FAST APPROXIMATE L0-NORM DECONVOLUTION USING STRUCTURED WAVELET DOMAIN PRIORS

Timothy D. Roberts, Nick Kingsbury

Signal Processing & Communication Lab
Department of Engineering
University of Cambridge, U.K.
Email: {tr331,ngk10}@cam.ac.uk

ABSTRACT

In this paper, we propose a new algorithm to solve linear inverse problems using an approximate l_0 penalty on overlapped groups of wavelet coefficients, and apply this to the deconvolution problem specifically. Prior work has shown the improvements gained from using group-sparse penalties over coefficient-sparse penalties for deconvolution and compressed sensing. Instead of minimizing an l_1-norm, as done in prior work, we instead design an overlapping group prior which utilizes the Gaussian scale mixture model, and use this to promote l_0 sparsity. Using a Bayesian argument, we derive a novel convex penalty function, which is a reweighted l_2 approximation to the l_0-norm that can be efficiently minimized. We show that the new group-sparse algorithm produces superior deconvolution results compared to the same algorithm utilizing an unstructured coefficient-sparse penalty.

Index Terms— structured models, deconvolution, iterative algorithms, sparsity.

1. INTRODUCTION

Deconvolution of natural images is a well-studied, ill-posed, linear inverse problem. Natural images are sparsely represented by wavelets [1], and superior deconvolution results have been obtained using this knowledge by promoting solutions with sparse wavelet-domain representations [2, 3]. Furthermore, signal models which exploit overlapped parent-child groupings of these sparse wavelet coefficients are potentially advantageous in efficiently solving deconvolution problems, and these topics are considered here.

Owing to the ability of wavelet transforms to localize edges, coefficient magnitudes exhibit strong statistical dependencies across scales [4]. Early work in exploiting this structured sparsity employed the hidden Markov model [5–7]. While this model is effective for capturing some of the non-Gaussian behavior of wavelet magnitudes and promoting the persistence of magnitudes across scales, it requires learning the dependencies using iterative reweighting schemes. This motivates locally-convex quadratic penalties which can be more efficiently solved.

Rao *et al* [8] have employed a convex overlapped group lasso (l_1) penalty and demonstrated significant improvements over coefficient-wise penalties for deconvolution and compressed sensing problems. In [9], Zhang *et al* have shown improved sparse recovery results over the l_1 sparse penalty using the modified subband-adaptive iterative shrinkage/threshold (MSIST) algorithm, which minimizes the reweighted l_2 approximation to the l_0-norm of wavelet coefficient magnitudes. For this reason, we propose a group sparse solver which incorporates an approximate l_0 group sparse penalty, and apply this to the deconvolution problem, building on some of the strengths in both [8] and [9].

In [8], a simple parent-child, overlapping grouping policy is used to facilitate the modeling of groups so that optimization can be performed on a pre-defined structure. Overlapping group structures are useful when the precise group structure or the degree of persistence of wavelet magnitudes across scales is not known *a priori* [10]. Since the problem of finding the group structure which constitutes the sparse approximation of a signal is NP hard in general [11], we adopt the overlapping group approach in which the interscale group dependencies are gradually learned, starting with a sensible pre-defined group structure.

In [12], Zhang *et al* have shown promising deconvolution results using the MSIST algorithm. This algorithm is derived from Bayesian principles, utilizing a Gaussian scale mixture (GSM) prior, which is a powerful model for capturing the statistics of wavelet coefficients [13–15].

Combining Bayesian wavelet-based image deconvolution using the GSM prior, quadratic penalties, and group sparsity, we propose a new locally-convex group-sparse solver. Unlike the work in [8], we promote sparse solutions by approximating the l_0-norm. Additionally, instead of promoting l_0 sparsity at the coefficient-level as in [12], we sparsify at the group level, leading to better estimates of the sparse approximation of the underlying signal. We have chosen the dual-tree complex wavelet transform (DT-$\mathbb{C}$WT) as our sparsifying transformation, because it has the advantages of approximate shift-invariance and directional selectivity for a limited transform domain redundancy. By incorporating the strong parent-child dependency structure of these complex wavelet coefficient magnitudes into our prior model, we show that we can find group-sparse solutions via fast quadratic minimization.

In sec. 2.1, we introduce the GSM model and the MSIST penalty. We define the group sparse model in sec. 2.2, and extend the MSIST penalty to derive the group sparse MSIST algorithm in sec. 2.3. In sec. 3 we show that, by imposing a simple grouping policy, we can obtain superior results to those obtained without a group structure, using the same basic algorithm and parameters. In sec. 4 we conclude with suggested extensions of this research.

2. METHODS

2.1. Gaussian Scale Mixture (GSM) Model

The GSM model assumes that the statistics of an image can be described locally, so that each wavelet coefficient follows an indepen-

dent Gaussian process with spatially varying statistical parameters. Given an unobserved signal $\mathbf{x} \in \mathbb{R}^N$ and an additive white Gaussian noise process $\mathbf{n} \sim \mathrm{N}(0, \nu^2)$, we observe

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

via matrix $\mathbf{\Phi} = \mathbf{H}\mathbf{W}^T$. Usually $\mathbf{H}$ is a convolution matrix, and $\mathbf{W}^T$ represents an inverse wavelet transform. From a Bayesian point of view, we can reconstruct the image from the observation model, or likelihood $\mathrm{p}(\mathbf{y}|\mathbf{x})$, and a GSM prior on the signal, $\mathrm{p}(\mathbf{x})$:

$$\mathrm{p}(\mathbf{y}|\mathbf{x}) \propto \exp\left(\frac{-\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2}{2\nu^2}\right) \tag{2}$$

$$\mathrm{p}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{S}\mathbf{x}\right) \tag{3}$$

where $\mathbf{S}$ is a diagonal matrix formed from the j-indexed vector $\mathbf{s}$ of inverse variance estimates $\frac{1}{v_j}$ of each element x_j in $\mathbf{x}$.

Since we are using the DT-$\mathbb{C}$WT, which transforms real image pixels into complex wavelet coefficients, the redundant tight-frame representation of the transform requires the coefficients to be split into separate real and imaginary parts in $\mathbf{x}$, to avoid the need for the awkward "real-part" operator in the inverse transform $\mathbf{W}^T$. This can be elegantly handled by regarding the real and imaginary components of each complex coefficient as non-overlapping groups of size 2, and we have incorporated this into our previous work on sparsity-regularized deconvolution [9, 12]. In the present context, we obtain estimates of each v_j by taking the mean of the squares of the real and imaginary parts of each complex coefficient and also including ϵ^2, an estimate of the variance of the noise present in the wavelet coefficients. Hence,

$$v_j = \mathbb{E}\left[|x_j|^2\right] \approx \frac{1}{2}(x_{j,\mathrm{Re}}^2 + x_{j,\mathrm{Im}}^2) + \epsilon^2 \tag{4}$$

The maximum *a posteriori* (MAP) estimate $\mathbf{x}_{\mathrm{MAP}}$ obtained by maximizing the posterior $\mathrm{p}(\mathbf{x}|\mathbf{y})$, or by minimizing its negative logarithm, is given by:

$$\begin{aligned}\mathbf{x}_{\mathrm{MAP}} &= \mathrm{argmin}_{\mathbf{x}}[-\ln(\mathrm{p}(\mathbf{y}|\mathbf{x})\mathrm{p}(\mathbf{x}))] \\ &= \mathrm{argmin}_{\mathbf{x}}\left[\frac{\|\mathbf{\Phi}\mathbf{x} - \mathbf{y}\|_2^2}{2\nu^2} + \frac{1}{2}\mathbf{x}^T\mathbf{S}\mathbf{x}\right]\end{aligned} \tag{5}$$

The expression in eq. (5) can be minimized using fast threshold Landweber iterations [3]. This is the MSIST algorithm, which minimizes the data fidelity term while promoting sparsity in an approximate l_0 sense, if the terms in $\mathbf{S}$ are equal to the reciprocal of the expected variance of each term in $\mathbf{x}$. We now consider how a group-sparse penalty may be incorporated into MSIST, and introduce a separate model for the prior in eq. (3) which incorporates group structure, producing more robust estimates of the v_j in eq. (4).

2.2. Overlapping Group Model

In this section, we introduce the notation and terminology for the overlapping group sparse model. In this setup, x_k can belong to two or more separate *groups*, but a problem exists when we wish to use the statistics of those groups to update the current estimate of x_k. As in [8], a strategy we use for dealing with overlapping groups is to introduce a new vector $\hat{\mathbf{x}} \in \mathbb{R}^M$ of latent variables comprising multiple copies of each x_k. An advantage of this duplication strategy is that it does not require duplicating the columns of $\mathbf{\Phi}$ in eq. (1) for each of the latent variables, and thus minimizes computational complexity. Generally $M \geq N$, with equality for the special case when each coefficient comprises a non-overlapping group by itself. The size M of the latent variable vector depends on the grouping policy. We may also consider clustering elements of $\hat{\mathbf{x}}$ which correspond to each element x_k of $\mathbf{x}$. We define the following terms:

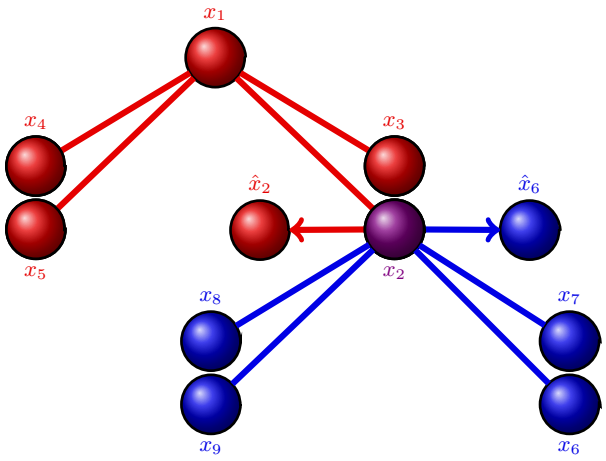

Fig. 1. Diagram showing quadtree hierarchy of x and $\hat{x}$ with a parent+4children overlapping group policy. Groups $g_1 = \{1, 2, 3, 4, 5\}$ and $g_2 = \{2, 6, 7, 8, 9\}$ are shown using red and blue undirected edges, respectively. Cluster elements $\hat{x}_2$ and $\hat{x}_6$ belong to groups $\hat{g}_1 = \{1, 2, 3, 4, 5\}$ and $\hat{g}_2 = \{6, 10, 11, 12, 13\}$, respectively, which aren't fully shown.

- *Group* refers to a subvector $\mathbf{x}_{g_i}$ of $\mathbf{x}$, or a subvector $\hat{\mathbf{x}}_{\hat{g}_i}$ of $\hat{\mathbf{x}}$, indexed by the sets of indices $g_i \subseteq \{1 \ldots N\}$ and $\hat{g}_i \subseteq \{1 \ldots M\}$ for $i \in \{1 \ldots N_g\}$, where N_g is the number of groups. The groups g_i may be overlapping, whereas the groups $\hat{g}_i$ may not. Two groups from the parent+4children overlapping group policy are shown in fig. 1 using red and blue undirected edges.
- *Cluster* refers to a subvector $\hat{\mathbf{x}}_{G_k}$ of $\hat{\mathbf{x}}$, indexed by the set of indices $G_k \subseteq \{1 \ldots M\}$ for $k \in \{1 \ldots N\}$ such that all elements of $\hat{\mathbf{x}}_{G_k}$ correspond to the multiple copies of a single element x_k of $\mathbf{x}$. One cluster of size 2 for the overlapped variable x_2 is shown in fig. 1 using arrows.

2.3. Group Sparse MSIST Algorithm

In this section, we consider sparsifying at the group level while imposing a *cluster fidelity constraint* in order to keep the *master* close to the mean of its *duplicate* elements.

First, adopting a Bayesian interpretation, we estimate $\mathbf{x}$ and $\hat{\mathbf{x}}$ jointly, maximizing the posterior $\mathrm{p}(\mathbf{x}, \hat{\mathbf{x}}|\mathbf{y})$.

$$\begin{aligned}[\mathbf{x}, \hat{\mathbf{x}}]_{\mathrm{MAP}} &= \mathrm{argmin}_{\mathbf{x},\hat{\mathbf{x}}}[-\ln(\mathrm{p}(\mathbf{y}|\mathbf{x}, \hat{\mathbf{x}})\mathrm{p}(\mathbf{x}, \hat{\mathbf{x}}))] \\ &= \mathrm{argmin}_{\mathbf{x},\hat{\mathbf{x}}}[-\ln(\mathrm{p}(\mathbf{y}|\mathbf{x}, \hat{\mathbf{x}})\mathrm{p}(\mathbf{x}|\hat{\mathbf{x}})\mathrm{p}(\hat{\mathbf{x}}))]\end{aligned} \tag{6}$$

For all $k \in \{1 \ldots N\}$, we assume the conditional distribution of each x_k is normal, with mean $\bar{x}_k$ and variance $v_{G,k}$. Hence

$$\mathrm{p}(x_k|\hat{x}_{G_k}) = \frac{1}{\sqrt{2\pi v_{G,k}}}\exp\left(-\frac{(x_k - \bar{x}_k)^2}{2v_{G,k}}\right) \tag{7}$$

where $\bar{x}_k$, the *cluster mean*, and $v_{G,k}$, the *cluster variance*, are estimated from $\hat{\mathbf{x}}_{G_k}$ by

$$\bar{x}_k = \frac{1}{|G_k|} \sum_{c \in G_k} \hat{x}_c \tag{8}$$

$$v_{G,k} = \frac{1}{|G_k|} \sum_{c \in G_k} |\hat{x}_c - \bar{x}_k|^2 \tag{9}$$

Under an approximate group-independence assumption, we can obtain the group prior:

$$\mathrm{p}(\hat{\mathbf{x}}) \approx \prod_{i \in \{1 \ldots N_g\}} \mathrm{p}(\hat{\mathbf{x}}_{\hat{g}_i}) \propto \prod_{i \in \{1 \ldots N_g\}} \prod_{k \in \hat{g}_i} \exp\left(-\frac{\hat{x}_k^2}{2 v_{g,i}}\right) \tag{10}$$

where $v_{g,i}$ is estimated from $\hat{x}_{\hat{g}_i}$. We also assume an independent prior $\mathrm{p}(v_{g,i}|\epsilon^2) \propto \exp(-\epsilon^2/2v_{g,i})$ for each $v_{g,i}$ as in [9], so that

$$v_{g,i} = \mathbb{E}\left[|\hat{x}_{\hat{g}_i}|^2\right] = \frac{1}{2|\hat{g}_i|} \|\hat{\mathbf{x}}_{\hat{g}_i}\|_2^2 + \epsilon^2 \tag{11}$$

Next, we form diagonal matrices $\bar{\mathbf{S}}$ ($N \times N$) and $\hat{\mathbf{S}}$ ($M \times M$) from vectors $\bar{\mathbf{s}}$ and $\hat{\mathbf{s}}$. The entries $\bar{s}_k$ of N-length vector $\bar{\mathbf{s}}$ are the cluster inverse variance estimates $\frac{1}{v_{G,k}}$, and the entries $\hat{s}_i$ of N_g-length vector $\hat{\mathbf{s}}$ are the group inverse variance estimates $\frac{1}{v_{g,i}}$. When $N_g < M$, $\hat{\mathbf{S}}$ is formed from an expanded version of $\hat{\mathbf{s}}$ so that $\hat{s}_i$ is repeated to give $\hat{\mathbf{S}}_{j,j} = \hat{s}_i$ for each $j \in g_i$. Thus we have the following penalty function derived from eq. (6):

$$\begin{aligned} J(\mathbf{x}, \mathbf{z}, \hat{\mathbf{x}}, \hat{\mathbf{S}}, \bar{\mathbf{x}}, \bar{\mathbf{S}}) = & \|\mathbf{\Phi x} - \mathbf{y}\|_2^2 \\ & + \nu^2 \left[(\mathbf{x} - \bar{\mathbf{x}})^T \bar{\mathbf{S}} (\mathbf{x} - \bar{\mathbf{x}}) - \ln(|\bar{\mathbf{S}}|)\right] \\ & + \nu^2 \left[\hat{\mathbf{x}}^T \hat{\mathbf{S}} \hat{\mathbf{x}} - \ln(|\hat{\mathbf{S}}|) + \epsilon^2 \mathrm{tr}(\hat{\mathbf{S}})\right] \\ & + (\mathbf{x} - \mathbf{z})^T \mathbf{\Lambda}_\alpha (\mathbf{x} - \mathbf{z}) - \|\mathbf{\Phi}(\mathbf{x} - \mathbf{z})\|_2^2 \end{aligned} \tag{12}$$

The penalty function in eq. (6) includes the *cluster fidelity constraint* (second line) and group sparsity penalty (third line). In practice we estimate $\bar{\mathbf{x}}$ and $\bar{\mathbf{S}}$ from $\hat{\mathbf{x}}$, but for generality eq. (12) is left as a function of all three quantities. The Majorization-Minimization (MM) terms (fourth line) are not part of the Bayesian derivation, but serve to remove the difficult cross-coupling term $\mathbf{x}^T \mathbf{\Phi}^T \mathbf{\Phi x}$ which results from expanding the first line. The MM term is strictly positive, and causes $J(\cdot)$ to upper-bound the MAP penalty (the first three lines), as long as the diagonal matrix $\mathbf{\Lambda}_\alpha$ is chosen such that $(\mathbf{\Lambda}_\alpha - \mathbf{\Phi}^T \mathbf{\Phi})$ is positive definite [3, 12].

In order to minimize eq. (12) we consider estimating the quantities $\bar{\mathbf{x}}$ and $\bar{\mathbf{S}}$ in terms of $\hat{\mathbf{x}}$. The vector of cluster means $\bar{\mathbf{x}} \in \mathbb{R}^N$ can be written using an explicit matrix form:

$$\bar{\mathbf{x}} = \mathbf{\Lambda}_{|G_k|} \mathbf{D} \hat{\mathbf{x}} = \mathbf{A} \hat{\mathbf{x}} \tag{13}$$

where $\mathbf{D}$ is an $N \times M$ matrix which performs the sum in eq. (8), whose $(k, m)^{\text{th}}$ element is given by

$$D_{k,m} = \begin{cases} 1, & \text{if } \hat{x}_m \in \hat{\mathbf{x}}_{G_k} \\ 0, & \text{otherwise} \end{cases} \quad \text{for } k \in \{1 \ldots N\} \text{ and } m \in \{1 \ldots M\} \tag{14}$$

and $\mathbf{\Lambda}_{|G_k|}$ is an $N \times N$ diagonal matrix of $\frac{1}{|G_k|}$.

In practice we can achieve a similar effect of reducing the second line in eq. (12) by minimizing the l_2 norm using a user-defined Lagrangian parameter τ^2. In this way, we can control the relative strength of the closeness of x_k to the mean of $\hat{\mathbf{x}}_{G_k}$ using τ^2 while encouraging sparseness at the group level using ν^2. In practice, we find that using τ^2 instead of $\bar{\mathbf{S}}$ produces better deconvolution results. Additionally, this simplification avoids computing eq. (9) on each iteration. Substituting $\tau^2 \mathbf{I}$ for $\nu^2 \bar{\mathbf{S}}$ in eq. (12), we have

$$\begin{aligned} J(\mathbf{x}, \mathbf{z}, \hat{\mathbf{x}}, \hat{\mathbf{S}}, \bar{\mathbf{x}}, \bar{\mathbf{S}}) = & \|\mathbf{\Phi x} - \mathbf{y}\|_2^2 \\ & + \tau^2 \|\mathbf{x} - \bar{\mathbf{x}}\|_2^2 + \text{const} \\ & + \nu^2 \left(\hat{\mathbf{x}}^T \hat{\mathbf{S}} \hat{\mathbf{x}} - \ln(|\hat{\mathbf{S}}|) + \epsilon^2 \mathrm{tr}(\hat{\mathbf{S}})\right) \\ & + (\mathbf{x} - \mathbf{z})^T \mathbf{\Lambda}_\alpha (\mathbf{x} - \mathbf{z}) - \|\mathbf{\Phi}(\mathbf{x} - \mathbf{z})\|_2^2 \end{aligned} \tag{15}$$

Using eq. (13) where appropriate, we derive update rules to minimize eq. (15) at each iteration n by differentiating $J(\cdot)$ with respect to each variable in turn and setting the derivatives to 0:

$$\begin{aligned} \frac{\partial J(\cdot)}{\partial \mathbf{x}} &= 2\left[-\mathbf{\Phi} \mathbf{y}^T + \mathbf{\Lambda}_\alpha (\mathbf{x} - \mathbf{z}) + \mathbf{\Phi}^T \mathbf{\Phi z} + \tau^2 (\mathbf{x} - \bar{\mathbf{x}})\right] \\ \mathbf{x}_{n+1} &= \left(\mathbf{\Lambda}_\alpha + \tau^2\right)^{-1} \left[\mathbf{\Phi} \mathbf{y}^T + (\mathbf{\Lambda}_\alpha - \mathbf{\Phi}^T \mathbf{\Phi}) \mathbf{z}_n + \tau^2 \bar{\mathbf{x}}_n\right] \end{aligned} \tag{16}$$

$$\begin{aligned} \frac{\partial J(\cdot)}{\partial \mathbf{z}} &= 2(\mathbf{\Lambda}_\alpha - \mathbf{\Phi}^T \mathbf{\Phi})(\mathbf{x} - \mathbf{z}) \\ \mathbf{z}_{n+1} &= \mathbf{x}_{n+1} \end{aligned} \tag{17}$$

$$\begin{aligned} \frac{\partial J(\cdot)}{\partial \hat{\mathbf{x}}} &= 2\left[-\mathbf{A}^T \tau^2 (\mathbf{x} - \mathbf{A}\hat{\mathbf{x}}) + \nu^2 \hat{\mathbf{S}} \hat{\mathbf{x}}\right] \\ \hat{\mathbf{x}}_{n+1} &= \left[\tau^2 \mathbf{A}^T \mathbf{A} + \nu^2 \hat{\mathbf{S}}_n\right]^{-1} \tau^2 \mathbf{A}^T \mathbf{x}_n \end{aligned} \tag{18}$$

$$\begin{aligned} \frac{\partial J(\cdot)}{\partial \hat{s}_i} &= v_{g,i} - \frac{1}{\hat{s}_i} \\ \hat{s}_{i,n+1} &= \frac{1}{v_{g,i,n}}, \quad \forall\, i \in \{1 \ldots N_g\} \end{aligned} \tag{19}$$

Let $\hat{\tilde{\mathbf{S}}}$ be the $M \times M$ block-diagonal matrix $\left[\tau^2 \mathbf{A}^T \mathbf{A} + \nu^2 \hat{\mathbf{S}}_n\right]$ in eq. (18). The $(p, q)^{\text{th}}$ entry of each of N $|G_k| \times |G_k|$ blocks, $\hat{\tilde{\mathbf{S}}}_k$, of $\hat{\tilde{\mathbf{S}}}$ is given by

$$\hat{\tilde{\mathbf{S}}}_{k,(p,q)} = \begin{cases} \frac{\tau^2}{|G_k|^2} + \nu^2 \hat{s}_{i,(p)}, & \text{if } p = q \\ \frac{\tau^2}{|G_k|^2}, & \text{otherwise} \end{cases} \tag{20}$$

where $\hat{s}_{i,(p)}$ is the group inverse variance estimate corresponding to the $(p)^{\text{th}}$ element of $\hat{\mathbf{x}}_{G_k}$. Thus, the update of $\hat{\mathbf{x}}$ in eq. (18) can be accomplished via

$$\hat{\mathbf{x}}_{n+1} = \tau^2 \hat{\tilde{\mathbf{S}}}^{-1} \mathbf{A}^T \mathbf{x}_n \tag{21}$$

which requires N small matrix inverse operations and M vector-multiply operations.

3. RESULTS

In this section, we test our claim that the algorithm which incorporates the group sparse prior in eq. (10) and minimizes the penalty in eq. (15), referred to as Group MSIST (MG), outperforms the coefficient-sparse MSIST algorithm, referred to as Coefficient MSIST (MC), which solves eq. (5). We have used the parent+4children grouping, as in fig. 1, for these experiments.

Using the standard Cameraman test image, we have simulated various BSNR levels and blurring operations and compared with

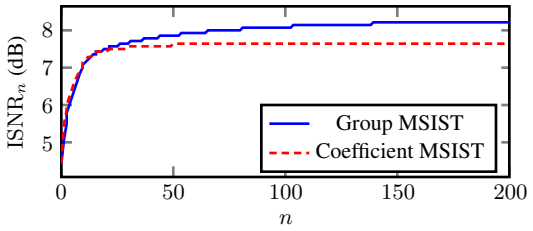

Fig. 2. Average ISNR$_n$ (dB) on the 40dB BSNR (9×9 uniform blur) Cameraman for Group MSIST and Coefficient MSIST over 30 noise realizations.

previously reported MSIST deconvolution results [12]. For these tests, we have used the 9×9 uniform and the 15×15 Gaussian ($\sigma = 1$) blur kernels for $\mathbf{H}$ and the DT-$\mathbb{C}$WT as our redundant sparsifying transform $\mathbf{W}$ to produce $\mathbf{\Phi}$, and added Gaussian noise $\mathbf{n}$ to produce the observation $\mathbf{y}$, as in eq. (1). Additionally, we've used the same values for $\mathbf{\Lambda}_\alpha$ as reported in [12] for the 9×9 uniform blur test case. By varying the nominal noise variance ν^2, we've produced several blurred signal-to-noise ratio (BSNR) test scenarios. BSNR is computed using

$$\text{BSNR} = 10\log_{10} \frac{\|\mathbf{H}\mathbf{u}_r - \overline{\mathbf{H}\mathbf{u}_r}\|_2^2}{N_p \nu^2} \tag{22}$$

where $\mathbf{u}_r = \mathbf{W}^T\mathbf{x} \in \mathbb{R}^{N_p}$ is the original ground truth signal before applying the over-complete transform $\mathbf{W}$ and $\overline{\mathbf{H}\mathbf{u}_r}$ is the mean of $\mathbf{H}\mathbf{u}_r$.

We've seeded both algorithms with $\mathbf{x}_0$ obtained by using the regularized Wiener filter:

$$\mathbf{x}_0 = \mathbf{W}\left[(\mathbf{H}^T\mathbf{H} + 10^{-3}\nu^2\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}\right] \tag{23}$$

and on each iteration, we've computed the improvement in signal-to-noise ratio (ISNR) via

$$\text{ISNR}_n = 10\log_{10} \frac{\|\mathbf{y} - \mathbf{u}_r\|_2^2}{\|\mathbf{W}^T\mathbf{x}_n - \mathbf{u}_r\|_2^2} \tag{24}$$

and use this to compare the performance of our algorithm to the MC benchmark.

For almost all test cases, incorporating the group sparse penalty leads to improved deconvolution results in terms of visual quality and final ISNR. Fig. 3 shows the final visual result obtained from applying the MC and MG algorithms for 200 iterations. In fig. 2 and tables 1 and 2, we show average ISNR values at iteration n obtained from repeating our experiments over 30 noise realizations. We've ensured that the parameters ν and ϵ for the coefficient MSIST and group MSIST experiments were the same for each test scenario. In particular, we've employed geometric homotopy continuation rules for these parameters with a shrinkage factor of 0.8 per iteration, so that the reweighted l_2 penalty $\hat{\mathbf{x}}^T\hat{\mathbf{S}}\hat{\mathbf{x}}$ more closely approximates the l_0-norm as ϵ is decreased, which is explained in detail in [12]. For all of these experiments, we have used $\tau = 0.1$.

While this algorithm generally requires N block inverses, with this chosen grouping policy the maximum block size is 2×2 ($G_k \leq 2$), so that the matrix inverse has an explicit form which permits efficient vector multiply operations. Thus, the per-iteration computational overhead is not significantly greater for MG than MC.

(a) Ground truth. (b) Blurred, noisy. (c) MC: 7.64dB. (d) MG: 8.22dB.

Fig. 3. Visual deconvolution results on 40dB BSNR (9×9 uniform blur) Cameraman, with ISNR$_{200}$ values.

n	20dB		40dB		50dB	
	MC	MG	MC	MG	MC	MG
10	2.58	**2.68**	**7.12**	7.09	8.76	**9.61**
30	2.99	**3.20**	7.53	**7.66**	10.29	**10.34**
50	3.19	**3.33**	7.60	**7.86**	10.60	**10.62**
70	3.31	**3.36**	7.63	**7.98**	10.67	**10.80**
100	**3.40**	3.36	7.64	**8.08**	10.68	**10.97**

Table 1. Average ISNR$_n$ (dB) on the Cameraman for the Group MSIST (MG) and Coefficient MSIST (MC) over 30 noise realizations using the 9×9 uniform blur for 20dB, 40dB, and 50dB BSNR test scenarios. For MC, 20dB and 50dB BSNR results are those reported in [12].

n	40dB		50dB	
	MC	MG	MC	MG
10	**4.50**	3.74	**8.96**	8.80
30	**6.34**	6.32	**9.64**	9.63
50	6.76	**6.89**	9.84	**9.98**
70	6.84	**7.05**	9.90	**10.14**
100	6.84	**7.14**	9.91	**10.26**

Table 2. Average ISNR$_n$ (dB) on the Cameraman for the Group MSIST (MG) and Coefficient MSIST (MC) over 30 noise realizations using the 15×15 Gaussian blur, $\sigma = 1$ for 40dB and 50dB BSNR test scenarios.

4. CONCLUSION

In this paper, we've shown improved deconvolution results by incorporating group sparsity using a Bayesian argument. These results come at the added per-iteration expense of doing N small inverses as part of the variable duplication approach. However, when small cluster sizes are employed, this expense can be minimized. Future work may include faster approximation of these inverses, or employing a strategy which simply optimizes the coefficient-wise variances in the standard GSM model using group information, potentially avoiding the need for latent variables. Using other conjugate priors, such as the Gamma prior, and variational Bayesian methods for estimating the wavelet coefficient densities, may lead to additional gains. The ISNR results for this and other sparsity-based methods do not compare favorably with the state-of-the-art results of BM3D [16]. However, relatively efficient algorithms which do not rely on block matching and also leverage the strengths of statistical modelling and wavelet domain thresholding, such as SURE-LET [17], show promise for general linear inverse problems and tractable extensions to three-dimensional deconvolution.

5. REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, Academic Press, 3rd edition, 2008.

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

[3] C. Vonesch and M. Unser, "A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution," *Image Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 539–549, 2008.

[4] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 57, no. 9, pp. 3488–3497, 2009.

[5] J.K. Romberg, M.B. Wakin, H. Choi, and R.G. Baraniuk, "A geometric hidden markov tree wavelet model," in *Optical Science and Technology, SPIE's 48th Annual Meeting*. International Society for Optics and Photonics, 2003, pp. 80–86.

[6] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *Signal Processing, IEEE Transactions on*, vol. 46, no. 4, pp. 886–902, 1998.

[7] M.F. Duarte, M.B. Wakin, and R.G. Baraniuk, "Wavelet-domain compressive signal reconstruction using a hidden markov tree model," in *Acoustics, Speech and Signal Processing, 2008. IEEE International Conference on*. IEEE, 2008, pp. 5137–5140.

[8] N.S. Rao, R.D. Nowak, S.J. Wright, and N.G. Kingsbury, "Convex approaches to model wavelet sparsity patterns," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 1917–1920.

[9] Y. Zhang and N. Kingsbury, "Fast L0-based sparse signal recovery," in *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*. IEEE, 2010, pp. 403–408.

[10] S.D. Babacan, S. Nakajima, and M.N. Do, "Bayesian group-sparse modeling and variational inference," *Submitted to IEEE Transactions on Signal Processing*, 2012.

[11] L. Baldassarre, N. Bhan, V. Cevher, and A. Kyrillidis, "Group-sparse model selection: Hardness and relaxations," *arXiv preprint, arXiv:1303.3207*, 2013.

[12] Y. Zhang and N. Kingsbury, "Improved bounds for modified subband-adaptive iterative shrinkage/thresholding algorithms," *Image Processing, IEEE Transactions on*, vol. 22, no. 4, pp. 1373–1381, 2013.

[13] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *Image Processing, IEEE Transactions on*, vol. 12, no. 11, pp. 1338–1351, 2003.

[14] P. de Rivaz and N. Kingsbury, "Bayesian image deconvolution and denoising using complex wavelets," in *Image Processing, 2001. Proceedings. 2001 International Conference on*. IEEE, 2001, vol. 2, pp. 273–276.

[15] Y. Zhang and N. Kingsbury, "A Bayesian wavelet-based multi-dimensional deconvolution with sub-band emphasis," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 3024–3027.

[16] Aram Danielyan, "BM3D frames and variational image deblurring," *Image Processing, IEEE . . .*, vol. 2011, no. 213462, pp. 2006–2011, 2012.

[17] Feng Xue, Florian Luisier, and Thierry Blu, "Multi-Wiener SURE-LET deconvolution.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, no. 5, pp. 1954–68, May 2013.