

# A Fast Thresholded Landweber Algorithm for Wavelet-Regularized Multidimensional Deconvolution

Cédric Vonesch, *Student Member, IEEE*, and Michael Unser, *Fellow, IEEE*

**Abstract**—We present a fast variational deconvolution algorithm that minimizes a quadratic data term subject to a regularization on the  $\ell^1$ -norm of the wavelet coefficients of the solution. Previously available methods have essentially consisted in alternating between a Landweber iteration and a wavelet-domain soft-thresholding operation. While having the advantage of simplicity, they are known to converge slowly. By expressing the cost functional in a Shannon wavelet basis, we are able to decompose the problem into a series of subband-dependent minimizations. In particular, this allows for larger (subband-dependent) step sizes and threshold levels than the previous method. This improves the convergence properties of the algorithm significantly. We demonstrate a speed-up of one order of magnitude in practical situations. This makes wavelet-regularized deconvolution more widely accessible, even for applications with a strong limitation on computational complexity. We present promising results in 3-D deconvolution microscopy, where the size of typical data sets does not permit more than a few tens of iterations.

**Index Terms**—Deconvolution, fast, fluorescence microscopy, iterative, nonlinear, sparsity, 3-D, thresholding, wavelets,  $\ell^1$ -regularization.

## I. INTRODUCTION

### A. Motivation

**D**URING the past decade, biological imaging has been revolutionized by the widespread availability of novel fluorescence labeling techniques and the development of advanced 3-D optical microscopy [33]. As biological research moves towards molecular scale, the constraints in terms of resolution and light efficiency are becoming more and more stringent. In this context, scientists are increasingly relying on computational methods to take full advantage of their instrumentation [6], [16], [17], [25].

The use of deconvolution to enhance micrographs is commonly referred to as *deconvolution microscopy* [23], [34]. The results are the most striking in 3-D, as demonstrated by the pioneering work of Agard and Sedat [1], which had a significant impact on the field. The main challenge in 3-D resides in the

size of typical data sets, which severely restricts the computational complexity of numerical restoration methods. With the present state of computer hardware, one cannot afford more than the equivalent of a few tens of FFTs on a complete 3-D stack. This explains why most software packages for deconvolution microscopy are still based on relatively standard methods, at least when compared to the state-of-the-art in 2-D imaging [3], [27]. For instance, the Tikhonov–Miller and Richardson–Lucy algorithms are widely used; they require two (respectively four) FFTs per iteration, so that they are usually stopped after 30 to 40 iterations.

Wavelet regularization is a relatively recent technique in the area of deconvolution. Several research groups have independently proposed a restoration procedure that alternates between wavelet-coefficient thresholding and a Landweber update. Some of the ideas of the method, which we shall call the thresholded Landweber (TL) algorithm, can be traced back to the work of Starck *et al.* [29], [30]. Figueiredo and Nowak presented the first formal derivation of the TL algorithm using a statistically based penalized likelihood formulation [14]. The algorithm was then rederived in a deterministic variational framework and generalized to a more general class of linear operators by Daubechies *et al.* [11]; note that these authors also provided a general convergence proof that can be readily transposed to the algorithm that we present in this paper. Bect *et al.* considered a similar wavelet-regularized cost function, but with an additional total variation term [4]. Most recently, Chaux *et al.* extended this type of approach to frame-based wavelet regularization [9].

The TL algorithm offers attractive features for 3-D deconvolution microscopy. First, it belongs to the class of nonlinear wavelet-based methods, which have already proven to be very efficient for image-restoration applications such as denoising. In a series of 2-D deconvolution experiments (see [14]), the TL algorithm was actually shown to perform best against several other wavelet-based deconvolution methods. Second, the TL algorithm is simple to implement. Since it is iterative (unlike other wavelet-based methods, e.g., the ones described in [12], [19], and [26]), additional constraints such as positivity can be easily incorporated. Finally, it requires only two FFTs and two wavelet transforms per iteration. This represents an intermediate computational complexity compared to the standard algorithms mentioned before.

The only weak point of the TL algorithm is its slow convergence speed. This is especially true at high SNR levels, which may require several hundred iterations. The primary purpose of

Manuscript received June 5, 2007; revised November 27, 2007. This work was supported in part by the Hasler Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stanley J. Reeves.

The authors are with the Biomedical Imaging Group, EPFL, CH-1015 Lausanne, Switzerland (e-mail: cedric.vonesch@epfl.ch; michael.unser@epfl.ch).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2008.917103

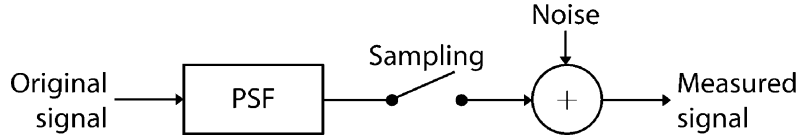


Fig. 1. Image-formation and measurement model.

this paper is to propose a variation of the TL algorithm that minimizes the same functional but that is substantially faster computationally. Our efforts are directed towards making wavelet regularization a tractable alternative for deconvolving large data sets. In particular, we demonstrate its feasibility on a 3-D stack of widefield fluorescence micrographs. To our knowledge, this is the first practical implementation of wavelet-based 3-D deconvolution microscopy, and the perspectives in this area appear to be quite promising.

### B. Image-Formation Model

In this paper, we are concerned with the recovery of a multi-dimensional signal from its noisy observation through a convolutive imaging system.

Fig. 1 shows a model of the image-acquisition chain that applies, for example, to fluorescence microscopy. Generally speaking, we are interested in  $d$ -dimensional signals. The observed image is the  $d$ -dimensional convolution of an original image (the characteristic function of the object of interest) with the impulse response (or point spread function, PSF) of the imaging system. In practice, we only have access to a finite number of samples of the image and these measurements are corrupted by noise.

In the sequel, to simplify the presentation, we will adopt a purely discrete description of the image-formation process. We shall assume that

$$\mathbf{y} = \mathbf{H}\mathbf{x}_{\text{orig}} + \mathbf{b}$$

where  $\mathbf{x}_{\text{orig}}$  and  $\mathbf{y}$  are vectors containing uniform samples of the original and measured signals, respectively;  $\mathbf{b}$  represents the measurement error. These vectors have  $N = N_1 \times N_2 \times \dots \times N_d$  components, where  $N_k$  stands for the number of samples along dimension  $k$ .  $\mathbf{H}$  is a square, (block-) circulant matrix that approximates the convolution with the PSF in Fig. 1.

### C. Deconvolution as a Variational Problem

The recovery of  $\mathbf{x}_{\text{orig}}$  from  $\mathbf{y}$  is an *ill-posed problem* [5]. To obtain an estimate with reasonable accuracy and robustness to noise, one must often rely on prior information about the original signal. The estimation of  $\mathbf{x}_{\text{orig}}$  is then formulated as an optimization problem which incorporates this prior information. The quality of a given estimate  $\mathbf{x}$  is typically measured by a *cost function* of the form

$$J(\mathbf{x}) = J_{\text{data}}(\mathbf{x}) + \lambda J_{\text{reg}}(\mathbf{x}). \quad (1)$$

The *data term*  $J_{\text{data}}(\mathbf{x})$  quantifies the “prediction error” with respect to the measurements. The *regularization term*  $J_{\text{reg}}(\mathbf{x})$  is designed to penalize an estimate that would not exhibit the expected properties. The *regularization parameter*  $\lambda$  balances the

contribution of both terms. In practice, the value of  $\lambda$  needs to be adjusted appropriately; this is either done empirically by trial and error or by using some data-driven cross-validation method [5], [32]. The problem of deconvolution is then to find an estimate that minimizes the cost function  $J(\mathbf{x})$ . This functional can also be interpreted as a (negative) log-likelihood in a Bayesian statistical framework, and deconvolution can then be seen as a maximum *a posteriori* (MAP) estimation problem [5], [14].

In this paper, we will consider cost functions that favor wavelet expansions with a small number of nonzero coefficients. This is typical of objects that are piecewise smooth [22]. The property of natural images having a sparse wavelet expansion is heavily exploited in compression standards such as JPEG2000 [10]. Here, we will concentrate on the case of an orthonormal wavelet decomposition, which will be characterized by an (orthonormal) decomposition matrix  $\mathbf{W}$ . In our notation, the matrix-vector product  $\mathbf{w}_{\mathbf{x}} = \mathbf{W}\mathbf{x}$  yields the coefficients of  $\mathbf{x}$  in the wavelet basis, and  $\mathbf{W}^T \mathbf{W}\mathbf{x}$  reconstructs the signal from these coefficients. We implicitly include the (coarsest-scale) scaling functions when referring to the wavelet basis, unless specified otherwise. The cost function is then defined by

$$J_{\text{data}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \quad \text{and} \quad J_{\text{reg}}(\mathbf{x}) = \|\mathbf{W}\mathbf{x}\|_1. \quad (2)$$

Here, the data term measures the residual in the image-domain using the (squared) Euclidian norm  $\|\cdot\|_2$ . The regularization term is the sum of the absolute values of the wavelet coefficients, which is denoted with the  $\ell_1$ -norm  $\|\mathbf{w}_{\mathbf{x}}\|_1$ . In contrast with traditional quadratic regularization, the  $\ell_1$ -norm leads to a nonlinear deconvolution algorithm. Compared to the  $\ell_2$  norm, it puts less weight on coefficients that are greater than 1, and more weight on coefficients that are smaller than 1. This tends to favor signals that are “sparse” solutions with a few large wavelet coefficients.

### D. Organization of the Paper

The remainder of the paper is organized as follows. In Section II, we present the basic building blocks of the algorithm and relate them to classical image processing tasks. In Section III, we recall the formalism of Daubechies *et al.* leading to the existing thresholded Landweber algorithm. In Section IV, we derive a faster algorithm that minimizes the same functional in the case of Shannon wavelets. The last section is devoted to numerical experiments that illustrate the performance of the proposed algorithm, on both simulated data and real 3-D microscopy data.

## II. PRELIMINARIES

There are two particular cases of (1) and (2) that provide the basic building blocks of the thresholded Landweber algorithm.

### A. Nonregularized Case

The first one corresponds to  $\lambda = 0$ . The cost function then reduces to

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2. \quad (3)$$

A minimizer of this expression is the classical *least squares* estimator, which is given by  $\mathbf{x} = \mathbf{H}^\dagger \mathbf{y}$ , where<sup>1</sup>

$$\mathbf{H}^\dagger = \lim_{c \rightarrow 0, c > 0} (\mathbf{H}^T \mathbf{H} + c\mathbf{I})^{-1} \mathbf{H}^T$$

is the pseudoinverse of  $\mathbf{H}$ . It is the solution with the smallest (Euclidian) norm. In practice,  $\mathbf{H}^\dagger \mathbf{y}$  can be simply computed in the frequency domain (i.e., in the DFT basis), where  $\mathbf{H}^\dagger$  is diagonalized.

However, this direct method is prone to noise amplification because the matrix  $\mathbf{H}^T \mathbf{H}$  may have small but nonzero eigenvalues. In order to delay this amplification process, the minimization of (3) is often performed iteratively. Starting from an initial estimate  $\mathbf{x}^{(0)}$ , one constructs a sequence  $(\mathbf{x}^{(n)})_{n \in \mathbb{N}}$  that converges to a minimizer of  $J(\mathbf{x})$ . A gradient descent on (3) with a fixed step size  $\tau$  yields the following update formula:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \tau \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}). \quad (4)$$

This algorithm is usually named after *Landweber* [20]. For an in-depth investigation of this very basic procedure—the grandfather of iterative deconvolution—we refer to [5].

### B. The Pure Denoising Case

The second particular case occurs when  $\mathbf{H} = \mathbf{I}$ . The cost function becomes

$$J(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1. \quad (5)$$

This corresponds to a pure denoising problem, where we want to recover the original signal from the nonblurred but noisy measurement  $\mathbf{y}$  (see [8] for a complete mathematical treatment). When the wavelet basis is orthonormal—as in our case—the minimization of this expression is straightforward, since the quadratic term can be written in the wavelet domain (Parseval relation)

$$\begin{aligned} J(\mathbf{x}) &= \|\mathbf{W}\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1 \\ &= \sum_{n=1}^N |w_{y_n} - w_{x_n}|^2 + \lambda |w_{x_n}| \end{aligned} \quad (6)$$

where  $w_{y_n}$  and  $w_{x_n}$  are the wavelet coefficients of  $\mathbf{y}$  and  $\mathbf{x}$ , respectively. In this form, the cost function is completely decoupled, which implies that its minimization can be achieved in a coefficient-wise fashion. The solution involves the *soft-thresholding* function

$$\mathcal{T}_{\lambda/2}(w) = \text{sgn}(w)(|w| - \lambda/2)_+ \quad (7)$$

where  $(\cdot)_+$  is the positive-part function

$$(t)_+ = \begin{cases} t, & \text{if } t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

<sup>1</sup>In the case of complex vectors/matrices,  $T$  will denote Hermitian transposition.

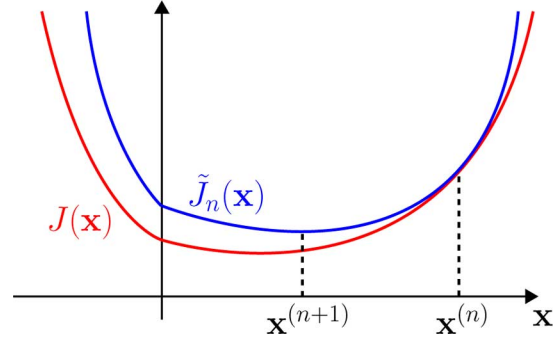


Fig. 2. Bound-optimization principle: an estimate of the minimizer of the original cost function is constructed by minimizing an auxiliary functional. The process is repeated iteratively.

The minimizer of (5) is simply given by

$$\mathbf{x} = \mathbf{W}^T \mathcal{T}_{\lambda/2} \{\mathbf{W}\mathbf{y}\}$$

where  $\mathcal{T}_{\lambda/2}\{\cdot\}$  denotes a component-wise application of the soft-thresholding function. Note that this formalism can be adapted to complex wavelet decompositions by replacing  $\text{sgn}(w)$  by  $\exp(i \arg w)$  in (7). Interestingly, this algorithm was first proposed empirically for noise reduction in magnetic resonance imaging [35]. It has become very popular in the field, following the impulsion of Donoho and Johnstone who justified it on solid statistical grounds [13]. Several authors have also proposed a Bayesian interpretation of the method, see, e.g., [2], [24], and [28].

### C. General Principle of Bound Optimization Algorithms

Except for simplified cases such as those just mentioned, the functional  $J(\mathbf{x})$  defined by (2) cannot be minimized directly by a coefficient-wise rule. This is primarily due to the convolution matrix  $\mathbf{H}$ , which makes the wavelet coefficients of  $\mathbf{x}$  interdependent, hence precluding a simple coefficient-wise rule. Instead, one can use an iterative minimization technique as proposed in [4], [11], and [14]. Interestingly, this involves a judicious combination of the two aforementioned methods. In the next section, we shall briefly review the derivation of Daubechies *et al.*; it is based on a *bound optimization* approach [18], [21], which essentially consists in replacing  $J(\mathbf{x})$  by a succession of auxiliary functionals that are easy to minimize.

Using the current estimate  $\mathbf{x}^{(n)}$ , the key idea is to construct an auxiliary functional  $\tilde{J}_n(\mathbf{x})$  with the following properties:

- when  $\mathbf{x} = \mathbf{x}^{(n)}$ ,  $\tilde{J}_n(\mathbf{x})$  coincides with  $J(\mathbf{x})$ ;
- when  $\mathbf{x} \neq \mathbf{x}^{(n)}$ ,  $\tilde{J}_n(\mathbf{x})$  upper-bounds  $J(\mathbf{x})$ .

As illustrated in Fig. 2 for  $N = 1$ , these properties guarantee that, by minimizing  $\tilde{J}_n(\mathbf{x})$ , we will also decrease  $J(\mathbf{x})$ . The general update equation is thus

$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x}} \tilde{J}_n(\mathbf{x}).$$

A general discussion of convergence criteria for bound optimization methods would be outside the scope of this paper; the reader is referred to [21] and the references therein. Moreover, as mentioned in the introduction, the paper by Daubechies *et al.*

contains a convergence proof that can be readily extended to cover both algorithms presented in this paper.

### III. EXISTING THRESHOLDED LANDWEBER ALGORITHM

#### A. Bound With Decoupled Wavelet Coefficients

Daubechies *et al.* [11] proposed to use functionals of the form

$$\tilde{J}_n(\mathbf{x}) = \alpha \left\| \mathbf{x}^{(n)} - \mathbf{x} \right\|_2^2 + J(\mathbf{x}) - \left\| \mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x} \right\|_2^2. \quad (8)$$

Here, the (real and positive) scalar  $\alpha$  must be chosen strictly larger than the spectral radius of  $\mathbf{H}^T\mathbf{H}$

$$\alpha > \rho(\mathbf{H}^T\mathbf{H}) \quad \text{where} \quad \rho(\mathbf{H}^T\mathbf{H}) = \max_{\|\mathbf{v}\|_2=1} \|\mathbf{H}\mathbf{v}\|_2^2.$$

Equivalently, since  $\mathbf{H}$  is a convolution matrix,  $\rho(\mathbf{H}^T\mathbf{H})$  is the largest squared modulus of the DFT coefficients of  $\mathbf{H}$ .

Let us discuss this choice in more details. First, the constraint on  $\alpha$  ensures that  $\tilde{J}_n(\mathbf{x})$  is a proper bound for  $J(\mathbf{x})$ .

*Property 1:* Assume that  $\alpha > \rho(\mathbf{H}^T\mathbf{H})$  holds in (8). Then  $\tilde{J}_n(\mathbf{x}) > J(\mathbf{x})$ , except at  $\mathbf{x} = \mathbf{x}^{(n)}$ , where  $\tilde{J}_n(\mathbf{x}) = J(\mathbf{x})$ .

*Proof:* The inequality  $\alpha > \rho(\mathbf{H}^T\mathbf{H})$  ensures that  $\alpha\mathbf{I} - \mathbf{H}^T\mathbf{H}$  is positive-definite. This means that, when  $\mathbf{x} \neq \mathbf{x}^{(n)}$ , we have

$$\begin{aligned} \alpha \left\| \mathbf{x}^{(n)} - \mathbf{x} \right\|_2^2 - \left\| \mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x} \right\|_2^2 \\ = \left( \mathbf{x}^{(n)} - \mathbf{x} \right)^T (\alpha\mathbf{I} - \mathbf{H}^T\mathbf{H}) \left( \mathbf{x}^{(n)} - \mathbf{x} \right) > 0 \end{aligned}$$

whence it follows that  $\tilde{J}_n(\mathbf{x})$  is a strict upper bound of  $J(\mathbf{x})$ , except at  $\mathbf{x} = \mathbf{x}^{(n)}$  where the equality is met. ■

Second, the specific form of (8) makes it easy to minimize  $\tilde{J}_n(\mathbf{x})$ , based on the following observation.

*Property 2:* Definition (8) is equivalent to

$$\tilde{J}_n(\mathbf{x}) = \alpha \left\| \mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T \left( \mathbf{y} - \mathbf{H}\mathbf{x}^{(n)} \right) - \mathbf{x} \right\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1 + c$$

where  $c$  is a constant that does not depend on  $\mathbf{x}$ .

*Proof:*

$$\begin{aligned} \tilde{J}_n(\mathbf{x}) &= \alpha \left\| \mathbf{x}^{(n)} - \mathbf{x} \right\|_2^2 + \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \\ &\quad - \left\| \mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x} \right\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1 \\ &= \alpha \left\| \mathbf{x}^{(n)} \right\|_2^2 - 2\alpha \Re \left\{ \mathbf{x}^T \mathbf{x}^{(n)} \right\} + \alpha \|\mathbf{x}\|_2^2 \\ &\quad + \|\mathbf{y}\|_2^2 - 2\Re \left\{ \mathbf{x}^T \mathbf{H}^T \mathbf{y} \right\} + \|\mathbf{H}\mathbf{x}\|_2^2 \\ &\quad - \|\mathbf{H}\mathbf{x}^{(n)}\|_2^2 + 2\Re \left\{ \mathbf{x}^T \mathbf{H}^T \mathbf{H}\mathbf{x}^{(n)} \right\} \\ &\quad - \|\mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1 \\ &= \alpha \|\mathbf{x}\|_2^2 - 2\alpha \\ &\quad \times \Re \left\{ \mathbf{x}^T \left( \mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T \left( \mathbf{y} - \mathbf{H}\mathbf{x}^{(n)} \right) \right) \right\} \\ &\quad + \lambda \|\mathbf{W}\mathbf{x}\|_1 + c' \end{aligned}$$

where  $c' = \alpha \|\mathbf{x}^{(n)}\|_2^2 + \|\mathbf{y}\|_2^2 - \|\mathbf{H}\mathbf{x}^{(n)}\|_2^2$  is a constant that does not depend on  $\mathbf{x}$ . We complete the proof by adding the constant  $\alpha \|\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})\|_2^2$  (which does not depend on  $\mathbf{x}$ ) so as to complete the quadratic term. ■

The above derivation reveals that the term  $\|\mathbf{H}\mathbf{x}\|_2^2$ , which hinders the direct minimization of  $J(\mathbf{x})$ , is canceled by the negative term in (8). As a result,  $\mathbf{x}$  is not premultiplied by  $\mathbf{H}$  anymore in the expression of Property 2. This means that the wavelet coefficients of  $\mathbf{x}$  are now completely decoupled and the minimization of  $\tilde{J}_n(\mathbf{x})$  involves a simple coefficient-wise operation in the wavelet domain.

#### B. Resulting Algorithm

To make this more apparent, we will denote by  $\mathbf{z}^{(n)}$  the expression  $\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$ ; note that this is exactly the update formula of the classical Landweber iteration (4), with step size  $\tau = \alpha^{-1}$ . Omitting  $c$  and dividing by the (positive) constant  $\alpha$ , the minimization of  $\tilde{J}_n(\mathbf{x})$  thus reduces to the minimization of

$$\left\| \mathbf{z}^{(n)} - \mathbf{x} \right\|_2^2 + \lambda\tau \|\mathbf{W}\mathbf{x}\|_1.$$

Now this is almost exactly the pure denoising functional of (5), with  $\mathbf{z}^{(n)}$  as the noisy signal and  $\lambda\tau$  as the regularization parameter. We have already seen that its minimizer is obtained via a simple soft-thresholding operation on the wavelet coefficients of  $\mathbf{z}^{(n)}$ . In the present case, the threshold level will be  $\lambda\tau/2$ .

To summarize, the resulting TL algorithm alternates between the following two steps, starting from an initial estimate  $\mathbf{x}^{(0)}$ :

- compute the Landweber iteration  $\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \tau\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$ , with step-size  $\tau$ ;
- perform the wavelet-domain denoising operation  $\mathbf{x}^{(n+1)} = \mathbf{W}^T \mathcal{T}_{\lambda\tau/2} \{ \mathbf{W}\mathbf{z}^{(n)} \}$ , with threshold level  $\lambda\tau/2$ .

### IV. FAST THRESHOLDED LANDWEBER ALGORITHM

Using the algorithm described above, Figueiredo and Nowak have reported numerical results that are competitive with state-of-the-art 2-D deconvolution methods [14]. However, the convergence of this algorithm can be rather slow, especially when one chooses a small regularization parameter  $\lambda$ . This is typically the case for higher signal-to-noise ratios (see the numerical examples in Section V), which still require regularization, but at a moderate level.

Starting from this section, we will assume that the regularization term of (2) is expressed in the Shannon wavelet basis (see [22, p. 223] or [31, p. 51]). This wavelet family allows for larger (subband-specific) step sizes and threshold levels, resulting in a significant speed-up.

#### A. Shannon Wavelet Basis

Besides being orthonormal, the main characteristic of Shannon wavelets is that their spectrum is ideal and disjoint across subbands. This proves to be especially convenient when dealing with convolution operators.

Let us denote by  $(W_j)_{j \in S}$  the different wavelet subspaces.<sup>2</sup> More precisely, our convention will be that there are  $j_{\max}$  wavelet subspaces, and that  $j = j_{\max} + 1$  corresponds to the coarsest-scale scaling function subspace. In other words, the indexing set is  $S = \{1, \dots, j_{\max} + 1\}$ . Fig. 3 illustrates the frequency support of these subspaces in the 1 D case, where  $j_{\max}$  corresponds to the number of scales of the decomposition. For

<sup>2</sup>We use the terms subband and subspace interchangeably.

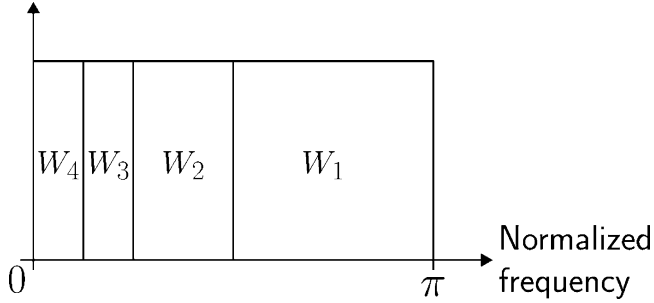


Fig. 3. Frequency support of the Shannon-wavelet subspaces in the 1-D case. The figure corresponds to  $j_{\max} = 3$ , which means that  $W_4$  represents the scaling function subspace.

higher dimensions, we use a separable extension of the Shannon wavelet basis. Note that in our finite-dimensional setting, Hermitian symmetry is violated at the mid frequency point, which implies that our basis elements are complex-valued.

We use the (boldface) symbol  $\mathbf{W}_j$  to denote the decomposition matrix that yields the coefficients of a signal in subspace  $W_j$ . The original signal can be reconstructed from its wavelet coefficients using the formula

$$\mathbf{x} = \sum_{j \in S} \underbrace{\mathbf{W}_j^T \mathbf{W}_j}_{\mathbf{P}_j} \mathbf{x}. \quad (9)$$

Note that we can use the transpose of  $\mathbf{W}_j$  because the Shannon wavelet basis is orthonormal. In this equation,  $\mathbf{W}_j^T \mathbf{W}_j \mathbf{x}_j$  represents the orthogonal projection of  $\mathbf{x}$  on  $W_j$ .

The following result states that, for the Shannon wavelet basis, the projection matrix  $\mathbf{P}_j = \mathbf{W}_j^T \mathbf{W}_j$  commutes with any convolution matrix.

*Property 3:* Let  $\mathbf{C}$  be a block-circulant matrix. Then, for the Shannon wavelet basis

$$\mathbf{C}\mathbf{P}_j = \mathbf{P}_j\mathbf{C}.$$

*Proof:* Let  $\mathbf{v}$  be an arbitrary vector. Applying (9) to  $\mathbf{C}\mathbf{P}_j\mathbf{v}$  yields

$$\mathbf{C}\mathbf{P}_j\mathbf{v} = \sum_{\ell \in S} \mathbf{P}_\ell \mathbf{C}\mathbf{P}_j\mathbf{v}.$$

Now recall that, with the Shannon wavelet basis, the subspaces  $(W_j)_{j \in S}$  have disjoint frequency supports. Because  $\mathbf{C}$  does not modify the frequency support of a signal (since it is diagonalized in the DFT basis), it must be that  $\mathbf{C}\mathbf{P}_j\mathbf{v} \in W_j$ . Therefore,  $\mathbf{P}_\ell \mathbf{C}\mathbf{P}_j\mathbf{v} = \mathbf{0}$  for  $\ell \neq j$  and

$$\mathbf{C}\mathbf{P}_j\mathbf{v} = \mathbf{P}_j \mathbf{C}\mathbf{P}_j\mathbf{v} = \mathbf{P}_j \mathbf{C} \left( \mathbf{v} - \sum_{\ell \in S \setminus \{j\}} \mathbf{P}_\ell \mathbf{v} \right).$$

From this relation, we can reapply the same argument to obtain  $\mathbf{C}\mathbf{P}_j\mathbf{v} = \mathbf{P}_j \mathbf{C}\mathbf{v}$  for every  $\mathbf{v}$ . This is equivalent to the commutativity of  $\mathbf{C}$  and  $\mathbf{P}_j$ . ■

## B. Derivation of a Subband-Adapted Bound

Our algorithm is based on the idea of cutting the cost function into subband-specific terms, thanks to the above commutativity property.

*Property 4:* When using the Shannon wavelet basis

$$J(\mathbf{x}) = \sum_{j \in S} \|\mathbf{P}_j \mathbf{y} - \mathbf{H}\mathbf{P}_j \mathbf{x}\|_2^2 + \lambda \|\mathbf{W}_j \mathbf{x}\|_1$$

where  $\mathbf{P}_j$  is the projection operator on the  $j$ th subband.

*Proof:* Using (9), we first observe that

$$\mathbf{y} - \mathbf{H}\mathbf{x} = \sum_{j \in S} \mathbf{P}_j \mathbf{y} - \mathbf{H}\mathbf{P}_j \mathbf{x}.$$

Property 3 tells us that we can rewrite the components of this sum as  $\mathbf{P}_j(\mathbf{y} - \mathbf{H}\mathbf{x})$ , which shows that they are mutually orthogonal (since the subspaces  $(W_j)_{j \in S}$  are orthogonal). This implies that we can separate the data term as follows:

$$\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 = \sum_{j \in S} \|\mathbf{P}_j \mathbf{y} - \mathbf{H}\mathbf{P}_j \mathbf{x}\|_2^2.$$

Combining this relation with the fact that  $\|\mathbf{W}\mathbf{x}\|_1 = \sum_{j \in S} \|\mathbf{W}_j \mathbf{x}\|_1$  yields the desired result. ■

The above result allows us to bound the cost function in a subband-specific manner. The idea is to apply the bound of Daubechies *et al.* to the individual terms of the sum in Property 4, leading to the following type of auxiliary functional:

$$\begin{aligned} \tilde{J}_n(\mathbf{x}) = J(\mathbf{x}) + \sum_{j \in S} \alpha_j \left\| \mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{P}_j \mathbf{x} \right\|_2^2 \\ - \left\| \mathbf{H}\mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{H}\mathbf{P}_j \mathbf{x} \right\|_2^2. \quad (10) \end{aligned}$$

The fundamental difference is that we now have subband-dependent constants  $(\alpha_j)_{j \in S}$ , which can be chosen significantly smaller than the constant  $\alpha$  of the standard algorithm. More precisely, for the sum in (10) to be positive, it is sufficient that, for every  $j$

$$\begin{aligned} \alpha_j > \rho_j(\mathbf{H}^T \mathbf{H}) \\ \text{where } \rho_j(\mathbf{H}^T \mathbf{H}) = \max_{\|\mathbf{v}\|_2=1, \mathbf{v} \in W_j} \|\mathbf{H}\mathbf{v}\|_2^2. \end{aligned}$$

We point out that the definition of  $\rho(\mathbf{H}^T \mathbf{H})$  (a squared matrix norm) is similar to that of  $\rho_j(\mathbf{H}^T \mathbf{H})$ . However, in the latter, the maximum is taken over subspace  $W_j$ . Therefore, we necessarily have

$$\rho_j(\mathbf{H}^T \mathbf{H}) \leq \rho(\mathbf{H}^T \mathbf{H}).$$

Equivalently, since  $\mathbf{H}$  is a convolution matrix,  $\rho_j(\mathbf{H}^T \mathbf{H})$  is the largest squared modulus of the DFT coefficients of  $\mathbf{H}$ , over the frequency support of subband  $W_j$ . This value can be much smaller than  $\rho(\mathbf{H}^T \mathbf{H})$  (the largest squared modulus over the whole spectrum). Fig. 4 compares choices for  $\alpha$  and  $(\alpha_j)_{j \in S}$  in a simplified 1-D situation.

When the previous condition is met,  $\tilde{J}_n(\mathbf{x})$  upper-bounds  $J(\mathbf{x})$ . Still,  $\tilde{J}_n(\mathbf{x})$  is easy to minimize (thanks to the commu-

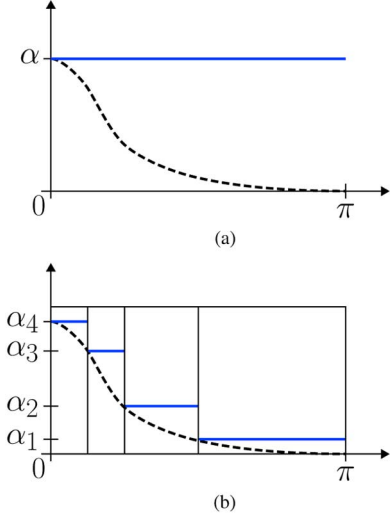


Fig. 4. Comparison of the constants  $\alpha$  and  $(\alpha_j)_{j \in S}$  for a fictitious 1-D low-pass convolution kernel. The dashed curve represents the squared modulus of its Fourier transform.

tativity stated in Property 3). This transposition of Properties 1 and 2 to our new auxiliary functional can be summarized as follows:

*Property 5:* Assume that  $\alpha_j > \rho_j(\mathbf{H}^T \mathbf{H})$  holds for every  $j \in S$  in definition (10). Then  $\tilde{J}_n(\mathbf{x}) > J(\mathbf{x})$ , except at  $\mathbf{x} = \mathbf{x}^{(n)}$ , where  $\tilde{J}_n(\mathbf{x}) = J(\mathbf{x})$ . Moreover, (10) is equivalent to

$$\tilde{J}_n(\mathbf{x}) = \sum_{j \in S} \alpha_j \left\| \mathbf{P}_j \left[ \mathbf{x}^{(n)} + \alpha_j^{-1} \mathbf{H}^T \left( \mathbf{y} - \mathbf{H} \mathbf{x}^{(n)} \right) \right] - \mathbf{P}_j \mathbf{x} \right\|_2^2 + \lambda \|\mathbf{W}_j \mathbf{x}\|_1 + c_j$$

where the constants  $(c_j)_{j \in S}$  do not depend on  $\mathbf{x}$ .

*Proof:* The definition of  $\rho_j(\mathbf{H}^T \mathbf{H})$  implies that

$$\alpha_j \left\| \mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{P}_j \mathbf{x} \right\|_2^2 - \left\| \mathbf{H} \mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{H} \mathbf{P}_j \mathbf{x} \right\|_2^2 \geq 0$$

for every  $j \in S$ . When  $\mathbf{x} \neq \mathbf{x}^{(n)}$ , there is at least one  $j$  for which this inequality is strict. If  $\mathbf{x} = \mathbf{x}^{(n)}$ , the equality is clear. This proves the first part of the property.

Applying Property 2 to every term of the sum in (10) yields

$$\tilde{J}_n(\mathbf{x}) = \sum_{j \in S} \alpha_j \left\| \mathbf{P}_j \mathbf{x}^{(n)} + \alpha_j^{-1} \mathbf{H}^T \left( \mathbf{P}_j \mathbf{y} - \mathbf{H} \mathbf{P}_j \mathbf{x}^{(n)} \right) - \mathbf{P}_j \mathbf{x} \right\|_2^2 + \lambda \|\mathbf{W}_j \mathbf{x}\|_1 + c_j.$$

Since we are using Shannon wavelets, we can use the commutativity of  $\mathbf{P}_j$  with the convolution matrices  $\mathbf{H}$  and  $\mathbf{H}^T$  to derive the second part of the property. ■

### C. Resulting Algorithm

Let us define the constants  $\tau_j = 1/\alpha_j$  and the intermediate signal

$$\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \sum_{j \in S} \tau_j \mathbf{P}_j \mathbf{H}^T \left( \mathbf{y} - \mathbf{H} \mathbf{x}^{(n)} \right).$$

Since the Shannon wavelet basis is orthonormal, we can replace  $\mathbf{P}_j$  by  $\mathbf{W}_j$  in the expression of Property 5. The minimization of  $\tilde{J}_n(\mathbf{x})$  then becomes equivalent to the minimization of

$$\sum_{j \in S} \alpha_j \left( \left\| \mathbf{W}_j \mathbf{z}^{(n)} - \mathbf{W}_j \mathbf{x} \right\|_2^2 + \lambda \tau_j \|\mathbf{W}_j \mathbf{x}\|_1 \right).$$

Each term of this sum depends on the wavelet coefficients of  $\mathbf{x}$  in a different subband. As a consequence, we can minimize them independently. Moreover, the individual terms have the same form as in (6), up to a (positive) multiplicative constant. Our new auxiliary functionals can thus be minimized using essentially the same algorithm as before—with the same cost per iteration—but with subband-dependent parameters.

The resulting “fast thresholded Landweber” (FTL) algorithm alternates between the following two steps.

- Compute the Landweber iteration

$$\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \sum_{j \in S} \tau_j \mathbf{P}_j \mathbf{H}^T \left( \mathbf{y} - \mathbf{H} \mathbf{x}^{(n)} \right)$$

with subband-dependent step sizes  $\tau_j$ .

- Perform the wavelet-domain denoising operation

$$\mathbf{x}^{(n+1)} = \sum_{j \in S} \mathbf{W}_j^T \mathcal{T}_{\lambda \tau_j / 2} \left\{ \mathbf{W}_j \mathbf{z}^{(n)} \right\}$$

with subband-dependent thresholds  $\lambda \tau_j / 2$ .

## V. NUMERICAL EXPERIMENTS

### A. Implementation Details and Computational Cost

The FTL algorithm is simple to implement, for it essentially consists in going back and forth between the frequency domain and the wavelet domain.

In the first step, one must start by computing the Landweber correction term  $\mathbf{H}^T (\mathbf{y} - \mathbf{H} \mathbf{x}^{(n)})$ . This is done in the DFT basis, where  $\mathbf{H}$  is diagonalized. The actual update should *a priori* be performed in the wavelet domain, since the step sizes are subband-dependent. However, in the particular case of a Shannon wavelet basis, it can be carried out in the frequency domain: it is equivalent to multiplying each frequency component of the correction term by  $\tau_j$ , where  $j$  refers to the wavelet subband that contains the considered frequency.

As described above, the second step just requires the application of a wavelet transform to  $\mathbf{z}^{(n)}$ , followed by a soft-thresholding and an inverse wavelet transform of the result. However, in practice—for both algorithms presented in this paper—we use the random-shift method described in [14]. The main motivation is to reduce unpleasant artifacts that are common side-effects of thresholding operations in nonredundant wavelet bases. Still, we would like to do so without having to resort to a fully shift-invariant (undecimated) transform. The compromise consists—for every iteration—in applying a random (circular) shift to the intermediate estimate  $\mathbf{z}^{(n)}$ , before performing its wavelet decomposition. After the thresholding and the inverse transform, the new estimate is shifted back to the original position. By doing this, the effect of thresholding is “averaged out” over

multiple shifts during the iteration process. Note that a shift in the image domain corresponds to a simple modulation in the frequency domain. This means that the random-shift method can be used at no significant additional cost.

The computational cost is, therefore, essentially the same as for the classical thresholded Landweber algorithm, which also requires two (one direct and one inverse) wavelet transforms per iteration. We use a frequency-domain implementation of the wavelet transform [7]. Thus, every iteration essentially amounts to computing two FFTs, which is no more than the standard algorithms mentioned in the introduction.

One last remark relating to the scaling-function subspace: to simplify the presentation, we have not distinguished it from the wavelet subspaces so far. However, the scaling-function coefficients are usually not included in the regularization term, which means that they are not thresholded in practice.

*B. Evaluation Protocol*

To compare the performance of the thresholded Landweber (TL) and the fast thresholded Landweber (FTL) algorithms, we performed several experiments on synthetic and real data. The literal way of assessing the convergence speedup of our algorithm would be to compare the intermediate estimates to the true minimizer of the cost function. Generally, however, this true minimizer is not available and its determination would involve an overwhelming amount of computations. Therefore, following the common practice in the field, we will use the quadratic error between the intermediate estimates and the original image (gold-standard). This is justified from at least two standpoints.

- The user standpoint: Ultimately, what counts in practice is the difference between the deconvolution result and the original signal.
- The theoretical standpoint: The true minimizer can be shown to be “close” to the original signal in the following sense. When the noise level and the regularization parameter tend to zero, the minimizer of the cost function tends to the original signal (see the paper by Daubechies *et al.* [11] for a rigorous formulation and proof).

For each experiment, we used the same regularization parameter for the TL and FTL algorithms. For the phantom experiments, this value was optimized to yield the result closest to the original signal after a large number of FTL iterations (so as to approach the true minimizer reasonably well). In the real data cases, it was adjusted empirically for best visual appearance.

*C. Visual Comparison in 2-D: Results for a Limited Number of Iterations*

In the first experiment, we took a  $512 \times 512$  image of a neuron cell acquired on a confocal microscope as our (ground-truth) original image. We then simulated an optical defocusing blur produced by a widefield microscope. We used a standard diffraction-limited PSF model [15], [33], which is represented in Fig. 5(b). Gaussian white noise was added to the result; the variance  $\sigma^2$  was set such that the *blurred signal-to-noise ratio* (BSNR) was equal to 40 dB. The BSNR is used to quantify the noise level with respect to the original signal strength, while

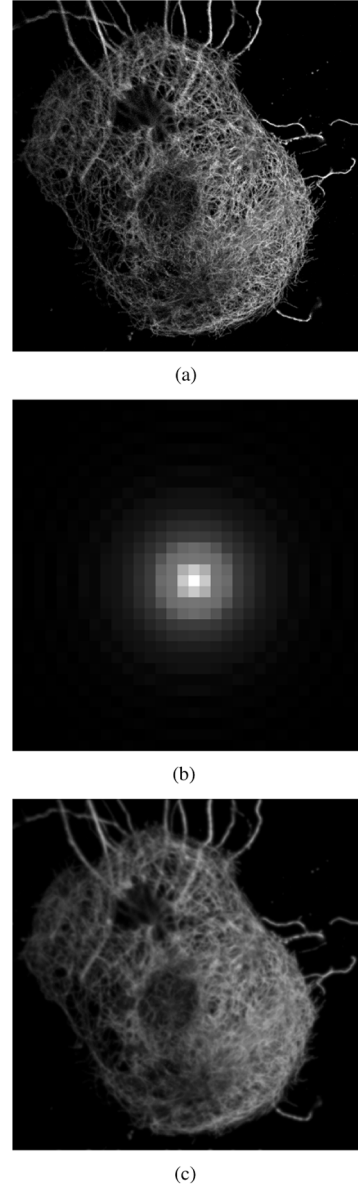


Fig. 5. (a) Original confocal image (courtesy of Bertrand Vilen, LNNME, EPFL) and (c) the simulated widefield image of Section V-C. The square root of the PSF is shown in (b), with a  $16 \times$  zoom factor compared to (a) and (c). (a) Original. (b) PSF. (c) Measured.

taking into account the attenuation introduced by the simulated imaging device. It is defined as

$$BSNR = 10 \log_{10} \left( \frac{\|\mathbf{H}\mathbf{x}_{orig}\|_2^2 - N \text{mean}(\mathbf{H}\mathbf{x}_{orig})^2}{N\sigma^2} \right).$$

Here,  $N$  is the total number of pixels per image and  $\text{mean}(\mathbf{H}\mathbf{x}_{orig})$  stands for the average gray level of the blurred original image.

Fig. 5 shows the original image (a) and the simulated measurement (c). Estimates at different stages of the TL and FTL iterations are displayed in Fig. 6. For both algorithms, we used the same initial estimate (the measured image  $\mathbf{y}$ ), the same number of decomposition levels ( $j_{max} = 5$ ) and the same regularization parameter. It is seen that the images obtained after ten or 30 iterations of the TL algorithm are less sharp than the one obtained

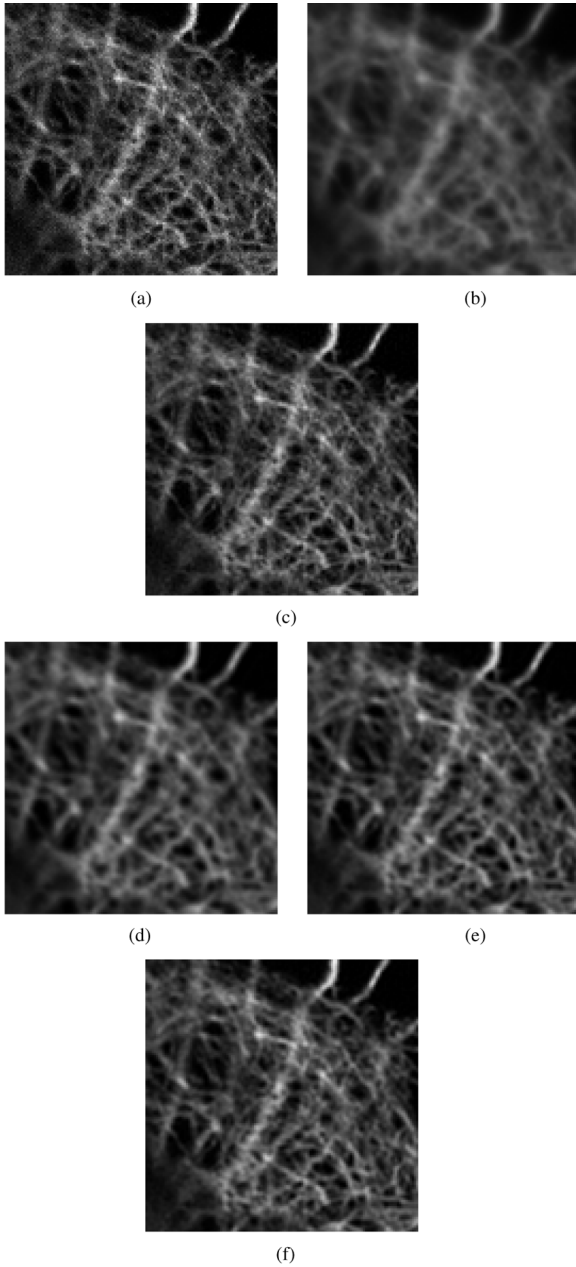


Fig. 6. Deconvolution results for the setup of Fig. 5: comparison on a  $128 \times 128$  region of interest. (a) Original. (b) Measured. (c) FTL, ten iterations. (d) TL, ten iterations. (e) TL, 30 iterations. (f) TL, 100 iterations.

with ten iterations of the FTL algorithm. This is especially true for the details inside the object which exhibit better contrast than in the latter. It is only after a larger number of iterations (100) that the output of the TL algorithm gets closer to the FTL result with ten iterations, at least from a visual standpoint. We did not include the results of the FTL algorithm with more than ten iterations, since the differences are not perceptible visually because the algorithm has already essentially reached convergence.

#### D. Quantitative Comparison in 2-D: Number of Iterations to Reach a Given Level of SER Gain

The above results strongly suggest that the FTL algorithm requires fewer iterations than the TL algorithm to reach a given

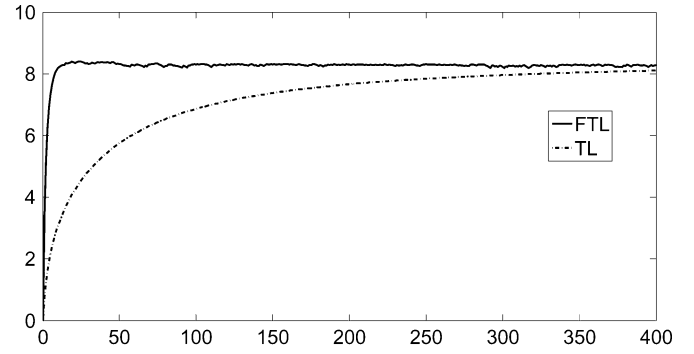


Fig. 7. SER gain as a function of the iteration number for the experiment of Section V-C.

level of restoration quality. Fig. 7 gives a quantitative insight: it shows the evolution of the *signal-to-error ratio* (SER) as a function of the iteration number. For an estimate  $\mathbf{x}$ , the SER is defined as

$$\text{SER}(\mathbf{x}) = 10 \log_{10} \left( \frac{\|\mathbf{x}_{\text{orig}}\|_2^2}{\|\mathbf{x} - \mathbf{x}_{\text{orig}}\|_2^2} \right).$$

For each estimate  $\mathbf{x}^{(n)}$ , we compare this figure to the SER of the measured signal, leading to the SER *gain*

$$\text{SERG}(\mathbf{x}^{(n)}) = \text{SER}(\mathbf{x}^{(n)}) - \text{SER}(\mathbf{y}).$$

One can observe that the FTL algorithm requires less than ten iterations to reach an improvement of 8 dB. This is roughly thirty times less than the TL algorithm, which takes about 300 iterations to reach the same level.

We performed a series of more extensive experiments on four standard test-images (Cameraman, MRI, House, Bird). We convolved these images with a  $9 \times 9$  uniform blur kernel and added white Gaussian noise to the results in order to replicate the experimental setup of Figueiredo and Nowak. We considered different noise levels (BSNR = 10, 20, 30, 40, 50 dB) and we averaged the SER gains over 30 noise realizations. For each test case, we used the same value of  $\lambda$  and the same initial estimate for both algorithms. The value of  $\lambda$  was optimized to yield the best possible SER gain after 300 iterations of the FTL algorithm. The initial estimate was obtained using the same Wiener-type filter as Figueiredo and Nowak [14]

$$\mathbf{x}^{(0)} = (\mathbf{H}^T \mathbf{H} + 10^{-3} \sigma^2 \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y}.$$

The results are summarized in Table I. For each image, we indicate the SER gain of the FTL algorithm after ten and 30 iterations. We also give the number of iterations of the TL algorithm required to reach the same SER improvement. The corresponding acceleration factors (number of TL iterations, divided by number of FTL iterations) are listed in the fifth and eighth column. The acceleration factors vary between 1.5 to 4.7 for low BSNR levels (10–20 dB) and 32.3 to 191.5 for high BSNR levels (40–50 dB).

Note that the accelerated algorithm that has been derived here is specific to Shannon wavelets and is not directly transposable



TABLE I  
NUMBER OF ITERATIONS REQUIRED TO REACH A GIVEN LEVEL OF SER GAIN (SERG)

|           | BSNR | SERGFTL  | Number of |               | SERGFTL  | Number of |               |
|-----------|------|----------|-----------|---------------|----------|-----------|---------------|
|           |      | 10 iter. | TL iter.  | Accel. factor | 30 iter. | TL iter.  | Accel. factor |
| Cameraman | 10   | 2.90     | 36        | 3.6           | 2.94     | 61        | 2.0           |
|           | 20   | 2.62     | 39        | 3.9           | 2.74     | 87        | 2.9           |
|           | 30   | 4.31     | 121       | 12.1          | 4.43     | 172       | 5.7           |
|           | 40   | 6.03     | 541       | 54.1          | 6.61     | 972       | 32.4          |
|           | 50   | 7.80     | 1876      | 187.6         | 8.38     | > 2000    | > 66.7        |
| MRI       | 10   | 3.68     | 40        | 4.0           | 3.87     | 99        | 3.3           |
|           | 20   | 4.53     | 47        | 4.7           | 4.81     | 91        | 3.0           |
|           | 30   | 6.79     | 167       | 16.7          | 6.98     | 211       | 7.0           |
|           | 40   | 7.94     | 724       | 72.4          | 9.55     | 1344      | 44.8          |
|           | 50   | 9.53     | 1915      | 191.5         | 10.70    | > 2000    | > 66.7        |
| House     | 10   | 3.73     | 22        | 2.2           | 3.82     | 58        | 1.9           |
|           | 20   | 3.81     | 27        | 2.7           | 3.99     | 49        | 1.6           |
|           | 30   | 6.25     | 170       | 17.0          | 6.39     | 199       | 6.6           |
|           | 40   | 7.30     | 559       | 55.9          | 9.07     | 969       | 32.3          |
|           | 50   | 7.49     | 1616      | 161.6         | 8.90     | > 2000    | > 66.7        |
| Bird      | 10   | 4.65     | 20        | 2.0           | 4.70     | 45        | 1.5           |
|           | 20   | 3.67     | 34        | 3.4           | 3.81     | 51        | 1.7           |
|           | 30   | 5.76     | 180       | 18.0          | 5.93     | 216       | 7.2           |
|           | 40   | 5.75     | 555       | 55.5          | 8.24     | 1004      | 33.5          |
|           | 50   | 5.40     | 1562      | 156.2         | 7.12     | > 2000    | > 66.7        |

to other types of basis functions. Nevertheless, the results obtained for the Cameraman image are comparable to those reported by Figueiredo and Nowak in terms of restoration quality,<sup>3</sup> even though these authors used different wavelets than ours. In fact, they observed that the performance of their algorithm was only very mildly dependent on the choice of a particular type of wavelet. Still, the present results suggest that there should be ways of accelerating the convergence with other types of wavelets as well, which calls for further investigation.

#### E. Visual Comparison in 3-D: Experiment on Fluorescence Microscopy Data

For the last experiment, we used 3-D image-stacks of a biological sample. The sample contained fibroblast cells stained with a DiO dye from Invitrogen Corporation (Carlsbad, CA). This dye is predominantly retained in the cell membrane. In addition, fluorescent microbeads were introduced into the sample medium. The dye and the microbeads were chosen to have approximately the same peak excitation and emission wavelengths; a 505–530 nm bandpass filter was used to delimit the detection spectrum.

The sample was observed on a Zeiss LSM 510 confocal microscope with a 63 $\times$ , 1.4 NA oil-immersion objective. We first acquired a stack of images with the pinhole completely open. In

this configuration, the confocal effect is not used and the system becomes essentially equivalent to a widefield microscope. This results in images with out-of-focus blur due to the poor localization of the widefield PSF along the  $z$ -dimension. The second set of images was acquired with the pinhole radius set to 1 Airy unit. In this configuration, much of the out-of-focus light is rejected by the system, resulting in significantly sharper images. We then used the confocal data set as a reference for comparing the performance of the TL and FTL algorithms on the widefield data set.

Maximum intensity projections (along the  $z$  axis) of both data sets are shown in Fig. 8. Below are the results of ten iterations of the TL and FTL algorithms on the widefield stack; this corresponds to roughly 1 min and 30 s of computation time on a 2.66-GHz Intel Xeon workstation. We used a 3-D version of the diffraction-limited model mentioned above (with parameters corresponding to manufacturer specification: NA = 1.4,  $n_1 = 1.518$ ) to generate the PSF and we took the measured widefield image as the initial estimate.

The TL algorithm cannot produce a visible deconvolution effect within ten iterations; its output is very similar to the original widefield data. On the other hand, the FTL algorithm rapidly produces an estimate that is significantly sharper. In particular, the thickness of the cell membranes is comparable to the confocal image. Also, the fluorescent microbeads are brighter than in the widefield image. The results were shown to biologists who were positively impressed.

<sup>3</sup>Table I of [14] indicates a gain of 6.33 dB at 40 dB BSNR, while our experiments yielded 6.03 and 6.61 dB after ten and 30 iterations of the FTL algorithm, respectively.

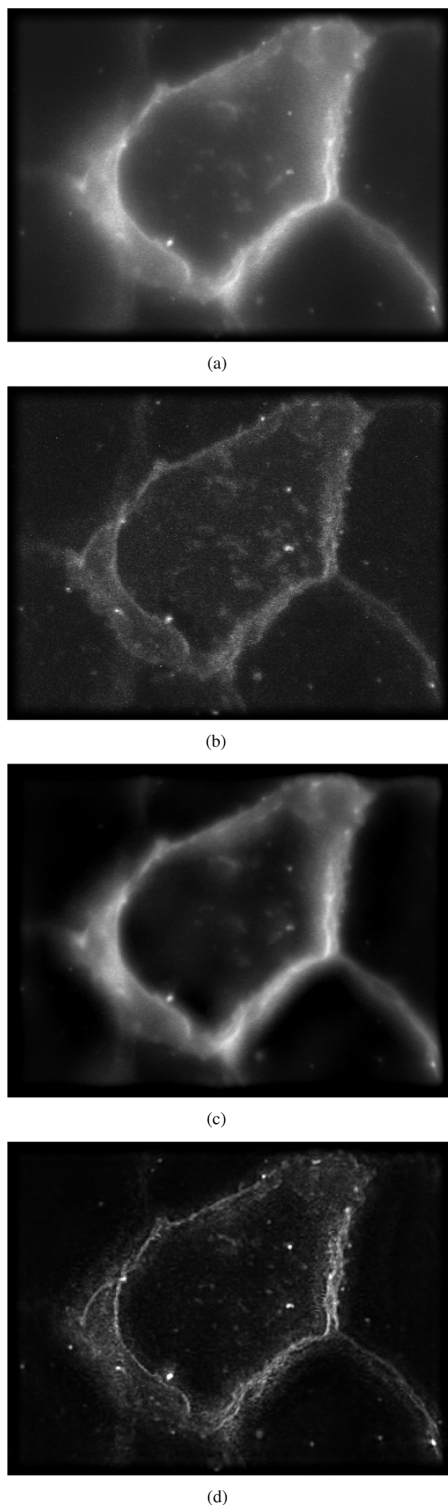


Fig. 8. Maximum intensity projections of  $384 \times 288 \times 16$  image stacks. (a) Widefield image and (b) confocal reference image of the same object. (c), (d) Results of the deconvolution algorithms applied to the widefield image: (c) TL, ten iterations; (d) FTL, ten iterations.

## VI. CONCLUSION

We have presented a fast algorithm for wavelet-regularized variational deconvolution. It is based on the use of a Shannon wavelet basis, which allows for a subband-adapted minimization strategy. As a result, the update introduced by each step of

the FTL algorithm is much more efficient than with the standard TL algorithm. For BSNR levels above 30 dB, we have obtained acceleration factors of one order of magnitude or better. This corresponds to relatively standard acquisition conditions for deconvolution microscopy, at least when considering fixed specimens.<sup>4</sup> Thanks to this substantial speed increase, we have demonstrated the feasibility of 3-D wavelet-based deconvolution microscopy on a widefield fluorescence image stack.

A direct extension of the proposed algorithm is to consider subband-dependent regularization parameters [11] and/or other sparsity-enforcing penalizations (that is, other thresholding functions, see [14]). The principle of the method could also be applied to other types of operators that are block-diagonal in the basis that is used for regularization. The possibilities of this type of procedure are numerous and our current experiments suggest that there is room for improvement, which calls for further investigations and more extensive comparative studies.

## ACKNOWLEDGMENT

The authors would like to thank M. Kropf (Cellular Neurobiology Laboratory, EPFL) for preparing the sample presented in Section V, as well as T. Laroche (BI-Op, EPFL) for his valuable help in acquiring the data sets. They would also like to thank B. Vilen (now with the Institute of Molecular Biophysics, FSU) for providing the image used for Figs. 5 and 6.

## REFERENCES

- [1] D. A. Agard and J. W. Sedat, "Three-dimensional architecture of a polytene nucleus," *Nature*, vol. 302, no. 5910, pp. 676–681, Apr. 1983.
- [2] A. Antoniadis and J. Fan, "Regularization of wavelet approximations," *J. Amer. Statist. Assoc.*, vol. 96, no. 455, pp. 939–967, Sep. 2001.
- [3] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Process. Mag.*, vol. 14, no. 2, pp. 24–41, Mar. 1997.
- [4] J. Beet, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A  $\ell^1$ -unified variational framework for image restoration," in *Proc. ECCV*, 2004, vol. 3024, pp. 1–13.
- [5] M. Bertero and P. Boccacci, "Introduction to inverse problems in imaging," *Institute of Physics Publishing*, 1998.
- [6] E. Betzig, G. H. Patterson, R. Sougrat, O. Wolf-Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, Sept. 2006.
- [7] T. Blu and M. Unser, "The fractional spline wavelet transform: Definition and implementation," in *Proc. 25th IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Jun. 2000, vol. 1, pp. 512–515.
- [8] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, "Non-linear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 319–335, Mar. 1998.
- [9] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inv. Probl.*, vol. 23, no. 4, pp. 1495–1518, Aug. 2007.
- [10] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: An overview," *IEEE Trans. Consum. Electron.*, vol. 46, no. 4, pp. 1103–1127, Nov. 2000.
- [11] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Aug. 2004.
- [12] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Appl. Comput. Harmon. Anal.*, vol. 2, pp. 101–126, 1995.

<sup>4</sup>Current cooled CCD cameras are capable of much wider dynamic ranges of 12 or even 16 bits per pixel (1 bit  $\approx$  6 dB).

- [13] D. D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Aug. 1994.
- [14] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
- [15] M. Gu, *Principles of Three-Dimensional Imaging in Confocal Microscopes*. Singapore: World Scientific, Dec. 1996.
- [16] M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, "T<sup>5</sup>M: 3D widefield light microscopy with better than 100 nm axial resolution," *J. Microsc.*, vol. 195, no. 1, pp. 10–16, Jul. 1999.
- [17] R. Heintzmann, T. M. Jovin, and C. Cremer, "Saturated patterned excitation microscopy (SPEM)—A novel concept for optical resolution improvement," *J. Opt. Soc. Amer. A*, vol. 19, no. 8, pp. 1599–1609, Aug. 2002.
- [18] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.
- [19] J. Kalifa and S. Mallat, "Thresholding estimators for linear inverse problems and deconvolutions," *Ann. Statist.*, vol. 31, no. 1, pp. 58–109, 2003.
- [20] L. Landweber, "An iterative formula for Fredholm integral equations of the first kind," *Amer. J. Math.*, vol. 73, no. 3, pp. 615–624, Jul. 1951.
- [21] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *J. Comput. Graph. Statist.*, vol. 9, no. 1, pp. 1–20, Mar. 2000.
- [22] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
- [23] J. G. McNally, T. Karpova, J. Cooper, and J. A. Conchello, "Three-dimensional imaging by deconvolution microscopy," *Methods*, vol. 19, no. 3, pp. 373–385, Nov. 1999.
- [24] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized gaussian and complexity priors," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 909–919, Apr. 1999.
- [25] M. Nagorni and S. W. Hell, "Coherent use of opposing lenses for axial resolution increase II. Power and limitation of nonlinear image restoration," *J. Opt. Soc. Amer. A*, vol. 18, no. 1, pp. 49–54, Jan. 2001.
- [26] R. Neelamani, H. Choi, and R. Baraniuk, "Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 418–433, Feb. 2004.
- [27] P. Sarder and A. Nehorai, "Deconvolution methods for 3-D fluorescence microscopy images," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 32–45, May 2006.
- [28] E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models Volume 141 of Lecture Notes in Statistics*. New York: Springer-Verlag, Jun. 1999, pp. 292–308.
- [29] J. L. Starck, D. L. Donoho, and E. J. Candes, "Astronomical image representation by the curvelet transform," *Astron. Astrophys.*, vol. 398, no. 2, pp. 785–800, Feb. 2003.
- [30] J.-L. Starck, M. K. Nguyen, and F. Murtagh, "Wavelets and curvelets for image deconvolution: A combined approach," *Signal Process.*, vol. 83, no. 10, pp. 2279–2283, Oct. 2003.
- [31] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Cambridge, MA: Wellesley-Cambridge, 1996.
- [32] C. Vogel, "Computational methods for inverse problems," *SIAM*, 2002.
- [33] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 20–31, May 2006.
- [34] W. Wallace, L. H. Schaefer, and J. R. Swedlow, "A working person's guide to deconvolution in light microscopy," *Biotechniques*, vol. 31, no. 5, pp. 1076–1097, Nov. 2001.
- [35] J. B. Weaver, Y. S. Xu, D. M. Healy, and L. D. Cromwell, "Filtering noise from images with wavelet transforms," *Magn. Res. Med.*, vol. 21, no. 2, pp. 288–295, Oct. 1991.



**Cédric Vonesch** (S'05) was born on April 14, 1981 in Obernai, France. He spent one year studying at the Ecole Nationale Supérieure des Télécommunications, Paris, France, and graduated from the Swiss Institute of Technology, Lausanne (EPFL), Switzerland, in 2004. He is currently pursuing the Ph.D. degree at the Biomedical Imaging Group, EPFL.

His research interests include multiresolution and wavelet analysis, inverse problems, and applications to bioimaging.



**Michael Unser** (M'89–SM'94–F'99) received the M.S. (summa cum laude) and Ph.D. degrees in electrical engineering from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 1981 and 1984, respectively.

From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda, MD. He is now Professor and Director of the Biomedical Imaging Group, EPFL. His main research topics are biomedical image processing, splines, and wavelets. He is

the author of over 120 published journal papers in these areas.

Dr. Unser has been actively involved with the IEEE TRANSACTIONS ON MEDICAL IMAGING, holding the positions of Associate Editor (1999–2002, 2006–present), member of steering committee, and Associate Editor-in-Chief (2003–2005). He has acted as an Associate Editor or member of the editorial board for eight more international journals, including the *IEEE Signal Processing Magazine*, the IEEE TRANSACTIONS ON IMAGE PROCESSING (1992–1995), and the IEEE SIGNAL PROCESSING LETTERS (1994–1998). He organized the first IEEE International Symposium on Biomedical Imaging (ISBI 2002). He currently chairs the technical committee of the IEEE-SP Society on Bio Imaging and Signal Processing (BISP), and well as the ISBI steering committee. He is the recipient of three Best Paper Awards from the IEEE Signal Processing Society.