Nonsmooth Convex Optimization for Structured Illumination Microscopy Image Reconstruction

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

▶ **To cite this version:**

Jérôme Boulanger, Nelly Pustelnik, Laurent Condat, Lucie Sengmanivong, Tristan Piolot. Nonsmooth Convex Optimization for Structured Illumination Microscopy Image Reconstruction. 2015. hal-01274259v2

HAL Id: hal-01274259

https://hal.science/hal-01274259v2

Preprint submitted on 19 Jun 2017 (v2), last revised 17 Sep 2019 (v4)

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

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

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 review its properties in various conditions. We then propose the minimization of nonsmooth convex functionals for the recovery of the unknown image and investigate several data–fitting and regularization terms in order to tackle reconstruction of noisy data. More specifically, we consider an original approach based on sparse local patch dictionaries for the regularization of the estimate. We demonstrate the good performance of the proposed approach on a test benchmark and perform some test experiments on images acquired on two different microscopes.

1 Introduction

Super–resolution approaches allows to go beyond the resolution of standard wide-field fluorescence microscopy therefore breaking the classical diffraction limit define by Abbe in 1873 [1]. They offer new opportunities for investigating molecular mechnisms in cell biology and have already provided new insight of the living cell. While not reaching the spatial resolution of electronic microscopy, their interest lies in their ability to provide functional information by the use of fluorescent tags and, importantly, follow them over time.

Structured illumination microscopy (SIM) is one of the recently proposed optical super–resolution methods compatible with time lapse imaging of several labels. Based on the illumination of a sample by a set of interference patterns it allows 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 by Fig. 2, where an acquired image has been reconstructed using a linear method [3]. While 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]. 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.

An elegant Bayesian approach was proposed for the reconstruction of SIM images [6]. This approach considers SIM reconstruction as an inverse problem and allows incorporating the point spread function (PSF) of the optical system in order to deconvolve and restore the image in a single step using the ℓ_2–norm of the Laplacian operator as a regularizer. If this prior leads to rather smooth estimates, a noticeable advantage of this formulation is that the number of phases is not constrained anymore and even more complex patterns can be considered. This idea was actually investigated in [7–9], where the sample is illuminated using the speckle produced by a coherent light source diffracted by a rough surface, at the cost of increasing the number of required images. Displacing this surface allows

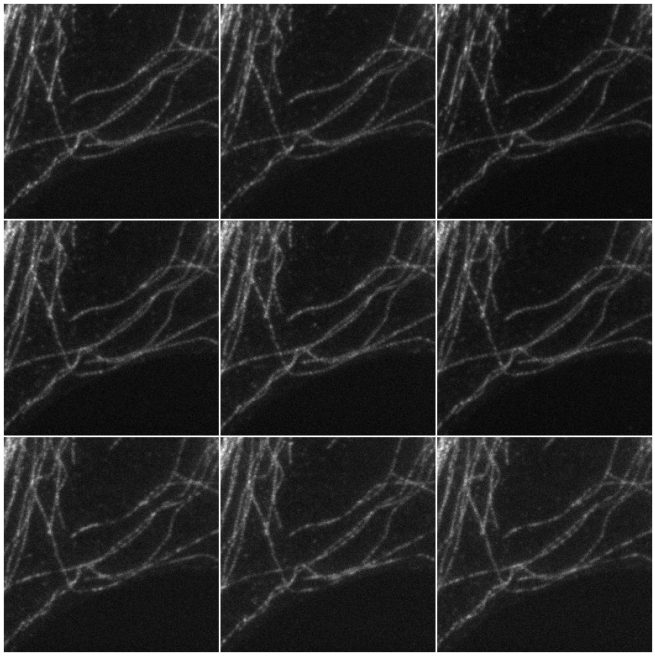

Figure 1: Example of real data. A Molecular Probe slide was imaged 9 times using a Nikon SIM microscope using a $100\times$ oil objective. The images represent a 256×256 region of 512×512 acquired images and display some labeled microtubules. The modulation pattern can be observed as a slight Moiré effect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

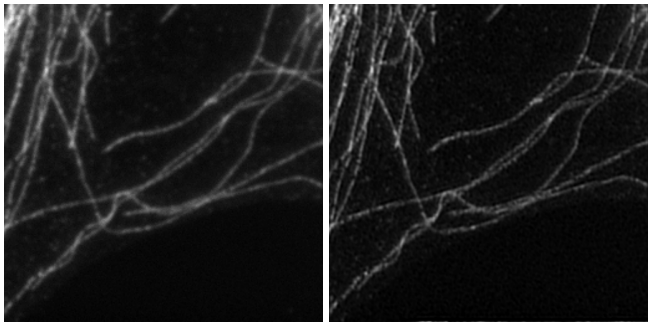

Figure 2: Reconstruction of the data displayed in Fig. 1. On the left the corresponding classical wide–field microscopy is obtained from the mean of the nine images. On the right, a linear least–squares reconstruction. The actual dimension of the image on the right is twice the size of the image of the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

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.

Formulating the 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. Nonsmooth cost functions (e.g. [10]) have been shown to display better performance in signal recovery than smooth penalty terms and this is specially true for super resolution as linear methods are known to compromise the resolution. We can note that total variation was shown to provide interesting results in the context of SIM image reconstruction in [11]. Recently two major directions have been explored for regularizing inverse problems. On one hand, higher derivatives such as Hessian matrix associated with the Schatten norm [12] have demonstrated very good performance in term of recovery and have been applied to Poisson image debluring [13]. Following a similar idea, combined first and second order approaches have been proposed for example in [14]. On the other hand, patch-based approaches have shown that a data driven and low level features such as blocks of pixels can be directly used to enforce a regularity of the solution. From the examplar-based methods for super-resolution [15] to the successful nonlocal means algorithm [16] for image denoising, patch–based approaches, though relatively simple, have shown to be very competitive. Their extension to a larger class of inverse problems has also been investigated in a variational framework in [17] and as

an extension of the total variation in [18–20]. The drawback of these approaches is the need of an oracle reference image to pre-compute a set of nonlocal weights. Besides, another class of patch-based approach rather extend the concept of image decomposition on a redundant frame basis learned from either a database of images or the image itself. In [21] a two step approach involving an orthogonal matching pursuit for dictionary learning is applied to image debluring. This approach has been extended to Poisson image deconvolution in the context of medical imaging in [22] in combination with total variation regularization with very good results. In these approaches, the joint estimation of the coefficients of the atoms and of the dictionary is not convex since the product of the two terms is involved in the minimization. Finally, adaptive patch–based approach extracting a collection of patches in every neighborhood of the image such as BM3D [23] have been also applied to inverse problem in e.g. [24]. Again the local grouping of patches is non–linear and the overall problem is then non–convex.

Most of these techniques have been developed for inverse problems, where the inversion of the forward operator can be done in Fourier space. In the case of structured illumination microscopy, such an approach is only possible in a few cases (see Section 3). In this respect, we cannot follow directly the approach developed for deconvolution problems especially when considering a Poisson noise model.

In this paper, after recalling the principle of SIM image reconstruction, we analyze the properties of the reconstruction process in a unified inverse problem framework and we show how standard SIM reconstruction methods can be formulated as suboptimal, particular instances of this setting. In order to better handle noisy data, we then consider convex non–linear variational approaches for SIM. Within a single primal–dual optimization framework able to deal with nonsmooth terms, we study the combination of different data fitting and regularization functionals. We also introduce a new patch–based approach involving the Schatten norm of a patch extraction operator leading to a convex nonsmooth patch-based approach for image restoration. We can note that this work is an extension of the contribution described in a conference proceeding [25].

The text of this article is organized as follows. We first propose in Section 3 a comprehensive analysis of the properties of linear least–squares reconstruction of SIM images in the case of sinusoidal modulations. This analysis also aims at being an introduction to structured illumination microscopy, for the reader unfamiliar with this modality. We also introduce a regularized least–squares approach in Section 4, which includes the approaches used for instance in commercial microscopes. We further consider in Section 5 nonlinear reconstruction methods taking into account the specificity of Poisson noise present in video–microscopy, while exploring different regularization approaches allowing to not compromise the image sharpness. Finally, in Section 6 we test the described 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. We finally define the proximity operator $\mathrm{prox}_f : \mathbb{R}^n \to \mathbb{R}^n$ of any closed proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ $\forall \mathbf{x} \in \mathbb{R}^n$ as :

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

2 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:

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

Table 1: Notations

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 of the algorithm	$\mathbf{S}_0$	dowsampling (matrix)
N	number of pixels of $\mathbf{x}$	$\mathbf{y}$	mesurements (stacked vector)
L	number of pixel of $\mathbf{y}$	$\mathbf{M}$	stacked modulations
K	number of modulations	$\mathbf{A}$	stacked point spread function
Q	number of cost term	$\mathbf{S}$	dowsampling (stacked matrix)
R	number of iterations	$\mathbf{W}$	diagonal weigth matrix
T	number of translation (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 phases matrix
μ_{DC}	dark current noise mean offset	$\mathbf{\Omega}$	stacked frequencies 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	$\mathbf{D}^2_{11}$	seconde order finite difference operator
τ, σ, ρ	algorithm parameters	$\mathbf{T}_q$	generic operator in the cost term
$\bar{\mathbf{x}}$	"true" value for $\mathbf{x}$	$\mathbf{T}_{\mathcal{D}}$	Stacked gradient operator
$\mathbf{x}$	unknown image (vector)	$\mathbf{T}_{\mathcal{L}}$	Laplacian operator
$\mathbf{y}_k$	mesurements (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

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 $\mathbf{M} = \left[\mathbf{M}_1, \cdots, \mathbf{M}_K\right]^{\mathrm{T}}$

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 wide–field microscopy by the auto–correlation of the pupil function as [6, 26]:

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

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. (1) as:

$$\bar{\mathbf{y}} = \mathbf{SAM}\bar{\mathbf{x}}. \tag{5}$$

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

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

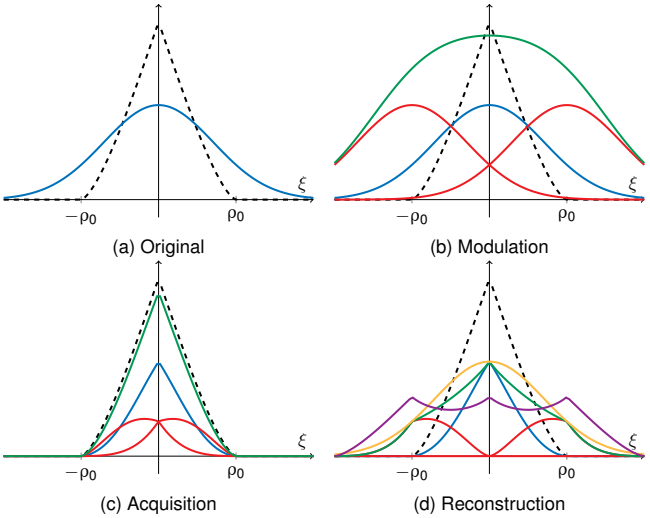

Figure 3: Principle of structured illumination microscopy illustrated in one dimension. (a) Spectrum of x in blue and the optical transfer function $\mathcal{A}_0$ in black (b) Spectrum of a modulated image $\mathbf{x}_n \cdot (1 + \cos(\omega n + \varphi))$ (green) as the sum of the three components 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 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 (purple) allows to recover the original image (yellow).

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 [27]:

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

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} characterizing the dark current and readout noise level. The offset term m_{DC} accounts for the baseline gray level that are characteristic of most CCD camera, while the variance of the additive Gaussian white noise summarizes several intensity independent noise 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 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 large 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 [28] would introduce nonlinearities, which would have a significant impact on the observation model (1) and make the reconstruction process intractable.

3 Least–squares Solution

3.1 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 substracted 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:

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

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

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

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

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.

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.

3.2 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,

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

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 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.

3.3 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:

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

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:

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

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:

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

3.4 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:

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

which leads to the least–squares solution:

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

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. [3,4,29] 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.

4 Regularized Least–Squares Solution

In the previous section, we assumed that the operator $\mathbf{SAM}$ was injective, but we know from Section 2 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:

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

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 $\mathbf{x}$ will ensure that the solution belong to the Sobolev space $\mathcal{H}^1$ 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 that is when $\mathbf{M}$ as in (13), we can obtain a closed form for the estimation of the SIM image:

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

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 approaches involving nonsmooth and non–necessarily finite criterion.

5 Nonsmooth Regularization

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

$$\hat{\mathbf{x}} \in \underset{\mathbf{x}\in C}{\operatorname{Argmin}} \sum_{q=1}^{Q} f_q(\mathbf{T}_q\mathbf{x}) \tag{24}$$

where f_q are convex, closed and proper functions from $\mathbb{R}^{M_q} \to \mathbb{R} \cup \{+\infty\}$, C is a nonempty closed convex subset of $\mathbb{R}^N$ 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 a regularization term. For instance, the least-square criterion (22) is a particular case of Eq. (24) with $Q = 2$, $f_1 = f_2 = \|\cdot\|^2$, $\mathbf{T}_1 = \mathbf{SAM}$ and $\mathbf{T}_2 = \mathbf{L}$. A usual choice for C is a constraint insuring non negativity.

When the involved functions are non-necessarily smooth, two main classes of algorithms can be derived to solve (24) and have been largely employed for solving inverse problems during the last decade: the alternating directions of multipliers method (ADMM) [30] or primal-dual proximal algorithms [31–35]. Both strategies have in common to split the processing of the $(\mathbf{f}_q)_{1\leq q\leq Q}$ and the $(\mathbf{T}_q)_{1\leq q\leq Q}$ and to rely on the computation of the proximity operator [36] of each f_q. The definition of proximity operator is provided in (1) and it is important to notice that a large number of closed form expression is known in the literature [37]. Some of them, useful for the study, will be provided below. The major difference between both strategies comes from the processing of the operators $(\mathbf{T}_q)_{1\leq q\leq Q}$. ADMM requires to compute $(\sum_{q=1}^{Q}\mathbf{T}_q^{*}\mathbf{T}_q)^{-1}$ while primal-dual strategies avoid such a step. Note that since in general the operator associated to SIM imaging is not directly invertible (cf. to Sections 3 and 4), ADMM would require an inner minimization procedure for the inversion of this operator. Consequently, we solve Eq. (24) using a primal–dual proximal algorithm from [31–35] (see Fig. 4) and consider several cases corresponding to the combination of function f_q and operator $\mathbf{T}_q$.

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 proximity operator [36]. 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 C will denote the non–negativity constraint and it is enforced directly on the iterate at step 8 of Fig. 4. Therefore, we can write $\hat{\mathbf{x}} \in \operatorname{Argmin}_{\mathbf{x}\geq\mathbf{0}} f_1(\mathbf{T}_1\mathbf{x}) + \lambda f_2(\mathbf{T}_2\mathbf{x})$ with λ the regularization parameter.

5.1 Possible choices for the data term $f_1 \circ \mathbf{T}_1$

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 [36] associated to f_{LS} is then [37] : $\forall\gamma > 0, \forall\mathbf{z} \in \mathbb{R}^{LK}, \operatorname{prox}_{\gamma f_{\mathrm{LS}}}(\mathbf{z}) = (\mathbf{z} + \gamma\ \mathbf{y})/(1+\gamma)$.

Require: $\mathbf{x}_0 \in \mathbb{R}^{N\times N}$, $\tau > 0$, $\sigma > 0$, $\rho > 0$
$\mathbf{z}_0 = 0$, $\mathbf{v}_{0,q} = \mathbf{x}_0$, $q \in 1, \dots, Q$
for $r \in 1, \dots, R$ **do**
  for $q \in 1, \dots, Q$ **do**
    $\mathbf{u}_{r,q} = \mathbf{v}_{r,q} + \sigma\, \mathbf{T}_q(2\mathbf{z}_r - \mathbf{x}_r)$
    $\mathbf{p}_{r,q} = \mathbf{u}_{r,q} - \sigma\, \mathrm{prox}_{f_q/\sigma}(\mathbf{u}_{r,q}/\sigma)$
    $\mathbf{v}_{r,q} = \rho\, \mathbf{p}_{r,q} + (1-\rho)\, \mathbf{v}_{r,q}$
  end for
  $\mathbf{z}_r = P_C\left(\mathbf{x}_r - \sum_{q=1}^{Q} \tau \mathbf{T}_q^* \mathbf{v}_{r,q}\right)$
  $\mathbf{x}_{r+1} = \rho\, \mathbf{z}_r + (1-\rho)\, \mathbf{x}_r$
end for

Figure 4: The primal dual minimization algorithm proposed in [34] allows to minimize the energy functional defined by Eq. (24) 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.

Kullback–Leibler divergence (KLD) Under a Poisson noise model assumption for the acquired data $\mathbf{y}$, the negative log–likelihood is given by the Kullback–Leilbler (KL) divergence [38]. The resulting function is the KL divergence defined as, for every $\mathbf{z} = (\mathbf{z}_n)_{1\leq n\leq LK}$, $f_{\mathrm{KL}}(\mathbf{z}) = \sum_{n=1}^{LK} f_{\mathrm{KL}}^{(n)}(\mathbf{z}_n)$ where the component–wise function is defined by [39] $\forall \mathbf{z}_n \in \mathbb{R}$:

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

In this configuration, the linear operator is $\mathbf{T}_{\mathrm{KL}} = \mathbf{SAM}$. The proximity operator is given component-wise for $n \in [1, LK]$ by $\forall \mathbf{z}_n \in \mathbb{R}$:

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

and the proximity operator $\mathrm{prox}_{\gamma f_{\mathrm{KL}}}(\mathbf{z}) = \left(\mathrm{prox}_{\gamma f_{\mathrm{KL}}}^{(n)}(\mathbf{z}_n)\right)_{1\leq n\leq LK}$ is obtained by applying Eq. (26) for each component of for the vector $\mathbf{z}$.

Often collected data show an offset and a gain such as described by Eq. (6). In the case where the variance of the Gaussian noise $\mathbf{n}$ is $\sigma_{DC} = 0$, we can do a variable change in order to take into account the sensor gain κ and offset m_{DC} such that the proximity operator becomes $\forall \mathbf{z}_n \in \mathbb{R}$:

$$\mathrm{prox}_{\gamma D_{\mathrm{KL}}}(\mathbf{z}_n) = \frac{1}{2}\left(\mathbf{z}_n - m_{DC} - \kappa\gamma + \frac{1}{\kappa}\sqrt{(\mathbf{z}_n - m_{DC} - \kappa\gamma)^2 + 4\gamma\kappa(\mathbf{y}_n - m_{DC})}\right) \tag{27}$$

Weighted least–squares (WLS) 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

$$(\forall \mathbf{z} \in \mathbb{R}^{LK}) \quad f_{\mathrm{WLS}}(\mathbf{z}) = \frac{1}{2}(\mathbf{z} - \mathbf{y})^{\mathrm{T}} \mathbf{W}^{-1} (\mathbf{z} - \mathbf{y}), \tag{28}$$

where $\mathbf{W}$ is a diagonal variance matrix with elements $(\mathbf{w}_n)_{1\leq n\leq LK}$. 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 [40] for more details):

$$\mathbb{V}\mathrm{ar}[\mathbf{y}_n] = \kappa \mathbb{E}[\mathbf{y}_n] + \sigma_{\mathrm{DC}}^2 - \kappa\, m_{DC} \tag{29}$$

using the model introduced in equation (6) and 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 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 $\frac{1}{\sqrt{20}}(\mathbf{D}_{11}^2\mathbf{y} + \mathbf{D}_{22}^2\mathbf{y})$) 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 $e_{DC} = \sigma^2_{\mathrm{DC}} - \kappa\, m_{DC}$. The variance-covariance matrix $\mathbf{Q}$ can then be approximated by a diagonal matrix whose elements are given by:

$$\mathbf{w}_n = \kappa \bar{\mathbf{y}}_n + e_{DC} \tag{30}$$

with $\bar{\mathbf{y}}_i \approx \mathbb{E}[\mathbf{y}_i]$ is the given by the estimation of the intensity by the median filter. The proximity operator associated to f_{WLS} is then

$$(\forall \gamma > 0)(\forall \mathbf{z} \in \mathbb{R}^{LK}) \quad \mathrm{prox}_{\gamma f_{\mathrm{WLS}}}(\mathbf{z}) = \Big((\mathbf{z}_n + \gamma\ \mathbf{w}_n \mathbf{y}_n)/(1 + \gamma \mathbf{w}_n) \Big)_{1 \leq n \leq LK}. \tag{31}$$

5.2 Possible choices for the regularization term $f_2 \circ \mathbf{T}_2$

Gradient squared ℓ_2–norm ($\|\nabla\|^2$) 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 4), we may still use the proposed approach. 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 $\mathbf{T}_D = [\mathbf{D}_1, \mathbf{D}_2]^{\mathrm{T}}$. The adjoint of this operator is then the divergence operator defined as $\mathbf{T}^*_D = \mathbf{D}_1^{\mathrm{T}}\mathbf{z}_1 + \mathbf{D}_2^{\mathrm{T}}\mathbf{z}_2$ 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, i.e. $f_D = \|\cdot\|^2$ whose proximity operator in this case is given by $\mathrm{prox}_{\gamma f_D}(\mathbf{z}) = \mathbf{z}/(1+\gamma)$ 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 $\mathbf{T}_{\mathcal{L}} = \mathbf{D}^2_{11} + \mathbf{D}^2_{22}$ where $\mathbf{D}^2_{11}$ and $\mathbf{D}^2_{22}$ 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 than 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 $\mathbf{x}$ [10]. Therefore, we can use this time the same operator $\mathbf{T}_D$ 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 $f_{\mathrm{TV}} = \|\cdot\|_{1,2}$ should be applied by considering the two gradient components as a vector [37]:

$$(\forall \mathbf{z} = [\mathbf{z}_1^{\top}, \mathbf{z}_2^{\top}]^{\top}) \in \mathbb{R}^{2N} \quad \|\mathbf{z}\|_{1,2} = \textstyle\sum_{n=1}^{N} \sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2}. \tag{32}$$

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

$$(\forall \mathbf{z}_n \in \mathbb{R}^2) \quad \mathrm{prox}_{\|\cdot\|_{1,2}}(\mathbf{z}_n) = \begin{cases} 1 - \dfrac{\gamma \mathbf{z}_n}{\sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2}}, & \sqrt{[\mathbf{z}_1]_n^2 + [\mathbf{z}_2]_n^2} \geq \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

Schatten norm of the Hessian operator ($S_p(\mathrm{T}_{\mathcal{H}})$) Recently a new regularization based on the Schatten norm of the Hessian operator has been proposed [12]. 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:

$$[\mathbf{T}_{\mathcal{H}}\mathbf{x}]_n = \begin{bmatrix} [\mathbf{D}^2_{11}\mathbf{x}]_n & [\mathbf{D}^2_{12}\mathbf{x}]_n \\ [\mathbf{D}^2_{12}\mathbf{x}]_n & [\mathbf{D}^2_{22}\mathbf{x}]_n \end{bmatrix} \tag{34}$$

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:

$$\mathbf{T}_{\mathcal{H}}^{*}\mathbf{z} = \mathbf{D}_{11}^{2*}\mathbf{z}_{11} + \mathbf{D}_{12}^{2*}(\mathbf{z}_{12} + \mathbf{z}_{21}) + \mathbf{D}_{22}^{2*}\mathbf{z}_{22} \tag{35}$$

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 $\mathbb{R}^N$.

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 $\mathbf{z}_n = \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{V}_n^{\mathrm{T}}$. Then, the associated proximity operator is given by [20]:

$$(\forall \mathbf{z}_n \in \mathbb{R}^{2\times 2}) \quad \mathrm{prox}_{\gamma\mathcal{S}_p}(\mathbf{z}_n) = \mathbf{U}_n \mathrm{prox}_{\gamma\|\cdot\|_p}(\mathbf{\Lambda}_n)\mathbf{V}_n^{\mathrm{T}}. \tag{36}$$

Nonlocal total variation (NLTV) The nonlocal total variation (NLTV) penalization was introduced in [18] and extended to various inverse problems in [19,41] 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 [20]. The operator associated to the NLTV regularization can be described as weighted non–local gradients defined as [42]:

$$[\mathbf{T}_{\mathrm{NL}}\mathbf{x}]_n = \begin{bmatrix} \left[\mathbf{W}_1(\mathbf{F}_1\mathbf{x} - \mathbf{x})\right]_n \\ \vdots \\ \left[\mathbf{W}_T(\mathbf{F}_T\mathbf{x} - \mathbf{x})\right]_n \end{bmatrix} \tag{37}$$

where for $t \in 1, \ldots, T$, we define some diagonal weight matrices as function 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 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 [43]. 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 $\mathbf{T}_{\mathrm{NL}}$ is defined by:

$$(\forall \mathbf{z} \in \mathbb{R}^{TN}) \quad \mathbf{T}_{\mathrm{NL}}^{*}\mathbf{z} = \sum_{t=1}^{T} \mathbf{W}_t(\mathbf{F}_t^{*} - \mathbf{I})\mathbf{z}_t \tag{38}$$

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:

$$(\forall \mathbf{z} \in \mathbb{R}^{TN}) \quad \|\mathbf{z}\|_{1,2} = \sum_{n=1}^{N}\left(\sum_{t=1}^{T}\mathbf{z}_{n,t}^2\right)^{\frac{1}{2}}. \tag{39}$$

The associated proximity operator is then defined by:

$$(\forall \mathbf{z}_n \in \mathbb{R}^{T}) \quad \mathrm{prox}_{\gamma\|\cdot\|_{1,2}}(\mathbf{z}_n) = \begin{cases} 1 - \frac{\gamma\mathbf{z}_n}{\sqrt{\sum_{t=1}^{T}[\mathbf{z}_t]_n^2}}, & \sqrt{\sum_{t=1}^{T}[\mathbf{z}_t]_n^2} \geq \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

Local patch dictionaries ($S_p(\mathrm{T}_{\mathcal{P}})$) In [44] 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 the context of image denoising in e.g., [45] as well as online [23] and more recently in [24] and [22] for the deconvolution of respectively natual and medical images. 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 patch dictionary. We consider in this paper the Schatten norm with $p = 1$ which can be seen as a relaxed version of the case $p = 0$ [46]. The associated proximity operator is then the same than the one defined for the Hessian based regularization (see above). However, we consider now an operator $\mathbf{T}_{\mathcal{P}}$ mapping 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:

$$\mathbf{T}_{\mathcal{P}}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_5 & x_6 \\ x_2 & x_3 & x_6 & x_7 \\ x_3 & x_4 & x_7 & x_8 \\ x_5 & x_6 & x_9 & x_{10} \\ x_6 & x_7 & x_{10} & x_{11} \\ x_7 & x_8 & x_{11} & x_{12} \\ x_9 & x_{10} & x_{13} & x_{14} \\ x_{10} & x_{11} & x_{14} & x_{15} \\ x_{11} & x_{12} & x_{15} & x_{16} \end{bmatrix} \tag{41}$$

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 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 exmaple, 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.

6 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:

$$\mathrm{PSNR}(\mathbf{x}, \bar{\mathbf{x}}) = 20 \log_{10} \frac{\sqrt{N}\|\bar{\mathbf{x}}\|_\infty}{\|\mathbf{x} - \bar{\mathbf{x}}\|_2} \tag{42}$$

6.1 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 "tubules" image generated by simulating several smoothed random walks originating from the center of the image. The three test images are displayed on Fig. 5. 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. 6.

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. 7 the extended resolution provided by the SIM imaging which also translates into a gain of PSNR in this simulation.

6.2 Algorithm and parameters

In our experiments, we use a single algorithm (described in Fig. 4) with the same parameters to minimized all the cost functions. It is clear that in some cases, this approach is sub-optimal. This is particularly true for linear

Figure 5: Test images used for the synthetic dataset.

Figure 6: 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.

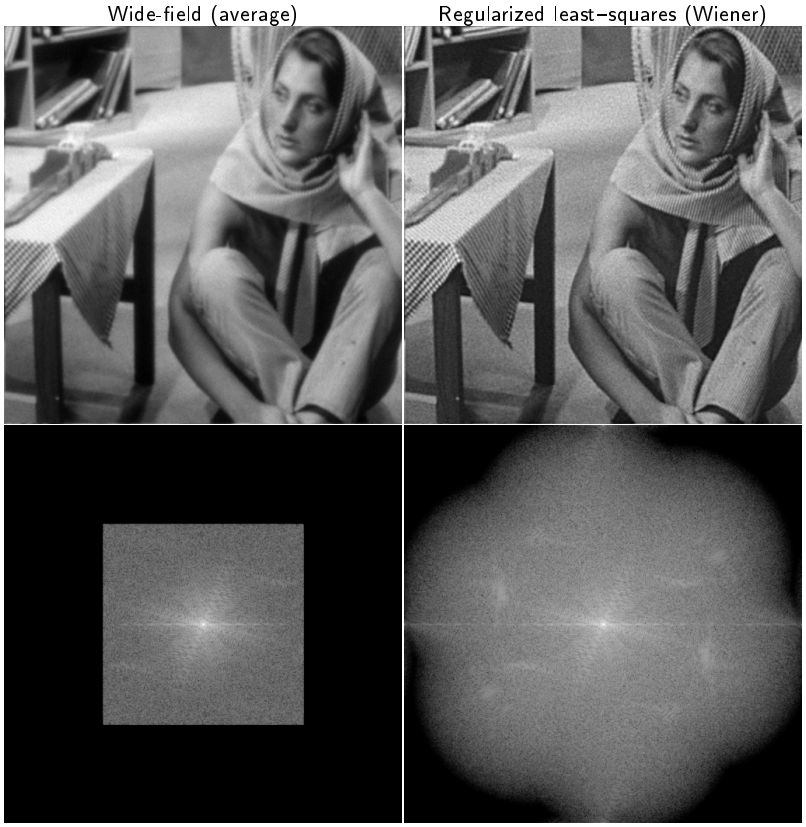

Figure 7: In the first column, a 512×512 wide–field image is obtained from the images displayed in Fig. 6 by computing an average of the images and up–sampling the result by Fourier padding. In the second column, a regularized (Wiener) least–squares estimate is displayed showing an improved resolution. 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.

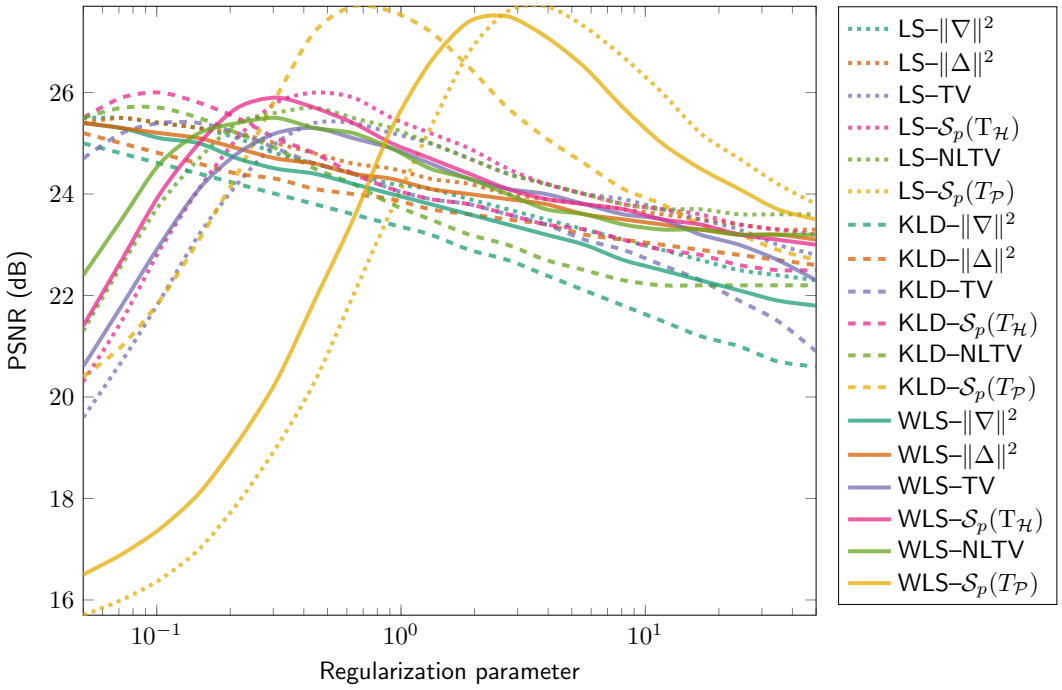

Figure 8: Evolution of the PNSR with respect to the regularization parameter for the "Barbara" test image and gain $\kappa = 1$.

cost functions. The parameters used for the algorithm are 500 iterations with no stopping criterion, $\tau = 10$ and $\rho = 0.75$. The parameter σ is defined as $\sigma = 1/(\tau L^2)$ where $L^2 = \|T_1\|^2 + \|T_2\|^2$. Those parameters were chosen to provide a reasonable compromise across all cost functions in term of convergence speed and stability.

6.3 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 ("Barbara", "Cameraman" and "Tubules") for 3 different gain value $\kappa \in [1, 2, 5]$ by minimizing the 18 resulting functionals using the described minimization algorithm (see Fig. 4). The "Tubules" image was generated by drawing smooth splines in order to emulate microtubules in a cell. The images were corrupted with noise as described in Eq. (6) with $m_{DC} = 100$ and $\sigma^2_{DC} = 2$. For each run, we use 500 iterations and we tested 20 values logarithmically spaced in the interval $[0.05, 50]$ 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.

In Fig. 8, as an example, the PSNR values obtained for each regularization parameter are displayed for the "Barbara" test image with $\kappa = 1$ test image. We can notice that those curves are relatively smooth indicating a good behavior of the optimization method for each of the functional given that each run corresponds to a different noise realization. We can already see on this particular example that the local patch dictionary regularization term $\mathcal{S}_p(T_\mathcal{P})$ shows a significant improvement.

In Fig. 9, the centered crop of the images corresponding to the best PSNR values from the curves shown in Fig. 8 are displayed along with the PSNR, the regularization parameter and the computation time. The stripe pattern have very high frequency and provide an insight on the potential gain in term of resolution provided by each of the 18 functionals. In correlation with the PSNR value, these stripes are much more better reconstructed with the *local patch dictionary* ($\mathcal{S}_p\mathrm{T}_\mathcal{P}$) regularization term while the *least square* (LS) data term shows the best PSNR out of the three tested data terms in this case. Note this is not always the case for the other regularization terms in this experiment. Finally, the computation time, which is deterministic as the number of iteration is the same for each test, is much more important for the *local patch dictionary* functional.

Table 2 summarizes the best PSNR values for 18 functionals obtained for each of the 9 test cases (3 images × 3 gain values). In each cases, the *local patch dictionary* ($\mathcal{S}_p(\mathrm{R}_\mathcal{P})$) term consistently outperforms the other regularization terms. In most cases, we can notice that when the gain increases, the KLD data term tends to perform better

Figure 9: Results for the image "Barbara" with $\kappa = 1$ and each of the 18 functionals. A center region of each estimate is displayed to better appreciate the details of the reconstruction.

Table 2: Performance results in term of PSNR for 18 functionnals tested on 3 test images.

			Barbara ($\kappa = 1$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	25.53	25.46	25.42	26.03	25.66	***27.72***
KLD	25.02	25.16	25.45	26.05	25.69	**27.68**
WLS	25.41	25.39	25.26	25.85	25.51	**27.49**
			Barbara ($\kappa = 2$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	24.67	24.60	24.59	24.97	25.00	**26.52**
KLD	24.70	24.66	24.68	25.06	25.07	***26.58***
WLS	24.58	24.52	24.44	24.79	24.84	**26.21**
			Barbara ($\kappa = 5$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	23.84	23.73	23.72	23.93	24.37	**24.90**
KLD	23.89	23.78	23.86	24.09	24.46	***25.10***
WLS	23.76	23.66	23.66	23.78	24.17	**24.61**
			Cameraman ($\kappa = 1$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	34.60	34.87	34.49	36.58	34.08	***36.59***
KLD	34.59	34.84	34.30	36.34	33.44	**36.56**
WLS	34.21	34.50	34.19	36.15	33.61	**36.03**
			Cameraman ($\kappa = 2$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	33.23	33.47	33.49	35.07	33.23	**35.27**
KLD	33.43	33.70	33.49	35.28	32.77	***35.60***
WLS	32.75	33.12	33.17	34.59	32.62	**34.75**
			Cameraman ($\kappa = 5$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	31.26	31.53	32.06	33.01	31.41	**33.49**
KLD	31.66	31.93	32.26	33.41	31.51	***33.95***
WLS	30.86	31.17	31.65	32.59	31.19	**32.79**
			Tubules ($\kappa = 1$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	30.86	30.85	30.52	31.51	31.56	***31.60***
KLD	30.55	30.65	30.32	31.28	31.16	**31.30**
WLS	30.78	30.82	30.36	31.38	31.33	**31.42**
			Tubules ($\kappa = 2$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	29.93	29.94	29.67	30.51	30.40	***30.64***
KLD	29.70	29.81	29.62	30.43	30.23	**30.53**
WLS	29.90	29.89	29.58	30.36	30.24	**30.44**
			Tubules ($\kappa = 5$)			
	$\|\nabla\|^2$	$\|\Delta\|^2$	TV	$\mathcal{S}_p(\mathrm{T}_\mathcal{H})$	NLTV	$\mathcal{S}_p(\mathrm{T}_\mathcal{P})$
LS	28.74	28.71	28.54	29.15	28.71	***29.36***
KLD	28.61	28.65	28.51	29.10	28.14	**29.31**
WLS	28.67	28.68	28.40	29.04	28.64	**29.14**

than the least square (LS) data term. The latter providing the best results for small gains $\kappa = 1$. This is particularly noticeable with the "Cameraman" standard test image which exhibits high contrasts between the subject (dark) and the scene (bright) while in the other cases, the optimizing the regularization parameters allows to adapt to the noise level which is more uniform in the image. Note that in [12] the minimization of the Schatten norm of the Hessian operator ($S_p(\mathrm{T}_\mathcal{P})$) is shown to be superior to other very competitive regularization terms such as the ℓ_1–norm of wavelet coefficients. And this experiment is comforting the idea that *local patch dictionary* regularization term is very effective at recovering structure of interest in images.

6.4 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.

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. 10 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 equiped 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 $\sim$ 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 outperforms the linear approach.

6.5 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 orientation but no phase shifts. This allow 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. 12 show that the sample is successfully recovered with PSNR of 26.75dB and that high frequency details are well estimated as shown on the powerspectrum on the second row.

7 Conclusion

In this paper, we have proposed an analysis of linear reconstruction methods for structured illumination microscopy and proposed a novel nonlinear approach based on the minimization of a nonsmooth convex functional. She have considered a very flexible framework and showed that it could accommodate the specificity of SIM image reconstruction. We have described the implementation details and extensively tested the derived functionals on synthetic and real data sets. The results show that the proposed approach leads to a significant improvement in terms of PSNR, especially when minimizing the Schatten norm of a local patch dictionary. Being able to better handle the noise perturbation allows to improve the resolution and the sensitivity of SIM images. We did not address the problem of the modulation parameter estimation, which is directly impacting the quality of the reconstructed images [47]; we leave this study for future work.

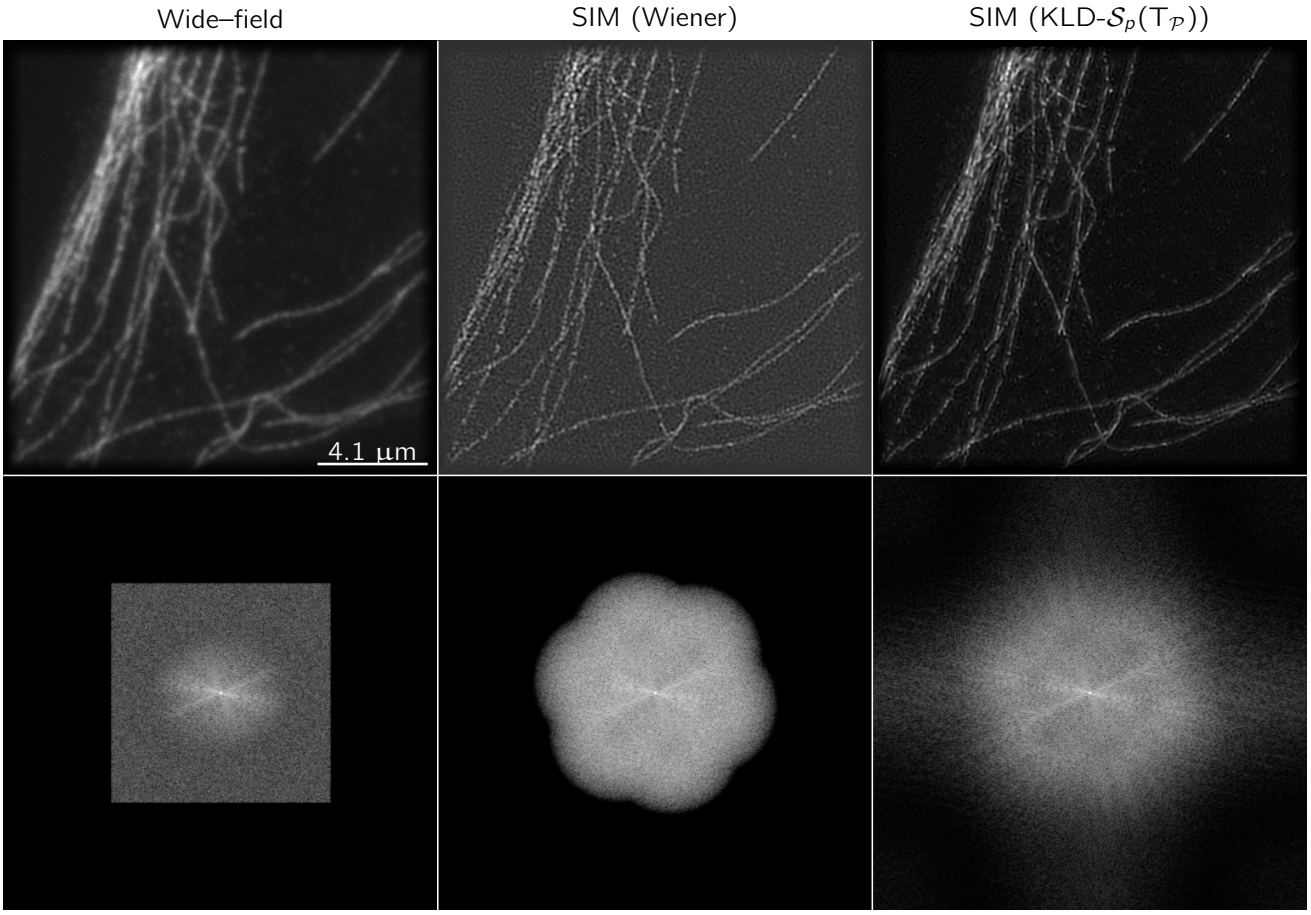

Figure 10: Reconstruction of acquired fluorescently labelled tubuline cell with the NSIM system. The structured illumination microscopy allow to reveal the crossing of fibers with more details than the wide-field image. The proposed approach allows to handle the noise and reduce the artifacts observed 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.

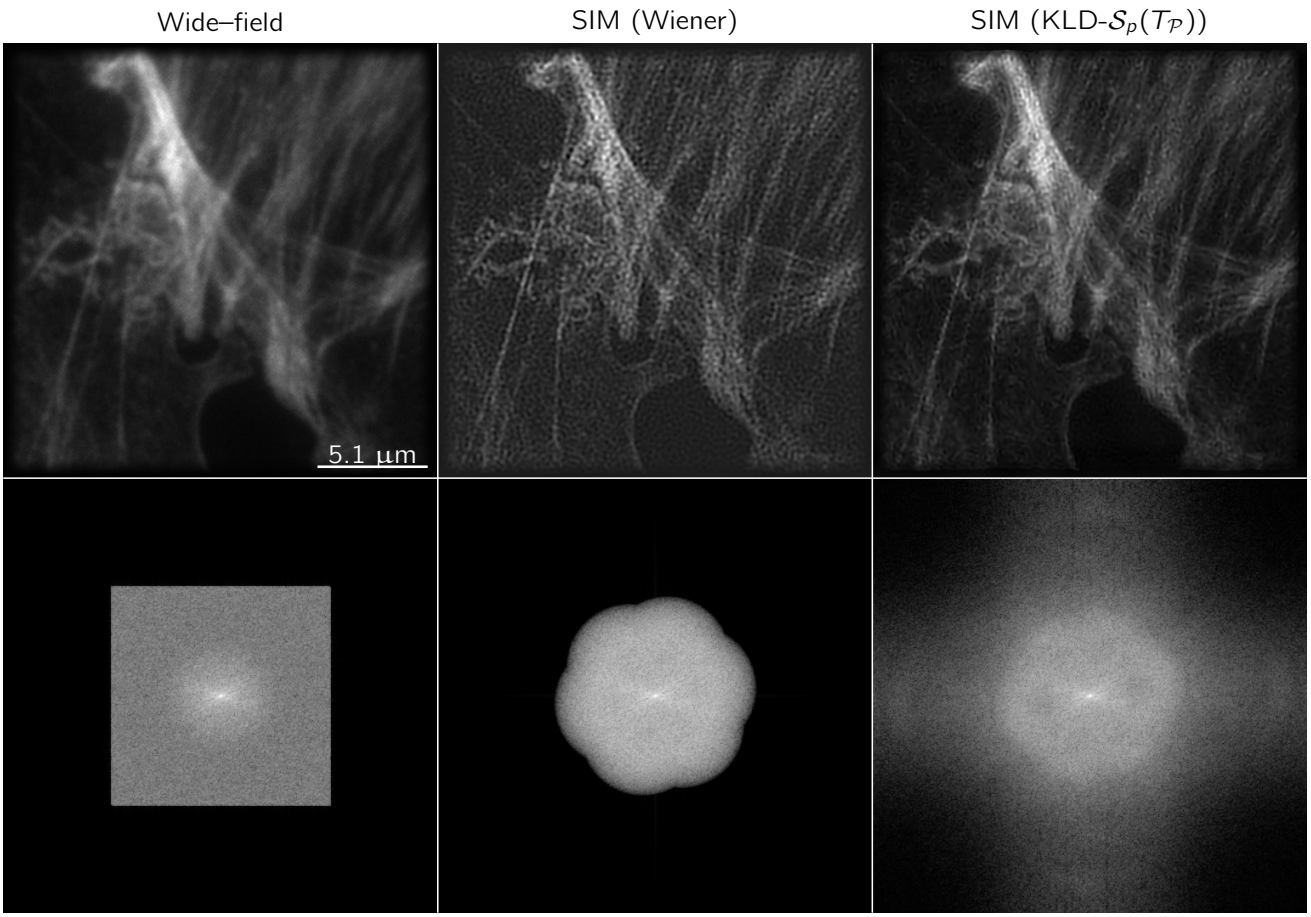

Figure 11: Reconstruction of acquired F-actine fluorescently labelled cell with the OMX setup. The fine and dense network structure of the actin cytosqueleton is better resolved when using the proposed approach.

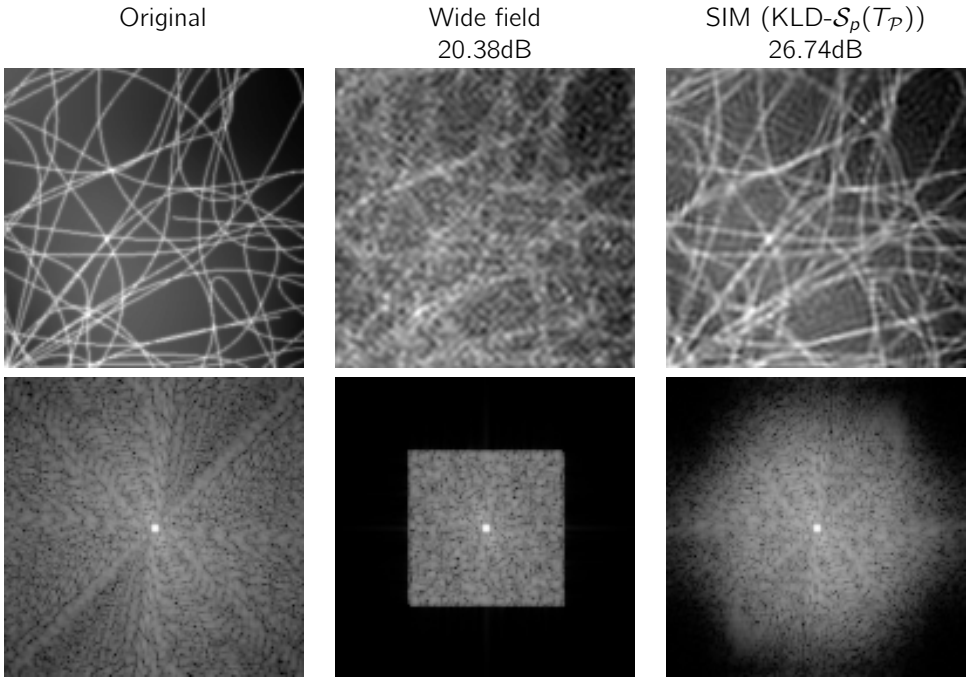

Figure 12: Reconstruction of a simulated 3 SIM images with a reduced number of images using only 3 modulation orientation and no phase shifts. The second row displays the corresponding powerspectrums.

8 Aknowledgments

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

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

[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] Lefkimmiatis S, Ward JP, Unser M. A Hessian Schatten-norm regularization approach for solving linear inverse problems. CoRR. 2012;22(5):1873–1888.

[13] Lefkimmiatis S, Unser M. 3D Poisson microscopy deconvolution with Hessian Schatten-norm regularization. In: ISBI; 2013. p. 161–164.

[14] Papafitsoros K, Schönlieb CB. A combined first and second order variational approach for image reconstruction. Journal of Mathematical Imaging and Vision. 2013 May;48(2):308–338.

[15] Freeman WT, Jones TR, Pasztor EC. Example-based super-resolution. IEEE Computer Graphics and Applications. 2002 Mar;22(2):56–65.

[16] Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005. vol. 2; 2005. p. 60–65.

[17] Kindermann S, Osher S, Jones P. Deblurring and denoising of images by nonlocal functionals. Multiscale Modeling & Simulation. 2005 Jan;4(4):1091–1115.

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

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

[20] 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.

[21] 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.

[22] 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.

[23] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing. 2007 Aug;16(8):2080–2095.

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

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

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

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

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

[29] 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.

[30] 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. DOI: 10.1007/978-3-319-41589-5_10.

[31] 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.

[32] 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.

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

[34] 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.

[35] 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.

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

[37] 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.

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

[39] 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.

[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] 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.

[42] 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;.

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

[44] Muresan DD, Parks TW. Adaptive principal components and image denoising. In: 2003 International Conference on Image Processing, 2003. ICIP 2003. Proceedings. vol. 1; 2003. p. 101–104.

[45] Mairal J, Elad M, Sapiro G. Sparse Representation for Color Image Restoration. IEEE Transactions on Image Processing. 2008 Jan;17(1):53–69.

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

[47] 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: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI); 2014. p. 604–607.