# A fast multilevel algorithm for wavelet-regularized image restoration

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

**Abstract**

We present a multi-level extension of the popular "thresholded Landweber" algorithm for wavelet-regularized image restoration that yields an order of magnitude speed improvement over the standard fixed-scale implementation. The method is generic and targeted towards large-scale linear inverse problems, such as 3D deconvolution microscopy.

The algorithm is derived within the framework of bound optimization. The key idea is to successively update the coefficients in the various wavelet channels using fixed, subband-adapted iteration parameters (step sizes and threshold levels). The optimization problem is solved efficiently via a proper chaining of basic iteration modules. The higher-level description of the algorithm is similar to that of a multigrid solver for PDEs, but there is one fundamental difference: the latter iterates though a sequence of multiresolution versions of the original problem, while, in our case, we cycle through the wavelet subspaces corresponding to the difference between successive approximations. This strategy is motivated by the special structure of the problem and the preconditioning properties of the wavelet representation. We establish that the solution of the restoration problem corresponds to a fixed point of our multilevel optimizer. We also provide experimental evidence that the improvement in convergence rate is essentially determined by the (unconstrained) linear part of the algorithm, irrespective of the type of wavelet. Finally, we illustrate the technique with some image deconvolution examples, including some real 3D fluorescence microscopy data.

**Keywords:** inverse problems, variational, reconstruction, restoration, deconvolution, non-linear, sparsity, $\ell_1$-regularization, wavelets, multiresolution, multiscale, multilevel, multigrid, fast, convergence acceleration, preconditioning, bound optimization, majorize-minimize, optimization transfer, surrogate optimization, 3D, microscopy, widefield, confocal, fluorescence.

The authors are with the Biomedical Imaging Group, EPFL, Lausanne, Switzerland. Contact information: Cédric Vonesch, EPFL-STI-IMT-LIB, BM 4.138 (Bâtiment BM), Station 17, CH-1015 Lausanne VD, Switzerland. Phone: +41 (0)21 693 51 36. Fax: +41 (0)21 693 37 01. Email: cedric.vonesch@epfl.ch, michael.unser@epfl.ch.

# I. Introduction

## A. Motivation and originality of the present work

Inverse problems arise in various imaging applications such as biomicroscopy [1], [2], medical imaging [3], [4] or astronomy [5], [6]. An increasingly important issue for reconstruction and restoration tasks is the mass of data that is now routinely produced in these fields. The instrumentation typically allows for high-dimensional and multimodal imaging, fostering the evolution of experimental practices towards more quantitative and systematic investigations. This trend will arguably persist over the forthcoming years, and, as a result, computation time will remain a serious bottleneck for restoration methods, despite the progress of computer hardware. In this context, advanced (non-linear) restoration methods that were developed for traditional 2D imaging cannot be applied directly; larger-scale problems require more efficient algorithmic implementations.

The concept of "sparsity" has drawn considerable interest recently, leading to a new and successful paradigm for the regularization of inverse problems. The main idea is to constrain the restored image to have only a few non-zero coefficients in a suitable transform domain. Based on this principle, a simple and elegant iterative algorithm—which we shall call the "thresholded Landweber" (TL) algorithm—was independently derived by several research groups [7], [8], [9]. The method has the advantage of being very general. However, it is known to converge slowly when applied to ill-conditioned inverse problems [10], [11], [12], which restrains its suitability for large data sets.

In this paper, we construct a multilevel version of the TL algorithm that is significantly faster; this allows us to apply the method for the restoration of real 3D multichannel fluorescence micrographs. To do so, we specifically consider the case where the sparsity constraint is enforced in the wavelet domain, which was shown to yield state-of-the-art results for 2D image restoration (see [7]). From a numerical standpoint, the advantage of using wavelet representations is twofold. First, their *tree structure* naturally leads to efficient computational schemes in the spirit of Mallat's Fast Wavelet Transform [13]. Second, the *spectral localization* properties of wavelets make them suitable for preconditioning, that is, for partly compensating the poor conditioning of the inverse problem.

The structure and the convergence speed of our multilevel algorithm make it comparable to multigrid schemes [14], [15]. These schemes belong to the most efficient known methods for the numerical resolution of partial differential equations; they are typically one order of magnitude faster than standard iterative methods. In fact, the connection between wavelet and multigrid theory was recognized early on [16], [17], [18], [19]. Surprisingly, though, the potential of wavelet-based multilevel methods for image restoration has hardly been exploited so far. An exception is the paper by Wang et al. [20], which is however restricted to linear restoration and based on a relatively empirical reformulation of the image-formation model in the wavelet domain.

Our approach is based on a non-quadratic variational formulation (leading to a non-linear restoration

method) and on the principle of bound optimization [21], [22]. This principle also underlies the derivation of [8] and is known under several alternative denominations, such as optimization transfer, surrogate functional optimization or majorize-minimize (MM) strategy.

Our method can be related to the family of "block-alternating MM algorithms" [23]. In the context of statistical signal processing, one of the earliest representatives of this family is the "Space-Alternating Generalized EM" (SAGE) algorithm of Fessler and Hero [24]. More recently, a bound-optimization approach was also used by Oh et al. to derive a multigrid inversion method for non-linear problems [25]. While the works [23], [24], [25] do not involve wavelets, the latter can be related to the so-called lazy wavelet transform [26], which itself corresponds to the anterior concept of hierarchical basis in the finite element and multigrid literature [27], [28]. Similarly, our work can be related to generalizations of the hierarchical-basis method [17], [29].

To achieve our goal, we construct a family of bounds that allow us to divide the original variational problem into a collection of smaller problems, corresponding to the different scales of the wavelet decomposition. The bounds can be made particularly tight for specific subbands. This leads to subband-dependent iteration parameters (step sizes and threshold levels), which are the key to faster convergence. The bound optimization framework provides a rationale for choosing these parameters in a consistent manner. At the same time, this framework is simple to deploy and guarantees that the underlying cost functional is monotonically decreased.

### *B. Image-formation model*

We will be concerned with the recovery of signals that are distorted by a linear measurement device and noise. Throughout the paper, we will use a discrete description where the measured signal is given by the algebraic relation

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

Here, the vector $\mathbf{x}_{\text{orig}} \in \mathbb{R}^N$ holds lexicographically ordered samples of the original $D$-dimensional signal ($N = N_1 \times N_2 \times \ldots \times N_D$ is the product of the number of samples along each dimension). $\mathbf{H}$ is a transform matrix modeling the image-formation device and $\mathbf{b}$ represents the noise component.

The estimation of the original signal $\mathbf{x}_{\text{orig}}$ from the measurement $\mathbf{y}$ is an ill-posed inverse problem [30]. Most approaches for overcoming this ill-posedness can be described in a variational framework, where one looks for an estimate that minimizes a predefined cost functional. This functional is typically the sum of a *data term* and a *regularization term*. Without going into the details of a Bayesian interpretation [5], [7], the former term enforces a certain level of consistency between the estimate and the measured signal (with respect to the image-formation model). The latter term prevents overfitting—and thus instability—by favoring estimates that are close to some desirable class of solutions (according to some regularity measure).

### C. Regularized inversion using a wavelet-domain sparsity constraint

The discovery that natural images can be well approximated using only a few large wavelet coefficients can be traced back to the seminal work of Mallat [13] and is, for example, exploited in the JPEG2000 compression format [31]. Following several recent works (see below), we will use a regularization term that promotes estimates with a sparse wavelet expansion; the data term will be a standard quadratic criterion.

In the sequel, we assume that the reader is familiar with the filter-bank implementation of the wavelet transform [26]. We will denote by $\mathbf{w}$ the vector that contains the coefficients of an estimate $\mathbf{x}$ in a preassigned wavelet basis; we shall refer to this basis as the *synthesis* wavelet basis. Introducing the *synthesis matrix* $\mathbf{W}$, whose columns are the elements of this basis, we can write that

$$\mathbf{x} = \mathbf{W}\mathbf{w}.$$

Later in this paper, we will also use the *analysis matrix* $\tilde{\mathbf{W}}$, whose columns are the elements of the dual wavelet basis. The perfect-reconstruction condition can be expressed as $\mathbf{W}\tilde{\mathbf{W}}^T = \mathbf{I}$, where $^T$ denotes transposition (or Hermitian transposition in the case of a complex wavelet transform). Note that the present formulation also includes the case of overcomplete wavelet representations ($\mathbf{W}$ and $\tilde{\mathbf{W}}$ are then non-square matrices).

With these notations, we consider that a solution to the inverse problem is given by $\mathbf{x} = \mathbf{W}\mathbf{w}$, where $\mathbf{w}$ minimizes the functional

$$\mathcal{C}(\mathbf{w}) = ||\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}||_2^2 + \lambda||\mathbf{w}||_1. \tag{1}$$

Here, $||\mathbf{w}||_1$ represents the $\ell_1$-norm of the wavelet coefficients, that is, the sum of their absolute values. Compared to the standard Euclidian norm (denoted by $||\cdot||_2$), the $\ell_1$-norm puts more weight on small coefficients, and less weight on large coefficients. Thus, depending on the magnitude of the regularization parameter $\lambda$, it favors estimates whose energy is mostly concentrated in a few large wavelet coefficients. Note that in general the coarsest-scale scaling-function coefficients are not included in the $\ell_1$ regularization term (see Section III-C for more details).

An algorithm for the minimization of (1) has been derived in [7], [8], [9], as well as in the earlier works [32], [33]. A similar procedure is also described in [34], [35]. The beauty of the method resides in its simplicity: it essentially consists in alternating between a Landweber iteration [36] and a wavelet-domain thresholding operation [37]—hence the name "thresholded Landweber" (TL) algorithm. When $\mathbf{H}$ is adequately normalized and $\mathbf{W}$ is orthonormal (implying that $\tilde{\mathbf{W}}^T = \mathbf{W}^T$), the TL algorithm can

be described by the recursive update rule[1]

$$\mathbf{x} \leftarrow \mathbf{W}\mathcal{T}_{\lambda/2}\Big\{\mathbf{W}^T\Big(\mathbf{x}+\mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{x})\Big)\Big\}, \tag{2}$$

starting from some arbitrary initial estimate $\mathbf{x}$. Here, $\mathcal{T}_\theta\{\cdot\}$ stands for a pointwise application of the well-known soft-thresholding function [38], which can be defined for $w \in \mathbb{C}$ as

$$\mathcal{T}_\theta(w) = \mathrm{sgn}(w)\max(|w|-\theta, 0) \qquad \text{where} \qquad \mathrm{sgn}(w) = \begin{cases} 0 & \text{if } w = 0; \\ \frac{w}{|w|} & \text{otherwise.} \end{cases}$$

The presence of $\mathcal{T}_{\lambda/2}$ in (2) guarantees that a certain fraction of wavelet coefficients will be set to zero, depending on the magnitude of the regularization parameter $\lambda$.

### *D. Recent relevant work and objectives of the paper*

The present work represents a substantial extension of a previous algorithm of ours [12], which was specific to convolutive image-formation operators and to a sparsity constraint in the (bandlimited) Shannon wavelet basis [26]. Here, the goal is to derive a comparably fast algorithm for an arbitrary wavelet basis, without making the assumption that the image-formation operator leaves the different subbands uncoupled. The approach described in the present paper differs fundamentally from [12] in that it is based on a sequential update of the wavelet subbands, instead of a parallel update. This requires a more sophisticated multilevel algorithm.

Similarly to what is done in some presentations of the multigrid methodology—where a "model problem" is often used to convey the intuition [15]—we will motivate and illustrate our approach in the context of deconvolution. In this case, $\mathbf{H}$ can be thought of as a (block-)circulant matrix corresponding to a given convolution kernel; our multilevel method is then particularly efficient, thanks to the shift-invariant structure of the wavelet subspaces. However, its principle can be applied to more general inverse problems. The most direct extension concerns inverse problems for which $\mathbf{H}^T\mathbf{H}$ can be approximated by a convolution matrix—specifically tomographic image-reconstruction, where $\mathbf{H}$ corresponds to a discretized Radon transform. The subclass of inverse problems involving a unitary image-formation operator (such that $\mathbf{H}^T\mathbf{H} = \mathbf{I}$)—e.g., denoising, reconstruction from K-space (frequency-domain) samples or digital holography microscopy [39]—may also benefit from the method. In the present work, we have tried to provide a general and modular pseudo-code description of the multilevel TL algorithm that is readily transposable to machine implementation.

[1]To keep the notations simple, we do not introduce a specific index to distinguish between the individual estimates. Instead we use the assignment operator "$\leftarrow$" whenever a quantity (such as the estimate) is updated. The algorithmic signification of this operator is that the expression on the right-hand side is evaluated and the result is stored in the left-hand side variable.

Several works have already extended the standard TL algorithm (which was originally formulated only for orthonormal bases) to more general decompositions, including overcomplete wavelet representations [10], [40]. Nevertheless, the principle and the convergence properties of the algorithm were not fundamentally changed in these settings (although [40] is based on a quite different proximal thresholding interpretation).

Faster methods for the minimization of (1) have only been proposed very recently. We are aware of two-step methods [41], [11], [42], line-search methods [43], [44], [45], coordinate-descent methods ([46] and also [43], [44]) and a domain-decomposition method [47]. The latter is based on a well-established concept from the finite-element literature, so that it is arguably the closest to our approach. However it is not specific to wavelets and relies entirely on dimension-reduction effects for decreasing the computational complexity.

The above methods differ with respect to the number and the determination of their step sizes. Among the fixed-step-size strategies, the domain-decomposition approach [47] uses the same step size for all subspaces, whereas the coordinate-descent methods described in [43], [44], [46] use step sizes that are adapted to each atom individually. The methods of Bioucas-Dias, Figueiredo and Nowak [41], [11], [42] have the advantage of simplicity, because they use only two iteration parameters that are also determined a priori (however these parameters may require some hand tuning based on the outcome of a small number of preliminary iterations). The principle of the line-search methods [43], [44], [45] is that the step sizes are adjusted depending on the context, which involves additional computations at every iteration. Our algorithm is somewhere in-between all these approaches: the step sizes are adjusted at the level of individual wavelet subbands, they can be precomputed for a given image-formation operator and wavelet family, and they remain fixed during the entire algorithm.

In summary, to the best of our knowledge, a wavelet-based multilevel method comparable to ours—which combines cyclic updates of the different resolution levels with the preconditioning effect of subband-specific iteration parameters—has not been proposed so far. Therefore, we have chosen to focus on the derivation and the experimental validation of our algorithm. A theoretical study of its convergence properties and a comparison with the aforementioned techniques is a research subject in its own right that will certainly be investigated in the future.

The remainder of the paper is organized as follows. In Section II, we revisit the derivation of the TL algorithm (2) which was presented in [8], introducing additional degrees of freedom into the bound optimization framework. This leads to our multilevel algorithm, described in Section III. Section IV is dedicated to numerical experiments.

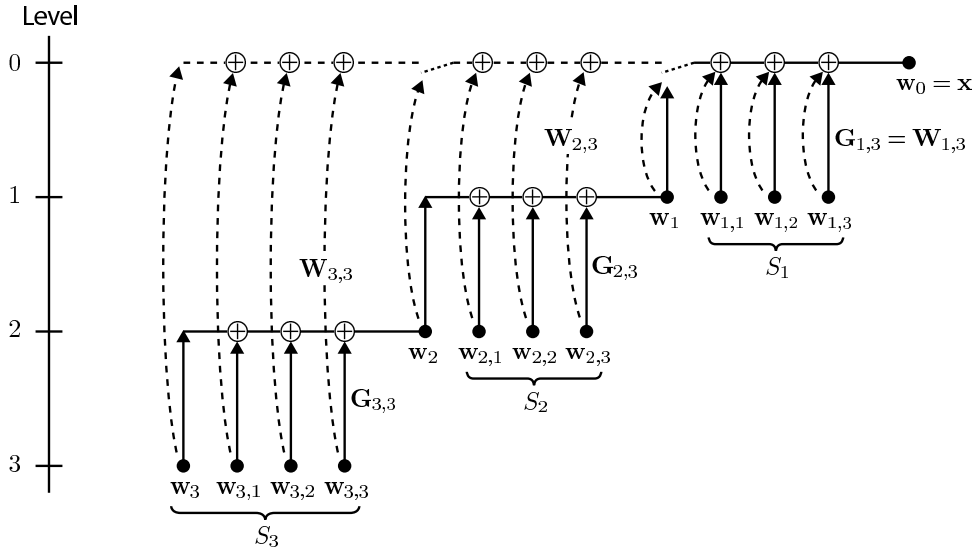



Fig. 1. Complementary notations reflecting the tree structure (notations associated with continuous lines) and the subspace structure (notations associated with dashed lines) of the wavelet representation. Here the number of decomposition levels is $J = 3$. The number of subbands is $M_j = 3$ at every scale $j$, which is typical for a 2D separable wavelet representation.

## II. The thresholded Landweber algorithm, revisited *(divide)*

### A. Notations

In this section, we will primarily be interested in the subspace structure of the wavelet representation. The tree-structure of the wavelet transform—that is, the embedding of the underlying scaling-function subspaces—will become important for the algorithmic considerations of the next section. To account for both aspects, we introduce the following notations, which are illustrated in Fig. 1. Throughout this paper, we shall use the terms "scale", "resolution level", "decomposition level" and "level" interchangeably.

- $J$: number of resolution levels of the wavelet representation ($j$: scale index).
- $M_j$: number of *wavelet* subbands at scale $j$, excluding the scaling-function subband ($m$: subband index).
- $s = (j, m)$: general subband index. Our convention will be that $(j, 0)$ corresponds to the scaling-function subband at scale $j$; however, for the sake of conciseness, we will often simply write $j$ instead of $(j, 0)$. The context will indicate whether we are referring to the scaling-function subband or to the decomposition level.
- $S_j = \{(j, m), m = 1, \ldots, M_j\}$: indexing set for all wavelet subbands at a scale $j \leq J - 1$. At the coarsest level, we include the scaling-function subband: $S_J = \{(j, m), m = 0, \ldots, M_J\}$.

- $S$: indexing set for all subbands produced by a $J$-scale decomposition (including the coarsest-scale scaling-function subband):
$$S = \bigcup_{j=1}^{J} S_j.$$
- $\mathbf{w}_s$: wavelet or scaling-function coefficients of the current estimate corresponding to subband $s$. $\mathbf{w}$ is the concatenation of $\mathbf{w}_s$ for every $s \in S$. Note that $\mathbf{w}_0$ is an alias for $\mathbf{x}$.
- $\mathbf{G}_{j,m}$: matrix corresponding to the reconstruction part (upsampling and filtering using the synthesis filters) of the $m^{\text{th}}$ channel of the filter bank at scale $j$, to go from $\mathbf{w}_{j,m}$ to $\mathbf{w}_{j-1}$.
- $\mathbf{W}_s$: "restriction" of the synthesis matrix $\mathbf{W}$ to subband $s$, such that
$$\mathbf{W}\mathbf{w} = \sum_{s \in S} \mathbf{W}_s \mathbf{w}_s. \tag{3}$$
More precisely, this is a cascade of upsampling and filtering operations defined recursively by
$$\begin{cases} \mathbf{W}_{1,m} = \mathbf{G}_{1,m}; \\ \mathbf{W}_{j,m} = \mathbf{W}_{j-1} \times \mathbf{G}_{j,m} \text{ for } j \geq 2. \end{cases} \tag{4}$$

### *B. Estimation of the cost functional using subband-dependent bounds*

Our algorithm is based on the availability of a wavelet-domain estimation of $\mathbf{H}$ that takes the following form: we assume that there are constants $(\alpha_s)_{s \in S}$ such that

$$\|\mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 \leq \sum_{s \in S} \alpha_s \|\mathbf{w}_s\|_2^2. \tag{5}$$

We shall assume for now that this inequality holds for an arbitrary vector of wavelet coefficients $\mathbf{w}$, and we shall revisit the derivation of the bound-optimization algorithm of Daubechies et al. [8]. Rather than directly considering the original cost functional $\mathcal{C}(\mathbf{w})$, the idea is to iteratively construct a series of auxiliary functionals that are easy to minimize.

Given an estimate of the minimizer of $\mathcal{C}(\mathbf{w})$, say $\mathbf{w}^{\text{old}}$, we define

$$\mathcal{A}(\mathbf{w}) = \sum_{s \in S} \left( \alpha_s \|\mathbf{w}_s^{\text{old}} - \mathbf{w}_s\|_2^2 \right) + \mathcal{C}(\mathbf{w}) - \|\mathbf{H}\mathbf{W}(\mathbf{w}^{\text{old}} - \mathbf{w})\|_2^2. \tag{6}$$

This functional has three important characteristics:

1) When $\mathbf{w} = \mathbf{w}^{\text{old}}$, $\mathcal{A}(\mathbf{w})$ takes the same value as $\mathcal{C}(\mathbf{w})$.
2) For all other values of $\mathbf{w}$, $\mathcal{A}(\mathbf{w})$ is an upper-bound of $\mathcal{C}(\mathbf{w})$, by virtue of (5).
3) $\mathcal{A}(\mathbf{w})$ admits a minimizer with a closed-form expression.

The first two properties imply that, if we find a new estimate $\mathbf{w}^{\text{new}}$ that minimizes (or at least decreases) $\mathcal{A}(\mathbf{w})$, we also decrease $\mathcal{C}(\mathbf{w})$. We simply have to observe that

$$\mathcal{C}(\mathbf{w}^{\text{new}}) \leq \mathcal{A}(\mathbf{w}^{\text{new}}) \leq \mathcal{A}(\mathbf{w}^{\text{old}}) = \mathcal{C}(\mathbf{w}^{\text{old}}).$$

The third property allows us to actually construct such a new estimate. It originates from the negative (rightmost) term in (6), which cancels out the coupling of the wavelet coefficients in $\mathcal{C}(\mathbf{w})$. As a result, the auxiliary functional can be rewritten as

$$\mathcal{A}(\mathbf{w}) = c + \sum_{s \in S} \alpha_s \Big( \|\mathbf{w}_s^{\text{old}} + \alpha_s^{-1} \underbrace{\mathbf{W}_s^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{\text{old}})}_{\mathbf{r}_s} - \mathbf{w}_s\|_2^2 + \lambda \alpha_s^{-1} \|\mathbf{w}_s\|_1 \Big), \tag{7}$$

where the constant $c$ does not depend on $\mathbf{w}$. This expression reveals that the auxiliary functional is essentially a weighted sum of "subfunctionals" that depend on distinct subbands. Furthermore, the wavelet coefficients appear to be completely decoupled in each subfunctional. This means that once we have computed $\mathbf{r}_s = \mathbf{W}_s^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{\text{old}})$ for every subband $s$, we can minimize each subfunctional using solely pointwise operations.

This minimization procedure can be related to two standard image-restoration methods. First, the computation of $\mathbf{w}_s^{\text{old}} + \alpha_s^{-1}\mathbf{r}_s$ may be seen as a wavelet-domain Landweber iteration [36], [30]: the wavelet decomposition of the "reblurred residual" $\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{\text{old}})$ serves as a correction-term, which is applied with a (subband-dependent) step size $\alpha_s^{-1}$. Let us point out, however, that the decomposition of the residual must be performed using the *synthesis* basis. Second, each subfunctional can be interpreted as a denoising functional where $\mathbf{w}_s^{\text{old}} + \alpha_s^{-1}\mathbf{r}_s$ represents the wavelet coefficients of a signal to be denoised and $\lambda\alpha_s^{-1}$ is a regularization parameter (again subband-dependent). The minimizer of such a functional is *unique* and is obtained by soft-thresholding the coefficients of the noisy signal, with a threshold level equal to half the regularization parameter [48].

### C. Relation with the standard thresholded Landweber algorithm

Iterating the previous minimization scheme produces a sequence of estimates that are guaranteed to monotonically decrease the cost functional. The procedure can be summarized by the following two-step update rule:

$$\begin{cases} \text{(a) For every } s \in S, \mathbf{r}_s \leftarrow \mathbf{W}_s^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}); \\ \text{(b) For every } s \in S, \mathbf{w}_s \leftarrow \mathcal{T}_{\lambda\alpha_s^{-1}/2}\{\mathbf{w}_s + \alpha_s^{-1}\mathbf{r}_s\}. \end{cases} \tag{8}$$

Note that the threshold levels must be adjusted proportionally to the inverse of the bound constants.

In particular, when the bounds are the same for all subbands ($\alpha_s = \alpha$ for every $s$), one obtains the standard "thresholded Landweber" (TL) algorithm. This algorithm uses the same step size ($\alpha^{-1}$) and the same threshold level ($\lambda\alpha^{-1}/2$) for all subbands. It is relatively easy to obtain an admissible value for $\alpha$ when $\mathbf{W}$ is an orthonormal matrix. We can then write that, for an arbitrary vector of wavelet coefficients $\mathbf{w}$,

$$\|\mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 \leq \rho(\mathbf{H}^T\mathbf{H})\|\mathbf{W}\mathbf{w}\|_2^2 = \rho(\mathbf{H}^T\mathbf{H})\|\mathbf{w}\|_2^2.$$

Here, $\rho(\mathbf{H}^T\mathbf{H})$ denotes the spectral radius of $\mathbf{H}^T\mathbf{H}$; when $\mathbf{H}$ is a convolution matrix, this is simply the maximum over the squared modulus of its frequency response. Thus, for (5) to hold, it is sufficient to

choose $\alpha_s = \rho(\mathbf{H}^T\mathbf{H})$ for every $s$. Note that (2), which corresponds to $\rho(\mathbf{H}^T\mathbf{H}) = 1$, is a space-domain reformulation of the TL algorithm that is made possible by using an orthonormal basis. This description is quite natural, since eventually we are interested in $\mathbf{x} = \mathbf{W}\mathbf{w}$.

However, we have already mentioned that the TL algorithm converges slowly, especially when the image-formation matrix $\mathbf{H}$ is ill-conditioned. This can be explained intuitively by the fact that using the same bound $\alpha$ for all subbands can only give a very limited account of the spectral characteristics of $\mathbf{H}$. The corresponding auxiliary functionals will thus be relatively poor approximations of the original cost functional, and many intermediate minimization steps will be required before getting a reasonable estimate of the minimizer.

### *D. The single-level thresholded Landweber algorithm*

Our motivation for introducing subband-dependent bounds is to design auxiliary functionals that better reflect the behavior of the underlying cost functional by exploiting the spectral localization properties of the wavelet basis. Specifically, we would like to use an estimate (5) that is tighter—i.e., that involves smaller constants $\alpha_s$—than the aforementioned bound for the standard TL algorithm.

In the sequel, $\rho_{s_2\,s_1}$ will denote the largest singular value of the matrix $\mathbf{W}_{s_2}^T\mathbf{H}^T\mathbf{H}\mathbf{W}_{s_1}$. In particular, $\rho_{s\,s} = \rho(\mathbf{W}_s^T\mathbf{H}^T\mathbf{H}\mathbf{W}_s)$ is the spectral radius of $\mathbf{W}_s^T\mathbf{H}^T\mathbf{H}\mathbf{W}_s$; when $\mathbf{H}$ is a convolution matrix, this is the upper Riesz bound of the *filtered version* of the wavelet that spans subspace $s$. Note that $\rho_{s\,s}$ can be significantly smaller than $\rho(\mathbf{H}^T\mathbf{H})$. As an intuitive example, one could imagine the case where $\mathbf{H}$ corresponds to a low-pass filter and $s$ is a high-frequency wavelet subband.

The quantity $\rho_{s\,s}$ is important because it represents a lower limit for $\alpha_s$. Indeed, for a vector $\mathbf{w}$ with a single non-zero wavelet subband, say $\mathbf{w}_s$, (5) reduces to $||\mathbf{H}\mathbf{W}_s\mathbf{w}_s||_2^2 \leq \alpha_s\,||\mathbf{w}_s||_2^2$. For this inequality to hold for every $\mathbf{w}_s$, we must choose $\alpha_s \geq \rho(\mathbf{W}_s^T\mathbf{H}^T\mathbf{H}\mathbf{W}_s)$.

A particular case arises when the subspaces spanned by the matrices $\mathbf{H}\mathbf{W}_s$ are mutually orthogonal. We can then use exactly the value $\alpha_s = \rho_{s\,s}$, since

$$||\mathbf{H}\mathbf{W}\mathbf{w}||_2^2 = \sum_{s\in S} ||\mathbf{H}\mathbf{W}_s\mathbf{w}_s||_2^2 \leq \sum_{s\in S} \rho_{s\,s}\,||\mathbf{w}_s||_2^2.$$

Our previous work [12] was based on the fact that the bandlimited Shannon wavelet basis exhibits this decorrelation property with respect to convolution operators. In such a situation we can directly apply algorithm (8).

When considering arbitrary wavelet families and image-formation operators, we must a priori bound numerous cross-subband correlation terms, since in general

$$||\mathbf{H}\mathbf{W}\mathbf{w}||_2^2 = \sum_{s_1\in S}\sum_{s_2\in S} \langle \mathbf{H}\mathbf{W}_{s_2}\mathbf{w}_{s_2}, \mathbf{H}\mathbf{W}_{s_1}\mathbf{w}_{s_1}\rangle. \tag{9}$$

This would require constants $\alpha_s$ that are significantly larger than $\rho_{s\,s}$. However, we can make the following observation: if we impose that $\alpha_s = \rho_{s\,s}$, inequality (5) remains valid for all vectors $\mathbf{w}$ that

have at most one non-zero subband. This means that (6) would define a valid upper-bound of the cost functional under the constraint that $\mathbf{w}$ and $\mathbf{w}^{\text{old}}$ differ by only one subband, that is, under the constraint that we update only one subband at a time.

In practice, owing to the structure of the wavelet representation, it is algorithmically more efficient to be able to update all subbands at a given scale simultaneously. We thus propose to replace (8) by

$$\begin{cases} \text{(a) For every } s \in S_j, \mathbf{r}_s \leftarrow \mathbf{W}_s^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}); \\ \text{(b) For every } s \in S_j, \mathbf{w}_s \leftarrow \mathcal{T}_{\lambda \alpha_s^{-1}/2}\{\mathbf{w}_s + \alpha_s^{-1}\mathbf{r}_s\}. \end{cases} \tag{10}$$

This choice only requires taking into account correlations between a small number of subbands (those located at the same scale), so that the resulting constants $\alpha_s$ are still close to $\rho_{s\,s}$. More precisely, the following property provides a valid upper-bound under the constraint that we update only a single scale.

*Property 1:* If we set

$$\alpha_s = \sum_{s_0 \in S_j} \rho_{s_0\,s} \tag{11}$$

for every $s$, then inequality (5) holds for an arbitrary vector of wavelet coefficients $\mathbf{w}$ satisfying the following constraint: there is a scale $j$ such that for all subbands $s \notin S_j$, $\mathbf{w}_s = \mathbf{0}$.

*Proof:* Equality (9) reduces to

$$\|\mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 = \sum_{s_1 \in S_j} \sum_{s_2 \in S_j} \langle \mathbf{H}\mathbf{W}_{s_2}\mathbf{w}_{s_2}, \mathbf{H}\mathbf{W}_{s_1}\mathbf{w}_{s_1} \rangle.$$

Combining this with the fact that, for every $\mathbf{w}_{s_1}$, $\mathbf{w}_{s_2}$,

$$|\langle \mathbf{H}\mathbf{W}_{s_2}\mathbf{w}_{s_2}, \mathbf{H}\mathbf{W}_{s_1}\mathbf{w}_{s_1} \rangle| \leq \rho_{s_2\,s_1} \, \|\mathbf{w}_{s_2}\|_2 \, \|\mathbf{w}_{s_1}\|_2, \tag{12}$$

we obtain

$$\|\mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 \leq \sum_{s_1 \in S_j} \sum_{s_2 \in S_j} \rho_{s_2\,s_1} \, (\|\mathbf{w}_{s_2}\|_2^2 + \|\mathbf{w}_{s_1}\|_2^2)/2 = \sum_{s \in S} \alpha_s \|\mathbf{w}_s\|_2^2.$$

■

Appendix A describes an algorithm for computing the constants $\rho_{s_0\,s}$ in the convolutive case; for more general operators one may use the power method [49]. Let us emphasize that, under the condition of Property 1, every application of (10) is guaranteed to decrease the auxiliary functional (and thus the original cost functional), despite the fact that only a subset of subbands is updated. This follows from (7), which shows that the minimization of the auxiliary functional can always be divided into a collection of subband-specific—hence independent—minimization problems. Of course, by letting $j$ vary at every iteration, we can successively update the subbands at all scales. The next section describes an efficient method for doing this.

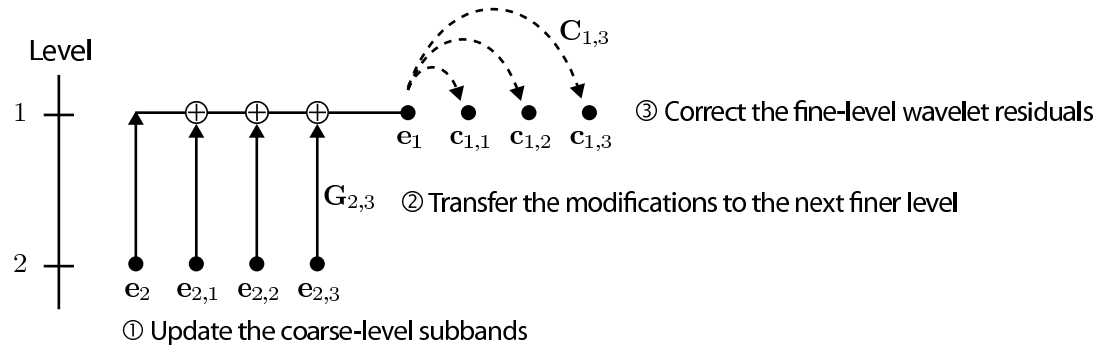



Fig. 2. Principle of the coarse-to-fine thresholded Landweber (CFTL) algorithm for a two-level decomposition.

## III. The multilevel thresholded Landweber algorithm *(conquer)*

The general idea behind our multilevel scheme is to interlace the computation of the residual and the minimization procedure. To give the reader the intuition of this principle, we focus on the description of a simplified strategy that consists in applying (10) successively from the coarsest-scale to the finest-scale subbands.

### A. A coarse-to-fine update strategy

Let $\mathbf{r} = \mathbf{W}^T\mathbf{H}^T(\mathbf{y}-\mathbf{HWw})$ be the residual corresponding to the current estimate. Assume that $\mathbf{w}$ is modified at scale $j$ by applying procedure (10). In general, this will imply a modification of the residual in all subbands (due to the matrix $\mathbf{W}^T\mathbf{H}^T\mathbf{HW}$, which couples the subbands). If the next iteration is performed at scale $j-1$, it is however not necessary to recompute the entire residual; instead, one can simply update the subbands $s \in S_{j-1}$. Denoting by $(\mathbf{e}_s)_{s\in S_j}$ the modifications that have been applied to the estimate, the corresponding correction that must be applied to the residual in a subband $s_0 \in S_{j-1}$ is

$$\mathbf{c}_{s_0} = -\mathbf{W}_{s_0}^T\mathbf{H}^T\mathbf{H}\sum_{s\in S_j}\mathbf{W}_s\mathbf{e}_s = -\mathbf{W}_{s_0}^T\mathbf{H}^T\mathbf{HW}_{j-1}\sum_{s\in S_j}\mathbf{G}_s\mathbf{e}_s.$$

The above equality stems from the cascade implementation of the wavelet transform—see (4). Thus, an updated version of the residual at scale $j-1$ is obtained as follows:

1) transfer all modifications to the scaling-function subband at the next finer scale $j-1$;
2) apply the "correction matrices" $\mathbf{C}_s = \mathbf{W}_s^T\mathbf{H}^T\mathbf{HW}_{j-1}$ for every $s \in S_{j-1}$;
3) subtract the results from the respective subbands.

This principle is illustrated in Fig. 2; its recursive application leads to the "coarse-to-fine thresholded Landweber" (CFTL) algorithm. A pseudo-code description is given below. Note that all modifications—including those from subbands located at coarser scales than the current scale $j$—are progressively

transfered to finer scales. The CFTL algorithm depends on the "single-level thresholded Landweber" (SLTL) procedure, which essentially corresponds to the updating rule (10). The only difference is that the modifications that are applied to the estimate are stored in intermediate variables so as to be able to update the residual. For simplicity, the variables $\mathbf{w}_s$, $\mathbf{e}_s$ and $\mathbf{r}_s$ are considered to be global (for every subband $s$) in all pseudo-code descriptions given in this paper.

We emphasize that the correction steps 2) and 3) above are the only additional operations compared to the standard TL algorithm. These steps should require little computational effort at coarse levels, thanks to the pyramidal structure of (decimated) wavelet representations. In other words, they can be implemented efficiently if the computational complexity of evaluating $\mathbf{H}^T\mathbf{H}$ (the forward image-formation model followed by the corresponding "back-projection") scales well with this data-size reduction.

A particular case arises when $\mathbf{H}^T\mathbf{H}$ is a convolution matrix. The shift-invariant structure of wavelet subspaces then implies that the correction steps essentially reduce to filtering operations. Here we refer to Appendix A, which also gives a recursive method for precomputing the correction filters: the procedure is akin to a wavelet decomposition of the convolution kernel corresponding to $\mathbf{H}^T\mathbf{H}$ and is easily implementable in the frequency domain (see also [50]). The correction steps can be implemented with a linear cost provided that we store the DFTs of the individual wavelet subbands; the actual wavelet coefficients are only needed for the thresholding operations and can be computed efficiently using the FFT algorithm. In terms of computational work, one iteration of our algorithm is therefore equivalent to two FFTs per subband, which amounts to two FFTs at the signal level (level 0). The overall complexity of a full coarse-to-fine run is thus on the same order as one run of the standard TL algorithm, which also requires two FFTs per iteration in the convolutive case.

With a slight anticipation of the next subsection, we conclude this part by noting that multigrid methodologies sometimes advocate the approximate resolution of coarse-level problems [15]. In the particular situations where the wavelet subbands are weakly coupled by the image-formation operator, we have indeed observed that the CFTL algorithm converges even if the correction steps are not applied; that is, if the residual is only updated at the beginning of the iteration loop. This amounts to applying (8) using fairly optimistic bound constants—without guarantee that the cost functional is monotonically decreased—and calls for further investigation. This approach may turn out to be useful when dealing with complex image-formation models that can not be evaluated easily at coarse levels.

### *B. A general multilevel scheme*

With the previous algorithm in mind, one can conceive of more general multilevel strategies for updating the different scales in a more flexible manner. In Appendix B, we provide a pseudo-code description of a method that is strongly inspired by the multigrid paradigm. However, there is one fundamental difference: traditional multigrid schemes typically cycle through nested subspaces corresponding

**Algorithm 1** SLTL($j$)

For every $s \in S_j$:

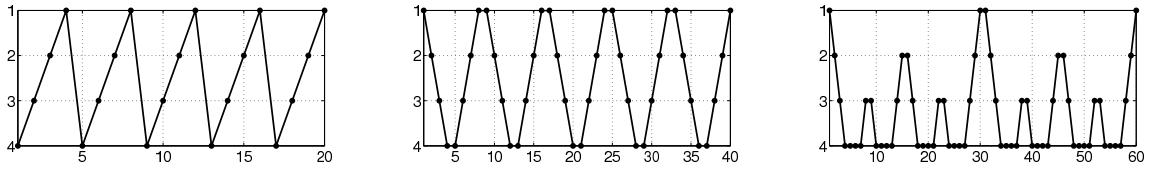
- $\mathbf{e}_s \leftarrow \mathbf{w}_s$
- $\mathbf{w}_s \leftarrow \mathcal{T}_{\lambda\alpha_s^{-1}/2}\left\{\mathbf{w}_s + \alpha_s^{-1}\mathbf{r}_s\right\}$
- $\mathbf{e}_s \leftarrow \mathbf{w}_s - \mathbf{e}_s$

**Algorithm 2** CFTL

- Initialization:
  - Choose some initial estimate $\mathbf{x}$
  - Compute its wavelet decomposition: $\mathbf{w} \leftarrow \tilde{\mathbf{W}}^T\mathbf{x}$
- Repeat $K$ times:
  - Compute the residual: for every $s \in S$, $\mathbf{r}_s \leftarrow \mathbf{W}_s^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x})$
  - Update the subbands from coarse to fine levels, i.e., for $j = J, J-1, J-2, \ldots, 1$:
    - Update the subbands at the current level: SLTL($j$)
    - Transfer the modifications to the next finer level: $\mathbf{e}_{j-1} \leftarrow \sum_{s \in S_j \cup \{j\}} \mathbf{G}_s\mathbf{e}_s$
    - If $j \geq 2$, correct the residual for the wavelet subbands at the next finer level: for every $s \in S_{j-1}$, $\mathbf{r}_s \leftarrow \mathbf{r}_s - \mathbf{C}_s\mathbf{e}_{j-1}$
  - Set $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{e}_0$
- Return $\mathbf{x}$

to increasingly coarse discretizations of the original inverse problem [51]. In the present context, we successively update the wavelet subbands at every scale; that is, we reinterpret the different scales of the wavelet transform as a multilevel representation of the inverse problem. The corresponding subspaces are not nested—they contain the oscillating components corresponding to the difference between successive coarse-level approximations. Incidentally, early attempts to apply the multigrid paradigm to image-restoration problems remained relatively unsuccessful because they were concentrating on slowly oscillating components [52], [53].

We have tried to specify the "multilevel thresholded Landweber" (MLTL) algorithm in a modular way that is relatively close to machine implementation. Its main building block, UpdateLevel($j$), depends on three parameters so as to be able to mimic typical multigrid schemes (see Fig. 3). The parameters are thus named $\eta_1$, $\eta_2$ and $\mu$, following the conventions of the multigrid literature [54], [15]. In the particular case $\eta_1 = 0$, $\eta_2 = 1$ and $\mu = 1$, one retrieves the coarse-to-fine update described in the preceding




(a) Coarse-to-fine ($\mu$=1; $\eta_1$=0; $\eta_2$=1) (b) V-cycles ($\mu = 1$; $\eta_1 = \eta_2 = 1$) (c) W-cycles ($\mu = 2$; $\eta_1 = \eta_2 = 1$)

Fig. 3. Examples of multigrid-like updating schemes made possible by the general MLTL algorithm. Each dot corresponds to an application of (10) at the corresponding level.

subsection (Fig. 3a). However, we should note that the MLTL algorithm is numerically (slightly) more stable, because the current estimate is explicitely reconstructed from its wavelet coefficients at every iteration.

The different modules of Appendix B can be summarized as follows:

- UpdateResidual($j$): updates the residual for the subbands at scale $j$ (if needed). Uses the correction principle of Fig. 2 if the wavelet subbands have not been modified so far. Otherwise the update is performed by temporarily moving up to the next finer scaling-function subband.
- UpdateLevel($j$): recursive procedure which
  - updates the subbands at coarser scales by calling itself $\mu$ times;
  - updates the subbands at scale $j$ by calling SLTL($j$) (the number of updates before and after a recursive call are fixed by $\eta_1$ and $\eta_2$ respectively).

  Note that the procedure must compute the current residual for the scaling-function subband at scale $j$ before calling itself. This is either done by applying the correction principle of Fig. 2 in the opposite direction (from the wavelet subbands to the scaling-function subband), or by going to the next finer scale using UpdateResidual($j$).
- MLTL: main routine that performs initialization tasks followed by several iterations of the update procedure. One may devise even more general—e.g. "full multigrid" [14]—schemes by adapting this routine.

### C. A fixed-point property

A comprehensive study of the convergence properties of the MLTL algorithm is well beyond the scope of the present work. In particular, obtaining *tight* theoretical convergence-rate estimates is a difficult problem even for linear subspace-correction methods [51]. In the next section, we thus propose a numerical study of the convergence rate of the MLTL algorithm, based on the following concise characterization of the minimizer(s) of the cost functional (1). We provide a proof in Appendix C for completeness (see also [8] and the general results in [55], [40]).

*Property 2:* $\mathbf{w}^*$ is a minimizer of $\mathcal{C}(\mathbf{w})$ if and only if it is a fixed point of the standard TL algorithm, that is, if and only if there is an arbitrary step size $\tau > 0$ such that $\mathbf{w}^* = \mathcal{T}_{\lambda\tau/2}\{\mathbf{w}^* + \tau\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)\}$. Furthermore the minimizer is unique if $\mathbf{W}^T\mathbf{H}^T\mathbf{H}\mathbf{W}$ is positive definite.

A similar property can be obtained for the MLTL algorithm. This ensures that we obtain a minimizer of the cost functional whenever the MLTL algorithm converges, which was always the case in our numerical experiments (see Section IV). In the sequel, $[\mathbf{v}]_n$ stands for the $n$-th component of a vector $\mathbf{v}$.

*Property 3:* $\mathbf{w}^*$ is a minimizer of $\mathcal{C}(\mathbf{w})$ if and only if it is a fixed point of the MLTL algorithm; that is, if and only if it is not modified by a sequence of successive applications of (10) at different scales, such that every subband is updated at least once.

*Proof:*

- *Necessary part:* $\mathbf{w}^*$ is assumed to be a minimizer of $\mathcal{C}(\mathbf{w})$.

  The characterization of Property 2 is equivalent to the following statement: there is a $\tau > 0$ such that, for every component $n$ of $\mathbf{w}^*$,

  $$\begin{cases} \text{either } [\mathbf{w}^*]_n = 0 \text{ and } |[\tau\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n| \leq \frac{\lambda\tau}{2}; \\ \text{or } [\mathbf{w}^*]_n \neq 0 \text{ and } [\tau\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n = \frac{\lambda\tau}{2}\mathrm{sgn}([\mathbf{w}^*]_n). \end{cases} \tag{13}$$

  Multiplying by a suitable constant shows that the step size $\tau > 0$ can actually be chosen arbitrarily. In particular, if $n$ corresponds to a wavelet coefficient of subband $s$, (13) also holds for $\tau = \tau_s$. Using this argument for all wavelet coefficients shows that $\mathbf{w}^*$ is a fixed point of (10) at any scale $j$.

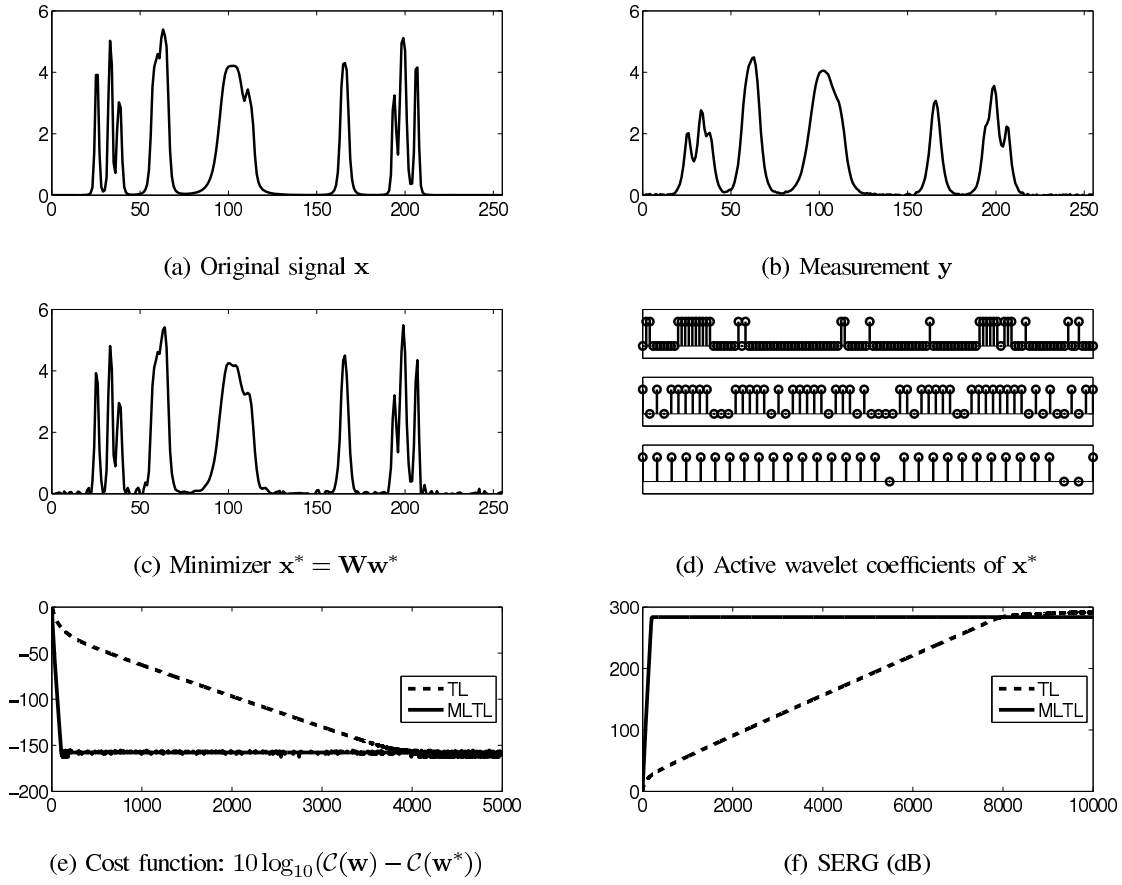
- *Sufficient part:* $\mathbf{w}^*$ is assumed to be invariant under a sequence of successive applications of (10), such that every scale is visited at least once.

  We first observe that, for a given scale $j$, (10) computes the minimizer of the auxiliary functional considered as a function of $(\mathbf{w}_s)_{s\in S_j}$ only. Because this minimizer is unique, the result of applying (10) is either to leave the estimate unchanged, or to strictly decrease the auxiliary functional—and thus the original cost functional by construction.

  The fixed-point assumption excludes the latter case. Therefore it must be that $\mathbf{w}^*$ is invariant under each individual application of (10). Let $n$ be an arbitrary component of $\mathbf{w}^*$, e.g. corresponding to a subband $s \in S_j$. Since we assume that (10) was applied at scale $j$ at least once, (13) holds with $\tau = \tau_s$; in fact, it holds for an arbitrary $\tau > 0$. Using this argument for all wavelet coefficients, one retrieves the characterization of Property 2.

■

Before proceeding to the experimental part of this work, we mention two straightforward extensions of the MLTL algorithm. First, it is clear that the algorithm (and the above results) can be extended




(a) Original signal $\mathbf{x}$

(b) Measurement $\mathbf{y}$

(c) Minimizer $\mathbf{x}^* = \mathbf{W}\mathbf{w}^*$

(d) Active wavelet coefficients of $\mathbf{x}^*$

(e) Cost function: $10\log_{10}(\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*))$

(f) SERG (dB)

Fig. 4. Experiment with a known minimizer (Symlet8, $\lambda = 0.005$).

to a cost functional with *subband-specific* regularization parameters $\lambda_s$. In particular, the coarsest-scale saling-function subband $\mathbf{w}_J$ is not thresholded in practice, i.e., $\lambda_J = 0$. Second, one can also replace the $\ell_1$-regularization in (1) by another coefficient-wise penalization in the wavelet-domain. This will essentially amount to changing the thresholding function; as long as the regularization term is convex, the same bound-optimization framework can be deployed.

## IV. Numerical experiments

In the experiments presented below, we use an $\ell_1$ norm for the regularization term. Unless specified otherwise, we use the same regularization parameter for all wavelet subbands; the scaling-function subband is never penalized.

### A. Asymptotic convergence (1D experiments)

To evaluate the convergence behavior of the MLTL algorithm, we designed an experiment where the true minimizer of the cost functional is used as a gold standard. Each test case was constructed as follows.

The standard "bumps" signal (Fig. 4a) is convolved with an $N$-periodic low-pass kernel defined by $h[n] \propto \exp(-|n|/2)$ for $n = -N/2, \ldots, N/2-1$; the corresponding convolution matrix $\mathbf{H}$ is normalized such that $\rho(\mathbf{H}^T\mathbf{H}) = 1$. White Gaussian noise is added to the result, so as to simulate a measurement $\mathbf{y}$. The standard TL algorithm (with a step size $\tau = 1$) is then initialized with this measurement and run for 50000 iterations in order to obtain our reference solution $\mathbf{w}^*$. Since $\mathbf{W}^T\mathbf{H}^T\mathbf{H}\mathbf{W}$ is positive-definite for the convolution kernel defined above (the smallest DFT coefficient of $\mathbf{H}^T\mathbf{H}$ being 0.06), Property 2 implies that this minimizer is unique. Figures 4b and 4c show an example of the measurement $\mathbf{y}$ and of the corresponding minimizer $\mathbf{x}^*$. Fig. 4d shows the locations of the non-zero wavelet coefficients of the solution (we use a 3-level wavelet decomposition and the upper plot represents the finest-resolution subband).

We then used this reference to compare the asymptotic behavior of the TL and MLTL algorithms. To this end, we performed a series of experiments where the algorithms are applied to the minimization of (1) and initialized with the measurement $\mathbf{y}$. Although the asymptotic convergence rates that are presented here may not be directly relevant to practical situations, they give a quantitative indication of the acceleration potential of the MLTL algorithm. Our asymptotic study required several thousand iterations of the TL and MLTL algorithms in various configurations, which is why we resorted to a small-scale problem ($D = 1$ and $N = 256$). The MLTL algorithm was used with the parameters $\mu = 1$, $\eta_1 = 0$ and $\eta_2 = 1$ (coarse-to-fine strategy). The computational cost of one complete MLTL iteration is then essentially the same as the cost of one TL iteration (each subband is updated once per iteration). This allows for a direct comparison of both algorithms in terms of number of iterations.

The decay of the cost functional towards its minimal value $\mathcal{C}(\mathbf{w}^*)$ is represented in Fig. 4e. This decay is only limited by the numerical precision of the computer environment (we used Matlab on a 64-bit Intel Xeon workstation). To reach this limit with the MLTL algorithm, the number of iterations is divided by more than 10 compared to the TL algorithm. We also display the distance between the estimate and the minimizer (Fig. 4f); here the "signal-to-error-ratio gain" is defined as $\mathrm{SERG} = 20\log_{10}(\|\tilde{\mathbf{W}}^T\mathbf{y} - \mathbf{w}^*\|_2/\|\mathbf{w} - \mathbf{w}^*\|_2)$. As expected, both algorithms converge to the minimizer, but the MLTL algorithm is again faster by more than one order of magnitude for reaching the level of numerical precision.

To obtain a more quantitative insight, we repeated the experiment in several test cases and computed the slope of the SERG curves between 100 and 250 dB. This measurement gives an estimate of the asymptotic convergence rate, in dB per iteration. The results are summarized in Table I, for various orthonormal wavelet bases and different values of the regularization parameter, corresponding to different noise levels (from top to bottom, the values of $\lambda$ correspond to BSNR noise levels of 60, 50, 40, 30, 20, 10 dB respectively—see [12] for the definition of BSNR).

For validation purposes, we computed a theoretical convergence-rate estimate in the case where $\lambda = 0$. With this particular choice, both algorithms reduce to linear (least squares) restoration procedures: since

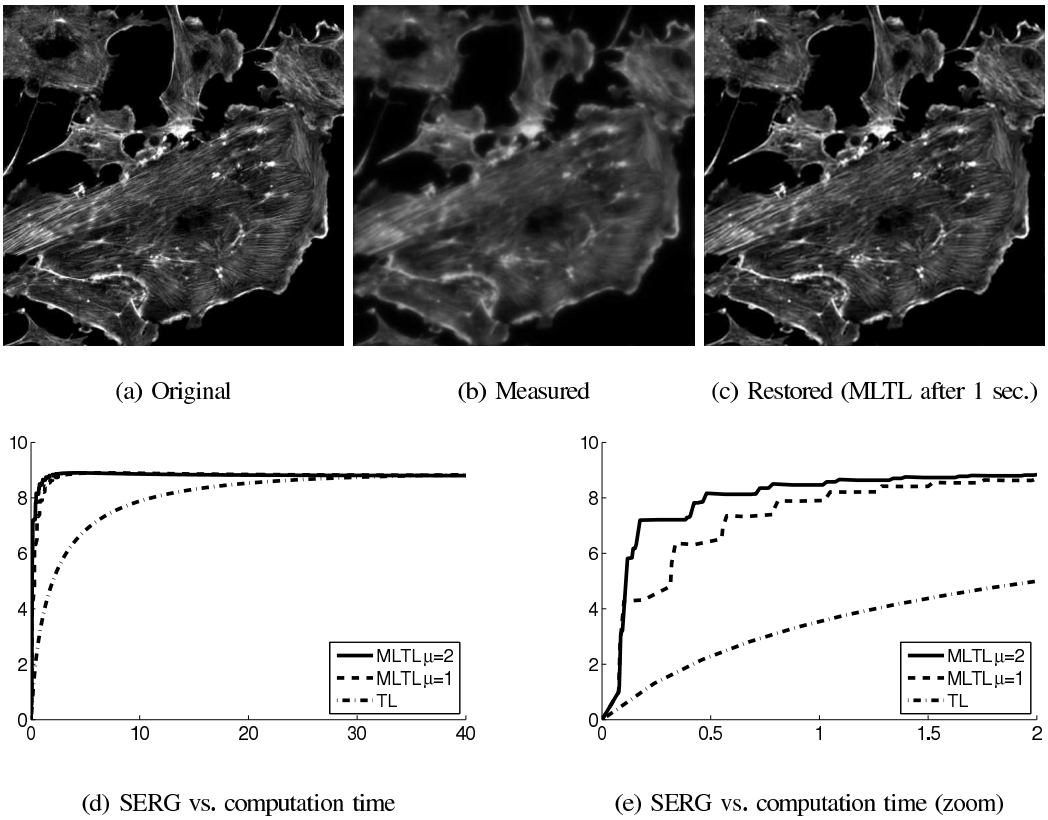
| $\lambda$ | Cv. rates | Algo. | Haar | Daub2 | CubicSpline | Sym8 | Shannon |
|---|---|---|---|---|---|---|---|
| 0 | Theoretical | TL | 0.031 | 0.031 | 0.031 | 0.031 | 0.031 |
| | | MLTL | 0.376 | 0.761 | 1.302 | 1.301 | 1.301 |
| 0 | Measured | TL | 0.032 | 0.032 | 0.032 | 0.032 | 0.032 |
| | | MLTL | 0.383 | 0.769 | 1.321 | 1.321 | 1.321 |
| 0.00025 | Measured | TL | 0.033 | 0.032 | 0.032 | 0.032 | 0.032 |
| | | MLTL | 0.393 | 0.803 | 1.324 | 1.331 | 1.335 |
| 0.002 | Measured | TL | 0.032 | 0.033 | 0.032 | 0.032 | 0.033 |
| | | MLTL | 0.408 | 0.818 | 1.336 | 1.326 | 1.379 |
| 0.01 | Measured | TL | 0.039 | 0.038 | 0.035 | 0.033 | 0.038 |
| | | MLTL | 0.570 | 1.178 | 1.347 | 1.368 | 1.486 |
| 0.05 | Measured | TL | 0.035 | 0.058 | 0.051 | 0.055 | 0.051 |
| | | MLTL | 0.433 | 1.317 | 1.559 | 1.512 | 1.595 |
| 0.25 | Measured | TL | 0.132 | 0.173 | 0.154 | 0.184 | 0.165 |
| | | MLTL | 1.054 | 1.570 | 1.837 | 2.013 | 1.978 |

TABLE I

CONVERGENCE RATES (IN DB PER ITERATION) FOR DIFFERENT VALUES OF THE REGULARIZATION PARAMETER AND VARIOUS ORTHONORMAL WAVELET BASES.

the thresholding step disappears, the TL algorithm reduces to the standard Landweber iteration and the MLTL algorithm corresponds to a wavelet-based multilevel implementation of the Landweber iteration. For this type of linear iterations, the asymptotic convergence rate can be estimated using the spectral radius of the so-called iteration matrix (see e.g. [54]). This spectral radius can be obtained directly for the TL algorithm because we consider orthonormal wavelet bases (the computation is essentially the same as in Section II-C). For the MLTL algorithm, the small dimension of the problem allows us to explicitly construct the iteration matrix in order to evaluate its spectral radius. The resulting theoretical convergence-rate estimates (expressed in dB per iteration for comparison purposes) are reported in the first row of Table I.

The theoretical and the measured values are in good agreement for $\lambda = 0$, suggesting that our experimental method for measuring the asymptotic convergence rate is reliable. The results for $\lambda \neq 0$ corroborate the former observation [7], [12] that the TL algorithm tends to converge faster for higher values of $\lambda$. This can be explained by the fact that the variational problem is more constrained, thus compensating for the unfavorable conditioning of the convolution kernel. Nevertheless, the convergence rates of the MLTL algorithm are consistently one order of magnitude larger than those of the TL algorithm. The figures suggest that the strongest acceleration is generally obtained for higher-order wavelets, which




(a) Original (b) Measured (c) Restored (MLTL after 1 sec.)

(d) SERG vs. computation time (e) SERG vs. computation time (zoom)

Fig. 5. Computation time comparison (in seconds).

can be related to their improved frequency selectivity. The Shannon wavelet basis provides perfect frequency selectivity, a property that was exploited in our previous work [12].

### *B. Computation time (2D experiments)*

In our second series of experiments, we evaluated the performance of the MLTL algorithm in terms of computation time. This type of assessment is most relevant in practical situations, but it depends on computer hardware parameters. Therefore we always provide a comparison with the standard TL algorithm.

We first simulated the effect of a defocusing blur on a 512×512 test image (Fig. 5a). We used a standard diffraction-limited point spread function (PSF) model for widefield fluorescence microscopy [1]. The result was then corrupted by additive white Gaussian noise with a BSNR of 40 dB (Fig. 5b). We restored this simulated measurement using the TL and MLTL algorithms. Both were initialized with the measurement. We used a separable orthonormalized cubic spline wavelet basis with four decomposition levels. The regularization parameter $\lambda = 0.2$ was the same for both algorithms; it was adjusted using multiple trials, so as to give the best restoration quality after the MLTL algorithm had converged. Figures

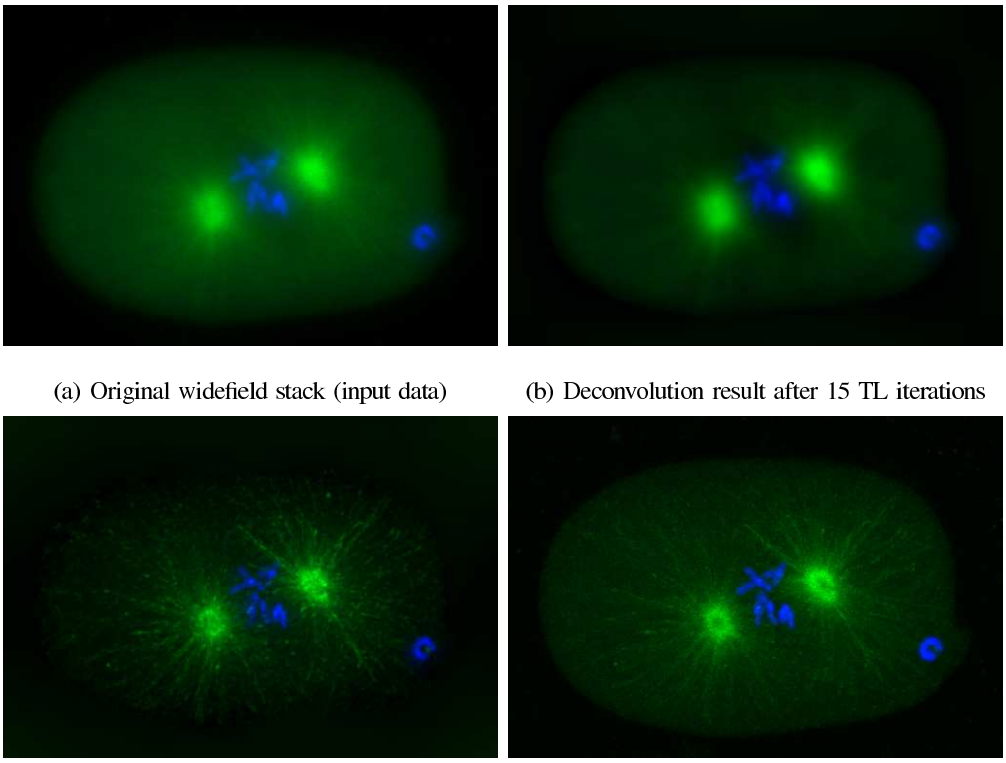
| Wavelet basis | MLTL SERG after 1 sec. | TL computation time to reach same SERG | MLTL SERG after 4 sec. | TL computation time to reach same SERG |
|---|---|---|---|---|
| Haar | 7.13 dB | 31.7 sec. | 7.53 dB | 41.6 sec. |
| 9/7 | 6.96 dB | 31.6 sec. | 7.30 dB | 46.0 sec. |
| Shannon | 6.50 dB | 29.5 sec. | 6.74 dB | 41.3 sec. |

TABLE II

COMPUTATION TIME REQUIRED TO REACH A GIVEN LEVEL OF RESTORATION QUALITY, FOR THE SECOND EXPERIMENT OF SECTION IV-B.

5d and 5e show the evolution of the restoration quality measure SERG $= 20\log_{10}(\|\mathbf{y}-\mathbf{x}_{\text{orig}}\|_2/\|\mathbf{x}-\mathbf{x}_{\text{orig}}\|_2)$, where $\mathbf{x}_{\text{orig}}$ stands for the original signal and $\mathbf{x}=\mathbf{W}\mathbf{w}$ is the estimate. One can observe that the coarse-to-fine MLTL algorithm requires 1 second of computations to reach an improvement of 8 dB (result shown in Fig. 5c). The TL algorithm needs approximately 10 seconds to reach the same figure. We found that the performance of the MLTL algorithm can be further improved by using $\mu = 2$, i.e., with a modified W-cycle iteration. This makes sense since natural images tend to have mostly low-frequency content; thus, iterating on coarse-scale subbands brings the largest improvement in the beginning, unless the algorithm is initialized with a very accurate estimate.

We used a similar protocol for the second part of our 2D experiments, where we replicated the standard test case used by Figueiredo and Nowak in [7] (Cameraman image convolved with a $9 \times 9$ uniform blur; additive white Gaussian noise with a BSNR of 40 dB; initialization with a Wiener-type estimate). In particular, we introduced a random shift of the estimate at the beginning of every TL and MLTL iteration; the authors found that this method gave optimal results with non-translation-invariant wavelet transforms. We present results for three wavelet bases, including biorthogonal 9/7 wavelets. We always used 3 decomposition levels and the same regularization parameter $\lambda = 0.04$. The second and fourth columns of Table II show the restoration quality of the MLTL algorithm after 1 and 4 seconds, respectively. The third and fifth columns give the minimum TL computation time that is required to reach the same restoration quality. Again, the MLTL algorithm provides an acceleration of roughly one order of magnitude. Our results confirm the superiority of the Haar basis among separable wavelet bases for 2D image restoration; this fact is already known from denoising applications [56]. In general, the ability to use other wavelet bases than the Shannon wavelet basis leads to substantial improvements over our previous work [12]. In summary, the MLTL algorithm can yield state-of-the-art results (similar to those obtained in [7]) in a substantially shorter time than the TL algorithm.



(a) Original widefield stack (input data)

(b) Deconvolution result after 15 TL iterations

(c) Deconvolution result after 15 MLTL iterations

(d) Confocal reference stack

Fig. 6. 3D deconvolution results (maximum-intensity projections of 512×352×96 image stacks).

### *C. Application to real fluorescence-microscopy data (3D experiments)*

To conclude this experimental part, we applied the MLTL algorithm to real 3D fluorescence microscopy data. Similarly to our experiment in [12], we acquired two image-stacks of the same sample (a C. Elegans embryo), one of them serving as a visual reference to assess the restoration quality. For the present work we acquired a much larger, two-channel data set.

Both data sets were acquired on a confocal microscope, which has the ability to reject out-of-focus light using a small aperture in front of the detector. This creates a relatively sharp but noisy image (Fig. 6d). When the aperture is opened the signal intensity is improved, but the measurement gets blurred by the contributions of defocused objects. This results in hazy images that are characteristic for widefield microscopes (Fig. 6a).

We applied the MLTL algorithm to the widefield-type stack, with three decomposition levels. To account for the anisotropic sampling scheme of the microscope, we used an orthonormalized linear spline wavelet for the X-Y dimensions, and a a Haar wavelet for the Z dimension. We kept the random-shift method of [7] and we used scale-dependent regularization parameters that were adjusted with the confocal stack as a visual reference. The channels were processed independently, using computer-generated PSFs

based on a three-dimensional version of the diffraction-limited model used in the previous subsection. The parameters of this model were adjusted according to manufacturer-provided specifications of the objective, the immersion oil and the fluorescent dyes (NA, refractive index, emission wavelength).

The result is shown in Fig. 6c: the restored image-stack provides significantly better contrast than the original widefield image, especially for the filaments (green channel). The chromosomes and their centromeres (blue channel) appear almost as sharp as in the confocal image.

For completeness, we have included the result of the TL algorithm after the same number of iterations (Fig. 6b). It is seen that the resulting image-stack is still very hazy; the TL algorithm fails to produce a visible deconvolution effect within the assigned budget of iterations. The computation time was on the order of 5 minutes for both algorithms.

## V. Discussion and conclusion

We have presented a wavelet-based multilevel image-restoration algorithm inspired from multigrid techniques. The method is one order of magnitude faster than the standard algorithm for sparsity-constrained restoration, whose results belong to the state-of-the-art in the field of image processing.

The MLTL algorithm allows for typical multigrid iteration schemes such as V-cycles and W-cycles. However, it differs from textbook multigrid schemes, which iterate on nested subspaces corresponding to different "resolution levels" of the inverse problem. Our algorithm, by contrast, iterates on the wavelet subbands, which are the complements of the standard multigrid spaces. Yet our algorithm takes advantage of the underlying multiresolution structure, which greatly contributes to the efficiency of the method. The other key point is the preconditioning effect of wavelets.

We have provided theoretical convergence rates for the linear parts of the TL and MLTL algorithms, giving a quantitative insight into the convergence acceleration of the latter. Our experimental results show that one can achieve the same kind of acceleration in the general non-linear case (with thresholding).

Our method is directly applicable to separable wavelet bases in arbitrary dimensions. We obtained promising results in the context of 3D fluorescence microscopy using such bases, extending our previous work [12]. Nevertheless we have tried to provide a sufficiently general description that should require little adaptation for more "exotic" wavelet representations, e.g. with quincunx subsampling schemes [57], or non-stationary refinement filters [58], [59]. The algorithm is readily implementable using our modular specification and standard wavelet-decomposition/reconstruction building blocks.

We are currently investigating the benefits of the MLTL algorithm for redundant wavelet representations. We have not specifically explored this possibility here because we were primarily interested in high-dimensional inverse problems that do typically not allow for redundant decompositions. We have already obtained promising results for medical applications (specifically fMRI signal restoration and tomographic image reconstruction) which will be the subject of forthcoming reports.

## ACKNOWLEDGMENTS

This work was funded in part by the Hasler Foundation.


## APPENDIX

### A. A method for precomputing the bound constants and the correction filters

To keep the presentation simple, we will consider the one-dimensional ($D = 1$) situation where $\mathbf{H}^T\mathbf{H}$ is a (positive) circulant matrix. Its eigenvalues (DFT coefficients) are real and positive and can thus be denoted $|\hat{h}[\nu]|^2$, $\nu = 0, \ldots, N-1$. Furthermore, we will consider a wavelet decomposition with a dyadic subsampling scheme. The method presented below can easily be extended to higher dimensions and more general wavelet representations.

To compute the bound constants defined in (11), one must essentially estimate the inner products of (12). These can be rewritten in the frequency domain as

$$\langle \mathbf{H}\mathbf{W}_{s_2}\mathbf{w}_{s_2}, \mathbf{H}\mathbf{W}_{s_1}\mathbf{w}_{s_1} \rangle = \frac{1}{N} \sum_{\nu=0}^{N-1} |\hat{h}[\nu]|^2 \, \hat{\psi}^*_{s_2}[\nu] \, \hat{w}^*_{s_2}[\nu] \, \hat{\psi}_{s_1}[\nu] \, \hat{w}_{s_1}[\nu],$$

where we use the following conventions.

- $\hat{w}_s[\nu]$ denotes the DFT of the wavelet coefficients corresponding to a subband $s$ at a given level $j$. Since we assume a dyadic subsampling scheme, $\hat{w}_s[\nu]$ can be seen as an $N_j$-periodic sequence, with $N_j = N/2^j$.
- $\hat{\psi}_s[\nu]$ denotes the DFT of the wavelet or scaling function that spans the subspace associated with subband $s$. Note that if we define $\hat{\psi}_0[\nu] = 1$, the discrete version of the standard scaling relation [26] can be stated as

$$\hat{\psi}_s[\nu] = \hat{g}_s[\nu]\hat{\psi}_{j-1}[\nu] \text{ for } s \in S_j \cup \{j\}, \tag{14}$$

  where $\hat{g}_s[\nu]$ is the $N_j$-periodic filter corresponding to $\mathbf{G}_s$ (see below).

Our multilevel method only requires the explicit value of the constants $\rho_{s_2\, s_1}$ when $s_2$ and $s_1$ are subbands located at the same level, that is, when $s_2, s_1 \in S_j \cup \{j\}$ for some $j$. In this case $\hat{w}_{s_2}[\nu]$ and $\hat{w}_{s_1}[\nu]$ have the same period $N_j$. We can thus write that

$$\langle \mathbf{H}\mathbf{W}_{s_2}\mathbf{w}_{s_2}, \mathbf{H}\mathbf{W}_{s_1}\mathbf{w}_{s_1} \rangle = \frac{1}{N_j} \sum_{\nu=0}^{N_j-1} \hat{w}^*_{s_2}[\nu] \, \hat{w}_{s_1}[\nu] \hat{c}_{s_2\, s_1}[\nu],$$

where

$$\hat{c}_{s_2\, s_1}[\nu] = \frac{1}{2^j} \sum_{k=0}^{2^j-1} |\hat{h}[\nu + kN_j]|^2 \, \hat{\psi}^*_{s_2}[\nu + kN_j] \, \hat{\psi}_{s_1}[\nu + kN_j]. \tag{15}$$

Inequality (12) is then obtained by defining $\rho_{s_2\, s_1} = \max_\nu |\hat{c}_{s_2\, s_1}[\nu]|$.

When $\mathbf{H}^T\mathbf{H}$ is a circulant matrix and $s \in S_j \cup \{j\}$, the matrix $\mathbf{C}_s = \mathbf{W}_s^T\mathbf{H}^T\mathbf{H}\mathbf{W}_j$ is also circulant and its DFT coefficients are precisely given by $\hat{c}_{s\,j}[\nu]$, for $\nu = 0, \ldots, N_j$. This property allows for

an efficient frequency-domain implementation of the residual correction steps in the CFTL and MLTL algorithms.

To prove the property, we introduce the general notation $\mathbf{C}_{s_2\,s_1} = \mathbf{W}_{s_2}^T \mathbf{H}^T \mathbf{H} \mathbf{W}_{s_1}$ for arbitrary subbands $s_2, s_1$. We also define $\mathbf{C}_{00} = \mathbf{H}^T\mathbf{H}$. We can then proceed by recurrence. For $j = 0$, $\mathbf{C}_{00}$ is circulant and its DFT coefficients are given by (15) with $s_2 = s_1 = 0$. For $j \geq 1$, we assume that $\mathbf{C}_{j-1\,j-1}$ is a circulant matrix whose DFT coefficients are $\hat{c}_{j-1\,j-1}[\nu]$. The cascade structure of the wavelet representation (4) implies that, for $s_2, s_1 \in S_j \cup \{j\}$,

$$\mathbf{C}_{s_2\,s_1} = \mathbf{G}_{s_2}^T \mathbf{C}_{j-1\,j-1} \mathbf{G}_{s_1}.$$

For a given subband $s$, the algorithmic interpretation of $\mathbf{G}_s$ is 1) dyadic upsampling, followed by 2) filtering with $\hat{g}_s[\nu]$. Its transpose $\mathbf{G}_s^T$ stands for 1) filtering with $\hat{g}_s[\nu]^*$, followed by 2) dyadic downsampling. Therefore, $\mathbf{C}_{s_2\,s_1}$ is also a circulant matrix with DFT coefficients

$$\frac{1}{2}\sum_{k=0}^{1} \hat{g}_{s_2}[\nu + kN_j]^*\, \hat{c}_{j-1\,j-1}[\nu + kN_j]\, \hat{g}_{s_1}[\nu + kN_j] = \hat{c}_{s_2\,s_1}[\nu]. \tag{16}$$

The equality stems from definition (15) for $\hat{c}_{j-1\,j-1}[\nu]$, and from the scaling relation (14); this completes the proof by recurrence.

Note that relation (16) provides a way to recursively compute the filters $\hat{c}_{s_2\,s_1}[\nu]$ and the corresponding constants $\rho_{s_2\,s_1}$ (with $s_2, s_1 \in S_j \cup \{j\}$).

### *B. Pseudo-code description of the general MLTL algorithm*

As in the previous subsection, we use the notation $\mathbf{C}_{s_2\,s_1} = \mathbf{W}_{s_2}^T \mathbf{H}^T \mathbf{H} \mathbf{W}_{s_1}$. The general MLTL algorithm uses both the matrices $\mathbf{C}_{j\,s}$ and $\mathbf{C}_{s\,j}$, for $s \in S_j \cup \{j\}$. It is useful to observe that $\mathbf{C}_{j\,s} = \mathbf{C}_{s\,j}^T$: in the convolutive case, this implies that $\hat{c}_{j\,s}[\nu] = \hat{c}_{s\,j}[\nu]^*$.

**Algorithm 3** UpdateResidual($j$)

- If $\mathbf{e}_s \neq \mathbf{0}$ for some $s \in S_j$:
  - $\mathbf{r}_{j-1} \leftarrow \mathbf{r}_{j-1} - \mathbf{C}_{j-1} \sum_{s \in S_j} \mathbf{G}_s \mathbf{e}_s$
  - For every $s \in S_j \cup \{j\}$, $\mathbf{r}_s \leftarrow \mathbf{G}_s^T \mathbf{r}_{j-1}$
- Otherwise, if $\mathbf{e}_j \neq \mathbf{0}$: for every $s \in S_j$, $\mathbf{r}_s \leftarrow \mathbf{r}_s - \mathbf{C}_{s\,j} \mathbf{e}_j$

### *C. Proof of Property 2*

A short computation reveals that

$$\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*) = \|\mathbf{H}\mathbf{W}(\mathbf{w} - \mathbf{w}^*)\|_2^2 - 2\mathfrak{Re}\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*), \mathbf{w} - \mathbf{w}^* \rangle + \lambda\|\mathbf{w}\|_1 - \lambda\|\mathbf{w}^*\|_1, \tag{17}$$

**Algorithm 4** UpdateLevel($j$)

- Initialization:
  - For every $s \in S_j \cup \{j\}$, $\mathbf{e}_s \leftarrow \mathbf{0}$
  - For every $s \in S_j \cup \{j\}$, $\mathbf{r}_s \leftarrow \mathbf{G}_s^T \mathbf{r}_{j-1}$
- Repeat $\mu$ times:
  - Repeat $\eta_1$ times:
    - UpdateResidual($j$)
    - SLTL($j$)
  - If $j < J$:
    - If $\mathbf{e}_s \neq \mathbf{0}$ for some $s \in S_j$:
      - If $\mathbf{e}_j \neq \mathbf{0}$, UpdateResidual($j$)
      - Otherwise, $\mathbf{r}_j \leftarrow \mathbf{r}_j - \sum_{s \in S_j} \mathbf{C}_{j\,s} \mathbf{e}_s$
    - $\mathbf{e}_j \leftarrow \mathbf{w}_j$
    - UpdateLevel($j+1$)
    - $\mathbf{e}_j \leftarrow \mathbf{w}_j - \mathbf{e}_j$
  - Repeat $\eta_2$ times:
    - UpdateResidual($j$)
    - SLTL($j$)
- $\mathbf{w}_{j-1} \leftarrow \sum_{s \in S_j \cup \{j\}} \mathbf{G}_s \mathbf{w}_s$

**Algorithm 5** MLTL

- Initialization:
  - Choose some initial estimate $\mathbf{x}$ and set $\mathbf{w}_0 \leftarrow \mathbf{x}$
  - Compute its wavelet decomposition (keeping the coarse approximations):
    for $j = 1, \ldots, J$, for every $s \in S_j \cup \{j\}$, $\mathbf{w}_s = \tilde{\mathbf{G}}_s^T \mathbf{w}_{j-1}$
- Repeat $K$ times:
  - $\mathbf{r}_0 \leftarrow \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{w}_0)$
  - UpdateLevel(1)
- Set $\mathbf{x} \leftarrow \mathbf{w}_0$ and return $\mathbf{x}$

where $\mathbf{w}$ and $\mathbf{w}^*$ are arbitrary vectors of wavelet coefficients.

- *Necessary part:* we assume that $\mathcal{C}(\mathbf{w}) \geq \mathcal{C}(\mathbf{w}^*)$ for every $\mathbf{w}$.
  - Suppose that $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n| > \lambda/2$ for some $n$. Given a real and strictly positive

constant $\epsilon$, we define the vector $\mathbf{w}$ by

$$[\mathbf{w}]_{n'} = \begin{cases} [\mathbf{w}^*]_{n'} + \epsilon \ \mathrm{sgn}\,[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n & \text{if } n' = n; \\ [\mathbf{w}^*]_{n'} & \text{otherwise.} \end{cases}$$

In view of (17), $\epsilon$ can always be chosen such that $\mathcal{C}(\mathbf{w}) < \mathcal{C}(\mathbf{w}^*)$, a contradiction. Therefore it must be that $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n| \leq \lambda/2$ for every $n$.

– Choosing $\epsilon \in ]0, 1]$ and inserting $\mathbf{w} = (1 - \epsilon)\mathbf{w}^*$ into (17) gives the necessary condition

$$\epsilon^2 ||\mathbf{H}\mathbf{W}\mathbf{w}^*||_2^2 + 2\epsilon \mathfrak{Re}\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*), \mathbf{w}^* \rangle - \epsilon\lambda||\mathbf{w}^*||_1 \geq 0.$$

If it were true that $2\mathfrak{Re}\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*), \mathbf{w}^* \rangle - \lambda||\mathbf{w}^*||_1 < 0$, we could find a sufficiently small $\epsilon$ such that this necessary condition is violated. Thus it must be that $2\mathfrak{Re}\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*), \mathbf{w}^* \rangle - \lambda||\mathbf{w}^*||_1 \geq 0$. Since $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n| \leq \lambda/2$ for every $n$, it follows that $[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n = \lambda/2 \ \mathrm{sgn}\,[\mathbf{w}^*]_n$ whenever $[\mathbf{w}^*]_n \neq 0$.

The combination of both results is equivalent to the fixed-point property.

- *Sufficient part:* $\mathbf{w}^*$ is assumed to be a fixed point of the TL algorithm.

  We use the same equivalence:

  – Since $[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n = \lambda/2 \ \mathrm{sgn}\,[\mathbf{w}^*]_n$ whenever $[\mathbf{w}^*]_n \neq 0$, we know that $2\mathfrak{Re}\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*), \mathbf{w}^* \rangle - \lambda||\mathbf{w}^*||_1 = 0$ and (17) reduces to

$$\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*) = ||\mathbf{H}\mathbf{W}(\mathbf{w} - \mathbf{w}^*)||_2^2 - 2\mathfrak{Re}\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*), \mathbf{w} \rangle + \lambda||\mathbf{w}||_1.$$

  – Since $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n| \leq \lambda/2$ for every $n$, it follows that

$$\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*) \geq ||\mathbf{H}\mathbf{W}(\mathbf{w} - \mathbf{w}^*)||_2^2 \geq 0.$$

  This also shows the unicity of the minimizer when $\mathbf{W}^T\mathbf{H}^T\mathbf{H}\mathbf{W}$ is positive definite.

## REFERENCES


[1] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser. The colored revolution of bioimaging. *IEEE Signal Processing Magazine*, 23(3):20–31, May 2006.

[2] P. Sarder and A. Nehorai. Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Processing Magazine*, 23(3):32–45, May 2006.

[3] A. K. Louis. Medical imaging: state of the art and future development. *Inverse Problems*, 8(5):709–738, October 1992.

[4] J. Beutel, H. L. Kundel, and R. L. Van Metter, editors. *Handbook of Medical Imaging: Physics and Psychophysics*, volume 1. SPIE Press, February 2000.

[5] R. Molina, J. Nunez, F. J. Cortijo, and J. Mateos. Image restoration in astronomy: a Bayesian perspective. *IEEE Signal Processing Magazine*, 18(2):11–29, March 2001.

[6] J. L. Starck, E. Pantin, and F. Murtagh. Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific*, 114(800):1051–1069, October 2002.

[7] M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, August 2003.

[8] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, August 2004.

[9] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle. A $\ell^1$-unified variational framework for image restoration. *Lecture Notes in Computer Science - Proceedings of ECCV 2004 Part IV*, 3024:1–13, 2004.

[10] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen. Tomographic inversion using $\ell_1$-regularization of wavelet coefficients. *Geophysical Journal International*, 170(1):359–370, July 2007.

[11] J. M. Bioucas-Dias and M. A. T. Figueiredo. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, December 2007.

[12] C. Vonesch and M. Unser. A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution. *IEEE Transactions on Image Processing*, 17(4):539–549, April 2008.

[13] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[14] W. Hackbusch. *Multi-grid methods and applications*, volume 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1985.

[15] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. Society for Industrial and Applied Mathematics, second edition, 2000.

[16] W. M. Lawton. Multiresolution properties of the wavelet Galerkin operator. *Journal of Mathematical Physics*, 32(6):1440–1443, June 1991.

[17] Z. Cai and W. E. Hierarchical method for elliptic problems using wavelet. *Communications in Applied Numerical Methods*, 8(11):819–825, November 1992.

[18] W. Dahmen and A. Kunoth. Multilevel preconditioning. *Numerische Mathematik*, 63(1):315–344, December 1992.

[19] W. L. Briggs and V. E. Henson. Wavelets and multigrid. *SIAM Journal on Scientific Computing*, 14(2):506–510, March 1993.

[20] G. Wang, J. Zhang, and G.-W. Pan. Solution of inverse problems in image processing by wavelet expansion. *IEEE Transactions on Image Processing*, 4(5):579–593, May 1995.

[21] K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, March 2000.

[22] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, February 2004.

[23] M. W. Jacobson and J. A. Fessler. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Transactions on Image Processing*, 16(10):2411–2422, October 2007.

[24] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing*, 42(10):2664–2677, October 1994.

[25] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. A general framework for nonlinear multigrid inversion. *IEEE Transactions on Image Processing*, 14(1):125–140, January 2005.

[26] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.

[27] H. Yserentant. On the multi-level splitting of finite element spaces. *Numerische Mathematik*, 49:379–412, 1986.

[28] R. E. Bank, T. F. Dupont, and H. Yserentant. The hierarchical basis multigrid method. *Numerische Mathematik*, 52(4):427–458, July 1988.

[29] P. S. Vassilevski and J. Wang. Stabilizing the hierarchical basis by approximate wavelets I: Theory. *Numerical Linear Algebra with Applications*, 4(2):103–126, March/April 1998.

[30] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. Institute of Physics Publishing, 1998.

[31] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 still image coding system: an overview. *IEEE Transactions on Consumer Electronics*, 46(4):1103–1127, November 2000.

[32] C. De Mol and M. Defrise. A note on wavelet-based inversion algorithms. In M. Zuhair Nashed and O. Scherzer, editors, *Inverse problems, image analysis and medical imaging*, volume 313 of *Contemporary Mathematics*, pages 85–96, Providence, RI, November 2002. American Mathematical Society. Proceedings of the Special Session on Interaction of Inverse Problems and Image Analysis. AMS Joint Mathematics Meetings. New Orleans, LA. January 10-13, 2001.

[33] R. D. Nowak and M. A. T. Figueiredo. Fast wavelet-based image deconvolution using the EM algorithm. In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers. Pacific Grove, CA. November 4-11, 2001.*, volume 1, 2001.

[34] J. L. Starck, D. L. Donoho, and E. J. Candès. Astronomical image representation by the curvelet transform. *Astronomy and Astrophysics*, 398(2):785–800, February 2003.

[35] J.-L. Starck, M. K. Nguyen, and F. Murtagh. Wavelets and curvelets for image deconvolution: a combined approach. *Signal Processing*, 83(10):2279–2283, October 2003.

[36] L. Landweber. An iterative formula for Fredholm integral equations of the first kind. *American Journal of Mathematics*, 73(3):615–624, July 1951.

[37] J. B. Weaver, Y. S. Xu, D. M. Healy, and L. D. Cromwell. Filtering noise from images with wavelet transforms. *Magnetic Resonance in Medicine*, 21(2):288–295, October 1991.

[38] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[39] M. Liebling and M. Unser. Autofocus for digital fresnel holograms by use of a fresnelet-sparsity criterion. *Journal of the Optical Society of America A*, 21(12):2424–2430, December 2004.

[40] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4):1495–1518, August 2007.

[41] J. M. Bioucas-Dias. Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors. *IEEE Transactions on Image Processing*, 15(4):937–951, April 2006.

[42] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16(12):2980–2991, December 2007.

[43] M. Elad. Why simple shrinkage is still relevant for redundant representations? *IEEE Transactions on Information Theory*, 52(12):5559–5569, December 2006.

[44] M. Elad, B. Matalon, and M. Zibulevsky. Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. *Applied and Computational Harmonic Analysis*, 23(3):346–367, November 2007.

[45] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 2007(4):586–597, December 2007.

[46] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *Annals of Applied Statistics*, 1(2):302–332, December 2007.

[47] M. Fornasier. Domain decomposition methods for linear inverse problems with sparsity constraints. *Inverse Problems*, 23(6):2505–2526, December 2007.

[48] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335, March 1998.

[49] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[50] T. Blu and M. Unser. The fractional spline wavelet transform: Definition and implementation. In *Proceedings of the Twenty-Fifth IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume I, pages 512–515, June 2000.

[51] J. Xu. The method of subspace corrections. *Journal of Computational and Applied Mathematics*, 128(1-2):335–362, March 2001.

[52] K. Zhou and C. K. Rushforth. Image restoration using multigrid methods. *Applied Optics*, 30(20):2906–2912, July 1991.

[53] T.-S. Pan and A. E. Yagle. Numerical study of multigrid implementations of some iterative image reconstruction algorithms. *IEEE Transactions on Medical Imaging*, 10(4):572–288, December 1991.

[54] W. Hackbusch. *Iterative Solution of Large Sparse Systems of Equations*, volume 95 of *Applied Mathematical Sciences*. Springer-Verlag, 1994.

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

[56] T. Blu and F. Luisier. The SURE-LET approach to image denoising. *IEEE Transactions on Image Processing*, 16(11):2778–2786, November 2007.

[57] D. van de Ville, T. Blu, and M. Unser. Isotropic polyharmonic B-splines: Scaling functions and wavelets. *IEEE Transactions on Image Processing*, 14(11):1798–1813, November 2005.

[58] I. Khalidov and M. Unser. From differential equations to the construction of new wavelet-like bases. *IEEE Transactions on Signal Processing*, 54(4):1256–1267, April 2006.

[59] C. Vonesch, T. Blu, and M. Unser. Generalized Daubechies wavelet families. *IEEE Transactions on Signal Processing*, 55(9):4415–4429, September 2007.