Nonsmooth Convex Optimization for Structured Illumination Microscopy Image Reconstruction

Jérôme Boulanger, Nelly Pustelnik, Laurent Condat, Lucie Sengmanivong, Tristan Piolot

Nonsmooth Convex Optimization for Structured Illumination Microscopy Image Reconstruction

Jérôme Boulanger, Nelly Pustelnik, Laurent Condat, Lucie Sengmanivong and Tristan Piolot

In this paper, we propose a new approach for structured illumination microscopy image reconstruction. We first introduce the principles of this imaging modality and describe the forward model. We then propose the minimization of nonsmooth convex objective functions for the recovery of the unknown image. In this context, we investigate two data-fitting terms for Poisson-Gaussian noise and introduce a new patch-based regularization method. This approach is tested against other regularization approaches on a realistic benchmark. Finally, we perform some test experiments on images acquired on two different microscopes.

1. Introduction

Context Superresolution approaches allow us to go beyond the resolution of standard wide-field fluorescence microscopy, therefore breaking the classical diffraction limit defined by Abbe in 1873 [1]. Structured illumination microscopy (SIM) is one of the recently proposed optical superresolution methods compatible with time lapse imaging of several labels. Based on the illumination of a sample by a set of interference patterns, this technique makes it possible to typically increase the resolution of the microscope by a factor of two [2, 3]. The resulting sinusoidal modulations of the fluorophore excitation signal lead to frequency shifts in the Fourier domain, which bring inaccessible frequencies within the scope of the optical transfer function of the microscope. An example of acquired raw data is depicted in Fig. 1. Once post-processed, the acquired images show an increased resolution, as illustrated in Fig. 2, where an acquired image has been reconstructed using a linear method [3]. Several studies have investigated the properties of such reconstruction algorithms and provided solutions for artifact reduction [4, 5]. However, like in many optical microscopy approaches, the photon counting process leads to noisy data, compromising the quality and the resolution of the final images. Therefore, the development of reconstruction methods less sensitive to noise and able to deal with the specificity of the structure of the reconstruction problem is crucially needed.

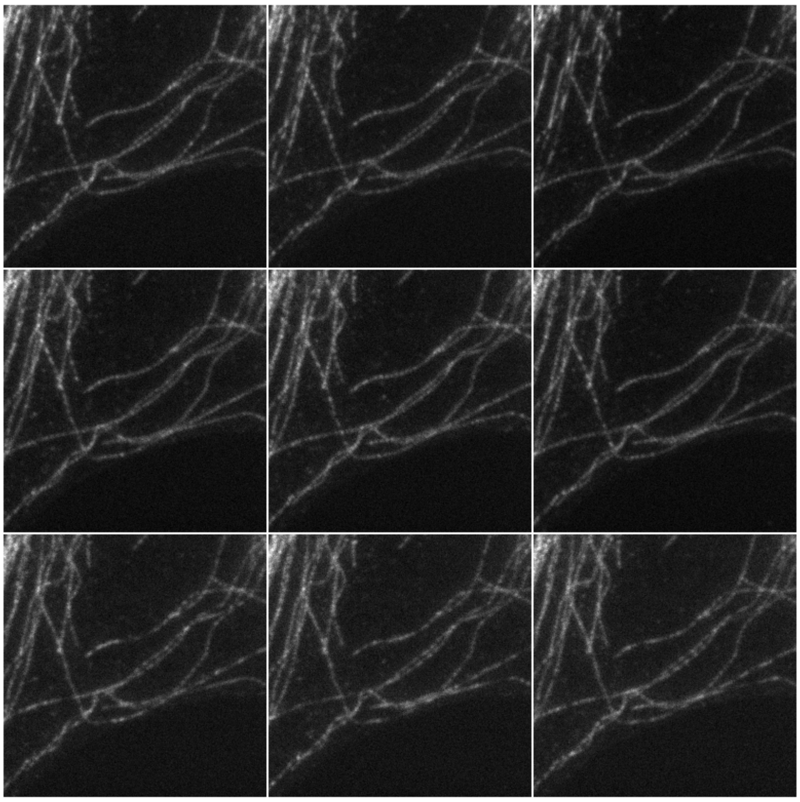

Figure 1: Example of real data. A Molecular Probe slide was imaged 9 times using a Nikon SIM microscope using a $100\times$ oil objective. The images represent a 256×256 region of 512×512 acquired images and display some labeled microtubules. The modulation pattern can be observed as a slight Moiré effect on the object.

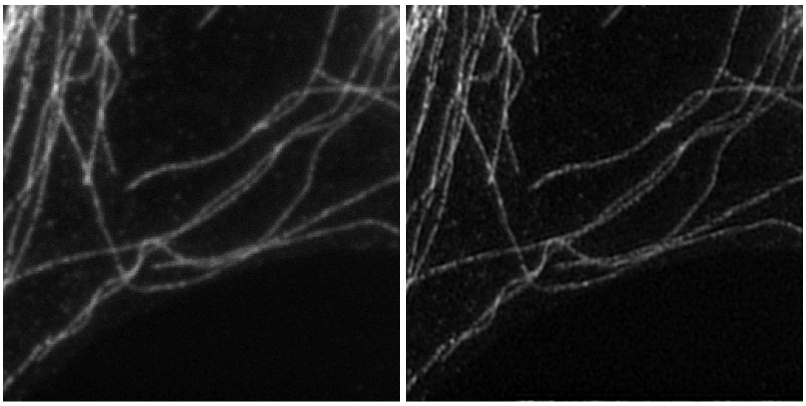

Figure 2: Reconstruction of the data displayed in Fig. 1. On the left the corresponding classical wide–field microscopy is obtained from the mean of the nine images. On the right, a linear least-squares reconstruction. The actual dimension of the image on the right is twice the size of the image of the left.

Related work While Wiener filtering remains the main reconstruction approach for SIM, the problem was recast in [6] as a more general inverse problem, allowing more complex illumination pattern to be considered [7–9]. Several regularization approaches have been also explored, in [6, 10] the ℓ_2 norm of the Laplacian operator is considered and total-variation (TV) was explored in [11] to deal with low signal to noise ratios. If [12] makes the underlying assumption of the Poisson noise model, none of these approaches consider a more accurate Poisson–Gaussian noise model. Moreover, none of the recent regularization methods, such as the Schatten norm of the Hessian operator [13], nonlocal total variation (NLTV) [14–16], global patch dictionaries [17,18] or local patch dictionaries [19] have been applied to structured illumination reconstruction, therefore limiting the final performance of this superresolution technique in its ability to discriminate fine structures of interest.

Contribution & organization of the article We propose here a reconstruction method taking into account the Poisson–Gaussian distribution of the noise and relying on a new regularization approach based on learning local dictionaries of patches in a convex setting. The minimization of the resulting cost function is performed using a versatile primal–dual optimization method. An extensive comparison with alternative regularization approaches is provided and we detail the implementation aspects of the tested regularization approaches in an unified way. Note that this work is an extension of the method published in a conference proceedings [20].

We will first recall the image formation problem in Section. 2.2, and further introduce the proposed reconstruction scheme based on Poisson-Gaussian approximation and local dictionaries of patches in Section 3. A performance evaluation on a synthetic dataset is then detailed in Section 4.2 and the reconstruction of acquired data is finally analyzed in Section 4.4. Finally, in A, an analysis of the least-squares solution underlying the issue associated to the reconstruction of SIM data is presented. In B, we recall the implementation details for the tested regularization cost terms and the needed tools for their minimization.

2. Presentation of the problem

2.1. Notations

In this article, $\mathbf{I}_N$ denotes the identity operator/matrix of size $N \times N$; when the size is not mentioned, it should be clear from the context. $\cdot^*$ denotes the adjoint of an operator; when the operator is assimilated to its representation matrix, with real entries, $\cdot^* = \cdot^{\mathrm{T}}$, the transpose operation. In the following, $\otimes$ denotes the Kronecker product and $\cdot^\dagger$, the Moore–Penrose pseudo-inverse. Notations used in this article are listed in Table 6.

2.2. Forward problem

Let us consider a set of K noise-free images $\bar{\mathbf{y}}_k$ with $k = 1, \ldots, K$:

$$\bar{\mathbf{y}}_k = \mathbf{S}_0 \mathbf{A}_0 \mathbf{M}_k \bar{\mathbf{x}}, \tag{1}$$

where $\bar{\mathbf{x}}$ is the unknown two–dimensional image defined on a regular grid of size $N_1 \times N_2$ and represented in a vectorized form by a vector of size $N = N_1 N_2$. $\mathbf{M}_k$, $\mathbf{A}_0$ and $\mathbf{S}_0$ are three linear operators represented by matrix multiplications and corresponding to modulation, convolution and down-sampling, respectively.

The modulation operator $\mathbf{M}_k$ performs a pixelwise multiplication by a pattern image $\mathbf{m}_k$, so that $\mathbf{M}_k = \mathrm{diag}(\mathbf{m}_k)$. In structured illumination microscopy, modulations often result from the interference of two or three coherent laser beams [2,21] and can be represented by a sinusoidal pattern defined for each point of coordinates $(n_1, n_2) \in N_1 \times N_2$ as:

$$[\mathbf{m}_k]_{n_1,n_2} = 1 + \alpha_k \cos(n_1\omega_{1,k} + n_2\omega_{2,k} + \varphi_k), \tag{2}$$

where α_k is the amplitude of the modulation, $\omega_{1,k}$ and $\omega_{2,k}$ are the modulation frequencies and φ_k a phase. However, one can devise other light structuring strategies such as a set of scanning points [22], or random illumination [7], often at the expense of the number of required images. In the following, we will stack all the modulations $\mathbf{M}_k$ in the matrix $\mathbf{M} = \left[\mathbf{M}_1, \cdots, \mathbf{M}_K\right]^{\mathrm{T}}$.

The convolution operator $\mathbf{A}_0$ models the point-spread function of the acquisition system, represented as a pseudo-circulant $N \times N$ matrix. In the sequel, we will use the notation $\mathbf{A} = \mathbf{I}_K \otimes \mathbf{A}_0$ to represent the convolution of all modulated images. Moreover, when approximating the optical microscope by a perfect diffraction-limited 2D imaging system, we can model the optical transfer function (OTF) in widefield microscopy by the auto-correlation of the pupil function as [6,23]:

$$\mathcal{A}_0(\varrho) = \begin{cases} \frac{2}{\pi}\left(\arccos\left(\frac{\varrho}{2\varrho_0}\right) - \frac{\varrho}{2\varrho_0}\sqrt{1-\left(\frac{\varrho}{2\varrho_0}\right)^2}\right) & \varrho \leq \varrho_0, \\ 0 & \text{otherwise,} \end{cases}, \tag{3}$$

where $\varrho = \sqrt{\xi_1^2 + \xi_2^2}$ is the amplitude of the frequencies in polar coordinates and ϱ_0 the cutoff frequency. A profile of the OTF $\mathcal{A}_0(\varrho)$ is depicted in dashed black in Fig. 3. We can note that for any pair of signals whose spectrum only differs for frequencies greater than ϱ_0, both signals will be equal when viewed through the optical system. We therefore cannot assume the operator $\mathbf{A}_0$ to be injective.

The down-sampling operator $\mathbf{S}_0$ represented by a matrix of size $L \times N$, where typically $L = N/4$, leads to down-sampling of a factor 2 in each dimension. Note that the images could be sampled at a higher rate at the acquisition time, however this would compromise the field of view and increase the noise level, since the number of photon per pixel would also decrease. In the rest of the text, down-sampling for the set of K images is represented by the operator $\mathbf{S} = \mathbf{I}_K \otimes \mathbf{S}_0$.

To summarize our forward imaging model, we can now conveniently rewrite Eq. (9) as:

$$\bar{\mathbf{y}} = \mathbf{SAM}\bar{\mathbf{x}}. \tag{4}$$

where $\bar{\mathbf{y}} = [\bar{\mathbf{y}}_1, \cdots, \bar{\mathbf{y}}_K]^{\mathrm{T}}$ is the stack of noise-free images.

The principle of SIM imaging in the case of sinusoidal modulations is illustrated in Fig. 3. It depicts how the modulations amount to a shift in the Fourier domain (Fig. 3.b), that makes it possible for the optical system to capture information at frequencies above the cutoff frequency ϱ_0 (Fig. 3.c). By shifting back these components individually, a high resolution image is recovered. However, in order to obtain this highly resolved image (Fig. 3.d yellow curve)

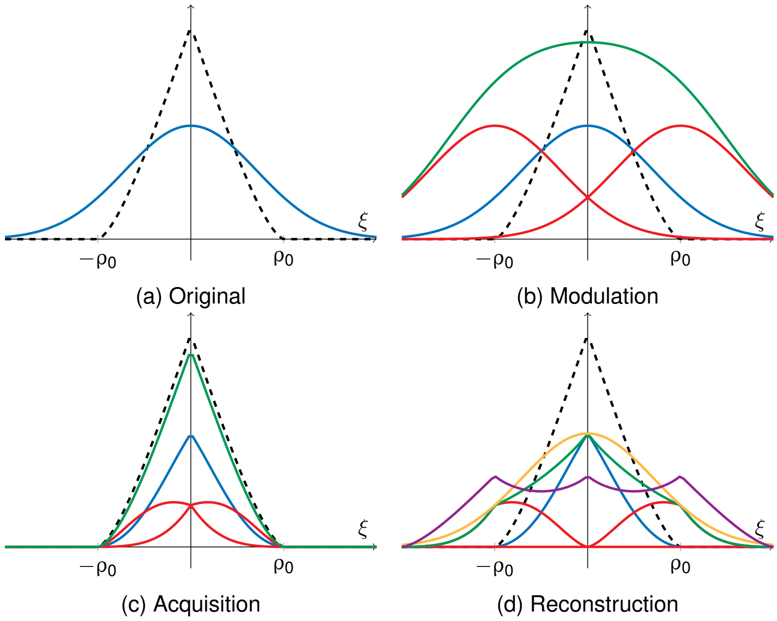

Figure 3: Principle of structured illumination microscopy illustrated in one dimension. (a) Spectrum of x in blue and the optical transfer function $\mathcal{A}_0$ in black (b) Spectrum of a modulated image $\mathbf{x}_n \cdot (1 + \cos(\omega n + \varphi))$ (green) as the sum of the three components 1, $e^{\pm i(\omega \xi + \varphi)}$ (resp. blue and red) (c) Spectrum of the sum (green) and of the individual components (blue and green) after being filtered by the OTF of the optical system (d) Reconstruction of a superresolved image obtained by shifting the modulated components (red) and summing them (green). Finally, normalizing the components by taking into account the shape of the OTF (purple) allows us to recover the original image (yellow).

a normalization step equivalent to the ratio of the demodulated images (green curve) with the shifted OTFs (purple curve) is necessary, at the risk of amplifying the noise present in the acquired data.

Indeed, the acquired images are actually degraded by some random noise due to the photo–electron counting process (shot noise) and the thermal agitation of the electrons (dark current and readout noise). To take into account those degradations, a general noise model can be written as [24]:

$$\mathbf{y} = \kappa\, \mathbf{p} + \mathbf{n}, \tag{5}$$

where κ is the overall gain of the acquisition system, $\mathbf{p}$ is a vector of Poisson distributed random variables of parameter $(\bar{\mathbf{y}} - m_{\mathrm{DC}})/\kappa$ accounting for the shot noise and $\mathbf{n}$ a vector of Normally distributed random variables of mean m_{DC} and variance σ^2_{DC}. The offset term m_{DC} accounts for the baseline gray level that are characteristic of the sensor, while the variance σ^2_{DC} of the additive white Gaussian noise summarizes several intensity-independent noise contributions such as dark current and readout noise. This formulation ensures that $\lim_{N_\ell \to \infty} \frac{1}{N_\ell} \sum_{\ell=1}^{N_\ell} (\mathbf{y})_\ell \to \bar{\mathbf{y}}$ for N_ℓ different statistical realizations of the random vector $(\mathbf{y})_\ell$. The resulting distribution $\mathbf{y}$ is then the convolution of a Poisson distribution and a Normal distribution. Note that a direct use of a variance stabilization transform [25] would introduce nonlinearities, which would have a significant impact on the observation model (9) and make the reconstruction process intractable. The additive white Gaussian noise model cannot deal with variations of the noise level, especially considering that an image often has a high dymanic range (16bit), and a pure Poisson approximation does not account for the presence of additional readout noise. Therefore, we consider in Section 3.2, which are able to capture the specificities of joint Poisson-Gaussian noise.

2.3. Regularized least-squares solution

Despite its suboptimality in the case of Poisson–Gaussian noise and its detrimental effect on the final image resolution, the first method for the reconstruction of structured illumination data [21] is based on a regularized least-squares approach [26] (See A):

$$\hat{\mathbf{x}} = \operatorname*{Argmin}_{\mathbf{x}} \tfrac{1}{2}\|\mathbf{y} - \mathbf{SAMx}\|^2 + \lambda\|\mathbf{Lx}\|^2, \tag{6}$$

for some linear operator $\mathbf{L}$ and regularization parameter $\lambda > 0$. Different operators $\mathbf{L}$ will then correspond to different well known regularization methods. In particular, when $\mathbf{L} = \mathbf{I}_N$, then Eq. (6) corresponds to Wiener regularization, choosing $\mathbf{L} = [\mathbf{D}_1, \mathbf{D}_2]^{\mathrm{T}}$ as the forward finite differences of $\mathbf{x}$ will favor a smooth solution and finally, using the $\mathbf{L} = \mathbf{D}_{11}^2 + \mathbf{D}_{22}^2$ as the Laplacian was used in [6] for SIM image reconstruction with $\mathbf{D}_{11}$ and $\mathbf{D}_{22}$ the second order derivatives along the horizontal and vertical directions.

For general modulation $\mathbf{M}$ the pseudoinverse solving the least-squares problem would be intractable (See A) and a minimization algorithms such as the conjugate gradient are needed [6,7]. However, in the specific case of separable sinusoidal modulations, that is when $\mathbf{M}$ takes the form defined in (25),we can obtain a closed form for the estimation of the SIM image:

$$\mathbf{x}^{\dagger} = (\mathbf{SA\Omega\Omega^*A^*S^*} + \lambda\mathbf{L^*L})^{-1}\mathbf{\Omega^*A^*S^*P}_1(\mathbf{I}_K \otimes \mathbf{\Phi}_0^{\dagger})\mathbf{P}_2\mathbf{y}, \tag{7}$$

where P_1 and P_2 are permutation matrices and † denotes the pseudoinverse. This approach is related to the original reconstruction proposed by [3] with $\mathbf{L}$ equal to the identity. Equation (7) amounts indeed to the separation of the differents modulation components by solving N small (3×3) linear system of dimension $(\mathbf{P}_1(\mathbf{I}_K \otimes \mathbf{\Phi}_0^{\dagger})\mathbf{P}_2)$, the remodulation of these components $(\mathbf{\Omega}^*)$ and a Wiener filter which can be performed in Fourier space. An apodization term defined in the Fourier domain as a triangle function, as an approximation of the OTF, is often used in order to reduce high frequency noise. In the following section, we propose to explore an alternative approach to handle the presence of Poisson–Gaussian noise and regularization based on local dictionaries of patches.

3. Proposed approach

In this work, we focus on a general framework aiming to deal with possibly nonfinite data fidelity terms and nonsmooth regularization terms [27]. We formulate the estimation procedure as a minimization involving a sum of Q cost terms defined by:

$$\hat{\mathbf{x}} \in \operatorname*{Argmin}_{\mathbf{x} \in C} \sum_{q=1}^{Q} f_q(\mathbf{T}_q\mathbf{x}), \tag{8}$$

where f_q are convex, closed and proper functions [28] from $\mathbb{R}^{M_q}$ to $\mathbb{R} \cup \{+\infty\}$, C is a nonempty closed convex subset of $\mathbb{R}^N$ (e.g. nonnegative solutions) and $\mathbf{T}_q$ operators represented as matrices of size $M_q \times N$. The cost terms $f_q(\mathbf{T}_q\mathbf{x})$ corresponding either to a data fidelity term or to a regularization term.

3.1. Primal–dual proximal minimization

When the involved functions are non-necessarily smooth, two main classes of algorithms can be derived to solve (8) and have been largely employed for solving inverse problems during the last decade: the alternating directions method of multipliers (ADMM) [29] or the Chambolle–Pock algorithm also known as primal–dual hybrid gradient (PDHG) [30–34]. Both strategies have in common to split the processing of the $(\mathbf{f}_q)_{1\le q\le Q}$ and the $(\mathbf{T}_q)_{1\le q\le Q}$ and to rely on the computation of the proximity operator [35] of each f_q. We define the proximity operator $\mathrm{prox}_f : \mathbb{R}^n \to \mathbb{R}^n$ of any closed proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ as:

$$(\forall \mathbf{x} \in \mathbb{R}^n) \quad \mathrm{prox}_f(\mathbf{x}) = \underset{\mathbf{y}\in\mathbb{R}^n}{\arg\min} \left(\frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + f(\mathbf{y}) \right) \tag{9}$$

Note that a large number of closed form expressions are known in the literature [28]. Some of them, useful for the study, are recalled in B. The major difference between both strategies comes from the way the operators $(\mathbf{T}_q)_{1\le q\le Q}$ are handled. The ADMM requires to compute $(\sum_{q=1}^{Q} \mathbf{T}_q^*\mathbf{T}_q)^{-1}$ while the PDHG strategies avoid such a step. Note that since in general the operator associated to SIM imaging is not directly invertible (cf. A),the ADMM would require an inner minimization procedure (*e.g.* the conjugate gradient) for the inversion of this operator. Finally, we can note that the PDHG can be formulated as a preconditioned ADMM [27]. Consequently, we solve Eq. (8) using the PDHG algorithm [30,33,34] (see Fig. 4) and consider several cases corresponding to the combination of function f_q and operator $\mathbf{T}_q$. The algorithm has four parameters: the number of iteration R, τ, σ and the acceleration $\rho \in]0,2[$. In practice, we use $R = 500$, $\rho \approx 2$ and $\sigma = 1/(\tau L)$ where $L = \sqrt{\sum_{q=1}^{Q} \|T_q\|^2}$ [33]. The last parameter τ is set to 1 by default in our experiments but could be adapted for each objective function. Note that this parameter will only change the convergence rate.

In the next section, we will explicit the cost terms $f_q(\mathbf{T}_q\mathbf{x})$ corresponding to the proposed approach, while B details the other regularization terms tested in Section 4.2. We give the expression of the function f_q and its associated proximity operator [35], and describe the operator $\mathbf{T}_q$ and its ajoints when needed. As a convention, we denote $\mathbf{z} = \mathbf{T}_q\mathbf{x}$ the vector of length M_q in the image space of $\mathbf{T}_q$. In practice, we will later consider only the combination of one data term along with one regularization term, while C will denote the nonnegativity constraint, which will be enforced directly on the iterates of the algorithm, see Step 8 in Fig. 4. Therefore, we can write the estimate as: $\hat{\mathbf{x}} \in \mathrm{Argmin}_{\mathbf{x}\ge\mathbf{0}} f_1(\mathbf{T}_1\mathbf{x}) + f_2(\mathbf{T}_2\mathbf{x})$.

3.2. Poisson–Gaussian approximation

Handling Poisson–Gaussian noise is challenging as the resulting probability density function (p.d.f) is the convolution of the Poisson and Gaussian densities. Consequently, several strategies have been developed over the years to approximate the resulting p.d.f (See [36] for a

Require: $\mathbf{x}_0 \in \mathbb{R}^N$, $\mathbf{v}_{1,q} \in \mathbb{R}^{M_q}$, $\forall q \in 1, \dots, Q$, $\tau > 0$, $\sigma > 0$, $\rho > 0$

$\mathbf{x}_1 = \mathbf{z}_1 = \mathbf{x}_0$
for $r \in 1, \dots, R$ **do**
  for $q \in 1, \dots, Q$ **do**
    $\mathbf{u}_{r+1,q} = \mathbf{v}_{r,q} + \sigma\, \mathbf{T}_q(2\mathbf{z}_r - \mathbf{x}_{r-1})$
    $\mathbf{p}_{r+1,q} = \mathbf{u}_{r+1,q} - \sigma\, \mathrm{prox}_{f_q/\sigma}(\mathbf{u}_{r+1,q}/\sigma)$
    $\mathbf{v}_{r+1,q} = \rho\, \mathbf{p}_{r+1,q} + (1-\rho)\, \mathbf{v}_{r,q}$
  end for
  $\mathbf{z}_{r+1} = P_C\left(\mathbf{x}_r - \sum_{q=1}^{Q} \tau \mathbf{T}_q^* \mathbf{v}_{r+1,q}\right)$
  $\mathbf{x}_{r+1} = \rho\, \mathbf{z}_{r+1} + (1-\rho)\, \mathbf{x}_r$
end for

Figure 4: The primal–dual minimization algorithm proposed in [33] allows us to minimize the energy functional defined by Eq. (8) given that the proximity operator of the function f_q and the operator $\mathbf{T}_q$ and its adjoint $\mathbf{T}_q^*$ are defined. We can notice that this algorithm does not require the direct inversion of these operators.

recent review). The different approximations are more or less precise, depending on the relative amount of Gaussian and Poisson noise, and are more are less numerically tractable. Here, we propose to consider two approximations: the shifted Poisson model and the heteroscedastic Gaussian model approximation. One can notice a degree of symmetry between these two approaches, as the first one approximates Gaussian noise as Poisson noise with shifted intensity while the second one approximate Poisson noise by Gaussian noise with variance depending on the intensity of the signal.

Shifted Poisson model Under a purely Poisson noise model assumption for the acquired data $\mathbf{y}$, the negative log-likelihood would coincide up to a constant with the Kullback–Leibler (KL) divergence also called I-divergence [37] and can be expressed as $\mathbf{z} = (\mathbf{z}_n)_{1\le n\le LK} \mapsto f_{\text{Poisson}}(\mathbf{z}) = \sum_{n=1}^{LK} f_{\text{Poisson}}^{(n)}(\mathbf{z}_n)$, where the component-wise function is defined by [38]:

$$(\forall \gamma > 0, \forall \mathbf{z}_n \in \mathbb{R}) \quad f_{\text{Poisson}}^{(n)}(\mathbf{z}_n) = \begin{cases} \mathbf{z}_n - \mathbf{y}_n \log \mathbf{z}_n, & \mathbf{z}_n, \mathbf{y}_n > 0, \\ \mathbf{z}_n, & \mathbf{z}_n > 0 \text{ and } \mathbf{y}_n = 0, \\ \infty, & \text{otherwise.} \end{cases} \tag{10}$$

In this configuration, the linear operator is $\mathbf{T}_{\text{Poisson}} = \mathbf{SAM}$. The proximity operator is given component-wise for $n \in [1, LK]$ by:

$$(\forall \mathbf{z}_n \in \mathbb{R}) \quad \mathrm{prox}_{\gamma f_{\text{Poisson}}^{(n)}}(\mathbf{z}_n) = \frac{1}{2}\left(\mathbf{z}_n - \gamma + \sqrt{(\mathbf{z}_n - \gamma)^2 + 4\gamma \mathbf{y}_n}\right) \tag{11}$$

and the proximity operator $\mathrm{prox}_{\gamma f_{\text{Poisson}}}(\mathbf{z}) = \left(\mathrm{prox}_{\gamma f_{\text{Poisson}}^{(n)}}(\mathbf{z}_n)\right)_{1\le n\le LK}$ is obtained by applying Eq. (11) to each component of the vector $\mathbf{z}$.

In the case of Poisson–Gaussian noise, we can shift the Poisson likelihood as proposed by [39]. However, we would like to take into account the full model that we proposed in Eq. (5) with a Gaussian noise $\mathbf{n} \sim \mathcal{N}(m_{\text{DC}}, \sigma_{\text{DC}}^2)$ and a gain κ. In order to do so, we seek a transformation of the form $(\mathbf{y} - b)/a$ with $(a, b) \in \mathbb{R}^2$ so that the two first moments are matching after transformation $\mathbb{E}[\frac{\mathbf{y}-b}{a}] = \mathbb{V}ar\left[\frac{\mathbf{y}-b}{a}\right]$ in order to satisfy the intrinsic property of Poisson random variable. Choosing $a = \kappa$ leads to $b = m_{\mathrm{DC}} - \sigma^2_{\mathrm{DC}}/\kappa$ and we can define then, the shifted Poisson data-fitting term as:

$$(\forall \mathbf{z}_n \in \mathbb{R}) \quad f^{(n)}_{\text{shifted-Poisson}}(\mathbf{z}_n) = f^{(n)}_{\text{Poisson}}\left(\frac{\mathbf{z}_n - b}{a}\right) \tag{12}$$

and the associated proximity operator component-wise for $n \in [1, LK]$ as:

$$(\forall \mathbf{z}_n \in \mathbb{R}) \quad \mathrm{prox}_{\gamma f^{(n)}_{\text{shifted}-\text{Poisson}}}(\mathbf{z}_n) = a\,\mathrm{prox}_{\gamma f^{(n)}_{\text{Poisson}}}\left(\frac{\mathbf{z}_n - b}{a}\right) + b \tag{13}$$

using these two constants for the shifted Poisson approximation.

Heteroscedastic Gaussian noise model As an approximation of the Poisson–Gaussian noise model, a weighted least-squares data term can be used to take into account the dependency between the variance of the noise level and the intensity of the signal. The weighted least-squares can be written as

$$(\forall \mathbf{z} \in \mathbb{R}^{LK}) \quad f_{\mathrm{WLS}}(\mathbf{z}) = \frac{1}{2}(\mathbf{z} - \mathbf{y})^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{z} - \mathbf{y}), \tag{14}$$

where $\mathbf{W}$ is a diagonal variance matrix of size $LK \times LK$ with elements $(\mathbf{w}_n)_{1 \le n \le LK}$ and the associated proximity operator is given by

$$(\forall \gamma > 0)(\forall \mathbf{z} \in \mathbb{R}^{LK}) \quad \mathrm{prox}_{\gamma f_{\mathrm{WLS}}}(\mathbf{z}) = \left(\frac{\mathbf{w}_n \mathbf{z}_n + \gamma\,\mathbf{y}_n}{\mathbf{w}_n + \gamma}\right)_{1 \le n \le LK}. \tag{15}$$

Given the noise model defined by Eq. (5) the variance at each point n is given by:

$$\mathbb{V}\mathrm{ar}[\mathbf{y}_n] = \kappa \mathbb{E}[\mathbf{y}_n] + \sigma^2_{\mathrm{DC}} - \kappa\, m_{DC}, \tag{16}$$

with $\mathbb{E}[\mathbf{y}_n]$ and $\mathbb{V}\mathrm{ar}[\mathbf{y}_n]$ the expectation and variance of the random variable $\mathbf{y}_n$. The weights are consequently given by:

$$\mathbf{w}_n = \kappa \bar{\mathbf{y}} + \sigma^2_{\mathrm{DC}} - \kappa\, m_{DC} \tag{17}$$

where $\bar{\mathbf{y}}$ can be approximated by its noisy counterpart $\mathbf{y}$.

Noise parameter estimation If the CCD or sCMOS sensor have not been calibrated and the parameters κ, m_{DC} and σ_{DC} are unknown, we can follow the procedure described in [40] to estimate the required parameters. The variance of the noise $\mathbb{V}\mathrm{ar}[\mathbf{y}_n]$ is estimated locally using a maximum of absolute deviation filter (MAD) computed on the pseudo-residuals (normalized Laplacian $\frac{1}{\sqrt{20}}(\mathbf{D}^2_{11}\mathbf{y} + \mathbf{D}^2_{22}\mathbf{y})$) of the image while the mean is estimated using a median filter. The linear regression allows us then to estimate the gain κ and the noise variance at the origin $e_{DC} = \sigma^2_{\mathrm{DC}} - \kappa\, m_{DC}$. Interestingly, these two parameters are sufficient to fully determine the noise model for both approximation.

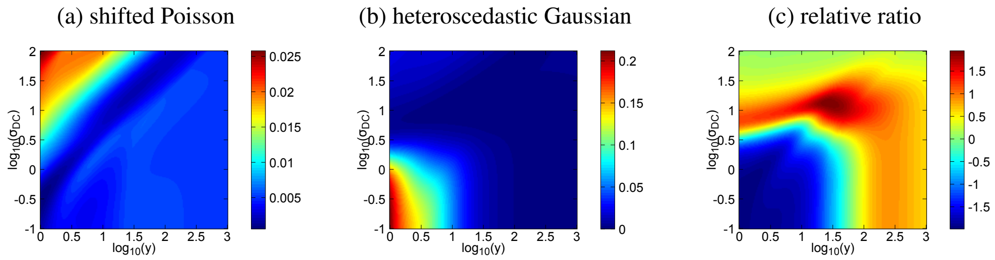

Figure 5: Poisson–Gaussian noise approximation error measured as the Hellinger distance between the exact likelihood and either shifted Poisson (a) or the heteroscedastic Gaussian (b) models. In (c) the relative ratio of the two errors highlights in blue the domain in the $(\mathbf{y}, \sigma_{\mathrm{DC}})$ space where the shifted Poisson model outperforms the heteroscedastic approximation.

Comparison of the associated likelihoods In order to gain some insight into these two approximations of the Poisson–Gaussian noise model, we can analyze the associated likelihoods. We present in Fig. 5 the Hellinger distance between the likelihoods of the exact model and the two approximations, in (a) and (b), as a function of the photo-electron count $\bar{\mathbf{y}}$ and readout noise σ_{DC} simplifying without loss of generality setting $\kappa = 1$ and $m_{\mathrm{DC}} = 0$ for simplicity without loss of generality. The relative ratio shows that the likelihood of the shifted Poisson model is closer to the exact likelihood when the number of photons $\bar{\mathbf{y}}$ is approximately below 50 and the standard deviation of the readout noise is approximately below 5. Then a transition zone shown in red indicates that the heteroscedastic likelihood is closer to the exact model. Increasing further the photon count and the readout noise finally seems to stifle the difference between these two models as the relative difference tends to zero (lime green color). This analysis suggests that the approximation that should be used could be selected according to the regime where the data have been acquired.

3.3. Regularization by local dictionaries of patches

We propose here to adapt the idea of online learning of sparse local dictionaries of patches in the context of inverse problem regularization by considering the nuclear norm of a patch extraction operator $\mathbf{T}_{\mathcal{P}}$. This operator $\mathbf{T}_{\mathcal{P}}$ maps all the $N_p \times N_p$ patches in each neighborhoods of dimension $N_w \times N_w$ into a matrices of dimension $N_p^2 \times N_w^2$. Dimension of $z = \mathbf{T}_{\mathcal{P}}(x)$ can be then represented as a 4D array, where a $N_p^2 \times N_w^2$ collection of patches corresponds to each 2D point of the image space. The adjoint of this operator is the projection of the overlapping collection patches onto the image. Note that the operator $\mathbf{T}_{\mathcal{P}}$ does not depend on the content of $\mathbf{x}$ but is only parametrized by the windows and patch dimensions. As an illustration, let us

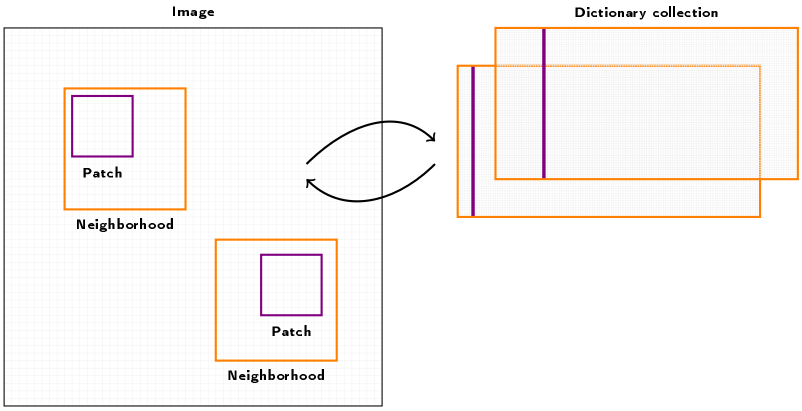

Figure 6: Illustration of the local dictionary of patch extraction. Neighborhood (16×16 pixels) depicted in orange are sampled every 8 pixels in each directions. For each neighborhood, each patches (violet) are collected and vectorized to form a matrix: the dictionary. Consequently, the image is represented as a collection of dictionaries.

consider the case of a 4×4 image and patches of size 2×2. Then the operator is:

$$\mathbf{T}_{\mathcal{P}}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_5 & x_6 \\ x_2 & x_3 & x_6 & x_7 \\ x_3 & x_4 & x_7 & x_8 \\ x_5 & x_6 & x_9 & x_{10} \\ x_6 & x_7 & x_{10} & x_{11} \\ x_7 & x_8 & x_{11} & x_{12} \\ x_9 & x_{10} & x_{13} & x_{14} \\ x_{10} & x_{11} & x_{14} & x_{15} \\ x_{11} & x_{12} & x_{15} & x_{16} \end{bmatrix}$$

and corresponds to the 9 possible translations of the patch of 4 elements, picking values within an image represented as a vector of 16 elements. The patch dictionaries are highly redundant and for computational efficiency, only a fraction of the possible neighborhoods can be considered by shifting the patch extraction window from its half in both directions. As an example, for a 512×512 images, the operator will map to a 128×128 field of 16×25 matrices corresponding to dictionaries of patches of size 4×4 extracted from neighborhoods of size 8×8. More generally, the total size of the collection of local dictionaries is given by $\frac{2n}{N_w}(N_w - N_p + 1)^2 N_p^2$. In order to better illustrate this approach, the diagram shown in Fig. 6 represents the extraction of two dictionaries associated to the neighborhoods of two selected points in the image.

The nuclear norm $\|\mathbf{z}_n\|_*$ of $\mathbf{z}_n \in \mathbb{R}^{2\times 2}$ is defined as the ℓ_1–norm of the diagonal matrix $\mathbf{\Lambda}_n$ such that $\mathbf{z}_n = \mathbf{U}_n \mathbf{\Lambda}_n \mathbf{V}_n^{\mathrm{T}}$. Then, the associated proximity operator is given by [16]:

$$(\forall \mathbf{z}_n \in \mathbb{R}^{2\times 2}) \quad \mathrm{prox}_{\gamma\|\cdot\|_*}(\mathbf{z}_n) = \mathbf{U}_n \mathrm{prox}_{\gamma\|\cdot\|_1}(\mathbf{\Lambda}_n)\mathbf{V}_n^{\mathrm{T}}. \tag{18}$$

The nuclear norm can be seen as a relaxed version of the case ℓ_0-norm of the eigenvalues [41]. Note that the Schatten norm $\mathcal{S}_p$ of the Hessian operator with $p = 1$ described in B and introduced in [13] is identical to the nuclear norm.

Patch	Neighborhood	Dictionary	λ	PSNR (dB)	Elapsed time (s)
4×4	8×8	100	1.0	14.25	105
8×8	16×16	648	2.0	14.34	106
16×16	32×32	4624	2.0	14.11	375

Table 1: Influence of patch and neighborhood size. The dictionary size factor is a multiple of the initial image size.

4. Results

4.1. Influence of the patch and neighborhood size

We proposed here to evaluate the influence of the patch and neighborhood size of the proposed patch-based regularization approach. Table 1 gives the dictionary size factor, PSNR and elapsed time for 3 combinations of patch and neighborhood size. We can observe that the two first are equivalent in term of computation time, with the second one providing a slightly better PSNR. Using a combination of larger patch and neighborhood increases by a factor 3 the computation time while not improving the PSNR. Finally, we can note the dependency of the optimal regularization parameter on the patch and neighborhood size. In the following, we will use a 8×8 patch size and a 16×16 neighborhood size.

4.2. Evaluation of data fitting term and regularization term

In order to evaluate the proposed reconstruction method, we consider alternative regularization approaches corresponding to different choices for the function f_2 and the operator $\mathbf{T}_2$. The regularization cost functions are the following

- the squared ℓ_2-norm of the gradient operator ($f_2(\,\cdot\,) = \|\cdot\|^2$, $\mathbf{T}_2 = [\mathbf{D}_1, \mathbf{D}_2]^T$),
- the squared ℓ_2-norm of the Laplacian operator ($f_2(\,\cdot\,) = \|\cdot\|^2$, $\mathbf{T}_2 = \mathbf{D}_{11}^2 + \mathbf{D}_{22}^2$) [6],
- the total variation as the ℓ_1-norm of the gradient operator ($f_2(\,\cdot\,) = \|\cdot\|_1$, $\mathbf{T}_2 = [\mathbf{D}_1, \mathbf{D}_2]^T$) [42, 43],
- the nonlocal total variation as the ℓ_1-norm of the weighted nonlocal finite difference operator ($f_2(\,\cdot\,) = \|\cdot\|_1$, $\mathbf{T}_2 = \mathbf{T}_{\mathrm{NL}}$) [14],
- the Schatten norm (with $p = 1$, that is the nuclear norm) of the Hessian operator with ($f_2(\,\cdot\,) = \mathcal{S}_1(\,\cdot\,)$, $\mathbf{T}_2 = \mathbf{T}_{\mathcal{H}}$) [13].

The details of the associated proximity operators and the definition of the linear operators are given in B. Note that only the three first regularization terms have been previously tested in this context, while the NLTV and Schatten norm of the Hessian operator have not been applied to SIM image reconstruction. Although none of these regularization terms have been tested in the context of Poisson–Gaussian noise model for SIM image reconstruction, we can note that they have been considered in the context image deconvolution under a Poisson-Gaussian noise model but with a different algorithm [36].

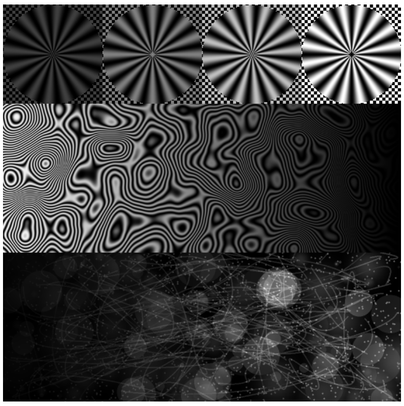

Image A (resolution test)

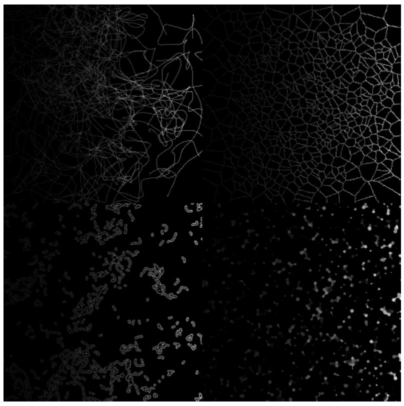

Image B (realistic sample)

Figure 7: Test images used for evaluating the data-fitting and regularization terms.

We generated two test images (See Fig. 7). The first one aims at providing a visual assessment of the performance of the reconstruction in term of resolution, by integrating details at various frequencies. To check the data fitting under the Poisson–Gaussian noise assumption, we took care to integrate a large range of gray levels and to prevent any bias towards a particular regularization term, various realistic textures (points, blobs, lines, disks) are also used. The second image emulates several realistic intracellular features of interest for single cell imaging such as microtubules, endoplasmic reticulum, mitochondria and vesicles. The dynamic range of both images is $[0, 255]$. We simulated structured illumination images by using a down-sampling factor of 2 and a cut-off frequency $\rho_0 = 1.53\,\text{pixel}^{-1}$. The modulations are composed of 3 equispaced phases and 3 equispaced angles with a frequency of 1 pixel^{-1}. The images were finally corrupted by noise, as described in Eq. (5), with $\kappa = 2$, $m_{DC} = 0$ and $\sigma_{DC} = 10$. For each run, we used 500 iterations and we tested 20 values logarithmically spaced in the interval $[0.01, 100]$ for the regularization parameters. We used the PSNR as a criterion in order to select the best image among the 20 results. All the implementation has been done using the MATLAB programming language.

To evaluate the performances of the reconstruction methods, we have considered the PSNR and the SSIM criteria. The PSNR being very sensitive to bias, we used a linear regression between the original image and the estimate to remove any systematic trend. More precisely the PSNR is defined as: $\text{PSNR}(\bar{\mathbf{x}}, \mathbf{x}) = 20 \log_{10}(255/\|\bar{\mathbf{x}} - (\mathbf{x} - c_0)/c_1)\|^2$ where the coefficient c_0 and c_1 are estimated by by minimizing $\|\hat{\mathbf{x}} - (c_0 + c_1\mathbf{x})\|^2$.

Figure 8 displays the evolution of both criteria as a function of the regularization parameter. We can notice that if the difference between the two data-fitting terms are more apparent in term of PSNR than in term of SSIM while the optimal regularization parameter seems to be consistent between the two performance measures. Figure 4.2 shows the images corresponding to the best PSNR for the 12 cost functions for both test images with the PSNR values, the corresponding SSIM, the regularization parameter and the elapsed time. PSNR and SSIM values seem to correlate with the visual inspection of the images. In particular, high resolution information in the middle row of the image A seems to be better retrieved when using the proposed

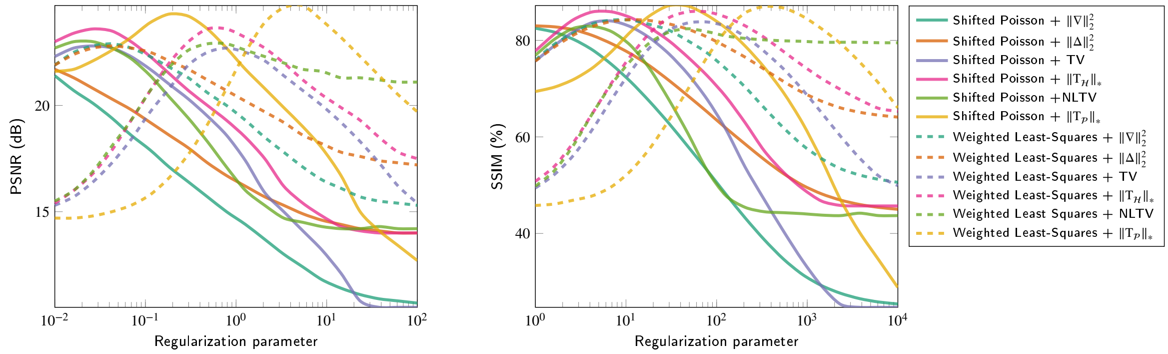

Figure 8: Evolution of the PSNR and the SSIM criterion as a function of the regularization parameters for the test image A.

Table 2: Best PSNR (dB) / SSIM (%) performance for both test images.

	$\|\nabla\|_2^2$	$\|\Delta\|_2^2$	TV	$\|T_{\mathcal{H}}\|_*$	NLTV	$\|T_{\mathcal{P}}\|_*$
Shifted Poisson	21.35 / 82.53	21.73 / 82.96	22.79 / 84.00	23.58 / 85.94	23.02 / 82.95	24.28 / **87.18**
Weighted Least-Squares	22.88 / 84.20	22.83 / 84.30	22.66 / 83.84	23.61 / 85.94	22.94 / 82.36	**24.67** / 87.00
	$\|\nabla\|_2^2$	$\|\Delta\|_2^2$	TV	$\|T_{\mathcal{H}}\|_*$	NLTV	$\|T_{\mathcal{P}}\|_*$
Shifted Poisson	26.70 / 85.42	26.86 / 85.89	27.99 / 87.59	28.38 / 88.08	27.86 / 86.52	**28.53** / 88.42
Weighted Least-Squares	27.82 / 86.86	27.79 / 86.97	27.68 / 87.81	28.03 / 88.51	27.12 / 85.57	28.16 / **88.61**

regularization approach. Note that the proposed regularization approach leads to significant computation times. On the one hand, all other approaches involve the optimized Matlab imfilter function, while on the other hand, the proposed approach requires many loops which are known to be slow in these condition. An implementation in another programming language using multi-threading would reduce the computational burden. Finally, best PSNR and SSIM values are displayed in Table 2 for each cost term and both images. We can see in bold that the best data fitting term would depend on the image and the performance measure while the proposed regularization consistently outperforms the other ones. Note that a comparison with recent methods taking advantage of the flexibility of the Bayesian framework for the formulation of the inverse problem [44] would be well suited for the reconstruction of blocky object, with sharp and sudden changes of intensity [45].

4.3. Modulation pattern

As described in [6], one advantage of considering the SIM image reconstruction as an inverse problem lies in the ability to reduce the number of acquired images. As an example, we can consider a set of 3 images with different modulation orientations but no phase shift. This allows us to effectively reduce the imaging speed and photo-toxicity which are both limiting factors in fluorescence light microscopy. This is a nonideal case as the sum of the modulation is not a uniform image and that therefore the noise is not spatially uniform. Nonetheless the results displayed in Fig. 10 show that the sample is successfully recovered with PSNR of 26.75dB and that high frequency details are well estimated as shown on the power-spectrum on the second

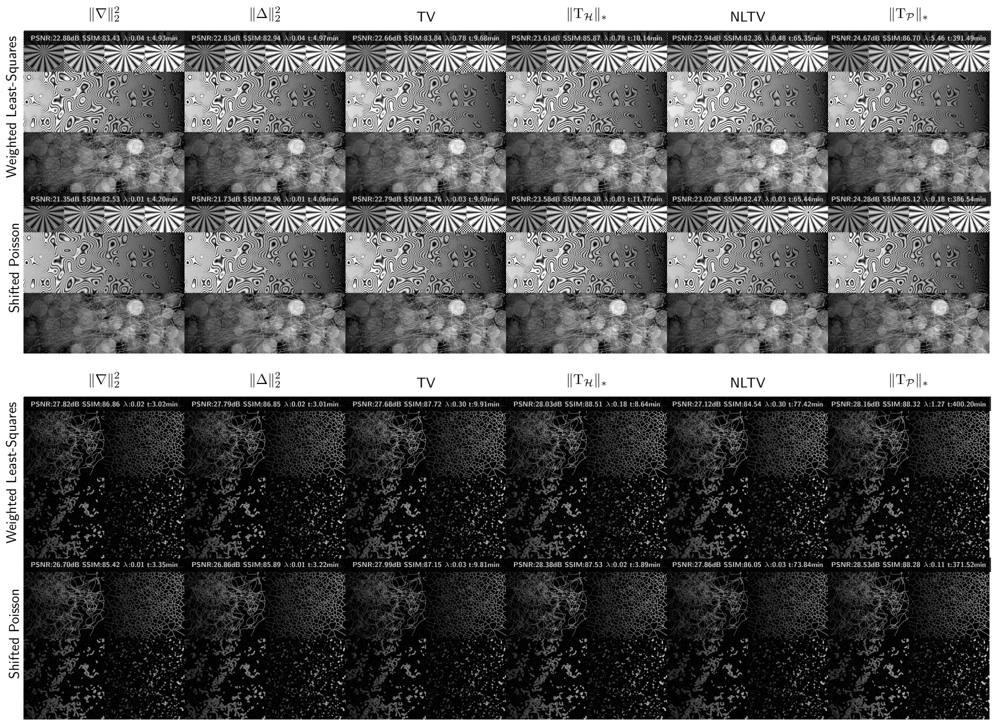

Figure 9: Reconstruction of the two test images with PSNR, SSIM, regularization parameter and elapsed time for the best PSNR. Images are displayed with gamma correction of 0.5 in order to better appreciate the details in the low end of the dynamic range.

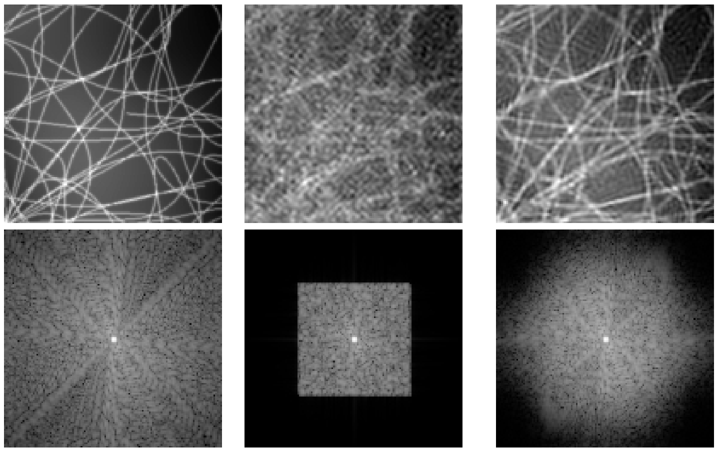

Figure 10: Reconstruction of a simulated 3 SIM images with a reduced number of images using only 3 modulation orientations and no phase shift. The second row displays the corresponding power spectra.

row.

4.4. Reconstruction of acquired data

We have tested the proposed approaches on acquired data. For this purpose, we used two commercial systems: the N–SIM from Nikon and the OMX from General Electrics. Both microscopes use a similar approach for performing SIM imaging and rely on the use of a diffraction grating which is optically conjugated with the object plane.

The N–SIM is equipped with a 100× (1.49 N.A.) objective and a 2.5× lense is set on the camera port. A Xion Ultra 897 EMCCD camera from Andor Technology Ltd was on the detection path leading to a pixel-size of ~ 64 nm in the final image. A FluoCell prepared slide #2 with BPAE cells with Mouse Anti-α-tubulin was imaged and the results obtained with the linear and the convex nonsmooth reconstruction with a Poisson data term and a *local patch dictionary* regularization are shown in Fig. 11 along, with the “wide–field” image obtained by averaging the nine acquired images. On this image, we can notice that filaments appear much thinner on the nonsmooth estimate than on the linear one. We can also observe that the power spectrum seems to have a larger support.

The OMX microscope is equipped with a 100× (1.4 N.A.) objective coupled with a 2× lense on the camera port. This time a Evolve 512 from Photometrics was used and the final pixel-size in the image is ~ 80 nm. A FluoCell prepared slide #1 with BPAEC cells with F-actin stained with Alexa fluor 488 phalloidin. Once again, both linear and the proposed nonsmooth convex reconstruction methods reveal an increased resolution. Varying the regularization parameter for the linear method does not allow to reduce noise without inducing a loss of resolution. The proposed method allows us to achieve a much better compromise in this respect and clearly outperforms the linear approach.

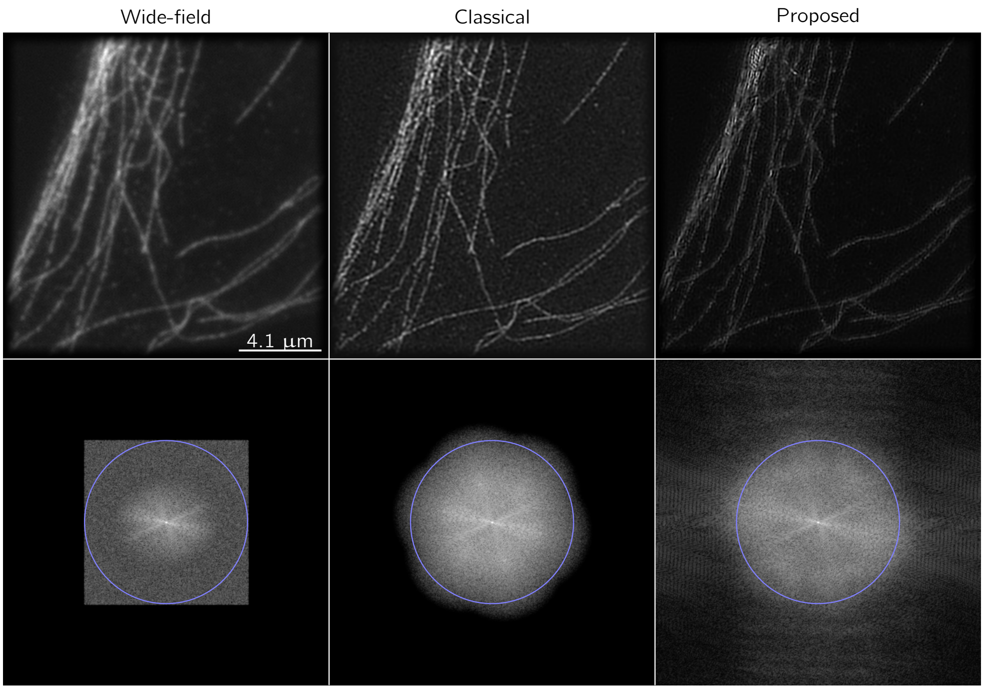

Figure 11: Reconstruction of acquired fluorescently labeled tubulin cell with a NSIM microscope. Structured illumination microscopy allows us to reveal the crossing of fibers with more details than the widefield image. The proposed approach is able to handle the noise and reduce the artifacts observed in the linear reconstruction (here the weighted least-square data fitting term was used). On the second row, the power spectrum is displayed as reveals the increased support in the frequency domain. The blue circle correspond to the resolution 128 μm.

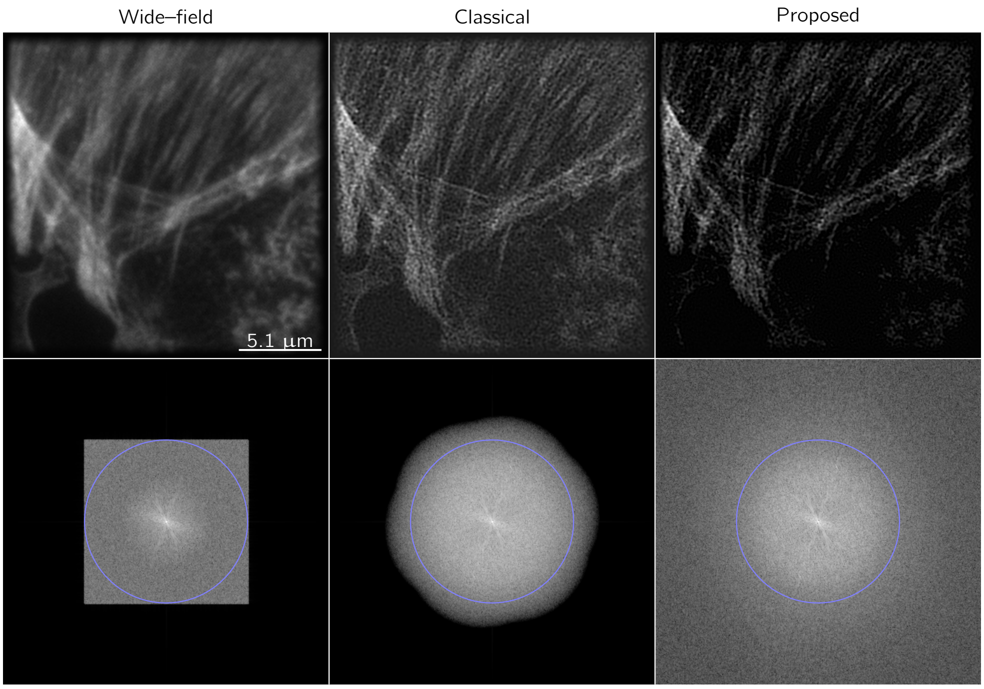

Figure 12: Reconstruction of acquired F-actin fluorescently labeled cell with the OMX setup. The fine and dense network structure of the actin cytosqueleton is better resolved when using the proposed approach. Corresponding power spectra are displayed in the second row.

5. Conclusion

We have proposed a new method for structured illumination image reconstruction. We have considered a primal–dual algorithm, which does not require the direct inversion of the forward operator, as this one is too large to be directly handled. In this framework, we have proposed a new regularization based on learning local patch dictionaries. Two valid approximations of the Poisson–Gaussian noise were tested and combined with several regularization to evaluate the performance of the proposed approach. The results show that the proposed approach leads to a significant improvement in terms of PSNR. Being able to better handle the noise perturbation make it possible to increase the resolution and the sensitivity of SIM images. We did not address the problem of the modulation parameter estimation, which can impact the quality of the reconstructed images [46]; we leave this study for future work. Finally, we have seen that the computation time associated to the proposed regularization are high. However, its implementation could be easily parallelized taking advantage of the multi-core architecture of modern CPUs and GPUs.

6. Acknowledgments

This work was funded by the CNRS PEPS grant “PROMIS”.

A. Least-squares solution

General Case Let us start by considering the additive Gaussian white noise approximation. In the following, we assume that the offset component m_{DC} has been subtracted from the data. In this case, maximum likelihood estimation of $\bar{\mathbf{x}}$ given the observed data $\mathbf{y}$ amounts to solve the least-squares problem:

$$\hat{\mathbf{x}} \in \operatorname*{Argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{SAMx}\|^2. \tag{19}$$

The solution of Eq. (19) is the pseudo-inverse:

$$\hat{\mathbf{x}} = (\mathbf{SAM})^{\dagger}\mathbf{y}, \tag{20}$$

which we can develop as:

$$\hat{\mathbf{x}} = (\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{SAM})^{-1}\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{y}. \tag{21}$$

We have

$$\begin{aligned} \mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{y} &= \mathbf{M}^*[(\mathbf{A}_0^*\mathbf{S}_0^*\mathbf{y}_1^{\mathrm{T}})^{\mathrm{T}} \ \cdots \ (\mathbf{A}_0^*\mathbf{S}_0^*\mathbf{y}_K^{\mathrm{T}})^{\mathrm{T}}]^{\mathrm{T}} & (22)\\ &= \textstyle\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{A}_0^*\mathbf{S}_0\, \mathbf{y}_k. & (23) \end{aligned}$$

Thus, we can notice that least-squares estimation reverts to applying the inverse operator $(\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{SAM})^{-1}$ to a *single* image, which is simply the sum of the $\mathbf{y}_k$, after they have been up-sampled, re-filtered and re-modulated.

Table 3: Detailed notations used in the paper.

n	index of the component of a vector (e.g. $\mathbf{x}$)	$\mathbf{I}_K$	$K \times K$ identity matrix
k	index for modulations	$\mathbf{M}_k$	modulations (diagonal matrix)
q	index of cost term (integer)	$\mathbf{A}_0$	point spread function (matrix)
r	iteration counter of an algorithm	$\mathbf{S}_0$	down-sampling (matrix)
N	pixel number of $\mathbf{x}$	$\mathbf{y}$	measurements (stacked vector)
L	pixel number of $\mathbf{y}$	$\mathbf{M}$	stacked modulations
K	number of modulations	$\mathbf{A}$	stacked point spread function
Q	number of cost term	$\mathbf{S}$	down-sampling (stacked matrix)
R	number of iterations	$\mathbf{W}$	diagonal weight matrix
T	number of translations (NLTV)	$\mathbf{p}$	number of photo electrons (stacked vector)
ϱ	radial frequency	$\mathbf{n}$	Gaussian white noise (stacked vector)
ξ_1, ξ_2	frequencies	$\mathbf{\Phi}_k$	phase matrix
κ	camera gain	$\mathbf{\Phi}$	stacked phase matrix
μ_{DC}	dark current noise mean offset	$\mathbf{\Omega}$	stacked frequency matrix
σ^2_{DC}	dark current noise variance	$\mathbf{L}$	Operator for linear regularized least-squares
ω_k	pulsation $2\pi\xi$ of the modulation	$\mathbf{D}_1$	finite difference operator
φ_k	phases of the modulation	$\mathbf{D}^2_{11}$	second order finite difference operator
τ, σ, ρ	algorithm parameters	$\mathbf{T}_q$	generic operator in the cost term
$\bar{\mathbf{x}}$	ground truth image $\mathbf{x}$	$\mathbf{T}_{\mathcal{D}}$	Stacked gradient operator
$\mathbf{x}$	estimated image (vector)	$\mathbf{T}_{\mathcal{L}}$	Laplacian operator
$\mathbf{y}_k$	measurements (vector)	$\mathbf{T}_{\mathcal{H}}$	Stacked Hessian operator
$\mathbf{z}$	the image of $\mathrm{T}_q\mathbf{x}$ (vector)	$\mathbf{T}_{\mathcal{P}}$	Patch extraction operator
$\mathbf{I}_K$	identity matrix	f_q	function in the cost term

Moreover, $\mathbf{A}^*\mathbf{S}^*\mathbf{S}\mathbf{A} = (\mathbf{I}_K \otimes \mathbf{A}_0^*\mathbf{S}_0^*)(\mathbf{I}_K \otimes \mathbf{S_0}\mathbf{A_0}) = \mathbf{I}_K \otimes \mathbf{A}_0^*\mathbf{S}_0^*\mathbf{S}_0\mathbf{A}_0$ is a block-diagonal matrix whose action is to apply the filter $\mathbf{A}_0^*\mathbf{S}_0^*\mathbf{S}_0\mathbf{A}_0$ to every image in the stack of K images. So, $(\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{A}\mathbf{S}\mathbf{M})^{-1} = \left(\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{A}_0^*\mathbf{S}_0^*\mathbf{S}_0\mathbf{A}_0\mathbf{M}_k\right)^{-1}$. In general, this inverse operator cannot be further simplified. It can be applied numerically using an iterative algorithm like the conjugate gradient (see [6]). In the remainder of this section, we discuss cases in which a decomposition is possible.

Case of annihilating modulations Let us assume that $\mathbf{S_0}\mathbf{A_0} = \mathbf{I}_N$. This would have little interest in practice as the resolution is not degraded, given that no low pass filtering is considered. However, we will see that it can help us formalizing the reconstruction algorithm proposed in [3]. In this specific case,

$$\hat{\mathbf{x}} = \mathbf{M}^\dagger \mathbf{y} = \Big(\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{M}_k\Big)^{-1}\Big(\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{y}_k\Big). \tag{24}$$

The operator $(\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{M}_k)^{-1}$ is a simple pixelwise division by the sum of the squared pixel values of the K modulation patterns $\mathbf{m}_k$. In the case where the modulation operator $\mathbf{M}$ is injective, then the division is well defined and the estimator exists. Moreover, considering sinusoidal modulation as defined by Eq. (2) and the specific case where the phase shift are defined as $\varphi_k = \pi k/K$ and the amplitudes are equal, such that $\alpha_K = \alpha$ for $k \in [1, \cdots, K]$, then we have $\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{M}_k = (1 + \alpha^2/2)K$. Therefore the reconstruction amounts to simply modulating the acquired images and normalizing by this constant. However, this ideal case is never encountered in practice.

Case of separable modulations with no blur operator Further on, let us consider the case where the modulation matrix can be decomposed as:

$$\mathbf{M} = \mathbf{P}_1\mathbf{\Phi}\mathbf{P}_2\mathbf{\Omega} \tag{25}$$

where matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are $KN \times KN$ permutation matrices, $\mathbf{\Phi}$ a full column rank matrix and $\mathbf{\Omega}$ a tight frame such that $\mathbf{\Omega}^*\mathbf{\Omega} = c\mathbf{I}_{KN}$. Then, the estimate $\hat{\mathbf{x}}$ can be obtained as:

$$\hat{\mathbf{x}} = (\mathbf{P}_1\mathbf{\Phi}\mathbf{P}_2\mathbf{\Omega})^\dagger \mathbf{y} = \tfrac{1}{c}\mathbf{\Omega}^*\mathbf{P}_2^*\mathbf{\Phi}^\dagger\mathbf{P}_1^*\mathbf{y} \tag{26}$$

Indeed, using the properties of the permutation matrices we have $\mathbf{P}_1^*\mathbf{P}_1 = \mathbf{I}_{KN}$ and since by definition, we have $\mathbf{\Phi}^\dagger\mathbf{\Phi} = \mathbf{I}_{KN}$ as well, it follows that:

$$\left(\tfrac{1}{c}\mathbf{\Omega}^*\mathbf{P}_2^*\mathbf{\Phi}^\dagger\mathbf{P}_1^*\right)\left(\mathbf{P}_1\mathbf{\Phi}\mathbf{P}_2\mathbf{\Omega}\right) = \tfrac{1}{c}\mathbf{\Omega}^*\mathbf{\Omega} = \mathbf{I}_{KN} \tag{27}$$

using the tight frame property of $\mathbf{\Omega}$.

This decomposition of the operator $\mathbf{M}$ is particularly well suited to describe the set of sinusoidal modulation for a single pattern frequency pair $(\omega_{1,k}, \omega_{2,k}) = (\omega_1, \omega_2)$ but with several phase shifts φ_k as defined by Eq. (2). In this context, the matrix $\mathbf{\Omega}$ is a of size $N\times 3$ with the following shape: $\mathbf{\Omega} = [\mathbf{a}, \mathbf{c}, \mathbf{s}]$ where the two vectors $\mathbf{c}$ and $\mathbf{s}$ are defined for $n = (n_1, n_2) \in N_1\times N_2$ by:

$$\begin{cases} \mathbf{a}_{n_1,n_2} & = & 1 \\ \mathbf{c}_{n_1,n_2} & = & \cos(n_1\omega_1 + n_2\omega_2) \\ \mathbf{s}_{n_1,n_2} & = & \sin(n_1\omega_1 + n_2\omega_2) \end{cases} \tag{28}$$

The corresponding $\mathbf{\Phi}$ matrix is then of the form $\mathbf{I}_K \otimes \mathbf{\Phi}_0$ with:

$$\mathbf{\Phi}_0 = \begin{bmatrix} 1 & \alpha_1\cos(\varphi_1) & -\alpha_1\sin(\varphi_1) \\ \vdots & \vdots & \vdots \\ 1 & \alpha_K\cos(\varphi_K) & -\alpha_K\sin(\varphi_K) \end{bmatrix} \tag{29}$$

and we have in this case the relationship: $\mathbf{\Phi}^\dagger = \mathbf{I}_N \otimes \mathbf{\Phi}_0^\dagger$. Finally, the least-squares estimate can be rewritten as:

$$\hat{\mathbf{x}} = \tfrac{1}{c}\mathbf{\Omega}^*\mathbf{P}_2^*\left(\mathbf{I}_K \otimes \mathbf{\Phi}_0^\dagger\right)\mathbf{P}_2\mathbf{y}. \tag{30}$$

Case of separable modulation with blur operator Now in order to understand the effect of the point spread function $\mathbf{A}_0$ on the reconstruction, we consider the case where we can write:

$$\mathbf{y} = \mathbf{P}_1\mathbf{\Phi}\mathbf{P}_2\mathbf{S}\mathbf{A}\mathbf{\Omega}\mathbf{x} \tag{31}$$

where again the matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ are $KN\times KN$ permutation matrices, $\mathbf{\Phi}$ a full column rank matrix and $\mathbf{\Omega}$ a tight frame such that $\mathbf{\Omega}^*\mathbf{\Omega} = c\mathbf{I}_{KN}$. The permutation of the operator $\mathbf{\Phi}$ and $\mathbf{SA}$ is possible when the matrix $\mathbf{\Phi}$ only operates on different images and not on pixels (see [4]). This is the case for sinusoidal modulations used in SIM and we have then:

$$\hat{\mathbf{x}} = (\mathbf{S}\mathbf{A}\mathbf{\Omega})^\dagger\mathbf{P}_1(\mathbf{I}_K \otimes \mathbf{\Phi}_0^\dagger)\mathbf{P}_2 \tag{32}$$

which leads to the least-squares solution:

$$\hat{\mathbf{x}} = (\mathbf{S}\mathbf{A}\mathbf{\Omega}\mathbf{\Omega}^*\mathbf{A}^*\mathbf{S}^*)^{-1}\mathbf{\Omega}^*\mathbf{A}^*\mathbf{S}^*\mathbf{P}_1(\mathbf{I}_K \otimes \mathbf{\Phi}_0^\dagger)\mathbf{P}_2\mathbf{y} \tag{33}$$

where the operator $\mathbf{S}\mathbf{A}\mathbf{\Omega}\mathbf{\Omega}^*\mathbf{A}^*\mathbf{S}^*$ can be inverted in Fourier domain and corresponds to a deconvolution by the sum of the modulated point spread function. Note that this direct solution is only valid for a limited type of modulations and it corresponds to the steps described in e.g. [3, 4, 47] where the modulation components are separated using the phase information by inverting the matrix $\mathbf{\Phi}_0$ and the resulting components are modulated (corresponding to the action of $\mathbf{\Omega}^*$) in order to shift back the frequencies components as described in Fig. 3.

B. Cost terms

We list here the implementation details for the other tested cost functions used in the numerical experiments.

Least-squares SIM (LS) When considering an additive Gaussian white noise model, the negative log-likelihood leads to a least-squares approach. The least-squares data term for SIM imaging is defined by $\frac{1}{2}\|\mathbf{y}-\mathbf{SAMx}\|_2^2$, corresponding to the combination of the function $f_{\mathrm{LS}} = \frac{1}{2}\|\cdot-\mathbf{y}\|_2^2$ and the linear operator $\mathbf{T}_{\mathrm{LS}} = \mathbf{SAM}$. The proximity operator [35] associated to f_{LS} is then [28] : $\forall\gamma > 0, \forall\mathbf{z} \in \mathbb{R}^{LK}, \mathrm{prox}_{\gamma f_{\mathrm{LS}}}(\mathbf{z}) = (\mathbf{z} + \gamma\ \mathbf{y})/(1+\gamma)$.

Gradient squared ℓ_2–norm ($\|\nabla\|^2$) While more efficient algorithms exist for minimizing the squared ℓ_2-norm of the gradient of $\mathbf{x}$, especially combined with a least-squares data term, we may still use the proposed approach. In this case, the operator is defined by the two first order derivatives along the horizontal $\mathbf{D}_1$ and vertical $\mathbf{D}_2$ directions, stacked together as $\mathbf{T}_{\mathcal{D}} = [\mathbf{D}_1, \mathbf{D}_2]^{\mathrm{T}}$. The adjoint of this operator is then the opposite of the divergence operator defined as $\mathbf{T}^*_{\mathcal{D}} = \mathbf{D}_1^{\mathrm{T}}\mathbf{z}_1 + \mathbf{D}_2^{\mathrm{T}}\mathbf{z}_2$ where $\mathbf{z}_1$ and $\mathbf{z}_2$ are the gradient components. The gradient $\mathbf{D}_1$ and are $\mathbf{D}_2$ computed using a forward finite differences scheme and their adjoints $\mathbf{D}_1^{\mathrm{T}}$ and $\mathbf{D}_2^{\mathrm{T}}$ are backward finite differences with Neumann boundary conditions in both cases. For Tikhonov regularization, the associated function is then the squared ℓ_2–norm, i.e. $f_{\mathcal{D}} = \|\cdot\|^2$ whose proximity operator in this case is given by $\mathrm{prox}_{\gamma f_{\mathcal{D}}}(\mathbf{z}) = \mathbf{z}/(1+\gamma)$ for every $\mathbf{z} \in \mathbb{R}^{2N}$.

Laplacian squared ℓ_2–norm ($\|\Delta\|^2$) A Laplacian squared ℓ_2–norm regularization was introduced in [6] for SIM image reconstruction. We can consider this regularization using the proposed minimization algorithm by combining the squared ℓ_2–norm with the Laplacian operator $\mathbf{T}_{\mathcal{L}} = \mathbf{D}_{11}^2 + \mathbf{D}_{22}^2$ where $\mathbf{D}_{11}^2$ and $\mathbf{D}_{22}^2$ are the second order derivatives in the horizontal and vertical directions. Note that the Laplacian operator is self-adjoint. Furthermore, we can use here the same function $f_{\mathcal{L}} = \|\cdot\|^2$ and the associated proximity operator as for Tikhonov regularization. Note that in the context of this study, unlike in [6], we do not consider the posterior mean estimate but only a maximum *a posteriori* (MAP) estimate.

Total variation (TV) The total variation seminorm can be defined as the ℓ_1–norm of the gradients of $\mathbf{x}$ [42, 43]. Therefore, we can use this time the same operator $\mathbf{T}_{\mathcal{D}}$ as for Tikhonov regularization, but with a different function f. Indeed, in order to achieve an isotropic total variation, a vectorial form of the ℓ_1–norm denoted by $f_{\mathrm{TV}} = \|\cdot\|_{1,2}$ should be applied, by considering the two gradient components as a vector [28]:

$$(\forall\mathbf{z} = [\mathbf{z}_1^\top, \mathbf{z}_2^\top]^\top) \in \mathbb{R}^{2N} \quad \|\mathbf{z}\|_{1,2} = \textstyle\sum_{n=1}^{N} \sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2}. \tag{34}$$

Then the proximity operator is applied component-wise for $n \in [1, N]$ as:

$$(\forall\mathbf{z}_n \in \mathbb{R}^2) \quad \mathrm{prox}_{\|\cdot\|_{1,2}}(\mathbf{z}_n) = \begin{cases} \mathbf{z}_n - \frac{\gamma\mathbf{z}_n}{\sqrt{[\mathbf{z}_1]_n^2+[\mathbf{z}_2]_n^2}}, & \sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2} \geq \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

Schatten norm of the Hessian operator ($\mathcal{S}_p(\mathrm{T}_{\mathcal{H}})$) Recently, a new regularization based on the Schatten norm of the Hessian operator has been proposed [13]. This approach has been developed in order to reduce the staircase artifacts observed with total variation regularization.

In order to include this regularization constraint, we consider the Hessian operator defined at each location $n \in \{1, \dots, N\}$ as:

$$[\mathbf{T}_{\mathcal{H}}\mathbf{x}]_n = \begin{bmatrix} [\mathbf{D}^2_{11}\mathbf{x}]_n & [\mathbf{D}^2_{12}\mathbf{x}]_n \\ [\mathbf{D}^2_{12}\mathbf{x}]_n & [\mathbf{D}^2_{22}\mathbf{x}]_n \end{bmatrix} \tag{36}$$

and composed of the second order derivative along horizontal, diagonal and vertical direction denoted respectively $\mathbf{D}^2_{11}$, $\mathbf{D}^2_{12}$ and $\mathbf{D}^2_{22}$. The adjoint of this operator is defined by:

$$\mathbf{T}^*_{\mathcal{H}}\mathbf{z} = \mathbf{D}^{2*}_{11}\mathbf{z}_{11} + \mathbf{D}^{2*}_{12}(\mathbf{z}_{12} + \mathbf{z}_{21}) + \mathbf{D}^{2*}_{22}\mathbf{z}_{22} \tag{37}$$

for every

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_{11} & \mathbf{z}_{12} \\ \mathbf{z}_{21} & \mathbf{z}_{22} \end{bmatrix}$$

where $\mathbf{z}_{11}$, $\mathbf{z}_{12} = \mathbf{z}_{21}$ and $\mathbf{z}_{22}$ represent the four components of the Hessian operator, each of size $\mathbb{R}^N$.

In a similar way to the nuclear norm, the Schatten norm $\mathcal{S}_p$ of $\mathbf{z}_n \in \mathbb{R}^{2\times 2}$ is defined as the ℓ_p–norm of the diagonal matrix $\mathbf{\Lambda}_n$ such that $\mathbf{z}_n = \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{V}_n^{\mathrm{T}}$, and the proximity operator has the following expression [16]:

$$(\forall \mathbf{z}_n \in \mathbb{R}^{2\times 2}) \quad \mathrm{prox}_{\gamma\mathcal{S}_p}(\mathbf{z}_n) = \mathbf{U}_n \mathrm{prox}_{\gamma\|\cdot\|_p}(\mathbf{\Lambda}_n)\mathbf{V}_n^{\mathrm{T}}. \tag{38}$$

Nonlocal total variation (NLTV) The nonlocal total variation (NLTV) penalization was introduced in [14] and extended to various inverse problems in [15,48] by considering differential operators defined on the graph associated to the sites of the image grid. It was also recently extended to multispectral images in [16]. The operator associated to the NLTV regularization can be described as weighted nonlocal gradients defined as [49]:

$$[\mathbf{T}_{\mathrm{NL}}\mathbf{x}]_n = \begin{bmatrix} \left[\mathbf{W}_1(\mathbf{F}_1\mathbf{x} - \mathbf{x})\right]_n \\ \vdots \\ \left[\mathbf{W}_T(\mathbf{F}_T\mathbf{x} - \mathbf{x})\right]_n \end{bmatrix} \tag{39}$$

where for $t \in 1, \dots, T$, we define some diagonal weight matrices as functions of the distance between patches $\mathbf{W}_t = \mathrm{diag}\left(\exp\left(-\frac{1}{\eta}\mathbf{B}(\mathbf{F}_t\tilde{\mathbf{x}} - \tilde{\mathbf{x}})^2\right)\right)$ with $\mathbf{F}_t$ a translation operator and $\mathbf{B}$ a convolution by a lowpass filter such as a box-filter, or a Gaussian filter and η a positive scalar. The image $\tilde{\mathbf{x}}$ can be obtained by minimizing the classical total variation for example. Note that the computation of the convolution could be done using separable recursive filters as proposed in [50]. However, since the estimation of the weights is performed only once, this step is not critical in terms of computation time. The T translations $\mathbf{F}_t$ are chosen so that they describe a square neighborhood of size $N_w \times N_w$ while the operator $\mathbf{B}$ corresponding to an image patch whose size $N_p \times N_p$ is given by the width of the support of the filter in the case of a box–filter. The adjoint of the operator $\mathbf{T}_{\mathrm{NL}}$ is defined by:

$$(\forall \mathbf{z} \in \mathbb{R}^{TN}) \quad \mathbf{T}^*_{\mathrm{NL}}\mathbf{z} = \sum_{t=1}^{T} \mathbf{W}_t(\mathbf{F}^*_t - \mathbf{I})\mathbf{z}_t, \tag{40}$$

where $\mathbf{F}_t^*$ with $t \in 1, \ldots, T$ are the translation with the corresponding opposite directions. The function associated to the NLTV regularization is a $\ell_{1,2}$–norm defined by:

$$(\forall \mathbf{z} \in \mathbb{R}^{TN}) \quad \|\mathbf{z}\|_{1,2} = \sum_{n=1}^{N} \left(\sum_{t=1}^{T} \mathbf{z}_{n,t}^2 \right)^{\frac{1}{2}}. \tag{41}$$

The associated proximity operator is then defined by:

$$(\forall \mathbf{z}_n \in \mathbb{R}^T) \quad \mathrm{prox}_{\gamma\|\cdot\|_{1,2}}(\mathbf{z}_n) = \begin{cases} z_n - \frac{\gamma \mathbf{z}_n}{\sqrt{\sum_{t=1}^T [\mathbf{z}_t]_n^2}}, & \sqrt{\sum_{t=1}^T [\mathbf{z}_t]_n^2} \geq \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

References

[1] Schermelleh L, Heintzmann R, Leonhardt H. A guide to super-resolution fluorescence microscopy. The Journal of Cell Biology. 2010 Jul;190(2):165–175.

[2] Heintzmann R, Cremer CG. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. In: SPIE Optical Biopsies and Microscopic Techniques III. vol. 3568; 1999. p. 185–196.

[3] Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. Journal of Microscopy. 2000 May;198(2):82–87.

[4] Schaefer LH, Schuster D, Schaffer J. Structured illumination microscopy: artefact analysis and reduction utilizing a parameter optimization approach. Journal of microscopy. 2004 Nov;216(2):165–174.

[5] O'Holleran K, Shaw M. Optimized approaches for optical sectioning and resolution enhancement in 2D structured illumination microscopy. Biomedical Optics Express. 2014;5(8):2580.

[6] Orieux F, Sepulveda E, Loriette V, Dubertret B, Olivo-Marin JC. Bayesian estimation for optimized structured illumination microscopy. IEEE Transactions on Image Processing. 2012 Feb;21(2):601–614.

[7] Mudry E, Belkebir K, Girard J, Savatier J, Moal EL, Nicoletti C, et al. Structured illumination microscopy using unknown speckle patterns. Nature Photonics. 2012;6(5):312–315.

[8] Min J, Jang J, Keum D, Ryu SW, Choi C, Jeong KH, et al. Fluorescent microscopy beyond diffraction limits using speckle illumination and joint support recovery. Scientific Reports. 2013;3.

[9] Negash A, Labouesse S, Sandeau N, Allain M, Giovannini H, Idier J, et al. Improving the axial and lateral resolution of three-dimensional fluorescence microscopy using random speckle illuminations. Journal of the Optical Society of America A. 2016 Jun;33(6):1089.

[10] Lukeš T, Křížek P, Švindrych Z, Benda J, Ovesný M, Fliegel K, et al. Three-dimensional super-resolution structured illumination microscopy with maximum a posteriori probability image estimation. Optics Express. 2014 Dec;22(24):29805–29817.

[11] Chu K, McMillan PJ, Smith ZJ, Yin J, Atkins J, Goodwin P, et al. Image reconstruction for structured-illumination microscopy with low signal level. Optics Express. 2014 Apr;22(7):8687.

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

[13] Lefkimmiatis S, Ward JP, Unser M. Hessian Schatten-norm regularization for linear inverse problems. IEEE Transactions on Image Processing. 2013 May;22(5):1873–1888.

[14] Gilboa G, Osher S. Nonlocal operators with applications to image processing. Multiscale Modeling & Simulation. 2008 Nov;7(3):1005–1028.

[15] Bougleux S, Peyré G, Cohen L. Non-local regularization of inverse problems. Inverse Problems and Imaging. 2011 May;5(2):511–530.

[16] Chierchia G, Pustelnik N, Pesquet-Popescu B, Pesquet JC. A nonlocal structure tensor-based approach for multicomponent image recovery problems. IEEE Transactions on Image Processing. 2014 Dec;23(12):5531–5544.

[17] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing. 2006 Nov;54(11):4311–4322.

[18] Ma L, Moisan L, Yu J, Zeng T. A dictionary learning approach for Poisson image deblurring. IEEE Transactions on Medical Imaging. 2013 Jul;32(7):1277–1289.

[19] Danielyan A, Katkovnik V, Egiazarian K. BM3D frames and variational image deblurring. IEEE Transactions on Image Processing. 2012 Apr;21(4):1715–1728.

[20] Boulanger J, Pustelnik N, Condat L. Non-smooth convex optimization for an efficient reconstruction in structured illumination microscopy. In: 11th International Symposium on Biomedical Imaging (ISBI'14); 2014. p. 995–998.

[21] Gustafsson MG. Extended resolution fluorescence microscopy. Current Opinion in Structural Biology. 1999 Oct;9(5):627–628.

[22] York AG, Parekh SH, Nogare DD, Fischer RS, Temprine K, Mione M, et al. Resolution doubling in live, multicellular organisms via multifocal structured illumination microscopy. Nature Methods. 2012 Jul;9(7):749–754.

[23] Goodman J W. Introduction to Fourier optics. McGraw-Hill Physical and Quantum Electronics Series. New York: McGraw-Hill; 1968.

[24] Starck JL, Murtagh FD, Bijaoui A. Image processing and data analysis: the multiscale approach. Cambridge University Press; 1998.

[25] Anscombe FJ. The transformation of Poisson, binomial and negative-binomial data. Biometrika. 1948 Dec;35(3/4):246–254.

[26] Bjorck A. Numerical Methods for Least Squares Problems. SIAM; 1996.

[27] Chambolle A, Pock T. An introduction to continuous optimization for imaging. Acta Numerica. 2016;25:161–319.

[28] Combettes PL, Pesquet JC. Proximal splitting methods in signal processing. In: Bauschke HH, Burachik RS, Combettes PL, Elser V, Luke DR, Wolkowicz H, editors. Fixed-Point Algorithms for Inverse Problems in Science and Engineering. vol. 49. New York, NY: Springer New York; 2011. p. 185–212.

[29] Burger M, Sawatzky A, Steidl G. First order algorithms in variational image processing. In: Glowinski R, Osher SJ, Yin W, editors. Splitting methods in communication, imaging, science, and engineering. Scientific Computation. Springer International Publishing; 2016. p. 345–407.

[30] Chambolle A, Pock T. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision. 2010 Dec;40(1):120–145.

[31] Combettes PL, Pesquet JC. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued and Variational Analysis. 2011 Aug;20(2):307–330.

[32] Vũ BC. A splitting algorithm for dual monotone inclusions involving cocoercive operators. Advances in Computational Mathematics. 2011 Nov;38(3):667–681.

[33] Condat L. A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. Journal of Optimization Theory and Applications. 2013 Aug;158(2):460–479.

[34] Komodakis N, Pesquet JC. Playing with Duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. IEEE Signal Processing Magazine. 2015 Nov;32(6):31–54.

[35] Moreau JJ. Proximité et dualité dans un espace hilbertien. Bulletin de la Société Mathématique de France. 1965;93:273–299.

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

[37] Csiszar I. A class of measures of informativity of observation channels. Periodica Mathematica Hungarica. 1972;2(1-4):191–213.

[38] Combettes PL, Pesquet J. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. IEEE Journal of Selected Topics in Signal Processing. 2007 Dec;1(4):564–574.

[39] Chakrabarti A, Zickler T. Image restoration with signal-dependent camera noise. arXiv:12042994 [cs, stat]. 2012 Apr;ArXiv: 1204.2994.

[40] Boulanger J, Kervrann C, Bouthemy P, Elbau P, Sibarita JB, Salamero J. Patch-based nonlocal functional for denoising fluorescence microscopy image sequences. IEEE Transactions on Medical Imaging. 2010 Feb;29(2):442–454.

[41] Cai JF, Candès EJ, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization. 2010;20(4):1956–1982.

[42] Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena. 1992 Nov;60(1–4):259–268.

[43] Condat L. Discrete total variation: new definition and minimization. SIAM Journal on Imaging Sciences. 2017 Jan;10(3):1258–1290.

[44] Kaipio J, Somersalo E. Statistical and computational inverse problems. Applied Mathematical Sciences. New York: Springer-Verlag; 2005.

[45] Calvetti D, Somersalo E. Hypermodels in the Bayesian imaging framework. Inverse Problems. 2008;24(3):034013.

[46] Condat L, Boulanger J, Pustelnik N, Sahnoun S, Sengmanivong L. A 2-D spectral analysis method to estimate the modulation parameters in structured illumination microscopy. In: 11th International Symposium on Biomedical Imaging (ISBI'14); 2014. p. 604–607.

[47] Gustafsson MGL, Shao L, Carlton PM, Wang CJR, Golubovskaya IN, Cande WZ, et al. Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. Biophysical Journal. 2008 Jun;94(12):4957–4970.

[48] Peyré G, Bougleux S, Cohen L. Non-local regularization of inverse problems. In: Forsyth D, Torr P, Zisserman A, editors. Computer Vision – ECCV 2008. No. 5304 in Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2008. p. 57–68.

[49] Chierchia G, Pustelnik N, Pesquet JC, Pesquet-Popescu B. Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools. Signal, Image and Video Processing. 2015;9(8):1737–1749.

[50] Condat L. A simple trick to speed up and improve the non-local means. Caen, France; 2010. research report hal-00512801.