

Improved Bounds for Subband-adaptive Iterative Shrinkage/Thresholding Algorithms

Yingsong Zhang, *Student Member, IEEE*, Nick Kingsbury, *Member, IEEE*,
Dept. Engineering, University of Cambridge

Abstract

This paper presents new methods for computing the step sizes of the subband-adaptive iterative shrinkage-thresholding algorithms proposed by Bayram & Selesnick [1] and Vonesch & Unser [12]. The method yields tighter wavelet-domain bounds of the system matrix, thus leading to improved convergence speeds. It is directly applicable to non-redundant wavelet bases and we also adapt it for the case of (redundant) frames. It turns out that the simplest and most intuitive setting for the step sizes that ignores subband aliasing is often satisfactory in practice. We show that our methods can be used to advantage with reweighted least squares penalty functions as well as L1 penalties. We emphasize that the algorithms presented here are suitable for performing inverse filtering on very large datasets, including 3D data, since inversions are applied only to diagonal matrices and fast transforms are used to achieve all matrix-vector products.

Index Terms

Deconvolution, Iterative algorithms, wavelets, multiresolution, sparsity.

I. INTRODUCTION

The inverse filtering (or deconvolution) problem aims to recover the true vector \mathbf{x} from the distorted and noisy observation vector \mathbf{y} ,

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where matrix \mathbf{H} represents the convolution and \mathbf{n} represents the noise. This problem is typically ill-posed and hence further *a priori* information is needed to provide regularization which reduces the uncertainty of the solution and prevents overfitting.

Wavelet-based methods have had a successful history in the field of image processing [1], [7], due to the fact that wavelets provide sparse representations for a wide range of images (i.e. many wavelet coefficients are close to zero). Applying wavelets to the deconvolution problem is never a trivial task, because convolution operators are generally quite difficult to represent in the wavelet domain, unlike the simple diagonalized representation in the Fourier domain [7]. Several research groups have independently proposed a forward-backward splitting procedure to circumvent this problem [4], [7]. Using the ℓ_1 norm as the regularization, their procedure alternates between a Landweber update and wavelet thresholding, and hence it is often called the Thresholded Landweber

(TL) algorithm [11] or the iterative shrinkage/thresholding (IST) algorithm [1]. The TL/IST algorithm offers computational advantages for the deconvolution of large images or 3D datasets, but its convergence speed is often slow.

The Fast Thresholded Landweber (FTL) [11] algorithm accelerates the convergence of the TL algorithm by iteratively updating the estimate in a subband adaptive fashion. The FTL was specific to the Shannon wavelet because it exploits the ideal spectral localization property of the Shannon wavelet to set the subband parameters to be as tight as possible. In [1], Bayram & Selesnick investigated the problem of generalizing to arbitrary wavelet frames by keeping to the all-subband-at-once structure of the original FTL, calling the resulting algorithm subband-adaptive IST (SISTA), and proposed a way to set the subband gain parameters. They also provided results suggesting that tighter subband parameters¹ tend to speed up convergence.

In this paper, we adopt a different approach from [1] to set the subband parameters (Section III). These parameters appear to be tighter than those proposed by Bayram & Selesnick's [1]. As a result, our modified algorithm exhibits faster convergence.

While this result works elegantly for orthonormal transforms, it cannot be generalized to redundant transforms directly, due to the difference between the range space of the redundant transform and the overall vector space. In Section III-B, we show that our proposed method can be applicable with redundant transforms which are formed from a union of orthonormal transforms. Unfortunately, with other types of redundant transform, we have to resort to a weaker argument, but we also derive a solution for this case.

We start, in Section II, by formulating the wavelet-based deconvolution problem with subband-separable regularizations, as in [13]. Though the regularization is not restricted to the ℓ_1 norm, the resulting iteration rule comprises the Landweber update and all-subband-at-once “denoising steps” as in SISTA. With the improved bounds, we therefore call our algorithm modified subband-adaptive iterative shrinkage/thresholding (MSIST) to differentiate it from the original SISTA.

Section IV provides evidence that supports our method of selecting subband parameters. The results of applying MSIST to several deconvolution problems are provided to demonstrate the speed-up effect of MSIST with different regularizations.

II. THE MSIST

We list the notation that we will use throughout the paper in Table I, and now introduce the basic algorithm. We consider a general minimization problem for the function:

$$\begin{aligned} F(\mathbf{w}) &= \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{M}\mathbf{w}\|^2 + \nu^2 \phi(\mathbf{w}) \\ &= \frac{1}{2} \|\mathbf{y} - \mathbf{H} \underbrace{\sum_{j \in S} \mathbf{M}_{(j)} \mathbf{w}_{(j)}}_{\mathbf{x}}\|^2 + \nu^2 \sum_{j \in S} \phi_j(\mathbf{w}_{(j)}) \end{aligned} \quad (2)$$

where $\phi_j(\mathbf{w}_{(j)})$ is the regularization function for subband j . The search for \mathbf{w} which minimizes (2) covers a broad range of wavelet based image/signal reconstruction and restoration (inverse filtering) problems. Many

¹We say the subband parameter is tight when the parameter is close to its lower limit for guaranteed convergence.

TABLE I
NOTATION FOR WAVELET ANALYSIS

Notation	Type	Representation
\mathbf{W}	matrix	forward wavelet transform.
\mathbf{w}	vector	wavelet coefficients of \mathbf{x} , $\mathbf{w} = \mathbf{W}\mathbf{x}$.
w_i	scalar	the i -th entry of \mathbf{w} .
j	integer	index of the subbands; $j = 0$ indexes the subband of the scaling function.
$\mathbf{W}_{(j)}$	matrix	forward wavelet transform in the given subband j , with all other subbands set to zero.
$\mathbf{w}_{(j)}$	vector	wavelet coefs in subband j ; $\mathbf{w}_{(j)} = \mathbf{W}_{(j)}\mathbf{x}$.
\mathbf{M}	matrix	inverse wavelet transform; $\mathbf{x} = \mathbf{M}\mathbf{w}$.
$\mathbf{M}_{(j)}$	matrix	inverse wavelet transform in the given subband j , with all other subbands set to zero.
$\mathbf{P}_{(j)}$	matrix	$\mathbf{P}_{(j)} = \mathbf{M}_{(j)}\mathbf{W}_{(j)}$, transfer fn. of subband j ; for perfect reconstruction, $\sum_{j=0}^J \mathbf{P}_{(j)} = \mathbf{I}$.

widely used sparse penalty functions are subband/subspace/group separable, such as the ℓ_p -norms ($0 \leq p \leq 1$), group lasso [14], weighted least squares, etc.

To be consistent with the prior work in [1], we follow the notation there. For a system with J subbands, we introduce the vector $\boldsymbol{\alpha} = [\alpha_0 \dots \alpha_J]$ and the diagonal operator $\Lambda_{\boldsymbol{\alpha}}$ that multiplies the j^{th} subspace/subband of \mathbf{w} by α_j , such that:

$$(\Lambda_{\boldsymbol{\alpha}}\mathbf{w})_{(j)} = \alpha_j \mathbf{w}_{(j)} \quad \text{for } j = 0 \dots J. \quad (3)$$

Let $\beta_j = \nu^2/\alpha_j$ and use subscript n to denote \mathbf{w} at the n th iteration, we are then able to state the MSIST algorithm as follows:

$$\left. \begin{aligned} \mathbf{b} &= \mathbf{w}_n + \Lambda_{\boldsymbol{\alpha}}^{-1} \mathbf{M}^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{x}) \\ (\mathbf{w}_{n+1})_{(j)} &= \text{prox}_{\beta_j \phi_j} \mathbf{b}_{(j)} \\ \mathbf{x} &= \mathbf{M}\mathbf{w}_{n+1}, \end{aligned} \right\} \quad (4)$$

where $\text{prox}_{\beta_j \phi_j}(\cdot)$ denotes the unique minimum

$$\arg \min_{\mathbf{w}_{(j)}} \left[\frac{1}{2} \|\mathbf{w}_{(j)} - \mathbf{b}_{(j)}\|^2 + \beta_j \phi_j(\mathbf{w}_{(j)}) \right]. \quad (5)$$

This is called the proximity operator in the literature [3] and is widely used [13]. The algorithm of (4) is the direct result of applying the majorization-minimization (MM) technique on (2). The MM technique was introduced for the linear inverse problem by Daubechies et al [4] and the general convergence property is also presented there and in [3]. The MM technique is very useful in dealing with the convolution operator in the deconvolution problem [6]. Instead of minimizing the target function $F(\mathbf{w})$ directly, the MM technique minimizes an easier surrogate function that upper bounds $F(\mathbf{w})$. In our case, the surrogate function is

$$G(\mathbf{w}, \mathbf{v}) = F(\mathbf{w}) + \frac{1}{2}(\mathbf{w} - \mathbf{v})^T \Lambda_{\boldsymbol{\alpha}}(\mathbf{w} - \mathbf{v}) - \frac{1}{2} \|\mathbf{H}\mathbf{M}(\mathbf{w} - \mathbf{v})\|_2^2. \quad (6)$$

Equation (4) is then obtained by alternately minimizing \mathbf{w} and \mathbf{v} (\mathbf{w} and \mathbf{v} correspond to \mathbf{w}_{n+1} and \mathbf{w}_n

respectively). For the MM algorithm to converge [3], [4], the surrogate function needs to satisfy that,

$$\begin{aligned} G(\mathbf{w}, \mathbf{v}) &> G(\mathbf{w}, \mathbf{w}), \text{ any } \mathbf{v} \neq \mathbf{w} \\ G(\mathbf{w}, \mathbf{w}) &= F(\mathbf{w}) \end{aligned} \quad (7)$$

This requirement is equivalent to the positive-definite condition:

$$\Lambda_\alpha - \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \succ 0 \quad (8)$$

where $\mathbf{A} \succ \mathbf{B}$ means that $\mathbf{u}^T \mathbf{A} \mathbf{u} > \mathbf{u}^T \mathbf{B} \mathbf{u}$ for any $\mathbf{u} \neq 0$. A more formal derivation is elaborated in [6] (replacing \mathbf{D} in [6] by Λ_α), and in [1] and [15] replacing the penalty function with a more general form.

In [3], Combettes & Wajs proved that (4) leads to a global minimizer of $F(\mathbf{w})$ if the Landweber step is non-expansive, and if $\phi(\cdot)$ is lower semi-continuous and convex. In [1] Bayram & Selesnick proved that the convergence rate of SISTA depends on the contraction rate of the Landweber update and the proximity operator. It is well known that the Landweber update converges faster if the spectral radius (largest eigenvalue) of $(\Lambda_\alpha - \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M})$ is smaller. The design of the subband adaptive algorithm aims to reduce the spectral radius, $\rho(\Lambda_\alpha - \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M})$, and hence speed up the convergence of the Landweber update.

It should be noted that the proximity operator does not always lead to a closed form solution. With the ℓ_1 norm and the zero-mean Gaussian log priors, the proximity operator is in closed form, but for a Generalized Gaussian distribution (GGD) and Gaussian scale mixture (GSM) log prior, for example, the solutions are not. Figueiredo et al in [6] discuss applying the MM techniques on such penalty functions. Even and subquadratic $\phi_j(\mathbf{w}_{(j)})$ can be majorized by even and quadratic functions. Alternate minimization then leads to an algorithm in the iterative reweighted least squares (IRLS) form. The IRLS algorithm in subband-adaptive form [15] will be given particular weights in the numerical result section, which demonstrates that IRLS can be used to minimize some difficult penalties, even non-convex ones. This is helpful when we want to consider a broader variety of penalties. The convergence analysis of IRLS is very difficult, but in practice, we have always observed convergence and high quality image restorations as long as the weights are properly initialized (all weights significantly non-zero).

III. α_j SELECTION

As stated above, Λ_α needs to be properly set such that (8) holds to ensure the convergence of the algorithm, and for maximum convergence speed, Λ_α should be an upper bound of $\mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}$ as tight as possible. This means that, for any vector $\mathbf{u} \neq 0$:

$$\mathbf{u}^T (\Lambda_\alpha - \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}) \mathbf{u} > 0. \quad (9)$$

Note that \mathbf{u} need not be within the range space of \mathbf{W} , if \mathbf{W} is a redundant frame. To make the argument in the following sections valid, we need to make the following clarification. We will use $\hat{\mathbf{u}}$ to denote $\mathbf{W}\mathbf{x}$, a wavelet-coefficient vector in the range space of \mathbf{W} , and \mathbf{u} to denote any vector that has the same dimension as $\hat{\mathbf{u}}$ and satisfies $\mathbf{x} = \mathbf{M}\mathbf{u}$. If \mathbf{W} is orthonormal, $\hat{\mathbf{u}} = \mathbf{u}$, otherwise in general $\hat{\mathbf{u}} \neq \mathbf{u}$. To keep the derivation simple, we first assume that \mathbf{W} is an orthonormal basis.

A. The orthonormal wavelet case

For orthonormal wavelets, $\hat{\mathbf{u}} = \mathbf{u}$ holds. From (9) the following expression must be positive for all non-zero \mathbf{u} :

$$\begin{aligned} & \mathbf{u}^T \Lambda_\alpha \mathbf{u} - \mathbf{u}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \mathbf{u} \\ &= \sum_{j=0}^J \alpha_j \mathbf{u}_{(j)}^T \mathbf{u}_{(j)} - \sum_{j=0}^J \mathbf{u}_{(j)}^T \mathbf{M}_{(j)}^T \mathbf{H}^T \mathbf{H} \mathbf{M}_{(j)} \mathbf{u}_{(j)} \\ & \quad - \sum_{j=0}^J \sum_{l \neq j} \mathbf{u}_{(j)}^T \mathbf{M}_{(j)}^T \mathbf{H}^T \mathbf{H} \mathbf{M}_{(l)} \mathbf{u}_{(l)} \end{aligned} \quad (10)$$

where we call $\mathbf{u}_{(j)}^T \mathbf{M}_{(j)}^T \mathbf{H}^T \mathbf{H} \mathbf{M}_{(j)} \mathbf{u}_{(j)}$ an inband component and $\mathbf{u}_{(j)}^T \mathbf{M}_{(j)}^T \mathbf{H}^T \mathbf{H} \mathbf{M}_{(l)} \mathbf{u}_{(l)}$ ($l \neq j$) a crossband component. Note that the crossband components represent the transmission of a signal through subband l of the inverse wavelet transform, blurring in the spatial domain by $\mathbf{H}^T \mathbf{H}$, and then transmitting through subband j ($j \neq l$) of the forward wavelet transform. Hence as long as the subbands are relatively non-overlapping in the frequency domain, the crossband summation term is likely to be significantly smaller than the inband summation term in (10), especially when bands j and l are non-adjacent.

Ignoring the crossband components, α_j can simply be chosen to be larger than $\rho(\mathbf{M}_{(j)}^T \mathbf{H}^T \mathbf{H} \mathbf{M}_{(j)})$ to ensure (10) is positive for all non-zero \mathbf{u} . Though the crossband summation complicates the situation, a potentially beneficial question to ask is “can the crossband summation be decomposed into only inband components?”. We achieve this as follows.

First we define

$$\Theta_0 = \mathbf{H}^T \mathbf{H} \quad (11)$$

and $\mathbf{P}_{(j)}$ as in Table I. We can then state the following theorem.

Theorem 1. Assume $\sum_{j=0}^J \mathbf{P}_{(j)} = \mathbf{I}$ with J being a positive finite number. For a given Hermitian matrix Θ , we introduce a positivity operator $P_+(\Theta)$ that sets every negative eigenvalue of Θ to 0. The matrix sequence $\{\Theta_k\}$ then defined by

$$\Theta_{k+1} = P_+(\Theta_k - \sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_k \mathbf{P}_{(j)}), \quad (12)$$

has the property $\lim_{k \rightarrow \infty} \Theta_k = 0$ and $\sum_k \mathbf{u}^T \Theta_k \mathbf{u}$ converges absolutely for any \mathbf{u} . Moreover,

$$\Theta_0 \preceq \sum_{j=0}^J \mathbf{P}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{P}_{(j)}, \quad (13)$$

where $\mathbf{A} \preceq \mathbf{B}$ means that $\mathbf{u}^T \mathbf{A} \mathbf{u} \leq \mathbf{u}^T \mathbf{B} \mathbf{u}$. (See Appendix A for the proof.)

Corollary 1. For orthonormal wavelets, $\mathbf{M}^T \Theta_0 \mathbf{M} \preceq \sum_{j=0}^J \mathbf{M}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{M}_{(j)}$. (See Appendix B for the proof.)

Because of Corollary 1, we can set

$$\alpha_j = \rho(\mathbf{M}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{M}_{(j)}) + \sigma \quad (14)$$

so as to ensure $\Lambda_\alpha \succ \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}$ (so that (10) > 0), with σ being a small constant.

B. Extension to redundant frames

First recall that we have defined $\hat{\mathbf{u}} = \mathbf{W}\mathbf{x}$ to be a wavelet coefficient vector in the range space of \mathbf{W} , and \mathbf{u} to be any vector that has the same dimension as \mathbf{u} and satisfies $\mathbf{x} = \mathbf{M}\mathbf{u}$. We then have

$$\mathbf{u}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \mathbf{u} = \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x} = \hat{\mathbf{u}}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \hat{\mathbf{u}}. \quad (15)$$

Therefore,

$$\begin{aligned} & \mathbf{u}^T \Lambda_\alpha \mathbf{u} - \mathbf{u}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \mathbf{u} \\ &= \underbrace{\mathbf{u}^T \Lambda_\alpha \mathbf{u} - \hat{\mathbf{u}}^T \Lambda_\alpha \hat{\mathbf{u}}}_{p_1} + \underbrace{\hat{\mathbf{u}}^T \Lambda_\alpha \hat{\mathbf{u}} - \hat{\mathbf{u}}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \hat{\mathbf{u}}}_{p_2} \end{aligned} \quad (16)$$

where p_2 is always non-negative if α_j is set according to (14). However, p_1 cannot always be nonnegative without any assumptions. Since the null space of a frame is orthogonal to its range space and $\mathbf{u} = \hat{\mathbf{u}} + \mathbf{u}_\perp$, $\hat{\mathbf{u}}$ will be the coefficient vector with the minimal ℓ_2 norm that satisfies $\mathbf{x} = \mathbf{M}\mathbf{u}$. Therefore, for any \mathbf{u} satisfying $\mathbf{x} = \mathbf{M}\mathbf{u}$, $\mathbf{u}^T \mathbf{u} \geq \hat{\mathbf{u}}^T \hat{\mathbf{u}}$ and hence $p_1 \geq 0$ if $\Lambda_\alpha = \alpha \mathbf{I}$. However when α_j is set differently for each subband, $p_1 \geq 0$ no longer necessarily holds. This means that directly applying (14) will not guarantee the positivity of (16); but we are able to provide some useful results for tight-frame transforms formed from a number of orthonormal transforms in parallel, as follows.

Let \mathbf{M}_l ($l = 1, \dots, L$) denote L parallel orthonormal transforms, where $\mathbf{M}_l \mathbf{M}_l^T = \mathbf{I}, \forall l$. The forward transforms are $\mathbf{W}_l = \mathbf{M}_l^T$. Hence

$$\begin{aligned} \hat{\mathbf{u}} &= \frac{1}{\sqrt{L}} [\mathbf{M}_1 \dots \mathbf{M}_L]^T \mathbf{x} \\ \mathbf{x} &= \frac{1}{\sqrt{L}} [\mathbf{M}_1 \dots \mathbf{M}_L] \mathbf{u} \end{aligned} \quad (17)$$

Let $\hat{\mathbf{u}}_l = \frac{1}{\sqrt{L}} \mathbf{M}_l^T \mathbf{x}$, so that $\hat{\mathbf{u}} = [\hat{\mathbf{u}}_1^T \dots \hat{\mathbf{u}}_L^T]^T$; and let \mathbf{u}_l be the equivalent components of \mathbf{u} . Because of Jensen's inequality

$$\begin{aligned} \mathbf{u}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \mathbf{u} &= \frac{1}{L} \left\| \sum_{l=1}^L \mathbf{H} \mathbf{M}_l \mathbf{u}_l \right\|_2^2 \\ &\leq \sum_{l=1}^L \|\mathbf{H} \mathbf{M}_l \mathbf{u}_l\|_2^2 \end{aligned} \quad (18)$$

We then have

$$\begin{aligned} & \mathbf{u}^T \Lambda_\alpha \mathbf{u} - \mathbf{u}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \mathbf{u} \\ & \geq \sum_{l=1}^L (\mathbf{u}_l^T \Lambda_l \mathbf{u}_l - \|\mathbf{H} \mathbf{M}_l \mathbf{u}_l\|_2^2) \end{aligned} \quad (19)$$

where Λ_l is the submatrix of Λ_α that corresponds to \mathbf{u}_l in \mathbf{u} . Because all of the \mathbf{u}_l are applied to orthonormal transforms, we can apply (14) on every orthonormal transform to ensure the positivity of $(\mathbf{u}_l^T \Lambda_l \mathbf{u}_l - \|\mathbf{H} \mathbf{M}_l \mathbf{u}_l\|_2^2)$ and thus ensure the positivity of $\mathbf{u}^T \Lambda_\alpha \mathbf{u} - \mathbf{u}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \mathbf{u}$, so that convergence of MSIST will be guaranteed.

For other types of tight frames, we have to consider a weaker argument for convergence. Instead of requiring $\Lambda_\alpha - \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}$ to be positive definite, we require

$$\mathbf{u}^T \Lambda_\alpha \mathbf{u} - \mathbf{u}^T \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M} \mathbf{u}^T > 0, \quad (20)$$

where $\mathbf{u} = \mathbf{w}_{n+1} - \mathbf{v}$, such that the cost function is monotonically reduced, $J(\mathbf{w}_{n+1}) \leq J(\mathbf{w}_n)$. With a properly chosen α that ensures the positivity of p_2 in (16), this is equivalent to

$$\mathbf{u}^T \Lambda_\alpha \mathbf{u} - \hat{\mathbf{u}}^T \Lambda_\alpha \hat{\mathbf{u}} \geq 0 \quad (21)$$

because of (16), where $\hat{\mathbf{u}} = \mathbf{W}\mathbf{M}\mathbf{u}$. Therefore, when applying MSIST with a redundant frame, we require an extra step to test whether (21) holds at each iteration. If it does not, we need to increase the α_j for that iteration, one subband at a time, until (21) is satisfied.

IV. NUMERICAL RESULTS

We note here that the normalized eigenvectors of shift-invariant transforms and blurring filters are the Fourier basis vectors. For shift-invariant systems (e.g. the undecimated wavelet transform), $\mathbf{P}_{(j)}^T \mathbf{H}^T \mathbf{H} \mathbf{P}_{(j)}$ is circulant and its DFT coefficients are real numbers and are also the eigenvalues of $\mathbf{P}_{(j)}^T \mathbf{H}^T \mathbf{H} \mathbf{P}_{(j)}$. For shift-variant systems, $\Theta_k (k \geq 1)$ and P_+ need to be evaluated explicitly, and the computation could be expensive for large $\mathbf{H}^T \mathbf{H}$. For ease of computation, we adopt some practical approximations.

A. Practical approximations

Because $\sum_{k=0}^{\infty} \Theta_k$ converges absolutely, a few (say $K + 1$) terms can approximate the right-hand term in (13) as follows:

$$\sum_{j=0}^J \mathbf{P}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{P}_{(j)} \approx \sum_{j=0}^J \mathbf{P}_{(j)}^T \left(\sum_{k=0}^K \Theta_k \right) \mathbf{P}_{(j)}. \quad (22)$$

The number of terms needed for a satisfactory approximation depends on how fast Θ_k converges. For shift-invariant systems, the computation is cheap and a large K can be used. For a shift-variant system with well-designed wavelets, the crossband summation in (10) does not play a significant role, so we expect Θ_k to converge quite quickly. Table II shows α_j of the critically sampled wavelet transform in the 1-dimensional case, calculated for the length-30 moving average filter by using different K for the approximation in (14). This filter is chosen as in [1] to provide directly comparable results. The limiting values show that our method clearly reaches a tighter bound than Bayram & Selesnick's [1], but the gap between the two methods shrinks when the wavelets have better spectral localization. With the Shannon wavelet (ideal localization), the two methods will produce identical results.

We also see that the approximated α_j with $K = 9$ for the db8 and db4 wavelet transforms are already very close to the limiting values. For db4 and db8 wavelets, it converges faster due to the much better selectivity of the wavelet subbands.

Combettes & Wajs's result [3], and, more directly, Bayram & Selesnick's result [1] assure convergence with a relaxed condition $2\Lambda_\alpha \succ \mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}$. This suggests that we can afford more losses in $\sum \Theta_k$ by using an even smaller K (typically only 0 or 1) and still assure convergence. In Table II, we note that α_j with $K = 0$ is larger than half of the limiting α_j for db8. Therefore, we suggest the following approximation for ease of computation, especially for the shift-variant wavelet transform with good frequency selectivity:

$$\alpha_j = \rho(\mathbf{M}_{(j)}^T \Theta_0 \mathbf{M}_{(j)}). \quad (23)$$

[1] and [12] propose fast algorithms for calculating $\rho(\mathbf{M}_{(j)}^T \Theta_0 \mathbf{M}_{(j)})$.

B. Applications to deconvolution

In this section, we show by examples that the subband-adaptive update rule of (4) is significantly more efficient than the standard TL algorithm. Note that Bayram & Selesnick have also demonstrated similar conclusions on 1-D and 2-D deconvolution problems in [1].

TABLE II

α FOR THE NORMALIZED LENGTH-30 MOVING AVERAGE FILTER WITH 6-LEVELS WITH DIFFERENT WAVELETS AND K . $\sigma = 1e - 5$.

	K	Level 1	2	3	4	5	6	0
db1	[1]	0.0779	0.1532	0.3023	0.5851	1.1558	1.3763	1.6780
	∞	0.0634	0.1112	0.1951	0.3385	0.5926	1.1175	1.0964
	9	0.0431	0.0902	0.1715	0.3059	0.5740	1.0791	1.0600
	1	0.0047	0.0139	0.0531	0.1645	0.4522	0.8383	1.0000
	0	0.0022	0.0044	0.0210	0.0915	0.3757	0.6878	1.0000
db4	[1]	0.0036	0.0121	0.0473	0.1646	0.7155	1.0694	1.1467
	∞	0.0031	0.0101	0.0406	0.1576	0.6192	0.9155	1.0003
	9	0.0029	0.0100	0.0340	0.1491	0.6106	0.9119	1.0000
	1	0.0023	0.0064	0.0238	0.0720	0.4982	0.8594	1.0000
	0	0.0022	0.0042	0.0185	0.0489	0.4524	0.8240	1.0000
db8	[1]	0.0028	0.0098	0.0342	0.0948	0.5950	0.9753	1.0558
	∞	0.0025	0.0093	0.0321	0.0754	0.5447	0.8544	1.0000
	9	0.0025	0.0092	0.0320	0.0719	0.5424	0.8524	1.0000
	1	0.0022	0.0060	0.0209	0.0518	0.4713	0.8361	1.0000
	0	0.0022	0.0048	0.0173	0.0456	0.4568	0.8317	1.0000

- [1] indicates that those subband parameters are calculated according to Bayram & Selesnick's paper on undecimated wavelet transform.
- The limiting value, denoted by ∞ , is obtained by running the algorithm until $\rho(\Theta_{k+1}) < 1e - 5$.

First we implemented a simple 1-D example with a 6-level critically sampled db4 transform. This example is taken from [1]². The signal is blurred by a length-30 moving average filter and with added Gaussian noise of variance 0.02. Hence in Figure 1 we compare the convergence speed of our α_j (dash-dot lines) with those calculated by Bayram & Selesnick (dash lines) [1] and basic TL. MSIST and SISTA are noticeably faster to converge than IST, but the differences between MSIST and SISTA are moderate because α computed on db4 with our proposed method and Bayram & Selesnick's method are very close as shown in Table II. The approximated α further speeds up the convergence, but again the differences among the compared approximated α are small, therefore we can opt for the simplest solution, i.e. our method with $K = 0$.

Secondly, we tested algorithms on the image deconvolution problem. We chose to use the DT CWT [9] as the analysis tool for two main reasons: (a) it has good frequency selectivity so we can expect Θ_k to converge quickly and hence we need only use $K = 0$ to calculate the α_j for each tree; and (b) it is an over-complete tight-frame wavelet transform consisting of 4 parallel trees, so we can test the theory in Section III-B in this example. In addition, the DT CWT is almost shift invariant, which reduces many of the artefacts of the critically sampled and much more shift-dependent DWT, and hence significantly enhances the wavelet-based processing [9].

For comparative purposes, we have performed a series of experiments on the standard test image, Cameraman. We convolved the image with a 9×9 uniform blur kernel. α computed by our method is shown numerically

²The authors would like to acknowledge Dr Bayram for generously allowing us access to his code.

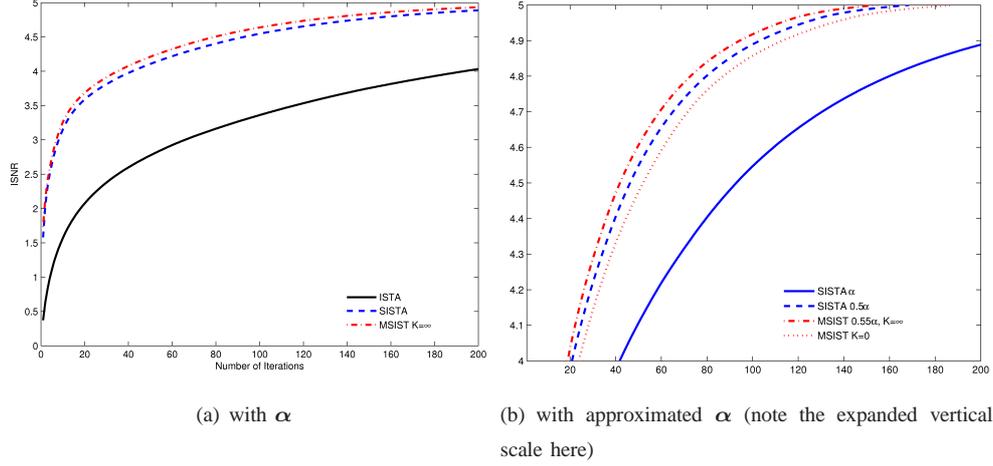


Fig. 1. Convergence speed of 1D signal with the blur kernel being a length-30 moving average filter. $\text{ISNR}(\mathbf{z}_n) = 10 \log_{10} \frac{\|\mathbf{z}_0 - \mathbf{x}\|_2^2}{\|\mathbf{z}_n - \mathbf{x}\|_2^2}$.

TABLE III

α USING THE $K = 0$ APPROXIMATION FOR THE 9×9 UNIFORM BLUR KERNEL AND DT CWT. THERE ARE 6 SUBBANDS FOR EACH LEVEL'S DT CWT DECOMPOSITION. 'LL' STANDS FOR THE SCALING SUBBAND AT LEVEL 4. $\sigma = 1e - 4$.

LL: $\alpha_0 = 1$						
Other subbands:						
	Subband 1	2	3	4	5	6
Level 1	0.0079	0.0002	0.0079	0.0079	0.0002	0.0079
Level 2	0.0265	0.0023	0.0265	0.0265	0.0023	0.0265
Level 3	0.1084	0.0288	0.1084	0.1084	0.0288	0.1084
Level 4	0.4997	0.3642	0.4997	0.4997	0.3642	0.4997

in Table III. We also compared against SISTA with $\frac{1}{2}\alpha$ computed according to [1]. We added white Gaussian noise to the blurred images and used the blurred signal-to-noise ratio (BSNR) to define the noise level over N pels:

$$\text{BSNR} = 10 \log_{10} \frac{\|\mathbf{H}\mathbf{x}_r - \overline{\mathbf{H}\mathbf{x}_r}\|_2^2}{N\nu^2} \quad (24)$$

where \mathbf{x}_r is the original reference image, $\overline{\mathbf{H}\mathbf{x}_r}$ denotes the mean of $\mathbf{H}\mathbf{x}_r$ and N is the pixel number. We adopted the improvement in signal-to-noise ratio (ISNR, equivalent to SERG in [11]) to evaluate each estimate \mathbf{z} of \mathbf{x}_r :

$$\text{ISNR}(\mathbf{z}) = 10 \log_{10} \left(\frac{\|\mathbf{y} - \mathbf{x}_r\|_2^2}{\|\mathbf{z} - \mathbf{x}_r\|_2^2} \right). \quad (25)$$

For each test case, we used the same initial estimate as in [11], which was obtained using the under-regularized Wiener-type filter:

$$\mathbf{z}_0 = (\mathbf{H}^T \mathbf{H} + 10^{-3} \nu^2 \mathbf{I})^{-1} \mathbf{H}^T \mathbf{y}. \quad (26)$$

Figure 2 compares the ISNR of MSIST to IST ($\alpha_j = \rho(\mathbf{H}^T \mathbf{H})$) and SISTA (with α_j computed as in [1]) with different penalty functions. The BSNR of the observations \mathbf{y} is 40dB. In each graph, the subband adaptive α_j of MSIST is as shown in Table III. Figure 2 (a) plots the results of the ℓ_1 -norm regularized algorithm, and

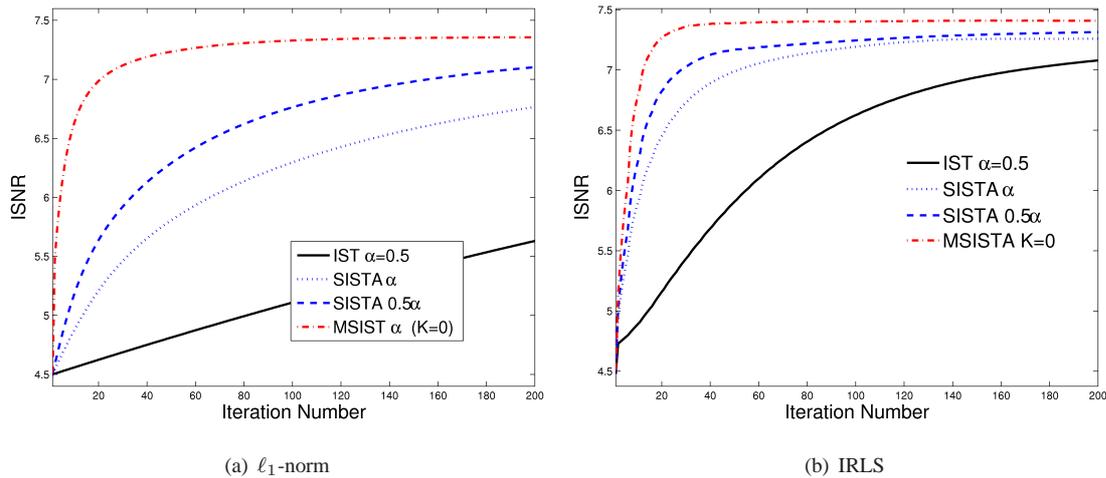


Fig. 2. ISNR versus iteration number, on Cameraman, BSNR = 40dB.

Figure 2 (b) plots the results of the iterative reweighted least-squares (IRLS)³ regularized algorithm.

We then considered another 2 different noise levels, BSNR = 20dB, 50dB; and averaged the ISNR results over 30 noise realizations. Results are summarized in Table IV, which shows that MSIST requires significantly fewer iterations to achieve a given quality of recovery under both ℓ_1 -norm and IRLS regularization when the noise is lower (50db).

We believe that the fast convergence of MSIST, shown in both Figure 1 and Figure 2, is directly dependent on how well the diagonal approximation to the blurring function, produced by Λ_α , approximates the true blurring function $\mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}$ in the wavelet domain. This in turn is related to the decorrelating properties of the chosen wavelet transform when applied to typical blurring operators \mathbf{H} . (Full decorrelation would result in a perfect diagonal representation being possible.) Hence proper choice of a good transform, when combined with expected forms of blurring, is an important factor in achieving good performance with this algorithm.

In the Cameraman example, another important observations is that IRLS reached a better ISNR than methods based on the ℓ_1 norm. This is presumably because IRLS minimizes a penalty that is closer to the ℓ_0 norm than the ℓ_1 norm does. It would also be possible to use MSIST-based IRLS to minimize ℓ_p ($0 < p < 1$) norms that could generate high-quality restorations. In our experiments, we found that having a whitening parameter ϵ in the in the IRLS weights, which slowly decreases from a relatively large value to a small value as the iterations proceed, is important for the IRLS to reach good solutions [2]. It has been observed by other authors that initializing each of the weights far from zero helps IRLS reach good results [6]. This explains why we need to set whitening parameter ϵ relatively large in the beginning, but there are currently no clear guidelines on how best to decrease it. Further work is needed here, but it is beyond the scope of this paper.

We have also applied the subband-adaptive IRLS algorithm successfully to a 3D microscopy dataset ($\sim 5\text{M}$ voxels) as shown in Figure 4. α_j in this example was again computed by (14) with $K = 0$. A similar result using an earlier version of MSIST was previously shown in [15]. This demonstrates that MSIST is suitable for

³The weights are set as $1/(|w_i|^2 + \epsilon)$ with w_i from the previous iteration. The corresponding penalty is element-wise $\log \frac{|w|^2 + \epsilon^2}{\epsilon}$, which is the log of the Cauchy Lorentz distribution. The Cauchy Lorentz distribution is very heavy-tailed and hence introduces sparsity.

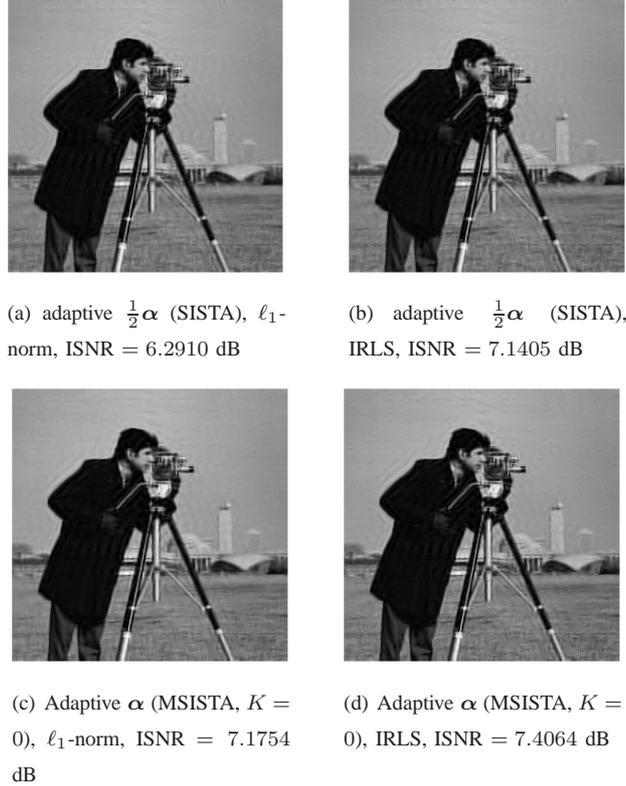


Fig. 3. The deconvolution results after 50 iterations within the range space, on Cameraman, BSNR = 40dB.

use on large datasets.

V. CONCLUSIONS

In this paper, we have considered ways to improve the estimation of subband dependent parameters α_j to further speed up SISTA. The proposed MSIST technique can be used straightforwardly on deconvolution problems with subband separable penalties and it can be expressed in a consistent form incorporating the Landweber update and denoising steps, with different forms of regularization.

Unlike Bayram & Selesnick's approach for calculating α_j , our method of computing the subband dependant parameters α_j is based on the geometric expansion of $\mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}$ on the orthonormal basis, and we obtain a Λ_α which appears to be tighter than than the one of [1]. We discuss this further in the appendix. By utilizing the result of Combettes & Wajs [3], we show that the simplest estimation that ignores the crossband components is sufficient to ensure convergence for typical wavelet bases, and that the convergence speed is good. More importantly, we consider the MSIST family of algorithms with *redundant* transforms and provide some useful results on setting the parameters in this case. It appears in our 1-D example that the convergence-speed improvement of MSIST with respect to the existing SISTA [1] is only small, but the improvement in the 2-D case is more significant. This is because our approximations are then tighter than those of SISTA and result in a smaller error between Λ_α and $\mathbf{M}^T \mathbf{H}^T \mathbf{H} \mathbf{M}$.

TABLE IV

ISNR RESULTS OVER 30 RANDOM REALIZATIONS OF NOISE. 'I' STANDS FOR THE NON-ADAPTIVE ALGORITHM IST; 'S' STANDS FOR SISTA WITH $\frac{1}{2}\alpha$ PROPOSED IN [1], 'M' STANDS FOR OUR PROPOSED ALGORITHM MSIST.

ℓ_1 norm						
BSNR	20 dB			50 dB		
Methods	I	S	M	I	S	M
10 iters	2.3183	2.5045	2.8604	6.2148	6.3137	7.8205
30 iters	2.4948	2.7633	2.9945	6.2798	6.5609	8.3353
50 iters	2.6113	2.8875	3.0131	6.3444	6.7872	8.6479
70 iters	2.6946	2.9622	3.0173	6.4088	6.9927	8.8562
100 iters	2.7829	3.0315	3.0217	6.5048	7.2638	9.0559

IRLS						
BSNR	20 dB			50 dB		
Methods	I	S	M	I	S	M
10 iters	2.3994	2.5749	2.5836	6.2691	6.8777	8.7598
30 iters	2.7973	2.9830	2.9897	6.3792	7.9687	10.2900
50 iters	2.9959	3.1867	3.1911	6.5393	8.8984	10.6011
70 iters	3.1129	3.3044	3.3077	6.7153	9.4842	10.6691
100 iters	3.2112	3.3997	3.4029	6.9744	9.9490	10.6827

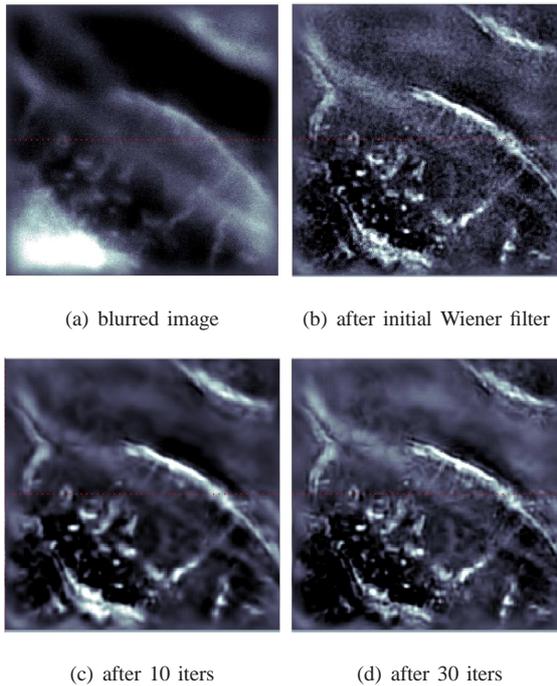


Fig. 4. One slice of a 3D fluorescence microscopy data set of size $256 \times 256 \times 81$ voxels.

APPENDIX A

PROOF FOR THEOREM 1

To prove theorem 1, we need the following result.

Let $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ denote the eigenvalues of the positive semi-definite (PSD) Hermitian matrix Θ , and

$\mathbf{u}_1, \dots, \mathbf{u}_n$ denote the corresponding eigenvectors. Therefore, for any vector \mathbf{x} :

$$\begin{aligned} \mathbf{x}^T \Theta \mathbf{x} &= \mathbf{x}^T \sum_{i=1}^n \lambda_i (\mathbf{u}_i^T \mathbf{x}) \mathbf{u}_i = \sum_{i=1}^n \lambda_i (\mathbf{u}_i^T \mathbf{x})^2 \\ &\geq \lambda_1 (\mathbf{u}_1^T \mathbf{x})^2 \end{aligned} \quad (27)$$

And, because $\sum_{j=0}^J \mathbf{P}_{(j)} = \mathbf{I}$,

$$\sum_{j=0}^J \mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{u}_i = \mathbf{u}_i^T \mathbf{u}_i = 1. \quad (28)$$

Using $\frac{1}{n} (\sum_{i=1}^n |x_i|^2) \geq (\frac{1}{n} \sum_{i=1}^n |x_i|)^2$, for any set \mathbf{x} , then gives

$$\begin{aligned} \sum_{j=0}^J \|\mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{u}_i\|^2 &\geq (\sum_{j=0}^J \mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{u}_i)^2 / (J+1) \\ &= 1 / (J+1). \end{aligned} \quad (29)$$

Proof:

1) Because Θ_0 is PSD and $P_+(\cdot)$ makes any other Θ_k PSD also, every $\mathbf{P}_{(j)}^T \Theta_k \mathbf{P}_{(j)}$ is PSD, and hence $\sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_{k-1} \mathbf{P}_{(j)}$ is also PSD.

Now we prove that $\sum_{k=0}^K \rho(\Theta_k)$ converges absolutely. This leads to the convergence of $\sum_{k=0}^K \mathbf{x}^T \Theta_k \mathbf{x}$ for any \mathbf{x} because $0 < \mathbf{x}^T \Theta_k \mathbf{x} \leq \rho(\Theta_k) \|\mathbf{x}\|_2^2$. If \mathbf{v}_k be the eigenvector corresponding to the largest eigenvalue of Θ_k , then:

$$\rho(\Theta_k) = \mathbf{v}_k^T \Theta_k \mathbf{v}_k = \mathbf{v}_k^T (\Theta_{k-1} - \sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_{k-1} \mathbf{P}_{(j)}) \mathbf{v}_k \quad (30)$$

Let $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ denote the eigenvalue of PSD Hermitian matrix Θ_{k-1} , and $\mathbf{u}_1, \dots, \mathbf{u}_n$ denote the corresponding eigenvectors. Let $\beta_i = \mathbf{u}_i^T \mathbf{v}_k$ and hence $\mathbf{v}_k = \sum_{i=1}^n \beta_i \mathbf{u}_i$. Therefore, we have

$$\mathbf{v}_k^T \Theta_{k-1} \mathbf{v}_k = \sum_{i=1}^n \lambda_i \beta_i^2 \quad (31)$$

and

$$\begin{aligned} \mathbf{v}_k^T \sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_{k-1} \mathbf{P}_{(j)} \mathbf{v}_k &= \sum_{j=0}^J \sum_{i=1}^n \lambda_i (\mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{v}_k)^2 \\ &= \sum_{i=1}^n \lambda_i \sum_{j=0}^J \|\mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{v}_k\|_2^2 \\ &\geq \sum_{i=1}^n \lambda_i \frac{1}{J+1} (\mathbf{u}_i^T \sum_{j=0}^J \mathbf{P}_{(j)} \mathbf{v}_k)^2 \\ &= \sum_{i=1}^n \lambda_i \frac{1}{J+1} (\mathbf{u}_i^T \mathbf{v}_k)^2 = \frac{1}{J+1} \sum_{i=1}^n \lambda_i \beta_i^2 \end{aligned} \quad (32)$$

Substituting the above two equations into (30), we have

$$\begin{aligned} \rho(\Theta_k) &\leq \frac{J}{J+1} \sum_{i=1}^n \lambda_i \beta_i^2 = \frac{J}{J+1} \mathbf{v}_k^T \Theta_{k-1} \mathbf{v}_k \\ &\leq \frac{J}{J+1} \rho(\Theta_{k-1}). \end{aligned} \quad (33)$$

Therefore, $\sum_{k=0}^K \rho(\Theta_k)$ converges absolutely.

2) Finally, we prove that

$$\Theta_0 \preceq \sum_{j=0}^J \mathbf{P}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{P}_{(j)}. \quad (34)$$

Because $\Theta_{n+1} \succeq \Theta_n - \sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_n \mathbf{P}_{(j)}$, we have

$$\Theta_n - \Theta_{n+1} \preceq \sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_n \mathbf{P}_{(j)}. \quad (35)$$

Therefore,

$$\sum_{n=0}^K (\Theta_n - \Theta_{n+1}) \preceq \sum_{n=0}^K \sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_n \mathbf{P}_{(j)}. \quad (36)$$

Letting $K \rightarrow \infty$ completes the argument. \blacksquare

Under a shift-invariant conjugate mirror system where $\mathbf{P}_{(j)} = \mathbf{W}_{(j)}^T \mathbf{W}_{(j)}$ can be diagonalized by the Fourier matrix, $\mathbf{P}_{(i)}^T \Theta \mathbf{P}_{(j)} + \mathbf{P}_{(j)}^T \Theta \mathbf{P}_{(i)}$ is also PSD for any Θ that is PSD and can be diagonalized by the Fourier matrix. This leads to $\Theta_{k+1} = \Theta_k - \sum_{j=0}^J \mathbf{P}_{(j)}^T \Theta_k \mathbf{P}_{(j)}$ and results in

$$\Theta_0 = \sum_j \mathbf{P}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{P}_{(j)}, \quad (37)$$

which means the rhs is the tightest upper bound for Θ_0 .

For the DWT, the above bound also appears to be tighter than the one in [1] in practice. Their system considers the upper bound

$$\begin{aligned} & \sum_i \sum_{j>i} 2\mathbf{x}^T \mathbf{P}_{(i)}^T \Theta_0 \mathbf{P}_{(j)} \mathbf{x} \\ & \leq_a \sum_i \sum_{j>i} 2 \|\mathbf{w}_{(i)}\| \rho(\mathbf{M}_{(i)}^T \Theta_0 \mathbf{M}_{(j)}) \|\mathbf{w}_{(j)}\| \\ & \leq_b \sum_i \sum_{j>i} \rho(\mathbf{M}_{(i)}^T \Theta_0 \mathbf{M}_{(j)}) (\|\mathbf{w}_{(i)}\|^2 + \|\mathbf{w}_{(j)}\|^2) \end{aligned} \quad (38)$$

For any given i and j and non-zero \mathbf{x} , the equality a holds only if $\mathbf{w}_{(i)} = \mathbf{W}_{(i)} \mathbf{x}$ and $\mathbf{w}_{(j)} = \mathbf{W}_{(j)} \mathbf{x}$ are both in the direction of the largest eigenvector of $\rho(\mathbf{M}_{(i)}^T \Theta_0 \mathbf{M}_{(j)})$ and $\mathbf{w}_{(i)} = k \mathbf{w}_{(j)}$ ⁴. This means we can swap $\mathbf{w}_{(i)}$ and $\mathbf{w}_{(j)}$, i.e. $\mathbf{w}_{(i)}^T \mathbf{M}_{(i)}^T \Theta_0 \mathbf{M}_{(j)} \mathbf{w}_{(j)} = \mathbf{w}_{(j)}^T \mathbf{M}_{(j)}^T \Theta_0 \mathbf{M}_{(i)} \mathbf{w}_{(i)}$, but the way that $\mathbf{w}_{(i)}$ is constructed ensures that $\mathbf{M}_{(j)} \mathbf{w}_{(i)} = 0$ for any $i \neq j$. Therefore, the inequality a is strict if there is any $\mathbf{M}_{(i)}^T \Theta_0 \mathbf{M}_{(j)} \neq 0$. In contrast to [1], our system upper bounds $2\mathbf{x}^T \mathbf{P}_{(i)}^T \Theta_0 \mathbf{P}_{(j)} \mathbf{x}$ by taking the negative eigenvalues out without changing the maximum positive values.

For the redundant frame formed by a union of orthonormal transforms where there is significant aliasing between subbands, the system in [1] considers $\rho(\mathbf{M}_{(i)}^T \Theta_0 \mathbf{M}_{(j)})$ from different orthonormal bases, while our system does not need to. This makes our parameter much tighter than theirs.

Note that the equality of (33) holds when the $\mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{v}_k$ equal each other for all j . For a well-designed wavelet system which has good frequency selectivity, there are normally only a few dominant $\mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{v}_k$ which makes $\sum_{j=0}^J (\mathbf{u}_i^T \mathbf{P}_{(j)} \mathbf{v}_k)^2$ much bigger than $(\mathbf{u}_i^T \sum_{j=0}^J \mathbf{P}_{(j)} \mathbf{v}_k)^2 / (J+1)$ and hence the convergence rate is much better than $J/(J+1)$ per iteration.

⁴There is one exception, when $\mathbf{P}_{(i)}^T \Theta_0 \mathbf{P}_{(j)} = 0$ for any i and j , equality a and b hold for whatever $\mathbf{w}_{(i)}$ and $\mathbf{w}_{(j)}$ are. This happens when Θ_0 is the identity matrix.

APPENDIX B
PROOF FOR COROLLARY 1

For any $\mathbf{x} = \mathbf{M}\mathbf{u}$,

$$\begin{aligned}
& \mathbf{u}^T \mathbf{M}^T \Theta_0 \mathbf{M} \mathbf{u} = \mathbf{x}^T \Theta_0 \mathbf{x} \\
& \leq \mathbf{x}^T \sum_{j=0}^J \mathbf{P}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{P}_{(j)} \mathbf{x} \\
& = \sum_{j=0}^J \hat{\mathbf{u}}_j^T \mathbf{M}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{M}_{(j)} \hat{\mathbf{u}}_j \\
& = \mathbf{u}^T \sum_{j=0}^J \mathbf{M}_{(j)}^T \left(\sum_{k=0}^{\infty} \Theta_k \right) \mathbf{M}_{(j)} \mathbf{u},
\end{aligned} \tag{39}$$

The last line of the above equation holds because the wavelet transform is orthonormal, i.e. $\hat{\mathbf{u}}_j = \mathbf{u}_j$.

REFERENCES

- [1] I. Bayram and I.W. Selesnick. A subband adaptive iterative shrinkage/thresholding algorithm. *Signal Processing, IEEE Transactions on*, 58(3):1131–1143, march 2010.
- [2] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pages 3869–3872, 2008.
- [3] P.L. Combettes and V.R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4):1164–1200, 2005.
- [4] 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.
- [5] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23:947, 2007.
- [6] M.A.T. Figueiredo, J.M. Bioucas-Dias, and R.D. Nowak. Majorization–minimization algorithms for wavelet-based image restoration. *Image Processing, IEEE Transactions on*, 16(12):2980–2991, 2007.
- [7] 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.
- [8] S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3rd edition, 2008.
- [9] 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.
- [10] I.W. Selesnick and M.A.T. Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7446, page 11, 2009.
- [11] 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.
- [12] C. Vonesch and M. Unser. A fast multilevel algorithm for wavelet-regularized image restoration. *Image Processing, IEEE Transactions on*, 18(3):509–523, 2009.
- [13] S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.
- [14] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.
- [15] Y. Zhang and N. Kingsbury. A Bayesian wavelet-based multidimensional deconvolution with sub-band emphasis. In *Engineering in Medicine and Biology Society, 2008.*, pages 3024–3027, 2008.



Zhang, Yingsong received the B.Sc. degree from mathematics department, East China Normal University in 1999, and M.Eng from Automation department, Tsinghua University, China, in 2007. She just finished her Ph.D. in sparse regulated inverse problem and its applications on image restoration, with Cambridge University, Cambridge, UK, in 2012.

Her current research interests include wavelet based methods for image restoration (denoising and deconvolution), sparse methods for signal recovery (compressive sensing).