Journal of Mathematical Imaging and Vision (2022) 64:968–992
https://doi.org/10.1007/s10851-022-01100-3

Image Reconstruction in Light-Sheet Microscopy: Spatially Varying Deconvolution and Mixed Noise

Bogdan Toader[1,3,4] · Jérôme Boulanger[2] · Yury Korolev[3] · Martin O. Lenz[1,5] · James Manton[2] · Carola-Bibiane Schönlieb[3] · Leila Mureşan[1,4,5]

Received: 8 August 2021 / Accepted: 23 April 2022 / Published online: 14 June 2022
© The Author(s) 2022

Abstract
We study the problem of deconvolution for light-sheet microscopy, where the data is corrupted by spatially varying blur and a combination of Poisson and Gaussian noise. The spatial variation of the point spread function of a light-sheet microscope is determined by the interaction between the excitation sheet and the detection objective PSF. We introduce a model of the image formation process that incorporates this interaction and we formulate a variational model that accounts for the combination of Poisson and Gaussian noise through a data fidelity term consisting of the infimal convolution of the single noise fidelities, first introduced in L. Calatroni et al. (SIAM J Imaging Sci 10(3):1196–1233, 2017). We establish convergence rates and a discrepancy principle for the infimal convolution fidelity and the inverse problem is solved by applying the primal–dual hybrid gradient (PDHG) algorithm in a novel way. Numerical experiments performed on simulated and real data show superior reconstruction results in comparison with other methods.

Keywords Deconvolution · Light-sheet microscopy · Poisson and Gaussian noise · Primal–dual hybrid gradient · Numerical methods

1 Introduction

Light-sheet microscopy is a fluorescence microscopy technique that enables volumetric imaging of biological samples at high frame rate with better sectioning and lower photo-toxicity in comparison with other fluorescent techniques. This is achieved by illuminating a thin slice of the sample using a sheet of light and detecting the emitted fluorescence from this plane with another objective perpendicular to the plane of the sheet. A schematic representation of a light-sheet microscope is shown in Fig. 1. Other microscopy techniques present certain disadvantages. For example, wide-field microscopy [1] illuminates the whole sample using a single objective and achieves only very limited sectioning, while confocal microscopy [1] allows improved sectioning by utilising a pinhole to discard out-of-focus light, at the cost of higher photo-toxicity and reduced frame rate. Light-sheet microscopy avoids these downsides by only selectively illuminating the slice of the sample being imaged. In this way, less photo-toxicity damage is induced and, therefore, imaging of living samples over a longer period of time is possible. The combination of lower photo-toxicity, better sectioning capabilities and faster image acquisition led to light-sheet microscopy being recognised as "Method of the Year" by Nature Methods in 2014 [2].

The focus of the present manuscript is on deconvolution techniques for light-sheet microscopy data. In this context, deconvolution refers to the computational method of reversing the effect of blurring in the image acquisition process due to the point spread function (PSF) of the microscope [3–5]. Specifically, the PSF of an imaging system repre-

✉ Bogdan Toader
bt382@cam.ac.uk

1 Cambridge Advanced Imaging Centre, University of Cambridge, Anatomy School, Downing Street, Cambridge CB2 3DY, UK

2 MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

3 Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

4 Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK

5 Sainsbury Laboratory, University of Cambridge, 47 Bateman Street, Cambridge CB2 1LR, UK

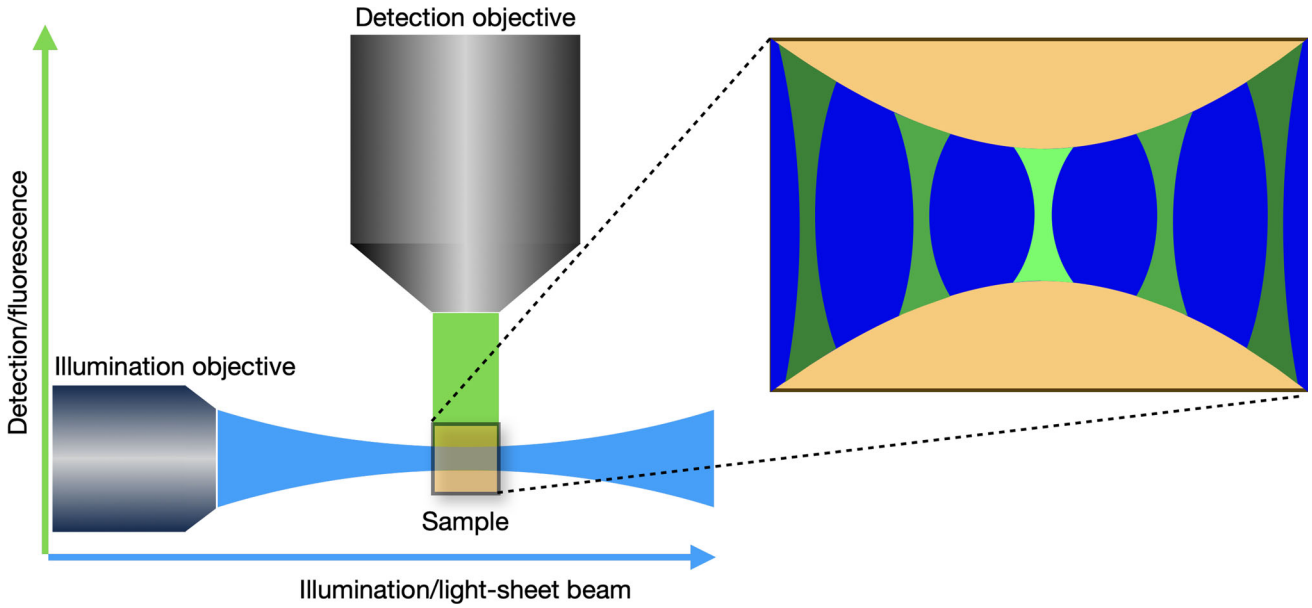


Fig. 1 Schematic of a light-sheet microscope, showing the illumination and the detection directions. The interaction of the light-sheet with the detection PSF leads to a spatially varying overall PSF and decreasing of the pixel intensities away from the centre in the horizontal direction

sents its response to a point object. In general, knowledge of the PSF can be modelled mathematically and calibrated using bead data (samples containing small spheres of known dimensions), and then used in the formulation of a forward model of the image formation, which can then be inverted, for example using optimisation methods, to reconstruct the original, deblurred object [6].

However, in the case of light-sheet microscopy, simply knowing or estimating the PSF of the detection objective is not sufficient, since the overall response of the system to a point source is also influenced by the excitation light-sheet used to illuminate the slice. The overall PSF could be approximated by the detection PSF in the region where the illumination sheet is focused. However, the detection PSF becomes more distorted and loses intensity away from the focus of the excitation light-sheet, an effect illustrated in Fig. 1, so this approach is not accurate. Therefore, we address this problem as a case of spatially varying deconvolution [7,8], where the variation of the system's overall PSF is determined by the interaction between the detection PSF and the light-sheet. We note that, in general, the detection PSF itself can be spatially varying due to optical aberrations in the sample, a problem that is not specific to light-sheet microscopy. We do not address this source of variability in this work, although such a spatially varying detection PSF could in principle be incorporated in our method.

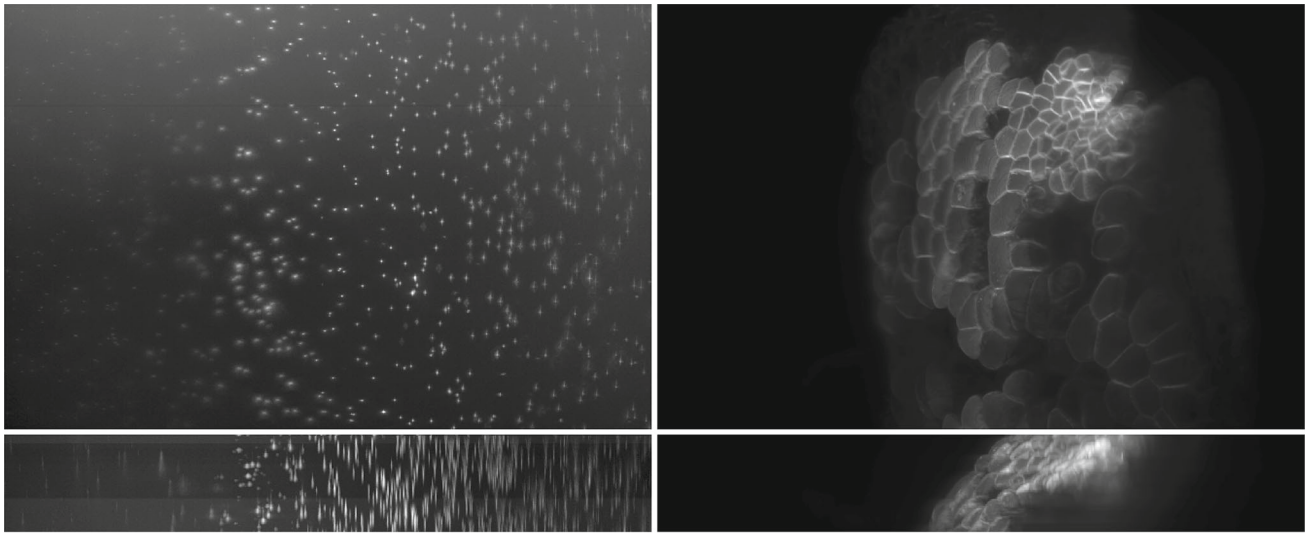
Two examples of acquired data are shown in Fig. 2. We can see in both cases the effect of the spatially varying light-sheet: the image is sharper in the centre and blurry on the sides, with the amount of blur growing with the horizontal distance from the centre. In addition, the fluorescence intensity of imaged beads in Fig. 2a is unevenly distributed despite imaging a homogeneous sample of beads, with the centre of the image being brighter than the left and right sides. The aim of our work is to correct these effects.

1.1 Contribution

We propose a method for deconvolution of 3D light-sheet microscopy data that takes into account the spatially varying nature of the PSF and is scalable to the dimensions typical to biological samples imaged using light-sheet microscopy—4.86 GB per 3D 16-bit stack of $2048 \times 2048 \times 580$ voxels.

Our approach is based on a new model for image formation that describes the interaction between the light-sheet and the detection PSF which replicates the physics of the microscope. Then, we formulate an inverse problem where the forward operator is given by model of the image formation process and which takes into account the degradation of the data by both Gaussian and Poisson noise as an infimal convolution (a concept that will be defined in Sect. 3) between an L^2 term and a Kullback–Leibler divergence term, following [9]. The proposed variational problem is solved by applying the Primal Dual Hybrid Gradient (PDHG) algorithm [6,10,11] in a novel way. Finally, we exploit the noise model to automatically tune the balance between the data fidelity and regularisation resorting to a discrepancy principle. We obtain convergence rates in a Bregman distance for the infimal convolution fidelity from [9] under a standard source condition.

In our numerical experiments, we first show how this method performs on simulated data, where the ground truth



(a) $0.5\mu m$ multi-colour Tetraspeck microspheres (slice) **(b)** Membrane labelled Marchantia (maximum intensity projection)

Fig. 2 Examples of light-sheet microscopy data of dimensions $665.6\,\mu\text{m} \times 665.6\,\mu\text{m}$: beads in **a** and Marchantia thallus in **b**. We show for both samples maximum intensity projections on the $x-y$ plane in the top row and on the $x-z$ plane in the bottom row, with the bead intensity shown in log scale for increased contrast. The effect of the light-sheet is visible along the horizontal direction (the x axis), as the image is sharp and has higher intensity in the centre, where the sheet is focused, while the quality of the image decreases away from the centre. The blurring effect of the light-sheet in the z direction is particularly noticeable in the $x-z$ projections in the bottom row. Another source of blur observed, especially in the bead image (left) is given by optical aberrations due to the sample imaging medium. The Marchantia image has been acquired using samples from Dr. Alessandra Bonfanti and Dr. Sarah Robinson using the genetic line provided by Prof. Sebastian Schornack and Dr. Giulia Arsuffi at the Sainsbury Laboratory Cambridge University

is known, then we apply our method to two examples of data from experiments: an image of fluorescent beads and a sample of Marchantia. In both cases, we see that the deconvolved images show improved contrast, while outperforming deconvolution using only the constant detection PSF.

1.2 Related Work

Before describing in more detail our approach to the deconvolution problem, we give a brief overview of the literature on spatially varying deconvolution in the context of microscopy and how our work relates to it.

Purely data-driven approaches estimate a spatially varying PSF in a low dimensional space (for scalability reasons) using bead images [7,8,12]. This is usually not application specific and can be included in a more general blind deconvolution framework. Similarly, the work in [13] involves writing the spatially varying PSF as a convex combination of spatially invariant PSFs. The algorithm alternates between estimating the image and estimating the PSF. In a similar vein, the authors of [14] approach the problem of blind deconvolution by defining the convolution operator using efficient matrix-vector multiplication operations. This decomposition is similar to the discrete formulation of our image formation model. These methods optimise over the (unknown) operator in addition to the unknown image. Related to these results is [15], where the authors consider the models from [12] and [14] under the assumption that the blurring operator is known and given as a sum of weighted spatially invariant operators. They exploit this structure of the operator and use a Douglas–Rachford-based splitting to solve the optimisation problem efficiently. A different data-driven approach is presented in [16], where a deep artificial neural network is used to learn the spatially varying PSF from simulated data obtained using a forward model of the microscope. While these approaches are more general than our method, we consider that using the knowledge of the image formation process in the forward model is advantageous for the reconstruction of light-sheet microscopy data.

A number of groups consider the problem of reconstruction from multiple views in the context of light-sheet microscopy. In [17], the problem of multi-view reconstruction under a spatially varying blurring operator for 3D light-sheet data is considered. They divide the image into small blocks where they perform deconvolution using spatially invariant PSFs estimated from beads (and interpolated PSFs in regions where there are no beads). In [18], the authors extend the Richardson–Lucy algorithm to the multi-view reconstruction problem in a Bayesian setting. While it allows for different PSFs for each view (estimated using beads), this work does not consider spatial variations of the PSF. While using data from multiple views improves the quality of the

reconstruction, these approaches are agnostic to the physics of the microscope.

The approach taken in [19,20] involves directly measuring the spatially varying PSF in different regions of the field of view using an additional hardware module installed with the microscope, and then deconvolving the image in each region using the measured PSF. In particular, [20] employs a sophisticated tiling-based deconvolution method based on the Richardson–Lucy algorithm and a formulation similar to a convolutional neural network in order to avoid artefacts usually caused by stitching tiles deconvolved with different PSFs.

Taking an approach similar in spirit to ours, the authors of [21] model the effective PSF of a light-sheet microscope, which is then plugged into a regularised version of the Richardson–Lucy algorithm for deconvolution. However, while they model the detection PSF and the light-sheet separately, they assume the effective PSF of the microscope is spatially invariant and the point-wise product of the two PSFs. In contrast, we do not take this simplifying step in our modelling, as we consider that the relationship between the two PSFs plays in important role in the resulting blur of the image.

The work of Guo et al. [22] uses a modified Richardson–Lucy algorithm implemented on GPU to improve the speed of convergence, further improved by the use of a deep neural network, which is a promising approach.

Moreover, in [23] the authors introduce an image formation model similar to the one described in the present manuscript. However, the regions of the resulting PSF where the light-sheet is out of focus are discarded, hence approximating the overall PSF with a constant PSF and then performing deconvolution using the ADMM algorithm. In Cueva et al. [24], a mathematical model which takes into account image fusion with two-sided illumination is derived from first principles. However, it is restricted to 2D and they do not apply the method to real data.

Lastly, regarding the mixed Gaussian–Poisson noise fidelity, our method follows the infimal convolution variational approach described in [9], with the additional light-sheet blurring operator. The same inverse problem, without the blurring operator, is solved in [25] albeit using an ADMM algorithm for the minimisation.

1.3 Paper Structure

The paper is organised as follows. In Sect. 2, we introduce a mathematical model of the image formation process in a light-sheet microscope. This model describes how the sample is blurred by the excitation illumination together with the detection objective PSF. Optical aberrations of the system are modelled using Zernike polynomials in the detection PSF, which we discuss in Sect. 2.3. In Sect. 3, we define the mathematical setting for the deconvolution problem and we state an inverse problem using a data fidelity as an infimal convolution of the individual Gaussian and Poisson data fidelities. We discuss convergence rates and a discrepancy principle for choosing the regularisation parameter in Sect. 2.1. In Sect. 4, we describe how PDHG is applied to this inverse problem, with details of the implementation of the proximal operator and the convex conjugate of the joint Kullback–Leibler divergence. Finally, we validate our method with numerical experiments both with simulated and real data in Sect. 5, before concluding and giving a few directions for future work in Sect. 6.

2 Forward Model

The first contribution of the current work is a model of the image formation process in light-sheet microscopy. By modelling the excitation light-sheet and the detection PSF separately and their interaction in a way that replicates the physics of the microscope, we are able to accurately simulate the spatially varying PSF of the imaging system. We then incorporate this knowledge as the forward model in an inverse problem, which we solve to remove the noise and blur in light-sheet microscopy data. In this section, we describe the image formation process and the PSF model.

2.1 Image Formation Model

A light-sheet propagated along the x direction is focused by the excitation objective at an axial position $z = z_0$ and the local light-sheet intensity l is modelled by the incoherent point spread function (PSF) of the excitation objective. The sample with local density of fluorophores u emits photons proportionally to the local intensity l of the light-sheet. These photons are then collected by a detection objective, whose action on the illuminated sample is modelled as a convolution with its PSF h. For clarity, see Fig. 3 for the directions of the axes. Finally, the sensor conjugated with the image plane z_0 collects photons and converts them to digital values for storage. Consequently, the recorded image is corrupted by a combination of Gaussian and Poisson noise. We can see here again how the local variation of the light-sheet will result in a spatially varying blur and spatially varying illumination intensity in the captured image. This process is then repeated for each z_0 to obtain the measured data f.

More specifically, we model u, f, l and h as functions defined on $\Omega \subset \mathbb{R}^3$, a rectangular domain of dimensions $\Omega_x \times \Omega_y \times \Omega_z$ (in μm) with $\Omega = [-\frac{\Omega_x}{2}, \frac{\Omega_x}{2}] \times [-\frac{\Omega_y}{2}, \frac{\Omega_y}{2}] \times [-\frac{\Omega_z}{2}, \frac{\Omega_z}{2}]$. For the sample u, the light-sheet l and the detec-

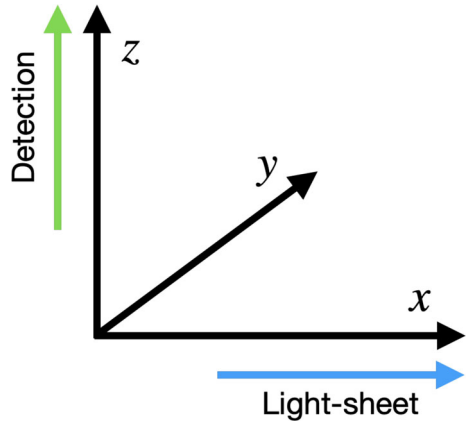


Fig. 3 Coordinate axes showing the light-sheet beam direction along the x axis and the detection direction along the z axis

tion objective PSF h, the measured data f is given by:

$$f(x,y,z) = \iiint u(s,t,w-z)l(s,t,w)h(x-s,y-t,-w)\mathrm{d}s\mathrm{d}t\mathrm{d}w. \tag{2.1}$$

The detection PSF h is given by

$$h(x,y,z) = \left| \iint g_\sigma * p_Z(\kappa_x,\kappa_y) e^{2i\pi z\sqrt{(n/\lambda_h)^2-\kappa_x^2-\kappa_y^2}} e^{2i\pi(\kappa_x x+\kappa_y y)} \mathrm{d}\kappa_x \mathrm{d}\kappa_y \right|^2 \tag{2.2}$$

and the light-sheet l is the y-averaged beam PSF l_{beam}:

$$l_{beam}(x,y,z) = \left| \iint p_0(\kappa_z,\kappa_y) e^{2i\pi x\sqrt{(n/\lambda_l)^2-\kappa_z^2-\kappa_y^2}} e^{2i\pi(\kappa_z z+\kappa_y y)} \mathrm{d}\kappa_z \mathrm{d}\kappa_y \right|^2, \tag{2.3}$$

where n is the refractive index, λ_h, λ_l are the wave lengths corresponding to the detection objective and light-sheet beam, respectively, and g_σ represents Gaussian blur. Lastly, $p_Z(\kappa_x,\kappa_y)$ and $p_0(\kappa_z,\kappa_y)$ are the pupil functions for the detection PSF and the light-sheet beam, respectively, both given by:

$$p_\varphi(x,y) = \begin{cases} e^{2i\pi\varphi} & \text{for } \sqrt{x^2+y^2} \le NA_i/\lambda_i, \\ 0 & \text{otherwise,} \end{cases} \tag{2.4}$$

for their respective wave lengths, $\lambda_i = \lambda_h$ or $\lambda_i = \lambda_l$, and numerical apertures, $NA_i = NA_h$ or $NA_i = NA_l$, where the phase for the light-sheet pupil p_0 is equal to zero and the phase for the detection PSF pupil p_Z is an approximation of the optical aberrations written as an expansion in a Zernike polynomial basis. The Gaussian blur g_σ in (2.2) is a technical detail that enables better fitting of the detection PSF h to the optical aberrations seen in bead data, an idea introduced in [26]. More details about the pupil functions and the aberration fitting using Zernike polynomials and the Gaussian blur g_σ will be given in Sect. 2.3. In general, the NA of the excitation sheet is much lower than the NA of the detection lens. We note that the overall process is not translation invariant and cannot be modelled by a convolution operator.

Note that both the detection PSF h and the light-sheet PSF have a similar formulation derived from:

$$\mathrm{PSF}(x,y,z) = \left| \iint p(\kappa_x,\kappa_y) e^{2i\pi z\sqrt{(n/\lambda_i)^2-\kappa_x^2-\kappa_y^2}} e^{2i\pi(\kappa_x x+\kappa_y y)} \mathrm{d}\kappa_x \mathrm{d}\kappa_y \right|^2, \tag{2.5}$$

which includes the pupil function for modelling aberrations and a defocus term before taking the Fourier transform (see, for example [27,28]). In addition, the actual light-sheet which illuminates a slice of the sample is obtained by rapid scanning of the illumination beam, which we model by y-averaging the illumination PSF l_{beam} given in (2.3) and repeating it in the y direction for the full length of the sample.

In practice, the image formation process modelled by (2.1) is discretised at the point of recording by the camera sensor in the xy plane and by the step size of the light-sheet in the z direction. If the camera has a resolution of $N_x \times N_y$ pixels and the light-sheet illuminates the sample at N_z distinct steps, the model (2.1) becomes:

$$\tilde{f}_{i,j,k} = \frac{1}{\tilde{C}} \sum_{i'=1}^{N_x} \sum_{j'=1}^{N_y} \sum_{k'=1}^{N_z} \tilde{l}_{i',j',k'} \tilde{u}_{i',j',k'-k} \tilde{h}_{i-i',j-j',k'}, \tag{2.6}$$

for all $i = 1,\ldots,N_x$, $j = 1,\ldots,N_y$, $k = 1,\ldots,N_z$, and a normalisation constant $\tilde{C}$, where $\tilde{u}, \tilde{f}, \tilde{l}, \tilde{h} \in \mathbb{R}^{N_x \times N_y \times N_z}$ are the discretised versions of u, f, l, h, respectively. Similarly, the sampling performed by the camera sensor leads to a discretisation of the Fourier space and the use of the discrete Fourier transform in the PSF and light-sheet models (2.2) and (2.3). Lastly, in our implementation we normalise $\tilde{h}$ so that $\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z} \tilde{h}_{i,j,k} = 1$ and choose the normalisation constant $\tilde{C}$ so that the norm of the resulting operator is equal to one.

2.2 Derivation of the Model

Let l, u, h be defined as in Sect. 2.1, with h and l centred at the origin and l translation invariant in the y direction and symmetric around the yz plane. For a fixed $z_0 \in [-\frac{\Omega_z}{2}, \frac{\Omega_z}{2}]$, we take the following steps, which replicate the inner workings of a light-sheet microscope:

1. Image the sample at $z = z_0$: centre the sample u at z_0 and multiply the result with the light-sheet l:

$$F(x, y, z; z_0) = u(x, y, z - z_0) \cdot l(x, y, z), \tag{2.7}$$

2. Convolve with the objective PSF h:

$$\begin{aligned} &C(x, y, z; z_0) \\ &\quad = F(x, y, z; z_0) * h(x, y, z) \\ &\quad = \iiint F(s, t, w; z_0) h(x - s, y - t, z - w) \mathrm{d}s\mathrm{d}t\mathrm{d}w, \end{aligned} \tag{2.8}$$

3. Slice at $z = 0$:

$$f(x, y, z_0) = [C(x, y, z; z_0)]_{z=0}, \tag{2.9}$$

which leads to:

$$\begin{aligned} f(x, y, z_0) = \iiint u(s, t, w - z_0) l(s, t, w) h(x - s, y \\ - t, -w) \mathrm{d}s\mathrm{d}t\mathrm{d}w. \end{aligned} \tag{2.10}$$

This is the same as model (2.1), where we substitute z for z_0. Note that, if there are no aberrations in h or other sources of asymmetry in the z direction, we could simply write $h(x - s, y - t, w)$ instead.

For a discretisation of the domain using a 3D grid with $N_x \times N_y$ pixels and N_z light-sheet steps, the forward model can be computed by following the three steps above for each $k = 1, \ldots, N_z$, where we perform the convolutions using the fast Fourier transform (FFT), resulting in a number of $\mathcal{O}(N_x N_y N_z^2 \log(N_x N_y N_z))$ operations.

Alternatively, we can rewrite the last integral above as:

$$f(x, y, z_0) = \int K(x, y, w) * h(x, y, -w) \mathrm{d}w, \tag{2.11}$$

where

$$K(x, y, w) = l(x, y, w) u(x, y, w - z_0), \tag{2.12}$$

and the convolution in (2.11) is a 2D convolution in (x, y):

$$\begin{aligned} &K(x, y, w) * h(x, y, -w) \\ &\quad = \iint K(s, t, w) h(x - s, y - t, -w) \mathrm{d}s\mathrm{d}t. \end{aligned} \tag{2.13}$$

In terms of numbers of FFTs performed on a discretised $N_x \times N_y \times N_z$ grid, this alternative formulation requires $\mathcal{O}(N_x N_y N_z^2 \log(N_x N_y))$ operations.

2.3 Point Spread Function Model

While both the light-sheet profile and the detection PSF are based on the same model of a defocused system (2.5) introduced in [27], note that our definition of h in (2.2) includes an additional convolution operation with a Gaussian g_σ and a pupil function p_Z with a nonzero phase. Let us turn to why this is the case.

It is well known that optical aberrations hamper results based on deconvolution with theoretical PSFs. In light-sheet microscopy, the effect of aberrations is more visible away from the centre, as shown for example in the bead image in Fig. 2, or in the more detailed example beads in Fig. 4. It is, therefore, required that we model the (spatially invariant) aberrations of the detection lens.

The general PSF model (2.5), with the phase of the pupil function equal to zero, does not take optical aberrations into account and therefore it is not an accurate representation of the objective PSF h. For example, a PSF calculated using (2.5) with zero phase of the pupil and the parameters of the detection objective, shown in Fig. 5, does not resemble the actual bead images in the data in Fig. 4.

There has been extensive work on the problem of phase reconstruction in the literature [26,29,30], but here we take a more straightforward approach using Zernike polynomials to include aberrations in the PSF [31], as follows. Let h_z be the objective PSF calculated using (2.5) with Zernike polynomials in the phase of the pupil function:

$$\begin{aligned} &h_z(x, y, z; c) \\ &\quad = \left| \iint p_Z(\kappa_x, \kappa_y; c) e^{2i\pi z \sqrt{(n/\lambda_h)^2 - \kappa_x^2 - \kappa_y^2}} e^{2i\pi(\kappa_x x + \kappa_y y)} \mathrm{d}\kappa_x \mathrm{d}\kappa_y \right|^2, \end{aligned} \tag{2.14}$$

where $p_z(\kappa_x, \kappa_y; c)$ is the pupil function with N_Z Zernike polynomials in the phase:

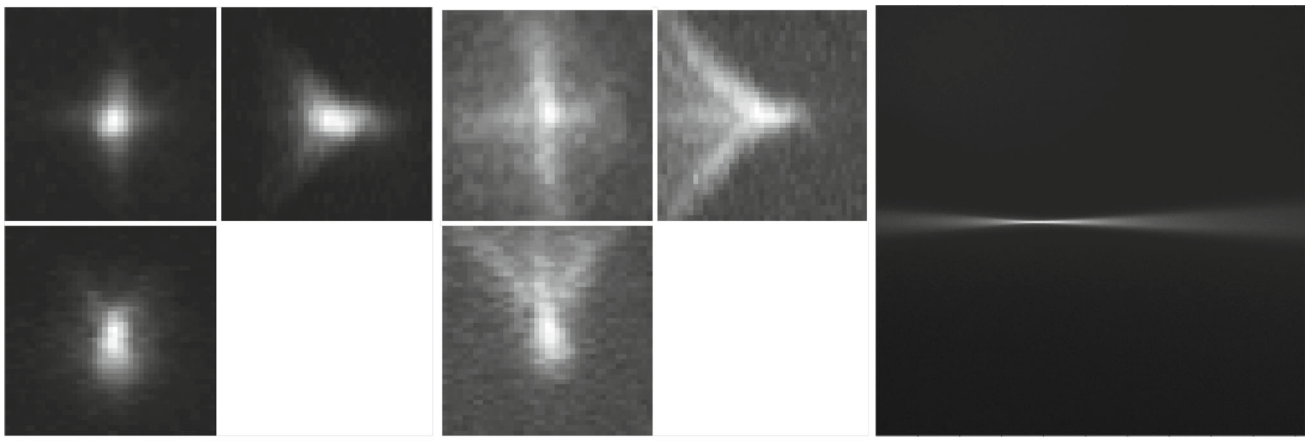
$$\begin{aligned} &p_z(\kappa_x, \kappa_y; c) \\ &\quad = \begin{cases} e^{2i\pi \sum_{j=1}^{N_Z} c_j Z_j(\kappa_x, \kappa_y)}, & \text{for } \sqrt{\kappa_x^2 + \kappa_y^2} \le NA/\lambda_h \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{2.15}$$

and $c = [c_1, \ldots, c_{N_z}]^T$ are coefficients corresponding to the polynomials for some integer $N_Z > 0$.

Moreover, let h_{zb} be the blurred PSF obtained by convolving h_z with a Gaussian g_σ with width σ:

$$h_{zb}(x, y, z; c, \sigma) = h_z(x, y, z; c) * g_\sigma. \tag{2.16}$$

This allows us to obtain a better approximation of the objective PSF [26]. The parameters c and σ are calculated by



(a) Central bead, with minimal aberration (maximum intensity projections)

(b) Peripheral bead, with strong aberration (maximum intensity projections)

(c) Light sheet profile (slice showing an $x - z$ plane)

Fig. 4 Examples of beads and light-sheet profile. The bead in **a** is cropped from the centre of Fig. 2a and the bead in **b** is cropped from the right-hand side of Fig. 2a. The maximum intensity projections are taken in the $x - y$ plane (top left), the $z - y$ plane (top right) and the $x - z$ plane (bottom left)

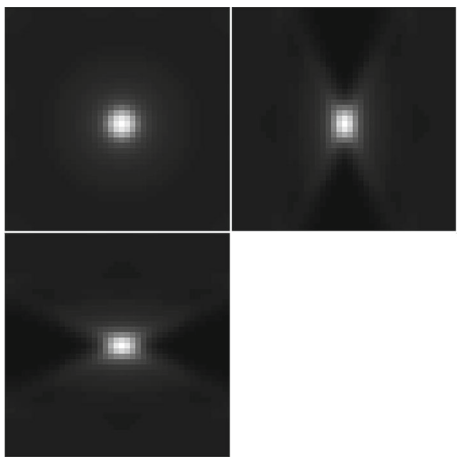


Fig. 5 Objective PSF used in our model, with no aberrations (maximum intensity projections taken in the same way as in Fig. 4)

solving the least-squares problem

$$\min_{c,\sigma} \quad \|h_{zb}(c,\sigma) * b - h_{bead}\|_2^2 \tag{2.17}$$

$$\text{subject to} \quad c \in [-B_Z, B_Z]^{N_Z}, \sigma > 0, \tag{2.18}$$

for some $B_Z > 0$, where h_{bead} is a bead image containing the aberrations that one wants to capture in the fitted detection PSF (for example the bead in Fig. 4b) and b is equal to one inside the sphere of the radius equal to the radius of the bead (a parameter that is provided) and zero outside the sphere. This takes into account the non-negligible size of the beads used to generate the data.

In the implementation of the fitting procedure, we normalise both the bead image h_{bead} and the simulated PSF h_{zb} by their maximum values before calculating their error, and we include two additional parameters, scaling and shift, to ensure a better fit of the intensity values (not shown here for simplicity of the presentation).

The best choice of the number of Zernike polynomial basis elements N_Z and the boundary B_Z of the coefficients c depend on the data h_{bead} and how well the fit is required to be in the deconvolution step. In general, at least the first 15 Zernike polynomials are needed to capture the main optical aberrations such as spherical and astigmatism. In our experiments, for the bead shown in Fig. 4b, we found that $N_Z = 15$ and $B = 3$ are an appropriate choice. The Zernike polynomials used and their resulting corresponding coefficients are shown in Table 1 and in Fig. 6. The resulting PSF is the detection PSF model (2.2) and is shown in Fig. 7.

Finally, we note that a more thorough approach to choosing the Zernike basis is possible, by using multiple bead images for fitting the Zernike coefficients and comparing the quality of the fit for different values of N_Z and B_Z. Alternatively, one could average multiple beads and perform the fitting procedure described above on the averaged bead. In both cases, it is worth mentioning that, since the optical aberrations vary within the sample image, we would only be able to fit the general shape of the PSF rather than the sharper features present in each bead, effectively fitting the low frequency information in the beads. In the end, this would achieve a similar effect to the Gaussian blur g_σ that we use in the fitting process. Moreover, one can employ more advanced techniques such as the ones described in [26,29,30] for estimating the pupil function, which can be plugged in to our image formation model. However, such an analysis focused on the pupil function is beyond the scope of the present work.

Table 1 The first 15 Zernike Polynomials (in polar coordinates) and their coefficients used in h_z

Z_j	Polynomial	c_j
Z_1	$\rho\cos\theta$	−0.7763
Z_2	$\rho\sin\theta$	−0.0460
Z_3	$2\rho^2-1$	−2.3608
Z_4	$\rho^2\cos 2\theta$	−1.3001
Z_5	$\rho^2\sin 2\theta$	0.2024
Z_6	$(3\rho^2-2)\rho\cos\theta$	−0.3999
Z_7	$(3\rho^2-2)\rho\sin\theta$	0.0348
Z_8	$6\rho^4-6\rho^2+1$	−1.2112
Z_9	$\rho^3\cos 3\theta$	−0.1521
Z_{10}	$\rho^3\sin 3\theta$	−0.0466
Z_{11}	$(4\rho^2-3)\rho^2\cos 2\theta$	−0.0930
Z_{12}	$(4\rho^2-3)\rho^2\sin 2\theta$	0.0427
Z_{13}	$(10\rho^4-12\rho^2+3)\rho\cos\theta$	−0.0117
Z_{14}	$(10\rho^4-12\rho^2+3)\rho\sin\theta$	−0.0581
Z_{15}	$20\rho^6-30\rho^4+12\rho^2-1$	−0.0633

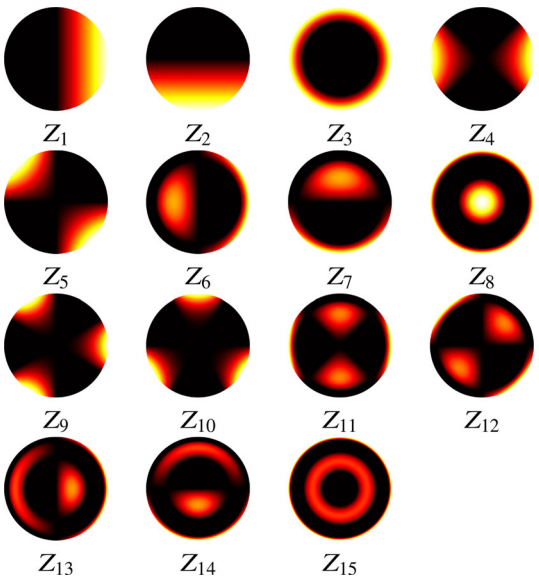


Fig. 6 The Zernike polynomials used in the PSF h_z, with image range $[-1, 1]$

2.4 Convolution with Spatially Varying Kernel

Having introduced the image formation model for a light-sheet microscope (2.1) as well as the models for the individual PSFs, it is worth expanding on the source of spatial variability that we tackle in this work.

First, note that with a change of variable $w \to w+z$, we can rewrite the model (2.1) as:

$$f(x,y,z) = \iiint u(s,t,w)l(s,t,w+z)h(x-s,y-t,-w-z)\mathrm{d}s\mathrm{d}t\mathrm{d}w,$$

so $f(x,y,z)$ is the convolution of $u(x,y,z)$ with the spatially varying kernel $\tilde{h}(x,y,z;s,t,w)$:

$$f(x,y,z) = \iiint u(s,t,w)\tilde{h}(x,y,z;s,t,w)\mathrm{d}s\mathrm{d}t\mathrm{d}w, \tag{2.19}$$

where

$$\tilde{h}(x,y,z;s,t,w) = l(s,t,w+z)h(x-s,y-t,-w-z) \tag{2.20}$$

gives the expression of a kernel $\tilde{h}(x,y,z;\cdot,\cdot,\cdot)$ which varies with its centre (x,y,z).

Therefore, the model presented in this section describes the spatial variation of the *overall* PSF of the system $\tilde{h}$, as a consequence of the interaction between the light-sheet beam PSF l and the detection PSF h. We highlight that l and h are not themselves spatially varying. By the process described in Sect. 2.2, this interaction (and spatial variability) is modelled explicitly. This is in contrast to approaches such as [16], where the spatial variability of the PSF is learned from the data and encoded in the black-box mechanism of an artificial neural network.

In practice, a second source of spatial variability of the PSF may be the detection PSF h, due to the optical aberrations that can vary within the sample image. As described in Sect. 2.3, in this work we do not account for this potential spatial variability of h, and we fit one pupil function to the bead data.

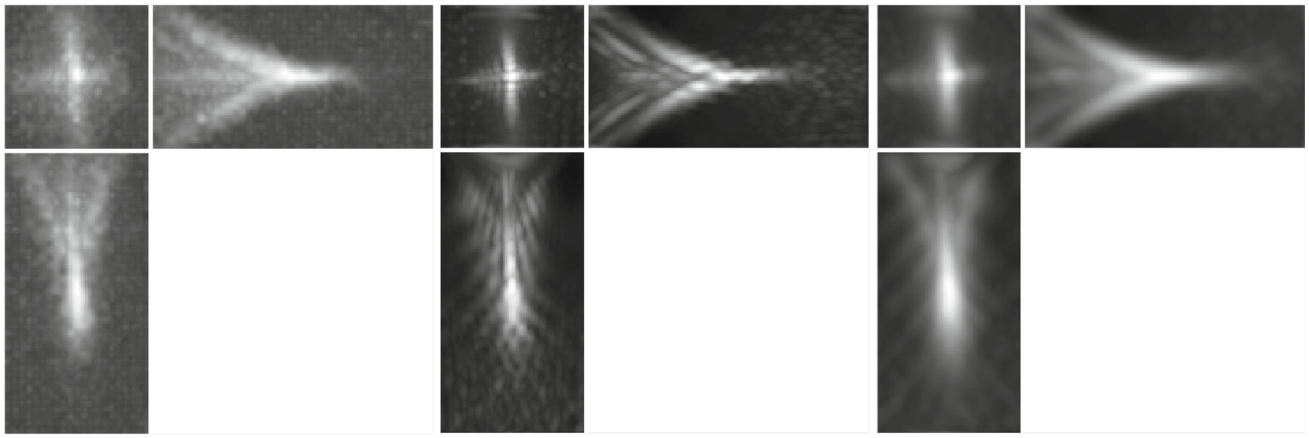
3 Inverse Problem

3.1 Problem Statement

In this section we formally state the inverse problem of deblurring a light-sheet microscopy image. Let $\Omega \subset \mathbb{R}^3$ be a bounded Lipschitz domain and let $L\colon \mathrm{L}^p(\Omega) \to \mathrm{L}^2(\Omega)$ be the forward operator defined by (2.1). Here $1 < p < 3/2$ is chosen such that the embedding of the BV space is compact [32]. Clearly, L is linear.

We consider the following inverse problem

$$Lu = \bar{f}, \tag{3.1}$$



(**a**) Bead image data (maximum intensity projections) (**b**) Fitted PSF h_z, no blur (maximum intensity projections) (**c**) Fitted PSF h with Gaussian blur (maximum intensity projections)

Fig. 7 Fitted PSF using Zernike polynomials. In panel **c**, we can see the benefits of using the Gaussian blur g_σ in obtaining an accurate approximation of the bead in **a**

where $\bar{f} \in \mathrm{L}^2(\Omega)$ is the exact (noise-free) data. As outlined in Sect. 2.1, the measurements in light microscopy are corrupted by a combination of Poisson and Gaussian noise. More precisely, the measurement is given by $f = v + w$, where $v \sim Pois(\bar{f})$ is a Poisson distributed random variable with mean $\bar{f}$ and w represents additive zero-mean Gaussian noise. We do not model Gaussian noise statistically and instead, in the spirit of (deterministic) variational regularisation, assume that $w \in \mathrm{L}^2(\Omega)$ is a fixed perturbation with $\|w\|_{\mathrm{L}^2(\Omega)} \leq \sigma_G$ for some known $\sigma_G > 0$. Poisson noise is typically modelled using the Kullback–Leibler divergence as the data fidelity term [33,34].

Let us give a brief justification of the inverse problem formulation described in this section [9,35], from a Bayesian perspective. First, by using the Poisson and Gaussian probability density functions, we have that $p(v|u) = \frac{(Lu)^v e^{-(Lu)}}{v!}$ and $p(f|v) = \frac{1}{\sqrt{2\pi}\sigma_G} e^{-\frac{1}{2}\left(\frac{f-v}{\sigma_G}\right)^2}$, and from Bayes' theorem and conditional probability:

$$p(u, v|f) = \frac{p(f|v)p(v|u)p(u)}{p(f)}, \tag{3.2}$$

where we used that $p(f|u, v) = p(f|v)$. Moreover, we assume that the prior is a Gibbs distribution $p(u) = e^{-\alpha \mathcal{J}(u)}$ for a convex functional $\mathcal{J}(u)$, which we will introduce later. To obtain a maximum *a posteriori* estimation of u and v (i.e. maximise the posterior distribution $p(u, v|f)$), we take the minimum of the negative log of (3.2) and, after discarding the denominator $p(f)$ and using the Stirling approximation for the factorial $\log v! = v \log v - v$, we obtain the minimisation problem:

$$\arg\min_{u,v} \alpha \mathcal{J}(u) + \frac{1}{2\sigma_G^2}\|f - v\|^2 + v \log \frac{v}{Lu} + Lu - v, \tag{3.3}$$

where the first term is the regularisation term and the remaining terms form the data fidelity term.

We will now describe the formal mathematical setting for (3.3) in the context of variational regularisation. This will allow us to show well-posedness of the model, establish convergence rates of the solution with respect to the noise in the measurements and to introduce a discrepancy principle for choosing the value of the regularisation parameter α.

First, note that in (3.3), we can perform the minimisation over v only on the data fidelity part of the objective, which can be written as an infimal convolution of the two separate Gaussian and Poisson fidelities. The infimal convolution of two functionals φ_1, φ_2 on L^2 is defined as[1]:

$$(\varphi_1 \square \varphi_2)(f) = \inf_{v \in \mathrm{L}^2(\Omega)} \{\varphi_1(f - v) + \varphi_2(v)\}, \tag{3.4}$$

for $f \in \mathrm{L}^2(\Omega)$. Therefore, we define the following data fidelity term, as proposed in [9]:

$$\Phi(\bar{f}, f) := \inf_{v \in \mathrm{L}^2_+(\Omega)} \left\{ \frac{1}{2}\|f - v\|^2_{\mathrm{L}^2} + D_{KL}(v, \bar{f}) \right\}, \tag{3.5}$$

for $f \in \mathrm{L}^2(\Omega)$ and $\bar{f} \in \mathrm{L}^1_+(\Omega)$, where $\mathrm{L}^{1,2}_+(\Omega)$ denotes the positive cone in $\mathrm{L}^{1,2}(\Omega)$ (that is, functions $f \in \mathrm{L}^{1,2}(\Omega)$ such that $f \geq 0$ a.e.) and D_{KL} denotes the Kullback–Leibler

[1] There is a geometric interpretation of the infimal convolution. Given two functions φ_1, φ_2 on Ω, the epigraph (the set of all points lying on or above the graph of a function) of their infimal convolution $\varphi_1 \square \varphi_2$ is the sum of the epigraphs of φ_1 and φ_2. This can be easily seen if we rewrite the definition (3.4) as $(\varphi_1 \square \varphi_2)(f) = \inf\{\varphi_1(u) + \varphi_2(v) | u, v \in \Omega, u + v = f\}$.

divergence, which we define as follows

$$D_{KL}(v,\bar{f}) := \begin{cases} \int_\Omega \left(\bar{f}(x) - v(x) + v(x)\log\frac{v(x)}{\bar{f}(x)}\right)\mathrm{d}x, \\ \quad \text{if } v,\bar{f} \geq 0 \wedge \int_\Omega v\mathrm{d}x = \int_\Omega \bar{f}\mathrm{d}x = 1, \\ +\infty, \qquad \text{otherwise,} \end{cases}$$

$$= \begin{cases} \int_\Omega v(x)(\log v(x) - \log \bar{f}(x))\mathrm{d}x, \\ \quad \text{if } v,\bar{f} \geq 0 \wedge \int_\Omega v\mathrm{d}x = \int_\Omega \bar{f}\mathrm{d}x = 1, \\ +\infty, \qquad \text{otherwise.} \end{cases} \tag{3.6}$$

We note that $\left|\int_\Omega v(x)\log v(x)\mathrm{d}x\right| < \infty$ for $v \in \mathrm{L}^2$, since L^2 is continuously embedded into the Orlicz space $\mathrm{L}\log\mathrm{L}$ of functions of finite entropy [36,37]

$$\mathrm{L}\log\mathrm{L}(\Omega) := \{f \in \mathrm{L}^1(\Omega) \colon \int_\Omega |f(x)|\,(\log|f(x)|)_+\mathrm{d}x < \infty\}, \tag{3.7}$$

where $(\cdot)_+ = \max\{\cdot, 0\}$ denotes the positive part.

A proof of the following result can be found in [9], but we provide it here for readers' convenience.

Proposition 3.1 (Exactness of the infimal convolution) *For any $\bar{f} \in \mathrm{L}^1_+$ such that $\int_\Omega \bar{f}\mathrm{d}x = 1$, there exists a unique solution $v^* = v^*(\bar{f})$ of (3.5), that is, the infimal convolution is exact. Moreover, the functional $\Phi(\bar{f},\cdot)\colon \mathrm{L}^2 \to \mathbb{R}_+ \cup \{+\infty\}$ is proper, convex and lower semicontinuous.*

Proof Fix $\bar{f} \in \mathrm{L}^1_+$ such that $\int_\Omega \bar{f}\mathrm{d}x = 1$. Then, (3.5) is the infimal convolution of the following two functionals on L^2

$$\varphi(g) := \chi_{\mathrm{L}^2_+}(g) + D_{KL}(g,\bar{f}),$$
$$\psi(g) := \frac{1}{2}\|g\|^2_{\mathrm{L}^2}, \quad g \in \mathrm{L}^2(\Omega),$$

where χ denotes the characteristic function. The function φ is proper, convex, non-negative and lower semicontinuous, while ψ is proper, convex, lower semicontinuous and coercive. Therefore, by [38, Prop. 12.14], the infimal convolution is exact and is itself a proper, convex and lower semicontinuous function. Uniqueness follows from strict convexity of ψ. □

Now we turn our attention to the lower semicontinuity of the functional $\Phi(\cdot, f)$ in its first argument.

Proposition 3.2 (Lower semicontinuity) *For any $f \in \mathrm{L}^2_+(\Omega)$ such that $\int_\Omega f\mathrm{d}x = 1$ the functional $\Phi(\cdot,f)\colon L^1(\Omega) \to \mathbb{R}_+ \cup \{+\infty\}$ is lower semicontinuous.*

Proof We have

$$\Phi(g,f) = \inf_{v\in\mathrm{L}^2_+(\Omega)} \left\{\frac{1}{2}\|f-v\|^2_{\mathrm{L}^2} + D_{KL}(v,g)\right\}$$
$$= \frac{1}{2}\left\|f - v^*(g)\right\|^2_{\mathrm{L}^2} + D_{KL}(v^*(g),g)$$
$$= \frac{1}{2}\left\|f - v^*(g)\right\|^2_{\mathrm{L}^2} + \\ + \int_\Omega v(x)(\log v(x) - \log\bar{f}(x))\mathrm{d}x + \chi_{\mathcal{C}}(g),$$

where $g \in \mathrm{L}^1(\Omega)$, $v^*(g)$ is as defined in Proposition 3.1 and $\mathcal{C} := \{g \in \mathrm{L}^1_+(\Omega)\colon \int_\Omega g\mathrm{d}x = 1\}$. The characteristic function is lower semicontinuous because $\mathcal{C}$ is closed in L^1 and the rest is lower semicontinuous by [9, Thm. 4.1]. □

The following fact is easily established.

Proposition 3.3 *The operator $L\colon \mathrm{L}^p(\Omega) \to \mathrm{L}^1(\Omega)$ defined in (2.1) is continuous for any $p \geq 1$. Moreover, if l and h are non-negative and have overlapping support:*

$$\mathrm{supp}(l) \cap \mathrm{supp}(h) \neq \varnothing,$$

then $\mathbf{1} \notin \mathcal{N}(L)$, where $\mathbf{1}$ is the constant one function and $\mathcal{N}(L)$ is the null space of L.

Proof By (2.1), we have

$$Lu(x,y,z) = \int_\Omega l(s,t,w)h(x-s,y-t,w)u(s,t,w-z)\mathrm{d}\mu_{stw},$$

where $\mathrm{d}\mu_{stw} := \mathrm{d}s\mathrm{d}t\mathrm{d}w$. Noting that the light-sheet PSF l and detection PSF h are bounded from above by some $C_1, C_2 > 0$, we have that:

$$\begin{aligned} &\|Lu\|_{\mathrm{L}^1} \\ &= \int_\Omega \left|\int_\Omega l(s,t,w)h(x-s,y-t,w)u(s,t,w-z)\mathrm{d}\mu_{stw}\right|\mathrm{d}\mu_{xyz} \\ &= \int_\Omega \Big|\int_\Omega l(s,t,w'+z)h \\ &\quad (x-s,y-t,w'+z)u(s,t,w')\mathrm{d}\mu_{stw'}\Big|\,\mathrm{d}\mu_{xyz} \\ &\quad (\text{by } w' = w - z) \\ &\leq C_1C_2\int_\Omega \left|\int_\Omega u(s,t,w')\mathrm{d}\mu_{stw'}\right|\mathrm{d}\mu_{xyz} \\ &= C_1C_2\,|\Omega|\left|\int_\Omega u(s,t,w')\mathrm{d}\mu_{stw'}\right| \\ &\leq C(p)\,\|u\|_{\mathrm{L}^p}, \end{aligned}$$

where in the last inequality we applied Hölder's inequality and $C(p)$ is a constant that depends on p (as well as $C_{1,2}$ and Ω). Hence, we obtain the first claim.

For the second claim, we observe that

$$L\mathbf{1}(x,y,z) = \int_\Omega l(s,t,w)h(x-s,y-t,w)\mathrm{d}\mu_{stw} \geq 0.$$

Consider

$$\int_\Omega L\mathbf{1}(x,y,z)\mathrm{d}\mu_{xyz} = \int_\Omega\int_\Omega l(s,t,w)h(x-s,y-t,w)\mathrm{d}\mu_{stw}\mathrm{d}\mu_{xyz}$$

and let $B_{l,h} \subset \operatorname{supp} l \cap \operatorname{supp} h$. Then, since both l and h are non-negative on Ω, from the last equality above we have that:

$$\int_\Omega L\mathbf{1}(x,y,z)\mathrm{d}\mu_{xyz} \geq \int_{B_{l,h}} l(x,y,z)h(x,y,z)\mathrm{d}\mu_{xyz} > 0,$$

which proves the second claim. □

Remark 3.1 Our setting with the measured data $f \in \mathrm{L}^2(\Omega)$ differs slightly from [9], where $f \in \mathrm{L}^\infty(\Omega)$ was assumed.

We will consider the following variational regularisation problem

$$\min_{u\in \mathrm{L}^p_+(\Omega)} \Phi(f, Lu) + \alpha\mathcal{J}(u), \tag{3.8}$$

where Φ is the infimal convolution fidelity as defined in (3.5), $\mathcal{J}\colon \mathrm{L}^p \to \mathbb{R}_+ \cup \{+\infty\}$ is a regularisation functional, $\alpha \in \mathbb{R}_+$ is a regularisation parameter and $1 < p < 3/2$. Without loss of generality, we assume that $\int_\Omega \bar{f}\mathrm{d}x = 1$.

As the regulariser $\mathcal{J}$, we choose the total variation [39]

$$\mathcal{J}(u) = \mathrm{TV}(u) := \sup_{\substack{\xi\in C_0^\infty(\Omega,\mathbb{R}^3)\\ \|\xi\|_\infty\leq 1}} \int_\Omega u \operatorname{div}(\xi)\mathrm{d}x.$$

By the Rellich–Kondrachov theorem, the space

$$\mathrm{BV}(\Omega) := \{u \in \mathrm{L}^1(\Omega)\colon \mathrm{TV}(u) < \infty\},$$
$$\|u\|_{\mathrm{BV}} := \|u\|_{\mathrm{L}^1} + \mathrm{TV}(u),$$

is compactly embedded into $\mathrm{L}^p(\Omega)$ for $1 \leq p < 3/2$ and continuously embedded into $\mathrm{L}^{3/2}(\Omega)$ since $\Omega \subset \mathbb{R}^3$. Therefore, we consider $\mathrm{TV}\colon \mathrm{L}^p \to \mathbb{R}_+ \cup \{+\infty\}$

$$\mathrm{TV}(u) := \begin{cases} \sup\limits_{\substack{\xi\in C_0^\infty(\Omega,\mathbb{R}^3)\\ \|\xi\|_\infty\leq 1}} \int_\Omega u \operatorname{div}(\xi)\mathrm{d}x, & u \in \mathrm{BV}(\Omega),\\ \infty, & u \in \mathrm{L}^p(\Omega)\backslash\mathrm{BV}(\Omega). \end{cases}$$

We will denote by $u^\dagger_{\mathrm{TV}}$ the TV-minimising solution of (3.1), i.e. a solution that satisfies

$$Lu^\dagger_{\mathrm{TV}} = \bar{f} \quad \text{and} \quad \mathrm{TV}(u^\dagger_{\mathrm{TV}}) \leq \mathrm{TV}(u) \text{ for all } u \text{ s.t. } Lu = \bar{f}.$$

The existence of such solution is obtained by standard arguments [40]. We will make the reasonable assumption that the TV-minimising solution is positive, i.e. $u^\dagger_{\mathrm{TV}} \geq 0$ a.e. Due to the positivity of the kernels involved in (2.1), it is clear that $u^\dagger_{\mathrm{TV}} \geq 0$ implies $Lu^\dagger_{\mathrm{TV}} = \bar{f} \geq 0$.

Since by Proposition 3.1 the infimal convolution (3.5) is exact, we can equivalently rewrite (3.8) as follows

$$\min_{\substack{u\in \mathrm{L}^p_+(\Omega)\\ v\in \mathrm{L}^2_+(\Omega)}} \frac{1}{2}\|f - v\|^2_{\mathrm{L}^2} + D_{KL}(v, Lu) + \alpha\mathcal{J}(u). \tag{3.9}$$

Existence of minimisers in (3.8) and (3.9) is obtained by standard arguments [9, Thm. 4.1].

Proposition 3.4 *Each of the optimisation problems* (3.8) *and* (3.9) *admits a unique minimiser.*

We will also need the following coercivity result.

Proposition 3.5 *The functional* $\Phi(f,\cdot)\colon \mathrm{L}^1(\Omega) \to \mathbb{R}_+ \cup \{+\infty\}$ *is strongly coercive with exponent* 2*, i.e. there exists a constant* $C > 0$ *such that*

$$\Phi(f, g) \geq C\,\|g - f\|^2_{\mathrm{L}^1}, \quad g \in \mathrm{L}^1(\Omega).$$

Proof Using Pinsker's inequality for the Kullback–Leibler divergence, we get

$$\begin{aligned}\Phi(f,g) &= \inf_{v\in \mathrm{L}^2_+} \frac{1}{2}\|v - f\|^2_{\mathrm{L}^2} + D_{KL}(v, g)\\ &\geq \inf_{v\in \mathrm{L}^2_+} \frac{1}{2}\|v - f\|^2_{\mathrm{L}^2} + \|g - v\|^2_{\mathrm{L}^1}\\ &\geq 2C \inf_{v\in \mathrm{L}^2_+} \|v - f\|^2_{\mathrm{L}^1} + \|g - v\|^2_{\mathrm{L}^1}\end{aligned}$$

for some $C > 0$. Note that Pinsker's inequality assumes that $f, g \geq 0$ and $\inf_\Omega f\mathrm{d}x = \int_\Omega g\mathrm{d}x = 1$, which we ensure by definition in (3.6).

Now, using the inequality $\frac{1}{2}(a+b)^2 \leq a^2 + b^2$ that holds for all $a, b \in \mathbb{R}$ and the triangle inequality, we obtain the claim

$$\begin{aligned}\Phi(f,g) &\geq C \inf_{v\in \mathrm{L}^2_+} \left(\|v - f\|_{\mathrm{L}^1} + \|g - v\|_{\mathrm{L}^1}\right)^2\\ &\geq C \inf_{v\in \mathrm{L}^2_+} \|v - f + g - v\|^2_{\mathrm{L}^1}\\ &= C\,\|g - f\|^2_{\mathrm{L}^1}.\end{aligned}$$

□

3.2 Convergence Rates

Our aim in this section is to establish convergence rates of minimisers of (3.8) as the amount of noise in the data decreases. But first we need to specify what we mean by the amount of noise in our setting.

We argue as follows. Since the noise in the measurement is generated sequentially, i.e. photo-electrons are first counted by the sensor leading to a Poisson noise and later they are collected by the electronic circuit generating an additive Gaussian noise, for any exact data $\bar{f}$ there exists $\bar{z} \sim Pois(\bar{f})$ such that $D_{KL}(\bar{z}, \bar{f}) \leq \gamma$, where $\gamma > 0$ depends on the exposure time t and vanishes as $t \to \infty$ [33]. Further, there exists $w \in \mathrm{L}^2(\Omega)$ with $\|w\|_{\mathrm{L}^2} \leq \sigma_G$ such that $f = \bar{z} + w$. Since $\bar{z} \geq 0$ is feasible in (3.5), we get the following upper bound on the fidelity term (3.5) evaluated at the measurement f and the exact data $\bar{f}$

$$\begin{aligned}\Phi(\bar{f}, f) &\leq \frac{1}{2}\|f - \bar{z}\|_{\mathrm{L}^2}^2 + D_{KL}(\bar{z}, \bar{f}) \\ &= \frac{1}{2}\|w\|_{\mathrm{L}^2}^2 + D_{KL}(\bar{z}, \bar{f}) \\ &\leq \frac{\sigma_G^2}{2} + \gamma. \end{aligned} \tag{3.10}$$

The standard tool for establishing convergence rates are Bregman distances associated with the regulariser $\mathcal{J}$. We briefly recall the necessary definitions.

Definition 3.1 Let X be a Banach space and $\mathcal{J}\colon X \to \mathbb{R}_+ \cup \{+\infty\}$ a proper convex functional. The generalised Bregman distance between $x, y \in X$ corresponding to the subgradient $q \in \partial\mathcal{J}(y)$ is defined as follows

$$D_{\mathcal{J}}^q(x, y) := \mathcal{J}(x) - \mathcal{J}(y) - \langle q, x - y\rangle.$$

Here $\partial\mathcal{J}(v)$ denotes the subdifferential of $\mathcal{J}$ at $y \in X$. If, in addition, $p \in \partial\mathcal{J}(x)$, the symmetric Bregman distance between $x, y \in X$ corresponding to the subgradients p, q is defined as follows

$$D_{\mathcal{J}}^{p,q}(x, y) := D_{\mathcal{J}}^q(x, y) + D_{\mathcal{J}}^p(y, x) = \langle p - q, x - y\rangle.$$

To obtain convergence rates, an additional assumption on the regularity of the TV-minimising solution, called the *source condition*, needs to be made. We use the following variant [41].

Assumption 3.1 (Source condition) There exists an element $\mu^\dagger \in L^\infty(\Omega)$ such that

$$q^\dagger := L^*\mu^\dagger \in \partial\mathcal{J}(u_{\mathrm{TV}}^\dagger).$$

3.2.1 Parameter Choice Rules

Let us summarise what we know about the fidelity function Φ as defined in (3.5), the regularisation functional TV and the forward operator L:

- $\Phi(f, \cdot)$ is proper, convex and coercive (Proposition 3.5) in $\mathrm{L}^1(\Omega)$;
- $\Phi(\cdot, \cdot)$ is jointly convex [42] and lower semicontinuous (Propositions 3.1 and 3.2);
- $\Phi(f, g) = 0$ if and only if $f = g$;
- $\mathrm{TV}\colon \mathrm{L}^1(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and lower semicontinuous [32] and its null space is given by $\mathcal{N}(\mathrm{TV}) = \mathrm{span}\{\mathbf{1}\}$, where $\mathbf{1}$ denotes the constant one function;
- TV is coercive on the complement of its null space in $\mathrm{L}^1(\Omega)$ [32];
- $L\colon \mathrm{L}^p(\Omega) \to \mathrm{L}^1(\Omega)$ is continuous and $\mathcal{N}(\mathrm{TV}) \cap \mathcal{N}(L) = \{0\}$ (Proposition 3.3).

Using these facts and slightly modifying the proofs from [43], we obtain the following

Theorem 3.2 (Convergence rates under a priori parameter choice rules) *Let assumptions made in Sect. 3.1 hold and let the source condition (Theorem 3.1) be satisfied at the* TV*-minimising solution* $u_{\mathrm{TV}}^\dagger$. *Let* $u_{\sigma_G,\gamma}$ *be a solution of* (3.8) *and let* α *be chosen such that*

$$\alpha(\sigma_G, \gamma) = O(\sigma_G + \sqrt{\gamma}).$$

Then,

$$D_{\mathrm{TV}}^{q^\dagger}(u_{\sigma_G,\gamma}, u_{\mathrm{TV}}^\dagger) = O(\sigma_G + \sqrt{\gamma}),$$

where $q^\dagger = L^*\mu^\dagger$ *is the subgradient from Theorem 3.1 and* $\sigma_G, \gamma > 0$ *are as defined in* (3.10).

Proof The proof is similar to [43, Thm. 3.9]. □

In a similar manner, we can obtain convergence rates for an a posteriori parameter choice rule known as the discrepancy principle [44–46]. Let f be the noisy data and $\delta > 0$ the amount of noise such that $\Phi(\bar{f}, f) \leq \delta$, where Φ is as defined in (3.5). In our case, $\delta = \frac{\sigma_G^2}{2} + \gamma$ by (3.10). The discrepancy principle amounts to selecting $\alpha = \alpha(f, \delta)$ such that

$$pg\alpha = \sup\{\alpha > 0\colon \Phi(Lu^\alpha, f) \leq \tau\delta\}, \tag{3.11}$$

where u^α is the regularised solution corresponding to the regularisation parameter α and $\tau > 1$ is a parameter.

Again, slightly modifying the proofs from [43], we obtain the following

Theorem 3.3 (Convergence rates under the discrepancy principle) *Let assumptions made in Sect. 3.1 hold and let the source condition (Theorem 3.1) be satisfied at the* TV*-minimising solution* $u_{\mathrm{TV}}^\dagger$. *Let* $u_{\sigma_G,\gamma}$ *be a solution of* (3.8) *with* α *chosen according to the discrepancy principle* (3.11). *Then,*

$$D_{\mathrm{TV}}^{q^\dagger}(u_{\sigma_G,\gamma}, u_{\mathrm{TV}}^\dagger) = O(\sigma_G + \sqrt{\gamma}),$$

where $q^\dagger = L^*\mu^\dagger$ is the subgradient from Theorem 3.1 and $\sigma_G, \gamma > 0$ are as defined in (3.10).

Proof The proof is similar to [43, Thm. 4.10]. □

4 Solving the Minimisation Problem

4.1 PDHG for Infimal Convolution Model

In practice, due to the joint convexity of the Kullback–Leibler divergence, we solve the minimisation problem (3.9), where we treat the reconstructed sample u and the Gaussian denoised image v jointly and, in addition, we impose lower and upper bound constraints on u and v by including the corresponding characteristic functions in the objective:

$$\min_{u,v} \alpha\,\mathrm{TV}(u) + \frac{1}{2\sigma_G^2}\|f - v\|_2^2 + D_{KL}(v, Lu) + \chi_{[l_1,l_2]^{2N}}([u, v]^T). \tag{4.1}$$

Note that the objective function in (4.1) is a sum of convex functions (the Kullback–Leibler divergence D_{KL} is jointly convex [47]), and therefore is itself convex. We then write the problem (4.1) as:

$$\min_w G(w) + \sum_{i=1}^m H_i(L_i w), \tag{4.2}$$

where we solve for $w = \begin{bmatrix} u \\ v \end{bmatrix}$, $m = 3$ and:

$$G(w) = \chi_{[l_1,l_2]^{2N}}\left(\begin{bmatrix} u \\ v \end{bmatrix}\right), \tag{4.3}$$

$$H_1(\cdot) = \frac{1}{2\sigma_G^2}\|\cdot - f\|_2, \qquad L_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}, \tag{4.4}$$

$$H_2(w) = D_{KL}(v, u), \qquad L_2 = \begin{bmatrix} L & 0 \\ 0 & 1 \end{bmatrix}, \tag{4.5}$$

$$H_3(\cdot) = \alpha\|\cdot\|_1, \qquad L_3 = \begin{bmatrix} \nabla_x & 0 \\ \nabla_y & 0 \\ \nabla_z & 0 \end{bmatrix}, \tag{4.6}$$

where L is the forward operator corresponding to the image formation model from Sect. 2.1.

Rather than solving the problem (4.2) directly, a common approach is to reformulate it as a saddle point problem using the Fenchel conjugate $G^*(y) = \sup_z \langle z, y\rangle - G(z)$. For proper, convex and lower semicontinuous function G, we have that $G^{**} = G$, so (4.2) can be written as the saddle point problem

$$\min_w \sup_{y_1,\ldots,y_m} G(w) + \sum_{i=1}^m \langle y_i, L_i x\rangle - H_i^*(y_i), \tag{4.7}$$

and by swapping the min and the sup and applying the definition of the convex conjugate G^*, one obtains the dual of (4.2):

$$\max_{y_1,\ldots,y_m} -G^*\left(-\sum_{i=1}^m L_i^* y_i\right) - \sum_{i=1}^m H_i^*(y_i). \tag{4.8}$$

The saddle point problem (4.7) is commonly solved using the primal–dual hybrid gradient (PDHG) algorithm [6,10,11], and by doing so, both the primal problem (4.2) and the dual (4.8) are solved. We apply the variant of PDHG from [48], which accounts for the sum of composite terms $H_i \circ L_i$. Given an initial guess for $(w_0, y_{1,0}, \ldots, y_{m,0})$ and the parameters $\sigma, \tau > 0$, and $\rho \in [\epsilon, 2-\epsilon]$ for some $\epsilon > 0$, each iteration $k \geq 0$ consists of the following steps:

1. $\tilde{w}_{k+1} := \mathrm{prox}_{\tau G}(w_k - \tau \sum_{i=1}^m L_i^* y_{i,k})$,
2. $w_{k+1} := \rho_k \tilde{w}_{k+1} + (1-\rho_k) w_k$,
3. $\forall i = 1, \ldots, m$:
   $\tilde{y}_{i,k+1} := \mathrm{prox}_{\sigma H_i^*}\left(y_{i,k} + \sigma L_i(2\tilde{w}_{k+1} - w_k)\right)$,
4. $\forall i = 1, \ldots, m$:
   $y_{i,k+1} := \rho \tilde{y}_{i,k+1} + (1-\rho) y_{i,k}$. (4.9)

where for a proper, lower semicontinuous, convex function G, $\mathrm{prox}_{\tau G}$ is its proximal operator, defined as:

$$\mathrm{prox}_{\tau G}(y) := \arg\min_x \left\{\frac{1}{2\tau}\|x - y\|_2^2 + G(x)\right\}. \tag{4.10}$$

The iterates $(w_k)_{k\in\mathbb{N}}$ and $(y_{i,k})_{k\in\mathbb{N}}$ $(i = 1, \ldots, m)$ are shown to converge if the parameters σ and τ are chosen such that $\sigma\tau\|\sum_{i=1}^m L_i^* L_i\| \leq 1$ (see [48, Theorem 5.3]). In step 3 in (4.9), we use Moreau's identity to obtain $\mathrm{prox}_{\sigma H_i^*}$ from $\mathrm{prox}_{H_i/\sigma}$:

$$\mathrm{prox}_{\sigma H_i^*}(y) + \sigma\,\mathrm{prox}_{H_i/\sigma}(y/\sigma) = y. \tag{4.11}$$

As a stopping criterion, one can use the primal–dual gap, i.e. the difference between the primal objective cost at the current iterate and the dual objective cost at the current (dual) iterate:

$$D_{pd}(w, y_1, \ldots, y_m)$$

$$= G(w) + \sum_{i=1}^{m} H_i(L_i w) + G^*(-\sum_{i=1}^{m} L_i^* y_i) + \sum_{i=1}^{m} H_i^*(y_i) \tag{4.12}$$

Due to strong duality, optimality is reached when the primal–dual gap is zero, so a practical stopping criterion is when the gap reaches a certain threshold set in advance.

Lastly, note that the optimisation is performed jointly over both u and v, which introduces a difficulty for the term $H_2(L_2 w)$ in Step 3 above, as this requires the proximal operator of the joint Kullback–Leibler divergence $D_{KL}(u, v)$. Similarly, the computation of the primal–dual gap in (4.12) requires the convex conjugate of the joint Kullback–Leibler divergence. We describe the details of these computations in Sects. 4.2 and 4.3, respectively.

4.2 Computing the Proximal Operator of the Joint Kullback–Leibler Divergence

When writing the optimisation problem in the form (4.2), it is common that the functions G and H_i $(i = 1, \ldots, m)$ are "simple", meaning that their proximity operators have a closed-form solution or can be easily computed with high precision. This is certainly true for G and H_1, but not obvious for the joint Kullback–Leibler divergence.

First, for discrete images $u = [u_1, \ldots, u_N]^T$, $[v_1, \ldots, v_N]^T$, the definition (3.6) becomes:

$$D_{KL}(v, u) = \sum_{j=1}^{N} u_j - v_j + v_j \log \frac{v_j}{u_j} \tag{4.13}$$

and then:

$$\begin{aligned}
\operatorname{prox}_{\gamma D_{KL}}(u^*, v^*) &= \arg\min_{u,v} D_{KL}(u, v) + \frac{1}{2\gamma} \left\| \begin{bmatrix} u \\ v \end{bmatrix} - \begin{bmatrix} u^* \\ v^* \end{bmatrix} \right\|_2^2 \\
&= \arg\min_{u,v} \sum_{j=1}^{N} u_j - v_j + v_j \log \frac{v_j}{u_j} \\
&\quad + \frac{1}{2\gamma} \left[(u_j - u_j^*)^2 + (v_j - v_j^*)^2 \right] \\
&= \sum_{j=1}^{N} \arg\min_{u_j, v_j} \Phi(u_j, v_j),
\end{aligned} \tag{4.14}$$

where we define the function Φ as:

$$\begin{aligned}
\Phi(u_j, v_j) &:= u_j - v_j + v_j \log \frac{v_j}{u_j} \\
&\quad + \frac{1}{2\gamma}[(u_j - u_j^*)^2 + (v_j - v_j^*)^2].
\end{aligned} \tag{4.15}$$

To find the minimiser of $\Phi(u_j, v_j)$, we let its gradient be equal to zero:

$$\begin{cases} \partial_{u_j} \Phi(u_j, v_j) = 0 \\ \partial_{v_j} \Phi(u_j, v_j) = 0 \end{cases} \iff \begin{cases} 1 - \frac{v_j}{u_j} + \frac{1}{\gamma}(u_j - u_j^*) = 0 \\ \log v_j - \log u_j + \frac{1}{\gamma}(v_j - v_j^*) = 0 \end{cases}. \tag{4.16}$$

In the second equation, we write u_j as a function of v_j, which we substitute in the first equation to obtain:

$$\begin{cases} 1 - e^{-\frac{1}{\gamma}(v_j - v_j^*)} + \frac{1}{\gamma}\left(v_j e^{\frac{1}{\gamma}(v_j - v_j^*)} - u_j^* \right) = 0 \\ u_j = v_j e^{\frac{1}{\gamma}(v_j - v_j^*)} \end{cases}. \tag{4.17}$$

The first equation is then solved using Newton's method, where the iteration is given by:

$$v_j^{(k+1)} = v_j^{(k)} - \frac{\gamma - \gamma e^{-\frac{1}{\gamma}(v_j^{(k)} - v_j^*)} + v_j^{(k)} e^{\frac{1}{\gamma}(v_j^{(k)} - v_j^*)} - u_j^*}{e^{-\frac{1}{\gamma}(v_j^{(k)} - v_j^*)} + (1 + \frac{1}{\gamma} v_j^{(k)}) e^{\frac{1}{\gamma}(v_j^{(k)} - v_j^*)}}. \tag{4.18}$$

4.3 Computing the Convex Conjugate of the Joint Kullback–Leibler Divergence

We compute the convex conjugate of the discrete joint Kullback–Leibler divergence $D_{KL}(v, u)$ in (4.13) for $u, v \in [l_1, l_2]^N$:

$$\begin{aligned}
D_{KL}^*(v^*, u^*) &= \sup_{v,u \in [l_1, l_2]^N} \left\langle \begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} u^* \\ v^* \end{bmatrix} \right\rangle - D_{KL}(v, u) \\
&= \sup_{v,u \in [l_1, l_2]^N} \sum_{j=1}^{N} u_j u_j^* + v_j v_j^* \\
&\quad - u_j + v_j - v_j \log \frac{v_j}{u_j} \\
&= \sum_{j=1}^{N} \sup_{v_j, u_j \in [l_1, l_2]} \Psi(v_j, u_j),
\end{aligned} \tag{4.19}$$

where Ψ is defined as:

$$\Psi(v_j, u_j) := u_j u_j^* + v_j v_j^* - u_j + v_j - v_j \log \frac{v_j}{u_j}. \tag{4.20}$$

To solve the optimisation problem on the last line in (4.19), we write the KKT conditions (where we use u, v instead of u_j, v_j to simplify the notation:

$$-\nabla \Psi(v, u) + \sum_{i=1}^{4} \mu_i \nabla g_i(v, u) = 0, \tag{4.21a}$$

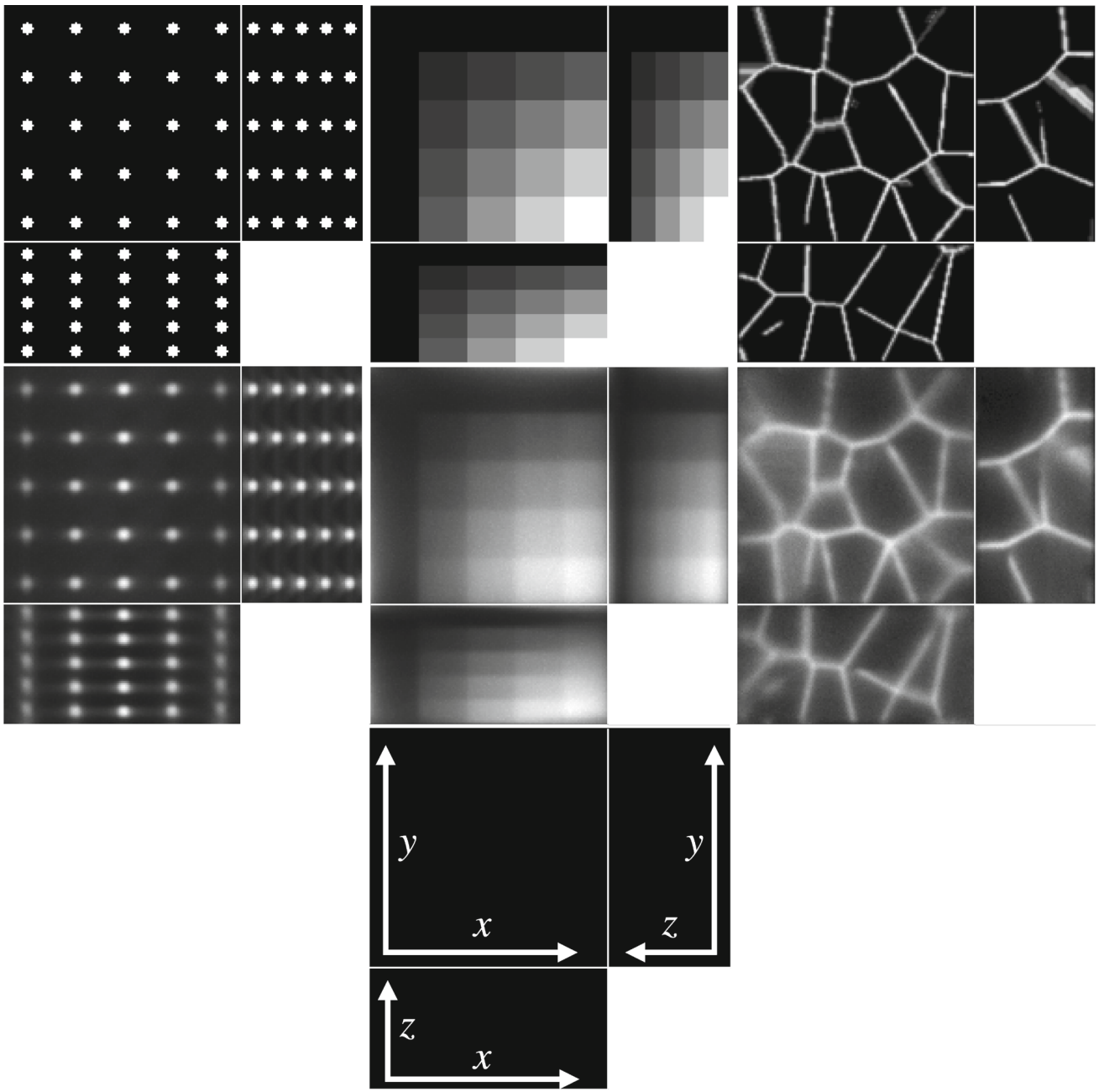


Fig. 8 Ground truth (top row) and measured images (middle row), shown using maximum intensity projections in each axis direction, except for the tissue images, where slices in each axis direction are shown. The axes of the plots are shown in the panel in the bottom row, with the missing axis in each panel being the direction in which the maximum intensity projection (or slice) is taken

$$g_i(v,u) \le 0, \quad \forall i = 1,\dots,4, \tag{4.21b}$$
$$\mu_i \ge 0, \quad \forall i = 1,\dots,4, \tag{4.21c}$$
$$\mu_i g_i(v,u) = 0, \quad \forall i = 1,\dots,4. \tag{4.21d}$$

where the functions g_i correspond to the bound constraints:

$$g_1(v,u) = u - l_2; \tag{4.22a}$$
$$g_2(v,u) = v - l_2; \tag{4.22b}$$
$$g_3(v,u) = -u + l_1; \tag{4.22c}$$
$$g_4(v,u) = -v + l_1; \tag{4.22d}$$

Noting that (4.21a) is equivalent to:

$$-u^* + 1 - \frac{v}{u} + \mu_1 - \mu_3 = 0, \tag{4.23a}$$
$$-v^* + \log v - \log u + \mu_2 - \mu_4 = 0, \tag{4.23b}$$

we solve the last two equations by using the complementarity conditions (4.21d) for cases when the Lagrange multipliers μ_i are zero or nonzero.

5 Numerical Results

In this section, we describe a number of numerical experiments that illustrate the performance of our deconvolution method. We start with four examples of simulated data, where we are able to quantify the reconstructed image in relation to the known ground truth image. Then, we show how our method performs on microscopy data, where we reconstruct an image of spherical beads and a sample of a Marchantia thallus. In the experiments with microscopy data, we compare our method with two standard approaches of performing shift-invariant deconvolution, one where the convolution kernel is the detection PSF and one where the convolution kernel is the point-wise multiplication of the detection PSF with the light-sheet.

5.1 Simulated Data

We consider four images of size $128\times125\times64$: a $5\times5\times5$ grid of beads where the effect of the light-sheet in the z coordinate and the shape of the objective PSF are noticeable, a piecewise constant image of "steps" where the Poisson noise affects each step differently based on intensity, and an image that replicates realistic biological samples of tissue. These are shown in the top row of Fig. 8.

To obtain the measured data, we proceed as follows. Given the ground truth image u_0, the forward operator described in Sect. 2.1 is applied to obtain the blurred image Lu_0. The parameters for the forward model are taken to be those of the microscope used in the experimental setup and are given in Table 2. Then, the image is corrupted with a mixture of Poisson and Gaussian noise. For the vectorised image Lu_0, at each pixel $i = 1, \ldots, N$, the Poisson noise component follows the Poisson distribution with parameter $(Lu_0)_i$ and the additive Gaussian component has zero mean and standard deviation $\sigma_G = 10$. The original image, which has intensity in [0, 1] is scaled so that the intensity of Lu_0 is in [0, 2000], to replicate a realistic scenario for the Poisson noise intensity. The resulting simulated measured data is shown in the bottom row of Fig. 8.

We compare the reconstruction obtained using the proposed approach, which we will refer to as LS-IC (light-sheet-infimal convolution), with the reconstructions obtained by using an L^2 data fidelity term instead of the infimal convolution term, or using a convolution operator corresponding to the objective PSF instead of the light-sheet forward model from Sect. 2.1. Specifically, we compare the solution of (4.1)

Table 2 Forward model parameters used in Sect. 5

Parameter	Value	Description, units
n	1.35	Refractive index
NA_h	1	Numerical aperture (objective lens)
NA_l	0.25	Numerical aperture (light-sheet)
λ_h	0.525	Wave length (objective lens), µm
λ_l	0.488	Wave length (light-sheet), µm
px_x	0.3250	Pixel size (x), µm
px_y	0.3250	Pixel size (y), µm
$step_z$	1	Light-sheet step size (z), µm

with the solutions to the following problems, all solved using PDHG as described in Sect. 4:

$$\min_u \alpha\,\mathrm{TV}(u) + \frac{1}{2\sigma_G^2}\|f - Hu\|_2^2 + \chi_{[0,B]^{2N}}([u, v]^T), \tag{PSF-L2}$$

$$\min_{u,v} \alpha\,\mathrm{TV}(u) + \frac{1}{2\sigma_G^2}\|f - v\|_2^2 + D_{KL}(v, Hu) + \chi_{[0,B]^{2N}}([u, v]^T), \tag{PSF-IC}$$

$$\min_u \alpha\,\mathrm{TV}(u) + \frac{1}{2\sigma_G^2}\|f - Lu\|_2^2 + \chi_{[0,B]^{2N}}([u, v]^T), \tag{LS-L2}$$

where H is the convolution operator with the detection objective PSF h_z as given in (2.14).

For each test image and each method above, the PDHG parameters ρ and σ used are given in Table 3 and τ is set to $\tau = 1/\sigma\|\sum_{i=1}^m L_i^* L_i\|$ to ensure convergence according to Theorem 5.3 in [48]. As a stopping criterion, we used the primal–dual gap (4.12), normalised by the number of pixels N and the dynamic range of the measured image f:

$$\tilde{D}_{pd} = \frac{D_{pd}}{N \cdot \max_{j=1,\ldots,N} f_j}, \tag{5.1}$$

with a threshold of 10^{-6} and a maximum number of 10,000 iterations.

The results of the four methods applied to the test images are given in Fig. 9 and quantitative results are given in Table 4. For each test image and each method, the regularisation parameter has been chosen to optimise the normalised l^2 error and the structural similarity index (SSIM), respectively.

Table 3 Values of the PDHG parameters ρ and σ used in the numerical experiments with simulated data

Method	LS-IC			LS-L2			PSF-IC			PSF-L2		
Image	Beads	Steps	Tissue	Beads	Steps	Tissue	Beads	Steps	Tissue	Beads	Steps	Tissue
ρ	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9
σ	0.0001	0.0001	0.00001	0.0001	0.001	0.0001	0.0001	0.0001	0.0001	0.0001	0.001	0.0001

We note that PSF-L2 and PSF-IC perform particularly poorly, highlighting the importance of an accurate representation of the image formation model instead of simply using the detection objective PSF as the forward operator. Comparing LS-IC and LS-L2, we see better results when using the infimal convolution data fidelity for the beads and the steps image, both visually and quantitatively. The deblurring is performed better on the beads image, while on the steps image we see a better denoising effect, especially along the edges in the image. For the tissue image, both fidelities give comparable results, but as we see in Fig. 10, when the ground truth is not known, choosing α using the discrepancy principle gives a better result for the infimal convolution model.

The reconstructions shown in Fig. 10 are obtained by applying the discrepancy principle corresponding to each method. For LS-IC, we choose a value of α such that it satisfies a variation of the discrepancy principle given in (3.11), where we enforce that the single noise fidelities are bounded by their respective noise bounds, rather than the sum of the fidelities being bounded by the sum of the noise bounds, as stated in (3.11). While both versions give good results, we found the former to give more accurate reconstructions. Here, the bound on the Poisson noise is set to $\frac{1}{2}$, motivated by the following lemma from [49], which gives the expected value of the Kullback–Leibler divergence:

Lemma 5.1 *Let Y_β be a Poisson random variable with expected value β and consider the function:*

$$F(Y_\beta) = 2\left\{Y_\beta \log\left(\frac{Y_\beta}{\beta}\right) + \beta - Y_\beta\right\}.$$

Then, for large β, the following estimate of the expected value of $F(Y_\beta)$ holds:

$$\mathbb{E}[F(Y_\beta)] = 1 + \mathcal{O}\left(\frac{1}{\beta}\right).$$

One last observation worth making about the results in Figs. 9 and 10 is about the square shape of the reconstructed beads (the first column of both figures). By looking carefully at the ground truth bead image in Fig. 8, one can see that the beads are almost square to begin with, due to the small dimensions of the image. The finer details that make them appear round are lost in the blurring process which, in combination with the total variation regulariser used in the deconvolution algorithm, leads to this detail not being present in the reconstruction, thus making them square. Moreover, the sharpening of their edges is an expected effect of the total variation regularisation, which could be avoided by using a different regularisation technique. However, this is beyond the scope of this article.

The experiments were run using Matlab version R2020b Update 2 (9.9.0.1524771) 64-bit in Scientific Linux 7.9 on a machine with Intel Xeon E5-2680 v4 2.40 GHz CPU, 256 GB memory and Nvidia P100 16 GB GPU. The running times, averaged over 5 runs for each method and each image, are given in Table 5.

5.2 Light-Sheet Data

In this section, we show the results of applying LS-IC to a cropped portion of the full resolution images in Fig. 2. Specifically, we select a cropped beads image of $1127 \times 111 \times 100$ voxels and a cropped Marchantia image of $1127 \times 156 \times 100$ voxels.

For comparison, we also run PSF-L2 on the same images. In addition, we run an alternative light-sheet deconvolution method, where we perform shift-invariant deconvolution using a PSF $\bar{h}$ obtained by point-wise multiplication of the detection PSF h_z in (2.14) and the light-sheet l, effectively clipping h by the width of the light-sheet. Therefore, the problem we solve, which we denote by PSF-L2-clip is:

$$\min_u \alpha\,\mathrm{TV}(u) + \frac{1}{2\sigma_G^2}\|f - \bar{H}u\|_2^2 + \chi_{[0,B]^{2N}}([u,v]^T), \tag{PSF-L2-clip}$$

where $\bar{H}$ is the convolution operator with the PSF $\bar{h} = h_z \cdot l$. A justification of this method is given by a simplified image formation model where we assume that the light-sheet has constant width (in the z direction) and constant intensity throughout the full sample, or in a region of interest where deconvolution is performed, as it is done for example in [21].

We run each method on both images for up to 6000 iterations, with a normalised primal–dual gap of 10^{-6} as a stopping criterion. The parameters for the image formation model used are the same as in Table 2 and the PDHG parameters are given in Table 6.

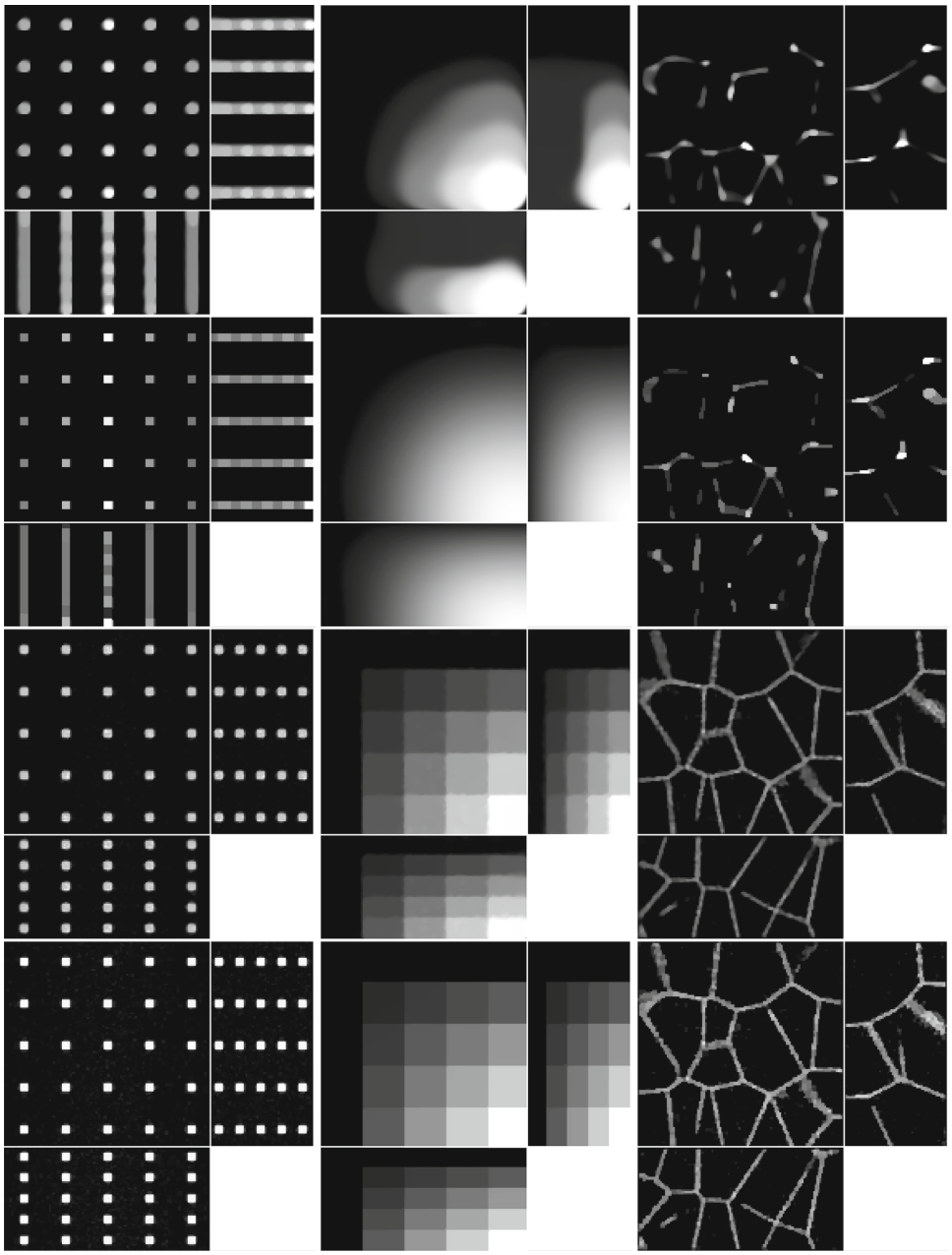


Fig. 9 Reconstruction on simulated data with regularisation parameter α such that best MSE is achieved for each method and each image. Shown as maximum intensity projections, except for tissue, where slices in each direction in the centre of the sample are shown. The axes are as shown in the bottom row of Fig. 8. *First row*: PSF-L2. *Second row*: PSF-IC. *Third row*: LS-L2. *Fourth row*: LS-IC

Table 4 Results of the numerical experiments on simulated data, with the regularisation parameter α chosen to optimise the normalised l_2 error and the SSIM, respectively

Image	Beads		Steps		Tissue	
Error metric	l_2	SSIM	l_2	SSIM	l_2	SSIM
PSF-L2	1.74	0.845	0.499	0.561	1.57	0.592
PSF-IC	1.54	0.844	0.324	0.659	1.65	0.582
LS-L2	0.282	0.982	0.055	0.971	0.301	0.951
LS-IC	0.258	0.983	0.012	0.998	0.349	0.931

The results of the deconvolution are shown in Figs. 11 and 13 for the beads image and the Marchantia image, respectively. In both figures, we first show the position of the light-sheet in the first row (due to the cropping, this is no longer centred), the measured data in the second row, followed by the PSF-L2, the PSF-L2-clip and the LS-IC reconstructions on the third, fourth and fifth rows, respectively. The regularisation parameter α was chosen in all four cases visually such that a balance is achieved between the amount of regularisation and the noise in the reconstruction.

In the beads image in Fig. 11, we note that LS-IC performs better than PSF-L2 and PSF-L2-clip at reversing the effect of the light-sheet. This is most obvious in the zy plane on the

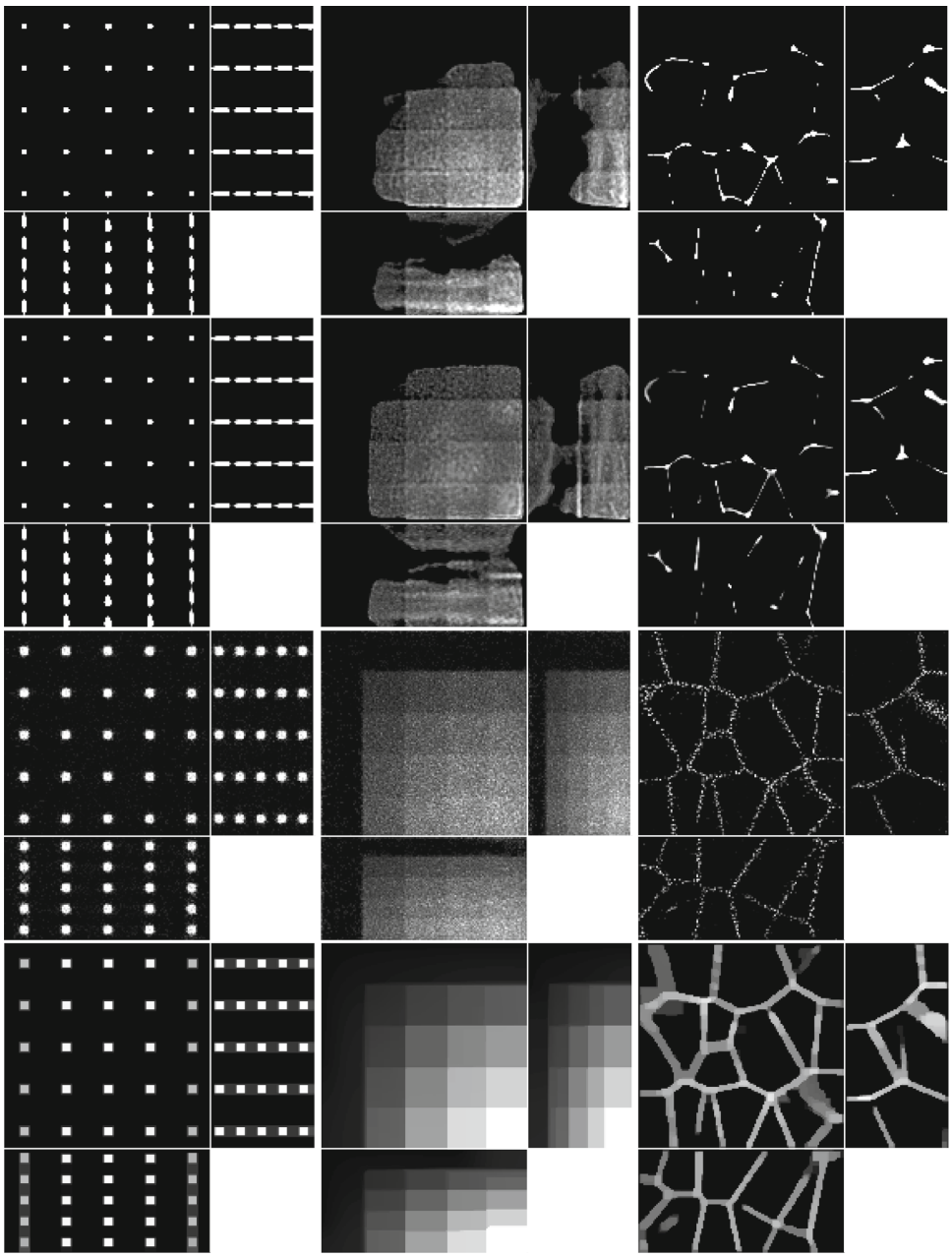


Fig. 10 Reconstruction on simulated data with regularisation parameter α chosen to satisfy the discrepancy principle (3.11). Shown as maximum intensity projections, except for tissue, where slices in each direction in the centre of the sample are shown. The axes are as shown in the bottom row of Fig. 8. *First row*: PSF-L2. *Second row*: PSF-IC. *Third row*: LS-L2. *Fourth row*: LS-IC

Table 5 Running times for each method and each simulated test image, averaged over 5 runs, in seconds

Image	Beads	Steps	Tissue
PSF-L2	233	1793	903
PSF-IC	689	1077	1805
LS-L2	2913	2194	2273
LS-IC	972	601	850

The minimisation is stopped when the primal–dual gap is lower than 10^{-6} or the maximum number of 10,000 iterations is reached

right-hand side of the image, where the length of the beads in the z direction has been reduced to a greater extent than in the PSF-L2 and the PSF-L2-clip reconstructions. In addition, the beads appear less blurry in the LS-IC reconstruction in the right-hand side of the xy plane. We also note that PSF-L2-clip fails to properly reverse the effects of the optical aberrations in the beads. This is not unexpected, as the information related to the aberrations is lost when the detection PSF is clipped by setting its upper and lower extremities to zero. The extent to which this happens depends on the width of the light-sheet: as the light-sheet becomes wider, the overall PSF approaches the detection PSF, in which case the deconvolved image will be the same as the reconstruction

Table 6 Values of the PDHG parameters ρ and σ used in the numerical experiments with real data

Method	LS-IC		PSF-L2		PSF-L2-clip	
Image	Beads	Marchantia	Beads	Marchantia	Beads	Marchantia
ρ	0.5	0.7	0.9	0.9	0.9	0.9
σ	0.0001	0.0001	0.01	0.001	0.01	0.001

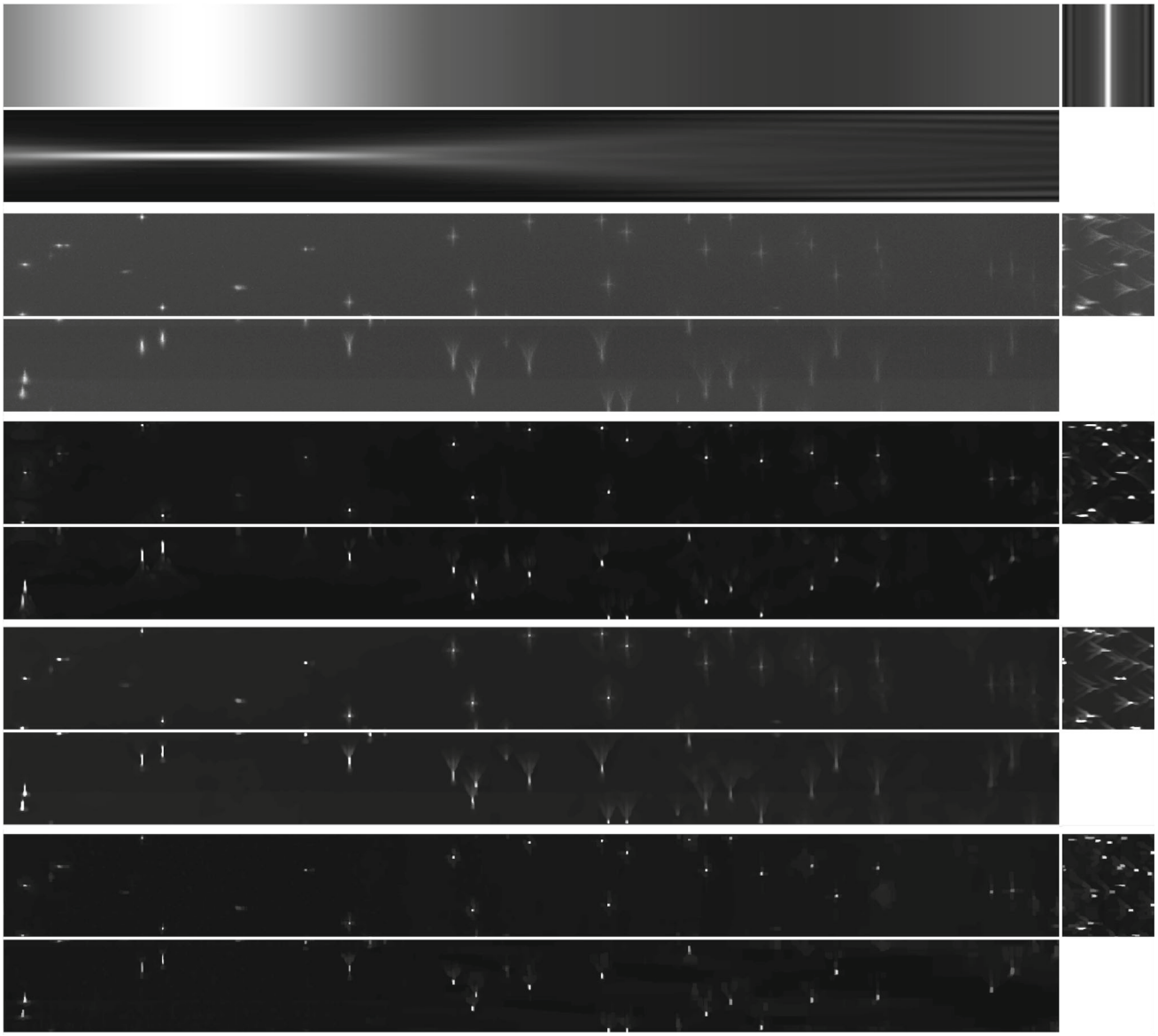


Fig. 11 Reconstruction results for the light-sheet bead image, shown as maximum intensity projections. The axes are as shown in the bottom row of Fig. 8. *First row*: The fitted light-sheet profile. *Second row*: The data. *Third row*: PSF-L2 with $\alpha = 0.1$. *Fourth row*: PSF-L2-clip with $\alpha = 0.7943$. *Fifth row*: LS-IC with $\alpha = 0.0046$

using PSF-L2. We show the bead images in 3D in Fig. 12, where the effect of the deconvolution in the z direction is more significant in the LS-IC reconstruction than in both the PSF-L2 and the PSF-L2-clip reconstructions, namely the beads are shorter in the z direction.

In the Marchantia reconstruction in Fig. 13, we see a similar effect of better sharpening in the z direction, most easily seen in the right-hand side and bottom projections in each panel (maximum intensity projections in the zy and the xz planes, respectively). In particular, we see additional artefacts in the PSF-L2-clip reconstruction: horizontal lines (parallel with the xy plane), likely due to the clipping of the detection PSF. Moreover, the 3D rendering of the Marchantia sample in Fig. 14 shows smoother cell edges in the LS-IC reconstruction compared to the other methods. Specifically, the PSF-L2 reconstruction contains reconstruction artefacts that are non-existent in the LS-IC reconstruction (indicated by the yellow arrows), while the PSF-L2-clip reconstruction contains areas where the blur has not been fully removed (for example at the same locations indicated by the yellow arrows), where the edges are not as sharp as in the LS-IC reconstruction.

Lastly, we reiterate that the strength of our proposed method is given by the physically accurate modelling of the interaction between the detection PSF and the light-sheet. This allows one to model the optical aberrations as part of the detection PSF (with no requirements on how this should be done), as well as the spatial dependence of the width and the intensity of the light-sheet and to combine them in an image formation model that does not require approximating using a light-sheet with constant width and intensity. As we see in the numerical experiments shown in this section, such approximation, while faster and less expensive computationally, leads to loss of information and results that are at most locally accurate.

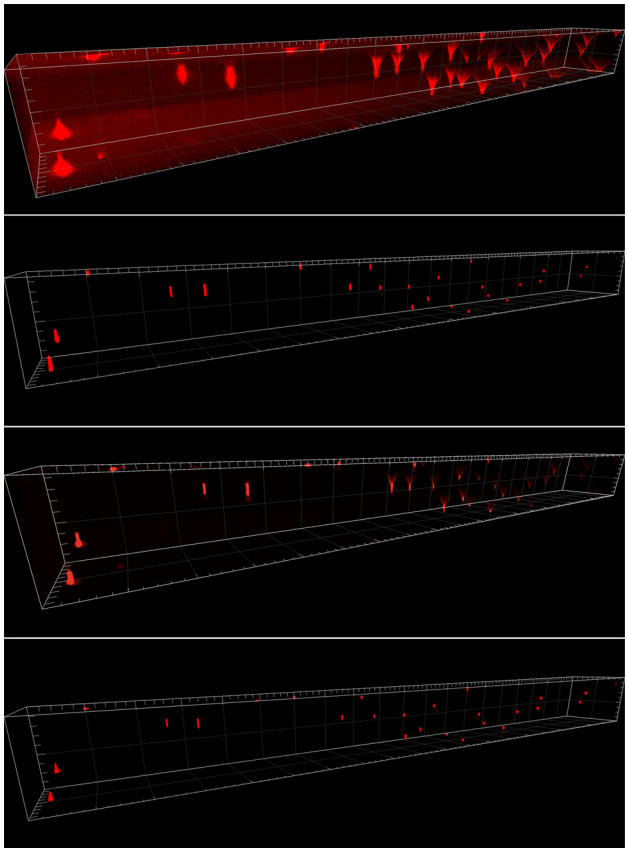


Fig. 12 3D rendering of the beads data and reconstruction images using Imaris Viewer 9.7.2. *First row*: The data. *Second row*: PSF-L2 with $\alpha = 0.1$. *Third row*: PSF-L2-clip with $\alpha = 0.7943$. *Fourth row*: LS-IC with $\alpha = 0.0046$

6 Conclusion

In this paper, we introduced a novel method for performing deconvolution for light-sheet microscopy. We start by modelling the image formation process in a way that replicates the physics of a light-sheet microscope, which is achieved by explicitly modelling the interaction of the illumination light-sheet and the detection objective PSF. Moreover, the optical aberrations in the system are modelled using a linear combination of Zernike polynomials in the pupil function of the detection PSF, fitted to bead data using a least squares procedure. We then formulate a variational model taking into account the image formation model as the forward operator and a combination of Poisson and Gaussian noise in the data. The model combines a total variation regularisation term and a fidelity term that is an infimal convolution between an L^2 term and the Kullback–Leibler divergence, introduced in [9]. In addition, we establish convergence rates with respect to the noise and we introduce a discrepancy principle for selecting the regularisation parameter α in the mixed noise setting. We solve the resulting inverse problem by applying the PDHG algorithm in a non-trivial way.

The results in the numerical experiments section show that our method, LS-IC, outperforms simpler approaches to deconvolution of light-sheet microscopy data, where one does not take into account the variability of the overall PSF introduced by the light-sheet excitation, or the combination of Gaussian and Poisson noise. In particular, numerical experiments with simulated data show superior reconstruction quality in terms of the normalised l^2 error and the structural similarity index, not only by optimising over the regularisation parameter α given the ground truth, but also with an a posteriori choice of α using the stated discrepancy principle. On bead data, the reconstruction obtained using LS-IC shows a more significant reduction of the blur in the z direction compared to PSF-L2, where the light-sheet variations and the Poisson noise are not taken into account. Moreover, reconstruction of a Marchantia sample with LS-

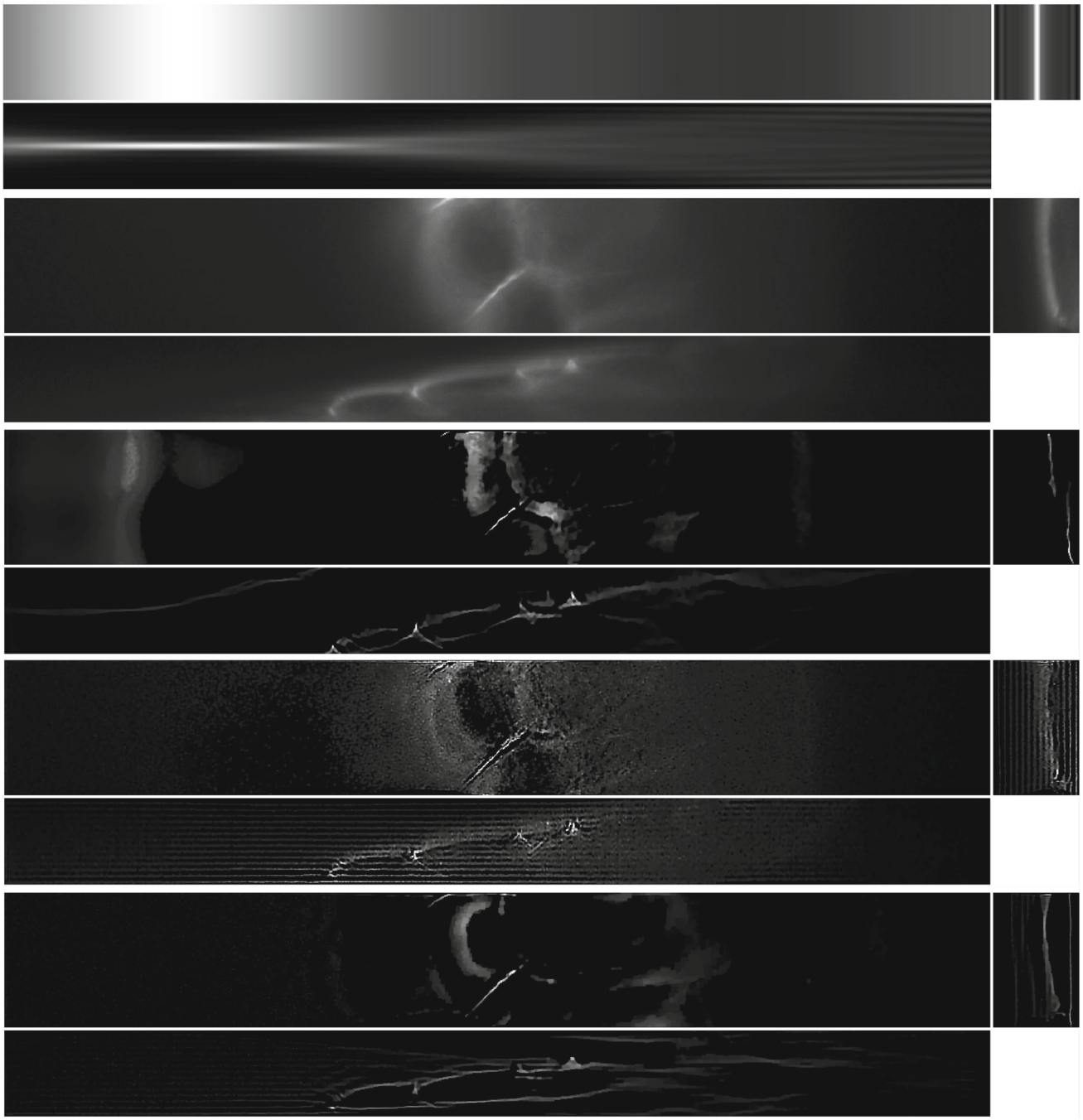


Fig. 13 Reconstruction results for the Marchantia sample, shown as slices in each direction in the centre of the sample. The axes are as shown in the bottom row of Fig. 8. *First row*: The fitted light-sheet profile. *Second row*: The data. *Third row*: PSF-L2 with $\alpha = 0.1$. *Fourth row*: PSF-L2-clip with $\alpha = 0.1$. *Fifth row*: LS-IC with $\alpha = 0.0005$

IC shows fewer artefacts than the PSF-L2 reconstruction, as well as sharper cell edges and smoother cell membranes.

Future work includes applying this technique to a broader range of samples and using it to answer questions of biological interest. To do so, we see a number of potential future directions that this work can take:

1. Adapting the discrepancy principle given in (3.11) for choosing the regularisation parameter α to real data sets, like the ones in Sect. 5.2.
2. Improving the running time of the method potentially by means of randomised approaches.
3. Investigating other regularisation terms.

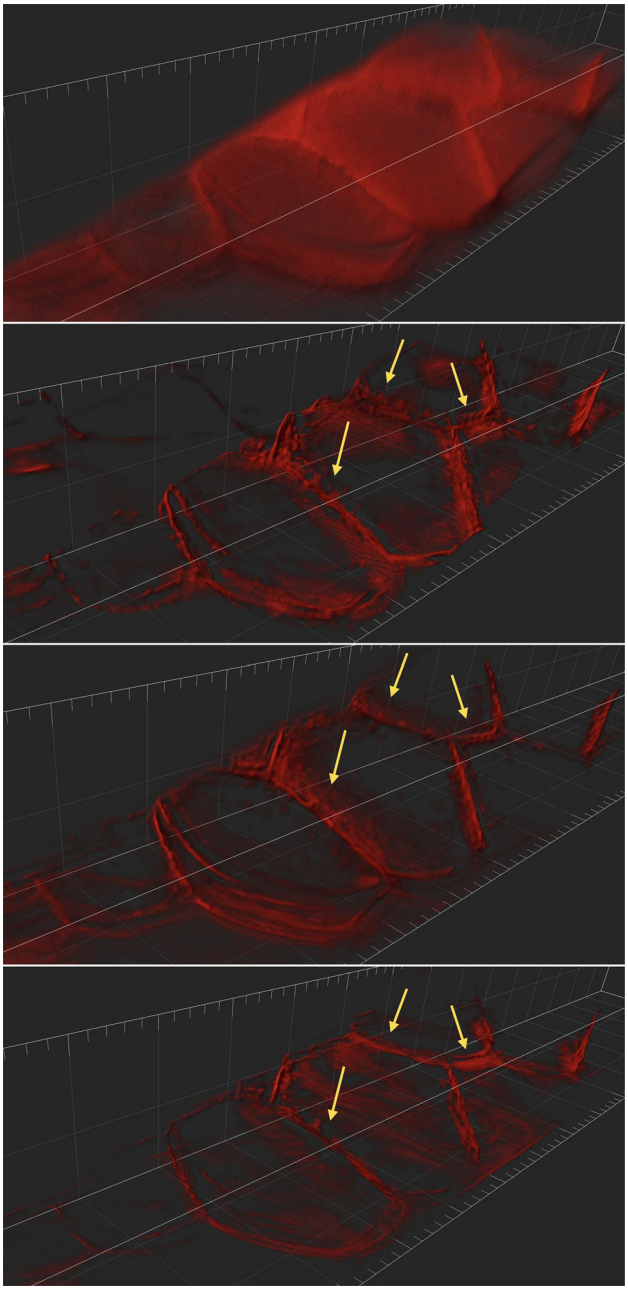


Fig. 14 3D rendering of the Marchantia data and reconstruction images using Imaris Viewer 9.7.2. *First row*: The data. *Second row*: PSF-L2 with $\alpha = 0.1$. *Third row*: PSF-L2-clip with $\alpha = 0.1$. *Fourth row*: LS-IC with $\alpha = 0.0005$

4. Making the technique available to other users as a more user-friendly tool.

Acknowledgements BT and LM gratefully acknowledge the funding by Isaac Newton Trust/Wellcome Trust ISSF/University of Cambridge Joint Research Grants Scheme and EPSRC EP/R025398/1. MOL and LM also thank the Gatsby Charitable Foundation for financial support. YK acknowledges financial support of the EPSRC (Fellowship EP/V003615/1), the Cantab Capital Institute for the Mathematics of Information at the University of Cambridge and the National Physical Laboratory. CBS acknowledges support from the Philip Leverhulme Prize, the Royal Society Wolfson Fellowship, the EPSRC Grants EP/S026045/1 and EP/T003553/1, EP/N014588/1, EP/T017961/1, the Wellcome Innovator Award RG98755, the Leverhulme Trust project Unveiling the invisible, the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 777826 NoMADS, the Cantab Capital Institute for the Mathematics of Information and the Alan Turing Institute. Imaging was performed at the Microscopy Facility of the Sainsbury Laboratory Cambridge University. We thank Dr. Alessandra Bonfanti and Dr. Sarah Robinson for providing the Marchantia sample and Prof. Sebastian Schornack and Dr. Giulia Arsuffi (Sainsbury Laboratory Cambridge University) for provision of the line of Marchantia used. We also acknowledge the support of NVIDIA Corporation with the donation of two Quadro P6000, a Tesla K40c and a Titan Xp GPU used for this research. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Pawley, J.: Handbook of Biological Confocal Microscopy, vol. 236. Springer, Berlin (2006)
2. Method of the year 2014. Nature Methods **12**(1), 1 (2015)
3. McNally, J.G., Karpova, T., Cooper, J., Conchello, J.A.: Three-dimensional imaging by deconvolution microscopy. Methods A Companion Methods Enzymol. **19**(3), 373–385 (1999)
4. Starck, J., Pantin, E., Murtagh, F.: Deconvolution in astronomy: a review. Publ. Astron. Soc. Pac. **114**(800), 1051–1069 (2002)
5. Sarder, P., Nehorai, A.: Deconvolution methods for 3-D fluorescence microscopy images. IEEE Signal Process. Mag. **23**(3), 32–45 (2006)
6. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numer. **25**, 161–319 (2016)
7. Denis, L., et al.: Fast approximations of shift-variant blur. Int. J. Comput. Vis. **115**(3), 253–278 (2015)
8. Debarnot, V., Escande, P., Weiss, P.: A scalable estimator of sets of integral operators. Inverse Probl. **35**(10), 105011 (2019)
9. Calatroni, L., De Los Reyes, J.C., Schönlieb, C.-B.: Infimal convolution of data discrepancies for mixed noise removal. SIAM J. Imaging Sci. **10**(3), 1196–1233 (2017)

10. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**(4), 1015–1046 (2010)
11. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
12. Nagy, J.G., O'Leary, D.P.: Restoring images degraded by spatially variant blur. SIAM J. Sci. Comput. **19**(4), 1063–1082 (1998)
13. Hadj, S.B., Blanc-Féraud, L., Aubert, G.: Space variant blind image restoration. SIAM J. Imaging Sci. **7**(4), 2196–2225 (2014)
14. Hirsch, M., Sra, S., Schölkopf, B., Harmeling, S.: Efficient filter flow for space-variant multiframe blind deconvolution. In: Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition. Max-Planck-Gesellschaft, Piscataway, NJ, USA, pp. 607–614. IEEE (2010)
15. O'Connor, D., Vandenberghe, L.: Total variation image deblurring with space-varying kernel. Comput. Optim. Appl. **67**(3), 521–541 (2017)
16. Yanny, K., Monakhova, K., Shuai, R.W., Waller, L.: Deep learning for fast spatially varying deconvolution. Optica **9**(1), 96–99 (2022)
17. Temerinac-Ott, M., et al.: Multiview deblurring for 3-D images from light-sheet-based fluorescence microscopy. IEEE Trans. Image Process. **21**(4), 1863–1873 (2012)
18. Preibisch, S., et al.: Efficient Bayesian-based multiview deconvolution. Nat. Methods **11**(6), 645–648 (2014)
19. Ancora, D., Furieri, T., Bonora, S., Bassi, A.: Spinning pupil aberration measurement for anisoplanatic deconvolution. Opt. Lett. **46**(12), 2884–2887 (2021)
20. Furieri, T., et al.: Aberration measurement and correction on a large field of view in fluorescence microscopy. Biomed. Opt. Express **13**(1), 262–273 (2022)
21. Becker, K., et al.: Deconvolution of light sheet microscopy recordings. Sci. Rep. **9**(1), 1–14 (2019)
22. Guo, M. et al.: Rapid image deconvolution and multiview fusion for optical microscopy. Nat. Biotechnol. (2020)
23. Zhang, Z., et al.: 3D Hessian deconvolution of thick light-sheet z-stacks for high-contrast and high-SNR volumetric imaging. Photon. Res. **8**(6), 1011–1021 (2020)
24. Cueva, E., et al.: Mathematical modeling for 2D lightsheet fluorescence microscopy image reconstruction. Inverse Probl. **36**(7), 075005 (2020)
25. Zhang, J., et al.: Bilinear constraint based ADMM for mixed Poisson–Gaussian noise removal. Inverse Probl. Imaging **15**(2), 339–366 (2021)
26. Hanser, B.M., Gustafsson, M.G., Agard, D.A., Sedat, J.W.: Phase-retrieved pupil functions in widefield fluorescence microscopy. J. Microsc. **216**(1), 32–48 (2004)
27. Stokseth, A.: Properties of a defocused optical system. J. Opt. Soc. Am. **59**(10), 1314–1321 (1969)
28. Soulez, F., Hiébaut, E.T., Ourneur, Y.T., Enis, L.D.: Déconvolution aveugle en microscopie de fluorescence 3D. GRETSI (2013)
29. Paxman, R.G., Schulz, T.J., Fienup, J.R.: Joint estimation of object and aberrations by using phase diversity. J. Opt. Soc. Am. A **9**(7), 1072 (1992)
30. Petrov, P.N., Shechtman, Y., Moerner, W.E.: Measurement based estimation of global pupil functions in 3D localization microscopy. Opt. Express **25**(7), 7945 (2017)
31. Wyant, J.C., Creath, K.: Basic wavefront aberration theory for optical metrology. Appl. Opt. Opt. Eng. **XI**, 11–53 (1992)
32. Burger, M., Osher, S.: A guide to the TV zoo. In: Burger, M., Osher, S. (eds.) Level-Set and PDE-based Reconstruction Methods. Springer, Berlin (2013)
33. Hohage, T., Werner, F.: Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data. Numer. Math. **123**(4), 745–779 (2013)
34. Hohage, T., Werner, F.: Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. Inverse Probl. **32**(9), 093001 (2016)
35. Lanza, A., Morigi, S., Sgallari, F., Wen, Y.-W.: Image restoration with Poisson–Gaussian mixed noise. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. **2**, 12–24 (2014)
36. Clason, C., Lorenz, D.A., Mahler, H., Wirth, B.: Entropic regularization of continuous optimal transport problems. J. Math. Anal. Appl. **494**(1), 124432 (2021)
37. Bennett, C., Sharpley, R.: Interpolation of Operators. Pure and Applied Mathematics, vol. 129. Academic Press, Boston (1988)
38. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
39. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1), 259–268 (1992)
40. Benning, M., Burger, M.: Modern regularization methods for inverse problems. Acta Numer. **27**, 1–111 (2018)
41. Burger, M., Osher, S.: Convergence rates of convex variational regularization. Inverse Probl. **20**(5), 1411 (2004)
42. Resmerita, E., Anderssen, R.S.: Joint additive Kullback–Leibler residual minimization and regularization for linear inverse problems. Math. Methods Appl. Sci. **30**(13), 1527–1544 (2007)
43. Bungert, L., Burger, M., Korolev, Y., Schönlieb, C.-B.: Variational regularisation for inverse problems with imperfect forward operators and general noise models. Inverse Probl. **36**(12), 125014 (2020)
44. Morozov, V.A.: On the solution of functional equations by the method of regularisation. Soviet Math. Dokl. **7**, 414–417 (1966)
45. Engl, H., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Springer, Berlin (1996)
46. Sixou, B., Hohweiller, T., Ducros, N.: Morozov principle for Kullback–Leibler residual term and Poisson noise. Inverse Probl. Imaging **12**(3), 607–634 (2018)
47. Lindblad, G.: Entropy, information and quantum measurements. Commun. Math. Phys. **33**(4), 305–322 (1973)
48. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**(2), 460–479 (2013)
49. Zanella, R., Boccacci, P., Zanni, L., Bertero, M.: Efficient gradient projection methods for edge-preserving removal of Poisson noise. Inverse Probl. **25**(4), 045010 (2009)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Bogdan Toader received a Ph.D. in mathematics from the University of Oxford in 2020. He joined the Cambridge Advanced Imaging Centre at the University of Cambridge for a postdoctoral position until 2021 and is currently a postdoctoral researcher at Yale University in the Department of Statistics and Data Science and the Quantitative Biology Institute. His interests include inverse problems, image processing, numerical analysis and their applications to science.

Jérôme Boulanger received a Ph.D. (2007) in Telecommunication and Signal Processing from the University of Rennes I, France. He joined the French National Centre for Scientific Research as a researcher scientist at the Institut Curie, Paris, France in 2011. Since 2015, he is an investigator scientist at the Laboratory of Molecular Biology at the Medical Research Council in Cambridge, UK.

Yury Korolev is an EPSRC postdoctoral fellow at the Department of Applied Mathematics and Theoretical Physics at the University of Cambridge and a Research Fellow at Hughes Hall. Prior to joining Cambridge, he has worked at the Universities of Lübeck and Münster. Yury's interests are applied analysis, mathematical foundations of machine learning, inverse problems and imaging.

Martin O. Lenz is a Senior Research Associate at the University of Cambridge with 17 years of experience designing bespoke optical instruments and more than 12 years developing high-end optical microscopes. He worked at some of the world's leading groups for application of super-resolution microscopy (P. French, Imperial College London) and IINS Bordeaux where he successfully implemented point-scanning STED super-resolution microscopy with single molecule-based super-resolution on the same microscope platform. In recent years, he has successfully developed multiple light-sheet microscopes for application in biological sciences at the Cambridge Advanced Imaging Centre. His expertise centres around hardware and software development for advanced light microscopy and integrating new and existing technologies for successfully pushing the boundaries of microscopy.

James Manton studied Natural Sciences and received BA, MRes and PhD degrees from the University of Cambridge. He is a Senior Investigator Scientist at the MRC Laboratory of Molecular Biology, where his research interests include live-cell microscopy and imaging-related inverse problems.

Carola-Bibiane Schönlieb graduated from the Institute for Mathematics, University of Salzburg (Austria) in 2004. From 2004 to 2005 she held a teaching position in Salzburg. She received her PhD degree from the University of Cambridge (UK) in 2009. After one year of postdoctoral activity at the University of Göttingen (Germany), she became a Lecturer at Cambridge in 2010, promoted to Reader in 2015 and promoted to Professor in 2018. Since 2011 she is a fellow of Jesus College Cambridge and since 2016 a fellow of the Alan Turing Institute, London. She currently is Professor of Applied Mathematics at the University of Cambridge, where she is head of the Cambridge Image Analysis group and co-Director of the EPSRC Cambridge Mathematics of Information in Healthcare Hub. Her current research interests focus on variational methods, partial differential equations and machine learning for image analysis, image processing and inverse imaging problems.

Leila Mureşan received a Ph.D. in Engineering (2010) from the Johannes Kepler University, Linz, Austria. She was a postdoctoral researcher at the Institute of Biology at the Ecole Normale Supérieure in Paris and at Centre de Génétique Moléculaire in Gif-sur-Yvette, France. In 2014, she joined the Cambridge Advanced Imaging Centre, where she focuses on light microscopy image analysis solutions. Since 2018, she is a Research Software Engineer EPSRC Fellow at the Department of Physiology Development and Neuroscience, University of Cambridge.