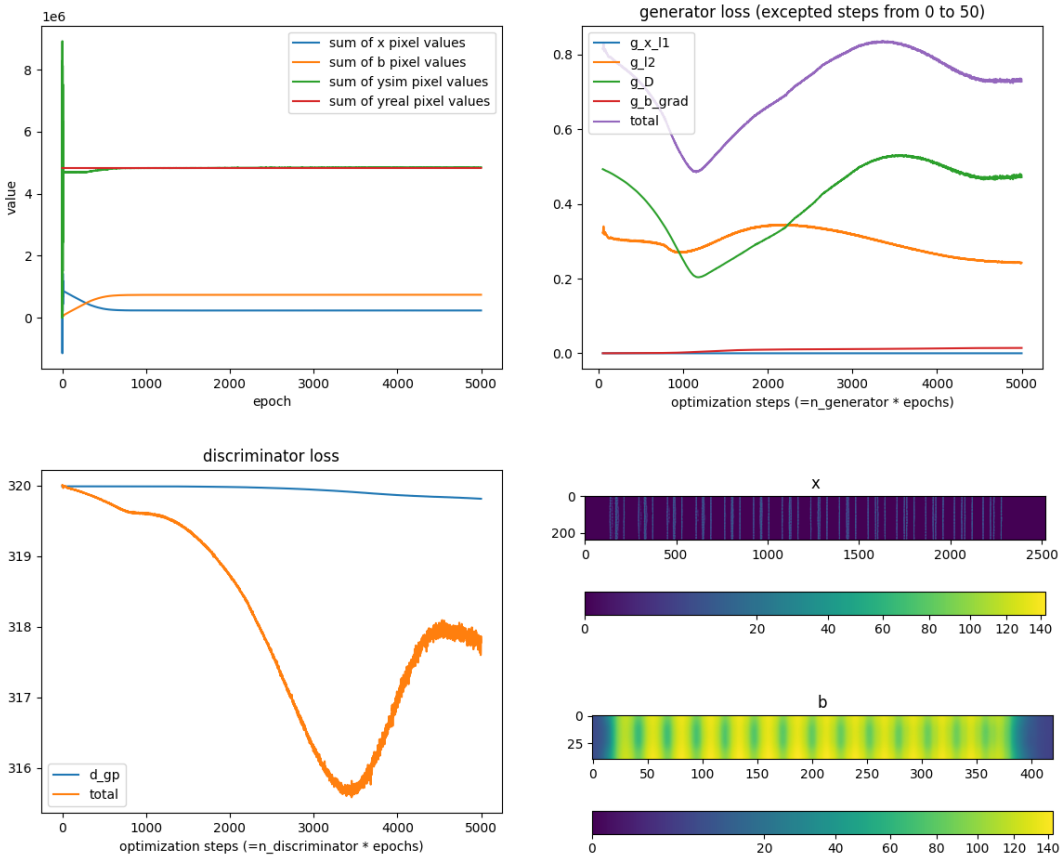
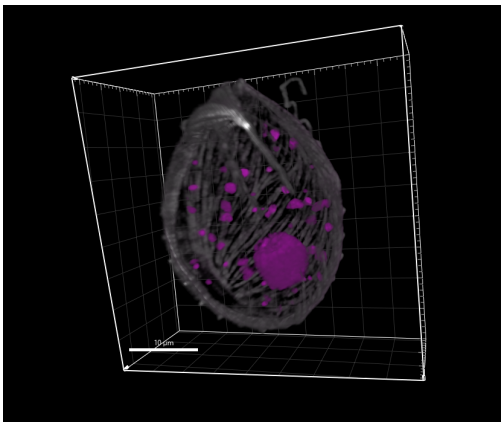
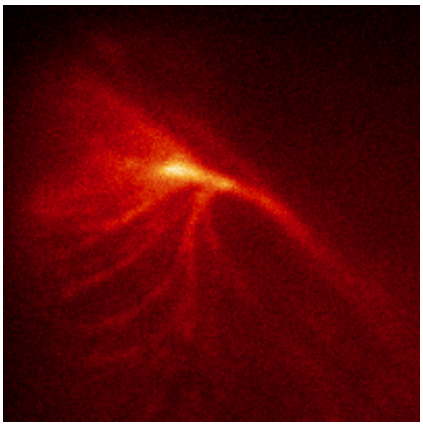


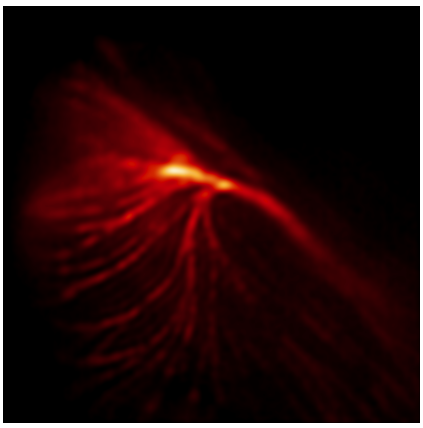

Figure 4.11 – Top left: sum of pixel values (i.e., total photon count) VS. iteration number. After about 1000 iterations the sum of pixels for simulated images ($\mathbf{y}^{\text{sim}}$) equals the sum of pixels for real input images ($\mathbf{y}^{\text{real}}$). Top right and bottom left: balance between simulator and discriminator training. Top right: contributions to generator loss function: discriminator distance, $\ell^2$ distance, regularizations on $\mathbf{b}$ and $\mathbf{x}$. Bottom left: discriminator loss function (orange) including the gradient penalty regularization (blue). Bottom right: super-resolution image $\mathbf{x}$ and reconstructed background $\mathbf{b}$.



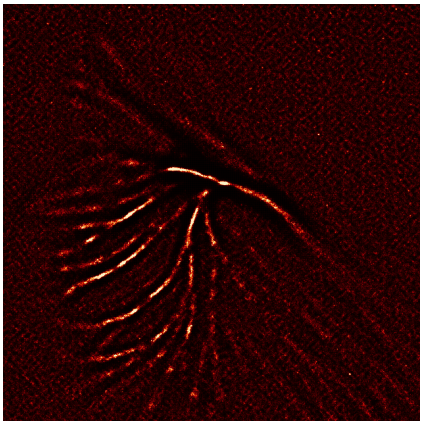
(a) Confocal 3D image of *Ostreopsis* with microtubules (white) and DNA (magenta).



(b) Low-resolution epifluorescent image (sample average) of the microtubules tip.



(c) FluoGAN reconstruction of Fig. 4.12b.



(d) SRRF reconstruction of Fig. 4.12b.

Figure 4.12 – Real *Ostreopsis cf Ovata* data.

## 4.4 Div-Blurring: Fluorescence image deconvolution microscopy via VAEs

We now present the method Diversity Deblurring (DivBlurring) that is an extension of the method DivNoising proposed by [Prakash et al., 2021]. The authors in DivNoising assume that images $\mathbf{y}$ have been created from a clean signal $\mathbf{x}$ via a measure or learned noise model (NM), i.e., $\mathbf{y} \sim p_{\text{NM}}(\mathbf{y}|\mathbf{x})$ and they aim at recovering the denoised image $\mathbf{x}$. In DivBlurring, in addition to the noise corruption in $\mathbf{y}$, we also consider blur degradations due to the convolution of the true signal $\mathbf{x}$ with the microscope PSF. We thus consider that images $\mathbf{y}$ have been created from a physical model (PM) and we write $\mathbf{y} \sim p_{\text{PM}}(\mathbf{y}|\mathbf{x}, \mathbf{\Psi})$, where $\mathbf{\Psi}$ is a known matrix modeling convolution operation. After applying the proposed method, we thus aim at reconstructing a deblurred and denoised image $\mathbf{x}$. Our method, similar to DivNoising, explicitly includes the simplified physical model in the decoder architecture of a traditional VAE. The DivBlurring architecture is reported in Figure 4.13.

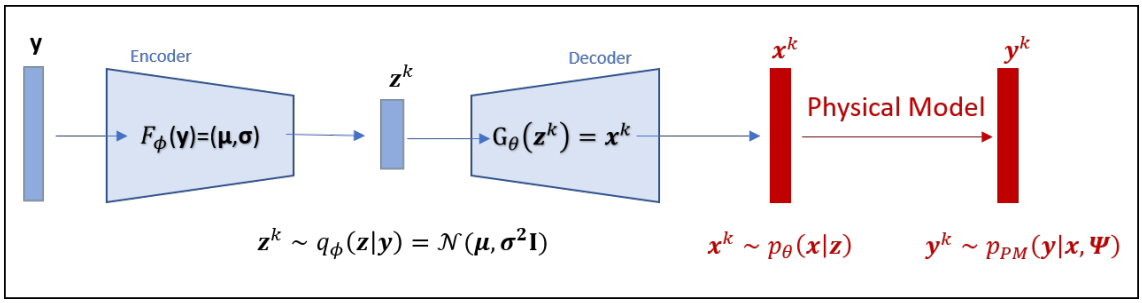



Figure 4.13 – DivBlurring Architecture. We color red the additional elements with respect to a normal VAE.

We describe in Section 4.4.1 the principles DivBlurring, then in Section 4.4.2 we show that it is possible to introduce in the loss function some additional priors on the solution that improve the reconstruction results and, finally, in Section 4.4.3 we present some preliminary results.

### 4.4.1 DivBlurring

We are using here a simplified model describing acquisitions $\mathbf{y} \in \mathbb{R}^{M^2}$ (column-wise vectorized form of image $\mathbf{Y} \in \mathbb{R}^{M \times M}$) in fluorescence microscopy that is given by:

$$\mathbf{y} = \mathbf{\Psi}\mathbf{x} + \mathbf{n} \tag{4.22}$$

where $\mathbf{x} \in \mathbb{R}^{M^2}$ is the "clean" vectorized image we are aiming to recover, $\mathbf{\Psi} \in \mathbb{R}^{M^2 \times M^2}$ is a given matrix used to describe the convolution operator defined by the system's PSF, while $\mathbf{n} \in \mathbb{R}^{M^2}$ is assumed, for simplicity, to be a vector of i.i.d. random variables following a Gaussian distribution with zero mean and variance equal to $\sigma_n^2$. We omit the down-sampling operator since in DivBlurring we will perform only the task of deconvolution and reconstruct images in the same grid with the one of acquisitions. To make explicit that images $\mathbf{y}$ have been generated by (4.22), called by us 'physical model', we write $\mathbf{y} \sim p_{\text{PM}}(\mathbf{y}|\mathbf{x}, \mathbf{\Psi})$.

In our framework, similar to [Prakash et al., 2021] from where we have borrowed many of the computations below, we assume that from a latent variable $\mathbf{z}$ we are able to generate clean

images $\mathbf{x}$ that, afterward, can be passed through the physical model leading to images $\mathbf{y}$. The main difference between our approach and the method DivNoising [Prakash et al., 2021] is that we consider a Physical Model (PM) while in DivNoising a Noise Model (NM) is considered. Figures 4.3 and 4.13 illustrate this easily.

In DivBlurring, the decoder describes a joint model between three variables as follows:

$$p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{x},\mathbf{z}) = p(\mathbf{y},\mathbf{x}|\mathbf{z})p(\mathbf{z}) = p(\mathbf{y}|\mathbf{x},\mathbf{z})p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})p(\mathbf{z}) \tag{4.23}$$

From Figure 4.13 it is easy to observe that $\mathbf{y}$ is conditionally independent of $\mathbf{z}$ once $\mathbf{x}$ is provided. Therefore, we can write:

$$p(\mathbf{y}|\mathbf{x},\mathbf{z}) = p_{\,\mathrm{PM}}(\mathbf{y}|\mathbf{x},\boldsymbol{\Psi}). \tag{4.24}$$

By further deterministically defining the clean image as $\mathbf{x} = g_{\boldsymbol{\theta}}(\mathbf{z})$, where $g_{\boldsymbol{\theta}}(\cdot)$ is the decoder network, and, subsequently, $p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$ as a Dirac distribution centered at $g_{\boldsymbol{\theta}}(\mathbf{z})$, we get:

$$p_{\boldsymbol{\theta}}(\mathbf{y},\mathbf{x},\mathbf{z}) = p_{\,\mathrm{PM}}\left(\mathbf{y}|\mathbf{x} = g_{\boldsymbol{\theta}}(\mathbf{z}),\boldsymbol{\Psi}\right)p(\mathbf{z}) \tag{4.25}$$

With simple computations, the reconstruction loss, given from (4.5) for conventional VAEs now becomes:

$$\mathcal{L}^{\text{R -DB}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) = -\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})}\{\log p_{\,\mathrm{PM}}\left(\mathbf{y}|\mathbf{x} = g_{\boldsymbol{\theta}}(\mathbf{z}),\boldsymbol{\Psi}\right)\} \tag{4.26}$$

Apart from this modification, the DivBlurring approach follows the standard VAE training procedure. The only modification consists in how the decoder distribution is modeled, so the KL divergence loss will remain the same (see (4.6), or under Gaussian assumptions (4.7)).

To further elaborate on the DivBlurring reconstruction loss $\mathcal{L}^{\text{R -DB}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$, we recall that we have assumed the noise component $\mathbf{n}$ in (4.22) to be white and Gaussian, with constant pixel variance equal to $\sigma_n^2$. Therefore, the log-likelihood $\log p_{\,\mathrm{PM}}(\mathbf{y}|\mathbf{x} = g_{\boldsymbol{\theta}}(\mathbf{z}),\boldsymbol{\Psi})$ is given by:

$$\log p_{\,\mathrm{PM}}(\mathbf{y}|\mathbf{x} = g_{\boldsymbol{\theta}}(\mathbf{z}),\boldsymbol{\Psi}) = -\frac{1}{2\sigma_n^2}\|\boldsymbol{\Psi}g_{\boldsymbol{\theta}}(\mathbf{z}) - \mathbf{y}\|^2 + c, \tag{4.27}$$

where $c \in \mathbf{R}_+$ is a constant. Aiming at minimizing the reconstruction loss, as indicated in Section 4.1.2, we can discard the constant $c$. From (4.26) and (4.27), the reconstruction loss is therefore given by:

$$\mathcal{L}^{\text{R-DB}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})}\{\frac{1}{2\sigma_n^2}\|\boldsymbol{\Psi}g_{\boldsymbol{\theta}}(\mathbf{z}) - \mathbf{y}\|^2\} \tag{4.28}$$

### 4.4.2 Regularization penalties

In order to avoid under-fitting and over-fitting, we use as for FluoGAN, regularization term enforcing physically-meaningful properties in the loss function. In formulas, we can therefore formulate a new loss function, as follows:

$$\mathcal{L}^{\text{DB}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) = \mathcal{L}^{\text{R-DB}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) + \mathcal{L}^{\text{KL}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) + \mathcal{R}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}), \tag{4.29}$$

where $\mathcal{L}^{\text{R-DB}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$ is the reconstruction loss defined for DivBlurring in (4.28), $\mathcal{L}^{\text{KL}}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$ is the KL loss defined in (4.7) under Gaussian assumptions and $\mathcal{R}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y})$ is a term favoring regularity on the solution $\mathbf{x} = g_{\boldsymbol{\theta}}(\mathbf{z})$. A regularization penalty promoting sparsity is:

$$\mathcal{R}_{\boldsymbol{\phi},\boldsymbol{\theta}}(\mathbf{y}) = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{y})}\{\lambda\|g_{\boldsymbol{\theta}}(\mathbf{z})\|_1\} \tag{4.30}$$

where $\lambda$ is the regularization parameter. We further would like to promote non-negativity. To do so, we are using a Rectified Linear Unit (ReLU) activation function as a final layer to the decoder.

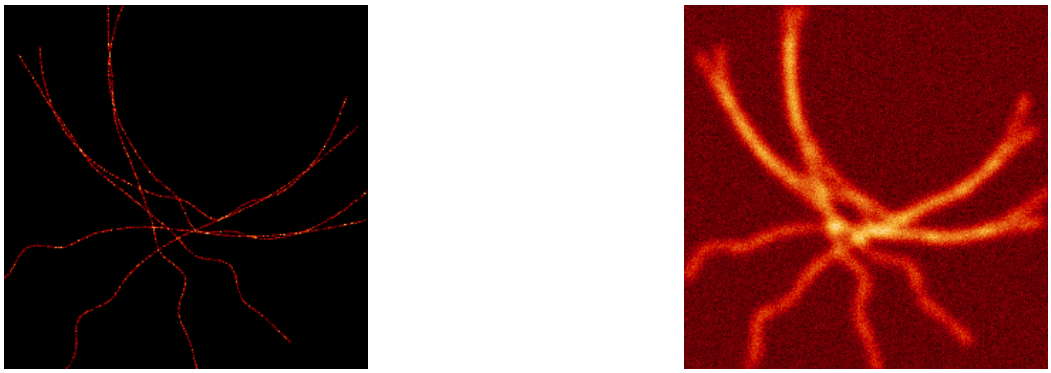
### 4.4.3 Preliminary Results

To train the model, we generated a synthetic dataset of fluctuating filaments. The training dataset contains 7000 blurred and noisy images. Each image corresponds to a different realization of the noise which we assume to be Additive White Gaussian (AWG) with constant variance. Data are blurred with a Gaussian PSF. The FWHM of the PSF is chosen to be equal to 221.8 nm, while the pixel size is equal to 25 nm. The testing dataset can be the same as the training one or different, although, corrupted with AWG noise of the same variance and blurred with the same PSF, as the data in the training set.

The selection of such a dataset arises from the fact that we want to be compatible with the acquisitions obtained by real microscopes. Using real microscopes, we are able to acquire more than one image from different parts of the sample or from different samples. Similarly, in our simulated data we consider 1000 images with the same support and additional sets of 1000 images with similar but not equal support (these images correspond to different parts of the sample).

DivBlurring uses the same fully convolutional network architecture with depth 2 as DivNoising, i.e. 2 down-sampling layers for the encoder network and 2 up-sampling layers for the decoder network. In total $36K$ parameters got assigned for both networks by considering input data of dimensions $256 \times 256$. After encoding, the resultant dimension of the bottleneck of the latent space is chosen to be equal to 64. The designed network is trained with a batch size equal to 32, an initial learning rate equal to 0.001, and 150 epochs.

In Figure 4.14, we plot the original ground truth image as well as an image of the noisy and blurred synthetic tubulin dataset to which we apply our method and which is generated using the simplified model (4.22), i.e., without degradation of background and Poisson noise. Note that we did not use the ground truth image in training the model as the method is unsupervised. We use it only for a comparison with our reconstructions.

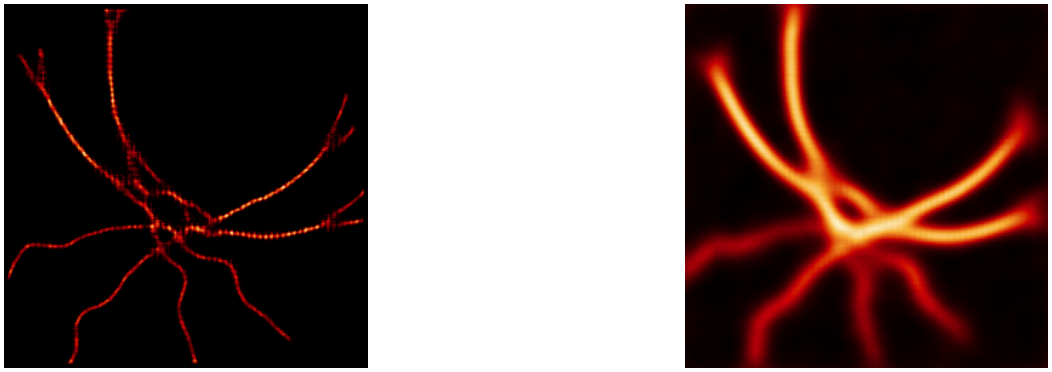


(a) (b)

Figure 4.14 – (a) The GT image, (b) A single frame of the temporal stack of diffraction-limited images

DivBlurring, similar to DivNoising [Prakash et al., 2021], is capable of generating not only one but multiple diverse samples from the posterior, $\mathbf{z}^k \sim q_\phi(\mathbf{z}|\mathbf{y})$, $k > 1$, giving rise to several estimations of deblurred/denoised images $\mathbf{x}^k$, $k > 1$. In order to compute a single final image we approximate the minimum mean square error (MMSE) estimate by simply averaging the images $\mathbf{x}^k$.

Averaging over 100 randomly selected samples $\{\mathbf{x}^k\}_{k=1}^{100}$ we are able to report in Figure 4.15a the reconstruction obtained by the method DivBlurring when applied to the blurred/noisy frame in Figure 4.14b. For comparison, in Figure 4.15b, we present the result of the method Divnoising applied to the same frame. Clearly, DivBlurring is able not only to denoise but also to deblur the given image, which is not the case, as expected, for DivNoising. The proposed method is still under investigation, with promising preliminary results. Finally, we noticed some repeated artifacts (see Figure 4.15a) that we would like to better investigate and correct.



(a) (b)

Figure 4.15 – (a) Reconstruction of 4.14b by DivBlurring (b) Reconstruction of 4.14b by DivNoising.

## 4.5 Discussion

In this chapter we focused on ways to solve inverse problems in fluorescence microscopy imaging applications using model-aware generative approaches.

Firstly, we presented FluoGAN, a novel framework for fluctuation-based super-resolution fluorescence microscopy combining the physical modeling of an optical system with data-driven adversarial learning. FluoGAN computes the desired super-resolved image along with a background image containing out-of-focus molecules, ambient fluorophores and other errors by comparing, in a suitable sense, the empirical distribution of observed data with the one of samples generated by a physically-grounded simulator. Due to its alternate minimization formulation, the proposed approach is flexible as it allows to the possible introduction of further regularization terms such as sparsity, smoothness, and non-negativity constraints. In comparison to standard GANs and inspired by recent work on Cryo-EM [Gupta et al., 2021], FluoGAN replaces the model-blind generator network with a simulator model encoding biophysical expertise in its structure and having as learnable parameters the desired quantities of interest. On simulated data, the proposed approach allows for achieving better resolutions than standard model-aware and state-of-the-art approaches. On real data, we first proved the efficiency of our algorithm on a 2D calibration slide used to access resolution power, then we validated FluoGAN on real challenging *Ostreopsis* data.

Next, we presented DivBlurring, an unsupervised method that includes the physical model of data acquisition in fluorescence microscopy in the decoder network of a standard VAE. The method has the advantage of generating more than one sample from a regularized latent space. We implemented the method following DivNoising [Prakash et al., 2021], a method performing

image denoising in microscopy images. Since DivBlurring is still a work in progress, as future work, we aim at exploiting the method more by performing super-resolution to both simulated and real microscopy images and using a more complex physical model describing the acquisition process better than the simplified physical model we are currently using (see (4.22)).

# CHAPTER 5

# Fluorescence image deconvolution using a Plug and Play Denoiser

*In this chapter we introduce the method Plug-and-Play COL0RME (PnP-COL0RME), an extension of the method COL0RME presented in Chapter 2 making use of deep learning for improving performance to challenging data. PnP-COL0RME is able to output an improved (deconvolved) image given as an input a temporal stack of diffraction-limited images acquired by conventional fluorescent microscopes. In Section 5.1 we detail the learning framework considered with a review of state-of-the-art PnP methods applied to image restoration problems, in Section 5.2 we introduce PnP-COL0RME, while in Section 5.3 we present some preliminary results of the proposed method on both simulated and real Ostreopsis data.*

---

**Related Publication:**

[Stergiopoulou et al., 2023]
V. Stergiopoulou, S. Mukherjee, L. Calatroni, L. Blanc-Féraud. "Fluctuation-based deconvolution in fluorescence microscopy using plug-and-play denoisers," to appear at the 9th International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), 2023.

## 5.1 Introduction to PnP methods

Plug-and-Play (PnP) methods appeared in the literature during the last decade. They are a very effective framework for solving inverse image restoration problems which, due to their structure, are also interpretable thanks to tools coming from optimization. They combine model-based and data-driven models, showing state-of-the-art performance in numerous applications. An overview of PnP approaches is provided in the survey paper [Kamilov et al., 2022].

### 5.1.1 Background

When dealing with ill-posed inverse problems in imaging (i.e., de-blurring, de-noising, super-resolution, inpainting, etc.), we typically aim to solve a regularized optimization problem in the form:

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^{L^2}} \mathcal{F}(\mathbf{x}) + \lambda\mathcal{R}(\mathbf{x}), \tag{5.1}$$

where $\mathcal{F}$ is a (smooth) data fidelity term and $\mathcal{R}$ a regularization term encoding prior knowledge on the solution $\hat{\mathbf{x}} \in \mathbb{R}^{L^2}$. Depending on the available prior information, different types of functions $\mathcal{R}$ can be used (see Section 1.3.3.1 for some examples). In most cases, $\mathcal{R}$ is non-smooth and proximal algorithms [Parikh and Boyd, 2014] are used for solving (5.1). Recalling Definition 1.1, the proximity operator of a convex, non-smooth function $\mathcal{R}$ with step-size $\tau > 0$ is given by:

$$\text{prox}_{\tau\mathcal{R}}(\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbb{R}^{L^2}} \mathcal{R}(\mathbf{x}) + \frac{1}{2\tau}\|\mathbf{z}-\mathbf{x}\|_2^2, \quad \mathbf{z} \in \mathbb{R}^{L^2}. \tag{5.2}$$

It is easy to notice that (5.2) corresponds to solving the problem of denoising an image $\mathbf{z}$ corrupted by an additive white Gaussian noise (AWGN) of constant variance equal to $\tau$. The observation that the proximal operator is indeed an image denoiser for AWGN inspired the authors of [Venkatakrishnan et al., 2013] to develop the framework of PnP priors, whose main idea consists in replacing, within a proximal algorithm, the operator $\text{prox}_{\tau\mathcal{R}}(\cdot)$ with a more general image denoiser $D(\cdot)$ corresponding to a regularization functional $\mathcal{R}$ whose definition is often not explicitly possible.

One of the main advantages of the PnP framework is that it allows complex imaging models to be used as advanced denoising models, e.g. denoisers parameterized by convolutional neural networks (CNNs) with impressive representational capabilities. In most cases, the image denoiser $D$ is trained to perform denoising on some pairs of clean-noisy images and can be used afterward in more-general image reconstruction problems (e.g. deblurring, super-resolution, etc.). Since the prior in PnP is learned on a task (image denoising) different from the final task (image reconstruction), PnP can be considered as a self-supervised learning framework [Kamilov et al., 2022].

### 5.1.2 State-of-the-art PnP methods

Current state-of-the-art denoisers include image-dependent filtering algorithms such as Block-Matching & 3D filtering (BM3D) [Dabov et al., 2007] or deep neural networks such as Denoising Convolutional Neural Networks (DnCNN) [Zhang et al., 2017] and Dilated-Residual U-Net (DRUNET) deep learning network [Zhang et al., 2021].

As already mentioned, these denoisers are used into iterative proximal schemes, such as the Proximal Gradient Descent (PGD) also called Forward-Backward Splitting

[Combettes and Wajs, 2005, Beck and Teboulle, 2009] that alternates between the gradient step on a differentiable data fidelity term $\mathcal{F}$ and application of the proximal operator of the regularizer $\mathcal{R}$. Other popular optimization schemes are the Alternate Directions Method of Multipliers (ADMM) [Boyd et al., 2011], the Douglas-Rachford Splitting (DRS) [Douglas and Rachford, 1956], and the Half-Quadratic Splitting (HQS) [Geman and Yang, 1995]. For each of these algorithms, it turns out that there is a corresponding PnP version obtained by associating the proximal map with the prior term and then replacing it with an off-the-shelf denoiser such as the ones above. PnP versions of proximal algorithms have been used to solve image restoration problems such as for example PnP-PGD in [Terris et al., 2020, Hurault et al., 2022a], PnP-ADMM and PnP-DRS in [Romano et al., 2017, Ryu et al., 2019, Hurault et al., 2022b] and PnP-HQS in [Zhang et al., 2017, Zhang et al., 2021, Cohen et al., 2021, Hurault et al., 2022a].

A recent breakthrough made in [Romano et al., 2017], consisted in the definition of explicit regularization by denoising (RED) approaches. RED is a framework where an explicit regularizer $\mathcal{R}(\mathbf{x})$ is constructed from a generic image denoiser $D$ by defining:

$$\mathcal{R}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T(\mathbf{x} - D(\mathbf{x})). \tag{5.3}$$

The great advantage of RED over other PnP methods is that it provides an explicit expression for the regularizer, which is implicit in most of the other approaches. In [Romano et al., 2017], it was further shown that under conditions on homogeneity, non-expansiveness and Jacobian symmetry, denoising can then be written as a gradient step, that is:

$$D(\mathbf{x}) = \mathbf{x} - \nabla\mathcal{R}(\mathbf{x}). \tag{5.4}$$

However, as shown in [Reehorst and Schniter, 2019], such conditions are unrealistic since widely-used denoisers, including BM3D, DnCNN and DRUNet, have non-symmetric Jacobians. Hence, RED does not in fact minimize an explicit regularization objective [Reehorst and Schniter, 2019]. In order to overcome this problem in [Cohen et al., 2021, Hurault et al., 2022a, Hurault et al., 2022b], the authors proposed to formulate, similar to RED, a gradient step denoiser of the form:

$$D_\varsigma(\mathbf{x}) = \mathbf{x} - \nabla\mathcal{R}_\varsigma(\mathbf{x}), \tag{5.5}$$

where $\mathcal{R}_\varsigma : \mathbb{R}^{L^2} \to \mathbb{R}$ is a scalar function defined by:

$$\mathcal{R}_\varsigma(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - N_\varsigma(\mathbf{x})\|^2, \tag{5.6}$$

with $N_\varsigma : \mathbb{R}^{L^2} \to \mathbb{R}^{L^2}$ being a neural network. The idea is thus to directly train the denoiser $D_\varsigma$ to denoise images corrupted by Gaussian noise of variance $\varsigma > 0$ as an explicit gradient step. Note the different notation $\varsigma$, here, for the noise variance, to not be confused with the variance of Gaussian noise in the measurements normally noted by $\sigma^2$ in this thesis. By inserting this denoiser (5.5) into well-known iterative schemes such as HQS, PGD, ADMM, and DRS, the authors are able to obtain iterative schemes converging to a stationary point of a function with an explicit (though complicated) expression.

## 5.2 PnP-COL0RME: Deconvolution using a PnP denoiser

We exploit the ideas above to define PnP-COL0RME, a method allowing for the amelioration of the support estimation (first) step of COL0RME, the method presented in Chapter 2. We recall that the drawback encountered in the use of the fully model-based approach we developed for support estimation, was the selection of an appropriate regularization penalty, which we choose to enforce sparsity in some sense (see Section 2.2 for more details). As discussed in Figure 2.19, real data results obtained were punctuated. The selection of a more appropriate regularization penalty modeling curvilinear structures led us to use PnP methods, due to their flexibility. As seen before, the regularizer is here parameterized by a deep neural network that can be trained on simulated data characterized by the desired structures of interest (e.g. filaments) and therefore capturing better the shape of the available data.

We recall that for the support estimation in COL0RME, the reformulated model in the covariance domain (2.12) was used. As input, the sample covariance matrix $\mathbf{R}_{\mathbf{y}}$, including auto-covariances and cross-covariances of the acquired data, was computed. Here, as a simplification and in order to benefit from easier calculations, we will only make use of its main diagonal, or similarly, of the auto-covariances.

### 5.2.1 Adaptation of COL0RME for temporal auto-covariances

In this section, we slightly reformulate problem (2.13) for the estimation of the support in COL0RME. We do so by considering in the forward model (2.12) only temporal auto-covariances, thus neglecting cross-terms. We start from the simplified model (2.6) describing acquisition $\mathbf{y}_t \in \mathbb{R}^{M^2}$ (a column-wise vectorization of image $\mathbf{Y}_t \in \mathbb{R}^{M\times M}$):

$$\mathbf{y}_t = \mathbf{\Psi}\mathbf{x}_t + \mathbf{b} + \mathbf{n}_t, \quad t \in \{1, \dots, T\} \tag{5.7}$$

where $\mathbf{\Psi} \in \mathbb{R}^{M^2\times L^2}$ is the (known) forward operator resulting from the composition of a down-sampling and a convolution operator and $\mathbf{b} \in \mathbb{R}^{M^2}$ is the background image. For all $t$, $\mathbf{n}_t \in \mathbb{R}^{M^2}$ is the noise component defined here as a matrix of i.i.d Gaussian elements with variance $s \geq 0$, $(n_t)_i \sim \mathcal{N}(0, s), \forall i \in \{1, ..., M^2\}$, and $\mathbf{x}_t \in \mathbb{R}^{L^2}$ is the clean image belonging in a $q$-times finer grid with $L = qM$.

Considering only temporal auto-covariances, we can proceed similarly as in Section 2.1.2. We first consider the discrete frames $\{\mathbf{y}_t\}_{t=1}^T$ as $T$ realizations of a random variable $\mathbf{y}$. The auto-covariance matrix of $\mathbf{y}$ is defined by:

$$\tilde{\mathbf{r}}_{\mathbf{y}} = \mathbb{E}_{\mathbf{y}}\{(\mathbf{y} - \mathbb{E}_{\mathbf{y}}\{\mathbf{y}\})^{\circ 2}\}. \tag{5.8}$$

where $\mathbb{E}_{\mathbf{y}}\{\cdot\}$ denotes the expected value computed w.r.t. the unknown law of $\mathbf{y}$, while $(\cdot)^{\circ 2}$ denotes an element-wise raise to the power of 2. Practically, we estimate $\tilde{\mathbf{r}}_{\mathbf{y}} \in \mathbb{R}^{M^2}$ by computing the empirical auto-covariance matrix:

$$\tilde{\mathbf{r}}_{\mathbf{y}} \approx \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{y}_t - \overline{\mathbf{y}})^2, \tag{5.9}$$

where $\overline{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t$ denotes the empirical temporal mean.

From (5.7) and (5.9), we can retrieve the following model between sample auto-covariances of data and desired signals:

$$\tilde{\mathbf{r}}_{\mathbf{y}} = \mathbf{\Psi}^2 \mathbf{r}_{\mathbf{x}} + \tilde{\mathbf{r}}_{\mathbf{n}}. \tag{5.10}$$

where $\mathbf{r}_{\mathbf{x}} \in \mathbb{R}^{L^2}$ and $\tilde{\mathbf{r}}_{\mathbf{n}} \in \mathbb{R}^{M^2}$ are the empirical auto-covariance matrices of $\{\mathbf{x}_t\}_{t=1}^T$ and $\{\mathbf{n}_t\}_{t=1}^T$, respectively. With the notation $\mathbf{\Psi}^2 \in \mathbb{R}^{M^2 \times L^2}$, we simply designate the matrix having the square of each element of matrix $\mathbf{\Psi} \in \mathbb{R}^{M^2 \times L^2}$. Finally, since the noise component $\mathbf{n}_t$ has i.i.d. elements of zero mean and constant variance $s$, it turns out that $\tilde{\mathbf{r}}_{\mathbf{n}} := s\mathbf{1}_{M^2}$.

Note that $\tilde{\mathbf{r}}_{\mathbf{y}} \in \mathbb{R}^{M^2}$ is in fact the result of the 2nd order SOFI method [Dertinger et al., 2009], i.e. a so-called SOFI image. In COL0RME, we used instead a matrix $\mathbf{R}_{\mathbf{y}} \in \mathbb{R}^{M^2 \times M^2}$ (vectorized in a vector $\mathbf{r}_{\mathbf{y}} \in \mathbb{R}^{M^4}$) which has as its main diagonal the SOFI image $\tilde{\mathbf{r}}_{\mathbf{y}}$, but showing also cross-covariances terms outside the main diagonal. Although using the full $\mathbf{R}_{\mathbf{y}}$ we are able to make better use of all the information we have, considering it to be diagonal is a fairly good simplification that allows benefiting from faster calculations. In future work, the whole matrix $\mathbf{R}_{\mathbf{y}}$ should be considered.

To find $\mathbf{r}_{\mathbf{x}}$, the support $\Omega := \{i : (\mathbf{r}_{\mathbf{x}})_i \neq 0\}$ and the noise variance $s$ using only temporal auto-covariances we consider the following minimization problem analogous to (2.13) presented in Chapter 2:

$$(\hat{\mathbf{r}}_{\mathbf{x}}, \hat{s}) \underset{\mathbf{r}_{\mathbf{x}} \geq 0,\ s \geq 0}{\arg\min} \ \frac{1}{2} \|\tilde{\mathbf{r}}_{\mathbf{y}} - \mathbf{\Psi}^{\mathbf{2}} \mathbf{r}_{\mathbf{x}} - s\mathbf{1}_{M^2}\|_2^2 + \lambda R(\mathbf{r}_{\mathbf{x}}), \tag{5.11}$$

where $\lambda > 0$ is a regularization parameter and $R(\cdot)$ is a regularization penalty to be defined.

We use the Alternate Minimization algorithm between $s$ and $\mathbf{r}_{\mathbf{x}}$ and we report the pseudo-code in Algorithm 4.

**Algorithm 4** Support Estimation of COL0RME (using only auto-covariances)

**Require:** $\tilde{\mathbf{r}}_{\mathbf{y}} \in \mathbb{R}^{M^2}, {\mathbf{r}_{\mathbf{x}}}^0 \in \mathbb{R}^{L^2}, \tau, \lambda > 0$

**repeat**

$s^{k+1} = \frac{1}{M^2} \mathbf{1}_{M^2}^T (\tilde{\mathbf{r}}_{\mathbf{y}} - \mathbf{\Psi}^{\mathbf{2}} {\mathbf{r}_{\mathbf{x}}}^k)$

$\mathbf{z}^{k+1} = {\mathbf{r}_{\mathbf{x}}}^k - \tau (\mathbf{\Psi}^{\mathbf{2}})^T (\tilde{\mathbf{r}}_{\mathbf{y}} - \mathbf{\Psi}^{\mathbf{2}} {\mathbf{r}_{\mathbf{x}}}^k - s^{k+1} \mathbf{1}_{M^2})$

${\mathbf{r}_{\mathbf{x}}}^{k+1} = \text{prox}_{\tau\lambda R}(\mathbf{z}^{k+1})$

**until** convergence

**return** $\Omega := \{i : (\mathbf{r}_{\mathbf{x}})_i \neq 0\}, s$

### 5.2.2 PnP-COL0RME

We can now replace the proximal step appearing in the Algorithm 4 with an off-the-shelf denoiser. The new method called PnP-COL0RME makes use of a proximal gradient step denoiser as proposed by Hurault *et al.* in [Hurault et al., 2022b]. In their paper, the authors showed that this choice is the proximal operator of a non-convex smooth function which allows the authors to derive convergence guarantees of the resulting PGD scheme. Using, thus, the denoiser $D_\varsigma$ given by (5.5), we can define Algorithm 5 which is used to estimate support $\Omega$ and noise variance $s$.

**Algorithm 5** Support Estimation via COL0RME-PnP

**Require:** $\tilde{\mathbf{r}}_\mathbf{y} \in \mathbb{R}^{M^2}, {\mathbf{r}_\mathbf{x}}^0 \in \mathbb{R}^{L^2}, \varsigma, \tau > 0$

**repeat**

$$s^{k+1} = \frac{1}{M^2}\mathbf{1}_{M^2}^T(\tilde{\mathbf{r}}_\mathbf{y} - \mathbf{\Psi^2}{\mathbf{r}_\mathbf{x}}^k)$$

$$\mathbf{z}^{k+1} = {\mathbf{r}_\mathbf{x}}^k - \tau(\mathbf{\Psi^2})^T(\tilde{\mathbf{r}}_\mathbf{y} - \mathbf{\Psi^2}{\mathbf{r}_\mathbf{x}}^k - s^{k+1}\mathbf{1}_{M^2})$$

$${\mathbf{r}_\mathbf{x}}^{k+1} = D_\varsigma(\mathbf{z}^{k+1})$$

**until** convergence

**return** $\Omega := \{i : (\mathbf{r}_\mathbf{x})_i \neq 0\}, s$

In [Hurault et al., 2022b] the authors trained the denoiser $D_\varsigma$ as a gradient descent step (see (5.5)) of a functional parameterized by a deep neural network (see (5.6)). Making use of work by Gribonval & Nikolova on the characterization of proximity operators [Gribonval and Nikolova, 2020], they proved that $D_\varsigma$ is the proximal operator of a function $\phi_\varsigma$, given by:

$$\phi_\varsigma(\mathbf{r}_\mathbf{x}) := \mathcal{R}_\varsigma(D_\varsigma^{-1}(\mathbf{r}_\mathbf{x})) - \frac{1}{2}\|D_\varsigma^{-1}(\mathbf{r}_\mathbf{x}) - \mathbf{r}_\mathbf{x}\|_2^2 \tag{5.12}$$

Consequently, the function that is minimized by Algorithm 5 is:

$$F_\varsigma(\mathbf{r}_\mathbf{x}, s) := f(\mathbf{r}_\mathbf{x}, s) + \phi_\varsigma(\mathbf{r}_\mathbf{x}), \tag{5.13}$$

with

$$f(\mathbf{r}_\mathbf{x}, s) = \frac{1}{2}\|\mathbf{r}_\mathbf{y} - \mathbf{\Psi^2}\mathbf{r}_\mathbf{x} - s\mathbf{1}_{M^2}\|_2^2. \tag{5.14}$$

In each iteration it is thus possible to evaluate the objective function which is given by:

$$F_\varsigma({\mathbf{r}_\mathbf{x}}^k, s^k) := f({\mathbf{r}_\mathbf{x}}^k, s^k) + \mathcal{R}_\varsigma(\mathbf{z}^k) - \frac{1}{2}\|\mathbf{z}^k - {\mathbf{r}_\mathbf{x}}^k\|_2^2,$$

after combining (5.12) and (5.13) and using $\mathbf{z}^k = D_\varsigma^{-1}({\mathbf{r}_\mathbf{x}}^k)$. The convergence property of the PGD scheme in the non-convex setting above and used here for the update of $\mathbf{r}_\mathbf{x}$ in Algorithm 5 is proved in Theorem 4.1 of [Hurault et al., 2022b].

Note that in (5.13), the regularization parameter $\lambda > 0$ that was used before to regulate the strength of the regularization has now vanished. The new hyperparameter $\varsigma > 0$ has now the role of tuning the regularization strength. Intuitively, $\varsigma$ encourages the denoiser to denoise images corrupted by white Gaussian noise of variance $\varsigma$. However, in practice, it is difficult to adjust such a parameter. A scaling parameter acting similarly to a regularization parameter (such as $\lambda$) can thus be introduced along with $\varsigma$.

### 5.2.3 PnP-COL0RME with scaling parameter

Following [Xu et al., 2020], we thus consider a further parameter in the definition of $D_\varsigma$, the so-called *denoiser scaling* parameter $\mu > 0$. We define a new scaled denoiser as:

$$D_{\mu,\varsigma}(\mathbf{z}) := \frac{1}{\mu}D_\varsigma(\mu\mathbf{z}), \tag{5.15}$$

where the scaled denoiser simply scales the old denoiser's input by a positive constant $\mu$ and its output by the inverse of the same constant.

Such a technique for scaling the denoiser, according to [Xu et al., 2020], provides a mechanism for adjusting the strength of the denoiser in a more intuitive way than tuning $\varsigma$. Note, however, that $\varsigma$ remains a parameter of the denoiser, but it is not the one that will be fine-tuned to adjust the denoiser strength.

First we replace in the Algorithm 5, the denoiser step which will now be:

$$\mathbf{r_x}^{k+1} = D_{\mu,\varsigma}(\mathbf{z}^{k+1}), \tag{5.16}$$

and result to Algorithm 6. By now interpreting the denoiser $D_\varsigma$ as the proximal operator of a function $\phi_\varsigma$, we can write:

$$D_\varsigma(\mathbf{z}) = \text{prox}_{\phi_\varsigma}(\mathbf{z}). \tag{5.17}$$

Following Proposition 1 in [Xu et al., 2020], it is possible to deduce:

$$D_{\mu,\varsigma}(\mathbf{z}) = \text{prox}_{\mu^{-2}\phi_\varsigma(\mu\cdot)}(\mathbf{z}). \tag{5.18}$$

Therefore, the objective function which is being minimized, now reads:

$$F_{\mu,\varsigma}(\mathbf{r_x}, s) := f(\mathbf{r_x}, s) + \frac{1}{\mu^2}\phi_\varsigma(\mu\mathbf{r_x}) \tag{5.19}$$

where the function $f(\cdot)$ is given by (5.14) and $\phi_\varsigma(\cdot)$ by (5.12).

**Algorithm 6** Support Estimation via COL0RME-PnP (scaling denoiser)

**Require:** $\tilde{\mathbf{r}}_\mathbf{y} \in \mathbb{R}^{M^2}, \mathbf{r_x}^0 \in \mathbb{R}^{L^2}, \varsigma, \tau, \mu > 0$

**repeat**

$s^{k+1} = \frac{1}{M^2}\mathbf{1}_{M^2}^T(\tilde{\mathbf{r}}_\mathbf{y} - \mathbf{\Psi^2}\mathbf{r_x}^k)$

$\mathbf{z}^{k+1} = \mathbf{r_x}^k - \tau(\mathbf{\Psi^2})^T(\tilde{\mathbf{r}}_\mathbf{y} - \mathbf{\Psi^2}\mathbf{r_x}^k - s^{k+1}\mathbf{1}_{M^2})$

$\mathbf{r_x}^{k+1} = D_{\mu,\varsigma}(\mathbf{z}^{k+1})$

**until** convergence

**return** $\Omega := \{i : (\mathbf{r_x})_i \neq 0\}, s$

In each algorithmic iteration the value of $F_{\mu,\varsigma}(\cdot,\cdot)$ is given by:

$$\begin{aligned} F_{\mu,\varsigma}(r_x{}^k, s^k) :=& f(\mathbf{r_x}^k, s^k) + \frac{1}{\mu^2}\left(\mathcal{R}_\varsigma(D_\varsigma^{-1}(\mu\mathbf{r_x}^k)) - \frac{1}{2}\|D_\varsigma^{-1}(\mu\mathbf{r_x}^k) - \mu\mathbf{r_x}^k\|_2^2\right) \\ =& f(\mathbf{r_x}^k, s^k) + \frac{1}{\mu^2}\left(\mathcal{R}_\varsigma(\mu\mathbf{z}^k) - \frac{1}{2}\|\mu\mathbf{z}^k - \mu\mathbf{r_x}^k\|_2^2\right) \\ =& f(\mathbf{r_x}^k, s^k) + \frac{1}{\mu^2}\mathcal{R}_\varsigma(\mu\mathbf{z}^k) - \frac{1}{2}\|\mathbf{z}^k - \mathbf{r_x}^k\|_2^2 \end{aligned} \tag{5.20}$$

where we used:

$$\mathbf{r_x}^k = D_{\mu,\varsigma}(\mathbf{z}^k) = \frac{1}{\mu}D_\varsigma(\mu\mathbf{z}^k) \Leftrightarrow D_\varsigma^{-1}(\mu\mathbf{r_x}^k) = \mu\mathbf{z}^k \tag{5.21}$$

Adapting the results of Theorem 4.1 in [Hurault et al., 2022b] to the rescaled case is a work in progress. However, we observe empirical convergence as indicated in Section 5.3.3.

Note that $\mu$ plays the role of a regularization parameter. Even if the relation between $\mu$ and the regularization parameter in front of the regularization ($\frac{1}{\mu^2}$) is not linear, $\mu$ still provides a more explicit mechanism to adjust the denoiser [Xu et al., 2020]. Results on support reconstructions after selecting different values for the $\mu$ parameter are available in Section 5.3.2.

## 5.3 Numerical results

In this section, we present some preliminary results obtained by using PnP-COL0RME for estimating the support of a super-resolved image using Algorithm 6. In Section 5.3.1 we describe in detail how we trained the denoiser, then in Section 5.3.2 we investigate the sensitivity of the method to the parameters $\mu$ and $\varsigma$ and, finally, in Sections 5.3.3 and 5.3.4 we report results on simulated and real data (high-density SMLM data and Ostreopsis data acquired by a confocal microscope). The effective use of the approach is confirmed on the simple task of deblurring (by setting the super-resolution factor $q$ equal to 1 and therefore $M = L$) while leaving the actual problem of super-resolution for future work.

### 5.3.1 Training the denoiser

To train the denoiser $D_\varsigma$ we created a dataset composed of clean and noisy image pairs. The structures appearing in this dataset must share the same geometry as the structures appearing in the images to restore. Differently from other methods, the proposed algorithm works with a model formulated in the covariance domain, so that the denoiser takes as an input noisy sample auto-covariance matrices of a fluctuating temporal sequence of images. Hence, to create the dataset we first started by creating different spatial patterns (thin filaments) shown in Figure 5.1 where the emitters have different positions in the continuous grid. Such patterns are the superposition, after rotations with different angles, of the ground truth spatial pattern provided in the MT0 microtubule training dataset uploaded for SMLM 2016 challenge [1]. Then, we used the fluctuation model discussed in [Girsault et al., 2016] to simulate temporal fluctuations and create a temporal stack of $T = 500$ frames for each spatial pattern. Two exemplar frames of one temporal stack of images are reported in Figures 5.2a and 5.2b. For each temporal stack of images, we could therefore calculate the temporal auto-covariance image (see Figure 5.2c) corresponding to one instance of the clean images $\mathbf{r}_{\mathbf{x}}^{\mathrm{GT}}$ in our dataset. To create now its noisy version we added Gaussian noise $\boldsymbol{\eta}$ with constant variance $\varsigma^2$, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \varsigma^2\mathbf{Id})$, with $\varsigma$ following a uniform distribution, $\varsigma \sim \mathcal{U}(\varsigma_1, \varsigma_2)$. We remark that since the noise in the covariance data comes from additive Gaussian noise on the individual frames, its actual distribution is, indeed, $\chi^2$. However, since the number of degrees of freedom is high (as $T = 500$), the distribution can be approximated by a Gaussian distribution. In our experiments, after normalising $\mathbf{r}_{\mathbf{x}}^{\mathrm{GT}}$ with maximum value equal to 1, we select $\varsigma_1 = \epsilon << 1$ and $\varsigma_2 = 50/255$.

The training was performed by following the procedure suggested in [Hurault et al., 2022b] and using the code available on the authors' GitHub repository [2]. For the neural network $N(\cdot)$ used to parameterize the denoiser (see (5.5), (5.6)), we used DRUNet, a CNN proposed in

1. https://srm.epfl.ch/Challenge/ChallengeSimulatedData
2. https://github.com/samuro95/Prox-PnP

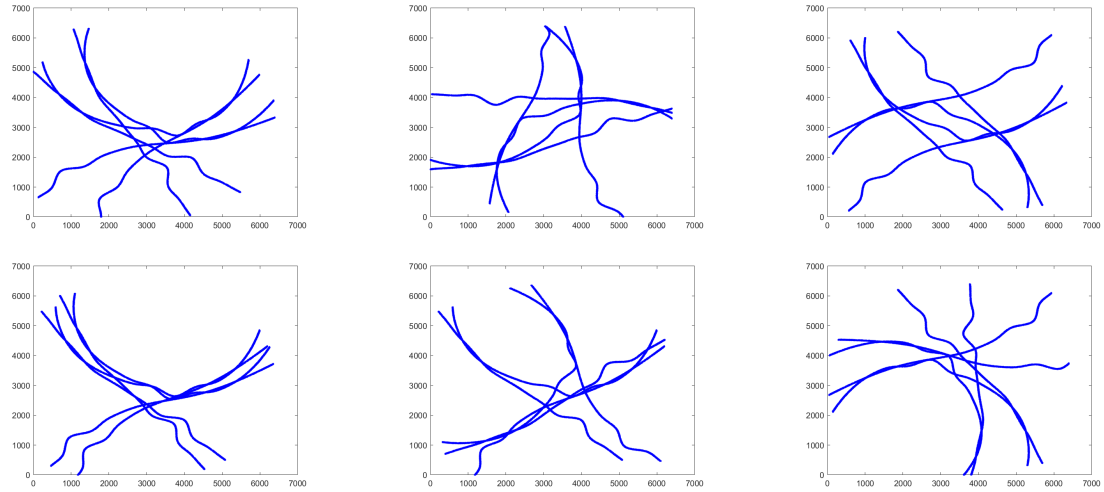
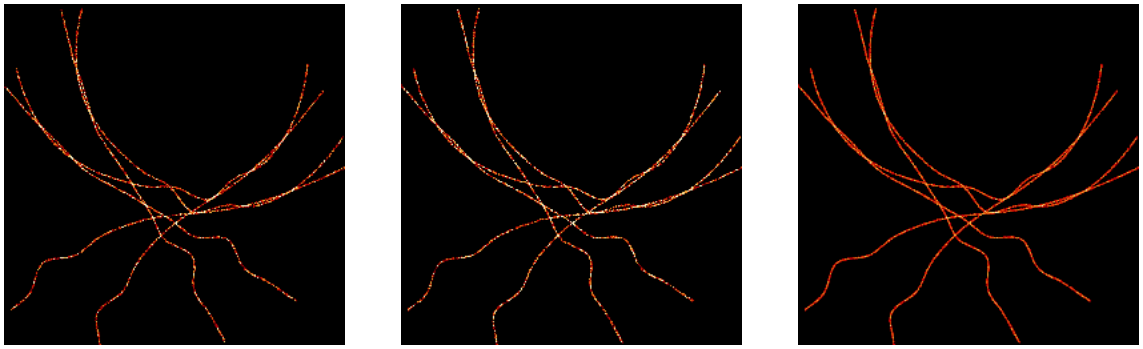


Figure 5.1 – Simulated spatial patterns



(a) (b) (c)

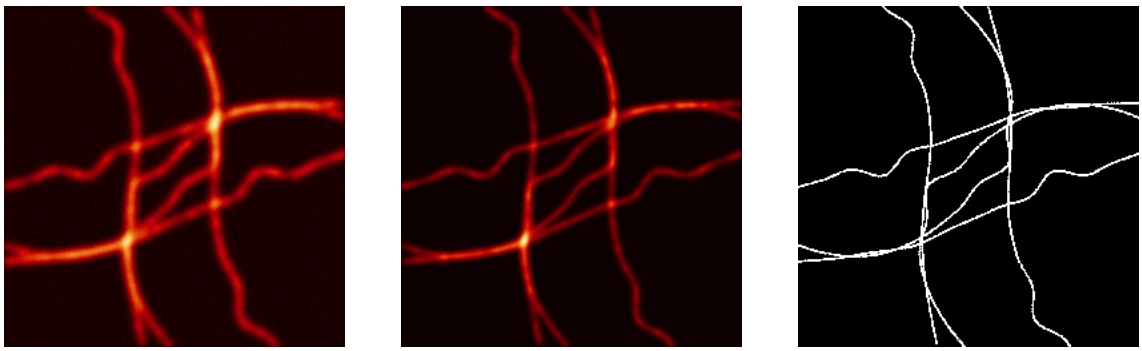
Figure 5.2 – (a-b) Two different frames of a simulated fluctuating stack made from the first spatial pattern from Figure 5.1, (c) The auto-covariance image $\mathbf{r}_{\mathbf{x}}^{\mathrm{GT}}$ estimated from the whole temporal sequence.

[Zhang et al., 2021]. For the training, we used 500 pairs of clean-noise auto-covariance images and 100 for validation. The network was trained using 1215 epochs via ADAM optimization and batch size equal to 16.

### 5.3.2 Sensitivity to denoising and scaling parameters

As discussed in Section 5.2.3, the performance of PnP-COL0RME depends on two hyper-parameters $\mu$ and $\varsigma$. In this section, we discuss how these two parameters affect the support reconstruction obtained by the method. For this reason, we take as an input the temporal auto-covariance image $\tilde{\mathbf{r}}_{\mathbf{y}}$ of a diffraction-limited (blurred) stack of images reported in Figure 5.3b, and we try to estimate a support close to the real one (ground truth support), shown in Figure 5.3c. For this example, neither background nor Poisson noise was considered for the generation of the data, while Gaussian noise was added. Reconstructions of PnP-COL0RME for different combinations

of $\mu$ and $\varsigma$ are reported in Figure 5.4. To make Figure 5.4 more readable, we increase the parameter $\mu$ along the horizontal axis (from left to right), while we increase the parameter $\varsigma$ along the vertical axis (from up to down).



(a) (b) (c)

Figure 5.3 – (a) A diffraction-limited frame generated by model (1.28) with no background and Poisson noise corruption, (b) The temporal auto-covariance image $\tilde{\mathbf{r}}_{\mathbf{y}}$ of the stack of diffraction-limited images, (c) The ground truth binary support image.

As we can see, the reconstructions we obtain are quite different: in some cases filaments are missing, while in others double filaments appear. Roughly, with a small $\mu$ it is possible to miss some filaments, while with a larger $\mu$ it is possible to have duplicate filaments. Being aware of such a pattern, we can increase $\mu$ until we are sure that we have included all filaments in the image.

### 5.3.3 Simulated data

We first apply PnP COL0RME to simulated data presented in Figure 5.5. The spatial pattern used here is the superposition, after rotations, with different angles of the ground truth spatial pattern provided in the MT0 microtubule training dataset uploaded for SMLM 2016 [3] which was previously used in Section 2.4.1. The PSF used to generate the data has an FWHM equal to 176.6 nm, the pixel size is equal to 25 nm, and the images have a size of $256 \times 256$ pixels. The support reconstruction obtained after setting the parameter $\mu$ equal to 1 and the parameter $\varsigma$ equal to $30/255$ is reported in Figure 5.5g. Along with the reconstructed image, we report a graph showing the evolution of the objective function $F_{\mu,\varsigma}$ in (5.20). Following [Hurault et al., 2022a, Hurault et al., 2022b], we present also the evolution of the quantity $\min_{i \leq k} \|\mathbf{r_x}^{i+1} - \mathbf{r_x}^{i}\|^2 / \|\mathbf{r_x}^{0}\|^2$ used in the non-convex setting to analyze the convergence rate of the algorithm, that is shown to be less than $\mathcal{O}(\frac{1}{k})$. Empirical convergence is observed for both quantities.

Thanks to its training, we observe that the proposed approach is able to capture the filaments' geometry fairly well. We observe that in comparison to the ground truth support in Figure 5.5c, the reconstruction in Figure 5.5g is rather accurate. For the evaluation of the localization precision the Jaccard Index (JI) has been used. It is a quantity in $[0, 1]$ computed as the ratio between correct detections (CD) and the total (correct, false negatives false positive) detections, i.e. JI $:= CD/(CD + FN + FP)$, up to a tolerance $\delta > 0$, measured in nm (see, e.g., [Sage et al., 2015]). For the reconstruction in Figure 5.5g, the tolerance precision was chosen $\delta = 40$ nm.

3. https://srm.epfl.ch/Challenge/ChallengeSimulatedData

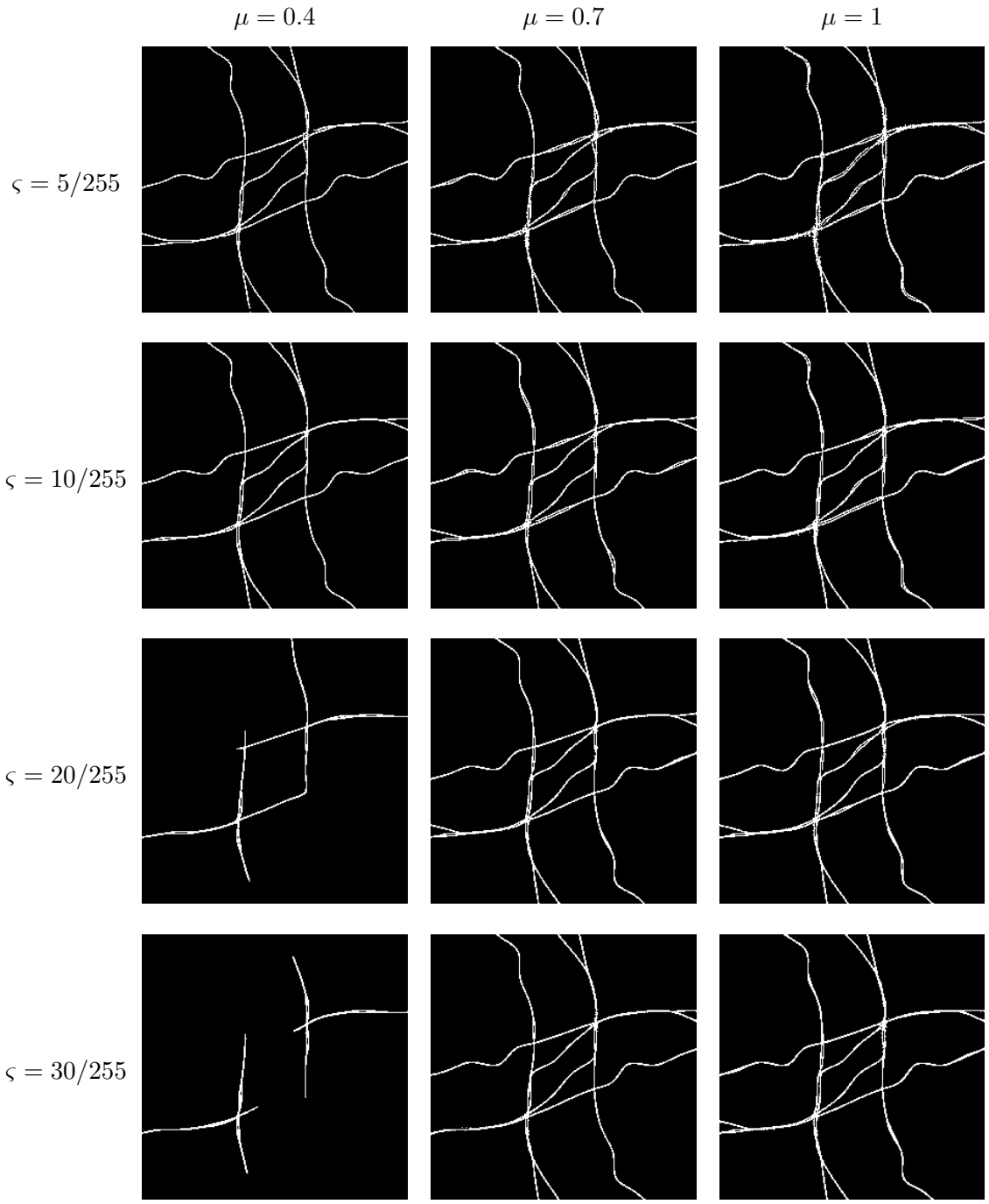
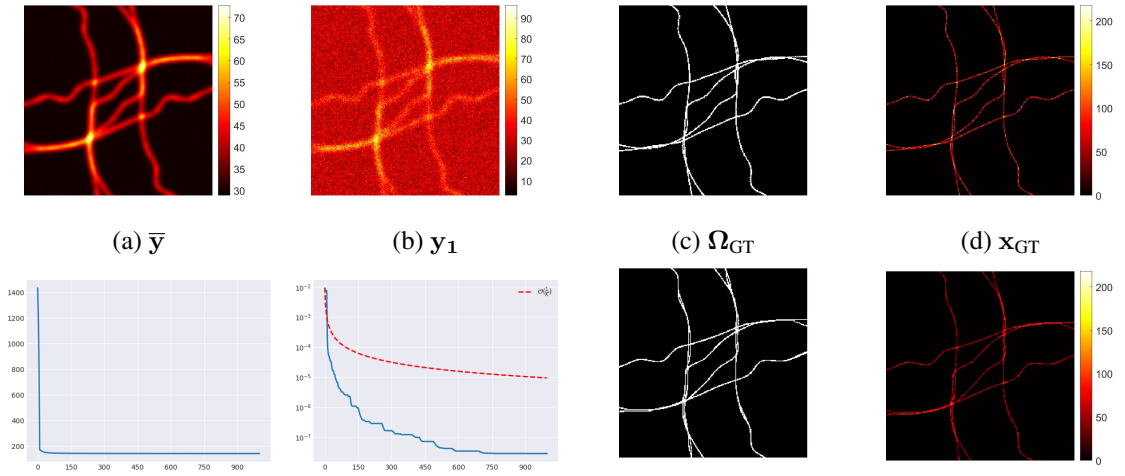



Figure 5.4 – PnP-COL0RME binary support image reconstruction of Figure 5.3b for different hyper-parameters $\mu$ and $\varsigma$.

Moreover, with the second step of the method COL0RME, intensities can also be estimated with high precision, see Figure 5.5h. However, for the challenging dataset in Figure 5.5 and avoiding fine-tuning the $\mu$ parameter, the appearance of small artifacts (e.g. incorrect duplication



(a) $\overline{\mathbf{y}}$ (b) $\mathbf{y_1}$ (c) $\mathbf{\Omega}_{\text{GT}}$ (d) $\mathbf{x}_{\text{GT}}$

(e) (f) (g) $\hat{\mathbf{\Omega}}$ (JI=0.67) (h) $\hat{\mathbf{x}}$ (PSNR=26.01)

Figure 5.5 – (a) Mean of the acquired temporal sequence, (b) First frame (c) The ground truth support (d) The ground truth intensity image (e) Evolution of cost function $F_{\mu,\varsigma}$ in (5.20) (f) the evolution of $\min_{i\leq k} \|\mathbf{r_x}^{i+1} - \mathbf{r_x}^{i}\|^2/\|\mathbf{r_x}^{0}\|^2$, in logarithmic scale, (g) Reconstructed support (h) Reconstructed intensity image.

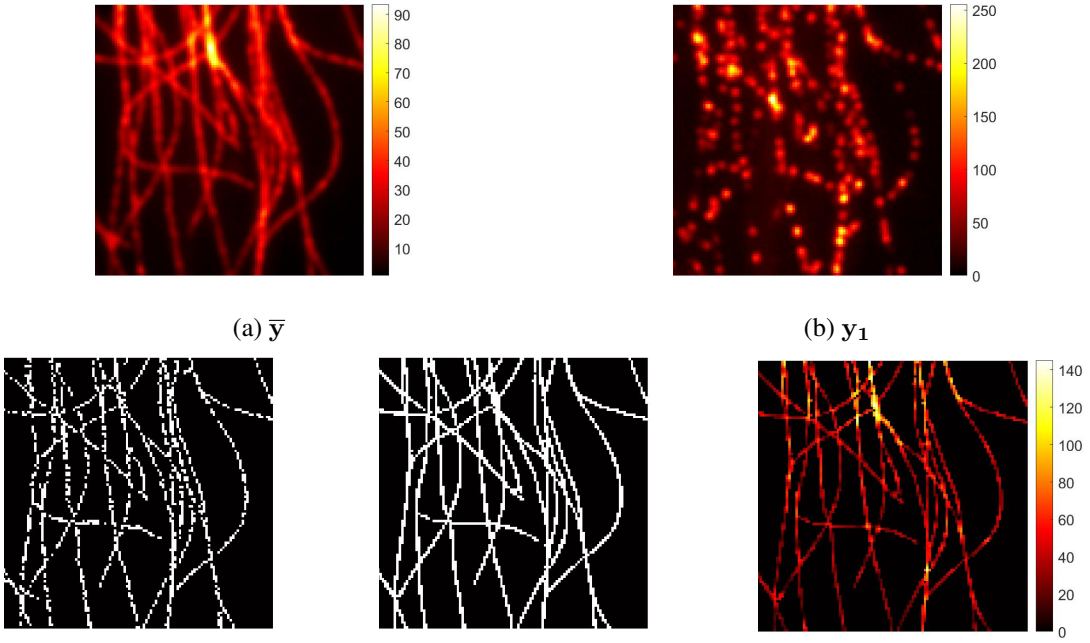
of filaments) due to the training dataset we built is observed. They could be potentially removed by retraining the model with more heterogeneous data.

### 5.3.4 Real data

We then applied the proposed approach to high-density SMLM acquisitions using a publicly available dataset created for the 2013 SMLM challenge [4], see Figure 5.6. Although in SMLM the molecules do not have a blinking behavior, but rather an on-to-off transition, we can consider blinking the temporal behavior of one pixel in high-density videos due to the presence of many molecules per pixel. The dataset contains $T = 500$ images, the PSF of the microscope used to acquire these data has an FWHM of 351.8 nm and the pixel size is equal to 100 nm. The support $\hat{\mathbf{\Omega}}$ computed by the model-based COL0RME approach (see Chapter 2) is compared to the one PnP-COL0RME variant of Algorithm 6 with $\mu = 1$. Since no ground truth is available for these data, no quantitative assessment can be computed, however, better continuation properties than COL0RME are observed.

Finally, we applied PnP-COL0RME on images of a real biological sample of the unicellular alga *Ostreopsis*. For this dataset, $T = 500$ images were acquired using a confocal microscope. The pixel size of the CCD camera used is equal to 70 nm, while the FWHM of the PSF has been estimated at 229 nm. Results for parameter $\varsigma = 10/255$ and for different parameters $\mu$ are available in Figure 5.7. We can see that we are able to reconstruct curvilinear structures, however, if we want to include all filaments in the image (increase the parameter $\mu$), double filaments instead of single filaments may appear.

4. https://srm.epfl.ch/Challenge/Challenge2013




(a) $\overline{\mathbf{y}}$ (b) $\mathbf{y_1}$

(c) $\hat{\mathbf{\Omega}}$ by COL0RME-$\ell_1$ (d) $\hat{\mathbf{\Omega}}$ by PnP-COL0RME (e) $\hat{\mathbf{x}}$

Figure 5.6 – HD-SMLM data: (first row) The temporal mean and the first frame of the acquired temporal sequence (second row) Support (COL0RME VS. PnP-COLORME) and intensity reconstruction, $\varsigma = 10/255$ and $\mu = 1$.

(a) (b) (c)

Figure 5.7 – Real Ostreopsis data (a) The temporal auto-covariance image $\tilde{\mathbf{r}}_{\mathbf{y}}$, (b) COL0RME PnP support reconstruction using $\varsigma = 10/255$ and $\mu = 0.8$, (c) PnP-COL0RME support reconstruction using $\varsigma = 10/255$ and $\mu = 0.9$.

## 5.4 Discussion

We present PnP-COL0RME, a method when an off-the-shelf denoiser is used in place of the proximity operator associated with a regularization functional, so as to improve the support estimation step in COL0RME. We choose to use the denoiser form proposed in [Hurault et al., 2022b]

due to its provable convergence guarantees. Following [Xu et al., 2020] we also introduced a scaling parameter that acts as a regularization weight able to adjust the strength of the denoiser.

Our results show that the geometry of specific structures (filaments) can be captured by suitable training. However, there is still room for further improvements in the method. First of all, we would like to select the parameter $\mu$ automatically so that the method is able to give accurate results without having to finely tune it. Furthermore, we would like to test the approach including also a down-sampling operator as well as to take into account the complete covariance matrix of our diffraction-limited images.

# CHAPTER 6

# Conclusions & Outlook

In this thesis, we focused on the development of new approaches based on optimization and learning for improving spatial resolution in fluorescence microscopy. Our main motivation is to implement methods that can be applied to images acquired by *conventional* fluorescence microscopes using *conventional* fluorescent dyes for studying living samples. We have mainly focused on improving the resolution in the lateral plane by taking advantage of the stochastic temporal fluctuations of individual fluorescent molecules, and we have also implemented a 3D super-resolution method by taking advantage of the MA-TIRF microscope. In the following, we summarize the contributions made in this thesis and discuss some future perspectives.

## 6.1 Summary of the main contributions

**COL0RME** We propose COL0RME, a method for lateral resolution enhancement implemented within a variational framework. It combines two steps, the support and the intensity estimation step, which are performed sequentially. The support estimation step allows for precise molecule localization, while the intensity estimation step computes real intensities on the pre-estimated support. To obtain a good support reconstruction we take advantage of the independent statistical behavior of standard fluorescent emitters by reformulating the problem using the covariance matrix of the acquisitions. In such covariance setting, we search for sparse solutions by considering convex and non-convex regularization penalties in the expression of the function we minimize. By introducing an additional variable to denote the variance of the noise in the minimization, we can incorporate a strategy for its estimation, which enables us to handle very noisy acquisitions. By incorporating selection strategies for the regularization parameters, we further make the method automatic. The second and simpler step is of great importance since intensity information is valuable for biological interpretation and for 3D image reconstruction (see 3D MA-TIRF COL0RME). COL0RME is the first method exploiting independent molecules' fluctuations capable of providing such information. These two innovative steps lead to accurate reconstructions with good levels of resolution of simulated and real 2D TIRF data. The reconstructions are comparable and mostly better than other state-of-the-art methods such as, e.g., SRRF [Gustafsson et al., 2016] and SPARCOM [Solomon et al., 2019].

**3D MA-TIRF COL0RME** Taking advantage of an MA-TIRF microscope, which can acquire images at different incident angles of the illumination beam, thus providing access to depth information, we develop 3D MA-TIRF COL0RME, an extension of COL0RME suited for this particular setting. Starting from a temporal stack of images acquired at different angles of incidence, the proposed method generates a high-resolution 3D image. For the implementation of the method, we successfully decouple the problem of finding the final super-resolved 3D image into

two sub-problems: first, the lateral super-resolved images for each incidence angle are obtained using COL0RME; then, they serve as input for an MA-TIRF reconstruction model, which improves the axial resolution. To our knowledge, this is the first 3D super-resolution approach for MA-TIRF acquisitions. The method has been verified on simulated and real MA-TIRF data and shows significantly improved spatial resolution.

Due to the interest in deploying data-driven approaches to better capture the distribution of the acquired data or the geometries of the structures they contain, while benefiting from the physical model to compute solutions with physical sense, we have implemented the following three hybrid approaches: FluoGAN, DivBlurring and PnP-COL0RME.

**FluoGAN** FluoGAN is a novel framework for fluctuation-based resolution enhancement in fluorescence microscopy. It combines the physical modeling of the optical system with data-driven adversarial learning. It computes the desired deblurred/super-resolved image by comparing, in an appropriate sense, the empirical distribution of the observed data with the one of samples generated by a physically-grounded simulator. It is fully unsupervised and has a similar architecture to GANs, the main difference being the replacement of the model-blind generator network with a simulator model encoding biophysical expertise in its structure and having as learnable parameters the quantities of interest. FluoGAN has been validated on simulated data, reaching better levels of resolution than standard model-based and state-of-the-art approaches, but also on a calibrated 2D sample and on difficult real Ostreopsis data.

**DivBlurring** The incorporation of the simplified physical model describing acquisitions by conventional fluorescence microscopes into the decoder of a conventional VAE has led to the DivBlurring approach. This is a new approach, closely related to another approach called DivNoising, which further takes into account the blurring due to the PSF of the microscope. Given a noisy and diffraction-limited observation, but also limiting the solution space with prior knowledge, DivBlurring is able to generate multiple and diverse samples from a distribution of possible clean images. As this is still a work in progress, only a few results on simulated data are reported in this manuscript.

**PnP-COL0RME** In the interest of obtaining a support reconstruction for COL0RME that is better adapted to specific sample geometries (e.g. filaments), we presented PnP-COL0RME. Following the plug-and-play (PnP) reconstruction framework, the structure of PnP-COL0RME is similar to the COL0RME support estimation step, where it replaces the proximity operator of the regularization term with an image denoiser (i.e. a pre-trained network). A challenging aspect of the approach is the training of the denoiser, which in this setting is applied to temporal auto-covariance matrices of simulated sequences of fluctuating fluorescent molecules. After suitable modeling simplifications, the proposed method is tested both on simulated and on real Ostreopsis fluorescence microscopy images for a simpler deblurring task, considering for the moment only temporal auto-covariances. The obtained results are promising, thus paving the way for the effective use of PnP-COL0RME in more challenging super-resolution problems.

## 6.2 Future Prospects

At the beginning of this thesis, we discussed variational approaches. Such approaches have the imaging model of fluorescence microscopy as a core element. By additionally incorporating prior knowledge into the reconstruction process, we demonstrated that variational approaches work well even under difficult imaging conditions. In parallel, data-driven approaches are able to extract information from real datasets that are otherwise very difficult to access. The path we followed in the second part of the thesis was to combine the two approaches (model-based and data-driven) and develop model-aware learning methods so that we can benefit from the best of both worlds.

Since we believe that such a pathway is likely to lead to powerful new developments in super-resolution fluorescence microscopy, we envisage that future research could focus on improving both modeling accuracy and learning. For example, in FluoGAN (see Section 4.3) a future prospect is to create an (even more) accurate model describing observations of fluctuating fluorescent emitters but also to make the learning approach more reliable by proving convergence guarantees. Similarly, for DivBlurring (see Section 4.4), more accurate modeling should be considered allowing for reconstructions in a finer grid than the one of the observations, but also a better investigation of the learning procedure (e.g. considering a different loss function to minimize or using a different architecture like a Hierarchical VAE proposed, e.g., in [Vahdat and Kautz, 2020]).

The idea of combining model-based and data-driven approaches can be well described, also, within the PnP framework. Thanks to its structure, it has two separate steps: a gradient step that takes into account the imaging model and a denoising step that is independent of the model and associated with knowledge (of the structure of the object to be imaged) that can be extracted from the data. The proposed method, PnP-COL0RME, with the promising preliminary results presented in Chapter 5, can be further improved in many different ways. One could, for instance, make better use of the information on the independent blinking behavior of the fluorescence emitters, i.e., considering the full covariance matrix of the observations, estimating an image with a finer grid than that of the acquisitions, and developing an automatic hyper-parameter selection strategy.

# Appendix

# A Proximal computations

Given the function $h : \mathbb{R}^{L^2} \to \mathbb{R}$, defined in (2.27), the proximal mapping of $h$ is a an operator given by:

$$\begin{aligned}\text{prox}_{h,\tau}(\mathbf{w}) &= \arg\min_{\mathbf{u}} \left( \frac{1}{2\tau}\|\mathbf{u}-\mathbf{w}\|_2^2 + h(\mathbf{w}) \right) \\ &= \arg\min_{\mathbf{u}} \left( \frac{1}{2\tau}\|\mathbf{u}-\mathbf{w}\|_2^2 + \frac{\alpha}{2}\left( \|\mathbf{I}_{\mathbf{\Omega}}\mathbf{u}\|_2^2 + \sum_{i=1}^{L^2} [\phi(u_i)]^2 \right) \right). \end{aligned} \tag{A.1}$$

The optimal solution $\hat{\mathbf{u}}$ ($\hat{\mathbf{u}} = \text{prox}_{h,\tau}(\mathbf{w})$), as the problem (A.1) is convex, is attained when:

$$\begin{aligned} \mathbf{0}_{L^2} &\in \nabla \left( \frac{1}{2\tau}\|\hat{\mathbf{u}}-\mathbf{w}\|_2^2 + \frac{\alpha}{2}\left( \|\mathbf{I}_{\mathbf{\Omega}}\hat{\mathbf{u}}\|_2^2 + \sum_{i=1}^{L^2} [\phi(\hat{u}_i)]^2 \right) \right), \\ \mathbf{0}_{L^2} &\in \frac{1}{\tau}(\hat{\mathbf{u}}-\mathbf{w}) + \alpha \left( \mathbf{I}_{\mathbf{\Omega}}\hat{\mathbf{u}} + \ [\phi(\hat{u}_i)\phi'(\hat{u}_i)]_{\{i=1,\dots,L^2\}} \right), \end{aligned} \tag{A.2}$$

where we denote: $[x_i]_{\{i=1,\dots,L^2\}} = [x_1\, x_2 \dots x_{L^2}]^\intercal$.

Starting from (2.23) we can compute $\phi' : \mathbb{R} \to \mathbb{R}_+$, as:

$$\phi'(z) := \begin{cases} 0 & \text{if } z \geq 0, \\ 1 & \text{if } z < 0, \end{cases} \qquad \forall z \in \mathbb{R}. \tag{A.3}$$

Given (A.3), we can write:

$$\mathbf{0}_{L^2} \in \frac{1}{\tau}(\hat{\mathbf{u}}-\mathbf{w}) + \alpha \left( \mathbf{I}_{\mathbf{\Omega}}\hat{\mathbf{u}} + [\phi(\hat{u}_i)]_{\{i=1,\dots,L^2\}} \right). \tag{A.4}$$

Exploiting component-wise, as problem (A.1) is separable with respect to both $\mathbf{x}$ and $\mathbf{w}$, and assuming $\hat{u}_i \geq 0$, the derivative computed at (A.4) vanishes for:

$$\hat{u}_i = \frac{1}{1+\alpha\tau\mathbf{I}_{\mathbf{\Omega}}(i,i)} w_i, \tag{A.5}$$

and it holds for $w_i \geq 0$. Similarly, for the case $\hat{u}_i < 0$, this analysis yields:

$$\hat{u}_i = \frac{1}{1+\alpha\tau(\mathbf{I}_{\mathbf{\Omega}}(i,i)+1)} w_i, \tag{A.6}$$

for $w_i < 0$.

So finally, the proximal operator is given by:

$$\left(\text{prox}_{h,\tau}(\mathbf{w})\right)_i = \text{prox}_{h,\tau}(w_i) = \begin{cases} \frac{w_i}{1+\alpha\tau\mathbf{I}_{\mathbf{\Omega}}(i,i)} & \text{if } w_i \geq 0, \\ \frac{w_i}{1+\alpha\tau(\mathbf{I}_{\mathbf{\Omega}}(i,i)+1)} & \text{if } w_i < 0. \end{cases} \tag{A.7}$$

In a similar way, we compute the proximal mapping of the function $\overline{h} : \mathbb{R}^{L^2} \to \mathbb{R}$, defined in (2.41), as follows:

$$\begin{aligned}\text{prox}_{\overline{h},\tau}(\mathbf{z}) &= \arg\min_{\mathbf{u}} \left( \frac{1}{2\tau}\|\mathbf{u}-\mathbf{z}\|_2^2 + \overline{h}(\mathbf{u}) \right) \\ &= \arg\min_{\mathbf{u}} \left( \frac{1}{2\tau}\|\mathbf{u}-\mathbf{z}\|_2^2 + \frac{\alpha}{2}\left( \|\mathbf{I}_{\boldsymbol{\Omega}}\mathbf{u}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu}\mathbf{u}\|_2^2 \right) \right). \end{aligned} \tag{A.8}$$

The optimal solution $\hat{\mathbf{u}}$ of (A.8) ($\hat{\mathbf{u}} = \text{prox}_{\overline{h},\tau}(\mathbf{z})$) is attained when:

$$\begin{aligned}\mathbf{0}_{L^2} &\in \nabla \left( \frac{1}{2\tau}\|\hat{\mathbf{u}}-\mathbf{z}\|_2^2 + \frac{\alpha}{2}\left( \|\mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{u}}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu}\hat{\mathbf{u}}\|_2^2 \right) \right), \\ \mathbf{0}_{L^2} &\in \frac{1}{\tau}(\hat{\mathbf{u}}-\mathbf{z}) + \alpha\left( \mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{u}} + \mathbf{I}_{\hat{\mathbf{x}}_\mu}\hat{\mathbf{u}} \right). \end{aligned} \tag{A.9}$$

By eliminating $\hat{\mathbf{u}}$ in the expression (A.9), we compute element-wise the proximal operator:

$$(\text{prox}_{\overline{h},\tau}(\mathbf{z}))_i = \text{prox}_{\overline{h},\tau}(z_i) = \frac{z_i}{1+\alpha\tau\left( \mathbf{I}_{\boldsymbol{\Omega}}(i,i) + \mathbf{I}_{\hat{\mathbf{x}}_\mu}(i,i) \right)}. \tag{A.10}$$

# B The minimization problem to estimate $\hat{\mathbf{x}}'_\mu$

Starting from the penalized optimization problem (2.21) and having $\mathbf{b}$ fixed, we aim to find a relation that contains the optimal $\hat{\mathbf{x}}_\mu$. While there are only quadratic terms, we proceed as follows:

$$\begin{aligned}\mathbf{0} &\in \nabla \left( \frac{1}{2}\|\boldsymbol{\Psi}\hat{\mathbf{x}}_\mu - (\overline{\mathbf{y}}-\mathbf{b})\|_2^2 + \frac{\mu}{2}\|\nabla\hat{\mathbf{x}}_\mu\|_2^2 + \frac{\alpha}{2}\left( \|\mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{x}}_\mu\|_2^2 + \sum_{i=1}^{L^2}[\phi((\hat{\mathbf{x}}_\mu)_i)]^2 \right) \right), \\ \mathbf{0} &\in \boldsymbol{\Psi}^\intercal(\boldsymbol{\Psi}\hat{\mathbf{x}}_\mu - (\overline{\mathbf{y}}-\mathbf{b})) + \mu\nabla^\intercal\nabla\hat{\mathbf{x}}_\mu + \alpha\left( \mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{x}}_\mu + [\phi((\hat{\mathbf{x}}_\mu)_i)\phi'((\hat{\mathbf{x}}_\mu)_i)]_{\{i=1,\dots,L^2\}} \right). \end{aligned} \tag{B.11}$$

Given (A.3) we can write:

$$\mathbf{0} \in \boldsymbol{\Psi}^\intercal(\boldsymbol{\Psi}\hat{\mathbf{x}}_\mu - \overline{\mathbf{y}} - \mathbf{b}) + \mu\nabla^\intercal\nabla\hat{\mathbf{x}}_\mu + \alpha\left( \mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{x}}_\mu + [\phi((\hat{\mathbf{x}}_\mu)_i)]_{\{i=1,\dots,L^2\}} \right). \tag{B.12}$$

Our goal is to compute $\hat{\mathbf{x}}'_\mu$, the partial derivative of $\hat{\mathbf{x}}_\mu$ w.r.t. $\mu$. So, we derive as follows:

$$\begin{aligned}&\frac{\partial}{\partial\mu}\left( \mathbf{0} \in \boldsymbol{\Psi}^\intercal(\boldsymbol{\Psi}\hat{\mathbf{x}}_\mu - \overline{\mathbf{y}} - \mathbf{b}) + \mu\nabla^\intercal\nabla\hat{\mathbf{x}}_\mu + \alpha\left( \mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{x}}_\mu + [\phi((\hat{\mathbf{x}}_\mu)_i)]_{\{i=1,\dots,L^2\}} \right) \right), \\ &\mathbf{0} \in \boldsymbol{\Psi}^\intercal\boldsymbol{\Psi}\hat{\mathbf{x}}'_\mu + \mu\nabla^\intercal\nabla\hat{\mathbf{x}}'_\mu + \nabla^\intercal\nabla\hat{\mathbf{x}}_\mu + \alpha\left( \mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{x}}'_\mu + [\phi'((\hat{\mathbf{x}}_\mu)_i)(\hat{\mathbf{x}}'_\mu)_i]_{\{i=1,\dots,L^2\}} \right). \end{aligned} \tag{B.13}$$

We define the matrix $\mathbf{I}_{\hat{\mathbf{x}}_\mu}$ such as:

$$\mathbf{I}_{\hat{\mathbf{x}}_\mu}(i,i) = \begin{cases} 0 & \text{if } (\hat{\mathbf{x}}_\mu)_i \geq 0, \\ 1 & \text{if } (\hat{\mathbf{x}}_\mu)_i < 0. \end{cases}$$

Now the vector $[\phi'((\hat{\mathbf{x}}_\mu)_i)(\hat{\mathbf{x}}'_\mu)_i]_{\{i=1,\dots,L^2\}}$, using further the equation (A.3), can be simply written as: $\mathbf{I}_{\hat{\mathbf{x}}_\mu}\hat{\mathbf{x}}'_\mu$ and then (B.13) becomes:

$$\mathbf{0} \in \boldsymbol{\Psi}^\intercal\boldsymbol{\Psi}\hat{\mathbf{x}}'_\mu + \mu\nabla^\intercal\nabla\hat{\mathbf{x}}'_\mu + \nabla^\intercal\nabla\hat{\mathbf{x}}_\mu + \alpha\left( \mathbf{I}_{\boldsymbol{\Omega}}\hat{\mathbf{x}}'_\mu + \mathbf{I}_{\hat{\mathbf{x}}_\mu}\hat{\mathbf{x}}'_\mu \right).$$

The minimization problem we should solve in order to find $\hat{\mathbf{x}}'_\mu$ thus is:

$$\hat{\mathbf{x}}'_\mu = \arg\min_{\mathbf{x}\in\mathbb{R}^{L^2}} \frac{1}{2}\|\boldsymbol{\Psi}\mathbf{x}\|_2^2 + \frac{\mu}{2}\|\nabla\mathbf{x} + \frac{1}{\mu}\nabla\hat{\mathbf{x}}_\mu\|_2^2 + \frac{\alpha}{2}\left( \|\mathbf{I}_{\boldsymbol{\Omega}}\mathbf{x}\|_2^2 + \|\mathbf{I}_{\hat{\mathbf{x}}_\mu}\mathbf{x}\|_2^2 \right). \tag{B.14}$$

# C Computing gradients for a Poisson random variable

The gradient of the expected value of a Poisson random variable with respect to its parameter(s) can be computed directly. Let $\mathbf{z} \in \mathbb{R}^m_{>0}$ and $\mathbf{q} \sim \text{Poisson}(\mathbf{z})$. For every component $i = 1, \dots, m$ $q_i$ is then a discrete random variable with univariate Poisson density given by $p(q_i = k|z_i) : k \mapsto \frac{1}{k!}e^{-z_i}z_i^k$, for $k \in \mathbb{N}$. Let now $f : \mathbb{R}^m_{>0} \to \mathbb{R}$ be a continuous and bounded function and $v_k$ be the function defined by $v_k : t \mapsto p(q_i = k|t) = \frac{1}{k!}e^{-t}\ t^k$. Clearly, $v_k$ is differentiable on $\mathbb{R}_{>0}$ for all $k \in \mathbb{N}$. We distinguish two cases:

— For $k \geq 1$ there holds:

$$v_k'(t) = \frac{-e^{-t}t^k}{k!} + \frac{e^{-t}kt^{k-1}}{k!} = v_{k-1}(t) - v_k(t) \tag{C.15}$$

— For $k = 0$ we have:

$$v_0'(t) = -e^{-t} = -v_0(t). \tag{C.16}$$

We would like to apply the dominated convergence theorem to the sequence $(v_k)_k$. For that, for every $i = 1, \dots, m$, let now be $a_i \in \mathbb{R}$ such that $a_i > \max(z_i, 1)$. For $x \in [0, a_i]$ we notice that the following properties hold:

— Since the sequence $(v_k(x))_k$ comes from a Poisson density, it has finite sum and for all $k \in \mathbb{N}$, $v_k$ is differentiable on $[0, a_i]$.

— For $k \geq 1$ there holds $v_k'(x) = \frac{1}{k!}e^{-x}x^{k-1}(k - x) < \frac{a_i^{k-1}}{(k-1)!}$. The sequence $(v_k'(x))_k$ is thus dominated by a summable sequence.

As a consequence of the dominated convergence theorem it is possible to switch the derivation and integral to obtain:

$$\frac{\partial}{\partial z_i}\mathbb{E}[f(\mathbf{q})] = \frac{\partial}{\partial z_i}\sum_{k=0}^{\infty}\mathbb{E}[f(\mathbf{q})|q_i = k]p(q_i = k|z_i) = \sum_{k=0}^{\infty}\mathbb{E}[f(\mathbf{q})|q_i = k]\frac{\partial}{\partial z_i}p(q_i = k|z_i)$$

We can now use the recursion formulas (C.15)-(C.16) for $v_k'$ as follows:

$$\begin{aligned}
\frac{\partial}{\partial z_i}\mathbb{E}[f(\mathbf{q})] &= \sum_{k=0}^{\infty}\mathbb{E}[f(\mathbf{q})|q_i = k]v_k'(z_i) \\
&= \mathbb{E}[f(\mathbf{q})|q_i = 0]v_0'(z_i) + \sum_{k=1}^{\infty}\mathbb{E}[f(\mathbf{q})|q_i = k](v_{k-1}(z_i) - v_k(z_i)) \\
&= -\mathbb{E}[f(\mathbf{q})|q_i = 0]v_0(z_i) - \sum_{k=1}^{\infty}\mathbb{E}[f(\mathbf{q})|q_i = k]v_k(z_i) \\
&+ \sum_{k=0}^{\infty}\mathbb{E}[f(\mathbf{q})|q_i = k+1]v_k(z_i)
\end{aligned}$$

Then a variable change is applied.

Denoting by $\mathbf{1_i} = \{\delta_{i,j}\}_{j=1}^n \in \mathbb{R}^n$ with $\delta_{i,j}$ being the Dirac delta function the vector of only zeros except a one in the $i$-th position, we can now write:

$$\frac{\partial}{\partial z_i}\mathbb{E}[f(\mathbf{q})] = -\sum_{k=0}^{\infty}\mathbb{E}[f(\mathbf{q})|q_i = k]p(k|z_i) + \sum_{k=0}^{\infty}\mathbb{E}[f(\mathbf{q} + \mathbf{1_i})|q_i = k]p(q_i = k|z_i)$$

Finally, the $i$-th component of the gradient of expected value of $f(\mathbf{q})$ can be simply written as a finite difference :

$$\frac{\partial}{\partial z_i}\mathbb{E}[f(\mathbf{q})] = \mathbb{E}[f(\mathbf{q}+\mathbf{1_i}) - f(\mathbf{q})] \tag{C.17}$$

This expression is easily interpreted: incremental increments of $z_i$ correspond to increments of 1 of the component $q_i$ in expectation. This is in fact a consequence of the definition of Poisson law, whose parameter equals its expected value.

Unlike other methods such as the one of score functions in [Mohamed et al., 2020], formula (C.17) can deal with the case $z_i = 0$. However, its application is computationally demanding: computing the expectancy in (C.17) by empirical means of $B$ realisations, requires in fact $B(m+1)$ evaluations of $f$. To reduce computations, and by assuming that $f$ is differentiable, we can make however the following approximation:

$$f(\mathbf{q}+\mathbf{1_i}) - f(\mathbf{q}) \simeq \frac{\partial f}{\partial q_i}(\mathbf{q}) \tag{C.18}$$

which can be plugged in (C.17) and estimated empirically by means of $B$ samples, thus finally getting:

$$\nabla_z \mathbb{E}[f(\mathbf{q})] \simeq \mathbb{E}[\nabla_q f(\mathbf{q})] \simeq \sum_{t=1}^{B} \nabla_q f(\mathbf{q}^t),$$

which provides a handy way of approximating the desired quantity.

**Remark :** The approximation (C.18) is valid only when $\frac{\partial f}{\partial q_i}$ does not change too much between $f(\mathbf{q})$ and $f(\mathbf{q}+\mathbf{1_i})$. If all the second derivatives of $f$ are bounded, then it is a sufficient condition to bound the second-order error of this approximation. However, in general we do not know whether $f$ is twice differentiable. In our context, this condition can thus be relaxed by assuming that we have a sufficiently fine quantisation compared to the variations of $\partial f/\partial q_i$.

# Bibliography


[Abbe, 1873] Abbe, E. (1873). Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für Mikroskopische Anatomie*, 9(1):413–468.

[Ambrosio et al., 2005] Ambrosio, L., Gigli, N., and Savare, G. (2005). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhäuser Basel.

[Arjovsky et al., 2017] Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR.

[Attouch et al., 2010] Attouch, H., Bolte, J., Redont, P., and Soubeyran, A. (2010). Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457.

[Axelrod, 1981] Axelrod, D. (1981). Cell-substrate contacts illuminated by total internal reflection fluorescence. *The Journal of cell biology*, 89(1):141—145.

[Axelrod, 2008] Axelrod, D. (2008). Chapter 7 total internal reflection fluorescence microscopy. In *Biophysical Tools for Biologists, Volume Two: In Vivo Techniques*, volume 89 of *Methods in Cell Biology*, pages 169–221. Academic Press.

[Beck and Teboulle, 2009] Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.

[Bect et al., 2004] Bect, J., Blanc-Féraud, L., Aubert, G., and Chambolle, A. (2004). A l1-unified variational framework for image restoration. In Pajdla, T. and Matas, J., editors, *Computer Vision - ECCV 2004*, pages 1–13, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Bertero and Boccacci, 1998] Bertero, M. and Boccacci, P. (1998). *Introduction to Inverse Problems in Imaging*. CRC Press.

[Betzig et al., 2006] Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., and Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645.

[Bond-Taylor et al., 2022] Bond-Taylor, S., Leach, A., Long, Y., and Willcocks, C. G. (2022). Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7327–7347.

[Bora et al., 2017] Bora, A., Jalal, A., Price, E., and Dimakis, A. G. (2017). Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 537–546. JMLR.org.

[Bora et al., 2018] Bora, A., Price, E., and Dimakis, A. G. (2018). AmbientGAN: Generative models from lossy measurements. In *International Conference on Learning Representations*.

[Born and Wolf, 1999] Born, M. and Wolf, E. (1999). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light (7th Edition)*. Cambridge University Press.

[Boulanger et al., 2014] Boulanger, J., Gueudry, C., Münch, D., Cinquin, B., Paul-Gilloteaux, P., Bardin, S., Guérin, C., Senger, F., Blanchoin, L., and Salamero, J. (2014). Fast high-resolution 3d total internal reflection fluorescence microscopy by incidence angle scanning and azimuthal averaging. *Proceedings of the National Academy of Sciences of the United States of America*, 111(48):17164—17169.

[Boyd et al., 2011] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122.

[Cachia et al., 2023] Cachia, M., Stergiopoulou, V., Calatroni, L., Schaub, S., and Blanc-Féraud, L. (2023). Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN). *Inverse Problems*, 39(5):054006.

[Candès et al., 2007] Candès, E., Wakin, M., and Boyd, S. (2007). Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905.

[Chambolle et al., 1998] Chambolle, A., DeVore, R. A., Lee, N., and Lucier, B. J. (1998). Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 7 3:319–35.

[Chambolle and Pock, 2011] Chambolle, A. and Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40.

[Cohen et al., 2021] Cohen, R., Blau, Y., Freedman, D., and Rivlin, E. (2021). It has potential: Gradient-driven denoisers for convergent solutions to inverse problems. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 18152–18164. Curran Associates, Inc.

[Combettes and Wajs, 2005] Combettes, P. L. and Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4:1168–1200.

[Condat, 2013] Condat, L. (2013). A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158.

[Cox et al., 2011] Cox, S., Rosten, E., Monypenny, J., Jovanovic-Talisman, T., Burnette, D., Lippincott-Schwartz, J., Jones, G., and Heintzmann, R. (2011). Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature methods*, 9:195–200.

[Dabov et al., 2007] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095.

[Dardikman-Yoffe and Eldar, 2020] Dardikman-Yoffe, G. and Eldar, Y. C. (2020). Learned sparcom: unfolded deep super-resolution microscopy. *Opt. Express*, 28(19):27736–27763.

[Daubechies et al., 2003] Daubechies, I., Defrise, M., and Mol, C. D. (2003). An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57.

[de Morais Goulart et al., 2019] de Morais Goulart, J. H., Blanc-Féraud, L., Debreuve, E., and Schaub, S. (2019). A study on tensor and matrix models for super-resolution fluorescence microscopy. In *CAMSAP 2019 - IEEE International Workshop on Computational Advances in*

*Multi-Sensor Adaptive Processing*, Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Le Gosier, Guadeloupe.

[Deng et al., 2014] Deng, Y., Sun, M., Lin, P., Ma, J., and Shaevitz, J. (2014). Spatial covariance reconstructive (score) super-resolution fluorescence microscopy. *PloS one*, 9:e94807.

[Denk et al., 1990] Denk, W., Strickler, J. H., and Webb, W. W. (1990). Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76.

[Dertinger et al., 2009] Dertinger, T., Colyer, R., Iyer, G., Weiss, S., and Enderlein, J. (2009). Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proceedings of the National Academy of Sciences, 106 (52)*, pages 22287–22292.

[Dertinger et al., 2010] Dertinger, T., Colyer, R., Vogel, R., Enderlein, J., and Weiss, S. (2010). Achieving increased resolution and more pixels with superresolution optical fluctuation imaging (sofi). *Opt. Express*, 18(18):18875–18885.

[Dos Santos et al., 2014] Dos Santos, M. C., Déturche, R., Vézy, C., and Jaffiol, R. (2014). Axial nanoscale localization by normalized total internal reflection fluorescence microscopy. *Optics letters*, 39(4):869–872.

[Douglas and Rachford, 1956] Douglas, J. and Rachford, H. H. (1956). On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439.

[Duff et al., 2021] Duff, M., Campbell, N. D. F., and Ehrhardt, M. J. (2021). Regularising inverse problems with generative machine learning models.

[Engl et al., 2000] Engl, H., Hanke, M., and Neubauer, A. (2000). *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands.

[Gale and Shapley, 1962] Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

[Gazagnes et al., 2017] Gazagnes, S., Soubies, E., and Blanc-Féraud, L. (2017). High density molecule localization for super-resolution microscopy using CEL0 based sparse approximation. In *ISBI 2017*, pages 28–31.

[Geissbuehler et al., 2012] Geissbuehler, S., Bocchio, N., Dellagiacoma, C., Berclaz, C., Leutenegger, M., and Lasser, T. (2012). Mapping molecular statistics with balanced super-resolution optical fluctuation imaging (bsofi). *Optical Nanoscopy*, 1.

[Geman and Yang, 1995] Geman, D. and Yang, C. (1995). Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7):932–946.

[Gfrerer, 1987] Gfrerer, H. (1987). An a posteriori parameter choice for ordinary and iterated tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Mathematics of Computation*, 49(180):507–522.

[Girsault et al., 2016] Girsault, A., Lukes, T., Sharipov, A., Geissbuehler, S., Leutenegger, M., Vandenberg, W., Dedecker, P., Hofkens, J., and Lasser, T. (2016). SOFI simulation tool: A software package for simulating and testing super-resolution optical fluctuation imaging. *PLOS ONE, 11 (9)*, (9):1–13.

[Goh et al., 2022] Goh, H., Sheriffdeen, S., Wittmer, J., and Bui-Thanh, T. (2022). Solving bayesian inverse problems via variational autoencoders. In Bruna, J., Hesthaven, J., and Zdeborova, L., editors, *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*, volume 145 of *Proceedings of Machine Learning Research*, pages 386–425. PMLR.

[Goncharova et al., 2020] Goncharova, A. S., Honigmann, A., Jug, F., and Krull, A. (2020). Improving blind spot denoising for microscopy. In Bartoli, A. and Fusiello, A., editors, *Computer Vision – ECCV 2020 Workshops*, pages 380–393, Cham. Springer International Publishing.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

[Gräf et al., 2005] Gräf, R., Rietdorf, J., and Zimmermann, T. (2005). *Live Cell Spinning Disk Microscopy*, pages 57–75. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Gribonval and Nikolova, 2020] Gribonval, R. and Nikolova, M. (2020). A characterization of proximity operators. *Journal of Mathematical Imaging and Vision*, 62.

[Griffa et al., 2010] Griffa, A., Garin, N., and Sage, D. (2010). Comparison of deconvolution software in 3D microscopy: A user point of view—Part 1. *G.I.T. Imaging & Microscopy*, 12(1):43–45.

[Gulrajani et al., 2017] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of Wasserstein GANs. *CoRR*, abs/1704.00028.

[Gupta et al., 2021] Gupta, H., McCann, M. T., Donati, L., and Unser, M. (2021). CryoGAN: A New Reconstruction Paradigm for Single-Particle Cryo-EM Via Deep Adversarial Learning. *IEEE Transactions on Computational Imaging*, 7.

[Gustafsson et al., 2008] Gustafsson, M., L.Shao, Carlton, P., Wang, C., Golubovskaya, I., Cande, W. Z., Agard, D., and Sedat, J. (2008). Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophysical Journal*, 94(12):4957–4970.

[Gustafsson, 2000] Gustafsson, M. G. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy, 198 (2)*.

[Gustafsson et al., 2016] Gustafsson, N., Culley, S., Ashdown, G., Owen, D. M., Pereira, P. M., and Henriques, R. (2016). Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature communications, 7(1)*, pages 12471–12471.

[Hadamard, 1902] Hadamard, J. (1902). Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52.

[Hansen, 2010] Hansen, P. C. (2010). *Discrete Inverse Problems: Insight and Algorithms*. Society for Industrial and Applied Mathematics, USA.

[Harke et al., 2008] Harke, B., Ullal, C., Keller, J., and Hell, S. (2008). Three-dimensional nanoscopy of colloidal crystals. *Nano. Lett*, pages 1309–1313.

[Hell and Wichmann, 1994] Hell, S. W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett., 19 (11)*, pages 780–782.

[Herman, 1998] Herman, B. (1998). *Fluorescence Microscopy*. Microscopy handbooks. BIOS Scientific Publ.

[Hestenes and Stiefel, 1952] Hestenes, M. R. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49:409–435.

[Huang et al., 2008] Huang, B., Wang, W., Bates, M., and Zhuang, X. (2008). Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 319(5864):810–813.

[Hurault et al., 2022a] Hurault, S., Leclaire, A., and Papadakis, N. (2022a). Gradient step denoiser for convergent plug-and-play. In *International Conference on Learning Representations*.

[Hurault et al., 2022b] Hurault, S., Leclaire, A., and Papadakis, N. (2022b). Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9483–9505. PMLR.

[Idier, 2008] Idier, J. (2008). *Bayesian Approach to Inverse Problems*. Digital signal and image processing series. ISTE.

[Idier et al., 2018] Idier, J., Labouesse, S., Allain, M., Liu, P., Bourguignon, S., and Sentenac, A. (2018). On the superresolution capacity of imagers using unknown speckle illuminations. *IEEE Transactions on Computational Imaging*, 4(1):87–98.

[Kamilov et al., 2022] Kamilov, U. S., Bouman, C. A., Buzzard, G. T., and Wohlberg, B. (2022). Plug-and-play methods for integrating physical and learned models in computational imaging. *arXiv preprint arXiv:2203.17061*.

[Kingma and Welling, 2014] Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. In Bengio, Y. and LeCun, Y., editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

[Kirshner et al., 2013] Kirshner, H., Aguet, F., Sage, D., and Unser, M. (2013). 3-d psf fitting for fluorescence microscopy: implementation and localization application. *Journal of Microscopy*, 249(1):13–25.

[Koulouri et al., 2021] Koulouri, A., Heins, P., and Burger, M. (2021). Adaptive superresolution in deconvolution of sparse peaks. *IEEE Transactions on Signal Processing*, 69:165–178.

[Laine et al., 2019] Laine, R. F., Tosheva, K. L., Gustafsson, N., Gray, R. D. M., Almada, P., Albrecht, D., Risa, G. T., Hurtig, F., Lindås, A.-C., Baum, B., Mercer, J., Leterrier, C., Pereira, P. M., Culley, S., and Henriques, R. (2019). Nanoj: a high-performance open-source super-resolution microscopy toolbox. *Journal of Physics D: Applied Physics*, 52(16):163001.

[Laville et al., 2021] Laville, B., Blanc-Féraud, L., and Aubert, G. (2021). Off-the-grid variational sparse spike recovery: Methods and algorithms. *Journal of Imaging*, 7:266.

[Laville et al., 2022] Laville, B., Blanc-Féraud, L., and Aubert, G. (2022). Off-the-grid covariance-based super-resolution fluctuation microscopy. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2315–2319.

[Lefkimmiatis et al., 2013] Lefkimmiatis, S., Ward, J. P., and Unser, M. (2013). Hessian schatten-norm regularization for linear inverse problems. *IEEE Transactions on Image Processing*, 22(5):1873–1888.

[Li and Vaughan, 2018] Li, H. and Vaughan, J. (2018). Switchable fluorophores for single-molecule localization microscopy. *Chemical Reviews*, 118.

[Lord Rayleigh, 1879] Lord Rayleigh, F. (1879). Xxxi. investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(49):261–274.

[Martin-Fernandez et al., 2013] Martin-Fernandez, M., Tynan, C., and Webb, S. (2013). A 'pocket guide' to total internal reflection fluorescence. *Journal of Microscopy*, 252(1):16–22.

[Minsky, 1961] Minsky, M. (1961). Microscopy apparatus.

[Mohamed et al., 2020] Mohamed, S., Rosca, M., Figurnov, M., and Mnih, A. (2020). Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62.

[Monga et al., 2021] Monga, V., Li, Y., and Eldar, Y. C. (2021). Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44.

[Mudry et al., 2012] Mudry, E., Belkebir, K., Girard, J., Savatier, J., Le Moal, E., Nicoletti, C., Allain, M., and Sentenac, A. (2012). Structured illumination microscopy using unknown speckle patterns. *Nature Photonics*, 6(5):312–315.

[Nesterov, 1983] Nesterov, Y. (1983). A method for solving the convex programming problem with convergence rate o($\frac{1}{k^2}$). *Proceedings of the USSR Academy of Sciences*, 269:543–547.

[Ochs et al., 2015] Ochs, P., Dosovitskiy, A., Brox, T., and Pock, T. (2015). On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision. *SIAM Journal on Imaging Sciences*, 8(1):331–372.

[Olarte et al., 2018] Olarte, O. E., Andilla, J., Gualda, E. J., and Loza-Alvarez, P. (2018). Light-sheet microscopy: a tutorial. *Adv. Opt. Photon.*, 10(1):111–179.

[Parikh and Boyd, 2014] Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Found. Trends Optim.*, 1(3):127–239.

[Prakash et al., 2021] Prakash, M., Krull, A., and Jug, F. (2021). Fully unsupervised diversity denoising with convolutional variational autoencoders. In *International Conference on Learning Representations*.

[Reehorst and Schniter, 2019] Reehorst, E. T. and Schniter, P. (2019). Regularization by denoising: Clarifications and new interpretations. *IEEE Transactions on Computational Imaging*, 5(1):52–67.

[Romano et al., 2017] Romano, Y., Elad, M., and Milanfar, P. (2017). The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844.

[Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268.

[Rust et al., 2006] Rust, M., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3:793–5.

[Ryu et al., 2019] Ryu, E., Liu, J., Wang, S., Chen, X., Wang, Z., and Yin, W. (2019). Plug-and-play methods provably converge with properly trained denoisers. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5546–5557, Long Beach, California, USA. PMLR.

[Sage et al., 2015] Sage, D., Kirshner, H., Pengo, T., Stuurman, N., Min, J., Manley, S., and Unser, M. (2015). Quantitative evaluation of software packages for single-molecule localization microscopy. *Nature methods, 12*.

[Sage et al., 2019] Sage, D., Pham, T.-A., Babcock, H., Lukes, T., Pengo, T., Chao, J., Velmurugan, R., Herbert, A., Agrawal, A., Colabrese, S., Wheeler, A., Archetti, A., Rieger, B., Ober, R., Hagen, G., Sibarita, J.-B., Ries, J., Henriques, R., Unser, M., and Holden, S. (2019). Super-resolution fight club: Assessment of 2d & 3d single-molecule localization microscopy software. *Nature Methods*, 16.

[Sanderson et al., 2014] Sanderson, M. J., Smith, I., Parker, I., and Bootman, M. D. (2014). Fluorescence microscopy. *Cold Spring Harbor Protocols*, 2014(10):pdb–top071795.

[Shah and Hegde, 2018] Shah, V. and Hegde, C. (2018). Solving linear inverse problems using gan priors: An algorithm with provable guarantees. In *2018 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2018 - Proceedings*, ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, pages 4609–4613. Institute of Electrical and Electronics Engineers Inc. Funding Information: This work was supported in part by grants from the National Science Foundation (NSF CCF-1566281) and NVIDIA. Publisher Copyright: © 2018 IEEE.; 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2018 ; Conference date: 15-04-2018 Through 20-04-2018.

[Shtengel et al., 2009] Shtengel, G., Galbraith, J., Galbraith, C., Lippincott-Schwartz, J., Gillette, J., Manley, S., Sougrat, R., Waterman, C., Kanchanawong, P., Davidson, M., Fetter, R., and Hess, H. (2009). Interferometric fluorescent super-resolution microscopy resolves 3d cellular ultrastructure. *Proceedings of the National Academy of Sciences*, 106(9):3125–3130.

[Solomon et al., 2019] Solomon, O., Eldar, Y. C., Mutzafi, M., and Segev, M. (2019). SPARCOM: Sparsity based super-resolution correlation microscopy. *SIAM Journal on Imaging Sciences, 12 (1)*, pages 392–419.

[Solomon et al., 2018] Solomon, O., Mutzafi, M., Segev, M., and Eldar, Y. C. (2018). Sparsity-based super-resolution microscopy from correlation information. *Opt. Express*, 26(14):18238–18269.

[Soubies, 2016] Soubies, E. (2016). *Sur quelques problèmes de reconstruction en imagerie MA-TIRF et en optimisation parcimonieuse par relaxation continue exacte de critères pénalisés en norme-l0*. Theses, Université Côte d'Azur.

[Soubies et al., 2019a] Soubies, E., Blanc-Féraud, L., Schaub, S., and Van Obberghen-Schilling, E. (2019a). Improving 3D MA-TIRF Reconstruction with Deconvolution and Background Estimation. In *IEEE International Symposium on Biomedical Imaging*, Venise, Italy.

[Soubies et al., 2015] Soubies, E., Blanc-Féraud, L., and Aubert, G. (2015). A continuous exact $\ell_0$ penalty (CEL0) for least squares regularized problem. *SIAM Journal on Imaging Sciences, 8 (3)*, pages 1607–1639.

[Soubies et al., 2019b] Soubies, E., Radwanska, A., Grall, D., Blanc-Féraud, L., Van Obberghen-Schilling, E., and Schaub, S. (2019b). Nanometric axial resolution of fibronectin assembly units achieved with an efficient reconstruction approach for multi-angle-tirf microscopy. *Scientific Reports*, 9(1).

[Stergiopoulou et al., 2022a] Stergiopoulou, V., Calatroni, L., de Morais Goulart, H., Schaub, S., and Blanc-Féraud, L. (2022a). COL0RME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation. *Biological Imaging*, 2:e1.

[Stergiopoulou et al., 2022b] Stergiopoulou, V., Calatroni, L., Schaub, S., and Blanc-Féraud, L. (2022b). 3D Image Super-Resolution by Fluorophore Fluctuations and MA-TIRF Microscopy Reconstruction (3D-COL0RME). In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4.

[Stergiopoulou et al., 2021] Stergiopoulou, V., de M. Goulart, J. H., Schaub, S., Calatroni, L., and Blanc-Féraud, L. (2021). COL0RME: Covariance-based $\ell_0$ super-resolution microscopy with intensity estimation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 349–352.

[Stergiopoulou et al., 2023] Stergiopoulou, V., Mukherjee, S., Calatroni, L., and Blanc-Féraud, L. (2023). Fluctuation-based deconvolution in fluorescence microscopy using plug-and-play denoisers. working paper or preprint.

[Terris et al., 2020] Terris, M., Repetti, A., Pesquet, J.-C., and Wiaux, Y. (2020). Building firmly nonexpansive convolutional neural networks. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8658–8662.

[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

[Tikhonov, 1963] Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*

[Vahdat and Kautz, 2020] Vahdat, A. and Kautz, J. (2020). Nvae: A deep hierarchical variational autoencoder. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA. Curran Associates Inc.

[Velasquez, 2021] Velasquez, D. (2021). *Cellular and molecular characterization of Ostreopsis cf. ovata cell cycle*. Theses, Sorbonne Université.

[Venkatakrishnan et al., 2013] Venkatakrishnan, S. V., Bouman, C. A., and Wohlberg, B. (2013). Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948.

[Xu et al., 2020] Xu, X., Liu, J., Sun, Y., Wohlberg, B., and Kamilov, U. S. (2020). Boosting the performance of plug-and-play priors via denoiser scaling. In *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pages 1305–1312. IEEE.

[Yahiatene et al., 2015] Yahiatene, I., Hennig, S., Müller, M., and Huser, T. (2015). Entropy-based super-resolution imaging(esi): From disorder to fine detail. *ACS Photonics*, 2.

[Zehni and Zhao, 2021] Zehni, M. and Zhao, Z. (2021). MSR-GaN: Multi-segment reconstruction via adversarial learning. *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*.

[Zhang et al., 2021] Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., and Timofte, R. (2021). Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Zhang et al., 2017] Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155.

# Publications & Talks

## Publications

### Journals

M. Cachia, **V. Stergiopoulou**, L. Calatroni, S. Schaub, L. Blanc-Féraud. "Fluorescence image deconvolution microscopy via generative adversarial learning (FluoGAN)," to appear in Inverse Problems, 2022.

**V. Stergiopoulou**, L. Calatroni, J. H. de Morais Goulart, S. Schaub, and L. Blanc-Féraud, "COL0RME: Super-resolution microscopy based on sparse blinking/fluctuating fluorophore localization and intensity estimation," Biological Imaging, vol. 2, p. e1, 2022, doi: 10.1017/S2633903X22000010.

### Conference Proceedings

**V. Stergiopoulou**, L. Calatroni, S. Schaub and L. Blanc-Féraud, "3D Image Super-Resolution by Fluorophore Fluctuations and MA-TIRF Microscopy Reconstruction (3D-COL0RME)," 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1-4, doi: 10.1109/ISBI52829.2022.9761572. **(Best-Paper Runner-Up Award)**

**V. Stergiopoulou**, J. H. de Morais Goulart, S. Schaub, L. Calatroni and L. Blanc-Féraud, "COL0RME: Covariance-Based l0 Super-Resolution Microscopy with Intensity Estimation," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 349-352, doi: 10.1109/ISBI48211.2021.9433976.

## Talks

| | |
|---|---|
| Oct. 2020 | Invited Oral Presentation at GdR MIA Thematic day on Non-Convex Sparse Optimization (ENSEEIHT, Toulouse, France) link |
| Oct. 2021 | Oral Presentation at 3IA PhD/Postdoc Seminar (Sophia Antipolis, France) |
| March 2022 | Oral Presentation at Non Conventional Optical Imaging Days - 17th edition (Institut Langevin, Paris, France) link |
| March 2022 | Invited Oral Presentation at 2022 SIAM Conference on Imaging Science (Virtual Conference) link |
| March 2022 | Oral Presentation at IEEE 19th International Symposium on Biomedical Imaging (Kolkata, India) link |
| June 2022 | Invited Oral Presentation at Cambridge Advanced Imaging Centre (Cambridge, UK) |
| June 2022 | Invited Oral Presentation at MRC Laboratory of Molecular Biology (Cambridge, UK) |
| Sep. 2022 | Oral Presentation at 1st French-Italian workshop on the Mathematics of Imaging, Vision and their Applications (Sophia Antipolis, France) link |
| Dec. 2022 | Oral Presentation at Mathematical Models for Plug-and-play Image Restoration (Paris, France) link |