FAST WAVELET-REGULARIZED IMAGE DECONVOLUTION

Cédric Vonesch and Michael Unser

Biomedical Imaging Group, EPFL, Lausanne, Switzerland

ABSTRACT

We present a modified version of the deconvolution algorithm introduced by Figueiredo and Nowak, which leads to a substantial acceleration. The algorithm essentially consists in alternating between a Landweber-type iteration and a wavelet-domain denoising step.

Our key innovations are 1) the use of a Shannon wavelet basis, which decouples the problem accross subbands, and 2) the use of optimized, subband-dependent step sizes and threshold levels.

At high SNR levels, where the original algorithm exhibits slow convergence, we obtain an acceleration of one order of magnitude. This result suggests that wavelet-domain ℓ_1-regularization may become tractable for the deconvolution of large datasets, e.g. in fluorescence microscopy.

Keywords: deconvolution, wavelets, iterative methods.

1. INTRODUCTION

During the past decade, biological imaging has been revolutionized by the rapid diffusion of novel fluorescent labeling techniques and digital microscopy [1]. Researchers in the life sciences are increasingly relying on 3D deconvolution to improve the resolution of their brightfield or confocal micrographs. Presently, the major challenge is the size of these data sets, which explains why commercial deconvolution packages for biomicroscopy employ relatively simple algorithms. Nonlinear deconvolution, based on more sophisticated regularization schemes, is still considered to be out of reach, because it is computationnally too demanding.

In this paper, we present preliminary results that demonstrate the feasibility of faster deconvolution algorithms for wavelet-domain ℓ_1-regularization. Our target is a computational cost corresponding to no more than 30 iterations of a classical Landweber or Richardson-Lucy algorithm, which constitutes the present tolerance limit for typical 3D data sets.

Consider the image-formation model in Fig. 1. The problem of deconvolution lies in the recovery of the characteristic function x_{orig} of an object which is observed through a convolutive imaging system. In microscopy, the measured image corresponds to a 3D convolution of x_{orig} with h, the impulse response of the system (aka point spread function); in addition, a noise component n corrupts the measurements.

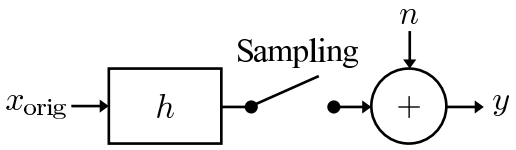

Fig. 1. The image-formation model.

We will assume that, after a suitable discretization, the image-formation process in d dimensions can be summarized by the matrix equation

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

Here, the boldface symbols $\mathbf{y}$, $\mathbf{x}_{\text{orig}}$ and $\mathbf{n}$ represent vectors of dimension N^d containing uniform samples of the signals y, x_{orig} and n respectively. We shall further assume periodic boundary conditions. This means that the matrix $\mathbf{H}$, which approximates the convolution with h in the continuous domain, is taken to be block-circulant.

It is well-known that the recovery of $\mathbf{x}_{\text{orig}}$ from $\mathbf{y}$ is an ill-posed problem [2]. Most approaches for overcoming this ill-posedness can be described in a variational framework, where the quality of an estimate $\mathbf{x}$ of $\mathbf{x}_{\text{orig}}$ is measured by a cost-function $J(\mathbf{x})$; the lower this cost function, the better the estimate. Typically $J(\mathbf{x})$ is composed of two terms: a *data term* that measures the mismatch between the estimate and the measured signal, and a *regularization term* that favors certain properties of the estimate, based on a priori knowledge of the original signal. The influence of both terms is balanced by a *regularization parameter* λ.

Recently, several research groups have advocated wavelet-domain regularization [3, 4, 5, 6], based on the assumption that many natural signals have a sparse wavelet expansion. In this framework, the functional takes the form

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

Here the data term is simply the squared Euclidian norm of the residual in the image domain. $\mathbf{W}$ is an orthonormal wavelet-transform matrix. The regularization term $\|\mathbf{W}\mathbf{x}\|_1$

This work was supported by the Hassler Foundation.

represents the sum of the absolute values of the wavelet coefficients of $\mathbf{x}$. This tends to favor estimates that can be represented with a small number of large wavelet coefficients.

The paper is organized as follows. In Section 2, we first recall the approach of Daubechies et al. [5] for the minimization of (1) — essentially the same algorithm is also derived in [4] and [6]. We then present a faster variant of this algorithm in Section 3. Numerical examples follow in Section 4.

2. THE THRESHOLDED LANDWEBER ALGORITHM

2.1. Bound optimization

Our method belongs to the category of so-called bound optimization algorithms [7]: rather than minimizing the cost function directly, we use a sequence of auxiliary functionals that are easier to minimize.

Similarly to Daubechies et al. [5], we will use functionals of the form

$$\tilde{J}(\mathbf{x},\mathbf{a}) = J(\mathbf{x}) + (\mathbf{x}-\mathbf{a})^T\mathbf{Q}(\mathbf{x}-\mathbf{a}) - \|\mathbf{H}\mathbf{x}-\mathbf{H}\mathbf{a}\|_2^2, \quad (2)$$

where $\mathbf{a}$ is a suitable vector. The negative term is introduced to cancel the coupling due to the convolution matrix $\mathbf{H}$, which hinders the direct minimization of $J(\mathbf{x})$. The formulation presented in [5] uses a matrix $\mathbf{Q}$ of the form $\alpha\mathbf{I}$, where α is a scalar and $\mathbf{I}$ is the identity matrix; but the concept can readily be extended to more general operators (see Remark 2.4 in [5]). The key-point is that $\mathbf{Q}$ be chosen such that:

1. $\mathbf{Q}$ is diagonalized in the wavelet basis; i.e., $\mathbf{Q} = \mathbf{W}^T\mathbf{D}\mathbf{W}$ where $\mathbf{D}$ is diagonal.
2. $\mathbf{Q}-\mathbf{H}^T\mathbf{H}$ is positive definite; i.e., $\mathbf{v}^T(\mathbf{Q}-\mathbf{H}^T\mathbf{H})\mathbf{v} > 0$ for any non-zero vector $\mathbf{v}$.

The first property ensures that $\mathbf{Q}$ does not introduce further coupling in the wavelet domain. The second property has two important consequences:

- Since $J(\mathbf{x})$ is convex, $\tilde{J}(\mathbf{x},\mathbf{a})$ is strictly convex; i.e., it admits a unique minimizer.
- For any given $\mathbf{a}$, $\tilde{J}(\mathbf{x},\mathbf{a})$ is a strict (except when $\mathbf{x}=\mathbf{a}$) upper-bound of $J(\mathbf{x})$. This implies that, if we are able to find the minimizer of $\tilde{J}(\mathbf{x},\mathbf{a})$ (and if this minimizer is different from $\mathbf{a}$), then we are also guaranteed to strictly decrease $J(\mathbf{x})$. If the minimizer is such that $\mathbf{x}=\mathbf{a}$, then we have also reached the minimum of $J(x)$.

Another important remark is that the closer $\mathbf{x}$ is to $\mathbf{a}$, the more $\tilde{J}(\mathbf{x},\mathbf{a})$ can be considered a good approximation of $J(\mathbf{x})$. This leads to the iterative algorithm

$$\mathbf{x}^{(n+1)} = \arg\min_{\mathbf{x}} \tilde{J}(\mathbf{x},\mathbf{x}^{(n)}), \quad (3)$$

whose convergence is established in [5]. This means that the distance $\|\mathbf{x}^{(n+1)}-\mathbf{x}^{(n)}\|_2$ between two successive minimizers of the functionals $\tilde{J}(\mathbf{x},\mathbf{x}^{(n)})$ tends to zero as $n\to\infty$, and essentially implies that their limit is the true minimizer of $J(\mathbf{x})$.

2.2. Daubechies et al.'s derivation revisited

Using $\mathbf{Q} = \mathbf{I}$ as a strict upper-bound of $\mathbf{H}^T\mathbf{H}$ is always possible, provided the largest eigenvalue of $\mathbf{H}^T\mathbf{H}$ is strictly smaller than 1. We shall assume this to be true for the remainder of this paper. With this choice of $\mathbf{Q}$, one can verify that the auxiliary functional defined by (2) is simply

$$\begin{aligned}\tilde{J}(\mathbf{x},\mathbf{a}) &= \|\,\mathbf{x} - \underbrace{(\mathbf{a}+\mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{a}))}_{\mathbf{z}}\,\|_2^2 + \lambda\|\mathbf{W}\mathbf{x}\|_1 + c \\ &= \|\,\mathbf{W}\mathbf{x}-\mathbf{W}\mathbf{z}\|_2^2 + \lambda\|\mathbf{W}\mathbf{x}\|_1 + c, \quad (4)\end{aligned}$$

where c is a constant with respect to $\mathbf{x}$. The second relation is provided by Parseval's equality, since we are considering an orthonormal wavelet basis.

The key point is that the minimization of (4) with respect to $\mathbf{x}$ (or $\mathbf{W}\mathbf{x}$) can be solved in a coordinate-wise fashion in the wavelet-domain. This is known to boil down to a soft-thresholding operation, with $\lambda/2$ as threshold level [8]. The term highlighted as $\mathbf{z}$ may be interpreted as a signal to be denoised. On the other hand, $\mathbf{z}$ corresponds exactly to the update formula of the classical Landweber algorithm [2], when replacing $\mathbf{a}$ by $\mathbf{x}^{(n)}$ according to (3).

Thus, starting from an initial estimate $\mathbf{x}^{(0)}$, the algorithm proposed by Daubechies et al. alternates between two steps:

- a Landweber iteration $\mathbf{x}^{(n+1)} \leftarrow \mathbf{x}^{(n)} + \mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{x}^{(n)})$;
- a denoising operation $\mathbf{x}^{(n+1)} \leftarrow \mathcal{T}_{\lambda/2}\left\{\mathbf{x}^{(n+1)}\right\}$.

Here $\mathcal{T}_{\lambda/2}$ is the (wavelet-domain) soft-thresholding operator.

More precisely, let us index the wavelet subspaces using a scale parameter $j \in [\![0, j_{\max}]\!]$. The corresponding basis elements will be denoted $\mathbf{w}_{j,k}$, where k is a translation parameter belonging to a set T_j. Note that we include the (coarsest-scale) scaling functions in this notation; by convention, they shall correspond to $j=0$. Then $\mathbf{x} = \sum_{j=0}^{j_{\max}}\sum_{k\in T_j} x_{j,k}\mathbf{w}_{j,k}$, where the $x_{j,k}$ are the wavelet (and scaling) coefficients of $\mathbf{x}$, and the soft-thresholding operator is

$$\begin{aligned}\mathcal{T}_{\lambda/2}\{\mathbf{x}\} = &\sum_{k\in T_0} x_{0,k}\mathbf{w}_{0,k} \\ &+ \sum_{j=1}^{j_{\max}}\sum_{k\in T_j} \operatorname{sgn}(x_{j,k})\left(|x_{j,k}|-\lambda/2\right)_+ \mathbf{w}_{j,k}.\end{aligned}$$

This definition can be adapted to complex wavelet decompositions by replacing $\operatorname{sgn}(x_{j,k})$ by $e^{i\arg(x_{j,k})}$.

3. AN ACCELERATED ALGORITHM

Using the previous algorithm, Figueiredo and Nowak have reported numerical results that are competitive with state-of-the-art deconvolution methods [4]. However, at high SNR levels, the convergence is particularly slow (see the numerical examples in Section 4).

In what follows, we construct upper-bounds that approximate the original cost function more closely. Thereby, we will need fewer auxiliary functionals (which means fewer iterations) for its minimization.

3.1. A tighter wavelet-domain bound for $\mathbf{H}^T\mathbf{H}$

From now on, we will assume that $\mathbf{w}_{j,k}$ corresponds to a Shannon wavelet basis [9]. Also, for a vector $\mathbf{v}$, we will denote $(\mathbf{v}_j)_{j\in[\![0,j_{\max}]\!]}$ its projection on the j^{th} wavelet subspace (respectively the scaling function subspace when $j=0$). $(S_j)_{j\in[\![0,j_{\max}]\!]}$ will stand for the frequency-support of the corresponding basis functions. Finally, we let $(\hat{h}[\boldsymbol{\nu}])_{\boldsymbol{\nu}\in[\![0,N-1]\!]^d}$ be the DFT coefficients of the filter that generates the circulant matrix $\mathbf{H}$. This means that the eigenvalues of $\mathbf{H}^T\mathbf{H}$ are the scalars $|\hat{h}[\boldsymbol{\nu}]|^2$. Then the following theorem holds:

Theorem 1. *For any vector* $\mathbf{v}$*,*

$$||\mathbf{H}\mathbf{v}||_2^2 \leq \sum_{j=0}^{j_{\max}} \left(\max_{\boldsymbol{\nu}\in S_j} |\hat{h}[\boldsymbol{\nu}]|^2\right) ||\mathbf{v}_j||_2^2.$$

Proof. Because their frequency-supports are disjoint, the vectors $(\mathbf{H}\mathbf{v}_j)_{j\in[\![0,j_{\max}]\!]}$ are mutually orthogonal. Therefore,

$$||\mathbf{H}\mathbf{v}||_2^2 = \sum_{j=0}^{j_{\max}} ||\mathbf{H}\mathbf{v}_j||_2^2.$$

Introducing the DFT coefficients $\hat{v}_j[\boldsymbol{\nu}]$ of $\mathbf{v}_j$ and applying Parseval's relation, it follows that

$$\begin{aligned}
||\mathbf{H}\mathbf{v}||_2^2 &= \frac{1}{N^d} \sum_{j=0}^{j_{\max}} \sum_{\boldsymbol{\nu}\in[\![0,N-1]\!]^d} \left|\hat{h}[\boldsymbol{\nu}]\hat{v}_j[\boldsymbol{\nu}]\right|^2 \\
&= \frac{1}{N^d} \sum_{j=0}^{j_{\max}} \sum_{\boldsymbol{\nu}\in S_j} \left|\hat{h}[\boldsymbol{\nu}]\hat{v}_j[\boldsymbol{\nu}]\right|^2 \\
&\leq \frac{1}{N^d} \sum_{j=0}^{j_{\max}} \left(\max_{\boldsymbol{\nu}\in S_j} |\hat{h}[\boldsymbol{\nu}]|^2\right) \sum_{\boldsymbol{\nu}\in S_j} |\hat{v}_j[\boldsymbol{\nu}]|^2.
\end{aligned}$$

Reapplying Parseval's relation in the opposite direction completes the proof. □

This result allows us to design an upper-bound of $\mathbf{H}^T\mathbf{H}$ that is tighter than the identity matrix. Indeed, in a given subband, the maximum value of $|\hat{h}[\boldsymbol{\nu}]|^2$ is potentially much smaller than 1. We can thus choose the eigenvalues $d_{j,k} = \mathbf{w}_{j,k}^T\mathbf{Q}\mathbf{w}_{j,k}$ in a subband-dependent manner, i.e., such that $d_{j,k} = d_j$. Provided that $d_j > \max_{\boldsymbol{\nu}\in S_j} |\hat{h}[\boldsymbol{\nu}]|^2$ for all $j \in [\![0, j_{\max}]\!]$, Theorem 1 guarantees that

$$||\mathbf{H}\mathbf{v}||_2^2 < \mathbf{v}^T\mathbf{Q}\mathbf{v} = \sum_{j=0}^{j_{\max}} d_j ||\mathbf{v}_j||_2^2$$

for any non-zero vector $\mathbf{v}$.

3.2. Summary of the proposed algorithm

Now, expanding (2) yields (up to an additive constant)

$$\tilde{J}(\mathbf{x},\mathbf{a}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + \lambda||\mathbf{W}\mathbf{x}||_1 - 2\mathbf{x}^T\left(\mathbf{Q}\mathbf{a} + \mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{a})\right).$$

In our case, this can be rewritten as a sum of terms that depend only on a specific wavelet subband[1]:

$$\sum_{j=0}^{j_{\max}} d_j \left(||\mathbf{x}_j - [\mathbf{a} + \tau_j\mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{a})]_j||_2^2 + \lambda\tau_j||\mathbf{W}\mathbf{x}_j||_1\right),$$

where $\tau_j = 1/d_j$. An important consequence is that we can minimize each term independently.

Even more importantly, the individual terms have the same form as in (4), up to a (positive) multiplicative factor. Therefore, they can be minimized using essentially the same algorithm as before — with the same cost per iteration — but with subband-dependent parameters. We now have to alternate between the following two steps, for each wavelet subband $j \in [\![0, j_{\max}]\!]$:

- a Landweber iteration with an adapted step size,

$$\mathbf{x}_j^{(n+1)} \leftarrow \mathbf{x}_j^{(n)} + \tau_j[\mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{x}^{(n)})]_j;$$

- a denoising operation with an adapted threshold,

$$\mathbf{x}_j^{(n+1)} \leftarrow \mathcal{T}_{\lambda\tau_j/2}\left\{\mathbf{x}_j^{(n+1)}\right\}.$$

In practice, the algorithm alternates between the frequency domain, for the Landweber update, and the wavelet domain, for the thresholding. Additionally, for both algorithms presented in this paper, we use the random shift method described in [4].

4. NUMERICAL RESULTS

We have tested our algorithm with the same synthetic setup as in Fig. 2 of [4], using Matlab. The 256×256 cameraman image was convolved with a 9×9 uniform blur kernel. Zero-mean white Gaussian noise with variance $\sigma^2 = 0.308$ was then added to the result. We used the same initial estimate as in [4]. The only difference is that, instead of Haar

[1] Here $[\mathbf{v}]_j$ stands for the projection on the j^{th} subspace of a vector $\mathbf{v}$.

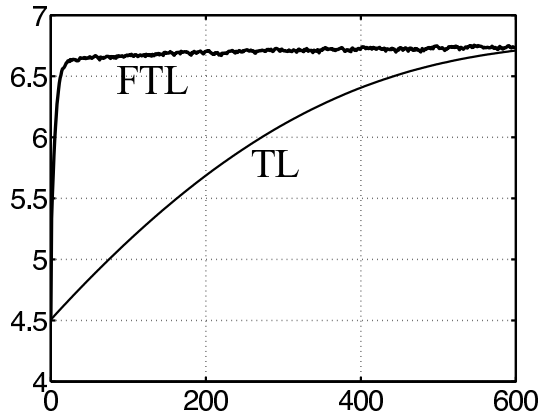

Fig. 2. SNR improvement in dB as a function of the iteration number, for the thresholded Landweber algorithm (TL) and its faster implementation (FTL).

wavelets, we use Shannon wavelets for regularization. Fig. 2 shows the SNR improvement curves during the execution of both algorithms. It is seen that a few tens of iterations of the fast thresholded Landweber (FTL) algorithm are almost equivalent to 600 iterations of the classical one (TL).

Table 1 presents more extensive results on how many iterations are needed for both algorithms to reach a given SNR level. The experiments were performed on the cameraman and MRI images, with the same blurring kernel, at various noise levels. The latter are expressed in terms of BSNR, that is, the signal to noise ratio with respect to the blurred version of the original signal. In each case, we have taken as reference the SNR improvements achieved after 10 and 30 iterations of the FTL algorithms. The results are promising and we are currently implementing the 3D version of the algorithm to test it on real micrographs.

Cameraman						
BSNR	SNRI	FTL	TL	SNRI	FTL	TL
20	2.36	10	98	2.39	30	198
30	2.78	10	23	2.85	30	83
40	4.77	10	113	4.83	30	134
50	6.75	10	505	7.29	30	818
MRI						
BSNR	SNRI	FTL	TL	SNRI	FTL	TL
20	3.64	10	58	3.88	30	200
30	5.79	10	44	5.90	30	55
40	8.64	10	290	8.71	30	304
50	9.58	10	851	11.81	30	> 1000

Table 1. Number of iterations required to reach a given level of SNR improvement (SNRI).

5. CONCLUSION

At high SNR levels, the accelerated thresholded Landweber algorithm is more than 10 times faster than the classical one. This makes a state-of-the-art regularization method more accessible for deconvolving large datasets, such as 3D time lapse images of biological samples. Moreover, the accelerated algorithm is simple to implement: compared to the existing one, it essentially consists in adapting the step size and thresholding parameter for each wavelet subband.

We have shown that the derivation of the bound optimization algorithm is relatively straightforward when the regularization term is expressed in a Shannon wavelet basis. Future work will be concerned with

1. the derivation of similar bounds for cost functions that are regularized in other wavelet bases;
2. the theoretical estimation of the achievable acceleration;
3. the 3D implementation and application to real data.

6. REFERENCES

[1] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 20–31, May 2006.

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

[3] J. L. Starck, D. L. Donoho, and E. J. Candès, "Astronomical image representation by the curvelet transform," *Astronomy and Astrophysics*, vol. 398, no. 2, pp. 785–800, February 2003.

[4] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, August 2003.

[5] 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*, vol. 57, no. 11, pp. 1413–1457, August 2004.

[6] 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*, vol. 3024, pp. 1–13, 2004.

[7] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, February 2004.

[8] 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*, vol. 7, no. 3, pp. 319–335, March 1998.

[9] G. Strang and T. Nguyen, *Wavelets and filter banks*, Wellesley-Cambridge Press, 1996.