HAL
open science

A Novel Variational Approach for Multiphoton Microscopy Image Restoration: from PSF Estimation to 3D Deconvolution

Julien Ajdenbaum, Emilie Chouzenoux, Claire Lefort, Ségolène Martin, Jean-Christophe Pesquet

To cite this version:

Julien Ajdenbaum, Emilie Chouzenoux, Claire Lefort, Ségolène Martin, Jean-Christophe Pesquet. A Novel Variational Approach for Multiphoton Microscopy Image Restoration: from PSF Estimation to 3D Deconvolution. 2023. hal-04296247

HAL Id: hal-04296247

https://hal.science/hal-04296247

Preprint submitted on 24 Nov 2023

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

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

A Novel Variational Approach for Multiphoton Microscopy Image Restoration: from PSF Estimation to 3D Deconvolution

Julien Ajdenbaum[1], **Emilie Chouzenoux**[1], **Claire Lefort**[2], **Ségolène Martin**[1*], **Jean-Christophe Pesquet**[1]

[1] Centre de Vision Numérique, Université Paris-Saclay, Inria, CentraleSupélec, Gif-sur-Yvette, France

[2] XLIM Research Institute, UMR CNRS 7252, Université de Limoges, Limoges, France

[*] Corresponding author

Abstract. In multi-photon microscopy (MPM), a recent *in-vivo* fluorescence microscopy system, the task of image restoration can be decomposed into two interlinked inverse problems: firstly, the characterization of the Point Spread Function (PSF) and subsequently, the deconvolution (i.e., deblurring) to remove the PSF effect, and reduce noise.

The acquired MPM image quality is critically affected by PSF blurring and intense noise. The PSF in MPM is highly spread in 3D and is not well characterized, presenting high variability with respect to the observed objects. This makes the restoration of MPM images challenging. Common PSF estimation methods in fluorescence microscopy, including MPM, involve capturing images of sub-resolution beads, followed by quantifying the resulting ellipsoidal 3D spot. In this work, we revisit this approach, coping with its inherent limitations in terms of accuracy and practicality. We estimate the PSF from the observation of relatively large beads (approximately 1μm in diameter). This goes through the formulation and resolution of an original non-convex minimization problem, for which we propose a proximal alternating method along with convergence guarantees.

Following the PSF estimation step, we then introduce an innovative strategy to deal with the high level multiplicative noise degrading the acquisitions. We rely on a heteroscedastic noise model for which we estimate the parameters. We then solve a constrained optimization problem to restore the image, accounting for the estimated PSF and noise, while allowing a minimal hyper-parameter tuning. Theoretical guarantees are given for the restoration algorithm.

These algorithmic contributions lead to an end-to-end pipeline for 3D image restoration in MPM, that we share as a publicly available Python software. We demonstrate its effectiveness through several experiments on both simulated and real data.

1. Introduction

Multi-photon microscopy (MPM) is a non-invasive laser imaging technique that selectively induces fluorescence in a thin plane through localized nonlinear excitation while avoiding excitation elsewhere [32]. This technique enables three-dimensional imaging with infrared light, reaching depths that are at least double compared to traditional single-photon microscopy [40, 11, 51]. Particularly, MPM contactless and non-invasive nature makes it suitable for imaging diverse objects, spanning from living organisms [77] to materials [46].

Though MPM offers significant advantages, its widespread adoption remains limited due to degradations affecting the images, especially blur and noise, requiring innovative computational solutions to unlock its full potential. Central to MPM imaging is the Point Spread Function (PSF). This function describes how the imaging system responds to a a Dirac impulse, in practice a single-point light object. In practical terms, this response appears as an extended spot in the image around the original object location, which is responsable for distortions and blurry aspects [72, 73, 29], particularly in the depth direction. As MPM typically operates as a non-coherent and space-invariant system, it is commonplace to model the captured observation as a convolution of the actual object with a blur kernel. In the following, we will use the terms *PSF* and *blur kernel* interchangeably as they refer to the same mathematical object. A major challenge with the PSF in MPM is its strong dependency on the medium [35], implying that it should be individually estimated for each new acquisition.

In MPM, the restoration process aims to recover the sought (non-degraded) object of interest, $\bar{x}\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$, jointly with an estimation of the Point Spread Function (PSF), $\bar{h}\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$, guided by the following linear observational model, at each location $\boldsymbol{u}$ of the 3D volume:

$$(\forall \boldsymbol{u} \in \mathbb{R}^3) \quad y(\boldsymbol{u}) = \mathcal{D}(\bar{h} \star \bar{x} + \alpha)(\boldsymbol{u}). \tag{1}$$

In (1), $y\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$ stands for the observation, $\alpha \in \mathbb{R}$ represents a background factor, and $\star$ the continuous 3D convolution operation, here to be assumed to be defined. Additionally, $\mathcal{D}$ refers to the noise model. In the computational imaging literature, there are two main ways of addressing the above inverse problem. Some methods directly tackle the *blind* inverse problem represented by (1), i.e. they simultaneously estimate both $\bar{x}$ and $\bar{h}$ [69, 59, 45, 27]. However, this approach can be computationally demanding due to its non-convex nature and potential underdetermined nature. Another strategy is to break down the problem into two successive inverse problems. In this two-step approach, one first performs a calibration step [48, 43, 36, 44, 60], providing an estimate of the PSF $\bar{h}$ in a controlled situation where $\bar{x}$ is a known (i.e., calibrated) entity. Then, utilizing the estimated PSF, one can infer an estimate of another, more complex, unknown object $\bar{x}$.

PSF calibration in MPM typically involves imaging sub-resolution fluorescent beads in a homogeneous medium [20, 34] or directly in the sample (*in situ*) [52, 34, 68]. The image of such a bead captured with the MPM device appears as a spread spot, which

is then fit with a shape function (e.g., Gaussian) to estimate the PSF characteristics [68, 75, 41, 17]. However, this method presents significant practical limitations. Firstly, handling beads as small as 0.2μm in diameter requires special care, as sub-micrometric objects are difficult to even observe with the microscope. As a result, one often injects hundreds of beads into the sample, often leading to beads aggregation that compromises the accurate completion of the PSF calibration protocol. Secondly, observing the impulse response of the instrument on such tiny beads requires the use of high laser power, which can potentially damage the sample itself. These limitations highlight the need for an alternative and a more practical protocol for PSF estimation in MPM.

Following PSF estimation, the subsequent task is image deblurring and denoising. Over the years, several computational solutions have been proposed to solve this inverse problem [24, 63, 61] in the context of microscopy. A common strategy is to minimize a cost function which balances a data-fitting term with a regularization term [33, 25, 30]. Yet, such an approach often suffers from a time-intensive parameter tuning phase, made even more challenging by subjective decision-making in the absence of a clear ground-truth. Further difficulties emerge from consideration on the nature of the noise in MPM. While most studies assume a standard Gaussian additive noise [34, 23], the noise in MPM images is more appropriately considered as multiplicative, as pointed out in [22, 57]. However, such a noise model can be tedious to deal with numerically at large-scale, often involving minimizing a non-smooth data-fidelity term [74, 16]. While this issue has been tackled in confocal microscopy [67, 74, 6], its exploration in MPM imaging remains scarce.

In this work, we present a novel approach for addressing the MPM image restoration inverse problem in an end-to-end fashion. Our comprehensive restoration pipeline revisits the conventional restoration protocol from acquisition to the final restored outcome. Within this pipeline, we introduce two major advancements:

(i) **PSF Calibration Step.** Our contributions regarding the estimation of the instrumental PSF are two-fold.
   - We introduce a novel optimization problem formulation for estimating the a Gaussian-shaped PSF assuming large diameter beads as opposed to sub-resolution beads.
   - Next, we address the PSF estimation inverse problem by deploying an alternating proximal approach whose convergence is proved.

(ii) **Image restoration Step.** Our contributions to the image restoration phase encompass two key elements.
   - First, we adopt an additive heteroscedastic noise model that approximates the Poisson-Gaussian noise. We then introduce a quantization-based image analysis method to estimate the noise model parameters.
   - Additionally, we address the parameter tuning issue raised in penalized restoration approaches, by introducing an original constrained formulation of the image restoration inverse problem. Our formulation introduces an

interpretable upper bound on the data fidelity term, making the restoration stage stable, reliable, and parameter-free. The problem resolution is then addressed with the local subspace Majorization-Minimization algorithm proposed in [18], for which we show the convergence in our context.

We demonstrate the effectiveness of our two-step restoration pipeline on both simulated and real-world data acquired with an Olympus XLPLN25XWMP two-photon microscope. The associated Python code is publicly shared‡, ensuring reproducible research.

The outline is as follows. Section 2 introduces our PSF estimation method, and validation experiments on synthetic data. In Section 3, we introduce our noise modeling and image restoration approaches, and we assess them on synthetic datasets. Lastly, in Section 4, we present the experimental evaluation of our complete pipeline to the imaging of real mouse muscle data.

2. A novel approach to PSF calibration using large beads

In this section, we introduce our novel approach for estimating the PSF from acquisition of calibrated fluorescent beads with a diameter of 1μm, larger than the microscope resolution.

2.1. Preparing the data

For performing the calibration stage, multiple fluorescent beads with known size, are observed under the microscope in one comprehensive 3D image. The PSF estimation aims at fitting a model on each individual bead. To do so, the initial large image is cropped, to form several volumes of interest each containing an isolated bead. The cropping process we propose consists in applying a standard Wiener filtering to reduce noise, followed by a binarization using a predefined threshold (in practice, a given pourcentage of the maximum intensity), and an automatic connected component labeling algorithm [70]. Each selected component then defines a cropped region, to which we apply the PSF estimation method. Under the assumption of stationarity of the PSF model within the observed volume, the final PSF estimate can be obtained by averaging the models fitted in each volume interest, to get a more accurate overall estimation. We now explicit our PSF estimation process, given a volume containing a single bead. An example of image of a polystyrene bead with 1 μm diameter (FluoSphere™ Carboxylate-Modified Microsphere, manufactured by Thermofisher, product number: F8823), acquired by our Olympus XLPLN25XWMP two-photon microscope is displayed in Figure 1.

‡ `https://github.com/SegoleneMartin/biphoton`

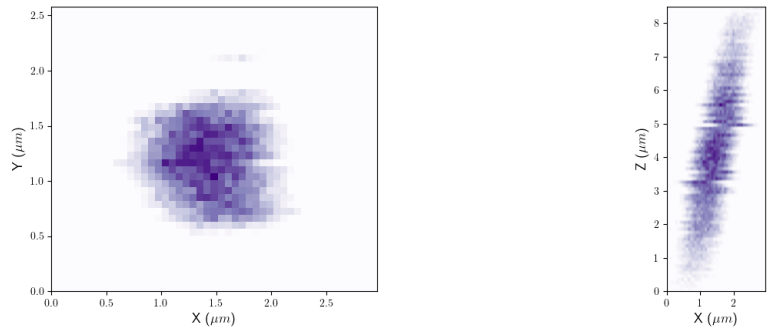

Figure 1. Two-photon acquisition of a polystyrene bead of 1μm of diameter. XY plane section (left) and XZ plane section (right). The acquisition was carried out with an excitation wavelength of 810nm. The voxel size was fixed to $0.049 \times 0.049 \times 0.1 \mu\text{m}^3$.

2.2. Modeling of the problem

Since we are in a calibration context, the true bead characteristics are known. In particular, the experimenter selects the bead diameter, and a spherical shape can be assumed. Moreover, the position of the bead can be easily deduced by a basic identification of the centroid of the bead image. This allows us to build a complete model for the observed object, up to a potential multiplicative factor in the pixel intensities.

For the remainder of the paper, let us introduce the following notations: $\|\cdot\|$ represents the Euclidean norm on $\mathbb{R}^N$, $\mathbf{1}_N$ the unit vector of $\mathbb{R}^N$, $\|\cdot\|_{\mathrm{F}}$ the Frobenius norm on $\mathbb{R}^{N\times N}$, $2^{\mathcal{C}}$ the set of all subsets of a given set $\mathcal{C}$, $\mathcal{S}_3$ the set of symmetric matrices in $\mathbb{R}^{3\times 3}$, $\mathcal{S}_3^+$ the set of symmetric positive semi-definite matrices in $\mathbb{R}^{3\times 3}$, $\mathcal{S}_3^{++}$ the set of symmetric positive definite matrices in $\mathbb{R}^{3\times 3}$, and $\boldsymbol{I}_3$ the identity matrix of $\mathbb{R}^3$.

Continuous Modeling Let $x\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$ represent a sphere with diameter τ, modeling the bead fluorescence. The bead is assumed to emit with a uniform intensity, so we set, for all $\mathbf{u} \in \mathbb{R}^3$, $x(\boldsymbol{u}) = 1$ if $\|\mathbf{u}\| \leq \tau$, and $x(\boldsymbol{u}) = 0$ otherwise. The ground truth florescence intensity is then scaled by a factor $\bar{\beta}_{\mathrm{c}} \in (0, +\infty)$, which is usually unknown.

Let $y\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$ be the image of the bead obtained with the MPM device. The observation model takes the form of (1). For the sake of simplicity, in this calibration step, we adopt an assumption of an additive i.i.d. Gaussian noise. Thus, we have:

$$(\forall \boldsymbol{u} \in \mathbb{R}^3) \quad y(\boldsymbol{u}) = \bar{\alpha} + \bar{\beta}_{\mathrm{c}}(\bar{h} \star x)(\boldsymbol{u}) + \nu(\boldsymbol{u}), \tag{2}$$

where ν is a function accounting for the Gaussian noise, $\bar{\alpha} \in \mathbb{R}$ is a background term, and $\bar{h} \in \mathrm{L}^1(\mathbb{R}^3)$ is a convolution kernel (i.e., the PSF) such that $\bar{h}(\boldsymbol{u}) \geq 0$ almost everywhere on $\mathbb{R}^3$ and satisfying

$$\int_{\mathbb{R}^3} \bar{h}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} = 1. \tag{3}$$

Given the observation model (2), the goal is to estimate the unknowns $\bar{\alpha}$, $\bar{\beta}_{\mathrm{c}}$, and $\bar{h}$.

Discrete Modeling MPM acquisitions are done point-by-point, on a voxel grid with a given resolution. Hence, one only has access to a sampling of y on a bounded discrete set Ω of $\mathbb{R}^3$, paved into $N \in \mathbb{N}$ voxels with mass centers $(\boldsymbol{\omega}_n)_{1\leq n\leq N} \in (\mathbb{R}^3)^N$. The discretized version of the continuous problem (2) then becomes:

$$\boldsymbol{y} = \bar{\alpha}\mathbf{1}_N + \bar{\beta}(\bar{\boldsymbol{h}} * \boldsymbol{x}) + \boldsymbol{\nu}, \tag{4}$$

where $\boldsymbol{y} = (y(\boldsymbol{\omega}_n))_{1\leq n\leq N} \in \mathbb{R}^N$ (resp. $\boldsymbol{x} = (x(\boldsymbol{\omega}_n))_{1\leq n\leq N} \in \mathbb{R}^N$ and $\boldsymbol{\nu} = (\nu(\boldsymbol{\omega}_n))_{1\leq n\leq N} \in \mathbb{R}^N$) is the discretization of y (resp. x and ν) on Ω, $\bar{\boldsymbol{h}} = (\bar{h}_n)_{1\leq n\leq N} \in \mathbb{R}^N$ is a discrete convolution kernel satisfying $\bar{\boldsymbol{h}} \in \Delta_N$ where Δ_N is the simplex of $\mathbb{R}^N$ defined as

$$\Delta_N = \left\{ \boldsymbol{h} = (h_n)_{1\leq n\leq N} \in \mathbb{R}^N \,\middle|\, (\forall n \in \{1,\ldots,N\})\ h_n \geq 0 \text{ and } \sum_{n=1}^{N} h_n = 1 \right\}, \tag{5}$$

and $*$ states for the discrete 3D convolution operator with appropriate padding (typically, circular padding). In addition, $\bar{\beta} > 0$ is a scaling factor for the discrete model, which differs from the one in the continuous model, $\bar{\beta}_{\mathrm{c}}$. Indeed, observe that $\bar{\boldsymbol{h}}$ is not, per se, a discretization of h. In particular, an implicit rescaling is assumed to ensure the sum of $\bar{\boldsymbol{h}}$ equals 1, for simplicity. The goal is thus to provide estimates of $\bar{\boldsymbol{h}}$, and the shift/scaling factors $(\bar{\alpha}, \bar{\beta})$, given the knowledge of the measurements $\boldsymbol{y}$ and the bead model $\boldsymbol{x}$.

2.3. Proposed minimization problem

Following [17], we introduce a prior so as to promote a PSF with a Gaussian shape. This leads to solving the following regularized variational problem:

$$\underset{\substack{\alpha\in\mathcal{A},\,\beta\in\mathcal{B},\\ \boldsymbol{h}\in\Delta_N,\,\boldsymbol{D}\in\mathcal{S}_3^+}}{\text{minimize}}\ \frac{1}{2}\|\boldsymbol{y} - \alpha\mathbf{1}_N - \beta(\boldsymbol{h} * \boldsymbol{x})\|^2 + \lambda\,\mathcal{KL}\left(\boldsymbol{h} \,\|\, \zeta\boldsymbol{g}(\boldsymbol{D} + \epsilon_1\boldsymbol{I}_3)\right) + \epsilon_2\|\boldsymbol{D}\|_{\mathrm{F}}^2. \tag{6}$$

Hereabove, $\epsilon_1 > 0$, so for every $\boldsymbol{D} \in \mathcal{S}_3^+$, matrix $\boldsymbol{D}+\epsilon_1\boldsymbol{I}_3 \in \mathcal{S}_3^{++}$, and $\boldsymbol{g}(\boldsymbol{D}+\epsilon_1\boldsymbol{I}_3) \in \mathbb{R}^N$ states for the discretization on Ω of the centered Gaussian density function with inverse covariance $\boldsymbol{D} + \epsilon_1\boldsymbol{I}_3$, corresponding to

$$(\forall \boldsymbol{S} \in \mathcal{S}_3^{++})\,(\forall n \in \{1,\ldots,N\}) \quad [\boldsymbol{g}(\boldsymbol{S})]_n = \sqrt{\frac{|\boldsymbol{S}|}{(2\pi)^3}}\exp\left(-\frac{1}{2}\boldsymbol{\omega}_n^\top\boldsymbol{S}\boldsymbol{\omega}_n\right). \tag{7}$$

Moreover, $\lambda > 0$ is a regularization parameter, and $\zeta \in (0, +\infty)$ is a known scaling parameter accounting for the measure of the discrete Gaussian function $\boldsymbol{g}(\boldsymbol{D} + \epsilon_1\boldsymbol{I}_3)$ on the grid Ω. When the grid is sufficiently fine, which we will assume throughout this discussion, ζ can be approximated by

$$\zeta \simeq r_X r_Y r_Z, \tag{8}$$

where $r = (r_X, r_Y, r_Z) \in (0, +\infty)^3$ represents the dimensions, in micrometers, of the voxels in the captured image (i.e., the image resolution). Sets $\mathcal{A} = [\alpha_-, \alpha_+]$ and $\mathcal{B} = [\beta_-, \beta_+]$ are real intervals corresponding to known bounds on α and β, respectively. $\mathcal{KL}$ denotes the Kullback-Leibler divergence between two discrete probability distributions, defined as

$$(\forall(\boldsymbol{p}, \boldsymbol{q}) \in (\Delta_N \cap (0, +\infty))^2) \quad \mathcal{KL}\left(\boldsymbol{p} \| \boldsymbol{q}\right) = \sum_{n=1}^{N} p_n \log\left(\frac{p_n}{q_n}\right). \tag{9}$$

Finally $\epsilon_2 \in (0, +\infty)$ is an arbitrarily small penalty parameter.

Let us denote, for conciseness:

$$(\forall \boldsymbol{h} \in \Delta_N)\,(\forall \boldsymbol{D} \in \mathcal{S}_3^+) \quad \Psi(\boldsymbol{h}, \boldsymbol{D}) = \mathcal{KL}\left(\boldsymbol{h} \,\|\, \zeta \boldsymbol{g}(\boldsymbol{D} + \epsilon_1 \boldsymbol{I}_3)\right). \tag{10}$$

To avoid potential numerical issues with the logarithmic term in $\mathcal{KL}$ divergence, we extend Ψ into a twice continuously differentiable function $\tilde{\Psi}$ on $\Delta_N \times \mathcal{S}_3$, defined as

$$\begin{aligned}(\forall \boldsymbol{h} \in \Delta_N)\,(\forall \boldsymbol{D} \in \mathcal{S}_3) \quad \tilde{\Psi}(\boldsymbol{h}, \boldsymbol{D}) = \sum_{n=1}^{N} \Big(& E(h_n) - h_n \ln \zeta \\ & + \frac{h_n}{2}\left(3\ln(2\pi) + \Phi(\boldsymbol{D}) + \boldsymbol{\omega}_n^\top (\boldsymbol{D} + \epsilon_1 \boldsymbol{I}_3)\boldsymbol{\omega}_n\right)\Big), \end{aligned} \tag{11}$$

where, for any $\boldsymbol{D} \in \mathcal{S}_3$ diagonalized as $\boldsymbol{D} = \boldsymbol{U}\mathrm{Diag}(\boldsymbol{s})\boldsymbol{U}^\top$ with $\boldsymbol{U}$ an orthogonal matrix of $\mathbb{R}^{3\times 3}$ and $\boldsymbol{s} \in \mathbb{R}^3$,

$$\Phi(\boldsymbol{D}) = \sum_{i=1}^{3} \phi(s_i), \tag{12}$$

with ϕ defined as

$$(\forall t \in \mathbb{R}) \quad \phi(t) = \begin{cases} -\ln(t + \epsilon_1) & \text{if } t \geq 0, \\ -\ln(\epsilon_1) - \epsilon_1^{-1} t + \epsilon_1^{-2} t^2 & \text{otherwise,} \end{cases} \tag{13}$$

and

$$(\forall u \in \mathbb{R}) \quad E(u) = \begin{cases} u \ln(u) & \text{if } u > 0, \\ 0 & \text{if } u = 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{14}$$

In order to model the constraints in Problem (6), we introduce, for any non-empty, closed, convex set $C \subseteq \mathbb{R}^n$, its indicator function ι_C, namely $\iota_C(x) = 0$ if $x \in C$, $\iota_C(x) = +\infty$ otherwise. In a nutshell, Problem (6) is equivalent to minimize, for $(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^N \times \mathcal{S}_3$,

$$\begin{aligned} F(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) = \frac{1}{2}\|\boldsymbol{y} - \alpha \boldsymbol{1}_N - \beta(\boldsymbol{h} * \boldsymbol{x})\|^2 + \lambda \tilde{\Psi}(\boldsymbol{h}, \boldsymbol{D}) \\ + \iota_{\mathcal{A}}(\alpha) + \iota_{\mathcal{B}}(\beta) + \iota_{\Delta_N}(\boldsymbol{h}) + \iota_{\mathcal{S}_3^+}(\boldsymbol{D} + \epsilon_1 \boldsymbol{I}_3) + \epsilon_2 \|\boldsymbol{D}\|_F^2. \end{aligned} \tag{15}$$

It will be established in Proposition 6 that, since $\epsilon_2 \in (0, +\infty)$, F admits a minimizer.

2.4. Minimization algorithm

The objective function in (15) is nonconvex. However, it is convex with respect to each variable. A common strategy for such scenarios consists of using alternating minimization techniques. At each iteration, we minimize F with respect to a variable while the other variables are held constant. This simple method, also known as Block Coordinate Descent or the nonlinear Gauss–Seidel method, has previously been utilized for PSF model fitting in microscopy, for instance in [39, 53]. Yet, closed-form updates are not always available. Moreover, convergence to an optimal solution using this method is only guaranteed when the partial functions on each variable and at each iteration are uniformly strongly convex [55, 4]. This stringent assumption is not met here. To ensure both efficiency and guarantees of convergence, a strategy enriched with proximal tools is preferable, as highlighted for instance in [17, 19, 8]. The particular structure of (15) suggests a hybrid proximal alternating scheme ([71, 58, 14]) that we present hereafter.

2.4.1. Preliminaries In this section, we give some mathematical definitions that will be useful in the subsequent parts of the paper. Let $\mathcal{H}$ be a Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$.

Definition 1 (Domain)**.** [3, Def.1.4] Let $f \colon \mathcal{H} \longrightarrow (-\infty, +\infty]$. The domain of f is the set of all points in $\mathcal{H}$ where f attains finite values. It is defined as $\mathrm{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$. A function f is said to be proper if its domain is nonempty. The set of proper, lower semi-continuous, convex functions of $\mathcal{H}$ is denoted by $\Gamma_0(\mathcal{H})$.

Next, we introduce the concept of the proximity operator. This tool will be at the core of our iterative algorithm.

Definition 2 (Proximity operator)**.** [3, Def.12.23] Let $f \in \Gamma_0(\mathcal{H})$. The proximity operator of f at $x \in \mathcal{H}$ is defined as $\mathrm{prox}_f(x) = \underset{u \in \mathcal{H}}{\mathrm{argmin}} f(u) + \frac{1}{2}\|u - x\|^2$. In particular, for a non-empty closed convex set C, the proximity operator of ι_C coincides with the orthogonal projection onto C.

One of the foundational concepts of convex analysis is the subdifferential, which generalizes the concept of a derivative.

Definition 3 (Moreau subdifferential)**.** [3, Def.16.1] Let $f \in \Gamma_0(\mathcal{H})$. The Moreau subdifferential of f, denoted by ∂f, is

$$\begin{aligned} \partial f : \mathcal{H} &\longrightarrow 2^{\mathcal{H}} \\ x &\longmapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H}) \langle y - x | u \rangle + f(x) \leq f(y)\}. \end{aligned} \tag{16}$$

When f is both convex and Fréchet differentiable, its subdifferential at any point x just contains its gradient.

Finally, we present some useful properties of the subdifferential and the proximity operator.

Proposition 1. *[3, Chap.16] Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.*

(i) If $\operatorname{int}(\operatorname{dom} g) \cap \operatorname{dom} f \neq \varnothing$ *or* $\operatorname{dom} g \cap \operatorname{int}(\operatorname{dom} f) \neq \varnothing$*, then*

$$\partial f + \partial g = \partial(f+g). \tag{17}$$

(ii) The following equivalence holds:

$$(\forall x \in \mathcal{H}) \quad p = \operatorname{prox}_f(x) \iff x - p \in \partial f(p). \tag{18}$$

With these mathematical concepts explained, we are now ready to get into more details of our algorithm.

2.4.2. Proposed algorithm Guided by the methodology in [71] and [58], we opt for an alternating minimization algorithm acting on the variables α, β, $\boldsymbol{h}$, and $\boldsymbol{D}$. At each iteration of the algorithm, we update a specific variable employing one of the three following distinct schemes applied to the partial function with respect to this variable: an exact update step, a proximal step, or a proximal linearized step (also called forward-backward, or proximal gradient).

Within our framework, the variables α and β undergo exact step updates, $\boldsymbol{D}$ is updated through a proximal point step, and $\boldsymbol{h}$ is updated through to a proximal linearized step. We provide the expressions for these updates in the section that follows. Our algorithm, called GENTLE (**G**aussian k**E**r**N**el fi**T**ting using **L**arge B**E**ads), is summarized in Algorithm 1, where we use the short notation

$$(\forall(\alpha,\beta,\boldsymbol{h},\boldsymbol{D}) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathcal{S}_3) \quad \tilde{F}(\alpha,\beta,\boldsymbol{h},\boldsymbol{D}) = \lambda\tilde{\Psi}(\boldsymbol{h},\boldsymbol{D}) + \iota_{\Delta_N}(\boldsymbol{h}), \tag{19}$$

and the stepsizes $(\gamma_h,\gamma_D) \in (0,+\infty)^2$. In addition, we denote $\boldsymbol{X} \in \mathbb{R}^{N\times N}$ the block Toepliz matrix such that, for all $(\boldsymbol{u},\boldsymbol{x}) \in (\mathbb{R}^N)^2$, $\boldsymbol{X}\boldsymbol{u} = \boldsymbol{x} * \boldsymbol{u}$.

Algorithm 1: GENTLE: **G**aussian k**E**r**N**el fi**T**ting using **L**arge B**E**ads

Inputs: Let $(\alpha^{(0)},\beta^{(0)},\boldsymbol{h}^{(0)},\boldsymbol{D}^{(0)}) \in \mathcal{A}\times\mathcal{B}\times\Delta_N\times\mathcal{S}_3^+$, $(\gamma_h,\gamma_D)\in(0,+\infty)^2$.

for $\ell = 0,1,\ldots$ **do**

$$\alpha^{(\ell+1)} = \operatorname*{argmin}_{\alpha\in\mathbb{R}} F(\alpha,\beta^{(\ell)},\boldsymbol{h}^{(\ell)},\boldsymbol{D}^{(\ell)})$$

$$\beta^{(\ell+1)} = \operatorname*{argmin}_{\beta\in\mathbb{R}} F(\alpha^{(\ell+1)},\beta,\boldsymbol{h}^{(\ell)},\boldsymbol{D}^{(\ell)})$$

$$\boldsymbol{h}^{(\ell+1)} = \operatorname{prox}_{\gamma_h\tilde{F}}\left(\boldsymbol{h}^{(\ell)} - \gamma_h\beta^{(\ell+1)}\boldsymbol{X}^\top(\boldsymbol{y} - \alpha^{(\ell+1)}\mathbf{1}_N - \beta^{(\ell+1)}\boldsymbol{X}\boldsymbol{h}^{(\ell)})\right)$$

$$\boldsymbol{D}^{(\ell+1)} = \operatorname{prox}_{\gamma_D F(\alpha^{(\ell+1)},\beta^{(\ell+1)},\boldsymbol{h}^{(\ell+1)},\cdot)}(\boldsymbol{D}^{(\ell)})$$

Return: α, β, $\boldsymbol{h}$, $\boldsymbol{D}$.

2.4.3. Expressions of the updates We now explicit the expressions for the four updates involved at each iteration of GENTLE.

Proposition 2 (Update on α)**.** *Let* $(\beta, \boldsymbol{h}, \boldsymbol{D}) \in \mathbb{R} \times \mathbb{R}^N \times \mathcal{S}_3$*. Then the minimizer of* $F(\cdot, \beta, \boldsymbol{h}, \boldsymbol{D})$ *at* α *is given by*

$$\alpha = \mathrm{proj}_{\mathcal{A}} \left(\frac{1}{N} (\boldsymbol{y} - \beta(\boldsymbol{h} * \boldsymbol{x}))^\top \mathbf{1}_N \right), \tag{20}$$

where $\mathrm{proj}_{\mathcal{A}}$ *is the orthogonal projection on the closed convex set* $\mathcal{A}$*, reading* $\mathrm{proj}_{\mathcal{A}}(\alpha) = \max(\alpha_-, \min(\alpha, \alpha_+))$*.*

Proof. Let $f = F(\cdot, \beta, \boldsymbol{h}, \boldsymbol{D})$. Then, for any $\alpha \in \mathbb{R}$,

$$\begin{aligned} 0 \in \partial f(\alpha) &\iff (\boldsymbol{y} - \beta(\boldsymbol{h} * \boldsymbol{x}))^\top \mathbf{1}_N - \alpha N \in \partial \iota_{\mathcal{A}}(\alpha) \\ &\iff \frac{1}{N} (\boldsymbol{y} - \beta(\boldsymbol{h} * \boldsymbol{x}))^\top \mathbf{1}_N - \alpha \in \partial \iota_{\mathcal{A}}(\alpha) \\ &\iff \alpha = \mathrm{proj} \left(\frac{1}{N} (\boldsymbol{y} - \beta(\boldsymbol{h} * \boldsymbol{x}))^\top \mathbf{1}_N \right), \end{aligned} \tag{21}$$

where we used Proposition 1(i) and (ii) □

Proposition 3 (Update on β)**.** *Let* $(\alpha, \boldsymbol{h}, \boldsymbol{D}) \in \mathbb{R} \times \mathbb{R}^N \times \mathcal{S}_3$*, and* $\gamma_\beta \in (0, +\infty)$*. Then the minimizer of of* $F(\alpha, \cdot, \boldsymbol{h}, \boldsymbol{D})$ *is given by*

$$\beta = \mathrm{proj}_{\mathcal{B}} \left(\frac{(\boldsymbol{y} - \alpha \mathbf{1}_N)^\top (\boldsymbol{h} * \boldsymbol{x})}{\|\boldsymbol{h} * \boldsymbol{x}\|^2} \right). \tag{22}$$

where $\mathrm{proj}_{\mathcal{B}}$ *is the orthogonal projection on the closed convex set* $\mathcal{B}$*, reading* $\mathrm{proj}_{\mathcal{B}}(\beta) = \max(\beta_-, \min(\beta, \beta_+))$*.*

We skip the proof which is similar to the previous one.

Proposition 4 (Update on $\boldsymbol{h}$)**.** *Let* $(\alpha, \beta, \boldsymbol{h}', \boldsymbol{D}) \in \mathbb{R}^2 \times \mathbb{R}^N \times \mathcal{S}_3$*, and* $\gamma_h \in (0, +\infty)$*. Then the proximity operator of* $\gamma_h \tilde{F}(\alpha, \beta, \cdot, \boldsymbol{D})$ *at* $\boldsymbol{h}'$ *is given by*

$$\mathrm{prox}_{\gamma_h \tilde{F}(\alpha, \beta, \cdot, \boldsymbol{D})}(\boldsymbol{h}') = (\rho^{-1} W(\rho \exp(w_n(\widehat{\mu})))_{1 \leq n \leq N}, \tag{23}$$

where W *denotes the Lambert-W function [21],*

$$\rho = \frac{1}{\lambda \gamma_h}, \tag{24}$$

and, for every $n \in \{1, \ldots, N\}$*,* w_n *is the function defined as*

$$(\forall \mu \in \mathbb{R}) \quad w_n(\mu) = -1 - c_n + \rho(h'_n - \mu) + \ln \zeta, \tag{25}$$

with

$$c_n = \frac{1}{2} \left(3 \log(2\pi) + \Phi(D) + \boldsymbol{\omega}_n^\top (\boldsymbol{D} + \epsilon_1 \boldsymbol{I}_3) \boldsymbol{\omega}_n \right). \tag{26}$$

Moreover, $\widehat{\mu} \in \mathbb{R}$ *is the unique zero of the function*

$$(\forall \mu \in \mathbb{R}) \quad \kappa(\mu) = \rho^{-1} \sum_{n=1}^{N} W(\rho \exp(w_n(\mu))) - 1. \tag{27}$$

Proof. The function to minimize is $\tilde{F}(\alpha, \beta, \zeta, \cdot, \boldsymbol{D}) + \frac{1}{2\gamma_h}\| \cdot - \boldsymbol{h}'\|^2$. It is equivalent to minimizing the function f defined by

$$(\forall \boldsymbol{h} \in (0, +\infty)^N) \quad f(\boldsymbol{h}) = \lambda\gamma_h \sum_{n=1}^{N} (h_n \ln h_n - h_n \ln \zeta + c_n h_n) + \iota_{\Delta_N}(\boldsymbol{h}) + \frac{1}{2}\|\boldsymbol{h} - \boldsymbol{h}'\|^2. \tag{28}$$

The Lagrangian function associated with the minimization of f reads

$$\begin{aligned}(\forall \boldsymbol{h} \in (0, +\infty)^N)(\forall \mu \in \mathbb{R}) \quad \mathcal{L}(\boldsymbol{h}, \mu) = \lambda\gamma_h \sum_{n=1}^{N} \Bigg(& h_n \ln h_n + h_n (-\ln \zeta + c_n) \\ & + \frac{1}{2\lambda\gamma_h}(h_n - h'_n)^2 \Bigg) + \mu \left(\sum_{n=1}^{N} h_n - 1 \right). \end{aligned} \tag{29}$$

Since Slater's condition obviously holds, there exists $\hat{\mu} \in \mathbb{R}$ such that $(\boldsymbol{h}, \hat{\mu})$ is a saddle point of $\mathcal{L}$ [5]. By Fermat's rule [3], $\boldsymbol{h}$ is thus obtained by finding a zero of the derivative of $\mathcal{L}(\cdot, \hat{\mu})$. This yields, for every $n \in \{1, \ldots, N\}$,

$$\begin{aligned} & \lambda\gamma_h \left(1 + \ln \hat{h}_n + c_n - \ln \zeta\right) + \hat{h}_n - h'_n + \hat{\mu} = 0, \\ \iff \quad & \ln(\hat{h}_n) + \rho \hat{h}_n = w_n(\hat{\mu}), \\ \iff \quad & \rho \hat{h}_n \exp(\rho \hat{h}_n) = \rho \exp(w_n(\hat{\mu})), \end{aligned} \tag{30}$$

where ρ is given by (24) and $w_n(\hat{\mu}) = -1 + \rho(h'_n - \hat{\mu}) - c_n + \ln \zeta$.

Finally, recalling that the W-Lambert function is such that $(\forall z \in \mathbb{R})\ W(z)\exp(W(z)) = z$, we deduce that

$$\hat{h}_n = \rho^{-1} W\left(\rho \exp(w_n(\hat{\mu}))\right). \tag{31}$$

In addition, one can obtain $\hat{\mu}$ from the linear equality constraint. This amounts to finding a zero of the function κ defined as

$$(\forall \mu \in \mathbb{R}) \quad \kappa(\mu) = \rho^{-1} \sum_{n=1}^{N} W\left(\rho \exp(w_n(\mu)\right) - 1. \tag{32}$$

It was shown in [17], relying on the properties of the W-Lambert function, that κ admits a unique zero. □

The Lambert-W function appearing in Proposition 4 is commonly in the expression of the proximity operators of entropic functions [50, 13]. While its evaluation, requiring the solution of a transcendental equation, can be efficiently achieved using a Newton-based method, the composition of W with the exponential can lead to arithmetic overflow for large inputs. To address this, we employ the asymptotic expansion $W(\exp(u)) \approx u - \log(u)$ for $u > 10^2$ [21].

Proposition 5 (Update on $\boldsymbol{D}$)**.** *Let* $(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) \in \mathbb{R}^2 \times (0, +\infty)^N \times \mathcal{S}_3$*, and* $\gamma_D \in (0, +\infty)$*. Then the proximity operator of* $\gamma_D F(\alpha, \beta, \boldsymbol{h}, \cdot)$ *at* $\boldsymbol{D}$ *is given by*

$$\mathrm{prox}_{\gamma_D F(\alpha,\beta,\boldsymbol{h},\cdot)}(\boldsymbol{D}) = \frac{1}{2}\boldsymbol{V}\mathrm{Diag}\left(\left(\max(\mu_i - \epsilon_1 + \sqrt{(\mu_i + \epsilon_1)^2 + 4m}, 0)\right)_{1\leq i\leq d}\right)\boldsymbol{V}^\top, \tag{33}$$

where $\boldsymbol{\mu} = (\mu_i)_{1\leq i\leq 3}$ *is a vector of eigenvalues of* $(2\epsilon_2\gamma_D + 1)^{-1}\boldsymbol{D} - \boldsymbol{S}$ *and* $\boldsymbol{V}$ *is a* 3×3 *orthogonal matrix such that* $(2\epsilon_2\gamma_D + 1)^{-1}\boldsymbol{D} - \boldsymbol{S} = \boldsymbol{V}\mathrm{Diag}(\boldsymbol{\mu})\boldsymbol{V}^\top$ *with*

$$\boldsymbol{S} = \frac{\gamma_D}{2(2\epsilon_2\gamma_D + 1)}\lambda \sum_{n=1}^{N} h_n \boldsymbol{\omega}_n \boldsymbol{\omega}_n^\top, \tag{34}$$

and $m = \frac{1}{2}\gamma_D\lambda$*.*

Proof. Direct extension of [17, Prop.4]. □

2.4.4. Convergence analysis We now establish the convergence of GENTLE. We first show that the considered minimization problem has at least one solution.

Proposition 6. *The cost function* F *defined is* (15) *is lower-bounded and admits a minimizer.*

Proof. It is clear that F is lower semi-continuous on $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^N \times \mathcal{S}_3$. Let us now show that F is coercive. Since $\mathcal{A}$, $\mathcal{B}$, and Δ_N are bounded subsets, it suffices to show that, for any $(\alpha, \beta, \boldsymbol{h}) \in \mathcal{A} \times \mathcal{B} \times \Delta_N$,

$$F(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) \underset{\substack{\|\boldsymbol{D}\|_{\mathrm{F}} \to +\infty \\ \boldsymbol{D}\in\mathcal{S}_3^+}}{\longrightarrow} +\infty. \tag{35}$$

The following lower bound holds for F:

$$\begin{aligned}(\forall \boldsymbol{D} \in \mathcal{S}_3) \quad F(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) &\geq \lambda c_1 + \frac{\lambda}{2}\sum_{n=1}^{N} h_n \left(3\ln(2\pi) + \Phi(\boldsymbol{D}) + \boldsymbol{\omega_n}^\top(\boldsymbol{D} + \epsilon_1\boldsymbol{I}_3)\boldsymbol{\omega_n}\right) \\ &\quad + \epsilon_2\|\boldsymbol{D}\|_F^2 \\ &\geq \lambda c_1 + \frac{3\lambda}{2}\ln(2\pi) + \frac{\lambda}{2}\Phi(\boldsymbol{D}) + \epsilon_2\|\boldsymbol{D}\|_F^2,\end{aligned} \tag{36}$$

where $c_1 = \inf\left\{\sum_{n=1}^{N} E(h_n) - h_n \ln\zeta \mid \boldsymbol{h} \in \Delta_N\right\}$ and we have used the fact that $\boldsymbol{D} + \epsilon_1\boldsymbol{I}_3 \succ 0$. Finally, given the definition of Φ in (12), it is clear the lower bound in (36) goes to $+\infty$ when $\|\boldsymbol{D}\|_{\mathrm{F}} \to +\infty$. Therefore, F admits a minimizer. □

Let us now establish the convergence of the iterates generated by Algorithm 1. We base our analysis on the convergence result stated in [71, Thm. 2].

Theorem 1. *Let* $(\alpha^{(0)}, \beta^{(0)}, \boldsymbol{h}^{(0)}, \boldsymbol{D}^{(0)}) \in \mathcal{A} \times \mathcal{B} \times \Delta_N \times \mathcal{S}_3^+$ *be an initial point. For every* $\ell \in \mathbb{N}$*, let* $\boldsymbol{t}^{(\ell)} = (\alpha^{(\ell)}, \beta^{(\ell)}, \boldsymbol{h}^{(\ell)}, \boldsymbol{D}^{(\ell)})$ *be the sequence generated by Algorithm 1. Assume that* $\gamma_h < \frac{2}{\overline{L}}$ *with* $\overline{L} = \beta_+^2 \max\{(|\operatorname{DFT}(\boldsymbol{x})_n|^2)_{1\leq n\leq N}\}$ *and* DFT *denotes the discrete Fourier transform. Assume that the grid is discretized finely enough such that the true bead* $\boldsymbol{x}$ *has at least one pixel with unit intensity. Then* $(\boldsymbol{t}^{(\ell)})_{\ell\in\mathbb{N}}$ *converges to a critical point of the objective function* (15)*,* $\hat{\boldsymbol{t}} = (\hat{\alpha}, \hat{\beta}, \hat{\boldsymbol{h}}, \hat{\boldsymbol{D}})$*.*

Proof. We show that the assumptions required by [71, Thm. 2] are satisfied.

- **Splitting of the objective function**. We first split the objective function F into 1) a coupling term $f\colon \mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathcal{S}_3$ which is block convex, differentiable with a Lipschitz continuous gradient on bounded subsets, 2) a separable sum of functions $g_\alpha \in \Gamma_0(\mathbb{R})$, $g_\beta \in \Gamma_0(\mathbb{R})$, $g_h \in \Gamma_0(\mathbb{R}^N)$ and $g_D \in \Gamma_0(\mathcal{S}_3)$ such that:

$$F(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) = f(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) + g_\alpha(\alpha) + g_\beta(\beta) + g_h(\boldsymbol{h}) + g_D(\boldsymbol{D}). \tag{37}$$

To do so, we define

$$(\forall(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}^N\times\mathcal{S}_3)$$
$$f(\alpha, \beta, \boldsymbol{h}, \boldsymbol{D}) = \frac{1}{2}\|\boldsymbol{y} - \alpha\boldsymbol{1}_N - \beta(\boldsymbol{h}*\boldsymbol{x})\|^2 + \frac{\lambda}{2}\sum_{n=1}^N h_n\left(\Phi(\boldsymbol{D}) + \boldsymbol{\omega}_n^\top(\boldsymbol{D} + \epsilon_1\boldsymbol{I}_3)\boldsymbol{\omega}_n\right), \tag{38}$$

$$g_\alpha = \iota_{\mathcal{A}}, \quad g_\beta = \iota_{\mathcal{B}}, \quad g_D = \iota_{\mathcal{S}_3^+} + \epsilon_2\|\boldsymbol{D}\|_{\mathrm{F}}^2, \tag{39}$$

and

$$(\forall \boldsymbol{h} \in \mathbb{R}^N) \quad g_h(\boldsymbol{h}) = \lambda\sum_{n=1}^N\left(E(h_n) - h_n\ln(\zeta) + \frac{3h_n}{2}\ln(2\pi)\right). \tag{40}$$

- **Properties of the partial functions**. We first need to check that the partial functions with respect to α and β are strongly convex along the iterates $(\boldsymbol{t}^{(\ell)})_{\ell\in\mathbb{N}}$, with modulus independent of ℓ. It is clear that it is the case for the partial function with respect to variable α. On the other hand, for any $\ell \in \mathbb{N}$, let us denote $F_{\mathsf{B}}^{(\ell)}$ the function $\beta \mapsto F(\alpha^{(\ell+1)}, \beta, \boldsymbol{h}^{(\ell)}, \boldsymbol{D}^{(\ell)})$. Then, since the bead $\boldsymbol{x}$ has at least one pixel value equal to 1, say at ω_i, $i \in \{1, \dots, N\}$,

$$\begin{aligned}\nabla^2 F_{\mathsf{B}}^{(\ell)}(\beta) &= \|\boldsymbol{h}^{(\ell)} * \boldsymbol{x}\|^2\\ &\geq \|\boldsymbol{h}^{(\ell)} * \boldsymbol{\delta}_i\|^2\\ &= \|\boldsymbol{h}^{(\ell)}\|^2,\end{aligned} \tag{41}$$

with $\boldsymbol{\delta}_i \in \mathbb{R}^N$ the image corresponding to one illuminated pixel with intensity equal to 1 at ω_i, and 0 elsewhere. The inequalitiy holds because $\boldsymbol{h}^{(\ell)} \in (0, +\infty)^N$. Moreover, since $\|\cdot\| \geq \frac{1}{N}\|\cdot\|_1$ and $\boldsymbol{h}^{(\ell)} \in \Delta_N$, we deduce that $\nabla^2 F_{\mathsf{B}}^{(\ell)}(\beta) \geq \frac{1}{N^2}$. Therefore, $F_{\mathsf{B}}^{(\ell)}$ is $\frac{1}{N^2}$-strongly convex. Secondly, for the proximal linearized step on variable $\boldsymbol{h}$, the step-size γ_h has to satisfy the condition:

$$\gamma_h < \frac{2}{L}, \tag{42}$$

where L is the Lipschitz constant of the gradient of the differentiable component $\boldsymbol{h} \longmapsto \frac{1}{2}\|\boldsymbol{y} - \alpha \mathbf{1}_N - \beta(\boldsymbol{h} * \boldsymbol{x})\|^2$. It can easily been shown that $L \leq \overline{L}$.

- **Boundedness of the sequence.** Let us demonstrate that the sequence $(\boldsymbol{t}^{(\ell)})_{\ell \in \mathbb{N}}$ is bounded. Since GENTLE alternates exact, proximal and proximal linearized updates, the sequence $(F(\boldsymbol{t}^{(\ell)}))_{\ell \in \mathbb{N}}$ is non-increasing. In addition, the objective function is coercive as seen in Proposition 6. This implies that $(\boldsymbol{t}^{(\ell)})_{\ell \in \mathbb{N}}$ is bounded.
- **Kurdyka-Łojasiewicz's inequality.** Lastly, the objective function F must satisfy the so-called Kurdyka-Łojasiewicz's inequality [49, 7]. This property is satisfied by a wide range of functions in the context of image processing. In the present case, we omit the proof that F satisfies the Kurdyka-Łojasiewicz's inequality as a very similar proof to the one in [17] could be driven.

Finally, applying [71, Thm. 2], we deduce that $(\boldsymbol{t}^{(\ell)})_{\ell \in \mathbb{N}}$ converges to a critical point of the objective function F. In practice, we assess the convergence using the stopping criterion $\|\boldsymbol{t}^{(\ell+1)} - \boldsymbol{t}^{(\ell)}\| \leq \varepsilon$ with $\varepsilon = 10^{-7}$.

□

2.5. Validation of the PSF calibration method

In this section, we present experiments conducted to validate the proposed PSF calibration method GENTLE. These experiments have been designed to test the performance of our method both in simulated scenarios and in real-life settings. In all our experiments, the regularization parameter λ is determined through an empirical grid search. Specifically, we select the λ that minimizes the criterion $\|\boldsymbol{y} - \hat{\alpha} - \hat{\beta}\boldsymbol{g}(\hat{\boldsymbol{D}} + \epsilon_1 \boldsymbol{I}_3) * \boldsymbol{x}\|^2$, given estimates $(\hat{\alpha}, \hat{\beta}, \hat{\boldsymbol{D}})$ produced by GENTLE for a certain λ. The other hyper-parameters are set to $(\alpha_-, \alpha_+) = (0, 1)$, $(\beta_-, \beta_+) = (0, 3)$, and $\epsilon_1 = \epsilon_2 = 10^{-6}$.

2.5.1. Experiments in simulated scenarios We simulate a 3D synthetic bead image $\boldsymbol{x} \in \mathbb{R}^N$ of 1μm of diameter on a regularly spaced grid with size $N = 40 \times 40 \times 80$ and voxel dimension $0.05 \times 0.05 \times 0.1\mu\text{m}^3$ (which is a typical resolution grid in MPM). The observation $\boldsymbol{y} \in \mathbb{R}^N$ is then simulated as $\boldsymbol{y} = \bar{\alpha} + \bar{\beta}\bar{\boldsymbol{h}} * \boldsymbol{x} + \boldsymbol{\nu}$, for fixed values of $\bar{\alpha} \in [0, +\infty)$, $\bar{\beta} \in [0, +\infty)$, $\bar{\boldsymbol{h}} \in \Delta_N$ and $\boldsymbol{\nu} \in \mathbb{R}^N$ the realization of a zero-mean Gaussian noise, with standard deviation chosen so as to obtain a given input signal-to-noise ratio (SNR).

We choose generalized exponential distributions shapes for the ground truth PSF $\bar{\boldsymbol{h}}$. This family of distributions not only includes Gaussian ones, but also effectively captures those with shorter or longer tails. In microscopy, the PSF often deviates from a ideal Gaussian form, optical aberrations being a notable factor influencing this shape variation. [62]. These distributions are uniquely defined by a matrix $\bar{\boldsymbol{S}} \in \mathcal{S}_3^+$ and a parameter $\eta > 0$, as

$$\bar{\boldsymbol{h}} \propto \left(\exp\left\{ -\frac{1}{2} \left(\boldsymbol{\omega}_n^\top (\bar{\boldsymbol{S}}) \boldsymbol{\omega}_n \right)^{\eta/2} \right\} \right)_{1 \leq n \leq N}, \tag{43}$$

where $\bar{\boldsymbol{h}}$ is normalized so that $\bar{\boldsymbol{h}} \in \Delta_N$. When $\eta = 2$, we retrieve a Gaussian distribution. The matrix $\bar{\boldsymbol{S}}$ can be designed to represent a kernel with a specific inclination and width. Formally, $\bar{\boldsymbol{S}}$ can be decomposed in an orthonormal basis so that it only depends on the second and third Euler angles $(\bar{\theta}, \bar{\varphi}) \in [0, \pi] \times [-\pi, \pi]$ and on the eigenvalues $\bar{\boldsymbol{s}} \in (0, +\infty)^3$ for each principal direction. Indeed, for symmetry reasons, the first Euler angle, corresponding to a rotation around the vertical axis, has no effect so can be set to 0. We thus adopt the notation $\bar{\boldsymbol{S}} = \boldsymbol{S}(\bar{\theta}, \bar{\varphi}, \bar{\boldsymbol{s}})$. GENTLE provides an estimate $\hat{\boldsymbol{D}}$ such that $\hat{\boldsymbol{D}} + \epsilon_1 \boldsymbol{I}_N$ approaches $\boldsymbol{S}(\bar{\theta}, \bar{\varphi}, \bar{\boldsymbol{s}})$.

We draw a comparison of our method GENTLE with another method, which will refer to as Nonlinear Least Squares (NLS) and which has been extensively used in MPM [26, 47, 76]. It corresponds to considering directly the following non-regularized and non-multiconvex problem:

$$\underset{\substack{\alpha \in \mathcal{A},\, \beta \in \mathcal{B}, \\ \theta \in [0,\pi],\, \varphi \in [-\pi,\pi],\, \boldsymbol{s} \in (0,+\infty)^3}}{\text{minimize}} \quad \frac{1}{2} \|\boldsymbol{y} - \alpha \mathbf{1}_N - \beta \boldsymbol{g}(\boldsymbol{S}(\theta, \varphi, \boldsymbol{s})) * \boldsymbol{x}\|^2, \tag{44}$$

where $\boldsymbol{g}(\boldsymbol{S}(\theta, \varphi, \boldsymbol{s}))$ is the normalized discrete Gaussian kernel on Ω defined with the Euler angles (θ, φ) and the eigenvalues $\boldsymbol{s}$. This problem can be tackled using a Levenberg-Marquardt solver [56]. Note that, despite its popularity, NLS has its shortcomings, notably it lacks convergence guarantees.

Our results are reported in Figure 2 using the Percent root-mean-square difference (PRD) metric, defined as $100 \times \cdot \|\hat{\alpha} + \hat{b}(\hat{\boldsymbol{h}} * \boldsymbol{x}) - \bar{\alpha} - \bar{b}(\bar{\boldsymbol{h}} * \boldsymbol{x})\| / \|\bar{\alpha} + \bar{b}(\bar{\boldsymbol{h}} * \boldsymbol{x})\|$. In the case of GENTLE, variable $\hat{\boldsymbol{h}}$ is obtained as an output, while for NLS, we set $\hat{\boldsymbol{h}} = \boldsymbol{g}(\boldsymbol{S}(\hat{\theta}, \hat{\varphi}, \hat{\boldsymbol{s}}))$. For this experiment, the true Euler angles were set to $(\bar{\theta}, \bar{\varphi}) = (5\pi/6, \pi/6)$, the true eigenvalues to $\bar{\boldsymbol{s}} = (138.6, 138.6, 3.2)$, and the voxels dimensions to $0.049 \times 0.049 \times 0.1 \mu\text{m}^3$. The results were averaged over 10 random noise realisations. As one can observe, NLS performance degrades significantly for values of the model exponent η lower or larger than 2, i.e., when there is a mismatch between the assumed Gaussian model and the true one. By contrast, although promoting Gaussian shapes using a regularization strategy, GENTLE manages to provide accurate and robust estimations even in the presence of non-Gaussian generalized exponential PSF shapes.

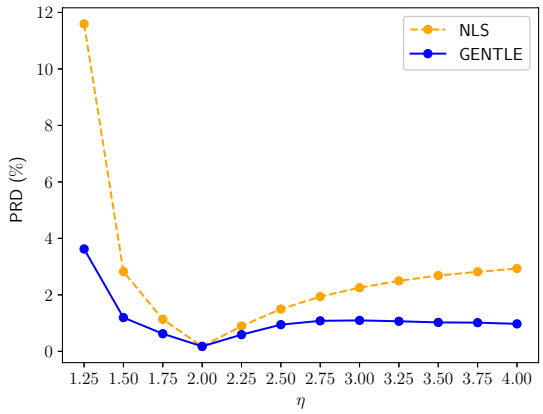

Figure 2. Percent root-mean-square difference (PRD) as a function of the parameter η, for both GENTLE and NLS, for a noise level corresponding to SNR=10dB.

2.5.2. Experiments on real beads in homogeneous medium The next step of our validation process consists in testing our method on real beads acquired under optimal conditions, namely, beads within in a homogenous medium. Such setting is typical in MPM device calibration, although it usually involves sub-resolution beads instead of 1μm diameter beads. The PSF depends on the system optical configuration, including the lens properties and acquisition settings [31, 20]. We evaluate the PSF estimation using the Full Width at Half Maximum (FWHM) as a metric, as it is commonly done in the microscopy community. Measured in micrometers for our study, the FWHM represents the width of a bell-shaped curve when it is at half peak value. In the ideal case of multiphoton acquisitions where the optics are assumed "perfect", the values of the FWHM along each principal axis X', Y' and Z' are [31]:

$$\mathrm{FWHM}_{X'} \simeq \mathrm{FWHM}_{Y'} \simeq \frac{0.7\lambda_{\mathrm{em}}}{\mathrm{NA}}, \tag{45}$$

$$\mathrm{FWHM}_{Z'} \simeq \frac{2.3\lambda_{\mathrm{em}} n_r}{\mathrm{NA}^2}, \tag{46}$$

where λ_{em} is the emission wavelength, n_r is the refractive index of the immersion medium and NA the numerical aperture. The above formula will serve as a reference to check the consistency of GENTLE results.

For our experiment, we imaged a distilled water solution ($n_r = 1.33$) containing fluorescent polystyrene microspheres with a diameter of $\tau = 1\mu$m, marked with a yellow-green fluorophore emitting at 515 nm. The acquisition was performed with an excitation wavelengh $\lambda_{\mathrm{exc}} = 810$ nm, a numerical aperture NA $= 1.05$, and a voxel size of $0.037 \times 0.037 \times 0.05\,\mu\mathrm{m}^3$.

In the obtained image, we were able to select four individual volumes of interest containing isolated beads, using our cropping procedure, on which we applied GENTLE with $\epsilon_1 = \epsilon_2 = 10^{-6}$, so yielding estimates $(\hat{\alpha}, \hat{\beta}, \boldsymbol{H}, \hat{\boldsymbol{D}})$. Then, performing a singular value decomposition of $\hat{\boldsymbol{D}} + \epsilon_1 \mathbf{I}_3$ and using the trigonometry formulas from [28], we

deduce the Euler angles $(\hat{\theta}, \hat{\varphi}) \in [0, \pi] \times [-\pi, \pi]$ and the eigenvalues $\hat{\boldsymbol{s}} = (\hat{s}_{X'}, \hat{s}_{Y'}, \hat{s}_{Z'}) \in (0, +\infty)^3$ along each principal axis (where re-ordering is performed to align the axis with the Cartesian grid (X, Y, Z)) The FWHM along each axis $A \in \{X', Y', Z'\}$ is obtained as $\text{FWHM}_A = 2\sqrt{2\ln(2)/\hat{s}_A}$.

As Figure 3 demonstrates, the estimated FWHM and angles are consistent across the four beads, validating the assumption of a stationary PSF in the observed field. Moreover, the average FWHM values are close to the theoretical expectations computed with (45) and (46), equal to 0.34μm along the axial axes, and 1.43μm in the depth direction, which validates our approach.

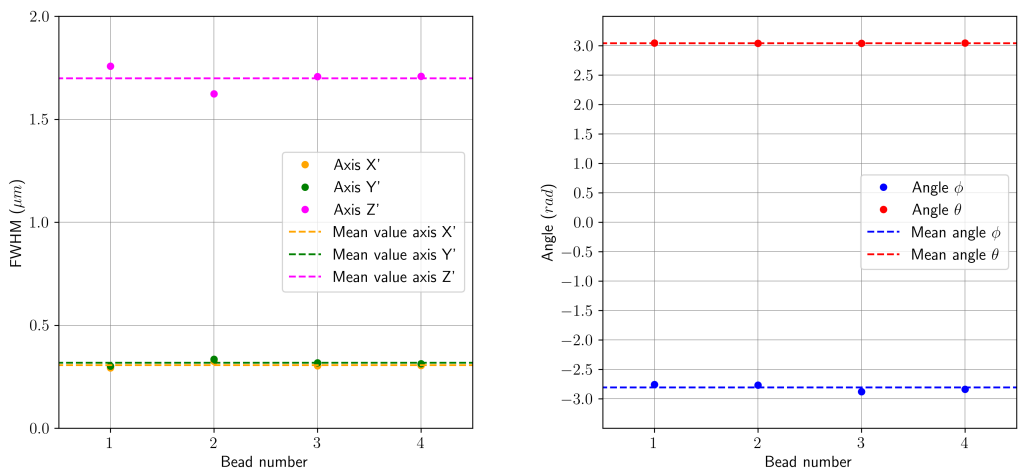

Figure 3. Estimated FWHM (left) in the 3 principal component directions $\{X', Y', Z'\}$, and associated Euler angles (right), for the four isolated beads.

3. Proposed solution for the inverse problem

We now shift our attention to the image restoration phase, where a non-calibrated and potentially complex, object, denoted by $\bar{\boldsymbol{x}} \in \mathbb{R}^M$, is observed, through the multi-photon microscope. The (degraded) observation is denoted $\boldsymbol{y} \in \mathbb{R}^M$. Note that the problem dimension M usually differs from the dimension associated with the prior PSF estimation problem, typically with $M > N$, as the object of interest is typically much more spread than a micrometric bead. The goal is to accurately reconstruct an estimate of $\bar{\boldsymbol{x}} \in \mathbb{R}^M$ given $\boldsymbol{y} \in \mathbb{R}^M$, and our knowledge of the PSF, deduced from the approach from Section 2. Note that, in a practical scenario, the calibration phase could have been performed at another grid resolution, which would require an interpolation step, before using the estimated PSF from GENTLE algorithm. For the sake of simplicity, we opt here for a simple (though common) setting where the PSF $\boldsymbol{h} \in \Delta_N$ has been estimated under the same voxel resolution than the acquisition of the new object. GENTLE method also provides the background parameter, $\alpha \in \mathbb{R}$. The inverse problem thus

reads,

$$y = \mathcal{D}(\boldsymbol{H}\bar{\boldsymbol{x}} + \alpha), \tag{47}$$

where $\mathcal{D}\colon \mathbb{R}^M \to \mathbb{R}^M$ represents the noise model and $\boldsymbol{H} \in \mathbb{R}^{M\times M}$ is the linear operator such that, for every $\boldsymbol{u} \in \mathbb{R}^M$, $\boldsymbol{H}\boldsymbol{u} = \boldsymbol{h} * \boldsymbol{u}$. Here, the convolution product is performed with zero-padding. While the calibration step is usually performed in ideal acquisition conditions, where the Gaussian additive noise assumption is sufficient, the situation differs in the image restoration stage (for example, *in vivo* acquisitions might require low laser power, and as such, leads to low-photon counting). It then becomes necessary to introduce a more realistic noise model, to guarantee high quality restored images. In the following we introduce the considered noise model, we propose a direct method for estimating noise parameters, as well as a variational approach paired with an algorithmic solution to address the inverse problem (47).

3.1. A heteroscedastic noise model for MPM

Many restoration techniques for MPM imaging assume a Gaussian homoscedastic noise (i.e., with constant variance across the whole image) [52, 23, 34]. While such assumption simplifies computational needs, it might not always align with the reality of fluorescence microscopy imaging. The Gaussian assumption generally holds true for high-exposure scenes with abundant photons per the central limit theorem. In contrast, multiphoton or confocal imaging often exhibits a Poisson or Poisson-Gaussian noise profile [22]. In [65], the authors illustrate the benefits of adopting a Poisson noise model over a Gaussian one in terms of restoration quality. However, handling such noise types can be computationally intensive, as it requires minimization of the non-smooth Poisson data-fidelity function [74, 9, 16]. The mixed Poisson-Gaussian noise model describes the corrupted image $\boldsymbol{y} = (y_m)_{1\leq m\leq M}$ as

$$(\forall m \in \{1, \dots, M\}) \quad y_m \sim \alpha + a\mathcal{P}(a^{-1}[\boldsymbol{H}\bar{\boldsymbol{x}}]_m) + \mathcal{N}(0, b), \tag{48}$$

with some noise parameters $a \in (0, +\infty)$ and $b \in (0, +\infty)$. Such noise consists of the mixture of a multiplicative noise with mean equal to the intensity of the pixel and a variance proportional to the intensity of the pixel up to the fixed scale a, and an additive Gaussian noise identically distributed in each pixel, corresponding to a background noise with variance b. However, as shown in [16], the resulting likelihood term is complicated, involving intractable series that need to be approximated, at the price of the computational time.

In the present work, we consider instead an additive heteroscedastic (i.e., with variance depending on the pixel) noise model approximating the Poisson-Gaussian noise, in the line of the method proposed in [38, 37] for low-exposure natural images. To transform the Poisson-Gaussian noise in (48) into a fully additive one, we use the following well-known result: if X is a random variable following a Poisson distribution $\mathcal{P}(\theta)$, with $\theta > 0$, then

$$\frac{1}{\sqrt{\theta}}(X - \theta) \xrightarrow[\theta\to+\infty]{\mathcal{L}} \bar{X}, \tag{49}$$

where $\bar{X} \sim \mathcal{N}(0,1)$ and $\mathcal{L}$ denotes the convergence in distribution. In other words, for large values of θ, the approximation $\mathcal{P}(\theta) \simeq \mathcal{N}(\theta,\theta)$ is valid. Therefore, assuming the multiplicative and background noises are independent, the final model of noise we consider is Gaussian, with a variance varying linearly with the intensity of the pixel, i.e.,

$$(\forall m \in \{1,\dots,M\}) \quad y_m = \alpha + [\boldsymbol{H}\bar{\boldsymbol{x}}]_m + w_m, \tag{50}$$

where $(w_m)_{1\leq m\leq M}$ are independent variables sampled according to

$$(\forall m \in \{1,\dots,M\}) \quad w_m \sim \mathcal{N}(0,\sigma^2([\boldsymbol{H}\bar{\boldsymbol{x}}]_m)), \tag{51}$$

and σ is the function corresponding to the standard deviation, defined as

$$(\forall t \in \mathbb{R}) \quad \sigma(t) = \begin{cases} \sqrt{at+b} & \text{if } t \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{52}$$

The noise model (51) depends on two parameters a and b, for which we present hereafter an estimation strategy.

3.2. Estimation of the noise parameters

We propose in Algorithm 2 our procedure for estimating heteroscedastic Gaussian noise parameters in 3D images. Our approach deviates from that proposed in [38]. In [38], the segmentation into non-overlapping level sets it performed through a wavelet decomposition of the noisy image, which we found too computationally intensive and unnecessary in our MPM context. Moreover, the protected toolbox associated with [38] is not tailored for 3D images.

Algorithm 2: Estimation of parameters a and b

Step 1: 3D image smoothing. Let $\boldsymbol{y} \in \mathbb{R}^M$ the observed 3D image. Convolve $\boldsymbol{y}$ with a normalized uniform kernel of size $s \times s \times s$, with $s \in \mathbb{N}^*$, yielding a smoothed image $\boldsymbol{y}_s$.

Step 2: Volume segmentation.

(i) Let $J \in \mathbb{N}^*$. Segment $\boldsymbol{y}_s$ into J distinct parts using the Lloyd-Max estimator [54] for optimal quantization, so as to minimize the mean squared error between the segmented image and $\boldsymbol{y}_s$.

(ii) Denote $\ell_1 \leq \ell_2 \leq \cdots \leq \ell_J$ the resulting J intensity levels.

(iii) Denote $S_1, S_2, \ldots, S_J$ the segmented parts of the image, corresponding to the intensity intervals $([\ell_{j-1}, \ell_j])_{1 \leq j \leq J}$.

Step 3: Intensity and variance estimation. For each segmented zone $(S_j)_{1 \leq j \leq J}$:

(i) Estimate the intensity $\hat{I}_j$ within S_j as:

$$\hat{I}_j = \frac{1}{|S_j|} \sum_{m \in S_j} y_m, \tag{53}$$

with $|S_j|$ denotes the number of pixels (i.e., the cardinal) in S_j.

(ii) Estimate the variance $\hat{\sigma}_j^2$ within S_j as:

$$\hat{\sigma}_j^2 = \frac{1}{|S_j|} \sum_{m \in S_j} \left(y_m - \hat{I}_j \right)^2. \tag{54}$$

Step 4: Linear regression. Obtain the coefficients (a, b) by conducting a least squares linear regression on the paired values $(\hat{I}_j, \hat{\sigma}_j^2)$ to fit Relation (52).

3.3. A constrained parameter-free deconvolution framework

After the noise modelling step, we move to the resolution of the inverse problem (47). A standard approach is to minimize a cost function composed of a data fidelity term, $f \colon \mathbb{R}^M \longrightarrow \mathbb{R}$, in conjunction with a regularization function, $g \colon \mathbb{R}^M \longrightarrow \mathbb{R}$, which can be expressed as

$$\underset{\boldsymbol{x} \in [0,+\infty)^M}{\text{minimize}} \quad f(\boldsymbol{x}) + \chi g(\boldsymbol{x}), \tag{55}$$

where $\chi > 0$ denotes a regularization parameter. The data fidelity term is designed to ensure that the reconstructed data aligns with the observation model. The regularization function introduces specific *a priori* characteristics to the reconstructed image, such as smoothness. The role of the regularization parameter is to balance the two terms. Its tuning can be based on image quality metrics such as the Mean Square Error or the SNR. However, when the ground truth or reference images are unavailable, this strategy becomes infeasible. Thus, image restoration often involves a time-consuming

adjustment of the regularization parameter, with the final decision largely influenced by the user's expertise and subjective judgment. While various statistical methods exist for estimating this parameter [2, 66], they are usually too computationally intensive for 3D data.

Instead, we propose to tackle the restoration challenge using a constrained approach (sometimes refereed to as the discrepancy principle). This strategy is reminiscent of those described in [10, 1, 42]. The data fidelity term is constrained so that it does not exceed a known (or easily estimated) value. Our formulation reads

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^M}{\text{minimize}} \quad g(\boldsymbol{x}), \\ &\text{subject to} \quad f(\boldsymbol{x}) \leq B \text{ and } \boldsymbol{x} \in [0,+\infty)^M, \end{aligned} \tag{56}$$

where $B > 0$ is a bound, either predetermined or deduced from the data. A main advantage of this formulation is that, by assuming an appropriate choice for the data fidelity function f, it is often feasible to derive a statistical-based upper bound for f, giving a good first trial value for B. It is worth noting that, if f and g are both convex, then for each $B > 0$, there usually exists a corresponding $\chi > 0$ such that (56) and (55) are equivalent. Let us now explicit our choices for f, g, and B.

3.4. Choices for the data-fidelity and regularization functions

Following our noise model (51), we set

$$(\forall \boldsymbol{x} \in \mathbb{R}^M) \quad f(\boldsymbol{x}) = \|\boldsymbol{W}(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y} + \alpha)\|^2. \tag{57}$$

with $\boldsymbol{W} \in \mathcal{S}_M^+$ a weighting matrix accounting for the heterodasticity of the noise. Here, we take

$$\boldsymbol{W} = \text{Diag}\left\{\left(\frac{1}{\sigma([\boldsymbol{H}\bar{\boldsymbol{x}}]_m)}\right)_{1\leq m\leq M}\right\}. \tag{58}$$

In practice, since $\boldsymbol{H}\bar{\boldsymbol{x}}$ is unknown, we approximate it by $\boldsymbol{y}_s$, a denoised version of $\boldsymbol{y}$ as defined in Algorithm 2. Because of the large number of pixels, the law of large numbers makes the following approximation valid: $f(\boldsymbol{x}) \approx M$, which suggests setting $B = M$ in (56).

For the regularization function g, we propose a re-weighted smooth total variation term, adjusted according to the voxel size along the axes X, Y, and Z. Specifically, given a smoothing parameter $\delta > 0$, function g is defined as

$$(\forall \boldsymbol{x} \in \mathbb{R}^M) \quad g(\boldsymbol{x}) = \sum_{m=1}^{M} \sqrt{\delta + \frac{1}{r_X}(\boldsymbol{G}_X\boldsymbol{x})_m^2 + \frac{1}{r_Y}(\boldsymbol{G}_Y\boldsymbol{x})_m^2 + \frac{1}{r_Z}(\boldsymbol{G}_Z\boldsymbol{x})_m^2}, \tag{59}$$

with $\boldsymbol{G} = [\boldsymbol{G}_X^\top \,|\, \boldsymbol{G}_Y^\top \,|\, \boldsymbol{G}_Z^\top]^\top \in \mathbb{R}^{3M\times M}$, where matrices $\boldsymbol{G}_X$, $\boldsymbol{G}_Y$, and $\boldsymbol{G}_Z$ represent discrete gradient operators along the axes X, Y, and Z, respectively.

3.5. Restoration algorithm

The resulting optimization problem (56) reads as the minimization of a smoothed convex function under convex constraints. An effcient Majorization-Minimization strategy [64], entitled P-MMS, was recently proposed in [18] to address such a class of optimization problem. Specifically tailored for large-scale constrained image processing problems such as ours, this algorithm features rapid execution times while providing theoretical convergence guarantees.

The central idea behind P-MMS is to use the external penalty principle to cope with the constraints. Namely, for the constraints $f(\boldsymbol{x}) \leq B$ and $\boldsymbol{x} \in [0, +\infty)^M$, we introduce corresponding penalty functions, denoted by R_1 and R_2:

$$(\forall \boldsymbol{x} \in \mathbb{R}^M) \quad R_1(\boldsymbol{x}) = \mathrm{d}^2_{\mathcal{B}(\boldsymbol{0}_M, B)} \left(\boldsymbol{W}(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y} + \alpha) \right), \tag{60}$$

and

$$(\forall \boldsymbol{x} \in \mathbb{R}^M) \quad R_2(\boldsymbol{x}) = \mathrm{d}^2_{[0,+\infty)^M} \left(\boldsymbol{x} \right), \tag{61}$$

where d_C is the distance to the set $C \subseteq \mathbb{R}^M$. Note that the gradient of R_1 and R_2 are easily computable since the projections onto the ball $\mathcal{B}(\boldsymbol{0}_M, B)$ and the nonnegative orthant $[0, +\infty)^M$ are closed form. A key assumption for P-MMS to be applicable is that every function in the problem admits a tangent quadratic upper bound at every point. In mathematical terms, for a differentiable function $\psi \colon \mathbb{R}^M \longrightarrow \mathbb{R}$, this requirement corresponds to the existence, for every $\boldsymbol{x}' \in \mathbb{R}^M$, of a curvature matrix $\boldsymbol{A}_\psi(\boldsymbol{x}') \in \mathcal{S}_M^+$, such that

$$(\forall \boldsymbol{x} \in \mathbb{R}^M) \quad \psi(\boldsymbol{x}) \leq \psi(\boldsymbol{x}') + \nabla\psi(\boldsymbol{x}')^\top(\boldsymbol{x} - \boldsymbol{x}') + (\boldsymbol{x} - \boldsymbol{x}')^\top \boldsymbol{A}_\psi(\boldsymbol{x}')(\boldsymbol{x} - \boldsymbol{x}'). \tag{62}$$

The existence of such quadratic upper bounds can be readily derived for each function g, R_1, and R_2 in our problem, using the descent Lemma [3, Lemma 2.64] or the half-quadratic majorization formula [15, Lemma 1]. This leads the following valid curvature matrices:

$$\boldsymbol{A}_g(\boldsymbol{x}) = \boldsymbol{G}^\top \mathrm{BDiag} \left\{ \left(\left(\delta + \frac{(\boldsymbol{G}_X \boldsymbol{x})_m^2}{r_X} + \frac{(\boldsymbol{G}_Y \boldsymbol{x})_m^2}{r_Y} + \frac{(\boldsymbol{G}_Z \boldsymbol{x})_m^2}{r_Z} \right)^{-1/2} \boldsymbol{I}_3 \right)_{1 \leq m \leq M} \right\} \boldsymbol{G}, \tag{63}$$

$$\boldsymbol{A}_{R_1}(\boldsymbol{x}) = \boldsymbol{H}^\top \boldsymbol{W}^\top \boldsymbol{W} \boldsymbol{H}, \quad \text{and} \quad \boldsymbol{A}_{R_2}(\boldsymbol{x}) = 2\boldsymbol{I}_M, \tag{64}$$

where BDiag stand for block diagonal matrix. Then, for any penalty parameter $\gamma > 0$, the penalized function

$$F_\gamma(\boldsymbol{x}) = g(\boldsymbol{x}) + \gamma(R_1(\boldsymbol{x}) + R_2(\boldsymbol{x})) \tag{65}$$

admits a quadratic tangent majorant at $\boldsymbol{x}$ uniquely defined by its curvature matrix:

$$\boldsymbol{A}_{F_\gamma}(\boldsymbol{x}) = \boldsymbol{A}_g(\boldsymbol{x}) + \gamma(\boldsymbol{A}_{R_1}(\boldsymbol{x}) + \boldsymbol{A}_{R_2}(\boldsymbol{x})). \tag{66}$$

We present in Algorithm 3 the P-MMS approach, to solve Problem (56). At each iteration of the inner algorithm, $\mathbf{x}_{k+1}$ is updated within the affine space defined by the directions $\{-\nabla F_{\gamma_j}(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}_{k-1}\}$, employing a Majorization-Minimization approach to determine the multidimensional step-size. A local version P-MMS^{loc} of this algorithm was developed which implements a trust-region strategy to accelerate further convergence. We redirect readers to [18] or to its publicly available code, for more details. The P-MMS algorithm benefits from the convergence guarantees given in Theorem 2.

Algorithm 3: P-MMS

Inputs: $(\gamma_j)_{j\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$, $(\varepsilon_j)_{j\in\mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$, $\boldsymbol{x}_0 \in \mathbb{R}^M$.
for $j = 0, 1, \ldots$ **do**
  `// find an (approximated) minimizer of` F_{γ_j}
  Set initial point $\mathbf{x}_0$,
  for $k = 1, \ldots$ **do**
    Construct subspace directions $\boldsymbol{D}_k = [-\nabla F_{\gamma_j}(\mathbf{x}_k), \mathbf{x}_k - \mathbf{x}_{k-1}]$,
    Compute $\boldsymbol{A}_{F_{\gamma_j}}(\mathbf{x}_k)$ according to (66),
    $\boldsymbol{u}_k = -\left[\boldsymbol{A}_{F_{\gamma_j}}(\mathbf{x}_k)\right]^{\dagger} \boldsymbol{D}_k^{\top} \nabla F_{\gamma_j}(\mathbf{x}_k)$,
    $\mathbf{x}_{k+1} = \mathbf{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k$,
    if $\|\nabla F_{\gamma_j}(\mathbf{x}_{k+1})\| < \varepsilon_j$ **then**
      exit loop `// stop inner algorithm if given precision` ε_j `on the norm of the gradient is reached`
      return $\mathbf{x}_{k+1}$.
  $\boldsymbol{x}_j = \mathbf{x}_{k+1}$.
return $\boldsymbol{x}_j$.

Theorem 2 (Convergence of the P-MMS algorithm)**.** *Assume the sequence of parameters $(\varepsilon_j)_{j\in\mathbb{N}}$ satisfies, for every $j \in \mathbb{N}$, $\varepsilon_j > 0$ and $\lim_{j\to+\infty} \varepsilon_j = 0$. Also assume that $(\gamma_j)_{j\in\mathbb{N}}$ is a nondecreasing sequence of positive reals and $\lim_{j\to+\infty} \gamma_j = +\infty$. Then, the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{N}}$ generated by Algorithm 3 is bounded and any of its cluster point is a solution to Problem* (56)*.*

Proof. It suffices to check that the assumptions for [18, Theorem 2] are satisfied.

- The functions g, R_1 and R_2 are differentiable,
- For every $\gamma > 0$, the function F_γ is coercive, convex, and satisfies the KŁ property.
- For every $\gamma > 0$, the curvature function defined in (66) is lower bounded independently of $\boldsymbol{x}$ and is continuous.

This concludes the proof. □

3.6. Validation of the restoration method on simulated data

To illustrate the performance of our heteroscedastic constrained formulation for the MPM image restoration tast, we first conducted an experiment using simulated data.

For this experiment, we chose an image of a fly brain from [12] as the object of interest $\bar{\boldsymbol{x}}$. This image, with dimensions $M = 128 \times 128 \times 40$, was artificially degraded through a convolution operator $\boldsymbol{H}$ mimicking the effect of a normalized 3D Gaussian kernel with inverse covariance matrix parameterized by the angles $\theta = 5\pi/6$ rad, $\varphi = 0$ rad, and eigenvalues $\boldsymbol{s} = (50, 50, 50)$. Moreover, heteroscedastic noise was introduced based on the noise model presented in (51), with $a = 0.01$ and $b = 10^{-5}$. The voxels dimensions were configured to $0.05 \times 0.05 \times 0.05\mu\text{m}^3$.

To provide a comprehensive evaluation, we compare our approach against the more traditional regularized approach in (55), where we choose the data-fidelity function f to fit homoscedastic Gaussian noise, setting $f(\boldsymbol{x}) = \|\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y} + \alpha\|^2$, and the smoothed TV regularization function g defined in (59).

For both models, the restoration was performed running the P-MMS algorithm. The parameter δ was set to 0.1 and, for the constrained problem, the bound B to the number of pixels, i.e., $B = M$. The penalty parameters $(\gamma_j)_{j\in\mathbb{N}}$ in Algorithm 3 were set to $\gamma_j = (2j)^2$ and the precision parameters $(\varepsilon_j)_{j\in\mathbb{N}}$ were chosen as $\varepsilon_j = 10^5/(\gamma_j)^{0.75}$. The algorithm was initialized with $\boldsymbol{x}_0 = \boldsymbol{y}$.

Figure 5 displays the SNR as a function of the iteration number of the P-MMS algorithm for the constrained formulation, confirming the theoretically convergence established in Proposition 2, and illustrating the fast stability after few dozens of iterations only. Figure 4 captures the performance of the two methodologies. It showcases the SNR of the restored images using the standard regularized approach as a function of the regularization parameter χ, as well as the SNR achieved using our parameter-free method. A notable observation is the important fluctuation in SNR values for the regularized method as χ gets closer to its optimal value. It is crucial to highlight that this optimal value of χ remains unknown in the real-world context of Multiphoton Microscopy (MPM), as the SNR metric is not available. In contrast, our constrained methodology consistently delivers superior SNR, underscoring the interest of a reliable noise model. The absence of any tuning parameter is furthermore a crucial advantage in real-life applications.

4. Application of the proposed pipeline to muscle tissue imaging

In this section, we apply the entire pipeline, encompassing both PSF estimation and deconvolution, to the restoration of a mouse muscle volume, characterized by its prominent myosin filaments, from MPM acquisitions. All mice were bred and housed in Limoges University's animal facility under controlled conditions (20°C, 12 hours light/12 hours dark cycle) with free access to standard mouse chow and tap water. Experimental procedures were carried out in accordance with the European

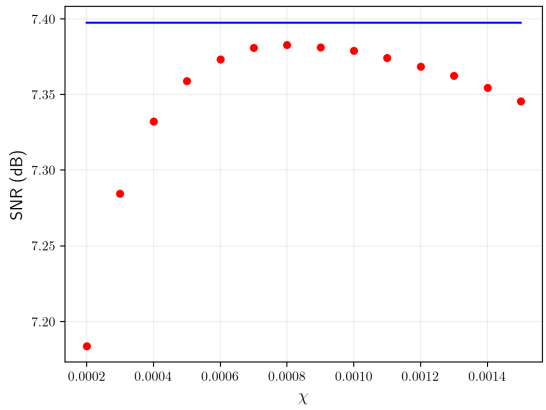

Figure 4. Comparison of the penalized least-squares approach (red dots) with the proposed constrained approach assuming a Gaussian heteroscedastic noise model (blue line), in terms of restored image SNR, in dB.

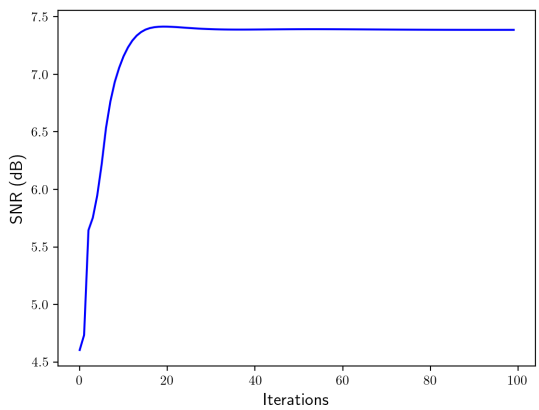

Figure 5. SNR evolution along the iterations j in Algorithm 3 for the resolution of the proposed constrained formulation.

legislation on animal experimentation (Directive 2010/63/UE), and approved by Ethical Committee no.033 (APAFIS #1903-2015091612088147 v2). Phenotypic and molecular analyses were performed on 12-week-old male animals.

4.1. Presentation of the experimental setup

A mouse muscle sample was immersed in a water solution. As a reference for the PSF estimation phase, polystyrene microspheres of 1μm diameter were integrated along the muscle's perimeter. We employed a bi-channel raw acquisition technique, combining both the Second Harmonic Generation signal from the myosin and the two-photon fluorescence from the microspheres. Thus, the PSF is calibrated in the same field of view, and under the same resolution, than the one where we observe the myosin organization. Such experimental setting is somewhat ideal, as the conditions for recording the PSF

match precisely those for capturing the image of the muscle sample. Given that the muscle assembly acts as an absorbing and scattering medium that distorts the optical wavefront, recording the PSF in the location where the image is restored enhances restoration accuracy. The microscopy acquisition was carried out with an excitation wavelength of 810nm. We used the same device than in our previous experiments, that is a water immersion objective tailored for multiphoton acquisitions, specifically the Olympus XLPLN25XWMP with a numerical aperture NA = 1.05. The selected voxel dimensions for the imaging process were $0.049 \times 0.049 \times 0.1\mu\text{m}^3$. The resulting volume was of dimension $M = 840 \times 840 \times 180$.

4.2. Results for PSF calibration step

We first conducted the PSF estimation on the channel of the acquired image that contains the beads, that is we perform an *in situ* PSF calibration within an heterogeneous medium. Given these conditions, we anticipate a notably higher dispersion compared to the optimal setting discussed in Section 2.5.2. We fitted a convolution kernel on 8 cropped volumes, containing isolated bead images, using GENTLE (Algorithm 1). The hyper-parameters of the PSF estimation model (6) were set identically to Section 2.5. Figure 6 presents a comparison of the estimated FWHMs along the principal axis and Euler angles for the 8 beads, using the GENTLE method. Estimations across different beads again exhibit consistency, confirming the PSF stationarity assumption in this experiment. In addition, we display in Figure 7 the intensity profiles of the observed beads and of their reconstitution $\hat{\alpha} + \hat{\beta}\hat{\boldsymbol{h}} * \boldsymbol{x}$, where $\boldsymbol{x}$ is the theoretical bead. The GENTLE method accurately fits the observation across all three axes and, in particular, fits asymmetrical shapes more effectively than NLS, thanks to its robust formulation.

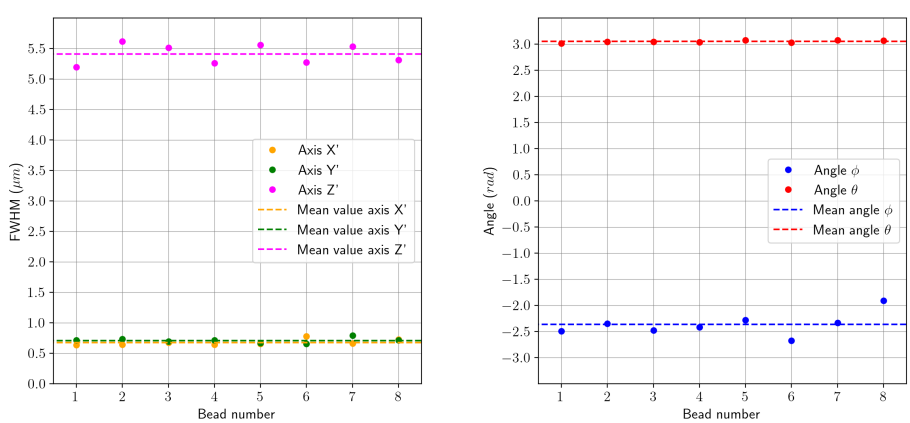

Figure 6. Estimated FWHM (left) in the 3 principal component directions denoted by X', Y' and Z', and angles (right) for the eight isolated beads.

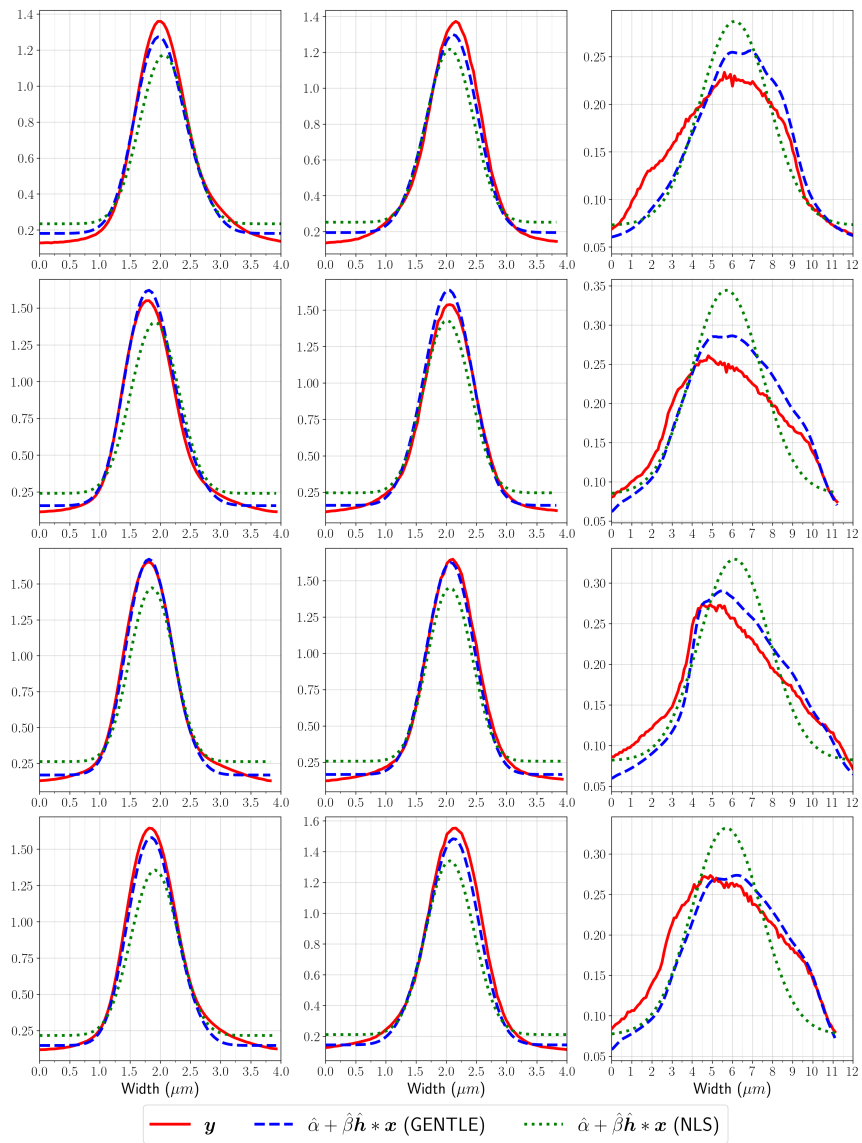

Figure 7. Profiles along axis X (left), Y (center) and Z (right) for 4 of the 8 isolated beads (rows). We compare the intensity profiles of the observation $\boldsymbol{y}$ (red), the reconstitution $\hat{\alpha} + \hat{\beta}\hat{\boldsymbol{h}} * \boldsymbol{x}$ obtained with GENTLE (blue) and the reconstitution obtained with NLS (green). The profiles for the remaining 4 beads, omitted for space reason, are consistent with the displayed ones.

4.3. Results for the restoration step

In the restoration phase, we first determined the noise parameters a and b across the entire volume using the methodology described in Algorithm 2 with $s = 5$ and $J = 25$. Figure 8 displays the linear regression corresponding to the fourth step of the algorithm. The close alignment of the points confirms the appropriateness of the heteroscedastic noise model proposed.

Subsequently, the standard deviations $(\sigma([\boldsymbol{H}\bar{\boldsymbol{x}}]_m))_{1 \leq m \leq M}$ were approximated using (52) with the estimated a, b, and a smoothed version of the image $\boldsymbol{H}\bar{\boldsymbol{x}}$ corresponding to

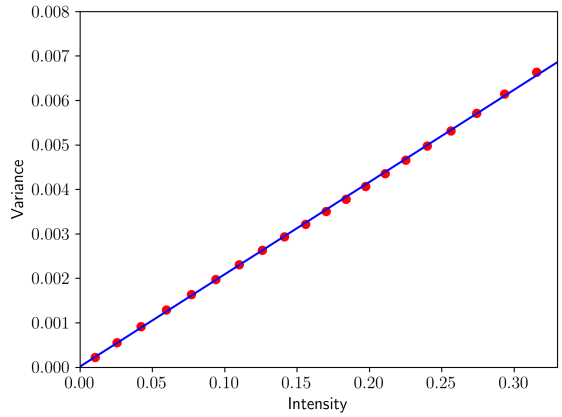

Figure 8. Linear regression (blue line) performed on the couples of values $(\hat{I}_j, \hat{\sigma}_j^2)$ (red dots) for the estimation of noise parameters $(a, b) \in \mathbb{R}^2$ on the mouse muscle 3D image.

the convolution of $\boldsymbol{y}$ with a $3 \times 3 \times 3$ uniform kernel. We solved Problem (56) by setting $B = M$ and $\delta = 0.1$, using Algorithm 3. The penalty parameters $(\gamma_j)_{j \in \mathbb{N}}$ were set to $\gamma_j = (1.5j)^{1.2}$ and the precision parameters $(\varepsilon_j)_{j \in \mathbb{N}}$ were chosen as $\varepsilon_j = 10^5 / (\gamma_j)^{0.5}$.

In Figure 9, we present slices from a portion of the raw image alongside corresponding slices from the restored image. The visual improvements are evident both in term of denoising and deblurring. In Figure 9(b) and (d), the presence of myofibrils appears clearly in the optical plan (XY) and in the optical axis (YZ). The myofiber structure is following the (OY) axis.

5. Conclusion

In this paper, we introduced an end-to-end approach for addressing the MPM image restoration problem. We decomposed the primary problem into two sub-problems: PSF estimation and image deconvolution, and for both, we proposed original formulation and optimization methods. Our results on simulated data, real beads imaged under optimal conditions, and mouse muscle samples, demonstrate the efficiency of our approach.

As highlighted in the paper, our method uses different noise models in its two distinct phases: initially a Gaussian model, and then a Poisson-Gaussian model. While the current setup has proven effective, considering the same heteroscedastic noise model throughout the methodology could be a promising next step to improve further the results.

Additionally, a future direction for this research could involve extending the PSF estimation strategy to other objects than beads. Given the generality of the method, it has potential applicability to any reference object with known geometrical structure.

On the restoration side, there is potential to investigate advanced regularization techniques to enhance visual outcomes as our framework is versatile, and not restricted to TV-based regularization. Lastly, the heteroscedastic noise model we proposed for

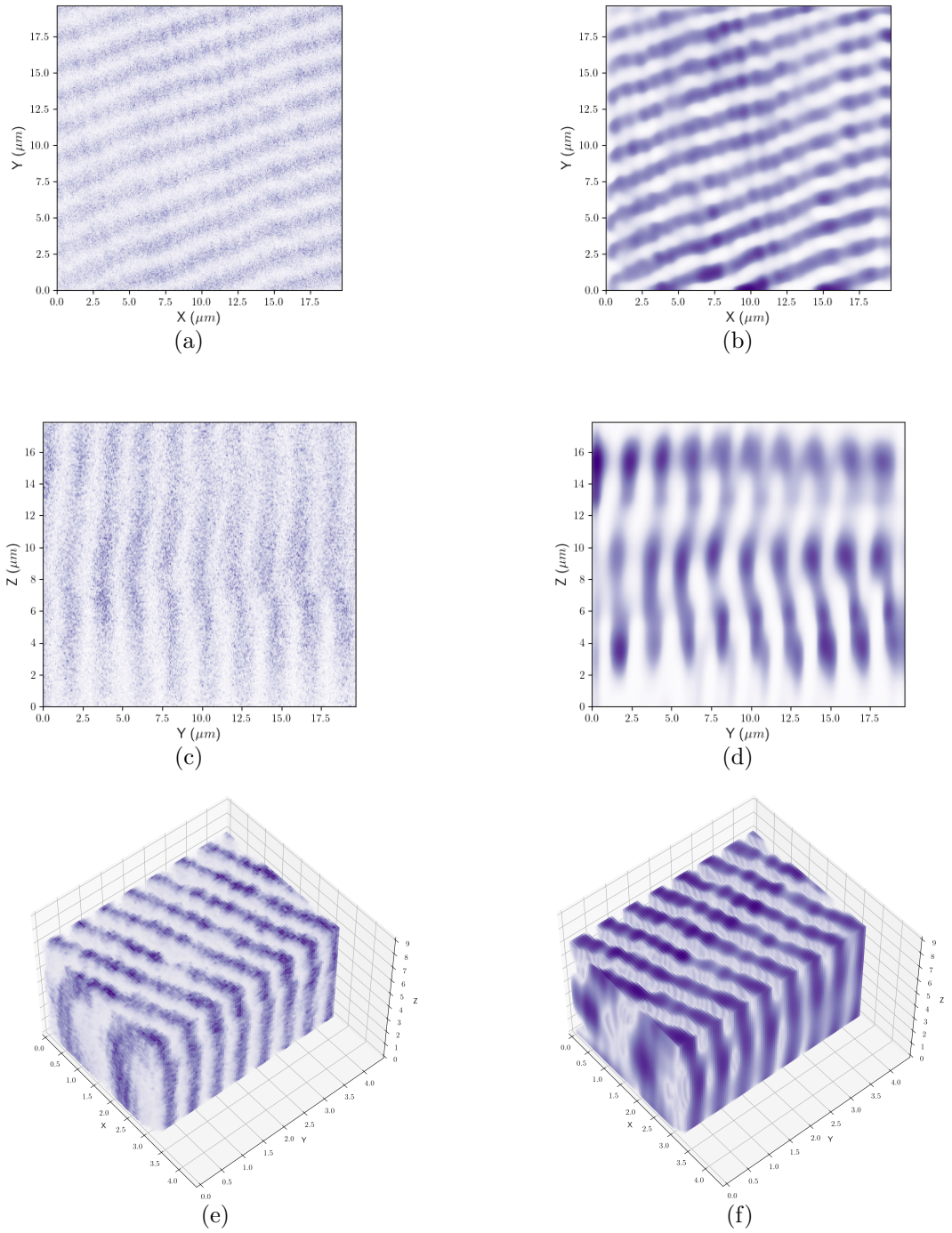

Figure 9. Raw image (left column) and restored image (right column) of a mouse muscle sample. For both 3D images, we display the same slice in the plane XY (first row) and YZ (second row), as well as a 3D visualization (third row).

MPM deserves deeper exploration. A study analyzing how parameters a and b fluctuate based on different optical settings and depth in the sample could provide insights on the physical properties of the noise.

Acknowledgements

This work was supported by the European Research Council Starting Grant MAJORIS ERC-2019-STG-850925 and by the ANR Research and Teaching Chair BRIDGEABLE in Artificial Intelligence.

The authors would like to thank Véronique Blanquet, Laetitia Magnol et Alexis Parenté for preparing the mouse muscle samples with microbeads.

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2010.

[2] J. Antoni, J. Idier, and S. Bourguignon. A Bayesian interpretation of the L-curve. *Inverse Problems*, 39(6):065016, 2023.

[3] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer New York, 2nd, corrected printing edition, 2019.

[4] A. Beck and L. Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM Journal on Optimization*, 23(4):2037–2060, 2013.

[5] D. Bertsekas. *Nonlinear Programming*. Athena Scientific Optimization and Computation Series. Athena Scientific, 1995.

[6] P. Bohra, D. Garg, K. S. Gurumoorthy, and A. Rajwade. Variance-stabilization-based compressive inversion under Poisson or Poisson–Gaussian noise with analytical bounds. *Inverse Problems*, 35(10):105006, 2019.

[7] J. Bolte, A. Daniilidis, and A. Lewis. The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[8] S. Bonettini, M. Prato, and S. Rebegoldi. A block coordinate variable metric linesearch based proximal gradient method. *Computational Optimization and Application*, 71(1):5–52, 2018.

[9] M. Carlavan and L. Blanc-Féraud. Sparse Poisson noisy image deblurring. *IEEE Transactions on Image Processing*, 21(4):1834–1846, 2011.

[10] R. E. Carrillo, J. D. McEwen, and Y. Wiaux. Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 426(2):1223–1234, oct 2012.

[11] V. E. Centonze and J. G. White. Multiphoton excitation provides optical sections from deeper within scattering specimens than confocal imaging. *Biophysical Journal*, 75(4):2015–2024, 1998.

[12] M. Chalvidal, E. Chouzenoux, J.-B. Fest, and C. Lefort. Block delayed Majorize-Minimize subspace algorithm for large scale image restoration. *Inverse Problems*, 39(4):044002, 2023.

[13] A. Cherni, É. Chouzenoux, and M.-A. Delsuc. Proximity operators for a class of hybrid sparsity + entropy priors application to dosy NMR signal reconstruction. In *2016 International Symposium on Signal, Image, Video and Communications (ISIVC)*, pages 120–125, Tunis, Tunisia, Nov 2016. IEEE.

[14] E. Chouzenoux, M.-C. Corbineau, J.-C. Pesquet, and G. Scrivanti. A variational approach for joint image recovery and feature extraction based on spatially varying generalised Gaussian models. https://arxiv.org/abs/2209.01375, Oct 2023.

[15] É. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A Majorize-Minimize subspace approach for $\ell_2 - \ell_0$ image regularization. *SIAM Journal on Imaging Sciences*, 6:563–591, 2013.

[16] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot. A convex approach for image restoration with exact Poisson–Gaussian likelihood. *SIAM Journal on Imaging Sciences*, 8(4):2662–2682, 2015.

[17] E. Chouzenoux, T. T.-K. Lau, C. Lefort, and J.-C. Pesquet. Optimal multivariate Gaussian fitting with applications to PSF modeling in two-photon microscopy imaging. *Journal of Mathematical Imaging and Vision*, 61(7):1037–1050, 2019.

[18] E. Chouzenoux, S. Martin, and J.-C. Pesquet. A local MM subspace method for solving constrained variational problems in image recovery. *Journal of Mathematical Imaging and Vision*, 08 2022.

[19] E. Chouzenoux, J.-C. Pesquet, and A. Repetti. A block coordinate variable metric forward–backward algorithm. *Journal of Global Optimization*, 66(3):457–485, 2016.

[20] R. W. Cole, T. Jinadasa, and C. M. Brown. Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control. *Nature Protocols*, 6(12):1929–1941, 2011.

[21] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5:329–359, 1996.

[22] M. Crivaro, H. Enjieu-Kadji, R. Hatanaka, S. Nakauchi, J. Bosch, J. Judin, J. Riera, and R. Kawashima. Multiphoton fluorescent images with a spatially varying background signal: a Ml deconvolution method. *Journal of Microscopy*, 242(3):311–324, 2011.

[23] A. Danielyan, Y.-W. Wu, P.-Y. Shih, Y. Dembitskaya, and A. Semyanov. Denoising of two-photon fluorescence images with block-matching 3D filtering. *Methods*, 68(2):308–316, 2014.

[24] L. Dao, B. Glancy, B. Lucotte, L.-C. Chang, R. S. Balaban, and L.-Y. HSU. A model-based approach for microvasculature structure distortion correction in two-photon fluorescence microscopy images. *Journal of Microscopy*, 260(2):180–193, 2015.

[25] J. B. De Monvel, E. Scarfone, S. Le Calvez, and M. Ulfendahl. Image-adaptive deconvolution for three-dimensional deep biological imaging. *Biophysical Journal*, 85(6):3991–4001, 2003.

[26] M. De Moraes Marim, B. Zhang, J.-C. Olivo-Marin, and C. Zimmer. Improving single particle localization with an empirically calibrated Gaussian kernel. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1003–1006. IEEE, 2008.

[27] V. Debarnot and P. Weiss. Deepblur: Blind identification of space variant PSF. In *Proceedings of the IEEE 18th International Symposium on Biomedical Imaging (ISBI 2021)*, 2021.

[28] D. Depriester. Computing Euler angles with Bunge convention from rotation matrix. 10.13140/RG.2.2.34498.48321/5, March 2018.

[29] A. Descloux, L. V. Amitonova, and P. W. H. Pinkse. Aberrations of the point spread function of a multimode fiber due to partial mode excitation. *Optics Express*, 24(16):18501–18512, 2016.

[30] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia. Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution. *Microscopy Research and Technique*, 69(4):260–266, 2006.

[31] A. Diaspro. *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances*. Alberto Diaspro, 2001.

[32] A. Diaspro, P. Bianchini, G. Vicidomini, M. Faretta, P. Ramoino, and C. Usai. Multi-photon excitation microscopy. *Biomedical Engineering Online*, 5:1–14, 2006.

[33] F. Difato, F. Mazzone, S. Scaglione, M. Fato, F. Beltrame, L. Kubínová, J. Janáček, P. Ramoino, G. Vicidomini, and A. Diaspro. Improvement in volume estimation from confocal sections after image deconvolution. *Microscopy Research and Technique*, 64(2):151–155, 2004.

[34] A. Doi, R. Oketani, Y. Nawa, and K. Fujita. High-resolution imaging in two-photon excitation microscopy using in situ estimations of the point spread function. *Biomedical Optics Express*, 9(1):202–213, 2018.

[35] C.-Y. Dong, K. Koenig, and P. So. Characterizing point spread functions of two-photon fluorescence microscopy in turbid medium. *Journal of Biomedical Optics*, 8(3):450–459, 2003.

[36] E. Dusch, T. Dorval, N. Vincent, M. Wachsmuth, and A. Genovesio. Three-dimensional point spread function model for line-scanning confocal microscope with high-aperture objective. *Journal of Microscopy*, 228(2):132–138, 2007.

[37] A. Foi. Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, 89(12):2609–2629, 2009.

[38] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.

[39] D. Fortun, P. Guichard, V. Hamel, C. O. S. Sorzano, N. Banterle, P. Gönczy, and M. Unser. Reconstruction from multiple particles for 3D isotropic resolution in fluorescence microscopy. *IEEE Transactions on Medical Imaging*, 37(5):1235–1246, 2018.

[40] W. Göbel, B. M. Kampa, and F. Helmchen. Imaging cellular network dynamics in three dimensions using fast 3D laser scanning. *Nature Methods*, 4(1):73–79, 2007.

[41] H. Guo. A simple algorithm for fitting a Gaussian function [DSP tips and tricks]. *IEEE Signal Processing Magazine*, 28(5):134–137, 2011.

[42] S. Harizanov, J.-C. Pesquet, and G. Steidl. Epigraphical projection for solving least squares Anscombe transformed constrained optimization problems. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 125–136, Berlin, Heidelberg, June 2013. Springer.

[43] J. A. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement*, 15:417, 1974.

[44] Y. Huang, E. Chouzenoux, and V. Elvira. Probabilistic modeling and inference for sequential space-varying blur identification. *IEEE Transactions on Computational Imaging*, 7:531–546, 2021.

[45] Y. Huang, E. Chouzenoux, and J.-C. Pesquet. Unrolled variational Bayesian algorithm for image blind deconvolution. *IEEE Transactions on Image Processing*, 32:430–445, 2022.

[46] M. Jonard, Y. Leventoux, M. Colas, S. Février, T. Mansuryan, J. Cornette, A. Tonello, K. Krupa, J.-R. Duclère, and C. Lefort. 3D multiphoton characterization of χ^2 nonlinearity induced in a multimode fiber through optical poling. In *2022 Conference on Lasers and Electro-Optics (CLEO)*, pages 1–2. IEEE, 2022.

[47] H. Kirshner, F. Aguet, D. Sage, and M. Unser. 3D PSF fitting for fluorescence microscopy: implementation and localization application. *Journal of Microscopy*, 249(1):13–25, 2013.

[48] J. E. Krist, R. N. Hook, and F. Stoehr. 20 years of Hubble space telescope optical modeling using Tiny Tim. In *Optical Modeling and Performance Predictions V*, volume 8127, pages 166–181. SPIE, 2011.

[49] K. Kurdyka. On gradients of functions definable in o-minimal structures. In *Annales de l'Institut Fourier*, volume 48, pages 769–783, 1998.

[50] M. Lapin, M. Hein, and B. Schiele. Analysis and optimization of loss functions for multiclass, top-k, and multilabel classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(7):1533–1554, 2017.

[51] A. M. Larson. Multiphoton microscopy. *Nature Photonics*, 5(1):1–1, 2011.

[52] C. Lefort, M. Chalvidal, A. Parenté, V. Blanquet, H. Massias, L. Magnol, and E. Chouzenoux. FAMOUS: a fast instrumental and computational pipeline for multiphoton microscopy applied to 3D imaging of muscle ultrastructure. *Journal of Physics D: Applied Physics*, 54(27):274005, 2021.

[53] J. Li, F. Xue, and T. Blu. Accurate 3D PSF estimation from a wide-field microscopy image. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 501–504. IEEE, 2018.

[54] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[55] Z.-Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals of Operations Research*, 46(1):157–178, 1993.

[56] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[57] S.-Y. Niu, L.-Z. Guo, Y. Li, Z. Zhang, T.-D. Wang, K.-C. Liu, Y.-J. Li, Y. Tsao, and T.-M. Liu. Boundary-preserved deep denoising of stochastic resonance enhanced multiphoton images. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–12, 2022.

[58] D. N. Phan and N. Gillis. An inertial block Majorization Minimization framework for nonsmooth nonconvex optimization. *Journal of Machine Learning Research*, 24:1–41, 2023.

[59] M. Prato, A. La Camera, S. Bonettini, and M. Bertero. A convergent blind deconvolution method for post-adaptive-optics astronomical imaging. *Inverse Problems*, 29(6):065017, 2013.

[60] D. K. Samuylov, P. Purwar, G. Székely, and G. Paul. Modeling Point Spread Function in fluorescence microscopy with a sparse Gaussian mixture: Tradeoff between accuracy and efficiency. *IEEE Transactions on Image Processing*, 28(8):3688–3702, 2019.

[61] P. Sarder and A. Nehorai. Deconvolution methods for 3D fluorescence microscopy images. *IEEE Signal Processing Magazine*, 23(3):32–45, 2006.

[62] S. Stallinga and B. Rieger. Accuracy of the Gaussian point spread function model in 2D localization microscopy. *Optics Express*, 18(24):24461–24476, 2010.

[63] F. Ströhl and C. F. Kaminski. A joint Richardson-Lucy deconvolution algorithm for the reconstruction of multifocal structured illumination microscopy data. *Methods and Applications in Fluorescence*, 3(1):014002, 2015.

[64] Y. Sun, P. Babu, and D. Palomar. Majorization-Minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, 65:794–816, Feb 2017.

[65] G. M. Van Kempen, L. J. van Vliet, P. J. Verveer, and H. T. van der Voort. A quantitative comparison of image restoration methods for confocal microscopy. *Journal of Microscopy*, 185(3):354–365, 1997.

[66] S. Vatankhah, R. A. Renaut, and V. E. Ardestani. Regularization parameter estimation for underdetermined problems by the χ^2 principle with application to 2D focusing gravity inversion. *Inverse Problems*, 30(8):085002, 2014.

[67] G. Vicidomini, P. Boccacci, A. Diaspro, and M. Bertero. Application of the split-gradient method to 3D image deconvolution in fluorescence microscopy. *Journal of Microscopy*, 234(1):47–61, 2009.

[68] M. Von Tiedemann, A. Fridberger, M. Ulfendahl, and J. B. De Monvel. Image adaptive point-spread function estimation and deconvolution for in vivo confocal microscopy. *Microscopy Research and Technique*, 69(1):10–20, 2006.

[69] S. V. Vorontsov and S. M. Jefferies. A new approach to blind deconvolution of astronomical images. *Inverse Problems*, 33(5):055004, 2017.

[70] K. Wu, E. Otoo, and A. Shoshani. Optimizing connected component labeling algorithms. In *Medical Imaging 2005: Image Processing*, volume 5747, pages 1965–1976. SPIE, 2005.

[71] Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6(3):1758–1789, 2013.

[72] P. Young, S. Clendenon, J. Byars, R. Decca, and K. Dunn. The effects of spherical aberration on multiphoton fluorescence excitation microscopy. *Journal of Microscopy*, 242(2):157–165, 2011.

[73] P. Young, S. Clendenon, J. Byars, and K. Dunn. The effects of refractive index heterogeneity within kidney tissue on multiphoton fluorescence excitation microscopy. *Journal of Microscopy*, 242(2):148–156, 2011.

[74] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero. Efficient gradient projection methods for edge-preserving removal of Poisson noise. *Inverse Problems*, 25(4):045010, 2009.

[75] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10):1819–1829, 2007.

[76] X. Zhu and D. Zhang. Efficient parallel Levenberg-Marquardt model fitting towards real-time automated parametric imaging microscopy. *PLOS One*, 8(10):e76665, 2013.

[77] W. R. Zipfel, R. M. Williams, and W. W. Webb. Nonlinear magic: multiphoton microscopy in the biosciences. *Nature Biotechnology*, 21(11):1369–1377, 2003.