Sparse Poisson Noisy Image Deblurring

Mikael Carlavan, Laure Blanc-Féraud

Sparse Poisson Noisy Image Deblurring

Mikael Carlavan and Laure Blanc-Féraud

Abstract—**Deblurring noisy Poisson images has recently been subject of an increasingly amount of works in many areas such as astronomy or biological imaging. In this paper, we focus on confocal microscopy which is a very popular technique for 3D imaging of biological living specimens which gives images with a very good resolution (several hundreds of nanometers), even though degraded by both blur and Poisson noise. Deconvolution methods have been proposed to reduce these degradations and we focus in this paper on techniques which promote the introduction of explicit prior on the solution. One difficulty of these techniques is to set the value of the parameter which weights the trade-off between the data term and the regularizing term. Actually, only few works have been devoted to the research of an automatic selection of this regularizing parameter when considering Poisson noise so it is often set manually such that it gives the best visual results. We present here two recent methods to estimate this regularizing parameter and we first propose an improvement of these estimators which takes advantage of confocal images. Following these estimators, we secondly propose to express the problem of Poisson noisy images deconvolution as the minimization of a new constrained problem. The proposed constrained formulation is well suited to this application domain since it is directly expressed using the anti log-likelihood of the Poisson distribution and therefore does not require any approximation. We show how to solve the unconstrained and constrained problem using the recent Alternating Direction technique and we present results on synthetic and real data using well-known priors such as Total Variation and wavelet transforms. Among these wavelet transforms, we specially focus on the Dual-Tree Complex Wavelet transform and on the dictionary composed of Curvelets and undecimated wavelet transform.**

Index Terms—**3D confocal microscopy deconvolution, regularizing parameter selection, discrepancy principle, alternating direction method, constrained minimization, Poisson noise.**

I. Introduction

DEBLURRING images corrupted by Poisson noise is a challenging process to which much research has been devoted as in astronomical or biological imaging. We focus in this paper on confocal microscopy imaging, introduced by M. Minksy in 1953 [26]. This technique is based on the principle of fluorescence and allows to observe inside living cells of the specimen by tagging the core, the membranes or others elements of the cells.

Confocal microscopy imaging offers several advantages over optical (wide-field) imaging such as a small depth-of-field, a reduction of out-of-focus blur and the ability of scanning 3D images. These advantages explain its quick grow in popularity during the last years. However, it suffers from two basic degradations: remaining blur and Poisson noise. If

we consider a discrete version of a specimen $\mathbf{x} \in \mathbb{R}^n$ (n being the number of voxels of the image) observed as an image $\mathbf{y} \in \mathbb{R}^n$ through an optical system with a Point Spread Function (PSF) h and corrupted by a Poisson noise process $\mathcal{P}$, then the image formation model can be written as [40]:

$$\mathbf{y} = \mathcal{P}(\mathbf{H}\mathbf{x} + \mathbf{b}), \qquad (1)$$

where $\mathbf{H} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ stands for the matrix notation of the convolution of the PSF h (we assume moreover $\mathbf{H}\mathbf{x} \geq 0 \, \forall \mathbf{x} \geq 0$) and $\mathbf{b} \in \mathbb{R}^n$ is a constant background. A good estimation of the PSF h is very important for any non-blind deconvolution algorithm. In this paper, we will use the model presented in [16], [39].

Using a bayesian approach, we want to retrieve the image $\mathbf{x}$ which maximizes the likelihood probability of (1). This probability can be expressed as:

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=0}^{n-1} \left(\frac{\left[(\mathbf{H}\mathbf{x} + \mathbf{b})^{\mathbf{y}}\right]_i \exp\left[-(\mathbf{H}\mathbf{x} + \mathbf{b})\right]_i}{\mathbf{y}_i!} \right). \quad (2)$$

Maximizing (2) with respect to $\mathbf{x}$ is equivalent to minimize $-\log p(\mathbf{y}|\mathbf{x})$ that is to minimize:

$$J_L(\mathbf{x}, \mathbf{y}) = \mathbf{1}^T (\mathbf{H}\mathbf{x} + \mathbf{b}) - \mathbf{y}^T \log(\mathbf{H}\mathbf{x} + \mathbf{b}), \qquad (3)$$

where $\mathbf{1}$ stands for a n-size vector whose components are all equal to 1. A popular algorithm to optimize (3), with respect to $\mathbf{x}$, in confocal microscopy is the Richardson-Lucy (RL) algorithm [24], [34]. This algorithm takes into account Poisson statistics of the photon counting noise, and implicitly imposes positivity constraint on the solution. This is, however, not sufficient to prevent from noise amplification during the deconvolution process due to the ill-posedness of this inverse problem and this algorithm is usually stopped after an arbitrary number of iterations.

Many authors favour instead the introduction of an explicit prior on the solution to regularize the ill-posed inverse problem and thus minimize a penalized likelihood as in [16], [17], [31]. Unfortunately, most of the refereed methods need to manually tune the regularizing parameter to control the weight of the prior. This approach is time consuming as it needs several resolution of the minization problem to find a result which is, after all, totally subjective.

The contributions of this work are as follows. We propose two revised estimation procedures for regularizing parameter when dealing with Poisson noise and l^1-norm regularization. We also propose a new constrained formulation of the optimization problem leading to simple parameter setting and we describe the Alternating Direction technique for both minimization formulations (constrained and unconstrained).

We evaluate image restoration using these parameter estimation procedures for several regularizations, Total Variation, Dual-Tree Complex Wavelet transform, dictionary composed of Curvelets and undecimated wavelet transform, on synthetic and real data.

The paper is organized as follows. In section II we present common priors used in confocal microscopy such as Total Variation and more recently priors using redundant wavelet transforms. We focus in this paper on the latter and present, still in section II, the Dual-Tree Complex Wavelet transform. In section III, we present the Alternating Direction technique and detail how to use this algorithm to solve both constrained and unconstrained problems. The first part of section IV is devoted to the introduction of two recent techniques to automatically select the regularizing parameter. We show how to compute more accurately these estimators by taking into account the physical properties of confocal images. From these estimation techniques, we propose a new constrained formulation for the resolution of the Poisson noisy images deconvolution problem. Finally, in the section V, we present results on 2D synthetic and 3D real data using the Total Variation, the Dual-Tree Complex Wavelet Transform and the dictionary composed of the Curvelets and the undecimated wavelet transform.

II. PRIORS FOR CONFOCAL MICROSCOPY

A. State of art

As discussed previously, many works promote the introduction of explicit priors on the solution to regularize the ill-posed inverse problem. Maximizing the a posteriori probability $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})\frac{p(\mathbf{x})}{p(\mathbf{y})}$, where $p(\mathbf{x})$ is the prior model on the object given by $p(\mathbf{x}) = \alpha \exp[-\tau J_R(\mathbf{x})]$ (α is a normalization constant and J_R is the regularizing term), is equivalent to solve:

$$\begin{aligned} \arg\min \quad & J(\mathbf{x},\mathbf{y}) := J_L(\mathbf{x},\mathbf{y}) + \tau J_R(\mathbf{x}) \ , \\ \text{subject to} \quad & \mathbf{x} \in \mathbb{R}^n \end{aligned} \quad (4)$$

τ being the regularizing parameter. To the best of our knowledge, the first regularizing term proposed in confocal microscopy was the Tikhonov-Miller regularization given by $J_R(\mathbf{x}) = \|\nabla\mathbf{x}\|_2^2$ [41]. This regularization is efficient to remove noise but its main drawback is that it smooths the edges and the details. To avoid this effect, the authors of [16] proposed to use instead the l^1-norm of the gradient leading to a well-known regularization in 2D image processing called the Total Variation (TV) [35]. The TV removes the noise while saving the discontinuities but smooths the details of the textures and the corners of the shapes.

Wavelets priors have been successfully used in 2D image processing to retrieve thin elements including textures ([25] and references therein). But it is only recently that these priors have been introduced in confocal microscopy [3], [4], [10]. These priors assume (and it is actually verified) that images have a compact representation (sparsity) in some wavelet basis. This, in the end, gives a good ability to remove the noise

from the image. This sparsity can be forced by using a l^1-norm term. For example, [10] used a wavelet Haar transform as a prior and [17] proposed to use a decomposition on a dictionary composed of an undecimated wavelet and a Curvelet transform. However, [17] only consider 2D images. On 3D images, using an undecimated wavelet transform leads to implementation issues as it involves an image which is $7L$ times higher than the image to restore (L being the number of decomposition levels). Thus, it is problematic in term of memory cost. Note that one can use transforms with limited redundancy such as [44]. We strongly believe however that $3D$ data can not handle this type of redundancy as it is highly consuming in term of computing time and memory cost.
Of course, the prior should be chosen according to the computing resources available. The purpose of this paper is not to do an exhaustive comparison of the different priors but only to present an unified framework for the deconvolution of Poissonian images in which we propose to use the DTCW transform as it seems to be a good trade-off between the computing resources needed and the quality of results.
Finally, let us note that [31] showed that these wavelets priors give better results when combined to the Total Variation.

B. Dual-Tree Complex Wavelet prior

As shown for example in [17], an undecimated wavelet transform and more generally the decomposition on dictionary using several wavelet transforms clearly improves the quality of the restored image. However this regularization technique for 3D images is really difficult to use as it needs a huge amount of memory. For this reason, we propose here to use the Dual-Tree Complex Wavelet Transform [36] which is an efficient wavelet transform with a reduced redundancy (8 in 3D). Unlike [17] where a prior expressed in the transform domain (synthesis prior) is used, we express our prior in the image domain (analysis prior), as it seems to give better results than a synthesis prior (at least for Gaussian noise [7], [18], [37]).

The DTCW transform uses two real trees combined to give complex coefficients (illustrated on the figure 1). The combination of these two trees offers several avantages including a translation and rotation *quasi*-invariant transform and a limited reduduncy.

The DTCW transform has been proposed in the domain of confocal microscopy in [30] but only for denoising. We will show here that we can include the deconvolution process in the algorithms. To improve the results, we also propose to use a subband-dependent regularization parameter. The proposed regularization writes:

$$J_R(\mathbf{x}) = \sum_{j=1}^{L} \alpha_j \|\mathbf{W}_j\mathbf{x}\|_1, \quad (5)$$

where L is the number of decomposition levels, $\mathbf{W}_j$ is the decomposition on the level j and α_j is the subband-dependent scale parameter. The DTCW transform, as common non-redundant wavelet transforms, uses filters normalized to
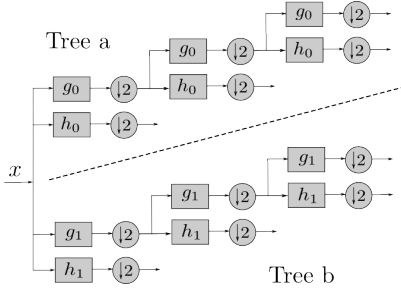
Fig. 1. Decomposition scheme of the DTCW transform for an 1D signal. This transform uses the two trees a and b in parallel. g_0 (respectively g_1) is the low-pass filters of the tree a (respectively tree b), h_0 (respectively h_1) is the high-pass filter of the tree a (respectively tree b). At each level, the details coefficients of each tree give the real and the imaginary part of the complex coefficients.

$\sqrt{2}$ and subsampling operators by factor 2 (see figure 1). Consequently, the scale of wavelet coefficients, following each dimension of the signal, is decreasing by a factor $\sqrt{2}$ at each decomposition level. As we deal with 3D data, the scale of the 3D coefficients is thus decreasing at each scale by a factor $2\sqrt{2}$, thus we will take:

$$\alpha_j = \left(2\sqrt{2}\right)^{-j}. \tag{6}$$

Even if different images have a different power law decrease of wavelet coefficients, this scaling is only meant to be quite general for $3D$ data and independent of the content of the image.

Low-pass coefficients are not included in the formulation (5). As there is no reason for these coefficients to be sparse, we set $\alpha_0 = 0$. By integrating the scaling dependance in the transform $\mathbf{W}_j$, (5) can be written as :

$$J_R(\mathbf{x}) = \|\mathbf{Wx}\|_1, \tag{7}$$

$\mathbf{W}$ standing for the whole transform which includes the weights α_j. Note that the tight-frame property of the initial DTCW is conserved as an operation to invert these weights is included in the computation of the adjoint operator $\mathbf{W}^*$. We present in the next section a state of art of minimization algorithms and the proposed algorithm to solve (4).

III. ALGORITHMS FOR CONFOCAL MICROSCOPY

A. State of art

The most widely used algorithm in confocal microscopy may be the Richardson-Lucy algorithm [24], [34]. Using the fact that the PSF is normalized, minimizing (3) leads to the RL algorithm (multiplicative form):

$$\mathbf{x}_{k+1} = \mathbf{x}_k \left\{ \mathbf{H}^* \left[\frac{\mathbf{y}}{\mathbf{Hx}_k + \mathbf{b}} \right] \right\}, \tag{8}$$

where $\mathbf{H}^*$ denotes the adjoint operator of $\mathbf{H}$. Here, multiplication and division must be understood as point to point operations. This algorithm has two interesting properties. It preserves the number of counts of the original object and has also the property the non-negativity: if the first estimate is positive, then the further estimates stay positive. This algorithm improves the quality of images, however it amplifies

the noise after several iterations [42].

Adding the Tikhonov-Miller regularization in the model leads to the following multiplicative algorithm:

$$\mathbf{x}_{k+1} = \frac{\mathbf{x}_k}{1 - 2\tau \operatorname{div}\left(\nabla \mathbf{x}_k\right)} \left\{ \mathbf{H}^* \left[\frac{\mathbf{y}}{\mathbf{Hx}_k + \mathbf{b}} \right] \right\}, \tag{9}$$

where ∇ is the gradient operator and div is the divergence operator (we will use the discretization proposed in [9] for the implementation of these operators). As discussed previously, using the l^2-norm on the gradient smooths the edges. The authors of [16] used instead a l^1-norm and obtained the following algorithm:

$$\mathbf{x}_{k+1} = \frac{\mathbf{x}_k}{1 - \tau \operatorname{div}\left(\frac{\nabla \mathbf{x}_k}{|\nabla \mathbf{x}_k|}\right)} \left\{ \mathbf{H}^* \left[\frac{\mathbf{y}}{\mathbf{Hx}_k + \mathbf{b}} \right] \right\}. \tag{10}$$

This algorithm, and more generally algorithms built under the multiplicative form of the Richardson-Lucy algorithm, may suffer from unstability. Even a small value of τ may result in a negative denominator in (9) or (10) breaking therefore the positivity property of the RL algorithm. [16] proposed to use the additive form of the RL algorithm to have a stable behaviour regarding to τ, that is to minimize (4) using a gradient descent. On this problem, both the l^1-norm and the logarithm (if $\mathbf{b} = 0$) have to be smoothed by adding a *small* constant ϵ, making the resulting algorithm very slow with a step descent of order $O(\epsilon^2)$. It can be accelerated using one of the framework of [5], [28], but even with these techniques, an accelerated algorithm on this problem can not give competitive computing time as the step descent is too small.

Several authors proposed to extend "well-known" 2D deconvolution algorithms to 3D confocal microscopy. For example, the authors of [10] proposed to use the "forward-backward" algorithm [11]. However, this algorithm can not be directly used here as the Poisson model leads to solve a problem which does not belong to the class of problems of this algorithm (the obtained criterion to minimize is convex but does not have the Lipschitz gradient property required by [11]). In consequence, the authors of [10] proposed to use a variance stabilizing transform (VST) on the data, the Anscombes [1] transform in that case, such that the Poisson noise is approximated as a Gaussian noise (thereby giving a minimizing criterion which has a Lipschitz gradient). [17] also proposed to use the Anscombes transform but refined the model such that the VST is taken into account in the data term and then solved the problem using an extension of the algorithm [10]. These VST may however not be efficient for images with a weak intensity as in confocal microscopy so [13] proposed a quadratic extension of the Poisson criterion such that "forward-backward" algorithm can be directly used on a Poisson model.

Non-iterative methods can also be used as in [33], where the algorithm used is the Tikhonov-regularized algorithm which leads to an explicit solution if the noise model is considered mainly to be Gaussian. This is actually verified

for biological images with high intensity as, in that case, the Poisson distribution is well approximated by a Gaussian distribution. This model may not be very efficient when dealing with biological images with weak intensity.

Recently, [14] introduced an algorithm which is able to minimize the sum of an arbitrary number of convex functions. [31] showed that this algorithm can be used on the Poisson deconvolution problem. [38] proposed also an efficient algorithm based on split Bregman techniques which really takes into account the Poisson noise statistics. These techniques consist in augmenting the size of the problem by adding several variables and then to solve the problem following each variable. This is actually closely related to the algorithm used in this paper, which is based on the Alternating Direction Method (ADM), and has also been recenlty proposed for the Poisson deconvolution problem in [20].

B. Alternating Direction Method

We propose to use an algorithm based on the alternating direction method (ADM) [19], [29]. A similar algorithm has been proposed recently in [20]. We recall the main ideas of the ADM in the following.

The ADM was initially proposed to solve the following problem:

$$\begin{aligned} \arg\min \quad & f_1(\mathbf{u}) + f_2(\mathbf{v}) \quad , \\ \text{subject to} \quad & \mathbf{A}\mathbf{u} + \mathbf{B}\mathbf{v} = \mathbf{a} \\ & \mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^m \end{aligned} \quad (11)$$

where:

- $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^m \to \mathbb{R}$ are two closed convex functions.
- $\mathbf{A} \in \mathbb{R}^{l \times n}$ and $\mathbf{B} \in \mathbb{R}^{l \times m}$ are two linear transforms.
- $\mathbf{a} \in \mathbb{R}^l$ is a given vector.

This algorithm is based on the minimization of the augmented Lagrangian. Using a Lagrange multiplier $\lambda \in \mathbb{R}^l$ for the linear constraint (11), the augmented Lagrangian writes:

$$\begin{aligned} \mathcal{L}(\mathbf{u}, \mathbf{v}, \lambda) = {} & f_1(\mathbf{u}) + f_2(\mathbf{v}) + \lambda^T (\mathbf{A}\mathbf{u} + \mathbf{B}\mathbf{v} - \mathbf{a}) \\ & + \frac{\beta}{2} \|\mathbf{A}\mathbf{u} + \mathbf{B}\mathbf{v} - \mathbf{a}\|_2^2, \end{aligned} \quad (12)$$

where β is a parameter which controls the linear constraint [21]. This algorithm consists in finding a saddle point of the augmented Lagrangian (thereby solving (11)), by minimizing it in an alternating way, subject to $\mathbf{u}, \mathbf{v}$, then to λ. The algorithm is given in algorithm 1.

This algorithm introduces a relaxation parameter ξ which has to belongs to $]0, \frac{\sqrt{5}+1}{2}[$ to ensure the convergence of the algorithm [21]. We will set this parameter to be equal to 1 in our experiments. β is the parameter which controls the constraint. The algorithm converges for $\forall \beta > 0$, however the speed of convergence strongly depends on this parameter. If β is small, the convergence of the algortihm will be fast but the linear constraint will take many more iterations to be respected. On the contrary if β is high, then the algorithm will

---

Algorithm 1: ADM to solve (11)

Data: Number of iterations N;
A starting point $\mathbf{u}^0 \in \mathbb{R}^n$;
A starting point $\mathbf{v}^0 \in \mathbb{R}^m$;
A starting point $\lambda^0 \in \mathbb{R}^l$;
Value of the parameters $\xi > 0$ and $\beta > 0$;
Result: $(\mathbf{u}^N, \mathbf{v}^N)$, an estimated solution of (11).
begin
  for *k from* 0 *to* $N-1$ **do**
    Step 1. $\mathbf{u}^{k+1} = \arg\min \quad \mathcal{L}(\mathbf{u}, \mathbf{v}^k, \lambda^k)$.
                    subject to $\mathbf{u} \in \mathbb{R}^m$
    Step 2. $\mathbf{v}^{k+1} = \arg\min \quad \mathcal{L}(\mathbf{u}^{k+1}, \mathbf{v}, \lambda^k)$.
                    subject to $\mathbf{v} \in \mathbb{R}^n$
    Step 3.
    $\lambda^{k+1} = \lambda^k + \beta\xi(\mathbf{A}\mathbf{u}^{k+1} + \mathbf{B}\mathbf{v}^{k+1} - \mathbf{a})$.
  end
end

---

be slow but the linear constraint will be quickly respected. Setting this parameter is actually an open problem and, for our expriments, we will set this parameter equal to 1.

We show in the next lines how the ADM algorithm can be used to solve the Poissonian deconvolution problem expressed in the unconstrained form. We will see in section IV-D how to apply this algorithm to the constrained optimization problem.

We recall that we want to minimize:

$$\begin{aligned} J(\mathbf{x}, \mathbf{y}) & := J_L(\mathbf{x}, \mathbf{y}) + \tau J_R(\mathbf{x}) \\ & = \mathbf{1}^T (\mathbf{H}\mathbf{x} + \mathbf{b}) - \mathbf{y}^T \log(\mathbf{H}\mathbf{x} + \mathbf{b}) + \tau\|\mathbf{W}\mathbf{x}\|_1. \end{aligned} \quad (13)$$

This function is a closed convex function and strictly convex if $\mathbf{y}_i > 0$ and if the intersection of the null spaces of J_L and J_R is zero [20]. As often mentioned, this type of problem is not straightforward to solve due to the l^1-norm non-differentiability and due to the presence of operators $\mathbf{H}$ and $\mathbf{W}$. Our problem is then to find:

$$\begin{aligned} \arg\min \quad & \mathbf{1}^T (\mathbf{H}\mathbf{x} + \mathbf{b}) - \mathbf{y}^T \log(\mathbf{H}\mathbf{x} + \mathbf{b}) + \tau\|\mathbf{W}\mathbf{x}\|_1 \\ \text{subject to} \quad & \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \geq 0 \end{aligned} \quad (14)$$

First, we can see that this problem is actually equivalent to:

$$\begin{aligned} \arg\min \quad & \mathbf{1}^T \mathbf{w} - \mathbf{y}^T \log(\mathbf{w}) + \tau\|\mathbf{z}\|_1 \quad . \\ \text{subject to} \quad & \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \geq 0 \\ & \mathbf{w} \in \mathbb{R}^n, \mathbf{w} = \mathbf{H}\mathbf{x} + \mathbf{b} \\ & \mathbf{z} \in \mathbb{R}^m, \mathbf{z} = \mathbf{W}\mathbf{x} \end{aligned} \quad (15)$$

We set:

$$\mathbf{u} = \begin{pmatrix} \mathbf{x} \\ \mathbf{w} \\ \mathbf{z} \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m, \quad \mathbf{a} = \begin{pmatrix} 0 \\ -\mathbf{b} \\ 0 \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m, \quad (16)$$

$$\mathbf{A} = -\mathbf{I}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H} \\ \mathbf{W} \end{bmatrix}, \quad (17)$$

$$f_1(\mathbf{u}) = \mathbf{1}^T\mathbf{w} - \mathbf{y}^T\log(\mathbf{w}) + \tau\|\mathbf{z}\|_1 + \chi_C(\mathbf{x}), \quad f_2(\mathbf{v}) = 0, \tag{18}$$

with χ_C being the indicator function on the non-empty convex set C:

$$\chi_C(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x}\in\mathbb{R}^n, \mathbf{x}_i \geq 0 \\ \infty & \text{otherwise} \end{cases}. \tag{19}$$

Then the problem (14) can be written as:

$$\begin{array}{cl} \arg\min & f_1(\mathbf{u}) \\ \text{subject to} & -\mathbf{u} + \mathbf{Bv} = \mathbf{a} \\ & \mathbf{u}\in\mathbb{R}^n, \mathbf{v}\in\mathbb{R}^m \end{array}. \tag{20}$$

We see that this formulation completely fits into the framework of the ADM method (11). The first step of the algorithm is to find:

$$\mathbf{u}^{k+1} = \begin{array}{cl} \arg\min & \mathcal{L}(\mathbf{u}, \mathbf{v}^k, \lambda^k) \\ \text{subject to} & \mathbf{u}\in\mathbb{R}^n \end{array}. \tag{21}$$

From (20) we can write the augmented Lagrangian as:

$$\mathcal{L}(\mathbf{u},\mathbf{v},\lambda) = f_1(\mathbf{u}) + \lambda^T(\mathbf{Bv}-\mathbf{u}-\mathbf{a}) + \frac{\beta}{2}\|\mathbf{Bv}-\mathbf{u}-\mathbf{a}\|_2^2. \tag{22}$$

Then (21) becomes:

$$\begin{aligned} \mathbf{u}^{k+1} &= \begin{array}{cl} \arg\min & f_1(\mathbf{u}) + \lambda^{k^T}(\mathbf{Bv}^k-\mathbf{u}-\mathbf{a}) \\ & +\frac{\beta}{2}\|\mathbf{Bv}^k-\mathbf{u}-\mathbf{a}\|_2^2 \\ \text{subject to} & \mathbf{u}\in\mathbb{R}^n \end{array} \\ &= \begin{array}{cl} \arg\min & \frac{1}{\beta}f_1(\mathbf{u}) + \frac{1}{2}\|\mathbf{Bv}^k-\mathbf{u}-\mathbf{a}+\frac{\lambda^k}{\beta}\|_2^2 \\ \text{subject to} & \mathbf{u}\in\mathbb{R}^n \end{array} \\ &= \text{prox}_{\frac{1}{\beta}f_1}\left(\mathbf{Bv}^k-\mathbf{a}+\frac{\lambda^k}{\beta}\right), \end{aligned} \tag{23}$$

where prox is the proximal operator defined by [15]:

$$\text{prox}_{\gamma f}(\mathbf{x}_0) = \begin{array}{cl} \arg\min & \gamma f(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}_0-\mathbf{x}\|_2^2 \\ \text{subject to} & \mathbf{x}\in\mathbb{R}^n \end{array}. \tag{24}$$

This proximal operator can be computed in closed-form for some functions f. We give here some examples [15]:

- If $f(x) = \|x\|_1$ then $\text{prox}_{\gamma f}(\mathbf{x}_0)$ is the soft-thresholding operator $\text{shrink}_\gamma(\mathbf{x}_0)$ of threshold γ given by:

$$\text{shrink}_\gamma(\mathbf{x}_0) = \text{sign}(\mathbf{x}_0)\max(|\mathbf{x}_0|-\gamma, 0). \tag{25}$$

- If $f(x) = \mathbf{1}^T\mathbf{x} - \mathbf{y}^T\log(\mathbf{x})$ then:

$$\text{prox}_{\gamma f}(\mathbf{x}_0) = \frac{1}{2}\left(\mathbf{x}_0 - \gamma + \sqrt{(\mathbf{x}_0-\gamma)^2 + 4\gamma\mathbf{y}}\right). \tag{26}$$

- If $f(x) = \chi_C(\mathbf{x})$ is the indicator function on a convex set C, then:

$$\text{prox}_{\gamma f}(\mathbf{x}_0) = \Pi_C(\mathbf{x}_0), \tag{27}$$

is the orthogonal projection on this set.

As the proximal operator is componentwise, from (23) we get that for any $\mathbf{u} = \begin{bmatrix}\mathbf{x}\\\mathbf{w}\\\mathbf{z}\end{bmatrix}\in\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^m$:

$$\text{prox}_{\frac{1}{\beta}f_1}(\mathbf{u}) = \begin{pmatrix}\max(\mathbf{x},0)\\\frac{1}{2}\left[\mathbf{w}-\frac{1}{\beta}+\sqrt{\left(\mathbf{w}-\frac{1}{\beta}\right)^2+\frac{4\mathbf{y}}{\beta}}\right]\\\text{sign}(\mathbf{z})\max(|\mathbf{z}|-\frac{\tau}{\beta},0)\end{pmatrix}. \tag{28}$$

The second step of the algorithm is to find:

$$\begin{aligned} \mathbf{v}^{k+1} &= \begin{array}{cl} \arg\min & \mathcal{L}(\mathbf{u}^{k+1}, \mathbf{v}, \lambda^k) \\ \text{subject to} & \mathbf{v}\in\mathbb{R}^n \end{array} \\ &= \begin{array}{cl} \arg\min & \lambda^{k^T}(\mathbf{Bv}-\mathbf{u}^{k+1}-\mathbf{a}) \\ & +\frac{\beta}{2}\|\mathbf{Bv}-\mathbf{u}^{k+1}-\mathbf{a}\|_2^2 \\ \text{subject to} & \mathbf{v}\in\mathbb{R}^n \end{array} \\ &= \begin{array}{cl} \arg\min & \|\mathbf{Bv}-\mathbf{u}^{k+1}-\mathbf{a}+\frac{\lambda^k}{\beta}\|_2^2 \\ \text{subject to} & \mathbf{v}\in\mathbb{R}^n \end{array}. \end{aligned} \tag{29}$$

Then the solution of (29) can be written as the solution of the following linear system:

$$\mathbf{B}^*\mathbf{Bv}^{k+1} = \mathbf{B}^*\left(\mathbf{u}^{k+1} - \frac{\lambda^k}{\beta} + \mathbf{a}\right), \tag{30}$$

which can always be solved with a conjugate gradient method since $(\mathbf{B}^*\mathbf{B})^* = \mathbf{B}^*\mathbf{B}$. However, using this technique, for each iteration of this inner loop, we have to compute $\mathbf{B}^*\mathbf{B}$. Even if the conjugate gradient loop only need 6 to 7 iterations to give a solution of precision 10^{-5}, this is really costly for a 3D image and leads to high computing time. But (30) can be exactly solved depending on the structure of the matrices $\mathbf{H}$ and $\mathbf{W}$. First, the convolution matrix $\mathbf{H}$ can, most of the time, be well implemented using the Fast Fourier Transform (FFT). Second, the matrix $\mathbf{W}$ has often a structure which also favour in the same way the computation of $\mathbf{W}^*\mathbf{W}$. For example, if $\mathbf{W}$ is the Total Variation, then $\mathbf{W}^*\mathbf{W}$ can also be computed using the FFT. But let us point out the great interest in using a (normalized) tight-frame (i.e. a transform $\mathbf{W}$ such that $\mathbf{W}^*\mathbf{W} = \mathbf{I}$) like the DTCW transform, Curvelets [6], undecimated wavelet transform, or both as in [17]. In that particular case, (30) simply writes:

$$\mathbf{v}^{k+1} = (\mathbf{H}^*\mathbf{H} + 2\mathbf{I})^{-1}\mathbf{B}^*\left(\mathbf{u}^{k+1} - \frac{\lambda^k}{\beta} + \mathbf{a}\right), \tag{31}$$

which can be easily computed using the FFT. The resulting algorithm is given in the algorithm 2.

Algorithm 2: ADM to solve (14)

Data: Number of iterations N;
A starting point $\mathbf{u}^0\in\mathbb{R}^n$;
A starting point $\mathbf{v}^0\in\mathbb{R}^m, \mathbf{t}^0 = \mathbf{Bv}^0 - \mathbf{a}\in\mathbb{R}^l$;
A starting point $\lambda^0\in\mathbb{R}^l$;
Value of the parameters $\xi > 0$ and $\beta > 0$;
Result: $\mathbf{v}^N$ an estimated solution of (14).
begin
  for k *from* 0 *to* $N-1$ **do**

    Step 1. $\mathbf{u}^{k+1} = \text{prox}_{\frac{1}{\beta}f_1}\left(\mathbf{t}^k + \frac{\lambda^k}{\beta}\right)$

    Step 2. $\mathbf{v}^{k+1} = (\mathbf{H}^*\mathbf{H} + \mathbf{W}^*\mathbf{W} + \mathbf{I})^{-1}\mathbf{B}^*\left(\mathbf{u}^{k+1} + \mathbf{a} - \frac{\lambda^k}{\beta}\right)$

    Step 3. $\mathbf{t}^{k+1} = \mathbf{Bv}^{k+1} - \mathbf{a}$
    Step 4. $\lambda^{k+1} = \lambda^k + \beta\xi(\mathbf{t}^{k+1} - \mathbf{u}^{k+1})$
  end
end

The interesting point is that this algorithm converges even if (30) is not solved exactly (to be more specific, this algorithm converges if the errors are summable [19]). In pratice, this algorithm converges to a solution of (14) in 200 iterations. However, the computing time strongly depends on the regularizing term. For example, using the TV regularization, this algorithm converges in 25 minutes (for an image with a size of $256 \times 256 \times 64$ voxels). With the DTCW regularization, the computed time is much longer and about 1h15 for the same number of iterations.

The main drawback of this algorithm, and more generally algorithms which use an augmentation of the size of the problem, is that it needs some memory. This allows, however, to get a reasonable computing time.

IV. REGULARIZATION PARAMETER SELECTION METHODS

A. State of art

In most of the deconvolution methods proposed in the literature, the regularizing parameter τ has to be chosen such that it gives the best qualitative results. However, the interpretation of an image may be difficult in biology for example, specially in the presence of artifacts. To overcome this problem, several authors proposed to handle the Poisson noise as a Gaussian noise in the restoration algorithm and/or to use regularization parameter selection methods originally designed for Gaussian noise. For example, [17] proposed to use the generalized cross validation (GCV) [22] on the Anscombe transform of the Poisson noise to estimate the regularizing parameter. [33] extended the work of [32] to an unbiaised estimator of the Mean Squared Error (MSE) for Poisson noise, leading to a pixel-dependent estimator. Consequently, they proposed to use a Tikhonov-regularized algorithm coupled with a least-square criterion for the data term (and thereby consider the noise to be Gaussian). The solution of the proposed cost function can be expressed in closed-form and thus allows the estimator to be easily implemented. Recently, [2] proposed to restore the image using an algorithm designed for Poisson noise and present a method to select the regularized parameter based on a Gaussian approximation of the noise. We detail this technique in this section as the Gaussian approximation is often made in the domain of Poisson deconvolution.

To the best of our knowledge, only [43] have proposed both a method to select the regularizing parameter especially designed to take into account the Poisson statistics of the noise and an algorithm which also deals with Poisson noise. We also recall their method.

The idea of *Bardsley et al.* [2] is to use a Taylor approximation to obtain a quadratic approximation of the term $J_L(\mathbf{x}, \mathbf{y})$ in (3) and to use a discrepancy principle on the approximation. First, they showed that if we consider J_L as a function of $\mathbf{x}$ and $\mathbf{y}$, then the Taylor approximation around the exact objects $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{b}$ writes:

$$J_L(\mathbf{x}, \mathbf{y}) \simeq J_L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \Psi_{\mathbf{y}}(\mathbf{x}), \qquad (32)$$

with:

$$\Psi_{\mathbf{y}}(\mathbf{x}) = \frac{1}{2}\| \left(\mathbf{H}\mathbf{x} - (\mathbf{y} - \mathbf{b}) \right) / \sqrt{\mathbf{y}} \|_2^2. \qquad (33)$$

Using the expected value function E (with respect to the distribution law of $\mathbf{y}$), it can be written that around $\bar{\mathbf{x}}$:

$$J_L(\bar{\mathbf{x}}, \mathbf{y}) \simeq J_L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + E(\Psi_{\mathbf{y}}(\mathbf{x})). \qquad (34)$$

Denoting $\mathbf{x}_\tau$ an estimated solution of (4) and using (32), it is reasonable to write:

$$J_L(\mathbf{x}_\tau, \mathbf{y}) \simeq J_L(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \Psi_{\mathbf{y}}(\mathbf{x}_\tau). \qquad (35)$$

By combining (34) and (35) we can say that a good value of τ is the one which verifies:

$$\Psi_{\mathbf{y}}(\mathbf{x}_\tau) = E(\Psi_{\mathbf{y}}(\bar{\mathbf{x}})). \qquad (36)$$

[2] showed that $E(\Psi_{\mathbf{y}}(\bar{\mathbf{x}}))$ can be well estimated using a common approximation [23]:

$$\mathbf{y} - \mathbf{b} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{e}, \qquad (37)$$

where $\mathbf{e}$ is a Gaussian random variable with mean 0 and multidimensional variance $\mathbf{y}$. If we set:

$$\mathbf{r}(\mathbf{x}) = \left(\mathbf{H}\mathbf{x} - (\mathbf{y} - \mathbf{b}) \right) / \sqrt{\mathbf{y}}. \qquad (38)$$

Then $\mathbf{r}(\bar{\mathbf{x}})$ is a Gaussian random variable with mean 0 and variance $\mathbf{I}$. In this case, a standard result gives:

$$\|\mathbf{r}(\bar{\mathbf{x}})\|_2^2 \sim \chi^2(n), \qquad (39)$$

where $\chi^2(n)$ is the chi-square distribution with n degree of freedom which has a mean equal to n. Using this result and (33)-(36) we get that a good value of τ verifies:

$$\Psi_{\mathbf{y}}(\mathbf{x}_\tau) \simeq \frac{n}{2}. \qquad (40)$$

Finally, the discrepancy principle proposed by [2] to find the regularizing parameter τ is the following:

$$\tau_{opt} = \begin{array}{c} \arg\min \\ \text{subject to} \end{array} \begin{array}{c} \left(\Psi_{\mathbf{y}}(\mathbf{x}_\tau) - \frac{n}{2} \right)^2 \\ \tau \in \mathbb{R}^+ \end{array}. \qquad (41)$$

In practice, the division by $\mathbf{y}$ in (33) is replaced by $\mathbf{H}\mathbf{x} + \mathbf{b}$ for better experimental results [2], so $\Psi_{\mathbf{y}}$ becomes:

$$\Psi_{\mathbf{y}}(\mathbf{x}) = \frac{1}{2}\| \left(\mathbf{H}\mathbf{x} - (\mathbf{y} - \mathbf{b}) \right) / \sqrt{\mathbf{H}\mathbf{x} + \mathbf{b}} \|_2^2. \qquad (42)$$

Finally, we would like to mention also the recent work of *Bertero et al.* [43] in which is introduced a discrepancy principle for Poisson noise. First, let us consider the following function:

$$F(\mathbf{y}_\lambda) = 2 \left(\mathbf{y}_\lambda \log \left(\frac{\mathbf{y}_\lambda}{\lambda} \right) + \lambda - \mathbf{y}_\lambda \right), \qquad (43)$$

where $\mathbf{y}_\lambda$ is a Poisson random variable with mean λ. Then, [43] showed that, for large λ:

$$E(F(\mathbf{y}_\lambda)) = 1 + O\left(\frac{1}{\lambda} \right). \qquad (44)$$

In the case of deconvolution, we have $\mathbf{y} = \mathcal{P}(\mathbf{Hx} + \mathbf{b})$ and thus $\mathbf{y}$ is a Poisson random variable with mean $\mathbf{Hx} + \mathbf{b}$. So, from this statement and (43), [43] defined:

$$\Upsilon_{\mathbf{y}}(\mathbf{x}) = \left[\mathbf{y}^T \log\left(\frac{\mathbf{y}}{\mathbf{Hx} + \mathbf{b}} \right) + \mathbf{1}^T(\mathbf{Hx} + \mathbf{b}) - \mathbf{1}^T\mathbf{y} \right], \tag{45}$$

and showed that by taking the expected value (following the distribution law of $\mathbf{y}$), one get that a good value of τ should verify:

$$\Upsilon_{\mathbf{y}}(\mathbf{x}_\tau) = E(\Upsilon_{\mathbf{y}}(\bar{\mathbf{x}})) \tag{46}$$

From (43)-(45), one get that:

$$E(\Upsilon_{\mathbf{y}}(\bar{\mathbf{x}})) \simeq \frac{n}{2}. \tag{47}$$

Thus, [43] proposed to select τ as the one which verifies:

$$\tau_{opt} = \begin{array}{c} \arg\min \\ \text{subject to} \end{array} \begin{array}{c} \left(\Upsilon_{\mathbf{y}}(\mathbf{x}_\tau) - \frac{n}{2} \right)^2 \\ \tau \in \mathbb{R}^+ \end{array}. \tag{48}$$

B. Comparisons

We compare the estimator (48) proposed by [43] and the estimator (41) proposed by [2] introduced in the beginning of this paragraph, on the blur and noisy synthetic images presented in figure 2. The algorithm used for this test is the algorithm 2, presented in the previous section, with the TV regularization. Values given by these two estimators are compared to the optimal value given by the minimum of *Mean Square Error* (MSE) which is most of the time the reference measure. It is defined as:

$$MSE(\mathbf{x}_\tau) = \|\bar{\mathbf{x}} - \mathbf{x}_\tau\|_2^2, \tag{49}$$

where $\bar{\mathbf{x}}$ is the true image.

Fig. 3. Comparaison of the two methods for the estimation of τ on the synthetic image (a) from the figure 2.

Fig. 4. Comparaison of the two methods for the estimation of τ on the synthetic image (b) from the figure 2.

But we also check that on some images we may have a huge difference with the value given by the minimization of the MSE. On the images (a) and (c), both methods (48) and (41) give values which are quite close to the value given by the minimization of the MSE. But for all images, these estimators give an estimated value of the regularizing parameter which is greater than the optimal value (in the MSE sense) and tend to overregularize the image.

We also check that these methods may drastically fail on some images, for example on the image (b) where we get an estimated regularizing parameter which is 20 times higher than the optimal value. Actually, our experimentations showed
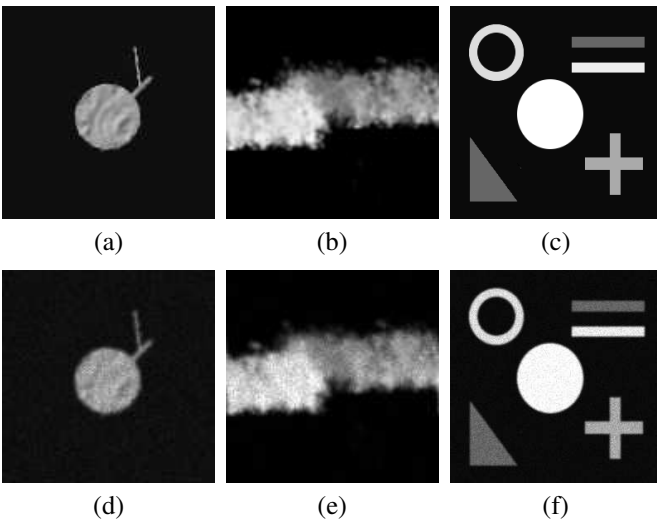

Fig. 2. Synthetic images. The first row is the original images and the second row is the degraded versions (Poisson noise and blur).

Results are shown in table I and figures 3, 4 and 5. We first see that both methods give similar values for all images.
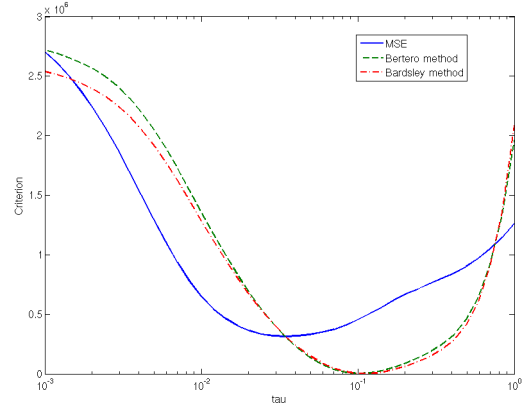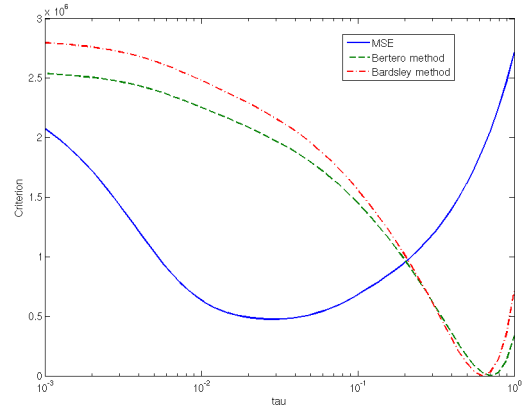
Image Criterion	(a)	(b)	(c)
Bardsley method (41)	0.119	0.626	0.142
Bertero method (45)	0.100	0.703	0.142
MSE (reference measure)	0.034	0.029	0.083

TABLE I
ESTIMATION OF THE REGULARIZING PARAMETER FOR THE TWO ESTIMATORS (41), (48) AND COMPARAISON TO THE REFERENCE MEASURE (49). IMAGES (A), (B) AND (C) ARE PRESENTED IN FIGURE 2.
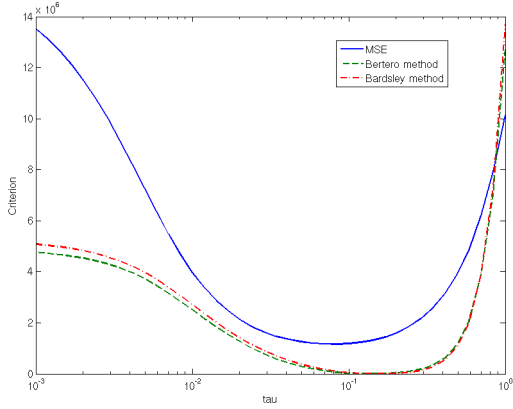
Fig. 5. Comparaison of the two methods for the estimation of τ on the synthetic image (c) from the figure 2.

that these estimators are generally more efficient on images having a background (that is with $\mathbf{b} > 0$) such as images (a) ($\mathbf{b} = 15$) and (c) ($\mathbf{b} = 1$) rather than on images with no brackground ($\mathbf{b} = 0$) such as image (b). For this reason, we propose a slight modification on both estimators.

C. Proposed modified estimators

We consider now that we do not have any background in the model (1) that is:

$$\mathbf{y} = \mathcal{P}(\mathbf{Hx}) \qquad (50)$$

This assumption is actually verified in many applications such as biology or astonomy imaging where background is often zero. The proposed modification mainly rely on the fact that the Poisson distribution is not defined for a zero mean, that is $\mathcal{P}(0) = 0$. It seems clear that for every voxel $\mathbf{x}_i$ inside a centered window (of the size of the kernel of the PSF) containing only 0 valued voxels, we can write that $\mathbf{y}_i = 0$. These pixels should then not be considered in the computation of the estimators (41) and (48) as they are not noisy (more precisely, there is not any pertubations on these pixels which change their values). This observation brings us to consider separately positive and null observed pixels $\mathbf{y}_i$ and to refine the computation of the estimators for each case. Note that this domain splitting actually constains a small approximation as we include in the null observed pixels the ones for which the real image is strictly positive (but small). Dealing with these pixels is however difficult without any knowledge on the real data.

If we consider the Gaussian estimator (41), we immediately see that, from (38), $[\mathbf{r}(\bar{\mathbf{x}})]_i$ is a Gaussian random variable with mean 0 but variance 0 when $\mathbf{y}_i = 0$. It seems thus more accurate to write that $\mathbf{r}(\bar{\mathbf{x}})$ is a Gaussian random variable with mean 0 but variance Σ with:

$$\Sigma = \begin{cases} 1 \text{ if } \mathbf{y}_i > 0, \\ 0 \text{ otherwise} \end{cases} . \qquad (51)$$

Then:

$$\|\mathbf{r}(\bar{\mathbf{x}})\|_2^2 \sim \chi^2(m), \qquad (52)$$

where $m = \#\{\mathbf{y}_i, \mathbf{y}_i > 0\}$. So the modified estimator writes:

$$\tau_{opt} = \begin{array}{c} \arg\min \\ \text{subject to} \end{array} \begin{array}{c} \left(\Psi_{\mathbf{y}}(\mathbf{x}_\tau) - \frac{m}{2}\right)^2 \\ \tau \in \mathbb{R}^+ \end{array} . \qquad (53)$$

where $\Psi_{\mathbf{y}}$ is defined in (42).

The same kind of modification can be applied to (48). If we consider that $0\log(0) = 0$, then $F(\mathbf{y}_\lambda) = 0$ in (43) for $\lambda = 0$. In consequence, we propose to change the estimator (48) to:

$$\tau_{opt} = \begin{array}{c} \arg\min \\ \text{subject to} \end{array} \begin{array}{c} \left(\Upsilon_{\mathbf{y}}(\mathbf{x}_\tau) - \frac{m}{2}\right)^2 \\ \tau \in \mathbb{R}^+ \end{array} . \qquad (54)$$

where $\Upsilon_{\mathbf{y}}$ is defined in (45). We have tested these estimators and the original estimators (41) and (48) on the image (b). Results are shown on figure 6 and table II. These modifications clearly improve the accuracy of the estimation (see table II) and therefore provide better results for images having a dark background as biological images for example. With these new estimators, we are actually very close to the value given by the minimization of the MSE. This modification has actually given very good results for biological images, for example.
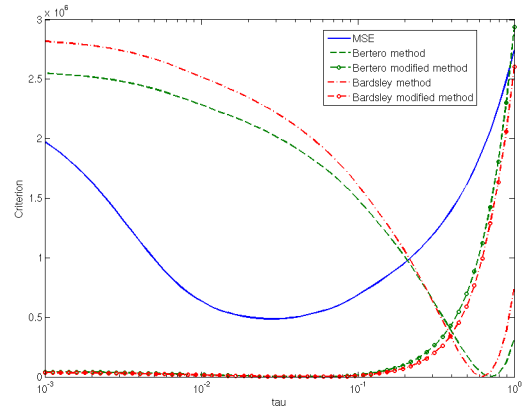

Fig. 6. Comparaison of the estimators (41) and (48) and their respective modification (53) and (54) for the estimation of τ on the synthetic image (b) from the figure 2.

Bardsley method (41)	0.626
Bardsley modified method (53)	0.053
Bertero method (48)	0.703
Bertero modified method (54)	0.047
MSE (reference measure)	0.024

TABLE II
ESTIMATION OF THE REGULARIZING PARAMETER FOR THE ESTIMATORS (41), (48), THEIR RESPECTIVE MODIFICATIONS (53), (54) AND COMPARAISON TO THE REFERENCE MEASURE (49) FOR THE IMAGE (B).

We have also observed that the accuracy of the estimators depends on how well the prior is modeled. Clearly images (a) and (c) are well adapted to a Total Variation prior while this prior may not suit to image (b). For this image, we changed the TV prior for the Dual-Tree Complex Wavelet prior, presented

in the section II-B, which is more efficient to represent the textures at the surface of the object. Results are shown on the figure 7 and table III. We can see that using a suitable prior for the image decreases the relative error of the regularizing parameter estimation from 42.5% to 19.4%. Thus, using an appropriate prior is also crucial to find a correct value of the regularizing parameter. For our experiments, we will use the modified estimator (54) to select the value of the regularizing parameter τ.
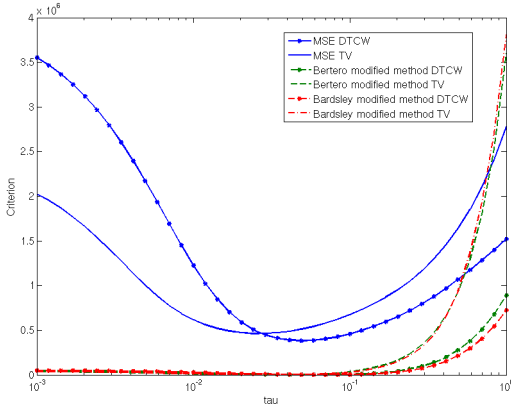

Fig. 7. Comparaison of the total variation prior and the complex wavelet prior for the estimation of τ on the synthetic image (b) from the figure 2.

Prior Criterion	TV	DTCW
Bardsley modified method (53)	0.034	0.070
Bertero modified method (54)	0.034	0.058
MSE (reference measure)	0.024	0.049

TABLE III
ESTIMATION OF THE REGULARIZING PARAMETER REGARDING TWO
DIFFERENT PRIORS (TV AND DTCW) FOR THE IMAGE (B).

D. A new contrained algorithm

The estimators (41) and (48) are highly time consuming as one need to solve several instances of (14) to find a good value of the regularizing parameter τ. We show in this part that we can reformulate the optimization problem (14) to a constrained problem giving, in one minimization instance, the same solution as the one obtained with a value of τ chosen by (54).

First it is easy to show that there exists $\sigma \in]0, +\infty[$ such that a solution of (14) for a given τ is the same as:

$$
\begin{aligned}
\mathbf{x}^* = \quad & \arg\min \quad \|\mathbf{W}\mathbf{x}\|_1 \\
& \text{subject to} \quad \mathbf{x} \in \mathbb{R}^n \\
& \qquad\qquad\quad J_L(\mathbf{x}, \mathbf{y}) \leq \sigma \\
& \qquad\qquad\quad \mathbf{x} \geq 0
\end{aligned}
\tag{55}
$$

The cost function (55) to minimize is a lower semi continuous function over a closed convex set. We assume this convex set to be non empty, so a solution $\mathbf{x}^*$ of (55) exists. We assume moreover that this solution is not trivial, that is to say this

solution is not in the null space of $\mathbf{W}$.

We show in the next line that the optimality conditions of the constrained problem (55) are, for $\sigma = \frac{m}{2}$, the same optimality conditions of the optimization problem (14) with a τ chosen by (54).

The Karush-Kuhn-Tucker optimality conditions of the problem (55) write (conditions on the positivity constraint have been omitted for the sake of simplicity):

$$
\begin{aligned}
\partial J_R(\mathbf{x}^*) + \tau^* \partial J_L(\mathbf{x}^*, \mathbf{y}) &\ni 0 \\
J_L(\mathbf{x}^*, \mathbf{y}) - \sigma &\leq 0 \\
\tau^* (J_L(\mathbf{x}^*, \mathbf{y}) - \sigma) &= 0 \\
\tau^* &\geq 0
\end{aligned}
\tag{56}
$$

The condition $\tau^* = 0$ could actually be remove as we assume the solution $\mathbf{x}^*$ not to be trivial. Therefore, the Karush-Kuhn-Tucker optimality conditions rewrite:

$$
\begin{aligned}
\partial J_R(\mathbf{x}^*) + \tau^* \partial J_L(\mathbf{x}^*, \mathbf{y}) &\ni 0 \\
J_L(\mathbf{x}^*, \mathbf{y}) &= \sigma \\
\tau^* &> 0.
\end{aligned}
\tag{57}
$$

On the other side, the optimality conditions of the problem (14) with a τ_{opt} (noted τ^* in the following) chosen by (54) can be written (remark that $J_L(\mathbf{x}, \mathbf{y}) = \Upsilon_{\mathbf{y}}(x)$):

$$
\begin{aligned}
\partial J_R(\mathbf{x}_{\tau^*}) + \tau^* \partial J_L(\mathbf{x}_{\tau^*}, \mathbf{y}) &\ni 0 \\
J_L(\mathbf{x}_{\tau^*}, \mathbf{y}) &= \frac{m}{2}
\end{aligned}
\tag{58}
$$

Therefore, if one set $\sigma = \frac{m}{2}$ in (55), then one get that the conditions (55) are the same optimality conditions of the unconstrained problem (14) where the optimal regularizing parameter τ^* has been chosen by (54). Hence, we propose to formulate the equivalent constrained problem as:

$$
\begin{aligned}
\mathbf{x}^* = \quad & \arg\min \quad J_R(\mathbf{x}) \\
& \text{subject to} \quad \mathbf{x} \in \mathbb{R}^n \\
& \qquad\qquad\quad J_L(\mathbf{x}, \mathbf{y}) \leq \frac{m}{2} \\
& \qquad\qquad\quad \mathbf{x} \geq 0
\end{aligned}
\tag{59}
$$

The criterion is convex so this problem can also be solved using the ADM method presented in the section III-B. We first formulate problem (59) as:

$$
\begin{aligned}
& \arg\min \quad \|\mathbf{z}\|_1 \\
& \text{subject to} \quad \mathbf{x} \in \mathbb{R}^n \\
& \qquad\qquad\quad \mathbf{x} \geq 0 \\
& \qquad\qquad\quad \Upsilon(\mathbf{w}) \leq \frac{m}{2} \\
& \qquad\qquad\quad \mathbf{w} \in \mathbb{R}^n, \mathbf{w} = \mathbf{H}\mathbf{x} + \mathbf{b} \\
& \qquad\qquad\quad \mathbf{z} \in \mathbb{R}^m, \mathbf{z} = \mathbf{W}\mathbf{x}
\end{aligned}
\tag{60}
$$

with:

$$
\Upsilon(\mathbf{x}) = \mathbf{1}^T(\mathbf{x}) - \mathbf{y}^T \log(\mathbf{x}) + \mathbf{y}^T \log(\mathbf{y}) - \mathbf{1}^T \mathbf{y}.
\tag{61}
$$

We set χ_C to be the indicator function on the non-empty convex set C:

$$
\chi_C(\mathbf{x}) = \begin{cases} 0 & \text{if} \quad \mathbf{x} \in \mathbb{R}^n, \mathbf{x}_i \geq 0 \\ \infty & \text{otherwise} \end{cases},
\tag{62}
$$

χ_K to be the indicator function on the non-empty convex set K defined by:

$$\chi_K(\mathbf{w}) = \begin{cases} 0 & \text{if} \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{w}_i > 0, \quad \Upsilon(\mathbf{w}) \leq \frac{m}{2} \\ \infty & \text{otherwise} \end{cases}, \tag{63}$$

and:

$$\mathbf{u} = \begin{pmatrix} \mathbf{x} \\ \mathbf{w} \\ \mathbf{z} \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m, \quad \mathbf{a} = \begin{pmatrix} 0 \\ -\mathbf{b} \\ 0 \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m, \tag{64}$$

$$\mathbf{A} = -\mathbf{I}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H} \\ \mathbf{W} \end{bmatrix}, \tag{65}$$

$$f_1(\mathbf{u}) = \|\mathbf{z}\|_1 + \chi_K(\mathbf{w}) + \chi_C(\mathbf{x}), \quad f_2(\mathbf{v}) = 0, \tag{66}$$

such that we still fit into the framework (11) of the ADM method:

$$\begin{array}{ll} \arg\min & f_1(\mathbf{u}) + f_2(\mathbf{v}) \\ \text{subject to} & \mathbf{Au} + \mathbf{Bv} = \mathbf{a} \\ & \mathbf{u} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^m \end{array}. \tag{67}$$

This formulation is very similar to the one presented in section III-B. Actually, only the proximal operator in (23) need to be changed. Indeed, for this problem, we have that for any $\mathbf{u} = \begin{bmatrix} \mathbf{x} \\ \mathbf{w} \\ \mathbf{z} \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$:

$$\text{prox}_{\frac{1}{\beta} f_1}(\mathbf{u}) = \begin{pmatrix} \max(\mathbf{x}, 0) \\ \Pi_K(\mathbf{w}) \\ \text{sign}(\mathbf{z}) \max(|\mathbf{z}| - \frac{1}{\beta}, 0) \end{pmatrix}, \tag{68}$$

where Π_K is the orthogonal projection on the convex set K. Even if we can not give a closed-form solution of this projection, we propose an iterative scheme to solve it. We recall that the orthogonal projection problem is to find:

$$\mathbf{w}^* = \Pi_K(\mathbf{w}_0) = \begin{array}{ll} \arg\min & \frac{1}{2}\|\mathbf{w} - \mathbf{w}_0\|_2^2 \\ \text{subject to} & \mathbf{w} \in \mathbb{R}^n \\ & \Upsilon(\mathbf{w}) \leq \frac{m}{2} \end{array}. \tag{69}$$

First notice that if $\Upsilon(\mathbf{w}_0) \leq \frac{m}{2}$ then $\mathbf{w}^* = \mathbf{w}_0$. Otherwise, there exists $\alpha \in]0, +\infty[$ such that:

$$\mathbf{w}^* = \begin{array}{ll} \arg\min & \frac{1}{2}\|\mathbf{w} - \mathbf{w}_0\|_2^2 + \alpha\Upsilon(\mathbf{w}) \\ \text{subject to} & \mathbf{w} \in \mathbb{R}^n \end{array} = \text{prox}_{\alpha\Upsilon}(\mathbf{w}_0)$$

$$= \frac{1}{2}\left[\mathbf{w}_0 - \alpha + \sqrt{(\mathbf{w}_0 - \alpha)^2 + 4\alpha\mathbf{y}} \right] = \Phi(\alpha). \tag{70}$$

The problem is thus to find α such that $\Upsilon(\Phi(\alpha)) \leq \frac{m}{2}$. Let us define:

$$f(\alpha) := \Upsilon(\Phi(\alpha)) - \frac{m}{2}. \tag{71}$$

It can be shown that f is a convex and decreasing function with respect to α. In order to find the root of the function f, we propose to use a Newton method and we only need to find $f'(\alpha)$. Simply remark that from the composition of functions, we have:

$$f'(\alpha) = \frac{\mathbf{1}^T}{2}\left\{ \left[\frac{\alpha - \mathbf{x}_0 + 2\mathbf{y}}{\sqrt{(\mathbf{x}_0 - \alpha)^2 + 4\alpha\mathbf{y}}} - 1 \right] \\ \left[1 - \frac{2\mathbf{y}}{\mathbf{x}_0 - \alpha + \sqrt{(\mathbf{x}_0 - \alpha)^2 + 4\alpha\mathbf{y}}} \right] \right\}. \tag{72}$$

The resulting algorithm is then given in the algorithm 3.

Algorithm 3: Newton method to solve (69)

Data: Number of iterations N;
Set $\alpha^0 = 0$;
Result: $\mathbf{w}^*$ an estimated of the solution of (69).
begin
    for k *from* 0 *to* $N-1$ **do**
        Step 1. $\quad \alpha^{k+1} = \alpha^k - \dfrac{f(\alpha^k)}{f'(\alpha^k)}$
    end
    $\mathbf{w}^* = \frac{1}{2}\left[\mathbf{w}_0 - \alpha^N + \sqrt{\left(\mathbf{w}_0 - \alpha^N\right)^2 + 4\alpha^N\mathbf{y}} \right]$
end

In all our simulations, we checked that 20 iterations are sufficient to get a machine precision.

Our simulations on this contrained problem show that we get the same result as the unconstrained problem (14), as we have shown in the beginning of section IV-D. This is a very important result as, in this case, we do not have to run algorithm 2 several times as it is needed when searching for the good value of the regularizing parameter τ_{opt} in (54). Moreover, the constrained problem (59) appears to be slightly faster to converge than the unconstrained problem (14) (see figure 8).

Finally, let us remark that we can also formulate a constrained problem from the results of the estimator (53):

$$\begin{array}{ll} \arg\min & \|\mathbf{Wx}\|_1 \\ \text{subject to} & \mathbf{x} \in \mathbb{R}^n \\ & \left\|(\mathbf{Hx} - (\mathbf{y} - \mathbf{b}))/\sqrt{\mathbf{y}}\right\|_2^2 \leq m \\ & \mathbf{x} \geq 0 \end{array}. \tag{73}$$

This formulation consider, however, the Poisson noise to be a weight Gaussian noise. A comparison of these two constrained problems is given in the last part of the next section.

V. RESULTS

A. Results on synthetic 2D data

We first compare the improvement of using wavelets as regularizing operator compared to the TV regularization on the synthetic images (a) and (b) of figure 2. Here, the regularization parameter τ has be chosen such that it minimizes the MSE, in order to evaluate the perfomances of each regularizing term. The results obtained on these images
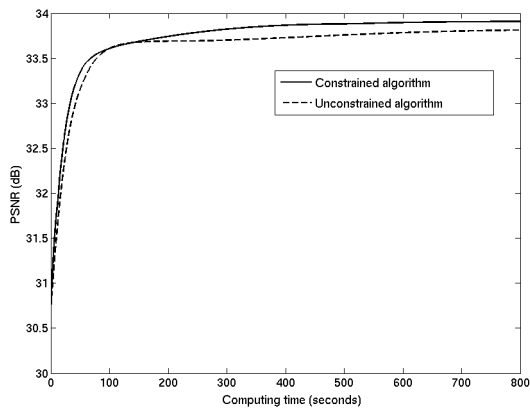
Fig. 8. Comparaison of the convergence of the constrained problem (59) and the unconstrained problem (14) (with a regularizing parameter given by (48)) on the synthetic image (a) from the figure 2.
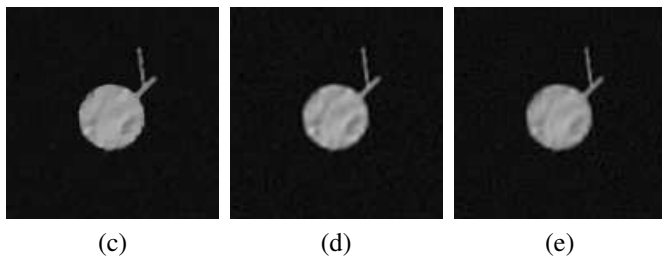

(a)   (b)

(c)   (d)   (e)

Fig. 9. Restoration of a blur and noisy synthetic image. (a) is the original image, (b) is the degraded version ($PSNR = 28.51 \ dB$), (c) is the result obtained with the Total Variation prior ($PSNR = 32.95 \ dB$), (d) is the result obtained with the proposed prior ($PSNR = 32.02 \ dB$), (e) is the result obtained with the prior proposed in [17] ($PSNR = 31.20 \ dB$).

are presented on figures 9 and 10.

On these images, wavelets priors allow to retrieve more details than the TV prior. Visually, we can see that the thin elements are better retrieved, in particular we can distinguish the details of the surface of the object. However, as wavelets regularizations, we may have some artifacts and contours may be slightly smoothed.

B. Results on real 3D data

We propose to test the algorithm on the restoration of the real images presented on figures 11 and 13. On each test, the regularizing parameter has be chosen using the method proposed in section IV-C. The microscope is a confocal/multi-photon Zeiss Axiovert 200M, with an internal magnification
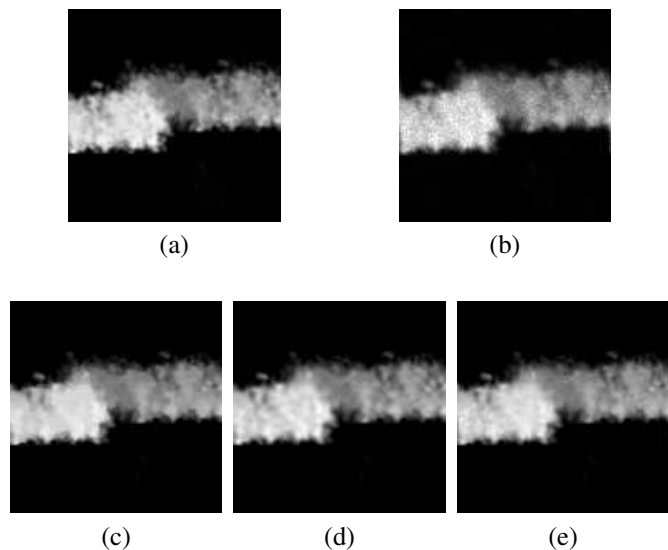

(a)   (b)

(c)   (d)   (e)

Fig. 10. Restoration of a blur and noisy synthetic image. (a) is the original image, (b) is the degraded version ($PSNR = 29.43 \ dB$), (c) is the result obtained with the Total Variation prior ($PSNR = 33.83 \ dB$), (d) is the result obtained with the proposed prior ($PSNR = 34.53 \ dB$), (e) is the result obtained with the prior proposed in [17] ($PSNR = 33.48 \ dB$).

(given by the manufacturer) of 3.3x. The objective is an immersion oil Apochromat[1] 40x for the first image, 63x for the second, with numerical aperture $NA = 1.4$. The oil refractive index is 1.518 (23° C). The acquisition software is Zeiss LSM 510 Meta.
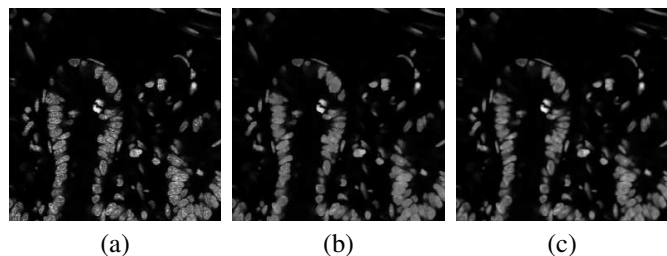

(a)   (b)   (c)

Fig. 11. Restoration of a sample of mouse intestin. (a) is the observed image, (b) is the result obtained with the TV prior and (c) is the result obtained with the DTCW prior.

Figures 11, 12 and 13 show the results obtained with the estimator (54) and the ADM algorithm regularized with the TV and DTCW priors on a sample of mouse intestine and on a bead. As discussed previously, it is difficult on $3D$ data to use a prior composed of several wavelet transforms. For this reason, we can not present, on these images, the result using the prior composed of the Curvelets and the undecimated wavelet transform as in [17].

First, we see on both figures 11 and 13 that the estimator (54) using the DTCW prior tends to oversmooth the image.

---

[1]that is an objective composed by a lens designed to bring light of three colours to the same focal point, thus reducing its chromatic aberration.
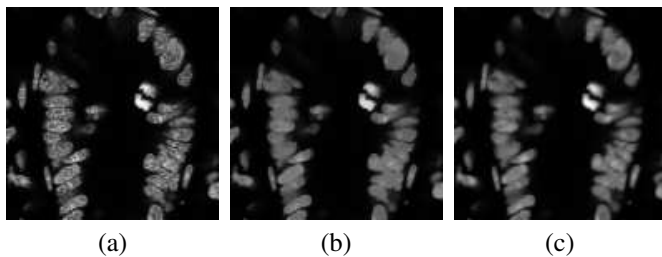
Fig. 12. Zoom on the restoration of the sample of mouse intestin. (a) is the observed image, (b) is the result obtained with the TV prior and (c) is the result obtained with the DTCW prior.
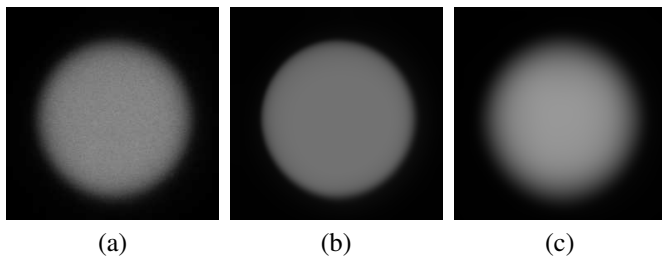

Fig. 13. Restoration of a bead. (a) is the original image, (b) is the result obtained with the TV prior and (c) is the result obtained with the DTCW prior.
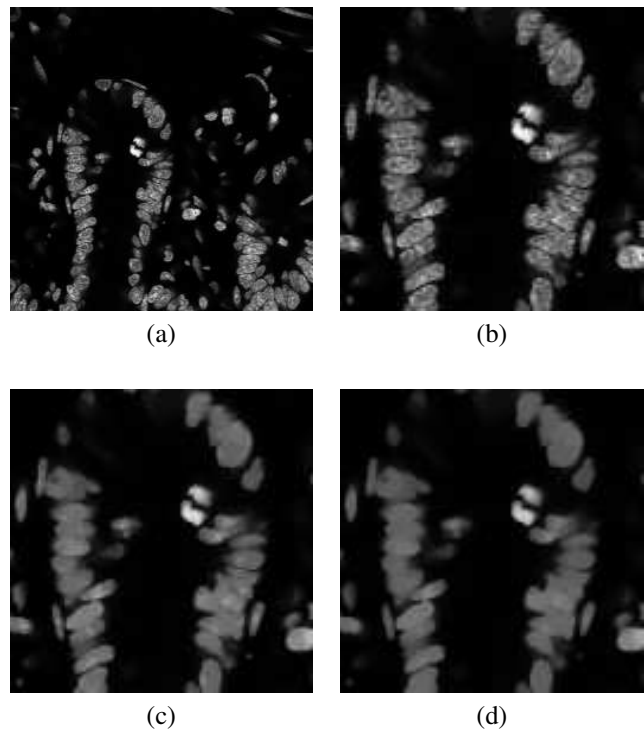

Fig. 14. Results of constrained formulations (59) and (73) on the sample of mouse intestin. (a) is the observed image, (b) is a zoom on the observation, (c) is the result with the constrained formulation (59) and (d) is the result with the constrained formulation (73)

This is quite common for this type of estimator, indeed most of the recent proposed estimator for Poisson deconvolution tends to oversmooth the image [2], [17], [43].

On figures 11 and 12 we see that the TV regularization smooths the textures of the cells, sticking it together forming a large pattern (right of the image). The DTCW prior allows to retrieve some details of the cells and preserves the space between it, even if the image retrieved is slightly smoothed. The DTCW prior is quite efficient and gives details of the inside of the cells (zoom on the figure 12).
The image presented in figure 11 is interesting as it contains many details that can be retrieved using the proposed prior. However, on a smooth object which does not contain many details (as the one presented in figure 13), the proposed prior does not bring much more information.

Finally, we compare the constrained problems (59) and (73) on the sample of mouse intestin (image (a) in figure 11) using the TV regularization. Results are shown in figure 14. We clearly see that the image retrieved with the formulation (59) is less smoothed than the one retrieved with the Gaussian approximation (73). We can distinguish more easily the details of the cells of the object.

VI. CONCLUSION

We have proposed a method for the deconvolution of images corrupted by blur and Poisson noise and applied it on confocal microscopy images. This method includes a wavelet prior which is well adapted to represent the thin structures of specimens in 3D and a method to estimate the value of the regularizing parameter based on a discrepancy principle. A new constrained minimization problem is introduced, allowing easy and fast parameter setting and gives efficient algorithm which is important regarding to the big volume of data coming from 3D confocal images. This method will be analyzed and tested on large data basis in order to estimate its range of validity.

REFERENCES

[1] F. J. Anscombe. The transformation, of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254, Dec. 1948.
[2] J. M. Bardsley and J. Goldes. Regularization parameter selection methods for ill-posed poisson maximum likelihood estimation. *Inverse Problems*, 25(9), 2009.
[3] B. Colicchio, E. Maalouf, O. Haeberlé, and A. Dieterlen Wavelet filtering applied to 3D wide field fluorescence microscopy deconvolution. *Proceedings PSIP'07*, Mulhouse, France, 2007.
[4] J. Boutet de Monvel, S. Le Calvez, and M. Ulfendahl Image Restoration for Confocal Microscopy: Improving the Limits of Deconvolution, with Application to the Visualization of the Mammalian Hearing Organ. *Biophysical Journal*, 80:2455-2470, May 2001.

[5] A. Beck and Teboulle M. A fast iterative shrinkage thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1), 2009.

[6] E. Candès, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006.

[7] M. Carlavan, P. Weiss, and L. Blanc-Féraud. Régularité et parcimonie pour les problèmes inverses en imagerie : algorithmes et comparaisons. *Traitement du Signal* 27(2):187-217, 2010.

[8] M. Carlavan, P. Weiss, L. Blanc-Féraud, and J. Zerubia. Complex wavelet regularization for solving inverse problems in remote sensing. In *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Cape Town, South Africa, Jul. 2009.

[9] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20(1-2):89–97, 2004.

[10] C. Chaux, L. Blanc-Féraud, and J. Zerubia. Wavelet-based restoration methods: application to 3d confocal microscopy images. In *Proc. SPIE Conference on Wavelets*, 2007.

[11] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A forward-backward algorithm for image restoration with sparse representations. In *Signal Processing with Adaptive Sparse Structured Representations (SPARS'05)*, pages 49–52, Rennes, France, Nov. 2005.

[12] C. Chaux, L. Duval, and J.-C. Pesquet. Image analysis using a dual-tree m-band wavelet transform. *Image Processing, IEEE Transactions on*, 15(8):2397–2412, Aug. 2006.

[13] C. Chaux, J.-C. Pesquet, and N. Pustelnik. Nested iterative algorithms for convex constrained image recovery problems. *SIAM Journal on Imaging Sciences*, 2(2):730–762, Jun. 2009.

[14] P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), 2008.

[15] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[16] N. Dey, L. Blanc-Féraud, C. Zimmer, Z. Kam, P. Roux, J. C. Olivo-Marin, and J. Zerubia. Richardson-lucy algorithm with total variation regularization for 3d confocal microscope deconvolution. *Microscopy Research Technique*, 69:260–266, 2006.

[17] F.-X. Dupé, J. Fadili, and J.-L. Starck. A proximal iteration for deconvolving poisson noisy images using sparse representations. *IEEE Transactions on Image Processing*, 18(2):310–321, Feb. 2009.

[18] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, 2007.

[19] E. Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. *UCLA Cam Report*, 2009.

[20] M. A. T. Figueiredo and J. M. Bioucas-Dias. Restoration of poissonian images using alternating direction optimization. *Submitted to IEEE Transactions on Image Processing*, Jan. 2010.

[21] R. Glowinski. *Numerical Methods for Nonlinear variation al Problems*. Springer-Verlag, 1984.

[22] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[23] A. Grinvald and I. Z. Steinberg. On the analysis of fluorescence decay kinetics by the method of least-squares. *Analytical Biochemistry*, 59(2):583–598, 1974.

[24] L. B. Lucy. An iterative technique for rectification of observed distributions. *The Astronomical Journal*, 79(6):745–765, 1974.

[25] S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 2008.

[26] M. Minsky. Memoir on inventing the confocal scanning microscope. *Scanning*, 10:128–138, 1988.

[27] V. A. Morozov. *Regularization Methods for Ill-Posed Problems*. CRC Press, 1993.

[28] Y. Nesterov. Gradient methods for minimizing composite objective function. *CORE Discussion Paper 2007/76*, 2007.

[29] M. Ng, P. Weiss, and X. M. Yuan. Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. *ICM Research Report*, Oct. 2009.

[30] G. Pons Bernad, L. Blanc-Féraud, and J. Zerubia. A restoration method for confocal microscopy using complex wavelet transform. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, Pennsylvania, USA, Mar. 2005.

[31] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Hybrid regularization for data restoration in the presence of Poisson noise. In *17th European Signal Processing Conference (EUSIPCO'09)*, Aug. 2009.

[32] S. Ramani, T. Blu, and M. Unser. Monte-Carlo SURE: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms. *IEEE Transactions on Image Processing*, 17(9):1540–1554, 2008.

[33] S. Ramani, C. Vonesch, and M. Unser. Deconvolution of 3d fluorescence micrographs with automatic risk minimization. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 732–735, May 2008.

[34] W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of Optical Society of America*, 62:55–59, 1972.

[35] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[36] I. W. Selesnick, R.G. Baraniuk, and N. G. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, Nov 2005.

[37] I. W. Selesnick and M. A. T. Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. volume 7446. SPIE, 2009.

[38] S. Setzer, G. Steidl, and T. Teuber. Deblurring poissonian images by split bregman techniques. *Journal of Visual Communication and Image Representation*, 21:193–199, 2010.

[39] C. J. R. Sheppard and C. J. Cogswell. Three-Dimensional Image Formation in Confocal Microscopy. *Journal of of Microscopy*, 159:179–194, Aug. 1990.

[40] G. M. P. van Kempen. *Image Restoration in Fluorescence Microscopy*. PhD thesis, Technische Universiteit Delft - Holland, Jan. 1999.

[41] G. M. P. van Kempen and L. J. Van Vliet. The influence of the regularization parameter and the first estimate on the performance of tikhonov regularized non-linear image restoration algorithms. *Journal of Microscopy*, 13:63–75, Apr. 2000.

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

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

[44] M.N. Do and M. Vetterli. Framing pyramids. *IEEE Transactions on Signal Processing*, 51(9):2329–2342, 2003.