HAL
open science

A proximal iteration for deconvolving Poisson noisy images using sparse representations

François-Xavier Dupé, Jalal M. Fadili, Jean-Luc Starck

To cite this version:

François-Xavier Dupé, Jalal M. Fadili, Jean-Luc Starck. A proximal iteration for deconvolving Poisson noisy images using sparse representations. IEEE Transactions on Image Processing, 2009, 18 (2), pp.310-321. 10.1109/TIP.2008.2008223 . hal-00264972v2

HAL Id: hal-00264972

https://hal.science/hal-00264972v2

Submitted on 27 Aug 2008

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.

A proximal iteration for deconvolving Poisson noisy images using sparse representations

F.-X. Dupé[a], M.J. Fadili[a] and J.-L. Starck[b]

[a] GREYC UMR CNRS 6072 [b] DAPNIA/SEDI-SAP CEA-Saclay

14050 Caen France 91191 Gif-sur-Yvette France

Abstract

We propose an image deconvolution algorithm when the data is contaminated by Poisson noise. The image to restore is assumed to be sparsely represented in a dictionary of waveforms such as the wavelet or curvelet transforms. Our key contributions are: First, we handle the Poisson noise properly by using the Anscombe variance stabilizing transform leading to a *non-linear* degradation equation with additive Gaussian noise. Second, the deconvolution problem is formulated as the minimization of a convex functional with a data-fidelity term reflecting the noise properties, and a non-smooth sparsity-promoting penalty over the image representation coefficients (e.g. ℓ_1-norm). An additional term is also included in the functional to ensure positivity of the restored image. Third, a fast iterative forward-backward splitting algorithm is proposed to solve the minimization problem. We derive existence and uniqueness conditions of the solution, and establish convergence of the iterative algorithm. Finally, a GCV-based model selection procedure is proposed to objectively select the regularization parameter. Experimental results are carried out to show the striking benefits gained from taking into account the Poisson statistics of the noise. These results also suggest that using sparse-domain regularization may be tractable in many deconvolution applications with Poisson noise such as astronomy and microscopy.

Index Terms

Deconvolution, Poisson noise, Proximal iteration, forward-backward splitting, Iterative thresholding, Sparse representations.

I. Introduction

Deconvolution is a longstanding problem in many areas of signal and image processing (e.g. biomedical imaging [1], [2], astronomy [3], remote-sensing, to quote a few). For example, research in astronomical image deconvolution has recently seen considerable work, partly triggered by the Hubble space telescope (HST) optical aberration problem at the beginning of its mission. In biomedical imaging, researchers are also increasingly relying on deconvolution to improve the quality of images acquired by confocal microscopes [2]. Deconvolution may then prove crucial for exploiting images and extracting scientific content.

There is an extensive literature on deconvolution problems. One might refer to well-known dedicated monographs on the subject [4]–[6]. In presence of Poisson noise, several deconvolution methods have been proposed such as Tikhonov-Miller inverse filter and Richardson-Lucy (RL) algorithms; see [1], [3] for a comprehensive review. The RL has been used extensively in many applications because it is adapted to Poisson noise. The RL algorithm, however, amplifies noise after a few iterations, which can be avoided by introducing regularization. In [7], the authors presented a Total Variation (TV)-regularized RL algorithm. In the astronomical imaging literature, several authors advocated the use of wavelet-regularized RL algorithm [8]–[10]. In the context of biological imaging deconvolution, wavelets have also been used as a regularization scheme when deconvolving biomedical images; [11] presents a version of RL combined with wavelets denoising, and [12] uses the thresholded Landweber iteration introduced in [13]. The latter approach implicitly assumes that the contaminating noise is Gaussian.

Other recent attempts for solving Poisson linear inverse problems is a Bayesian multi-scale framework proposed in [14] based on a multi-scale factorization of the Poisson likelihood function associated with a recursive partitioning of the underlying intensity. Regularization of the solution is accomplished by imposing prior probability distributions in a Bayesian paradigm and the maximum a posteriori solution is computed using the expectation-maximization algorithm. However, the multiscale analysis by the above authors is only tractable with the Haar wavelet. Similarly, the work in [15] on hard threshold estimators in the tomographic data framework has shown that for a particular operator (the Radon operator) an extension of wavelet-vaguelette decomposition (WVD) method [16] for Poisson data is theoretically feasible. But the authors do not provide any computational algorithm for computing the estimate. Inspired by the WVD method, the authors in [17] explored an alternative approach via wavelet-based decompositions combined with thresholding strategies that address adaptivity issues. Specifically, their framework extends the wavelet-Galerkin methods of [18] to the Poisson setting. In order to ensure the positivity of the estimated intensity, the log-intensity is expanded in a wavelet basis. This method is however limited to standard orthogonal wavelet bases.

In the context of deconvolution with Gaussian white noise, sparsity-promoting regularization over orthogonal wavelet coefficients has been recently proposed [13], [19], [20]. Generalization to frames was proposed in [21], [22]. In [23], the authors presented an image deconvolution algorithm by iterative thresholding in an overcomplete dictionary of transforms, and [24] designed a deconvolution method that combines both the wavelet and curvelet

transforms. However, sparsity-based approaches published so far have mainly focused on Gaussian noise.

In this paper, we propose an image deconvolution algorithm for data blurred and contaminated by Poisson noise. The Poisson noise is handled properly by using the Anscombe variance stabilizing transform (VST), leading to a *non-linear* degradation equation with additive Gaussian noise, see (1). The deconvolution problem is then formulated as the minimization of a convex functional combining a non-linear data-fidelity term reflecting the noise properties, and a non-smooth sparsity-promoting penalty over the representation coefficients of the image to restore. Such representations include not only the orthogonal wavelet transform but also overcomplete representations such as translation-invariant wavelets, curvelets or wavelets and curvelets. Since Poisson intensity functions are nonnegative by definition, an additional term is also included in the minimized functional to ensure the positivity of the restored image. Inspired by the work in [20], a fast proximal iterative algorithm is proposed to solve the minimization problem. Experimental results are carried out on a set of simulated and real images to compare our approach to some competitors. We show the striking benefits gained from taking into account the Poisson nature of the noise and the morphological structures involved in the image through overcomplete sparse multiscale transforms.

A. Relation to prior work

A naive solution to this deconvolution problem would be to apply traditional approaches designed for Gaussian noise. But this would be awkward as (i) the noise tends to Gaussian only for large mean intensities (central limit theorem), and (ii) the noise variance depends on the mean anyway. A more adapted way would be to adopt a bayesian framework with an appropriate anti-log-likelihood score—the negative of the log-likelihood function—to obtain a data fidelity term reflecting the Poisson statistics of the noise. The data fidelity term is derived from the conditional distribution of the observed data given the original image, which is known to be governed by physical considerations concerned with the data-acquisition device and the noise generating process (e.g. Poisson here). Unfortunately, doing so, we would end-up with a functional which does not satisfy a key property: the data fidelity term does not have a Lipschitz-continuous gradient as required in [20], hence preventing us from using the forward-backward splitting proximal algorithm to solve the optimization problem. To circumvent this difficulty, we propose to handle the noise statistical properties by using the Anscombe VST. Some previous authors [25] have already suggested to use the Anscombe VST, and then deconvolve with wavelet-domain regularization as if the stabilized observation were linearly degraded and contaminated by additive Gaussian noise. But this is not valid as standard results of the Anscombe VST lead to a non-linear degradation equation because of the square-root, see (1).

B. Organization of this paper

The organization of the paper is as follows: we first formulate our deconvolution problem under Poisson noise (Section II), and then recall some necessary material about overcomplete sparse representations (Section III). The core of the paper lies in Section IV, where we state the deconvolution optimization problem, characterize it and

solve it using monotone operator splitting iterations. We also focus on the choice of the two main parameters of the algorithm and propose some solutions. In Section V, experimental results are reported and discussed. The proofs of our main results are deferred to the appendix for the sake of presentation.

C. Notation and terminology

Let $\mathcal{H}$ a real Hilbert space, here a finite dimensional real vector space. We denote by $\|.\|$ the norm associated with the inner product $\langle .,. \rangle$ in $\mathcal{H}$, and I is the identity operator on $\mathcal{H}$. x and α are respectively reordered vectors of image samples and transform coefficients.

A real-valued function f is coercive, if $\lim_{\|x\| \to +\infty} f(x) = +\infty$. The domain of f is defined by $\operatorname{dom} f = \{x \in \mathcal{H} \ : \ f(x) < +\infty\}$ and f is proper if $\operatorname{dom} f \neq \emptyset$. We say that a real-valued function f is lower semi-continuous (lsc) if $\liminf_{x \to x_0} f(x) \geq f(x_0)$. Lower semi-continuity is weaker than continuity, and plays an important role for existence of solutions in minimization problems [26, Page 17]. $\Gamma_0(\mathcal{H})$ is the class of all proper lsc convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$. The subdifferential of a function $f \in \Gamma_0(\mathcal{H})$ at $x \in \mathcal{H}$ is the set $\partial f(x) = \{u \in \mathcal{H} | \forall y \in \mathcal{H}, f(y) \geq f(x) + \langle u, y - x \rangle\}$. An element u of ∂f is called a subgradient. A comprehensive account of subdifferentials can be found in [26].

An operator A acting on $\mathcal{H}$ is κ-Lipschitz continuous if $\forall x, y \in \mathcal{H}, \|\mathrm{A}(x) - \mathrm{A}(y)\| \leqslant \kappa \|x - y\|$ where κ is the Lipschitz constant. The spectral operator norm of A is given by $\|\mathrm{A}\|_2 = \max_{x \neq 0} \frac{\|\mathrm{A}x\|}{\|x\|}$.

We denote by $\imath_{\mathcal{C}}$ the indicator of the convex set $\mathcal{C}$: $\imath_{\mathcal{C}}(x) = \begin{cases} 0, & \text{if } x \in \mathcal{C} \\ +\infty, & \text{otherwise.} \end{cases}$. We denote by $\rightharpoonup$ the convergence.

II. Problem statement

Consider the image formation model where an input image of n pixels x is blurred by a point spread function (PSF) h and contaminated by Poisson noise. The observed image is then a discrete collection of counts $y = (y_i)_{1 \leqslant i \leqslant n}$ which are bounded, i.e. $y \in \ell_\infty$. Each count y_i is a realization of an independent Poisson random variable with a mean $(h \circledast x)_i$, where $\circledast$ is the circular convolution operator. Formally, this writes $y_i \sim \mathcal{P}\left((h \circledast x)_i\right)$. The deconvolution problem at hand is to restore x from the observed count image y.

A natural way to attack this problem would be to adopt a maximum a posteriori (MAP) bayesian framework with an appropriate likelihood function—the distribution of the observed data y given an original x—reflecting the Poisson statistics of the noise. But, as stated above, this would prevent us from using the forward-backward splitting proximal algorithm to solve the MAP optimization problem, since the gradient of the data fidelity term is not Lipschitz-continuous. Indeed, forward-backward iteration is essentially a generalization of the classical gradient projection method [27] for constrained convex optimization and monotone variational inequalities, and inherit restrictions similar to those methods. For such methods, Lipschitz continuity of the gradient is classical [27, Theorem 8.6-2]. The latter property is then crucial for the iterates in (13) to be determined uniquely, and for the forward-backward

splitting algorithm to converge; see Theorem 1 and also [28]. For this reason, we propose to handle the noise statistical properties by using the Anscombe VST [29] defined as

$$z_i = 2\sqrt{(h \circledast x)_i + \tfrac{3}{8}} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0,1), \tag{1}$$

where ε is an additive white Gaussian noise of unit variance[1]. In words, z is *non-linearly* related to x. In Section IV, we provide an elegant optimization problem and a fixed point algorithm taking into account such a non-linearity.

III. Sparse image representation

Let $x \in \mathcal{H}$ be an $\sqrt{n} \times \sqrt{n}$ image. x can be written as the superposition of elementary atoms φ_γ parameterized by $\gamma \in \mathcal{I}$ according to the following linear generative model :

$$x = \sum_{\gamma \in \mathcal{I}} \alpha_\gamma \varphi_\gamma = \Phi\alpha, \quad |\mathcal{I}| = L \geqslant n \ . \tag{2}$$

We denote by Φ the dictionary i.e. the $n \times L$ matrix whose columns are the generating waveforms $(\varphi_\gamma)_{\gamma \in \mathcal{I}}$ all normalized to a unit ℓ_2-norm. The forward (analysis) transform is then defined by a non-necessarily square matrix $\mathbf{T} = \Phi^{\mathrm{T}} \in \mathbb{R}^{L \times n}$ with $L \geqslant n$. When $L > n$ the dictionary is said to be redundant or overcomplete. In the case of the simple orthogonal basis, the inverse (synthesis) transform is trivially $\Phi = \mathbf{T}^{\mathrm{T}}$. Whereas assuming that Φ is a tight frame implies that the frame operator satisfies $\Phi\Phi^{\mathrm{T}} = c\mathrm{I}$, where $c > 0$ is the tight frame constant. For tight frames, the pseudo-inverse reconstruction (synthesis) operator reduces to $c^{-1}\Phi$. In the sequel, the dictionary Φ will correspond either to an orthobasis or to a tight frame of $\mathcal{H}$.

Owing to recent advances in modern harmonic analysis, many redundant systems, like the undecimated wavelet transform, curvelet, contourlet etc, were shown to be very effective in sparsely representing images. By sparsity, we mean that we are seeking for a good representation of x with only few significant coefficients.

In the rest of the paper, the dictionary Φ is built by taking union of one or several transforms, each corresponding to an orthogonal basis or a tight frame. Choosing an appropriate dictionary is a key step towards a good sparse representation, hence restoration. A core idea here is the concept of morphological diversity. When the transforms are amalgamated in the dictionary, they have to be chosen such that each leads to sparse representation over the parts of the image it is serving while being inefficient in representing the other image content. As popular examples, one may think of wavelets for smooth images with isotropic singularities [30, Section 9.3], curvelets for representing piecewise smooth C^2 images away from C^2 contours [31], [32], wave atoms or local DCT to represent locally oscillating textures [30], [33].

[1]Rigorously speaking, the equation is to be understood in an asymptotic sense.

IV. Sparse Iterative Deconvolution

A. Optimization problem

In this Section, we derive that the class of minimization problems we are interested in, see (5), can be stated in the general form :

$$\min_{\alpha \in \mathbb{R}^L} f_1(\alpha) + f_2(\alpha), \tag{3}$$

where $f_1 \in \Gamma_0(\mathbb{R}^L)$, $f_2 \in \Gamma_0(\mathbb{R}^L)$ and f_1 is differentiable with a κ-Lipschitz gradient. We denote by $\mathcal{M}$ the set of solutions of (3).

From (1), we immediately deduce the data fidelity term

$$F \circ \mathrm{H} \circ \Phi\ (\alpha), \text{ with} \tag{4}$$
$$F : \eta \in \mathbb{R}^n \mapsto \sum_{i=1}^{n} f(\eta_i), \quad f(\eta_i) = \frac{1}{2}\left(z_i - 2\sqrt{\eta_i + \tfrac{3}{8}}\right)^2,$$

where H denotes the (circular) convolution operator. From a statistical perspective, (4) corresponds to the anti-log-likelihood score. Note that for bias correction reasons [34], the value 1/8 may be used instead of 3/8 in (4). However, there are implications of this alternate choice on the Lipschitz constant in (8), and consequently it can be seen from Theorem 1 that this will have an unfavorable impact on the convergence speed of the deconvolution algorithm.

Adopting a bayesian framework and using a standard MAP rule, our goal is to minimize the following functional with respect to the representation coefficients α :

$$(\mathsf{P}_{\lambda,\psi}) : \min_{\alpha} J(\alpha)\ , \tag{5}$$
$$J : \alpha \mapsto \underbrace{F \circ \mathrm{H} \circ \Phi\ (\alpha)}_{f_1(\alpha)} + \underbrace{\imath_{\mathcal{C}} \circ \Phi\ (\alpha) + \lambda \sum_{i=1}^{L} \psi(\alpha_i)}_{f_2(\alpha)},$$

where we implicitly assumed that $(\alpha_i)_{1 \leqslant i \leqslant L}$ are independent and identically distributed with a Gibbsian density $\propto e^{-\lambda\psi(\alpha_i)}$. The penalty function ψ is chosen to enforce sparsity, $\lambda > 0$ is a regularization parameter and $\imath_{\mathcal{C}}$ is the indicator function of the convex set $\mathcal{C}$. In our case, $\mathcal{C}$ is the positive orthant. The role of the term $\imath_{\mathcal{C}} \circ \Phi$ is to impose the positivity constraint on the restored image because we are fitting Poisson intensities, which are positive by nature. We also define the set $\mathcal{C}' = \{\alpha | \Phi\alpha \in \mathcal{C}\}$, that is $\imath_{\mathcal{C}'} = \imath_{\mathcal{C}} \circ \Phi$.

From (5), we have the following,

Proposition 1.

(i) *f_1 is convex function. It is strictly convex if Φ is an orthobasis and $\ker(\mathrm{H}) = \emptyset$ (i.e. the spectrum of the PSF does not vanish within the Nyquist band).*

(ii) *The gradient of f_1 is*

$$\nabla f_1(\alpha) = \Phi^{\mathrm{T}} \circ \mathrm{H}^{\mathrm{T}} \circ \nabla F \circ \mathrm{H} \circ \Phi\ (\alpha)\ , \tag{6}$$

with

$$\nabla F(\eta) = \left(\frac{-z_i}{\sqrt{\eta_i + 3/8}} + 2 \right)_{1 \leqslant i \leqslant n} . \tag{7}$$

(iii) f_1 *is continuously differentiable with a* κ*-Lipschitz gradient where*

$$\kappa \leqslant \left(\tfrac{2}{3}\right)^{3/2} 4c \, \|\mathrm{H}\|_2^2 \, \|z\|_\infty < +\infty. \tag{8}$$

(iv) $(\mathsf{P}_{\lambda,\psi})$ *is a particular case of problem* (3).

A proof can be found in the appendix.

B. Characterization of the solution

Since J is coercive and convex, the following holds :

Proposition 2.

1) *Existence:* $(\mathsf{P}_{\lambda,\psi})$ *has at least one solution, i.e.* $\mathcal{M} \neq \emptyset$.
2) *Uniqueness:* $(\mathsf{P}_{\lambda,\psi})$ *has a unique solution if* Φ *is an orthobasis and* $\ker(\mathrm{H}) = \emptyset$*, or if* ψ *is strictly convex.*

C. Proximal iteration

We first define the notion of a proximity operator, which was introduced in [35] as a generalization of the notion of a convex projection operator.

Definition 1 (Moreau [35]). *Let* $\varphi \in \Gamma_0(\mathcal{H})$*. Then, for every* $x \in \mathcal{H}$*, the function* $y \mapsto \varphi(y) + \|x - y\|^2 / 2$ *achieves its infimum at a unique point denoted by* $\operatorname{prox}_\varphi x$*. The operator* $\operatorname{prox}_\varphi : \mathcal{H} \to \mathcal{H}$ *thus defined is the proximity operator of* φ*. Moreover,* $\forall x, p \in \mathcal{H}$

$$p = \operatorname{prox}_\varphi x \iff x - p \in \partial\varphi(p) \iff \langle y - p, x - p \rangle + \varphi(p) \leqslant \varphi(y) \ \forall y \in \mathcal{H}. \tag{9}$$

(9) *means that* $\operatorname{prox}_\varphi$ *is the resolvent of the subdifferential of* φ *[36]. Recall that the resolvent of the subdifferential* $\partial\varphi$ *is the single-valued operator* $J_{\partial\varphi} : \mathcal{H} \to \mathcal{H}$ *such that* $\forall x \in \mathcal{H}, x - J_{\partial\varphi}(x) \in \partial\varphi(J_{\partial\varphi}) \iff J_{\partial\varphi} = (\mathrm{I} - \partial\varphi)^{-1}$.

It will also be convenient to introduce the reflection operator $\operatorname{rprox}_\varphi = 2 \operatorname{prox}_\varphi - \mathrm{I}$.

For notational simplicity, we denote by Ψ the function $\alpha \mapsto \sum_i \psi(\alpha_i)$. Our goal now is to express the proximity operator associated to f_2, which will be needed in the iterative deconvolution algorithm. The difficulty stems from the definition of f_2 which combines both the 'positivity' constraint and the regularization. Unfortunately, we can show that even with a separable penalty $\Psi(\alpha)$, the operator $\operatorname{prox}_{f_2} = \operatorname{prox}_{\iota_C \circ \Phi + \lambda\Psi}$ has no explicit form in general, except the case where $\Phi = \mathrm{I}$. We then propose to replace explicit evaluation of $\operatorname{prox}_{f_2}$ by a sequence of calculations that activate separately $\operatorname{prox}_{\iota_C \circ \Phi}$ and $\operatorname{prox}_{\lambda\Psi}$. We will show that the last two proximity operators have closed-form expressions. Such a strategy is known as a splitting method of maximal monotone operators [36], [37]. As both

$\imath_{\mathcal{C}'}$ and Ψ belong to $\Gamma_0(\mathcal{H})$ and are non-differentiable, our splitting method is based on the Douglas-Rachford algorithm [28], [36], [37]. The following lemma summarizes our scheme.

Lemma 1. *Let Φ an orthobasis or a tight frame with constant c. Recall that $\mathcal{C}' = \{\alpha | \Phi\alpha \in \mathcal{C}\}$.*

1) *If $\alpha \in \mathcal{C}'$ then $\operatorname{prox}_{f_2}(\alpha) = \operatorname{prox}_{\lambda\Psi}(\alpha)$.*
2) *Otherwise, let $(\nu_t)_t$ be a sequence in $(0,1)$ such that $\sum_t \nu_t(1-\nu_t) = +\infty$. Take $\gamma^0 \in \mathcal{H}$, and define the sequence of iterates :*

$$\gamma^{t+1} = \gamma^t + \nu_t \left(\operatorname{rprox}_{\lambda\Psi + \frac{1}{2}\|\cdot - \alpha\|^2} \circ \operatorname{rprox}_{\imath_{\mathcal{C}'}} - \mathrm{I} \right)(\gamma^t), \tag{10}$$

where $\operatorname{prox}_{\lambda\Psi+\frac{1}{2}\|\cdot-\alpha\|^2}(\gamma^t) = \left(\operatorname{prox}_{\frac{\lambda}{2}\psi}\left((\alpha_i + \gamma_i^t)/2\right)\right)_{1 \leqslant i \leqslant L}$, $\mathcal{P}_{\mathcal{C}'} = \operatorname{prox}_{\imath_{\mathcal{C}'}} = c^{-1}\Phi^{\mathrm{T}} \circ \mathcal{P}_{\mathcal{C}} \circ \Phi + \left(\mathrm{I} - c^{-1}\Phi^{\mathrm{T}} \circ \Phi\right)$ *and $\mathcal{P}_{\mathcal{C}}$ is the projector onto the positive orthant $(\mathcal{P}_{\mathcal{C}}\eta)_i = \max(\eta_i, 0)$. Then,*

$$\gamma^t \rightharpoonup \gamma \text{ and } \operatorname{prox}_{f_2}(\alpha) = \mathcal{P}_{\mathcal{C}'}(\gamma). \tag{11}$$

The proof is detailed in the appendix. Note that when Φ is an orthobasis, $\mathcal{P}_{\mathcal{C}'} = \Phi^{\mathrm{T}} \circ \mathcal{P}_{\mathcal{C}} \circ \Phi$.

To implement the above iteration, we need to express $\operatorname{prox}_{\lambda\psi}$, which is given by the following result for a wide class of penalties ψ :

Lemma 2. *Suppose that ψ satisfies, (i) ψ is convex even-symmetric , non-negative and non-decreasing on $[0, +\infty)$, and $\psi(0) = 0$. (ii) ψ is twice differentiable on $\mathbb{R} \setminus \{0\}$. (iii) ψ is continuous on $\mathbb{R}$, it is not necessarily smooth at zero and admits a positive right derivative at zero $\psi'_+(0) = \lim_{h \to 0^+} \frac{\psi(h)}{h} > 0$. Then, the proximity operator of $\delta\Psi(\gamma)$, $\operatorname{prox}_{\delta\Psi}(\gamma)$ has exactly one continuous solution decoupled in each coordinate γ_i :*

$$\operatorname{prox}_{\delta\psi}(\gamma_i) = \begin{cases} 0 & \text{if } |\gamma_i| \leqslant \delta\psi'_+(0) \\ \gamma_i - \delta\psi'(\bar{\alpha}_i) & \text{if } |\gamma_i| > \delta\psi'_+(0) \end{cases} \tag{12}$$

A proof of this lemma can be found in [38]. A similar result also has recently appeared in [39]. Among the most popular penalty functions ψ satisfying the above requirements, we have $\psi(\alpha_i) = |\alpha_i|$, in which case the associated proximity operator is the popular soft-thresholding.

We are now ready to state our main proximal iterative algorithm to solve the minimization problem $(\mathsf{P}_{\lambda,\psi})$:

Theorem 1. *For $t \geq 0$, let $(\mu_t)_t$ be a sequence in $(0, +\infty)$ such that $0 < \inf_t \mu_t \leqslant \sup_t \mu_t < \left(\frac{3}{2}\right)^{3/2} / \left(2c\,\|\mathrm{H}\|_2^2\,\|z\|_\infty\right)$, let $(\beta_t)_t$ be a sequence in $(0,1]$ such that $\inf_t \beta_t > 0$, and let $(a_t)_t$ and $(b_t)_t$ be sequences in $\mathcal{H}$ such that $\sum_t \|a_t\| < +\infty$ and $\sum_t \|b_t\| < +\infty$. Fix $\alpha^0 \in \mathbb{R}^L$, for every $t \geqslant 0$, set*

$$\alpha^{t+1} = \alpha^t + \beta_t(\operatorname{prox}_{\mu_t f_2}\left(\alpha^t - \mu_t\left(\nabla f_1(\alpha^t) + b_t\right)\right) + a_t - \alpha^t) \tag{13}$$

where ∇f_1 and $\operatorname{prox}_{\mu_t f_2}$ are given by Proposition 1(ii) and Lemma 1. Then $(\alpha_t)_{t \geq 0}$ converges to a solution of $(\mathsf{P}_{\lambda,\psi})$.

This is the most general convergence result known on the forward-backward iteration. The name of the iteration is inspired by well-established techniques from numerical linear algebra. The words ”forward” and ”backward”

refer respectively to the standard notions of a forward difference (here explicit gradient descent) step and of a backward difference (here implicit proximity) step in numerical analysis. The sequences a_t and b_t play a prominent role as they formally establish the robustness of the algorithm to numerical errors when computing the gradient ∇f_1 and the proximity operator prox_{f_2}. The latter remark will allow us to accelerate the algorithm by running the sub-iteration (10) only a few iterations (see implementation details in IV-F).

For illustration, let's take Ψ as the ℓ_1 norm, in which case $\mathrm{prox}_{\lambda\Psi}$ is the component-wise soft-thresholding with threshold λ, $a_t = b_t \equiv 0$, $\beta_t \equiv 1$ and $\mu_t \equiv \mu$ in (13), and $\nu_t \equiv 1/2$ in (10). Thus, bringing all the pieces together, the deconvolution algorithm dictated by iterations (13) and (10) is summarized in Algorithm 1.

Algorithm 1

Task: Image deconvolution with Poisson noise, solve (5).

Parameters: The observed image counts y, the dictionary Φ, number of iterations N_{FB} in (13) and N_{DR} in sub-iteration (10), relaxation parameter μ, regularization parameter λ.

Initialization:

- Apply VST $z = 2\sqrt{y + 3/8}$.
- Initial solution $\alpha^0 = 0$.

Main iteration:

For $t = 0$ **to** $N_{\mathrm{FB}} - 1$,

- Compute blurred estimate $\eta^t = \mathrm{H}\Phi\alpha^t$.
- Compute 'residuals' $r^t = \left(\frac{-z_i}{\sqrt{\eta_i^t + 3/8}} + 2\right)_{1 \leqslant i \leqslant n}$.
- Move along the descent direction $\xi^t = \alpha^t + \mu\Phi^{\mathrm{T}}\mathrm{H}^{\mathrm{T}} r^t$.
- Initialize $\gamma^0 = \xi^t$, and start sub-iteration.
- **F**or $m = 0$ **to** N_{DR}-1,
  - Project γ^m orthogonally to $\mathcal{C}'$: $\zeta^m = c^{-1}\Phi^{\mathrm{T}}\left(\min(\Phi\gamma^m, 0)\right)$.
  - Update γ^{m+1} by soft-thresholding $\gamma^{m+1} = \mathrm{ST}_{\lambda/2}\left((\xi^t + \gamma^m)/2 - \zeta^m\right) + \zeta^m$.
- Update $\alpha^{t+1} = \gamma^{N_{\mathrm{DR}}} - c^{-1}\Phi^{\mathrm{T}}\left(\min(\Phi\gamma^{N_{\mathrm{DR}}}, 0)\right)$.

End main iteration

Output: Deconvolved image $x^\star = \Phi\alpha^{N_{\mathrm{FB}}}$.

D. Choice of μ

The relaxation (or descent) parameter μ has an important impact on the convergence speed of the algorithm. The upper-bound provided by Theorem 1, derived from the Lipschitz constant (8) is only a sufficient condition for (13) to converge, and may be pessimistic in some applications. To circumvent this drawback, Tseng proposed in [40] an extension of the forward-backward algorithm with an iteration to adaptively estimate a "good" value of μ. The

main result provided hereafter is an adaptation to our context to the one of Tseng [40]. We state it in full for the sake of completeness and the reader convenience.

Theorem 2. *Let $\mathcal{C}'$ as defined above (IV-A). Choose any $\alpha_0 \in \mathcal{C}'$. Let $(\mu_t)_{t\in\mathbb{N}}$ be a sequence such that $\forall t > 0, \mu_t \in (0,\infty)$. Let f_1 as defined in* (5)*. Then the sequence $(\alpha_t)_{t\in\mathbb{N}}$ of iterates*

$$\begin{aligned} \alpha_{t+\frac{1}{2}} &= \mathrm{prox}_{\lambda\Psi}\left(\alpha_t - \mu_t \nabla f_1(\alpha_t)\right) \ , \\ \alpha_{t+1} &= \mathcal{P}_{\mathcal{C}'}\left(\alpha_{t+\frac{1}{2}} - \mu_t\left(\nabla f_1(\alpha_{t+\frac{1}{2}}) - \nabla f_1(\alpha_t)\right)\right) \end{aligned} \tag{14}$$

converges to a minimum of J.

As ∇f_1 is Lipschitz-continuous, the update of the relaxation sequence μ_t is rather easy. Indeed, using an Armijo-Goldstein-type stepsize approach, we can compute and update μ_t at each iteration by taking μ_t to be the largest $\mu \in \{\sigma, \tau\sigma, \tau^2\sigma, \ldots\}$ satisfying

$$\mu \left\| \nabla f_1(\alpha_{t+\frac{1}{2}}) - \nabla f_1(\alpha_t) \right\| \leqslant \theta \left\| \alpha_{t+\frac{1}{2}} - \alpha_t \right\| \ , \tag{15}$$

where $\tau \in (0,1)$, $\theta \in (0,1)$ and $\sigma > 0$ are constants. $\tau = 1/2$ is a typical choice.

It is worth noting that for tight frames, this algorithm will somewhat simplify the computation of prox_{f_2}, removing the need of the Douglas-Rachford sub-iteration (10). But, whatever the transform, this will come at the price of keeping track of the gradient of f_1 at the points $\alpha_{t+\frac{1}{2}}$ and α_t, and the need to check (15) several times.

E. Choice of λ

As usual in regularized inverse problems, the choice of λ is crucial as it represents the desired balance between sparsity (regularization) and deconvolution (data fidelity). For a given application and corpus of images (e.g. confocal microscopy), a naive brute-force approach would consist in testing several values of λ and taking the best one by visual assessment of the deconvolution quality. However, this is cumbersome in the general case.

We propose to objectively select the regularizing parameter λ based on an adaptive model selection criterion such as the generalized cross validation (GCV) [41]. Other criteria are possible as well including the AIC [42] or the BIC [43]. GCV attempts to provide a data-driven estimate of λ by minimizing :

$$\mathrm{GCV}(\lambda) = \frac{\left\| z - 2\sqrt{\mathrm{H}\Phi\alpha^\star + \frac{3}{8}} \right\|^2}{(n - df)^2}, \tag{16}$$

where $\alpha^\star(z)$ denotes the solution arrived at by iteration (13) (or (14)), and df is the effective number of degrees of freedom.

Deriving the closed-form expression of df is very challenging in our case as it faces two main difficulties, (i) the observation model (1) is non-linear, and (ii) the solution $\alpha^\star(z)$ is not known in closed form but given by the iterative forward-backward algorithm.

Degrees of freedom is a familiar phrase in statistics. In (overdetermined) linear regression df is the number of estimated predictors. More generally, degrees of freedom is often used to quantify the model complexity of a

statistical modeling procedure. However, generally speaking, there is no exact correspondence between the degrees of freedom df and the number of parameters in the model. In penalized solutions of inverse problems where the estimator is linear in the observation, e.g. Tikhonov regularization or ridge regression in statistics, df is simply the trace of the so-called influence or the hat matrix. But in general, it is difficult to derive the analytical expression of df for general nonlinear modeling procedures such as ours. This remains a challenging and active area of research.

Stein's unbiased risk estimation (SURE) theory [44] gives a rigorous definition of the degrees of freedom for any fitting procedure. Following our notation, given the solution $\alpha^\star$ provided by our deconvolution algorithm, let $z^\star(z) = 2\sqrt{\mathrm{H}\Phi\alpha^\star(z) + 3/8}$ represent the model fit from the observation z. As $Z|\alpha \sim \mathcal{N}\left(2\sqrt{\mathrm{H}\Phi\alpha + 3/8}, 1\right)$, it follows from [45] that the degrees of freedom of our procedure is

$$df(\lambda) = \sum_{i=1}^{n} \mathrm{Cov}(z_i^\star(z), z_i) \ ,$$

a quantity also called the optimism of the estimator $z^\star(z)$. If the estimation algorithm is such that $z^\star(z)$ is almost-differentiable [44] with respect to z, so that its divergence is well-defined in the weak sense (as is the case if $z^\star(z)$ were uniformly Lipschitz-continuous), Stein's Lemma yields the so-called divergence formula

$$df(\lambda) = \sum_{i=1}^{n} \mathrm{Cov}(z_i^\star(z), z_i) = \mathrm{E}_Z\left[\mathrm{div}\left(z^\star(z)\right)\right] = \mathrm{E}_Z\left[\sum_{i=1}^{n} \frac{\partial z_i^\star(z)}{\partial z_i}\right] \ . \tag{17}$$

where the expectation E_Z is taken under the distribution of Z. The df is then the sum of the sensitivity of each fitted value with respect to the corresponding observed value. For example, the last expression of this formula has been used in [46] for orthogonal wavelet denoising. However, it is notoriously difficult to derive the closed-form analytical expression of df from the above formula for general nonlinear modeling procedures. To overcome the analytical difficulty, the bootstrap [47] can be used to obtain an (asymptotically) unbiased estimator of df. Ye [48] and Shen and Ye [49] proposed using a data perturbation technique to numerically compute an (approximately) unbiased estimate for df when the analytical form of $z^\star(z)$ is unavailable. From (17), the estimator of df takes the form

$$\begin{aligned} \widehat{df}(\lambda) &= \mathrm{E}_{V_0}\left[\left\langle v_0, \frac{z^\star(z+\tau v_0) - z^\star(z)}{\tau}\right\rangle\right], \ V_0 \sim \mathcal{N}(0, \mathrm{I}) \ , \\ &= \frac{1}{\tau^2}\int \langle v, z^\star(z+v)\rangle \, \phi(v; \tau^2\mathrm{I}) dv, \ V \sim \mathcal{N}(0, \tau^2\mathrm{I}) \ , \end{aligned} \tag{18}$$

where $\phi(v; \tau^2\mathrm{I})$ is the n-dimensional density of $\mathcal{N}(0, \tau^2\mathrm{I})$. It can be shown that this formula is valid if V is replaced by any vector of random of variables with finite higher order moments. The author in [48] proved that this is an unbiased estimate of df as $\tau \to 0$, that is $\lim_{\tau\to 0} \mathrm{E}_Z\left[\widehat{df}(\lambda)\right] = df(\lambda)$. It can be computed by Monte-Carlo integration with τ near 0.6 as devised in [48]. However, both bootstrap and Ye's method, although general and can be used for any $\Psi \in \Gamma_0(\mathbb{R}^L)$, are computationally prohibitive. This is the main reason we will not use them here.

Zou et al. [50] recently studied the degrees of freedom of the Lasso[2] in the framework of SURE. They showed that for any given λ the number of nonzero coefficients in the model is an unbiased and consistent estimate of df.

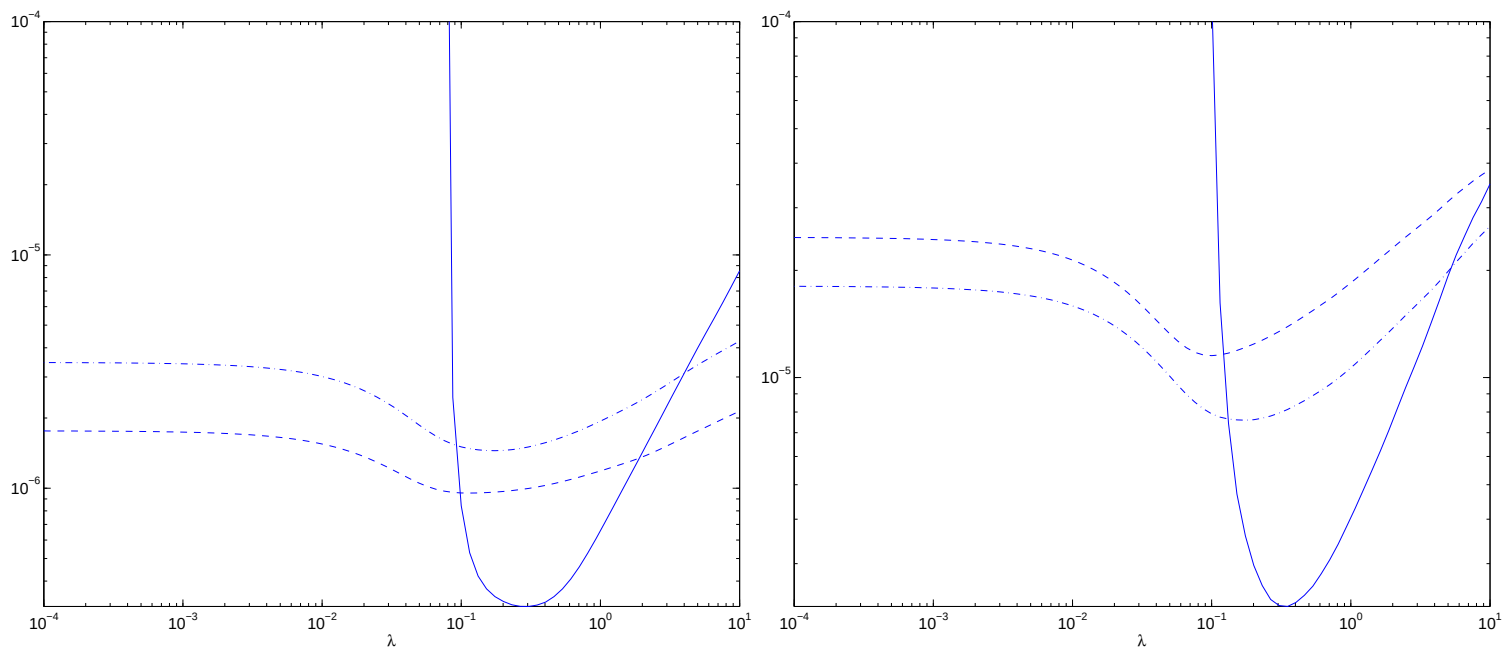
[2]The Lasso model correspond to the case of (5) where the degradation model in (1) is linear and Ψ is the ℓ_1-norm.

However, for their results to hold rigorously, the matrix $\mathrm{A} = \mathrm{H}\Phi$ in the Lasso must be over-determined $L < n$ with $\mathrm{Rank}(\mathrm{A}) = L$. Nonetheless, one can show that their intuitive estimator can be extended to the under-determined case (i.e. $L \geq n$) under the so-called (UC) condition of [51]; see Theorem 2 in that reference. This will yield an unbiased estimator of df, but consistency would be much harder to prove since it requires that the Gram matrix $\mathrm{A}^{\mathrm{T}}\mathrm{A}$ is *positive-definite* which only happens in the special case of Φ an orthogonal basis and $\ker(\mathrm{H}) = \emptyset$. Furthermore, even with the ℓ_1 norm, extending this simple estimator rigorously to our setting faces two additional serious difficulties beside underdeterminacy of A: namely the non-linearity of the degradation equation (1) and the positivity constraint in (5).

Following this discussion, it appears clearly that estimating df is either computationally intensive (bootstrap or perturbation techniques), or analytically difficult to derive. In this paper, in the same vein as [50], we conjecture that a simple estimator of df, is given by the cardinal of the support of $\alpha^\star$. That is, from (12)-(13)

$$\widehat{df}(\lambda) = \mathrm{Card}\ \left\{ i = 1, \ldots, L \ \middle|\ |\alpha_i^\star| \geq \lambda\mu \right\} \ . \tag{19}$$

With such simple formula on hand, expression of the model selection criteria GCV in (16) is readily available.



(a) (b)

Figure 1. GCV for the Cameraman (a) and the Neuron phantom (b). The translation-invariant discrete wavelet transform was used with the Cameraman image, and the curvelet transform with the Neuron phantom. The solid line represents the GCV, the dashed line the MSE and the dashed-dotted line the MAE.

Although this formula is only an approximation, in all our experiments, it performed reasonably well. This is testified by Fig. 1(a) and (b) which respectively show the behavior of the GCV as a function of λ for two images: the Cameraman portrayed in Fig. 4(a) and the Neuron phantom shown in Fig. 2(a). As the ground-truth is known in the simulation, we computed for each λ the mean absolute-error (MAE)—well adapted to Poisson noise as it

is closely related to the squared Hellinger distance [52]— as well as the mean square-error (MSE) between the deconvolved and true image. We can clearly see that the GCV reaches its minimum close to those of the MAE and the MSE. Even though the regularization parameter dictated by the GCV criterion is slightly higher than that of the MSE, which may lead to a somewhat over-smooth estimate.

F. Computational complexity and implementation details

The bulk of computation of our deconvolution algorithm is invested in applying Φ (resp. H) and its adjoint Φ^{T} (resp. H^{T}). These operators are never constructed explicitly, rather they are implemented as fast implicit operators taking a vector x, and returning Φx (resp. $\Phi^{\mathrm{T}} x$) and $\mathrm{H}x$ (resp. $\mathrm{H}^{\mathrm{T}} x$). Multiplication by H or H^{T} costs two FFTs, that is $2n \log n$ operations (n denotes the number of pixels). The complexity of Φ and Φ^{T} depends on the transforms in the dictionary: for example, the orthogonal wavelet transform costs $\mathcal{O}(n)$ operations, the translation-invariant discrete wavelet transform (TI-DWT) costs $\mathcal{O}(n \log n)$, the curvelet transform costs $\mathcal{O}(n \log n)$, etc. Let V_Φ denote the complexity of applying the analysis or synthesis operator. Define N_{FB} and N_{DR} as the number of iterations in the forward-backward algorithm and the Douglas-Rachford sub-iteration, and recall that L is the number of coefficients. The computational complexities of our iterations (13) and (14) are summarized below:

Algorithm	Computational complexity bounds	
	Φ orthobasis	Φ tight frame
(13)	$N_{\mathrm{FB}}\left(4n \log n + N_{\mathrm{DR}}\left(2V_\Phi + \mathcal{O}(n)\right)\right)$	$N_{\mathrm{FB}}\left(4n \log n + 2V_\Phi + N_{\mathrm{DR}}(2V_\Phi + \mathcal{O}(L))\right)$
(14)	$N_{\mathrm{FB}}\left(8n \log n + 2V_\Phi + \mathcal{O}(n)\right)$	$N_{\mathrm{FB}}\left(8n \log n + 6V_\Phi + \mathcal{O}(L)\right)$

The orthobasis case requires less multiplications by Φ and Φ^{T} because in that case, Φ is a bijective linear operator. Thus, the optimization problem (5) can be equivalently written in terms of image samples instead of coefficients, hence reducing computations in the corresponding iterations (13) and (14).

For our implementation, as in Algorithm 1, we have taken $a_t = b_t \equiv 0$ and $\beta_t \equiv 1$ in (13), and $\nu_t \equiv 1/2$ in (10). As the PSF h in our experiments is low-pass normalized to a unit sum, $\|\mathrm{H}\|_2^2 = 1$. Ψ was the ℓ_1-norm, leading to soft-thresholding. Furthermore, in order to accelerate the computation of prox_{f_2} in (13), the Douglas-Rachford sub-iteration (10) was only run once (i.e. $N_{\mathrm{DR}} = 1$) starting with $\gamma^0 = \alpha$. In this case, one can check that if $\gamma^0 \in \mathcal{C}'$, then this leads to the "natural" formula :

$$\mathrm{prox}_{f_2}(\alpha) = \mathcal{P}_{\mathcal{C}'} \circ \mathrm{prox}_{\frac{\lambda}{2}\Psi}(\alpha).$$

In our experimental studies, the GCV-based selection of λ was run using the forward-backward algorithm (13) which has a lower computational burden than (14) (see above table for computational complexities). Once λ was objectively chosen by the GCV procedure, the deconvolution algorithm was applied using (14) to exempt the user from the choice of the relaxation parameter μ.

V. Results

A. Simulated data

The performance of our approach has been assessed on several test images: a 128×128 neuron phantom [53], a 370×370 confocal microscopy image of micro-vessel cells [54], the Cameraman (256×256), a 512×512 simulated astronomical image of the Hubble Space Telescope Wide Field Camera of a distant cluster of galaxies [3]. Our algorithm was compared to RL with total variation regularization (RL-TV [7]), RL with multi-resolution support wavelet regularization (RL-MRS [9]), fast translation invariant tree-pruning reconstruction combined with an EM algorithm (FTITPR [55]) and the naive proximal method that would treat the noise as if it were Gaussian (NaiveGauss [12]). For all results presented, each algorithm was run with $N_{\mathrm{FB}} = 200$ iterations, enough to reach convergence. For all results below, λ was selected using the GCV criterion for our algorithm. For fair comparison to [12], λ was also chosen by adapting our GCV formula to the Gaussian noise.

Fig.2(a), depicts a phantom of a neuron with a mushroom-shaped spine. The maximum intensity is 30. Its blurred (using a 7×7 moving average) and blurred+noisy versions are in (b) and (c). With this neuron, and for NaiveGauss and our approach, the dictionary Φ contained the curvelet tight frame [32]. The deconvolution results are shown in Fig.2(d)-(h). As expected at this intensity level, the NaiveGauss algorithm performs quite badly, as it does not fit the noise model at this intensity regime. It turns out that NaiveGauss under-regularizes the estimate and the Poisson signal-dependent noise is not always under control. This behavior of NaiveGauss, which was predictable at this intensity level, will be observed on almost all tested images. RL-TV does a good job at deconvolution but the background is dominated by artifacts, and the restored neuron has staircase-like artifacts typical of TV regularization. Our approach provides a visually pleasant deconvolution result on this example. It efficiently restores the spine, although the background is not fully cleaned. RL-MRS also exhibits good deconvolution results. On this image, FTITPR provides a well smoothed estimate but with almost no deconvolution.

These qualitative visual results are confirmed by quantitative measures of the quality of deconvolution, where we used both the MAE and the traditional MSE criteria. At each intensity value, 10 noisy and blurred replications were generated and and the MAE was computed for each deconvolution algorithm. The average MAE over the 10 replications are given in Fig. 6 (similar results were obtained for the MSE, not shown here). In general, our algorithm performs very well at all intensity regimes (especially at medium to low). The NaiveGauss is among the worst algorithms at low intensity levels. Its performance becomes better as the intensity increases which was expected. RL-MRS is effective at low and medium intensity levels and is even better than our algorithm on the Cell image. RL-TV underperforms all algorithms at low intensity. We suspect the staircase-like artifacts of TV-regularization to be responsible for this behavior. At high intensity, RL-TV becomes competitive and its MAE comparable to ours.

The same experiment as above was carried out with the confocal microscopy cell image; see Fig. 3. In this experiment, the PSF was a 7×7 moving average. For the NaiveGauss and our approach, the dictionary Φ contained

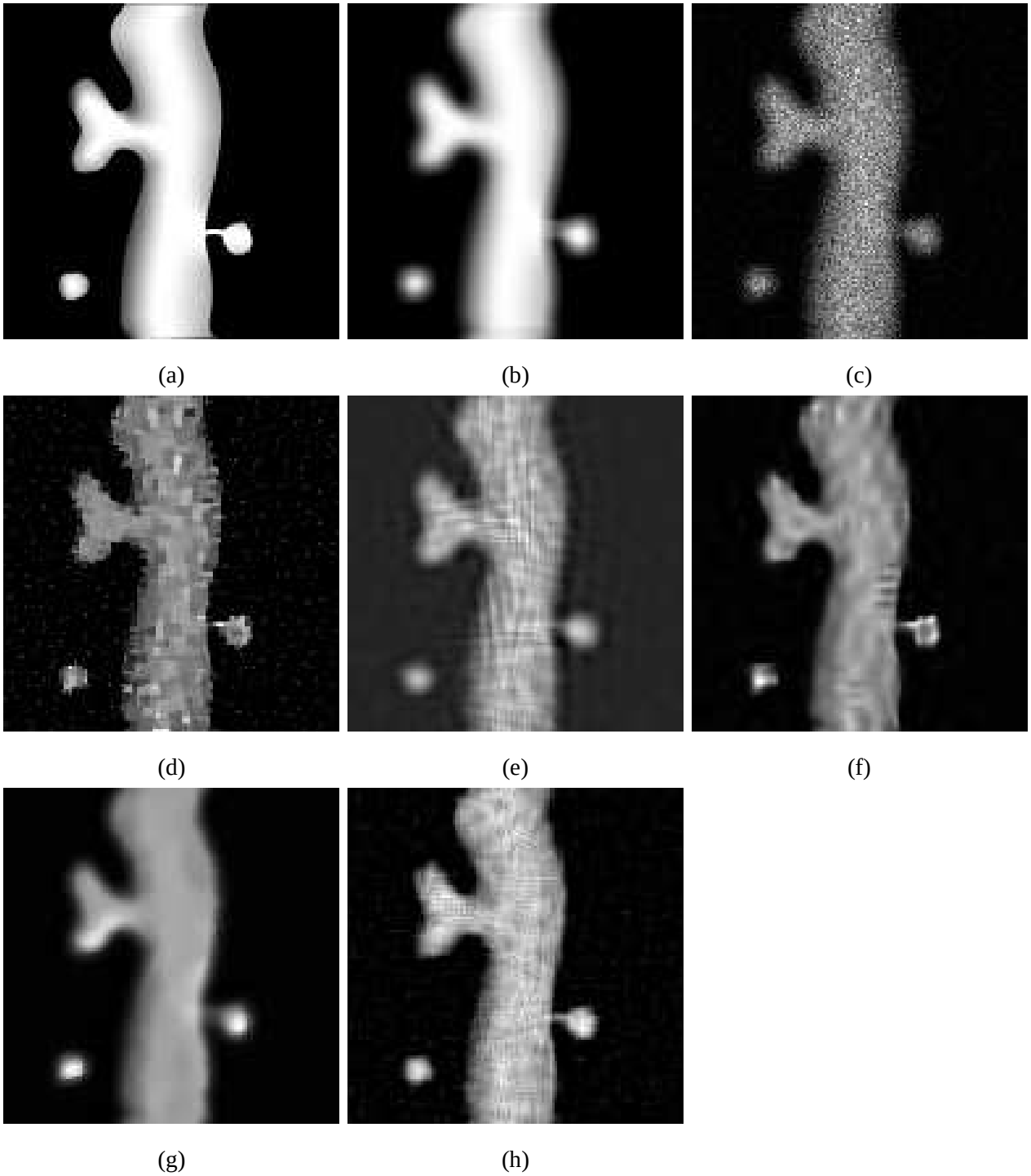


Figure 2. Deconvolution of a simulated neuron (Intensity $\leqslant$ 30). (a) Original, (b) Blurred, (c) Blurred&noisy, (d) RL-TV [7], (e) NaiveGauss [12], (f) RL-MRS [3], (g) FTITPR [55], (h) Our Algorithm.

the TI-DWT. NaiveGauss deconvolution result is spoiled by artifacts. RL-TV produces a good restoration of small isolated details but with a dominating staircase-like artifacts. FTITPR yields a somewhat oversmooth estimate, whereas our approach provides a sharper deconvolution result. This visual inspection is in agreement with the MAE measures of Fig. 6. In particular, one can notice that RL-MRS shows the best behavior, and the performance of our approach compared to the other methods on this cell image is roughly the same as on the previous neuron

image.

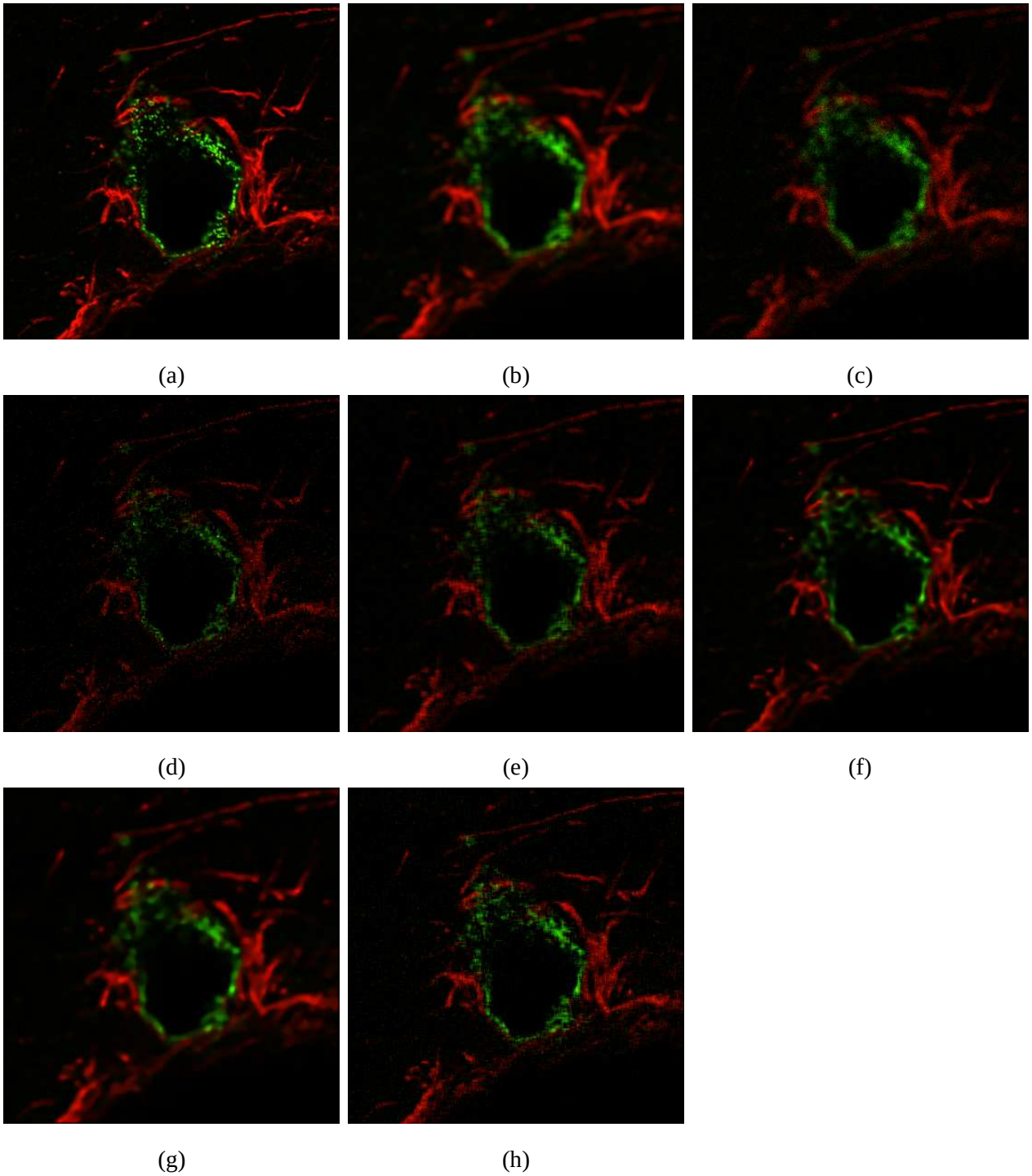


Figure 3. Deconvolution of a microscopy cell image (Intensity $\leqslant$ 30). (a) Original, (b) Blurred, (c) Blurred&noisy, (d) RL-TV [7], (e) NaiveGauss [12], (f) RL-MRS [3], (g) FTITPR [55] (h) Our Algorithm.

Fig.4(a) depicts the result of the experiment on the Cameraman with maximum intensity of 30. The PSF was the same as above. Again, the dictionary contained the TI-DWT frame. One may notice that the degradation in Fig.4(c) is quite severe. Our algorithm provides the most visually pleasing result with a good balance between regularization and deconvolution, although some artifacts are persisting. RL-MRS manages to deconvolve the image with more

artifacts than our approach, and suffers from a loss of photometry. Again, FTITPR gives an oversmooth estimate with many missing details. Both RL-TV and NaiveGauss yield results with many artifacts. This visual impression is in agreement with the MAE values in Fig. 6.

Figure 4. Deconvolution of the cameraman (Intensity $\leqslant$ 30). (a) Original, (b) Blurred, (c) Blurred&noisy, (d) RL-TV [7], (e) NaiveGauss [12], (f) RL-MRS [3], (g) FTITPR [55], (h) Our Algorithm.

To assess the computational cost of the compared algorithms, Tab. I summarizes the execution times on the Cameraman image with an Intel PC Core 2 Duo 2GHz, 2Gb RAM. Except RL-MRS which is written in C++, all

other algorithms were implemented in Matlab.

Method	Time (in s)
Our method	88
NaiveGauss	71
RL-MRS	99.5
RL-TV	15.5

Table I

EXECUTION TIMES FOR THE SIMULATED 256×256 CAMERAMAN IMAGE USING THE TI-DWT ($N_{\mathrm{FB}} = 200$).

The same experimental protocol was applied to a simulated Hubble Space Telescope wide field camera image of a distant cluster of galaxies portrayed in Fig.5(a). We used the Hubble Space Telescope PSF as given in [3]. The maximum intensity on the blurred image was 5000. For NaiveGauss and our approach, the dictionary contained the TI-DWT frame. For this image, the RL-MRS is clearly the best as it was exactly designed to handle Poisson noise for such images. Most faint structures are recovered by RL-MRS and large bright objects are well deconvolved. Our approach also yields a good deconvolution result and preserves most faint objects that are hardly visible on the degraded image. But the background is less clean than the one of RL-MRS. A this high intensity regime, NaiveGauss provides satisfactory results comparable to ours on the galaxies. FTITPR manages to properly recover most significant structures with a very clean background, but many faint objects are lost. RL-TV gives a deconvolution result comparable to ours on the brightest objects, but the background is dominated by spurious faint structures.

We also quantified the influence of the dictionary on the deconvolution performance on three test images. We first show in Fig. 7 the results of an experiment on a simulated 128×128 image, containing point-like sources (upper left), gaussians and lines. In this experiment, the maximum intensity of the original image is 30 and we used the 7×7 moving-average PSF. The TI-DWT depicted in Fig. 7(d) does a good job at recovering isotropic structures (point-like and gaussians), but the lines are not well restored. This drawback is overcome when using the curvelet transform as seen from Fig. 7(e), but as expected, the faint point-like source in the upper-left part is sacrificed. Visually, using a dictionary with both transforms seems to take the best of both worlds, see Fig. 7(g).

Fig. 8 shows the MAE—here normalized to the maximum intensity of the original image for the sake of legibility—as a function of the maximal intensity level for three test images: Neuron phantom, Cell and LinesGaussians. As above, three dictionaries were used: TI-DWT (solid line), curvelets (dashed line) and a dictionary built by merging both transforms (dashed-dotted line). For the Neuron phantom, which is piecewise-smooth, the best performance is given by the TI-DWT+curvelets dictionary at medium and high intensities. Even though the differences between dictionaries are less salient at low intensity levels. For the Cell image, which contains many localized structures, the TI-DWT seems to provide the best behavior, especially as the intensity increases. Finally, the behavior observed for the LinesGaussians image is just the opposite to that of the Cell. More precisely, the curvelets and TI-DWT+curvelets

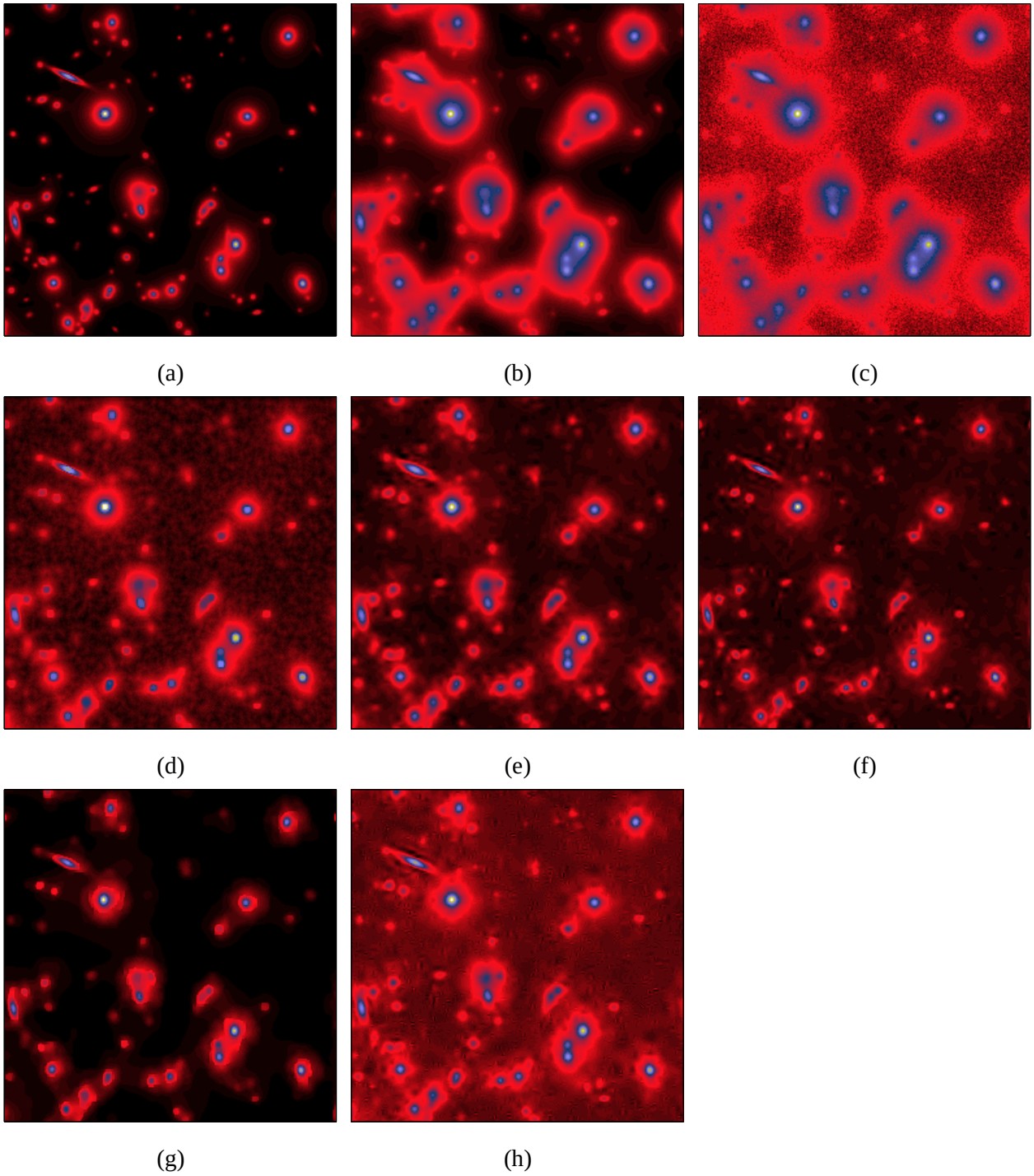
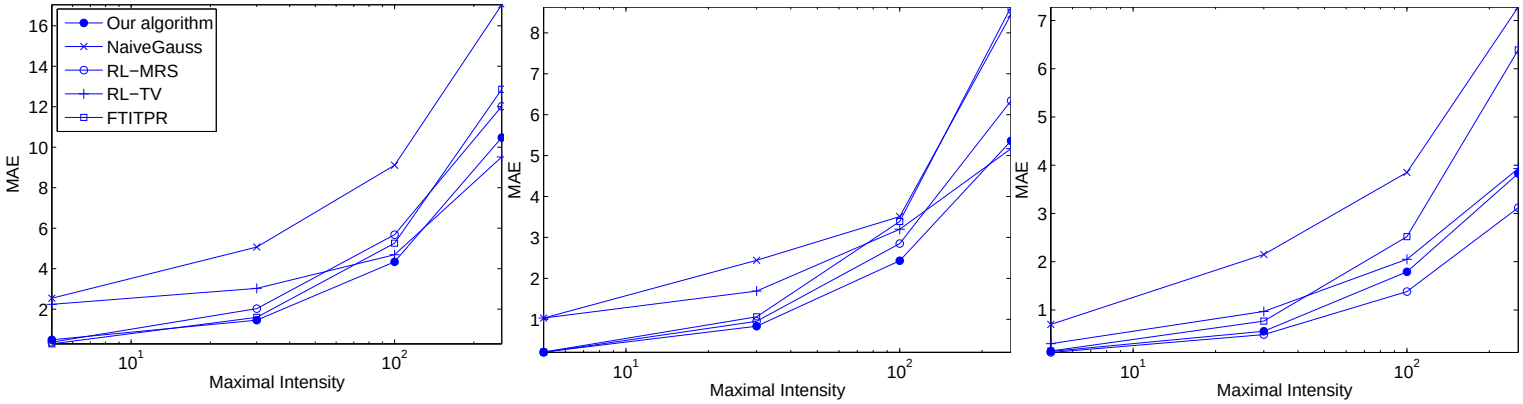


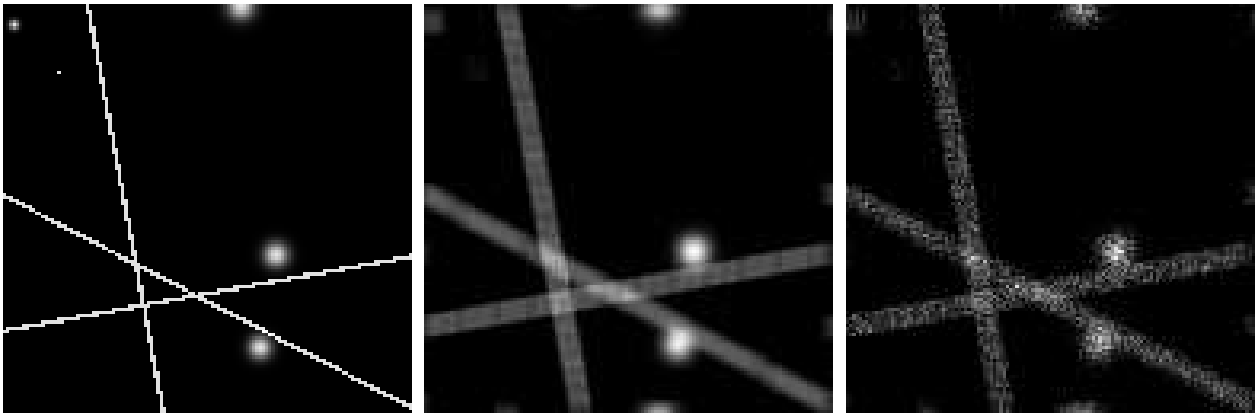
Figure 5. Deconvolution of the simulated sky. (a) Original, (b) Blurred, (c) Blurred&noisy, (d) RL-TV [7], (e) NaiveGauss [12], (f) RL-MRS [3], (g) FTITPR [55], (h) Our Algorithm.

dictionaries show the best performance with an advantage to the latter. However, this limited set of experiments does not allow to conclude that a dictionary built by amalgamating several transforms is the best strategy in general. Such a choice strongly depends on the image morphological content.

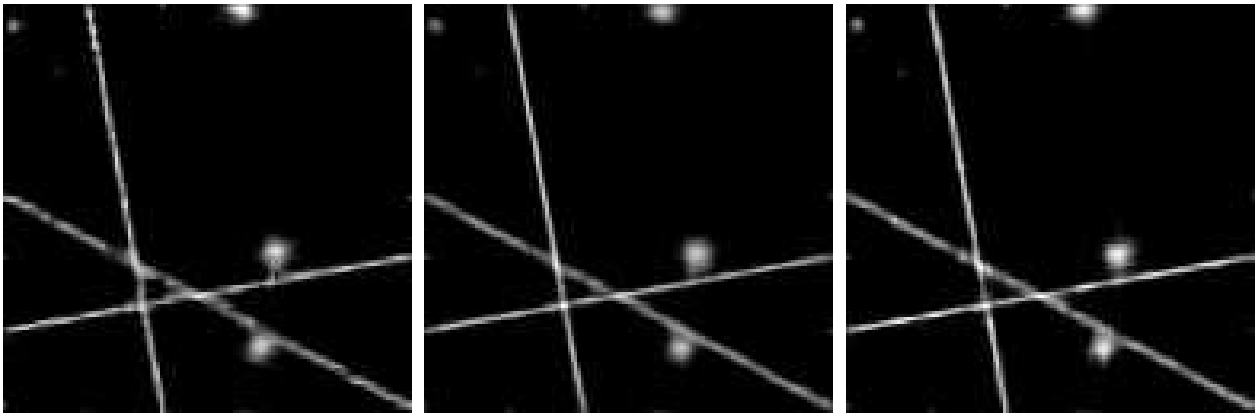


(a) (b) (c)

Figure 6. Average MAE of all algorithms as a function of the intensity level. (a) Cameraman, (b) Neuron phantom, (c) Cell.



(a) (b) (c)



(d) (e) (f)

Figure 7. Impact of the dictionary on deconvolution of the simulated LinesGaussians image with maximum intensity 30. (a) Original, (b) Blurred, (c) Blurred&noisy, (d) restored with TI-DWT, (e) restored with curvelets, (f) restored with a dictionary containing both transforms.

B. Real data

Finally, we applied our algorithm on a real 512×512 confocal microscopy image of neurons. Fig. 9(a) depicts the observed image[3] using the GFP fluorescent protein. The optical PSF of the fluorescence microscope was modeled

[3]Courtesy of the GIP Cycéron, Caen France.

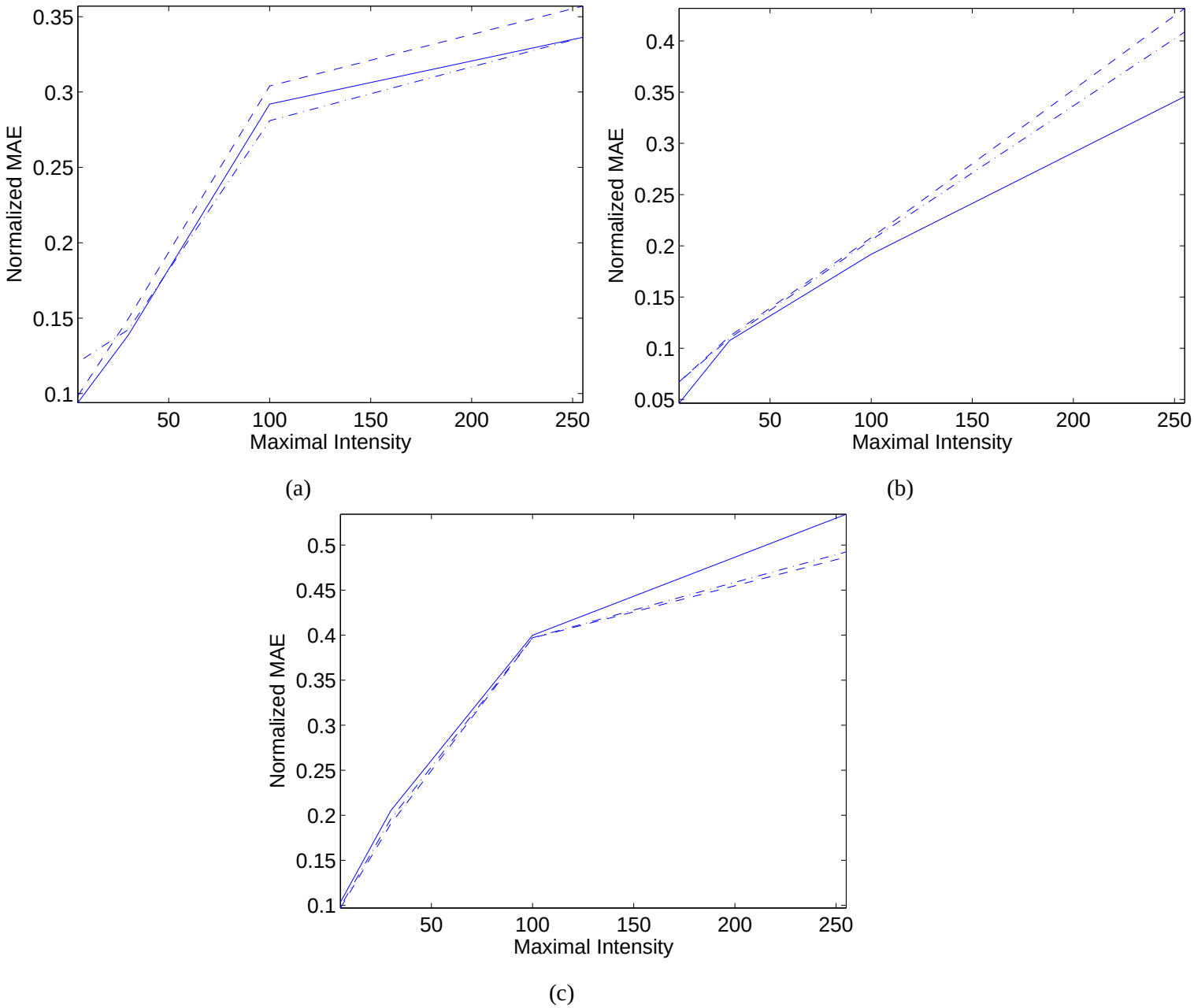
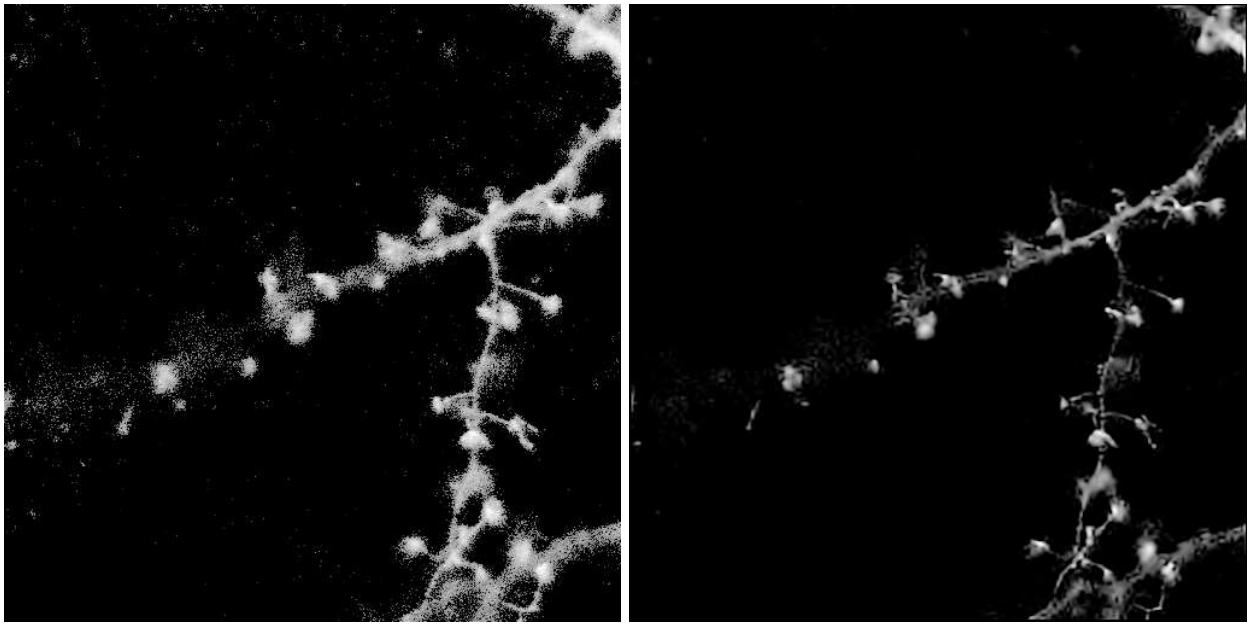


Figure 8. Impact of the dictionary on deconvolution performance as a function of the maximal intensity level for several test images: (a) Neuron phantom, (b) Cell and (c) LinesGaussians images. Solid line indicates the TI-DWT, dashed line corresponds to the curvelet transform, and dashed-dotted line to the dictionary built by merging both wavelets and curvelets.

using the gaussian approximation described in [56]. Fig. 9(b) shows the restored image using our algorithm with the wavelet transform. The images are shown in log-scale for better visual rendering. We can notice that the background has been cleaned and some structures have reappeared. The spines are well restored and part of the dendritic tree is reconstructed. However, some information can be lost (see tiny holes). We suspect that this result may be improved using a more accurate PSF model.



(a) (b)

Figure 9. Deconvolution of a real neuron. (a) Original noisy, (b) Restored with our algorithm

C. Reproducible research

Following the philosophy of reproducible research [57], a toolbox is made available freely for download at the first author's webpage `http://www.greyc.ensicaen.fr/~fdupe`. This toolbox is a collection of Matlab functions, scripts and datasets for image deconvolution under Poisson noise. It requires at least WaveLab 8.02 [57]. The toolbox implements the proposed algorithms and contains all scripts to reproduce most of the figures included in this paper.

VI. Conclusion

In this paper, a novel sparsity-based fast iterative thresholding deconvolution algorithm that takes account of the presence of Poisson noise was presented. The Poisson noise was handled properly. A careful theoretical study of the optimization problem and characterization of the iterative algorithm were provided. The choice of the regularization parameter was also attacked using a GCV-based procedure. Several experimental tests have shown the capabilities of our approach, which compares favorably with some state-of-the-art algorithms. Encouraging preliminary results were also obtained on real confocal microscopy images.

The present work may be extended along several lines. For example, it is worth noting that our approach generalizes straightforwardly to any non-linearity in (1) other than the square-root, provided that the corresponding data fidelity term as in (4) is convex and has a Lipschitz-continuous gradient. This is for instance the case if a generalization of the Anscombe VST [58] is applied to a Poisson plus Gaussian noise, which is a realistic noise model for data obtained from a CCD detector. For such a noise, one can easily show similar results to those proved in our work. In this paper, the simple expression of the degrees of freedom df was conjectured without a rigorous

proof. Deriving the exact analytical expression of df, if possible, needs further investigation and a very careful analysis that we leave for a future work. On the applicative side, the extension to 3D to handle confocal microscopy volumes is under investigation. Extension to multi-valued images is also an important aspect that will be the focus of future research.

REFERENCES

[1] P. Sarder and A. Nehorai, "Deconvolution Method for 3-D Fluorescence Microscopy Images," *IEEE Sig. Pro. Mag.*, vol. 23, pp. 32–45, 2006.

[2] J. Pawley, *Handbook of Confocal Microscopy*. Plenum Press, 2005.

[3] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. Springer, 2006.

[4] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Prentice-Hall, 1977.

[5] P. Jansson, *Image Recovery: Theory and Application*. New York: Academic Press, 1987.

[6] ——, *Deconvolution of Images and Spectra*. New York: Academic Press, 1997.

[7] N. Dey, L. Blanc-Féraud, C. Zimmer, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia., "A deconvolution method for confocal microscopy with total variation regularization," in *IEEE ISBI*, 2004.

[8] J.-L. Starck and F. Murtagh, "Image restoration with noise suppression using the wavelet transform," *Astronomy and Astrophysics*, vol. 288, pp. 343–348, 1994.

[9] J.-L. Starck, A. Bijaoui, and F. Murtagh, "Multiresolution support applied to image filtering and deconvolution," *CVGIP: Graphical Models and Image Processing*, vol. 57, pp. 420–431, 1995.

[10] G. Jammal and A. Bijaoui, "Dequant: a flexible multiresolution restoration framework," *Signal Processing*, vol. 84, pp. 1049–1069, 2004.

[11] J. B. de Monvel et al, "Image Restoration for Confocal Microscopy: Improving the Limits of Deconvolution, with Application of the Visualization of the Mammalian Hearing Organ," *Biophysical Journal*, vol. 80, pp. 2455–2470, 2001.

[12] C. Vonesch and M. Unser, "A fast iterative thresholding algorithm for wavelet-regularized deconvolution," *IEEE ISBI*, 2007.

[13] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraints," *Comm. Pure Appl. Math.*, vol. 112, pp. 1413–1541, 2004.

[14] R. Nowak and E. Kolaczyk, "A statistical multiscale framework for poisson inverse problems," *IEEE Transactions on Information Theory*, vol. 46, pp. 2794–2802, 2000.

[15] L. Cavalier and J.-Y. Koo, "Poisson intensity estimation for tomographic data using a wavelet shrinkage approach," *IEEE Transactions on Information Theory*, vol. 48, pp. 2794–2802, 2002.

[16] D. L. Donoho, "Nonlinear solution of inverse problems by wavelet-vaguelette decomposition," *Applied and Computational Harmonic Analysis*, vol. 2, pp. 101–126, 1995.

[17] A. Antoniadis and Bigot, "Poisson inverse problems," *Annals of Statistics*, vol. 34, pp. 1811–1825, 2006.

[18] A. Cohen, M. Hoffman, and M. Reiss, "Adaptive wavelet-galerkin methods for inverse problems," *SIAM J. Numerical Analysis*, vol. 42, pp. 1479–1501, 2004.

[19] M. Figueiredo and R. Nowak, "An em algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, pp. 906–916, 2003.

[20] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[21] G. Teschke, "Multi-frame representations in linear inverse problems with mixed multi-constraints," *Applied and Computational Harmonic Analysis*, vol. 22, no. 1, pp. 43–60, 2007.

[22] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inv. Prob.*, vol. 23, pp. 1495–1518, 2007.

[23] M. J. Fadili and J.-L. Starck, "Sparse representation-based image deconvolution by iterative thresholding," in *ADA IV*. France: Elsevier, 2006.

[24] J.-L. Starck, M. Nguyen, and F. Murtagh, "Wavelets and curvelets for image deconvolution: a combined approach," *Signal Processing*, vol. 83, pp. 2279–2283, 2003.

[25] C. Chaux, L. Blanc-Féraud, and J. Zerubia, "Wavelet-based restoration methods: application to 3d confocal microscopy images," in *SPIE Wavelets XII*, 2007.

[26] C. Lemaréchal and J.-B. Hiriart-Urruty, *Convex Analysis and Minimization Algorithms I*, 2nd ed. Springer, 1996.

[27] P. G. Ciarlet, *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson-Paris, 1985.

[28] P. L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, vol. 53, pp. 475–504, 2004.

[29] F. J. Anscombe, "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, vol. 35, pp. 246–254, 1948.

[30] S. G. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. Academic Press, 1998.

[31] E. J. Candès and D. L. Donoho, "Curvelets – a surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting: Saint-Malo 1999*, A. Cohen, C. Rabut, and L. Schumaker, Eds. Nashville, TN: Vanderbilt University Press, 1999.

[32] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul.*, vol. 5, pp. 861–899, 2005.

[33] L. Demanet and L. Ying, "Wave atoms and sparsity of oscillatory patterns," *Applied and Computational Harmonic Analysis*, vol. 23, no. 3, pp. 368–387, 2007.

[34] M. Jansen, "Multiscale poisson data smoothing," *J. Royal Stat. Soc. B.*, vol. 68, no. 1, pp. 27–48, 2006.

[35] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *CRAS Sér. A Math.*, vol. 255, pp. 2897–2899, 1962.

[36] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programming*, vol. 55, pp. 293–318, 1992.

[37] P.-L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[38] M. Fadili, J.-L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Computer Journal*, 2006, in press.

[39] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM Journal on Optimization*, vol. 18, no. 4, pp. 1351–1376, November 2007.

[40] P. Tseng, "A modified forward-backward splitting method for maximal monotone mappings," *SIAM J. Control & Optim.*, vol. 38, pp. 431–446, 2000.

[41] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[42] H. Akaike, B. N. Petrox, and F. Caski, "Information theory and an extension of the maximum likelihood principle," *Second International Symposium on Information Theory. Akademiai Kiado, Boudapest.*, pp. 267–281, 1973.

[43] G. Schwarz, "Estimation of the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[44] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 9, pp. 1135–1151, 1981.

[45] B. Efron, "How biased is the apparent error rate of a prediction rule," *Journal of the American Statistical Association*, vol. 81, pp. 461–470, 1981.

[46] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Processing*, vol. 56, no. 1, pp. 33–44, 1997.

[47] B. Efron, "The estimation of prediction error: covariance penalties and cross-validation," *Journal of the American Statistical Association*, vol. 99, pp. 619–642, 2004.

[48] J. Ye, "On measuring and correcting the effects of data mining and model selection," *Journal of the American Statistical Association*, vol. 93, pp. 120–131, 1998.

[49] X. Shen and J. Ye, "Adaptive model selection," *Journal of the American Statistical Association*, vol. 97, pp. 210–221, 2002.

[50] H. Zou, T. Hastie, and R. Tibshirani, "On the "degrees of freedom" of the lasso," *Annals of Statistics*, vol. 35, pp. 2173–2192, 2007.

[51] C. Dossal, "A necessary and sufficient condition for exact recovery by ℓ_1 minimization," 2007, preprint, hal-00164738. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00164738/en/

[52] A. Barron and T. Cover, "Minimum complexity density estimation," *IEEE Transactions on Information Theory*, vol. 37, pp. 1034–1054, 1991.

[53] [Online]. Available: Rebecca Willett homepage http://www.ee.duke.edu/ willett/

[54] [Online]. Available: ImageJ website http://rsb.info.nih.gov/ij/

[55] R. Willett and R. Nowak, "Fast multiresolution photon-limited image reconstruction," in *IEEE ISBI*, 2004.

[56] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope PSF models," *Applied Optics*, vol. 46, no. 10, pp. 1819–1829, 2007.

[57] J. Buckheit and D. Donoho, "Wavelab and reproducible research," in *Wavelets and Statistics*, A. Antoniadis, Ed. Springer, 1995.

[58] F. Murtagh, J.-L. Starck, and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support," *Astronomy and Astrophysics, Supplement Series*, vol. 112, pp. 179–189, 1995.

[59] P. L. Combettes and J.-. Pesquet, "A Douglas-Rachford splittting approach to nonsmooth convex variational signal recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, 2007.

APPENDIX

A. Proof of Proposition 1

Proof:

(i) f_1 is obviously convex, as Φ and H are bounded linear operators and f is convex.

(ii) The computation of the gradient of f_1 is straightforward.

(iii) For any $\alpha, \alpha' \in \mathcal{H}$, we have,

$$\left\|\nabla f_1(\alpha) - \nabla f_1(\alpha')\right\| \leqslant \|\Phi\|_2 \|\mathrm{H}\|_2 \left\|\nabla F \circ \mathrm{H} \circ \Phi(\alpha) - \nabla F \circ \mathrm{H} \circ \Phi(\alpha')\right\| . \tag{20}$$

The function $-\frac{z_i}{\sqrt{\eta_i+3/8}} + 2$ is one-to-one increasing on $[0, +\infty)$ with derivative uniformly bounded above by $\frac{z_i}{2}(8/3)^{3/2}$. Thus,

$$\begin{aligned} \left\|\nabla F \circ \mathrm{H} \circ \Phi(\alpha) - \nabla F \circ \mathrm{H} \circ \Phi(\alpha')\right\| &\leqslant \left(\frac{8}{3}\right)^{\frac{3}{2}} \frac{\|z\|_\infty}{2} \left\|\mathrm{H} \circ \Phi(\alpha) - \mathrm{H} \circ \Phi(\alpha')\right\| \\ &\leqslant \left(\frac{8}{3}\right)^{\frac{3}{2}} \frac{\|z\|_\infty}{2} \|\Phi\|_2 \|\mathrm{H}\|_2 \left\|\alpha - \alpha'\right\| . \end{aligned} \tag{21}$$

Using the fact that $\|\Phi\|_2^2 = \left\|\Phi\Phi^{\mathrm{T}}\right\|_2 = c$ for a tight frame, and z is bounded since $y \in \ell_\infty$ by assumption, we conclude that ∇f_1 is Lipschitz-continuous with the constant given in (8).

■

B. Proof of Proposition 2

Proof: The existence is obvious because J is coercive. If Φ is an orthobasis and $\ker(\mathrm{H}) = \emptyset$ then f_1 is strictly convex and so is J leading to a strict minimum. Similarly, if ψ is strictly convex, so is f_2, hence J. ■

C. Proof of Lemma 1

Proof:

1) Let $g : \gamma \mapsto \frac{1}{2}\|\alpha-\gamma\|^2 + \lambda\Psi(\gamma)$. From Definition 1, $\mathrm{prox}_{\lambda\Psi}(\alpha)$ is the unique minimizer of g, whereas $\mathrm{prox}_{f_2}(\alpha)$ is the unique minimizer of $g + \imath_{\mathcal{C}'}$. If $\alpha \in \mathcal{C}'$, then $\mathrm{prox}_{f_2}(\alpha)$ is also the unique minimizer of g as obviously $\imath_{\mathcal{C}'}(\alpha) = 0$ in this case. That is, $\mathrm{prox}_{f_2}(\alpha) = \mathrm{prox}_{\lambda\Psi}(\alpha)$.
2) Let's now turn to the general case. We have to find the unique solution to the following minimization problem:

$$\mathrm{prox}_{f_2}(\alpha) = \arg\min_{\gamma}\ g(\gamma) + \imath_{\mathcal{C}} \circ \Phi(\gamma) = \arg\min_{\gamma\in\mathcal{C}'}\ g(\gamma).$$

As both $\imath_{\mathcal{C}}$ and $g \in \Gamma_0\left(\mathbb{R}^L\right)$ but non-differentiable, we use the Douglas-Rachford splitting method [28], [36]. This iteration is given by:

$$\gamma^{t+1} = \gamma^t + \nu_t\left(\mathrm{rprox}_{\lambda\Psi+\frac{1}{2}\|.-\alpha\|^2} \circ \mathrm{rprox}_{\imath_{\mathcal{C}'}} - \mathrm{I}\right)(\gamma^t). \tag{22}$$

where the sequence ν_t satisfies the condition of the lemma. From [28, Corollary 5.2], and by strict convexity, we deduce that the sequence of iterates γ^t converges to a unique point γ, and $\mathcal{P}_{\mathcal{C}'}(\gamma)$ is the unique proximity point $\mathrm{prox}_{f_2}(\alpha)$.

It remains now to explicitly express $\mathrm{prox}_{\lambda\Psi+\frac{1}{2}\|.-\alpha\|^2}$ and $\mathrm{prox}_{\imath_{\mathcal{C}'}}$. $\mathrm{prox}_{\lambda\Psi+\frac{1}{2}\|.-\alpha\|^2}$ is the proximity operator of a quadratic perturbation of $\lambda\Psi$, which is related to $\mathrm{prox}_{\lambda\Psi}$ by:

$$\mathrm{prox}_{\lambda\Psi+\frac{1}{2}\|.-\alpha\|^2}(.) = \mathrm{prox}_{\frac{\lambda}{2}\Psi}\left(\frac{\alpha+.}{2}\right). \tag{23}$$

See [20, Lemma 2.6].

Using [59, Proposition 11], we have

$$\begin{aligned}\mathrm{prox}_{\imath_{\mathcal{C}}\circ\Phi} &= \mathrm{I} + c^{-1}\Phi^{\mathrm{T}} \circ (\mathcal{P}_{\mathcal{C}} - \mathrm{I}) \circ \Phi \\ &= c^{-1}\Phi^{\mathrm{T}} \circ \mathcal{P}_{\mathcal{C}} \circ \Phi + (\mathrm{I} - c^{-1}\Phi^{\mathrm{T}}\Phi).\end{aligned} \tag{24}$$

This completes the proof. ■

D. Proof of Theorem 1

Proof: The most general result on the convergence of the forward-backward algorithm is is due to [20, Theorem 3.4]. Hence, combining this theorem with Lemma 1, Lemma 2 and Proposition 1, the result follows. ■