Nonsmooth Convex Optimization for Structured Illumination Microscopy Image Reconstruction

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

Nonsmooth Convex Optimization for Structured Illumination Microscopy Image Reconstruction

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

***Abstract*—Our contribution is twofold: first, we investigate the properties of a linear reconstruction method in SIM, discussing its properties from an estimation point of view and its relationship with variational approaches. Second, capitalizing on this formalism, we adopt a variational approach taking into account the type of noise, typically Poissonian, associated to convex nonsmooth regularizing penalties, like the total variation, which have shown their efficiency to solve many imaging inverse problems. We detail the implementation of the proximal algorithm allowing to solve the problems exactly.**

***Index Terms*—Structured Illumination Microscopy (SIM), imaging inverse problem, image restoration, deconvolution, Poisson noise, proximal algorithm, total variation**

I. Introduction

Super–resolution approaches for fluorescence light microscopy are offering new opportunities for investigating molecular mechnisms in cell biology. While not reaching the spatial resolution of electronic microscopy, their interest lies in their ability to provide functional information by the use of fluorescent tag and to follow them over time [1].

Structured illumination microscopy (SIM) is one of the optical super–resolution methods which have been recently proposed [2], [3]. It is based on the illumination of the sample by a set of interference patterns. These sinusoidal modulations of the excitation of the fluorophores lead to frequency shifts in the Fourier domain which bring inaccessible frequencies within the scope of the optical transfer function leading to the well know Moiré effect. An example of acquired raw data is depicted in Fig. 1. Once post–processed, the acquired images show an increased resolution by typically a factor two as illustrated by Fig. 2 where an acquired image has been reconstructed using a linear method [3]. If this approach is very flexible with respect to the use of fluorescent tag, it requires at least nine images to recover the super–resolved image [2], [3]. Indeed, we will see that three phase shifts for each of the three angles are necessary to disentangle the three frequency components corresponding to a sinusoidal modulation. Several studies have investigated the properties of such reconstruction algorithms and provided solutions for the reduction of artifacts [4], [5].

Recently an elegant Bayesian approach has been proposed

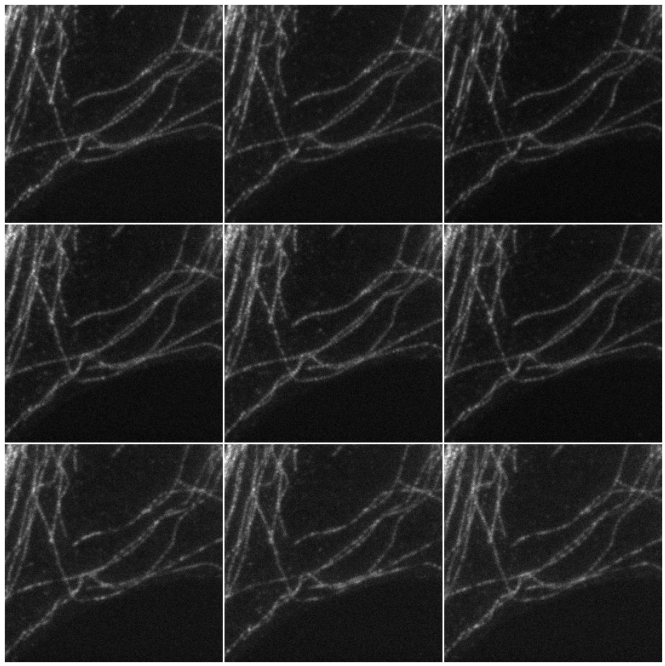

Fig. 1. Example of real data. A Molecular Probe slide was imaged 9 times using a Nikon SIM microscope using a $100\times$ oil objective. The images represent a 256×256 region of 512×512 acquired images and display some labeled microtubules. The modulation pattern can be observed as a slight Moiré effect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

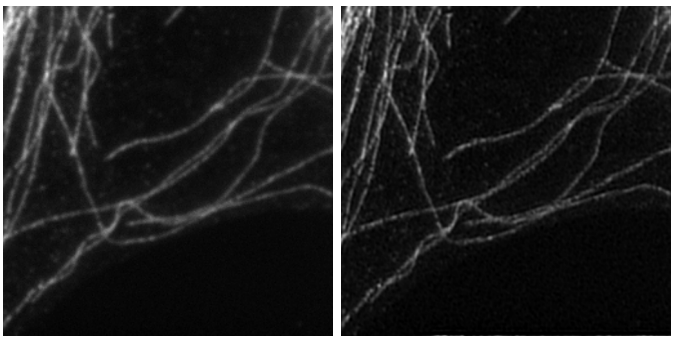

Fig. 2. Reconstruction of the data displayed in Fig. 1. On the left the corresponding classical wide–field microscopy is obtained from the mean of the nine images. On the right, a linear least–squares reconstruction. The actual dimension of the image on the right is twice the size of the image of the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

Jérôme Boulanger is with CNRS UMR144, 75248 Paris Cedex 05, France and Institut Curie, 75248 Paris Cedex 05, France (e-mail: jerome.boulanger@curie.fr).

Laurent Condat is with Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France and with CNRS, GIPSA-Lab, F-38000 Grenoble, France (e-mail: Laurent.Condat@gipsa-lab.fr)

Nelly Pustelnik is with Laboratoire de Physique ENS de Lyon, CNRS UMR5672, Université Lyon I, France (e-mail: nelly.pustelnik@ens-lyon.fr).

Lucie Sengmanivong is with Cell and Tissue Imaging Core Facility (PICT-IBiSA) and Nikon Imaging Centre@Institut Curie-CNRS, , UMR144, Institut Curie, Centre de Recherche, PSL Research University, Membrane Dynamics and Mechanics of Intracellular Signaling Laboratory, Paris Cedex 05, France

Tristan Piolot is with Institut Curie, 26 Rue d'Ulm, 75248, Paris Cedex 05, France and CNRS UMR3215, 75248, Paris Cedex 05, France and INSERM U934, 75248, Paris Cedex 05, France ...

for the reconstruction of SIM images [6]. This approach consider the SIM reconstruction as an inverse problem and allows the authors to incorporate the point spread function (PSF) of the optical system in order to deconvolve and restore the image in a single step using a ℓ_2–norm of the Laplacian operator as a regularizer. A noticeable advantage of this approach is that the number of phases is not constrained anymore and even more complex patterns can be considered. This idea was actually investigated in [7], [8], where the sample is illuminated using the speckle produced by a coherent light source diffracted by a rough surface. Displacing this surface allows to obtain a variation of the speckle and therefore of the random illumination pattern. A blind estimation procedure is then used to recover both the modulations and the image. Finally, a total variation constraint has also been explored in [9].

Formulating the issue of image reconstruction as an inverse problem often leads to a variational approach, where a functional defined as the sum of a data fitting term and a regularization term is minimized. The choice of these two terms is application–dependent and is the object of a vast literature.

The contributions in this article are the following:

- We reformulate the least–squares estimator and discuss its properties for different types of modulation patterns.
- We describe a regularized least–squares estimate.
- We apply a recent primal–dual approach for the minimization of convex nonsmooth functionals to SIM image reconstruction.
- We investigate the two-by-two combination of several data terms and regularization terms for SIM image reconstruction and provide the implementation details.
- We propose a new regularization approach for inverse problems based on the Schatten norm of a patch dictionary.
- We evaluate the proposed approach on simulated and real images.

This paper is organized as follow. We first propose in Section III a comprehensive analysis of the properties of linear least-squares reconstruction of SIM images in the case of sinusoidal modulations. We also derive a regularized least–squares approach in Section IV. We further consider in Section V nonlinear reconstruction methods taking into account the specificity of Poisson noise present in video–microscopy, while exploring different regularization approaches. Finally, in Section VI we test the proposed approaches on synthetic and real data.

In the sequel, $\mathbf{I}_N$ 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. To conclude, we define the proximal 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:

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

II. Structured Illumination Microscopy Model

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

$$\bar{\mathbf{y}}_k = \mathbf{S}_0 \mathbf{A}_0 \mathbf{M}_k \bar{\mathbf{x}} \tag{2}$$

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 $\mathbf{M}_k = \mathrm{diag}(\mathbf{m}_k)$. Traditionally, modulations are the result of interfering coherent laser beams and can be represented by a sinusoidal pattern defined for each points of coordinates $(n_1, n_2) \in N_1 \times N_2$ as:

$$[\mathbf{m}_k]_{n_1,n_2} = 1 + \alpha_k \cos(n_1\omega_{1,k} + n_2\omega_{2,k} + \varphi_k) \tag{3}$$

where α_k is the amplitude of the modulation, $\omega_{1,k}$ and $\omega_{2,k}$ are the modulations frequencies and φ_k a phase. In the following, we will stack all the modulations $\mathbf{M}_k$ in the matrix $\mathbf{M} = [\mathbf{M}_1, \cdots, \mathbf{M}_K]^{\mathrm{T}}$

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 wide–field microscopy by the auto–correlation of the pupil function as [6], [10]:

$$\mathcal{A}_0(\varrho) = \begin{cases} \frac{2}{\pi}\left(\arccos\left(\frac{\varrho}{2\varrho_0}\right) - \frac{\varrho}{2\varrho_0}\sqrt{1-\left(\frac{\varrho}{2\varrho_0}\right)^2} \right) & \varrho \le \varrho_0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $\varrho = \sqrt{\xi_1^2 + \xi_2^2}$ is the module of the frequencies in polar coordinate and ϱ_0 the cut–off 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 dimensions. In the rest of the text, down–sampling for the set of K images is represented by the operator $\mathbf{S} = \mathbf{I}_K \otimes \mathbf{S}_0$.

We can now conveniently rewrite Eq. (2) as:

$$\bar{\mathbf{y}} = \mathbf{S}\mathbf{A}\mathbf{M}\bar{\mathbf{x}}. \tag{5}$$

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

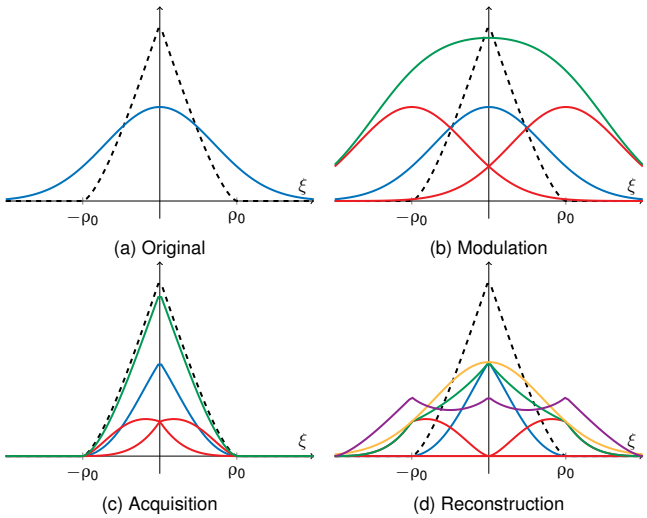

Fig. 3. Principle of structured illumination microscopy illustrated in one dimension. (a) Spectrum of x in blue and the optical transfer function $\mathcal{A}_0$ in black (b) Spectrum of a modulated image $\mathbf{x}_n \cdot (1 + \cos(\omega n + \varphi))$ (green) as the sum of the three components 1, $e^{\pm i(\omega\xi+\varphi)}$ (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 super–resolved image obtained by shifting the modulated components (red) and summing them (green). Finally, normalizing the components by tacking into account the shape of the OTF results allows to recover the original image (yellow).

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 possible for the optical system to capture information at frequencies above the cut–off 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) a normalization step equivalent to the ratio of the demodulated images (green curve) with the shifted OTFs (purple curve) is necessary and 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 and the thermal agitation of the electrons. To take into account those degradations, a general noise model can be written as [11]:

$$\mathbf{y} = \kappa\ \mathbf{p} + \mathbf{n} \tag{6}$$

where κ is the overall gain of the acquisition system, $\mathbf{p}$ is a vector of Poisson distributed random variables of parameter $(\bar{\mathbf{y}} - m_{\mathrm{DC}})/\kappa$ and $\mathbf{n}$ a vector of Normally distributed random variables of mean m_{DC} and variance σ^2_{DC}. This formulation insures 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 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. However, if the dynamic range of $\bar{\mathbf{y}}$ is limited and the number of photons high enough (say, ≥ 30) the additive Gaussian white noise model might be accurate enough to capture the signal degradation. In the other hand, if the electronic noise represented by $\mathbf{n}$ is negligible and assuming a unit gain $\kappa = 1$, the Poisson noise model is valid. Note that the use of a variance stabilization transform [12] would introduce non–linearities which would have a significant impact on the observation model (2) and make the reconstruction process intractable.

III. Least–squares Solution

A. General Case

Let us start by considering the additive Gaussian white noise approximation. In this case, maximum–likelihood estimation of $\bar{\mathbf{x}}$ given the observed data $\mathbf{y}$ amounts to solve the least–squares problem:

$$\hat{\mathbf{x}} \in \operatorname*{Arg\,min}_{\mathbf{x}} \|\mathbf{y} - \mathbf{SAMx}\|^2. \tag{7}$$

Hence,

$$\hat{\mathbf{x}} = (\mathbf{SAM})^\dagger \mathbf{y}. \tag{8}$$

If we recall the fact that the operator $\mathbf{SAM}$ is not injective for microscopy data then the least–squares solution is not unique. However, for the sake of the demonstration, let's assume that this is the case in this section (e.g. by assuming that the operator $\mathbf{A}_0$ is a Gaussian blur). Therefore, we can write:

$$\hat{\mathbf{x}} = (\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{SAM})^{-1}\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{y}. \tag{9}$$

We have

$$\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{y} = \mathbf{M}^*[(\mathbf{A}_0^*\mathbf{S}_0^*\mathbf{y}_1^{\mathrm{T}})^{\mathrm{T}}\ \cdots\ (\mathbf{A}_0^*\mathbf{S}_0^*\mathbf{y}_K^{\mathrm{T}})^{\mathrm{T}}]^{\mathrm{T}} \tag{10}$$

$$= \sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{A}_0^*\mathbf{S}_0\ \mathbf{y}_k. \tag{11}$$

Thus, it is remarkable 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.

Moreover, $\mathbf{A}^*\mathbf{S}^*\mathbf{SA} = (\mathbf{I}_K \otimes \mathbf{A}_0^*\mathbf{S}_0^*)(\mathbf{I}_K \otimes \mathbf{S}_0\mathbf{A}_0) = \mathbf{I}_K \otimes \mathbf{A}_0^*\mathbf{S}_0^*\mathbf{S}_0\mathbf{A}_0$ 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, $(\mathbf{M}^*\mathbf{A}^*\mathbf{S}^*\mathbf{ASM})^{-1} = \left(\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{A}_0^*\mathbf{S}_0^*\mathbf{S}_0\mathbf{A}_0\mathbf{M}_k\right)^{-1}$. 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.

B. 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,

$$\hat{\mathbf{x}} = \mathbf{M}^\dagger\mathbf{y} = \left(\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{M}_k\right)^{-1}\left(\sum_{k=1}^{K} \mathbf{M}_k^*\mathbf{y}_k\right). \tag{12}$$

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. (3) and the specific case where the phases 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 perfectly encountered in practice.

C. Case of separable modulations with no blur operator

Further on, let us consider the case where the modulation matrix can be decomposed as:

$$\mathbf{M} = \mathbf{P}_1 \mathbf{\Phi} \mathbf{P}_2 \mathbf{\Omega} \tag{13}$$

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:

$$\hat{\mathbf{x}} = (\mathbf{P}_1 \mathbf{\Phi} \mathbf{P}_2 \mathbf{\Omega})^\dagger \mathbf{y} = \tfrac{1}{c} \mathbf{\Omega}^* \mathbf{P}_2^* \mathbf{\Phi}^\dagger \mathbf{P}_2 \mathbf{y} \tag{14}$$

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:

$$\left(\tfrac{1}{c} \mathbf{\Omega}^* \mathbf{P}_2^* \mathbf{\Phi}^\dagger \mathbf{P}_1^*\right) (\mathbf{P}_1 \mathbf{\Phi} \mathbf{P}_2 \mathbf{\Omega}) = \tfrac{1}{c} \mathbf{\Omega}^* \mathbf{\Omega} = \mathbf{I}_{KN} \tag{15}$$

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 phases shifts φ_k as defined by Eq. (3). 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:

$$\left\{ \begin{array}{lcl} \mathbf{a}_{n_1,n_2} & = & 1 \\ \mathbf{c}_{n_1,n_2} & = & \cos(n_1\omega_1 + n_2\omega_2) \\ \mathbf{s}_{n_1,n_2} & = & \sin(n_1\omega_1 + n_2\omega_2) \end{array} \right. \tag{16}$$

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

$$\mathbf{\Phi}_0 = \begin{bmatrix} 1 & \alpha_1 \cos(\varphi_1) & -\alpha_1 \sin(\varphi_1) \\ \vdots & \vdots & \vdots \\ 1 & \alpha_K \cos(\varphi_K) & -\alpha_K \sin(\varphi_K) \end{bmatrix} \tag{17}$$

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:

$$\widehat{\mathbf{x}} = \tfrac{1}{c} \mathbf{\Omega}^* \mathbf{P}_2^* \left(\mathbf{I}_K \otimes \mathbf{\Phi}_0^\dagger\right) \mathbf{P}_2 \mathbf{y} \tag{18}$$

D. 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:

$$\mathbf{y} = \mathbf{P}_1 \mathbf{\Phi} \mathbf{P}_2 \mathbf{S} \mathbf{A} \mathbf{\Omega} \mathbf{x} \tag{19}$$

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:

$$\widehat{\mathbf{x}} = (\mathbf{S}\mathbf{A}\mathbf{\Omega})^\dagger \mathbf{P}_1 (\mathbf{I}_K \otimes \mathbf{\Phi}_0^\dagger) \mathbf{P}_2 \tag{20}$$

which leads to the least–squares solution:

$$\widehat{\mathbf{x}} = (\mathbf{S}\mathbf{A}\mathbf{\Omega}\mathbf{\Omega}^*\mathbf{A}^*\mathbf{S}^*)^{-1} \mathbf{\Omega}^* \mathbf{A}^* \mathbf{S}^* \mathbf{P}_1 (\mathbf{I}_K \otimes \mathbf{\Phi}_0^\dagger) \mathbf{P}_2 \tag{21}$$

where the operator $\mathbf{S}\mathbf{A}\mathbf{\Omega}\mathbf{\Omega}^*\mathbf{A}^*\mathbf{S}^*$ can be inverted in Fourier space and corresponds to the deconvolution by 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], [13] 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.

IV. Regularized Least–Squares Solution

In the previous section, we assumed that the operator $\mathbf{SAM}$ was injective, but we know from Section II that this assumption is not valid in microscopy. Moreover, even if the operator would be injective, this one would probably remain badly conditionned and the maximum likelihood estimator would amplify the noise leading to a non valuable estimate. A common approach is to regularize the problem by adding a constraint on the solution:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \tfrac{1}{2} \|\mathbf{y} - \mathbf{S}\mathbf{A}\mathbf{M}\mathbf{x}\|^2 + \lambda \|\mathbf{L}\mathbf{x}\|^2, \tag{22}$$

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

We saw that, in general, no closed form exists for the least–squares solution. This remains true for regularized least–squares and minimization algorithms such as the conjugate gradient are needed, as proposed in [6], [7].

However, in the specific case of separable sinusoidal modulations, we can obtain a closed form for estimation of the SIM image:

$$\mathbf{x}^\dagger = (\mathbf{S}\mathbf{A}\mathbf{\Omega}\mathbf{\Omega}^*\mathbf{A}^*\mathbf{S}^* + \lambda \mathbf{L}^*\mathbf{L})^{-1} \mathbf{\Omega}^* \mathbf{A}^* \mathbf{S}^* \mathbf{P}_1 (\mathbf{I}_K \otimes \mathbf{\Phi}_0^\dagger) \mathbf{P}_2 \tag{23}$$

This approach is related to the original reconstruction proposed by [3] when using $\mathbf{L}$ as the identity. However, an apodisation term defined in the Fourier domain 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 and Poisson–Gaussian noise via variational appraoches involving nonsmooth and non–necessarily finite criterion.

Require: $\mathbf{x}_0 \in \mathbb{R}^{N\times N}$, $\tau > 0$, $\sigma > 0$, $\rho > 0$
1: $\mathbf{z}_0 = 0$, $\mathbf{v}_{0,q} = \mathbf{x}_0$, $q \in 1, \ldots, Q$
2: **for** $r \in 1, \ldots, R$ **do**
3: **for** $q \in 1, \ldots, Q$ **do**
4: $\mathbf{u}_{r,q} = \mathbf{v}_{r,q} + \sigma\ \mathbf{T}_q(2\mathbf{z}_r - \mathbf{x}_r)$
5: $\mathbf{p}_{r,q} = \mathbf{u}_{r,q} - \sigma\ \mathrm{prox}_{f_q/\sigma}(\mathbf{u}_{r,q}/\sigma)$
6: $\mathbf{v}_{r,q} = \rho\ \mathbf{p}_{r,q} + (1-\rho)\ \mathbf{v}_{r,q}$
7: **end for**
8: $\mathbf{z}_r = \mathbf{x}_r - \sum_{q=1}^{Q} \tau \mathbf{T}_q^* \mathbf{v}_{r,q}$
9: $\mathbf{x}_{r+1} = \rho\ \mathbf{z}_r + (1-\rho)\ \mathbf{x}_r$
10: **end for**

Fig. 4. The primal dual minimization algorithm proposed in [16] allows to minimize the energy functional defined by Eq. (24) given that the proximal 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.

V. Nonsmooth Regularization

In order to explore non–finite data fidelity and nonsmooth regularization terms, we formulate the estimation procedure as a sum of Q cost terms defined by:

$$\hat{\mathbf{x}} \in \operatorname*{Arg\,min}_{\mathbf{x}} \sum_{q=1}^{Q} f_q(\mathbf{T}_q\mathbf{x}) \tag{24}$$

where f_q are functions from $\mathbb{R}^{M_q} \to \mathbb{R} \cup \{+\infty\}$ and $\mathbf{T}_q$ operators represented as matrices of size $M_q \times N$. We minimize this criterion using a primal–dual proximal algorithm from [14]–[17] (see Fig. 4) and consider several cases corresponding to the combination of function f_q and operator $\mathbf{T}_q$. Note that since in general the operator associated to SIM imaging is not directly invertible, primal approaches such as PPXA+ [18] and ADMM [19] would require an iterative minimization procedure for the inversion of this operator. The minimization procedure requires the specification of the proximal operators prox_{f_q} associated to the functions f_q and defined by Eq. (2). We will explicit now the list of cost term $f_q(\mathbf{T}_q\mathbf{x})$ corresponding either to a data fidelity term or a regularization term. We will give the expression for each case of the function f_q and its proximal operator. We will also 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 the non–negativity constraint is enforced directly on the iterate at step 9 of Fig. 4. Therefore, we can write $\hat{\mathbf{x}} \in \operatorname{Arg\,min}_{\mathbf{x}\geq 0} f_1(\mathbf{T}_1\mathbf{x}) + \lambda f_2(\mathbf{T}_2\mathbf{x})$ with λ the regularization parameter.

a) Least–squares SIM data term: 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 $f_{\mathrm{LS}}(\mathbf{z}) = \frac{1}{2}\|\mathbf{z} - \mathbf{y}\|_2^2$ and the linear operator $\mathbf{SAM}$. The proximal operator associated to f_{LS} is then [20] : $\forall\gamma > 0, \mathrm{prox}_{\gamma f_{\mathrm{LS}}}(\mathbf{z}) = (\mathbf{z} + \gamma\ \mathbf{y})/(1+\gamma)$ with $\mathbf{z} = \mathbf{SAMx}$.

b) Kullback–Leibler divergence SIM data term: Under a Poisson noise model assumption for the acquired data $\mathbf{y}$, the negative log–likelihood is given by the Kullback–Leilbler (KL) divergence [21]. The KL divergence between $\mathbf{y}$ and $\mathbf{z} = \mathbf{SAMx}$ is defined component–wise for $n \in [1, LK]$ by [22]:

$$D_{\mathrm{KL}}(\mathbf{z}_n) = \begin{cases} \mathbf{z}_n - \mathbf{y}_n \log \mathbf{z}_n, & \mathbf{z}_n, \mathbf{y}_n > 0 \\ \mathbf{z}_n, & \mathbf{z}_n > 0 \text{ and } \mathbf{y}_n = 0 \\ \infty, & \text{otherwise} \end{cases} \tag{25}$$

The function associated to the KL divergence for the vector $\mathbf{z}$ is then the sum over the component $f_{\mathrm{KL}}(\mathbf{z}) = \sum_{n=1}^{LK} D_{\mathrm{KL}}(\mathbf{z}_n)$. The proximal operator is given component-wise for $n \in 1, \ldots, LK$ by:

$$\mathrm{prox}_{\gamma D_{\mathrm{KL}}}(\mathbf{z}_n) = \frac{1}{2}\left(\mathbf{z}_n - \gamma + \sqrt{(\mathbf{z}_n - \gamma)^2 + 4\gamma\mathbf{y}_n}\right) \tag{26}$$

and the proximal operator $\mathrm{prox}_{\gamma f_{\mathrm{KL}}}(\mathbf{z})$ if obtained by applying Eq. (26) for each component of for the vector $\mathbf{z}$.

c) Weighted least–squares: 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

$$f_{\mathrm{WLS}}(\mathbf{x}) = \frac{1}{2}(\mathbf{z} - \mathbf{y})^{\mathrm{T}}\mathbf{Q}^{-1}(\mathbf{z} - \mathbf{y}),$$

where $\mathbf{Q}$ is a diagonal variance matrix. This relationship is linear for most CCD and CMOS sensors and can be estimated using a linear regression of the variance of the noise versus the intensity (see [23] for more details):

$$\mathbb{V}\mathrm{ar}[\mathbf{y}_i] = \kappa\mathbb{E}[\mathbf{y}_i] + \sigma_{\mathrm{DC}}^2 - \kappa\, m_{DC}$$

using the model introduced in equation (6) and with $\mathbb{E}[\mathbf{y}_i]$ and $\mathbb{V}\mathrm{ar}[\mathbf{y}_i]$ the expectation and variance of the random variable $\mathbf{y}_i$. The variance of the noise $\mathbb{V}\mathrm{ar}[\mathbf{y}_i]$ is estimated locally using a maximum of absolute deviation filter (MAD) computed on the pseudo–residuals (normalized Laplacian $\frac{1}{\sqrt{20}}(\mathbf{D}_{11}^2\mathbf{y} + \mathbf{D}_{22}^2\mathbf{y})$) of the image while the mean is estimated using a median filter. The linear regression allows then to estimate the gain κ and the noise variance at the origin $e_{DC} = \sigma_{\mathrm{DC}}^2 - \kappa\, m_{DC}$. The variance-covariance matrix $\mathbf{Q}$ can then be approximated by a diagonal matrix whose elements are given by:

$$\mathbf{Q}_{ii} = \kappa\bar{\mathbf{y}}_i + e_{DC}$$

with $\bar{\mathbf{y}}_i \approx \mathbb{E}[\mathbf{y}_i]$ is the given by the estimation of the intensity by the median filter.

d) Tikhonov regularization: While more efficient algorithms exist for minimizing the squared ℓ_2-norm of the gradients of $\mathbf{x}$ especially combined with a least–squares data term (See Section IV), it is possible to consider Tikhonov regularization in this framework. In this case, the operator is defined by the two first order derivative along the horizontal $\mathbf{D}_1$ and vertical $\mathbf{D}_2$ directions stacked together $\mathbf{T}_{\mathcal{D}} = [\mathbf{D}_1, \mathbf{D}_2]^{\mathrm{T}}$. The adjoint of this operator is then 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 difference scheme and their adjoints $\mathbf{D}_1^{\mathrm{T}}$ and $\mathbf{D}_2^{\mathrm{T}}$ are backward finite difference with in both cases Neumann

boundary conditions. For Tikhonov regularization, the associated function is then the squared ℓ_2–norm whose proximal operator in this case is given by $\text{prox}_{\gamma\|\cdot\|_2^2}(\mathbf{z}) = \mathbf{z}/(1+\gamma)$ with $\mathbf{z} = [\mathbf{D}_1, \mathbf{D}_2]^{\mathrm{T}}\mathbf{x}$.

e) Laplacian squared ℓ_2–norm regularization: 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 $f_{\gamma\|\cdot\|_2^2}$ and the associated proximal operator than for Tikhonov regularization but this time with $\mathbf{z} = (\mathbf{D}_{11}^2 + \mathbf{D}_{22}^2)\mathbf{x}$. 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.

f) Total variation regularization: The total variation seminorm can be defined as the ℓ_1–norm of the gradients of $\mathbf{x}$ [24]. Therefore, we can use this time the same operator $\mathbf{T}_{\mathcal{D}}$ than 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 $\|\cdot\|_{1,2}$ should be applied by considering the two gradient components as a vector [25]:

$$\|\mathbf{z}\|_{1,2} = \textstyle\sum_{n=1}^{N} \sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2} \tag{27}$$

with $\mathbf{z}_1 = \mathbf{D}_1\mathbf{x}$ and $\mathbf{z}_2 = \mathbf{D}_2\mathbf{x}$. Then the proximal operator is applied component–wise for $n \in [1, N]$ as:

$$\text{prox}_{\|\cdot\|_{1,2}}(\mathbf{z}_n) = \begin{cases} 1 - \frac{\gamma \mathbf{z}_n}{\sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2}}, & \sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2} \geq \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

g) Schatten norm of the Hessian operator: Recently a new regularization based on the Schatten norm of the Hessian operator has been proposed [26]. 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 $n \in 1, \ldots, N$ as:

$$[\mathbf{T}_{\mathcal{H}}\mathbf{x}]_n = \begin{bmatrix} [\mathbf{D}_{11}^2\mathbf{x}]_n & [\mathbf{D}_{12}^2\mathbf{x}]_n \\ [\mathbf{D}_{12}^2\mathbf{x}]_n & [\mathbf{D}_{22}^2\mathbf{x}]_n \end{bmatrix} \tag{29}$$

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

$$\mathbf{T}_{\mathcal{H}}^*\mathbf{z} = \mathbf{D}_{11}^{2*}\mathbf{z}_{11} + \mathbf{D}_{12}^{2*}(\mathbf{z}_{12} + \mathbf{z}_{21}) + \mathbf{D}_{22}^{2*}\mathbf{z}_{22} \tag{30}$$

where $\mathbf{z}_{11}$, $\mathbf{z}_{12} = \mathbf{z}_{21}$ and $\mathbf{z}_{22}$ represent the four components of the Hessian operator.

The Schatten norm $\mathcal{S}_p$ of the $[\mathbf{T}_{\mathcal{H}}\mathbf{x}]_n$ is defined as the ℓ_p–norm of the diagonal matrix $\mathbf{\Lambda}_n$ such that $\mathbf{z}_n = [\mathbf{T}_{\mathcal{H}}\mathbf{x}]_n = \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{V}_n^{\mathrm{T}}$. Then, the associated proximal operator is given by [27]:

$$\text{prox}_{\gamma\mathcal{S}_p}(\mathbf{z}_n) = \mathbf{U}_n \text{prox}_{\gamma\|\cdot\|_p}(\mathbf{\Lambda}_n)\mathbf{V}_n^{\mathrm{T}}. \tag{31}$$

h) Non-local total variation: The non–local total variation (NLTV) penalization was introduced in [28] and extended to various inverse problems in [29], [30] by considering differential operator defined on the graph associated to the sites of the image. It was also recently extended to multi–spectral images in [27]. The operator associated to the NLTV regularization can be described as weighted non–local gradients defined as [31]:

$$[\mathbf{T}_{\text{NL}}\mathbf{x}]_n = \begin{bmatrix} [\mathbf{W}_1(\mathbf{F}_1\mathbf{x} - \mathbf{x})]_n \\ \vdots \\ [\mathbf{W}_T(\mathbf{F}_T\mathbf{x} - \mathbf{x})]_n \end{bmatrix} \tag{32}$$

where for $t \in 1, \ldots, T$, we define some diagonal weight matrices as function of the distance between patches $\mathbf{W}_t = \text{diag}\left(\exp\left(-\frac{1}{\eta}\mathbf{B}(\mathbf{F}_t\tilde{\mathbf{x}} - \tilde{\mathbf{x}})^2\right)\right)$ with $\mathbf{F}_t$ a translation operator and $\mathbf{B}$ a convolution by a low pass 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 a separable recursive filters as proposed in [32]. However, since the estimation of the weights is performed only once this step is not critical in term 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 $\mathbf{T}_{\text{NL}}$ is defined by:

$$\mathbf{T}_{\text{NL}}^*\mathbf{z} = \sum_{t=1}^{T} \mathbf{W}_t(\mathbf{F}_t^* - \mathbf{I})\mathbf{z}_t \tag{33}$$

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 vectorial ℓ_1–norm defined by:

$$\|\mathbf{z}\|_{1,T} = \sum_{n=1}^{N} \left(\sum_{t=1}^{T} \mathbf{z}_{n,t}^2\right)^{\frac{1}{2}} \tag{34}$$

with $\mathbf{z}_{n,t} = [\mathbf{W}_t(\mathbf{F}_t\mathbf{x} - \mathbf{x})]_n$.

The associated proximal operator is then defined by:

$$\text{prox}_{\gamma\|\cdot\|_{1,T}}(\mathbf{z}_n) = \begin{cases} 1 - \frac{\gamma \mathbf{z}_n}{\sqrt{\sum_{t=1}^{T} [\mathbf{z}_t]_n^2}}, & \sqrt{\sum_{t=1}^{T} [\mathbf{z}_t]_n^2} \geq \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

i) Local patch dictionaries: In [33] an adaptive patch dictionary learning was proposed to denoised images based on the principal component analysis (PCA) of patches lying in a window. Patch dictionaries were also exploited in an off–line fashion with great success in e.g., [34] as well as online [35]. We propose here to adapt the idea of online sparse local patch dictionary learning in the context of inverse problem regularization by combining a patch extraction operator $\mathbf{T}_{\mathcal{P}}$ and the Schatten norm. Indeed with some approximation the Schatten norm $\mathcal{S}_p$ with $p = 0$ of the local patch dictionary is equivalent to the PCA of the dictionary. The operator $\mathbf{T}_{\mathcal{P}}$ maps all the $N_p \times N_p$ patches in neighborhoods of dimension $N_w \times N_w$ into a matrices of dimension $N_p^2 \times N_w^2$. The adjoint

of this operator is the projection of the patches onto the image. Note that the operator $\mathbf{T}_{\mathcal{P}}$ does not depends on the content of $\mathbf{x}$ but is only defined by the windows and patches. As an illustration let us 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. 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.

VI. Numerical experiments

In this section we will test the described methods on synthetic and real data set. For synthetic data sets, we propose to use the peak signal to noise ratio (PNSR) defined by:

$$\text{PSNR}(\mathbf{x}, \bar{\mathbf{x}}) = 20 \log_{10} \frac{\sqrt{N}\|\bar{\mathbf{x}}\|_\infty}{\|\mathbf{x} - \bar{\mathbf{x}}\|_2} \tag{36}$$

A. Generation of synthetic data

In order to generate a synthetic dataset, we use a 512×512 "barbara" and "cameraman" standard test image and a synthetic 512×512 "tubule" image generated by simulating several smoothed random walks originating from the center of the image. For each of these two test images the **SAM** operator is applied with a down–sampling of factor 2 for $\mathbf{S}_0$ and a cut–off frequency of $\varrho_0 = 1.53$ pixel^{-1} for $\mathbf{A}_0$ as defined in Eq. (4). The modulations are composed of 3 equi–spaced phases and 3 equi–spaced angles with a frequency of 0.9 pixel^{-1}. We simulate a Poisson noise on the resulting 9 images whose original dynamic range is in the interval $[0, 128]$. The obtained data set is displayed in Fig. 5.

In order to evaluate the gain of the SIM modality versus the standard wide–field imaging, we can compute the image that would have been observed by a standard "wide-field" microscope by taking the average of these images. We can also applied the regularized (Wiener) least–squares approach as proposed in the literature [3]. We can see on Fig. 6 the extended resolution provided by the SIM imaging which also translates into a gain of PSNR in this simulation.

B. Evaluation of data fitting term and regularization term

In order to evaluate the different pairwise combinations of the 3 data fitting and 6 regularization terms, we propose to reconstruct the three test images by minimizing the 18 resulting functionals using the described minimization algorithm (see

Fig. 5. Set of nine 256×256 simulated SIM images from the standard "barbara" test image. We can observe that the modulations pattern produces a Moiré effect revealing the underlying high frequencies.

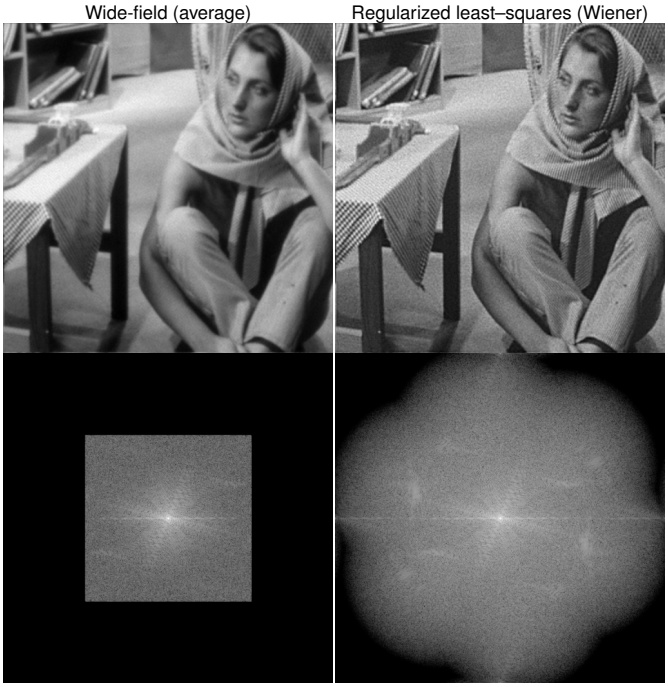

Fig. 6. In the first column, a 512×512 wide–field image is obtained from the images displayed in Fig. 5 by computing an average of the images and up–sampling the result by Fourier padding. In the second column, a regularized (Wiener) least–squares estimate is displayed showing an improved resolution. The PSNR of the wide–field image is 23.13dB while the PSNR of the SIM reconstruction is 23.94dB. Note that the regularization parameter has been selected to maximize the PSNR.

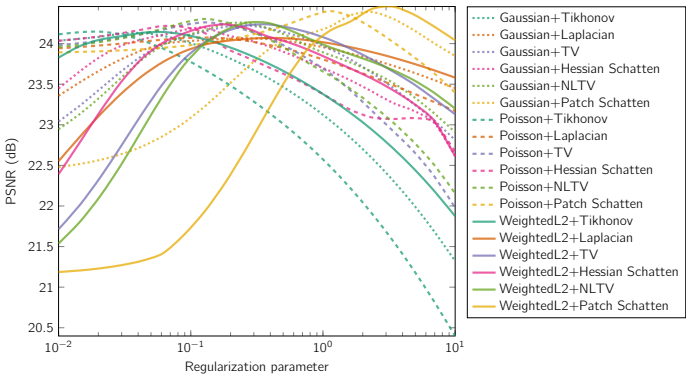

Fig. 7. Evolution of the PNSR with respect to the regularization parameter for the "tubules" synthetic test image.

Fig. 4). We use a maximum of 500 iterations and stop the algorithm if the relative norm of two consecutive estimates is below 10^{-5}. All the implementation has been done using the Matlab programming language. Results are summarized from Fig. 8 and 9 as well as on Table I, these one were obtained by maximizing the PSNR over a range of 20 values in the interval $[0.01, 10]$. In Fig. 7, the PSNR values obtained for each regularization parameter are displayed for the "tubules" test image. We can observe that for each data term, the PSNR tends toward the same value when the regularization parameter tends to 0, as expected.

We have tested 3 data terms against the Poisson noise model. The PSNR obtained using a Kullback–Leibler divergence was performing better in 14 cases out of 18. The "cameraman" image displays a large constrast between the coat of the character and the sky and grass leading to large variation of noise level was particularly penalizing more the two other approximations and the Poisson noise model gave better results for all regularization terms.

On the other hand, the local patch dictionnary outperforms all other reguarization approaches in term of PSNR as the cost of higher computation times. We can also observe that the NLTV improves over the TV in 7 cases over 9. However, it is possible that better tuning of the η parameter could probably "rescue" these two particular cases. We can also compare the two methods based on second order derivatives: the ℓ_2–norm of the Laplacian operator and the Schatten norm of the Hessian operator. The later performed better for the standard test images but not for the "tubules" synthetic image. The full Hessian matrix allowing probably to better capture the higher complexity of the natural images. Note that in [26] the minimization of the Hessian Schatten norm is shown to outperform other very competitive regularization terms such as the ℓ_1–norm of wavelet coefficients.

C. Reconstruction of acquired data

We have tested the proposed approached 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.

TABLE I
PERFORMANCE RESULTS IN TERM OF PSNR FOR 18 FUNCTIONNALS TESTED ON 3 TEST IMAGES.

	Tikhonov	Laplacian	TV	Hessian	NLTV	Patch
Barbara						
Gaussian	24.13	24.07	24.22	24.20	24.26	24.39
Poisson	24.15	24.07	24.26	24.22	24.30	24.40
Weighted	24.14	24.07	24.24	24.24	24.26	24.46
Cameraman						
	Tikhonov	Laplacian	TV	Hessian	NLTV	Patch
Gaussian	31.91	32.12	33.56	33.86	33.67	33.97
Poisson	32.12	32.35	33.84	33.99	34.06	34.27
Weighted	32.07	32.26	33.63	33.88	34.48	34.19
Tubules						
	Tikhonov	Laplacian	TV	Hessian	NLTV	Patch
Gaussian	34.03	34.22	33.77	34.18	33.86	34.35
Poisson	34.29	34.37	33.72	34.34	33.59	34.58
Weighted	34.14	34.22	33.51	34.17	33.26	34.45

The N–SIM is equipped with a $100\times$ (1.49 N.A.) objective and a $2.5\times$ 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 dictionnary regularization are shown in Fig. 10 along with the "wide–field" image obtained by averageing the nine acquired images. On this image we can notice that the filament 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 equiped with a $100\times$ (1.4 N.A.) objective coupled with a $2\times$ lense on the camera port. This time a Evolve 512 from Photometrics was used and the final pixel–size in the image is $\sim$ 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 to achieve a much better compromise in this respect and clearly outperform the linear approach.

VII. CONCLUSION

In this paper, we have proposed an analysis of linear reconstruction methods for structured illumination microscopy and proposed a novel approach based on the minimization of nonsmooth convex functionals. We have described the implementation details and tested the resulting functionals on synthetic and real data sets. The results show that the proposed approach especially when minimizing the Schatten norm of a patch dictionnary lead to a significative improvement in term of PSNR. Able to better handle the noise perturbation allows to improve the resolution and the sensitivity of SIM images. Finally, we believe that the proposed approach could be useful

Fig. 8. Comparison of 18 functionals on the "barbara" standard test images. For each case, the regularization parameter (in a $\log_{10}$ basis), the number of iterations and the computation times and the PNSR are indicated. The displayed image and values correspond to the maximum PSNR on a range of regularization parameters with in the interval $[\log_{10}(-2), \log_{10}(1)]$.

Fig. 9. Figure 8 continued.

to solve inverse problems where the forward operator cannot be easily inverted.

VIII. Aknowledgments

This work was funded by the CNRS PEPS grant "PROMIS".

References

[1] L. Schermelleh, R. Heintzmann, and H. Leonhardt, "A guide to super-resolution fluorescence microscopy," *The Journal of Cell Biology*, vol. 190, no. 2, pp. 165–175, Jul. 2010.

[2] R. Heintzmann and C. G. Cremer, "Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating," in *SPIE Optical Biopsies and Microscopic Techniques III*, vol. 3568, 1999, pp. 185–196.

[3] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *Journal of microscopy*, vol. 198, no. 2, pp. 82–87, May 2000.

[4] L. H. Schaefer, D. Schuster, and J. Schaffer, "Structured illumination microscopy: artefact analysis and reduction utilizing a parameter optimization approach," *Journal of microscopy*, vol. 216, no. 2, pp. 165–174, Nov. 2004.

[5] K. O'Holleran and M. Shaw, "Optimized approaches for optical sectioning and resolution enhancement in 2d structured illumination microscopy," *Biomedical Optics Express*, vol. 5, no. 8, p. 2580, 2014.

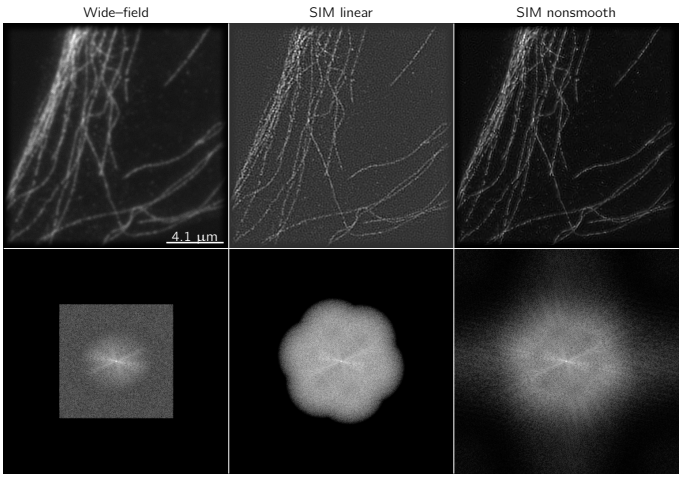

Fig. 10. Reconstruction of acquired fluorescently labelled tubuline cell with the NSIM system. The structured illumination microscopy allow to reveal the crossing of fibers with more details than the wide-field image. The proposed approach allows to handle the noise and reduce the artifacts observed in the linear reconstruction. On the second row, the power spectrum is displayed as reveal the increased support in the frequency domain. The blue circle correspond to the resolution 110 µm.

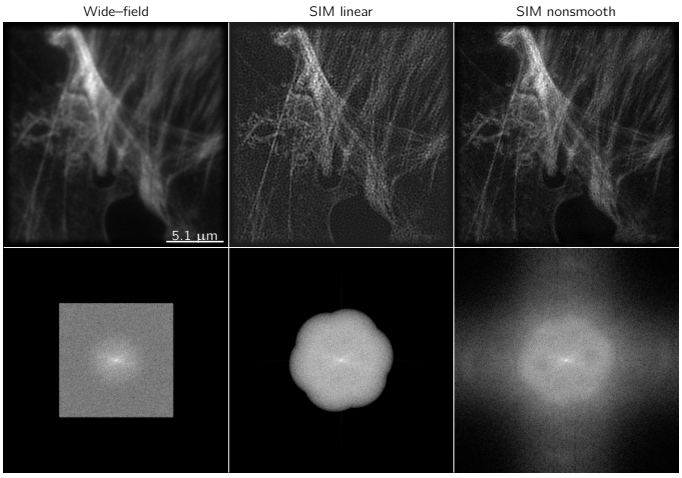

Fig. 11. Reconstruction of acquired F-actine fluorescently labelled cell with the OMX setup. The fine and dense network structure of the actin cytosqueleton is better resolved when using the proposed approach.

[6] F. Orieux, E. Sepulveda, V. Loriette, B. Dubertret, and J.-C. Olivo-Marin, "Bayesian Estimation for Optimized Structured Illumination Microscopy," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 601–614, Feb. 2012.

[7] E. Mudry, K. Belkebir, J. Girard, J. Savatier, E. L. Moal, C. Nicoletti, M. Allain, and A. Sentenac, "Structured illumination microscopy using unknown speckle patterns," *Nature Photonics*, vol. 6, no. 5, pp. 312–315, 2012.

[8] J. Min, J. Jang, D. Keum, S.-W. Ryu, C. Choi, K.-H. Jeong, and J. C. Ye, "Fluorescent microscopy beyond diffraction limits using speckle illumination and joint support recovery," *Scientific Reports*, vol. 3, 2013.

[9] K. Chu, P. J. McMillan, Z. J. Smith, J. Yin, J. Atkins, P. Goodwin, S. Wachsmann-Hogiu, and S. Lane, "Image reconstruction for structured-illumination microscopy with low signal level," *Optics Express*, vol. 22, no. 7, p. 8687, Apr. 2014.

[10] J. Goodman, W, *Introduction To Fourier Optics*, ser. McGraw-Hill Physical and Quantum Electronics Series. New York: McGraw-Hill, 1968.

[11] J.-L. Starck, F. D. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, May 1998.

[12] F. J. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, Dec. 1948.

[13] M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, "Three-Dimensional Resolution Doubling in Wide-Field Fluorescence Microscopy by Structured Illumination," *Biophysical Journal*, vol. 94, no. 12, pp. 4957–4970, Jun. 2008.

[14] A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, Dec. 2010.

[15] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, Nov. 2011.

[16] L. Condat, "A Primal–Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, Aug. 2013.

[17] N. Komodakis and J.-C. Pesquet, "Playing with Duality: An Overview of Recent Primal-Dual Approaches for Solving Large-Scale Optimization Problems," *arXiv:1406.5429 [cs, math]*, Jun. 2014, arXiv: 1406.5429.

[18] J.-C. Pesquet and N. Pustelnik, "A Parallel Inertial Proximal Optimization Method," *Pacific Journal of Optimization*, vol. 8, no. 2, pp. 273–305, 2012.

[19] M. Burger, A. Sawatzky, and G. Steidl, "First order algorithms in variational image processing," *arXiv:1412.4237 [cs, math, stat]*, Dec. 2014, arXiv: 1412.4237.

[20] P. L. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York, NY: Springer New York, 2011, vol. 49, pp. 185–212.

[21] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, "Richardson-Lucy algorithm with total variation regularization for 3d confocal microscope deconvolution," *Microscopy Research and Technique*, vol. 69, no. 4, pp. 260–266, Apr. 2006.

[22] P. Combettes and J. Pesquet, "A Douglas-Rachford Splitting Approach to Nonsmooth Convex Variational Signal Recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, Dec. 2007.

[23] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 442–454, Feb. 2010.

[24] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

[25] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York: Springer-Verlag, 2010.

[26] S. Lefkimmiatis, J. P. Ward, and M. Unser, "A Hessian Schatten-Norm Regularization Approach For Solving Linear Inverse Problems," *CoRR*, vol. 22, no. 5, pp. 1873–1888, 2012.

[27] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, "A Nonlocal Structure Tensor-Based Approach for Multicomponent Image Recovery Problems," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5531–5544, Dec. 2014.

[28] G. Gilboa, J. Darbon, S. Osher, and T. F. Chan, "Nonlocal convex functionals for image regularization," UCLA CAM, Tech. Rep. Report 06-57, Oct. 2006.

[29] G. Peyré, S. Bougleux, and L. Cohen, "Non-local Regularization of Inverse Problems," in *Computer Vision – ECCV 2008*, ser. Lecture Notes in Computer Science, D. Forsyth, P. Torr, and A. Zisserman, Eds. Springer Berlin Heidelberg, 2008, no. 5304, pp. 57–68.

[30] G. Peyré, S. Bougleux, and L. D. Cohen, "Non-local Regularization of Inverse Problems," *Inverse Problems and Imaging*, vol. 5, no. 2, pp. 511–530, 2011.

[31] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, "Epigraphical splitting for solving constrained convex formulations of inverse problems with proximal tools," *Signal, Image and Video Processing*, 2015.

[32] L. Condat, "A simple trick to speed up and improve the non-local means," Caen, France, Tech. Rep. hal-00512801, 2010.

[33] D. Muresan and T. Parks, "Adaptive principal components and image denoising," in *2003 International Conference on Image Processing, 2003. ICIP 2003. Proceedings*, vol. 1, Sep. 2003, pp. 101–104.

[34] J. Mairal, M. Elad, and G. Sapiro, "Sparse Representation for Color Image Restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, Jan. 2008.

[35] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.