A fast iterative thresholding algorithm for wavelet-regularized deconvolution

Cédric Vonesch and Michael Unser

Biomedical Imaging Group, EPFL, Lausanne, Switzerland

ABSTRACT

We present an iterative deconvolution algorithm that minimizes a functional with a non-quadratic wavelet-domain regularization term. Our approach is to introduce subband-dependent parameters into the bound optimization framework of Daubechies et al.; it is sufficiently general to cover arbitrary choices of wavelet bases (non-orthonormal or redundant). The resulting procedure alternates between the following two steps:

1. a wavelet-domain Landweber iteration with subband-dependent step-sizes;
2. a denoising operation with subband-dependent thresholding functions.

The subband-dependent parameters allow for a substantial convergence acceleration compared to the existing optimization method. Numerical experiments demonstrate a potential speed increase of more than one order of magnitude. This makes our "fast thresholded Landweber algorithm" a viable alternative for the deconvolution of large data sets. In particular, we present one of the first applications of wavelet-regularized deconvolution to 3D fluorescence microscopy.

Keywords: deconvolution, fast, wavelets, sparsity, subband-dependent, thresholding, Landweber, iterative, bound optimization.

1. INTRODUCTION

Wavelets have proven to be a very successful tool for the estimation of signals that are corrupted by noise. The first example of image denoising based on wavelet coefficient thresholding came from medical imaging.[1] Later, the asymptotic performance of this approach was demonstrated in a statistical framework.[2] Another interpretation of wavelet thresholding can be given in a variational setting[3] and is summarized below.

Let us denote by $\mathbf{y}$ a vector containing noisy measurements of some original signal $\mathbf{x}_{\text{orig}}$. Let $\mathbf{W}$ be a matrix that reconstructs a signal from its wavelet coefficients. To estimate $\mathbf{x}_{\text{orig}}$, it is proposed to compute $\mathbf{x} = \mathbf{W}\mathbf{w}$, where the wavelet coefficients $\mathbf{w}$ minimize the cost function

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

Here, $\|\mathbf{y} - \mathbf{W}\mathbf{w}\|_2^2$ is the squared Euclidian norm of the "data mismatch", while $\|\mathbf{w}\|_1$ represents the sum of the absolute values of the wavelet coefficients of the estimate. The latter term, whose influence can be controlled by the scalar λ, favors estimates with a sparse wavelet expansion. This constraint makes sense because many "natural signals" exhibit this property. When $\mathbf{W}$ is orthonormal, Parseval's relation can be used to rewrite (1) as

$$J(\mathbf{w}) = \|\mathbf{W}^T\mathbf{y} - \mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1, \tag{2}$$

where ^{T} denote Hermitian transposition. Because the components of $\mathbf{w}$ are decoupled in this expression, it can be minimized using a simple pointwise processing. More precisely, the minimization of (1) boils down to soft-thresholding $\mathbf{W}^T\mathbf{y}$ (the wavelet coefficients of $\mathbf{y}$) with threshold level $\lambda/2$. This leads to

$$\mathbf{x} = \mathbf{W}\mathcal{T}_{\lambda/2}\{\mathbf{W}^T\mathbf{y}\}, \tag{3}$$

where $\mathcal{T}_{\lambda/2}\{\ \}$ indicates a componentwise application of the soft-thresholding function; for real coefficients, it is defined by $\mathcal{T}_{\lambda/2}(w) = \text{sgn}(w)\,(|w| - \lambda/2)_+$.

A similar approach can also be deployed for the regularization of inverse problems. Consider the image formation equation $\mathbf{y} = \mathbf{H}\mathbf{x}_{\text{orig}} + \mathbf{b}$, where $\mathbf{b}$ represents a noise contribution and $\mathbf{H}$ is a matrix that models the linear distortions introduced by a measurement device. In the sequel, we will assume that $\mathbf{H}$ is a block-circulant matrix corresponding to the impulse response of a shift-invariant system. The restoration problem then becomes a joint deblurring and denoising problem, and the functional takes the form

$$J(\mathbf{w}) = \|\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1. \tag{4}$$

If $\lambda = 0$, the functional reduces to the quadratic data-term. Its (least-squares) minimizer can be obtained directly provided that one can compute the Moore-Penrose inverse of $\mathbf{H}$. However — to balance the noise-amplification that is inherent to this solution — the least-squares estimate is usually constructed iteratively. The well-known Landweber iteration[4] can be seen as a gradient-based minimization of the data term, with a fixed step size τ. Starting from an initial estimate $\mathbf{x}^{(0)}$, we recursively set

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \tau\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}). \tag{5}$$

Several research groups[5–8] have independently derived an algorithm to handle the case where $\lambda \neq 0$. The idea is essentially to alternate between (3) and (5) (note that the threshold level is adapted, though):

$$\mathbf{x}^{(n+1)} = \mathbf{W}\mathcal{T}_{\lambda\tau/2}\{\mathbf{W}^T[\mathbf{x}^{(n)} + \tau\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})]\}. \tag{TL}$$

Figueiredo and Nowak have shown that this "thresholded Landweber" (TL) algorithm yields results that are comparable to the best deconvolution methods available.[5] However, its convergence is rather slow, requiring several hundreds of iterations at high SNR levels. This is due to the fact that the step size τ (and thus the threshold level $\lambda\tau/2$) must be chosen relatively small. This considerably limits the effect of each iteration on the current estimate.

In this paper, we propose a generalized version of the TL algorithm that provides significantly faster convergence. We revisit the bound optimization formulation of Daubechies et al. in Section 2, leading to an algorithm with subband-dependent parameters. The performance of the method is illustrated by the experimental results of Section 3.

2. BOUND OPTIMIZATION, REVISITED

Similarly to Daubechies et al.,[7] we replace the minimization of $J(\mathbf{x})$ by the successive minimization of a series of auxiliary functionals. However, our approach is different in at least two aspects. First, it is sufficiently general to account for arbitrary wavelet bases (non-orthonormal or even redundant). Second, it takes into account the subband structure of the wavelet basis, thus allowing for auxiliary functionals that are closer to the original cost function; as a result, we need fewer iterations for its minimization.

We consider auxiliary functionals that depend on the current estimate, say $\mathbf{w}^{(n)}$, as follows:

$$J_{\text{n}}(\mathbf{w}) = \sum_{j\in S}\left(\alpha_j\|\mathbf{w}_j^{(n)} - \mathbf{w}_j\|_2^2\right) + J(\mathbf{w}) - \|\mathbf{H}\mathbf{W}(\mathbf{w}^{(n)} - \mathbf{w})\|_2^2. \tag{6}$$

In this definition, the subscript j refers to the different wavelet subbands (which are indexed by the set S). In particular, the vector $\mathbf{w}_j$ contains the coefficients of $\mathbf{w}$ corresponding to the j-th wavelet subband. The constants α_j are chosen such that the auxiliary functional is an upper-bound of the original cost function, that is to say, $J_n(\mathbf{w}) \geq J(\mathbf{w})$ for every $\mathbf{w}$. Moreover, we clearly have equality at the current estimate, i.e., $J_n(\mathbf{w}^{(n)}) = J(\mathbf{w}^{(n)})$. These properties guarantee that, by minimizing $J_n(\mathbf{w})$, we will also decrease $J(\mathbf{w})$. Indeed, if $\mathbf{w}^{(n+1)}$ minimizes $J_n(\mathbf{w})$, we have that

$$J(\mathbf{w}^{(n+1)}) \leq J_n(\mathbf{w}^{(n+1)}) \leq J_n(\mathbf{w}^{(n)}) = J(\mathbf{w}^{(n)}).$$

This leads to an iterative (so-called *bound optimization*) algorithm of the form

$$\mathbf{w}^{(n+1)} = \arg\min_{\mathbf{w}} J_n(\mathbf{w}).$$

The advantage of using (6) as an auxiliary functional is that one can derive an explicit expression of its minimizer. Let us introduce the reconstruction matrices $\mathbf{W}_j$ such that $\mathbf{W}\mathbf{w} = \sum_{j \in S} \mathbf{W}_j \mathbf{w}_j$. Then a simple computation reveals that

$$J_n(\mathbf{w}) = C + \sum_{j \in S} \alpha_j \Big(\|\mathbf{w}_j^{(n)} + \alpha_j^{-1} \mathbf{W}_j^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{(n)}) - \mathbf{w}_j\|_2^2 + \lambda \alpha_j^{-1} \|\mathbf{w}_j\|_1 \Big),$$

where C is a constant that does not depend on $\mathbf{w}$, and the transpose of $\mathbf{W}_j$ plays the role of a decomposition matrix for subband j. If we omit the constant, $J_n(\mathbf{w})$ is a (positively) weighted sum of "subfunctionals" that depend on different subbands. Therefore, each subband can be optimized independently. Moreover, each subfunctional has the same form as (2), which means that we can use essentially the same thresholding operation as before for its minimization. The main difference is that $\mathbf{W}^T\mathbf{y}$ has been replaced by $\mathbf{w}_j^{(n)} + \alpha_j^{-1} \mathbf{W}_j^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{(n)})$, which can be interpreted as a wavelet-domain Landweber iteration with step size $\tau_j = \alpha_j^{-1}$. Also, the threshold level must be adapted ($\lambda \tau_j / 2$). This leads to the following iterative procedure on the wavelet subbands of the estimate:

$$\mathbf{w}_j^{(n+1)} = \mathcal{T}_{\lambda \tau_j / 2} \{ \mathbf{w}_j^{(n)} + \tau_j \mathbf{W}_j^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{(n)}) \}. \tag{FTL}$$

The standard TL algorithm can be obtained as a particular case of the previous derivation, using the same constant α for all subbands in (6). By adapting the constants α_j, we are able to use larger step sizes and threshold levels. Thus we call the above procedure the fast thresholded Landweber (FTL) algorithm.

When $\mathbf{W}$ corresponds to a Shannon wavelet basis, we retrieve the fast algorithm that we have previously derived.[9] An important difference is that the above description is also applicable to redundant decompositions. The advantage of using a Shannon wavelet basis is that the different subbands are completely decoupled when computing the wavelet-domain Landweber update. This choice also greatly simplifies the determination of suitable constants α_j. The preliminary results presented in the next section were obtained with this wavelet basis.

3. EXPERIMENTAL RESULTS

We first performed a synthetic experiment. The standard MRI image was blurred with a Gaussian filter of parameter $\sigma = 2$ and corrupted by white Gaussian noise (see Fig. 1). We then compared the convergence speed of the TL and FTL algorithms, for $\lambda \simeq 0.0199$. We used cycle spinning for both algorithms, as in the work of Figueiredo and Nowak.[5] After five iterations, the result of the FTL algorithm seems to be closer to the original image. This is confirmed by the quantitative measurements presented in Fig. 2, which shows the SER gain (in dB) as a function of the number of iterations. After roughly 50 iterations, the FTL algorithm achieves a gain of 8 dB, a figure that would take more than 500 iterations with the TL algorithm.

In our second experiment we deconvolved a 3D image stack obtained in widefield fluorescence microscopy. The data was actually obtained using a confocal microscope with the pinhole opened to the maximum, so as

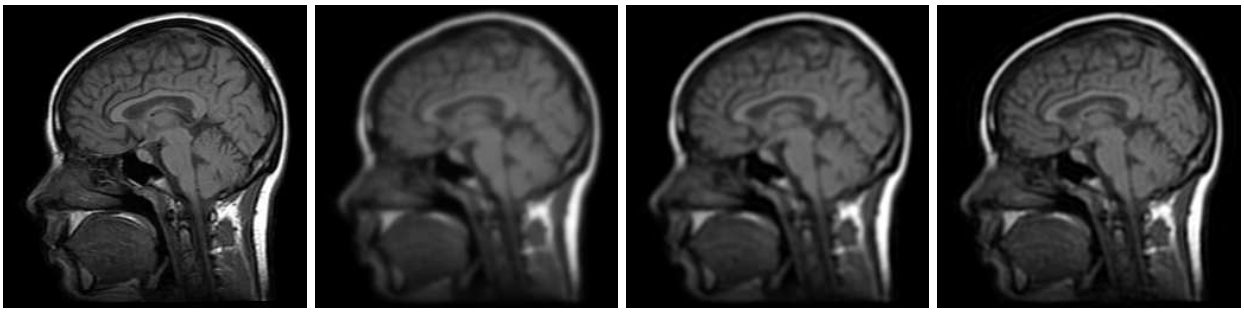

Figure 1. (From left to right) The original MRI image; blurred and noisy version (BSNR = 40 dB); deconvolution result after 5 iterations of the TL algorithm; result after 5 FTL iterations.

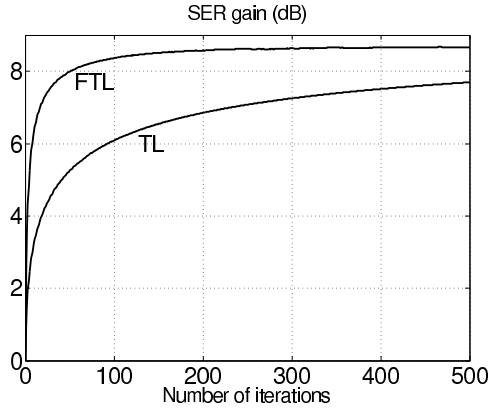

Figure 2. The gain in signal-to-error ratio (SER) as a function of the number of iterations.

to simulate a widefield system. This allowed us to produce a confocal reference image of the same object (by closing the pinhole to 1 Airy unit), with considerably reduced out-of-focus blur.

We observed a biological preparation containing fibroblast cells whose actin filaments were labeled with a phalloidin marker (Alexa488). Fig. 3 shows maximum intensity projections of a $384 \times 288 \times 32$ volume of interest. The center images are deconvolution results after 10 iterations of the TL and FTL algorithms, respectively. While the former still contains a lot of out-of-focus blur, the latter is clearly sharper. The spatial contrast of the FTL result may actually be compared to the confocal image of the same region (on the far right). However, the confocal image is relatively noisy because the signal intensity is reduced by the small pinhole configuration. Let us mention that the confocal image was taken before the widefield image; thereby, potential bleaching effects would mainly affect the widefield image. Also, we used the same laser power and scan frequency in both cases.

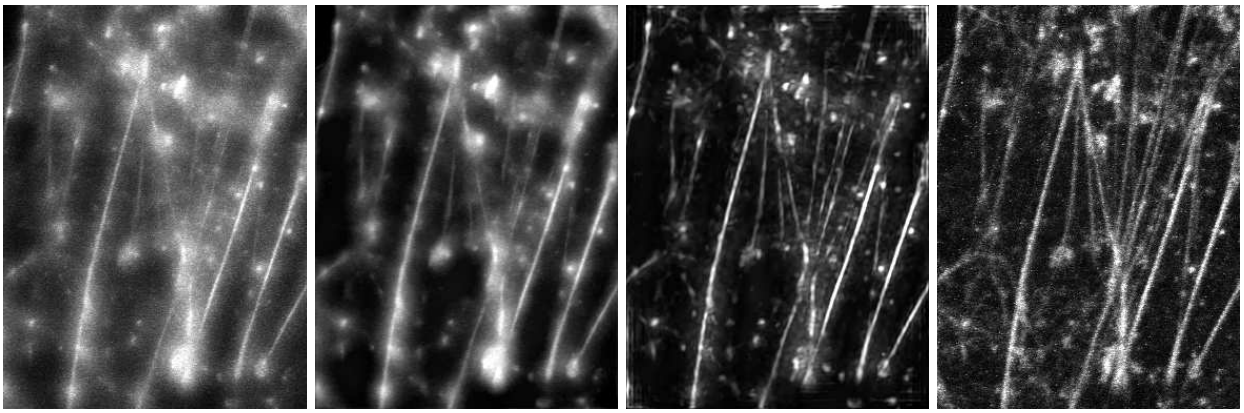

Figure 3. (From left to right) Original widefield image; deconvolution result after 10 TL iterations; result after 10 FTL iterations; confocal reference.

4. CONCLUSION

The fast thresholded Landweber (FTL) algorithm can provide an acceleration of more than one order of magnitude compared to the standard TL method. This considerably broadens the potential application field of wavelet-regularized deconvolution. A very promising area is 3D deconvolution microscopy, where the computational cost of restoration algorithms is severely limited due to the size of the data sets.

ACKNOWLEDGMENTS

This work was funded in part by the Hassler Foundation. The first author would like to thank Jessica Dessimoz for preparing the sample for the second experiment.

REFERENCES

1. 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**, pp. 288–295, October 1991.
2. D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika* **81**, pp. 425–455, August 1994.
3. 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**, pp. 319–335, March 1998.
4. M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*, Institute of Physics Publishing, 1998.
5. M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing* **12**, pp. 906–916, August 2003.
6. J.-L. Starck, M. K. Nguyen, and F. Murtagh, "Wavelets and curvelets for image deconvolution: a combined approach," *Signal Processing* **83**, pp. 2279–2283, October 2003.
7. 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**, pp. 1413–1457, August 2004.
8. J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A ℓ^1-unified variational framework for image restoration," *Lecture Notes in Computer Science - Proceedings of ECCV 2004 Part IV* **3024**, pp. 1–13, 2004.
9. C. Vonesch and M. Unser, "Fast wavelet-regularized image deconvolution," in *Proceedings of the 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Washington, DC, April 12-15, 2007*, pp. 608–611, April 2007.