A Bayesian wavelet-based multidimensional deconvolution with sub-band emphasis

Yingsong Zhang, Nick Kingsbury
Signal Processing & Communication Group, Dept. of Engineering, University of Cambridge
Cambridge CB2 1PZ, UK

***Abstract*— This paper proposes a new algorithm for wavelet-based multidimensional image deconvolution which employs subband-dependent minimization and the dual-tree complex wavelet transform in an iterative Bayesian framework. In addition, this algorithm employs a new prior instead of the popular ℓ_1 norm, and is thus able to embed a learning scheme during the iteration which helps it to achieve better deconvolution results and faster convergence.**

I. Introduction

In this paper we consider improved methods for image deconvolution with application to 3D datasets, such as fluorescence microscopy data.

Traditional frequently used approaches to deconvolution are based on the least squares (LS) formulation using the differences between the measured and deconvolved images as the cost function during optimization. It is well known that the LS formulation usually leads to an ill-posed problem that needs extra regularization. Under appropriate assumptions, the regularized deconvolution problem can be derived using a Bayesian framework and controlled by different prior assumptions.

Wavelet regularization is a relatively recent technique in the area of deconvolution. Figueiredo and Nowak in [3] define the basic idea of wavelet regularized deconvolution, which alternates between wavelet-coefficient shrinkage and a Landweber update. Daubechies et al. in [1] provided a general convergence proof of the algorithm. Recently, Vonesch & Unser explored subband-dependent minimization and proposed the fast thresholding Landweber (FTL) algorithm[6], with Shannon wavelets.

Based on the work of Daubechies et al.[1] and Vonesch & Unser[6], we formulate the wavelet-regularized subband-dependent minimization problem with a different prior instead of the traditional ℓ_1 norm. Hence, in our algorithm, we are able to update the prior during deconvolution which is shown to give better results in our experiments.

In addition, we choose to use the dual-tree complex wavelet transform[4] (DT $\mathbb{C}$WT) instead of the Shannon wavelet[6]. The DT $\mathbb{C}$WT is different from the Shannon wavelet because it is:

1) compactly supported in the spatial domain;
2) a redundant frame instead of an orthonormal basis;

We thank Thorlabs Ltd, UK, who provided the data which produced the images in fig. 3.
Thanks for CSC scholarship.

but equivalent to the Shannon wavelet in that it is:

1) a tight frame that conserves energy;
2) shift-invariant so that the wavelet transform commutes with the blurring operator.

Although redundancy requires more memory and computation, we still prefer the DT $\mathbb{C}$WT, since its compact-support property in the spatial domain offers the following advantages:

1) effects of coefficient shrinkage or other modification remain strongly localised;
2) it allows very efficient implementation with separable spatial domain filters;
3) it allows spatially varying algorithms for optimal processing of data with spatially varying statistics.

II. Image-formation model and Bayesian wavelet deconvolution

Normally, the image measurement process can be represented by a known stationary linear filter followed by the addition of white noise of zero mean and variance σ^2:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors containing the original image and the observed image, respectively; $\mathbf{n}$ represents independent white noise. These vectors have $N = N_1 \times N_2 \times \ldots \times N_K$ components, where N_k stands for the number of samples along dimension k. $\mathbf{H}$ is a square, block-circulant matrix that approximates the convolution(blurring function).

It is well known that we can get the MAP estimate of $\mathbf{x}$ by minimising a cost function

$$J(\mathbf{x}) = \frac{1}{2}\|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2 - \sigma^2 \log(p(\mathbf{x})) \tag{2}$$

where $\log(p(\mathbf{x}))$ corresponds to the prior expectations about image structure[2].

These expectations can most easily be formulated in the wavelet domain, since the wavelet transform of images tends to be sparse (i.e. many coefficients are close to zero)[3].

We choose to formulate this problem with the wavelet coefficients $\mathbf{w}$ of the dual-tree complex wavelet transform (DT $\mathbb{C}$WT) of $\mathbf{x}$ due to the transform's good signal energy compaction and shift invariant properties.

DT $\mathbb{C}$WT is an overcomplete wavelet transform, which is implemented in K dimentions by 2^K wavelet filterbanks in parallel. The coefficients produced by these filterbanks are

combined to give the real and imaginary parts of complex coefficients $\tilde{w}_i$.

By treating the real and imaginary parts of the wavelet coefficients separately, we use real matrices $\mathbf{W}$ and $\mathbf{M}$ to represent the forward and inverse DT $\mathbb{C}$WT, and a column vector $\mathbf{w}$ to denote the wavelet coefficients of the transform of the image $\mathbf{x}$, such that $\mathbf{w} = \mathbf{W}\mathbf{x}$; and $\mathbf{x} = \mathbf{M}\mathbf{w}$ is the image reconstructed from $\mathbf{w}$. We arrange the entries of the $\mathbf{w}$ such that $w_{2k-1} = Re(\tilde{w}_k),\ w_{2k} = Im(\tilde{w}_k)$.

In our prior model, we assume the wavelet coefficients are independently distributed according to Gaussian distributions with zero mean and spatially adaptable variance[2]; and that the real and imaginary parts of a given DT $\mathbb{C}$WT coefficient are independently drawn from the same Gaussian distribution. Hence, following Wang *et al*[7], the prior pdf $p(\omega)$ is proportional to $\exp\{-\frac{1}{2}\mathbf{w}^T\tilde{\mathbf{A}}\mathbf{w}\}$ where $\tilde{\mathbf{A}}$ is a diagonal matrix with $\tilde{A}_{ii}^{-1}$ being the expected variance of w_i[2].

For simplicity, let $A_{ii} = \sigma^2\tilde{A}_{ii} = \sigma^2/\mathrm{var}(w_i)$. Therefore the cost function in terms of wavelet coefficients can be written as:

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{H}\mathbf{M}\mathbf{w} - \mathbf{y}\|^2 + \frac{1}{2}\mathbf{w}^T\mathbf{A}\mathbf{w} \tag{3}$$

III. Algorithm

The above cost function(3) results in a correction direction

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \mathbf{M}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{M}\mathbf{w}) - \mathbf{A}\mathbf{w} \tag{4}$$

whose term, $\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{M}\mathbf{w})$, actually blurs the differences between the observed image $\mathbf{y}$ and the estimated observed image, $\mathbf{H}\mathbf{M}\mathbf{w}$. This in fact prevents the fine details of the image from being recovered since the PSF is often a low-pass filter. Therefore we need to emphasise the sub-bands which contain the fine details of the differences.

A. Derivation of a sub-band pre-emphasis bound

Following Vonesch & Unser [6], we denote by $(\mathbf{W}_j)_{j\in S}$ and $(\mathbf{M}_j)_{j\in S}$ the different decomposition and reconstruction wavelet subspaces (subbands) respectively, where $S = 0, 1, \ldots J$, J is the number of wavelet subspaces, 0 indexes the scaling function subspace. And specifically, regarding DT $\mathbb{C}$WT, $\mathbf{W}_j$ and $\mathbf{M}_j$ are masked versions of $\mathbf{W}$ and $\mathbf{M}$, respectively, such that $\mathbf{W} = \sum_{j\in S}\mathbf{W}_j$ and $\mathbf{M} = \sum_{j\in S}\mathbf{M}_j$. For simplicity, we denote $\mathbf{W}_j\mathbf{x}$ as column vector $\mathbf{w}_j$, with non-zero coefficients only in subband j.
Additionally, we assume that the real and imaginary parts of the transform's outputs are treated as separate coefficients so that $\mathbf{W}_j$ and $\mathbf{M}_j$ are real matrices. And thus for a real image $\mathbf{x}$, $\mathbf{w}_j = \mathbf{W}_j\mathbf{x}$ and $\mathbf{M}_j\mathbf{w}_j$ are real.
Since the DT $\mathbb{C}$WT is a tight frame[4], it yields:

$$\mathbf{x} = \sum_{j\in S}\mathbf{M}_j\mathbf{w}_j = \sum_{j\in S}\underbrace{\mathbf{M}_j\mathbf{W}_j}_{\mathbf{P}_j}\mathbf{x} \tag{5}$$

$$\|\mathbf{x}\|^2 = \|\mathbf{w}\|^2 = \sum_{j\in S}\|\mathbf{w}_j\|^2 \tag{6}$$

$$\mathbf{M}_j = \mathbf{W}_j^T \tag{7}$$

At the same time, we can obtain:

$$\mathbf{P}_j\mathbf{H} = \mathbf{H}\mathbf{P}_j \tag{8}$$

because the DT $\mathbb{C}$WT is shift invariant, and thus $\mathbf{P}_j = \mathbf{M}_j\mathbf{W}_j$ acts just like a spatially invariant filter, which enables it to commute with $\mathbf{H}$.

By defining

$$(\mathbf{A}_j)_{ii} = \begin{cases} A_{ii}, & \text{if } w_i \text{ is in subband } j \\ 0, & \text{otherwise} \end{cases}$$

We obtain

$$J(\mathbf{w}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\underbrace{\mathbf{M}\mathbf{w}}_{\mathbf{x}}\|^2 + \sum_{j\in S}\frac{1}{2}\mathbf{w}^T\mathbf{A}_j\mathbf{w} \tag{9}$$

$$= \frac{1}{2}\|\mathbf{y} - \mathbf{H}\underbrace{\sum_{j\in S}\mathbf{M}_j\mathbf{w}_j}_{\mathbf{x}}\|^2 + \frac{1}{2}\sum_{j\in S}\mathbf{w}_j^T\mathbf{A}_j\mathbf{w}_j \tag{10}$$

The above result contains a difficult term for minimization of $J(\mathbf{w})$:

$$\|\mathbf{H}\mathbf{x}\|^2 = (\sum_{j\in S}\mathbf{M}_j\mathbf{w}_j)^T\mathbf{H}^\mathbf{T}\mathbf{H}(\sum_{j\in S}\mathbf{M}_j\mathbf{w}_j),$$

which gives a coupled system of nonlinear equations of $\mathbf{w}_j$ [1]. In order to eliminate this term from (10), we apply Daubechies' idea of an upper-bound function[1], to give the following auxiliary function:

$$\bar{J}_n(\mathbf{w}) = J(\mathbf{w}) + \frac{1}{2}\sum_{j\in S}\alpha_j\|\mathbf{W}_j\mathbf{x}^{(n)} - \mathbf{W}_j\mathbf{x}\|^2 - \frac{1}{2}\|\mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x}\|^2 \tag{11}$$

where $\mathbf{x} = \sum_{j\in S}\mathbf{M}_j\mathbf{w}_j$, $\mathbf{x}^{(n)}$ is the nth iteration of $\mathbf{x}$ and the α_j must be properly chosen to make sure that $\bar{J}(\mathbf{w}) - J(\mathbf{w}) \geq 0$. If we let $\epsilon = \mathbf{x}^{(n)} - \mathbf{x}$, this inequality is equivalent to

$$\sum_{j\in S}\alpha_j\|\mathbf{W}_j\epsilon\|^2 \geq \|\mathbf{H}\epsilon\|^2 \tag{12}$$

and so

$$\sum_{j\in S}\alpha_j\epsilon^T\underbrace{\mathbf{W}_j^T\mathbf{W}_j}_{\mathbf{P}_j}\epsilon \geq \epsilon^T\mathbf{H}^T\mathbf{H}\epsilon \tag{13}$$

1) The value of $\boldsymbol{\alpha_j}$:
Since $\mathbf{P}_j$, and $\mathbf{H}$ are convolution matrices (shift invariant), we may let $\epsilon_{\underline{k}} = Ce^{j\underline{\omega}^T\underline{k}}$ and analyse this in the frequency domain. Let $\mathbf{P}_j(\underline{\omega})$ be the gain of $\mathbf{P}_j$ and $|\mathbf{H}(\underline{\omega})|^2$ be the gain of $\mathbf{H}^T\mathbf{H}$ at frequency $\underline{\omega}$.

Thus

$$\sum_{j\in S}\alpha_j\mathbf{P}_j(\underline{\omega}) \geq |\mathbf{H}(\underline{\omega})|^2 \qquad \forall\underline{\omega} \tag{14}$$

Since $\mathbf{P}_j(\underline{\omega}) \geq 0$ and $\sum_{j\in S}\mathbf{P}_j(\underline{\omega}) = 1$ for all $\underline{\omega}$ and $\mathbf{P}_j(\underline{\omega}) \approx 1$ at the center frequency of each subband where all the other subbands have approximately zero gain, we may find an approximate value for α_j by setting it equal to $|\mathbf{H}(\underline{\omega}_j)|^2/\mathbf{P}_j(\underline{\omega}_j)$, where $\underline{\omega}_j$ is the centre frequency of subband j. The above analysis assumes that $|\mathbf{H}(\underline{\omega})|^2$ is smooth enough to be approximated by $\sum_{j\in S}\alpha_j\mathbf{P}_j(\underline{\omega})$.
For rapid convergence [6] shows that each α_j should be as small as (13) allows.

2) The resulting algorithm:

Let $\mathbf{w}_j^{(n)} = \mathbf{W}_j \mathbf{x}^{(n)}$, and rewrite $2\bar{J}_n(\mathbf{w})$ as

$$\begin{aligned} 2\bar{J}_n(\mathbf{w}) &= \|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 + \sum_{j\in S}\alpha_j\|\mathbf{w}_j^{(n)}-\mathbf{w}_j\|^2 \\ &\quad - \|\mathbf{H}\mathbf{x}^{(n)}-\mathbf{H}\mathbf{x}\|^2 + \sum_{j\in S}\mathbf{w}_j^T\mathbf{A}_j\mathbf{w}_j \\ &= -2\mathbf{y}^T\mathbf{H}\mathbf{x} + \sum_{j\in S}\alpha_j\|\mathbf{w}_j^{(n)}-\mathbf{w}_j\|^2 + \\ &\quad 2(\mathbf{H}^T\mathbf{H}\mathbf{x}^{(n)})^T\mathbf{x} + \sum_{j\in S}\mathbf{w}_j^T\mathbf{A}_j\mathbf{w}_j + C(\mathbf{x}^{(n)},\mathbf{y}) \\ &= -2(\mathbf{H}^T\mathbf{y})^T\sum_{j\in S}\mathbf{M}_j\mathbf{w}_j + \sum_{j\in S}\alpha_j\|\mathbf{w}_j^{(n)}-\mathbf{w}_j\|^2 \\ &\quad + 2(\mathbf{H}^T\mathbf{H}\mathbf{x}^{(n)})^T\sum_{j\in S}\mathbf{M}_j\mathbf{w}_j \\ &\quad + \sum_{j\in S}\mathbf{w}_j^T\mathbf{A}_j\mathbf{w}_j + C(\mathbf{x}^{(n)},\mathbf{y}) \end{aligned} \tag{15}$$

where $C(\mathbf{x}^{(n)},\mathbf{y}) = \|\mathbf{y}\|^2 - \|\mathbf{H}\mathbf{x}^{(n)}\|^2$ is a function that does not depend on $\mathbf{w}_j$.

Equation (15) is clearly a quadratic equation in $\mathbf{w}_j$. Its global minimum, given $\mathbf{x}^{(n)}$, is achieved when $\partial\bar{J}_n(\mathbf{w})/\partial\mathbf{w}_j = 0$, which results in

$$(\alpha_j\mathbf{I}+\mathbf{A}_j)\mathbf{w}_j = \alpha_j\mathbf{w}_j^{(n)} + \mathbf{W}_j\mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{x}^{(n)}), \quad \forall j \tag{16}$$

Hence we get the new estimates of $\mathbf{w}_j$ and $\mathbf{x}$

$$\mathbf{w}_j^{(n+1)} = (\alpha_j\mathbf{I}+\mathbf{A}_j)^{-1}\left(\alpha_j\mathbf{w}_j^{(n)} + \mathbf{W}_j\mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{x}^{(n)})\right) \tag{17}$$

$$\mathbf{x}^{(n+1)} = \sum_{j\in S}\mathbf{M}_j\mathbf{w}_j^{(n+1)} \tag{18}$$

B. Updating the prior $\mathbf{A}$

The iteration rule of (17) and (18) can also be divided into two building blocks:

- the Landweber iteration:
$$\mathbf{z}_j^{(n)} = \alpha_j\mathbf{w}_j^{(n)} + \mathbf{W}_j\mathbf{H}^T(\mathbf{y}-\mathbf{H}\mathbf{x}^{(n)})$$
- the projection operation that actually projects the new coefficients back into the space dominated by the prior A:
$$\mathbf{w}_j^{(n)} = (\alpha_j\mathbf{I}+\mathbf{A}_j)^{-1}\mathbf{z}_j^{(n)}$$

We can expect a good estimate of $\mathbf{w}$ if we have an accurate prior, because the prior $\mathbf{A}$ represents the image structure, and also because the prior helps to suppress the noise and artifacts which are unlikely to be part of the image structure.

In general, we don't know the exact information about the prior. Thus a coarse estimate of $\mathbf{A}$ from the blurred image is often used instead. The wavelet variance estimates are given by the energy of the wavelet coefficients of an estimate of the original image; and the diagonal entries A_{ii} are the reciprocals of the estimated variances.

Since the estimated image is often contaminated by artifacts and noise, this straightforward approach often fails to obtain a satisfactory estimate of $\mathbf{A}$. A heuristic alternative approach is to update the estimate of $\mathbf{A}$ as the restored image structure gets closer to the true one.

An intuitive approach to overcome such difficulties is to denoise the coefficients before we calculate $\mathbf{A}$. Because $\mathbf{w}$ is sparse, we may therefore estimate $\hat{\mathbf{w}}$ by wavelet shrinkage of $\mathbf{w}$, and we use $\sigma^2/\hat{w}_i^2$ for each term A_{ii} in $\mathbf{A}$.

IV. Experimental Results

In our experiments, we choose to use the bivariate shrinkage denoising rule with the DT $\mathbb{C}$WT as described in [5] in our estimation of $\mathbf{A}$. This is because this shrinkage algorithm exploits the interscale dependencies between wavelet coefficients while being relatively simple.

For comparative purposes, we performed a series of experiments on 2 standard test images (Cameraman, House). We convolved these images with a 9×9 uniform blur kernel, and added white Gaussian noise to the results in order to replicate the experimental setup of Vonesch & Unser[6]. We use the blurred signal-to-noise ratio (BSNR) to define the noise level over N pels:

$$\mathrm{BSNR} = 10\log_{10}\frac{\|\mathbf{H}\mathbf{x}\|^2 - N(\overline{\mathbf{H}\mathbf{x}})^2}{N\sigma^2} \tag{19}$$

We considered 5 different noise levels: BSNR = 10dB, 20dB, 30dB, 40dB, 50dB. We also adopted improvement in signal-to-noise ratio (ISNR, same as SERG in [6]) to evaluate each estimate $\mathbf{x}^{(n)}$:

$$\mathrm{ISNR}(\mathbf{x}^{(n)}) = 10\log_{10}\left(\frac{\|\mathbf{y}-\mathbf{x}\|^2}{\|\mathbf{x}^{(n)}-\mathbf{x}\|^2}\right) \tag{20}$$

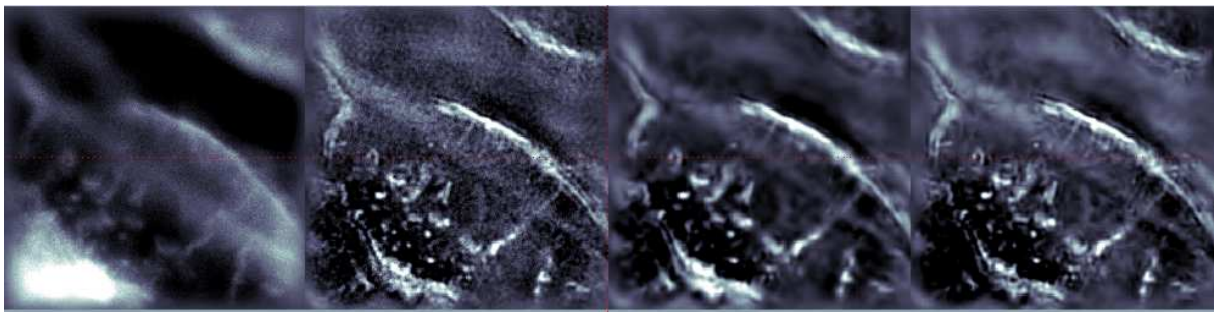
For each test case, we used the same initial estimate as [6], which was obtained using the under-regulized Wiener-type filter:

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

We averaged the ISNR results over 30 noise realizations, and summarize them in Table I. The numerical results of the FTL method are quoted from Vonesch & Unser[6].

When the image House was used and BSNR was set to 50dB, we found that the algorithm became less efficient because σ is so small that it has weakened the prior. Hence, we used $2\sigma^2$ instead of σ^2 on the same images and got the data marked by * in the table. We also amplified σ^2 a little to $1.1\sigma^2$ when BSNR was set to 40dB, and got the data marked by †. This correponds to the need for a small amount of regularization even in the absence of additive noise. We show blurred and restored images in Fig.1, and convergence curves in Fig.2.

To demonstrate the effect of updating matrix $\mathbf{A}$, we ran the algorithm while holding $\mathbf{A}$ constant from the begining, and compared the ISNR with those obtained when updating $\mathbf{A}$ after each iteration. It is clear that updating of the prior knowledge does help to improve the performance. We have also applied the algorithm successfully to 3D microscopy data as shown in Fig. 3.



(a) observed image (b) after intitial Wiener filter (c) after 10 iters (d) after 30 iters

Fig. 3. One slice of a 3D fluorescence microscopy data set of size $256 \times 256 \times 81$.

TABLE I

NUMBERS OF ITERATIONS TO REACH A GIVEN LEVEL OF ISNR

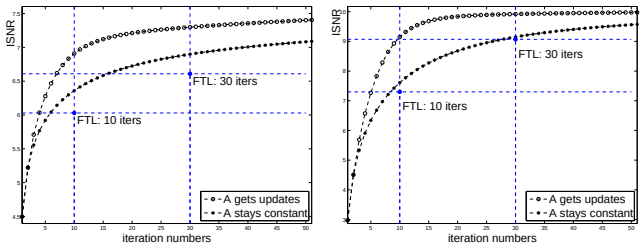
	BSNR	BWSE 2 iter.	BWSE 10 iter.	FTL 10 iter.	BWSE 30 iter.	FTL 30 iter.
Cameraman	10dB	2.44	2.82	2.90	2.98	2.94
	20dB	2.34	2.57	2.62	2.76	2.74
	30dB	4.31	4.57	4.31	4.79	4.43
	40dB	5.66	6.90	6.03	7.24	6.61
	50dB	6.62	7.86	7.80	9.23	8.38
House	10dB	3.54	4.34	3.73	4.75	3.82
	20dB	4.11	4.45	3.81	4.95	3.99
	30dB	5.70	7.18	6.25	7.59	6.39
	40dB	4.48 4.93†	7.13 8.14†	7.30	9.67 9.86†	9.07
	50dB	4.50 5.49*	6.56 8.08*	7.49	8.60 10.73*	8.90

Note: number marked by * or † is the results when A is amplified a little. BWSE is the acronym for our proposed algorithm Bayesian wavelet subband emphasis.

(a) blurred image (b) after initial Wiener filter

(c) after 10 iters (d) after 30 iters

Fig. 1. Experiment on Cameraman with white additive noise 40dB. The images were obtained by the proposed algorithm with an evolving $\mathbf{A}$.



(a) (b)

Fig. 2. ISNR as a function of the iteration number when applying the proposed algorithm. Blue-filled circles are the markers for the results of FTL algorithm. $-o-$ marks the ISNR curve when $\mathbf{A}$ gets updates while $-*-$ marks the ISNR curve when $\mathbf{A}$ stays constant. (a) the result of Cameraman with BSNR=40dB; (b) the result of House with BSNR = 40dB.

V. CONCLUSION

We have shown how the fast thresholding Landweber algorithm of [6] can be modified to use the redundant DT $\mathbb{C}$WT[4] instead of Shannon wavelets, and how the algorithm can be improved by updating the regularizing prior matrix $\mathbf{A}$ using the bivariate shrinkage rule[5]. The DT $\mathbb{C}$WT gives more compact support than the Shannon wavelet, and improved results over those reported in [6].

REFERENCES

[1] 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, 2004.

[2] P. de Rivaz and N. Kingsbury. Bayesian image deconvolution and denoising using complex wavelets. *Image Processing, 2001. Proceedings. 2001 International Conference on*, 2, 2001.

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

[4] I.W. Selesnick, R.G. Baraniuk, and N.G. Kingsbury. The dual-tree complex wavelet transform. *Signal Processing Magazine, IEEE*, 22(6):123–151, 2005.

[5] L. Sendur and I.W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 50(11):2744–2756, 2002.

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

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