CONVEX APPROACHES TO MODEL WAVELET SPARSITY PATTERNS

Nikhil S. Rao, Robert D. Nowak, Stephen J. Wright

University of Wisconsin-Madison

Nick G. Kingsbury

Department of Engineering,
University of Cambridge

ABSTRACT

Statistical dependencies among wavelet coefficients are commonly represented by graphical models such as hidden Markov trees (HMTs). However, in linear inverse problems such as deconvolution, tomography, and compressed sensing, the presence of a sensing or observation matrix produces a linear mixing of the simple Markovian dependency structure. This leads to reconstruction problems that are non-convex optimizations. Past work has dealt with this issue by resorting to greedy or suboptimal iterative reconstruction methods. In this paper, we propose new modeling approaches based on group-sparsity penalties that leads to convex optimizations that can be solved exactly and efficiently. We show that the methods we develop perform significantly better in deconvolution and compressed sensing applications, while being as computationally efficient as standard coefficient-wise approaches such as lasso.

Index Terms— wavelet modeling, deconvolution, compressed sensing

1. INTRODUCTION

Statistical dependencies among wavelet coefficients are commonly represented by trees or graphical models such as hidden Markov trees (HMTs) [8]. HMTs provide superior denoising results compared to independent coefficient-wise thresholding/shrinkage methods, like the lasso [14]. Fast exact and/or approximate inference algorithms exist in many situations, but not all. In linear inverse problems (e.g., deconvolution, tomography, and compressed sensing) the presence of a sensing/observation matrix can linearly mix the Markovian dependency structure so that simple and exact inference algorithms no longer exist. Past work has dealt with this issue by resorting to greedy or suboptimal iterative reconstruction methods such as those based on belief propagation [9], iterative re-weighting [6], or variants of the Orthogonal Matching Pursuit [2, 1] (see also [12, 7]). In this paper, we propose a new modeling approach based on group-sparsity penalties that leads to convex optimizations that can be solved exactly and efficiently. Our results show that the approach performs much better in deblurring and compressed sensing applications, while being as computationally efficient as standard coefficient-wise approaches. Our work uses the group lasso with overlap formulation introduced in [3], which we further modify to better represent dependencies among wavelet coefficients.

We motivate our problem in section 2. In section 3, we explain how we model the wavelet transform coefficients into overlapping groups. Section 4 outlines the experiments we performed, and the results obtained. We conclude the paper in section 5.

2. PROBLEM FORMULATION

Consider a linear observation model which can represent blurring, tomographic projection, or compressed sensing:

$$y = L\,x\, +\, w\,,$$

where y is the measured data, L is a linear observation operator, x is the image to be reconstructed, and w is additive Gaussian noise. Throughout the paper we will assume a standard matrix-vector representation: the image is represented as a column vector and linear operators are represented as matrices. Images typically have approximately sparse representations in the wavelet domain, and this has led to many approaches to image reconstruction that attempt to exploit this property [5, 13]. For example, the standard ℓ_2/ℓ_1 or *lasso* [14] reconstruction problem is written as

$$\widehat{\theta} \quad := \quad \arg\min_\theta \left\{ \frac{1}{2}\|y - A\theta\|_2^2\, +\, \lambda_l\|\theta\|_1 \right\} \qquad (1)$$

where $A := LW$, the composition of L and the inverse wavelet transform W, θ denotes a set of wavelet coefficients, and $\lambda_l > 0$ is a regularization parameter that balances the tradeoff between fitting to the data y and minimizing the ℓ_1 norm of θ (which serves as a surrogate for sparsity).

The lasso penalty reflects the fact that the wavelet coefficients are approximately sparse, but in reality not all patterns of sparsity are equally plausible/probable. A commonly observed effect is the persistence of large (or small) wavelet coefficient across scales due to the localized nature of edges. Many models have been proposed to represent such patterns, and in particular tree-structured models have been among the most successful and widely used (e.g., hidden Markov trees[8]). Tree-structured models admit very efficient estimation procedures based on pruning or message-passing algorithms in denoising applications where A is the identity operator, but when A is not identity such simple strategies can no longer be applied. In fact, in general the optimization problem resulting from tree models is non-convex (unlike equation (1) above) and so exact solutions are difficult or almost impossible to obtain. This issue has been addressed by resorting to greedy or suboptimal iterative reconstruction procedures [9, 6, 12, 7, 2, 1].

This motivates the main idea and contribution of this paper. While tree models can represent the patterns of sparsity (or approximate sparsity) in the wavelet coefficients of natural images, they are not the only way to capture such effects. We are particularly interested in modeling the so-called "parent-child dependency" [8], which is used in reference to the persistence of large/small wavelet coefficients across scales. Specifically, if a wavelet coefficient at a certain spatial location and scale is large/small, then its "neighboring" coefficients at roughly the same location but finer or coarser scales tend

to be large/small. The term parent-child refers to a pair of coefficients at a certain location and adjacent scales. Our goal is to exploit the fact that the coefficients in each pair are typically both large or both small (or zero) in magnitude. This can be accomplished using an overlapping-group penalty function [3, 11, 10] that generalizes the lasso in a way that captures parent-child dependencies while at the same time retaining its convex nature.

Fig. 1 depicts wavelet quadtree structures and example of parent-child groups. Each (non-leaf) coefficient has an associated orientation (horizontal, vertical or diagonal) and four child coefficients of the same orientation at the finer scale below it. Many options exist for grouping parents with children, and two options are depicted in the figure. Many other grouping schemes are also possible in our framework (e.g., including "grandparent" coefficients as well), but we will not explore such extensions in this paper.
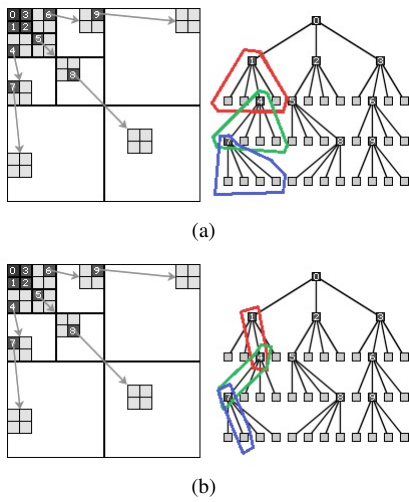

(a)

(b)

Fig. 1. Quadtree corresponding to the 2-d DWT. At each scale, parent coefficients can be grouped with child coefficients. All four children may be grouped together with the parent (above) or the parent can be grouped with each child individually (below).

The main contributions of the paper are threefold: 1) We introduce a new approach to representing the wavelet coefficient sparsity patterns commonly observed in natural images; 2) We adapt and extend recently proposed methods for group lasso with overlaps in order to take advantage of these sparsity patterns in linear inverse problems using simple convex optimization techniques; 3) We demonstrate fast and efficient reconstruction in image deblurring and compressed sensing and significant improvements in reconstruction error relative to the lasso.

3. CONVEX GROUP REGULARIZERS

To encourage solutions that have wavelet coefficient sparsity patterns reflective of the parent-child group structure and persistence of large/small coefficients across scales we apply the group lasso regularization [4]:

$$\widehat{\theta}_{gl} := \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2}\|y - A\theta\|_2^2 + \lambda_g \sum_{g \in \mathcal{G}} \|\theta_g\|_2 \right\}, \quad (2)$$

where $\mathcal{G}$ denotes the collection of all parent-child coefficient groups and g denotes one such group. The quantity θ_g is the subvector of θ of the coefficients in group g. The group lasso penalty enforces group sparsity by setting whole groups to be zero if the ℓ_2 norm of the group is small relative to their importance in the data-fitting term. In most applications of the group lasso, the groups are assumed to be disjoint, but in our case they are overlapping and hence the penalty terms are coupled. Because of this coupling the standard group lasso optimization strategies (e.g., [13]) cannot be directly applied. We offer two approaches to deal with this issue.

3.1. Variable Replication Approach

One way to deal with overlapping groups is to introduce replicates of each coefficient so each group involving a certain coefficient has its own "copy" of it. This "decouples" the overlapping groups from each other. This approach was proposed and analyzed in [3]. The replication of variables (and subsequently the columns of the matrix A) results in a formulation that can be expressed as

$$\widehat{\theta}_{OGLR} := \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2}\|y - \tilde{A}\tilde{\theta}\|_2^2 + \lambda_{rep} \sum_{\tilde{g} \in \tilde{\mathcal{G}}} \|\tilde{\theta}_{\tilde{g}}\|_2 \right\}, \quad (3)$$

where $\widetilde{\theta}$ is the extended vector with replicates and $\tilde{A}$ is a matrix obtained by replicating the corresponding columns of A. Because the penalty function is now separable, computationally efficient iterative shrinkage/thresholding methods can be applied [13].

3.2. Constraint-Based Approach

The overlap group lasso with the replication strategy treats each group independently of each other. This means that, we can have a grandparent - parent coefficient group selected, but not the parent - child group, violating the persistence of wavelet transforms across multiple scales [8]. This motivates the use of a penalty that tends to cause all the coefficients in a location across scales to have a similar value. To this end, we modify equation (3) as

$$\widehat{\theta}_{OGL} \quad := \underset{\theta}{\operatorname{argmin}} \left\{ \frac{1}{2}\|y - A\theta\|_2^2 + \lambda_{ogl} \sum_{\tilde{g} \in \tilde{\mathcal{G}}} \|\tilde{\theta}_{\tilde{g}}\|_2 . \right.$$
$$\left. + \frac{1}{2}\tau^2 \sum_{i=1}^n \sum_{j \in J_i} (\theta_i - \theta_i^{(j)})^2 \right\} \quad (4)$$

where θ_i is the "master copy" of the i-th coefficient and $\theta_i^{(j)}$ denote the copies of it that appear in the group penalty terms. J_i is the set of replicated variables of θ_i. Setting $\tau > 0$ large forces the replicated copies to agree, yielding a solution to the group lasso in (2). This encourages a stronger degree of persistence across scales. To the best of our knowledge, this approach has not been previously proposed for the overlapping group lasso problem. Note that the additional quadratic penalty can be combined with the quadratic data-fitting term to obtain a quadratic plus separable group sparsity penalty, and we can then directly apply standard solvers such as [13]. Henceforth we use L to denote lasso, OGL to denote the Overlap Group Lasso (corresponding to (2) and equation (4) with $\tau \gg 1$) and OGLR to denote the overlap group lasso with replication (equation (3)).

3.3. Sparsity Patterns and Penalties

To demonstrate the effect of the group penalties in contrast to the usual ℓ_1 penalty, we compare their values for the 'cameraman' image. Fig. 2 shows the wavelet coefficients for the image in the standard organization (left) and in a randomized organization (right). The ℓ_1 norm is invariant to the randomization, but the group penalties do change because parent-child dependencies are not preserved. To quantify the degree to which the group penalties encourage parent-child persistence, we compute the ratio of the group penalties for the left image to the randomly organized image on the right. The group penalties are larger when the parent-child relationships are lost due to randomization, and so the ratios are less than 1 for the group penalties. The OLG penalty ratio is smaller than the OLGR (with replication) penalty, indicating that the OLG penalty in (2) more strongly favors the structure in this image compared to the OLGR penalty (3) or the ℓ_1 penalty in the lasso. Next we show experimental evidence that OLG produces better reconstructions.

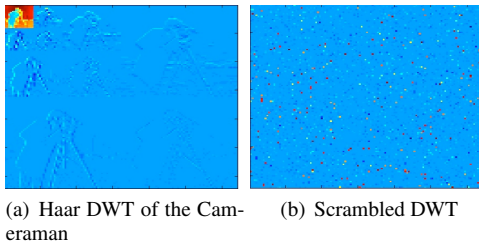(a) Haar DWT of the Cam-
eraman
(b) Scrambled DWT

Fig. 2. Ratio of sparsity penalty of (a) to (b): lasso 1.00; group lasso 0.70; group lasso with replication 0.85. The group penalties significantly favor the structured sparsity pattern of (a)

4. EXPERIMENTS AND RESULTS

We evaluate the proposed approaches on 1-dimensional signals, toy images (shown in Fig. 3), and a real image. We used SpaRSA [13], modified to suit the overlapping groups scenario and the associated modified case, to solve equations (3) and (4). We used the Haar wavelet basis as the sparsity inducing transform. Groups were defined according to the methods explained in section 2.
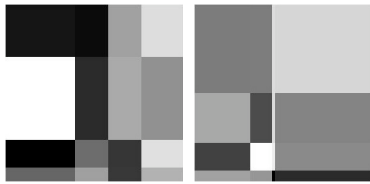

Fig. 3. Sample of toy images used for testing

To illustrate the potential of the proposed methods we first consider compressed sensing and deconvolution results for the cameraman image, resized to 128×128. Fig. 4(a) and Fig. 4(b) show the compressed sensing results. The image was undersampled using a random *iid* gaussian matrix, using only 800 samples for every 64×64 subimage. Fig. 4(c) and Fig. 4(d) show the results of deblurring the image, blurred with a gaussian kernel of variance 1. The samples in both cases were corrupted by WGN of variance 1.
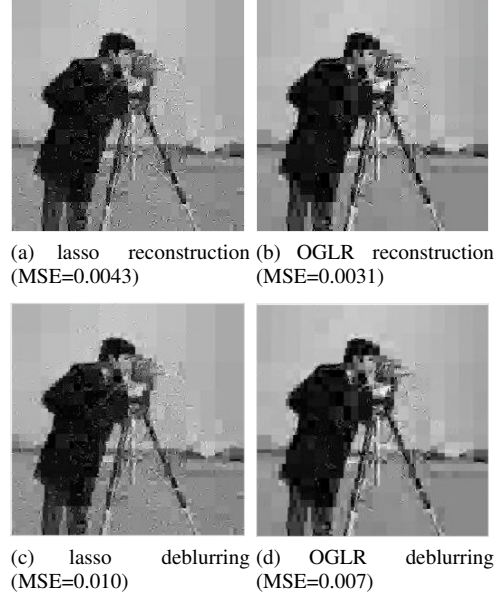

(a) lasso reconstruction
(MSE=0.0043)
(b) OGLR reconstruction
(MSE=0.0031)

(c) lasso deblurring
(MSE=0.010)
(d) OGLR deblurring
(MSE=0.007)

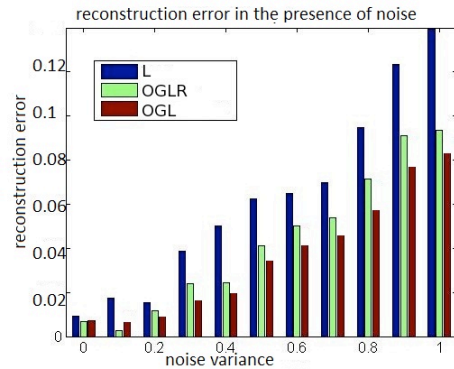**Fig. 4**. Performance on the cameraman image

Fig. 5. Effect of varying the noise variance. Groups are formed according to Fig. 1(a)

4.1. Varying the noise

To evaluate the improvements of the group penalties we first consider performance relative to the signal-to-noise ratio. Our first set of experiments involved testing the recovery scheme in a compressed sensing framework (using an iid gaussian measurement matrix of size 800×4096). The noise variance was varied from 0 to 1 in steps of 0.1, and we measured the mean reconstruction error $||\theta^* - \hat{\theta}||_2^2$. Results obtained using the toy images are plotted in Fig. 5 and Fig. 6. The results are averaged over 100 independent trials at each noise level. A random 'toy' image similar to those in Fig. 3 was generated for each trial. We employed a grid search to pick the best value of λ and τ (wherever applicable). Note that OLG and OLGR produces better results than standard lasso.
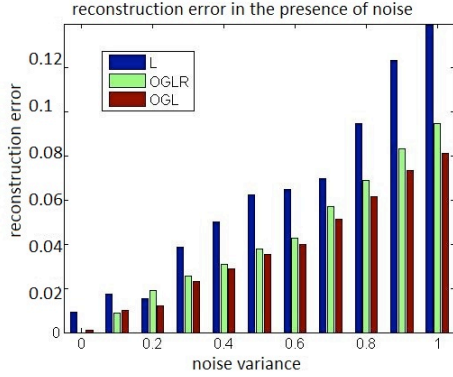
Fig. 6. Effect of varying the noise variance. Groups are formed according to Fig. 1(b)
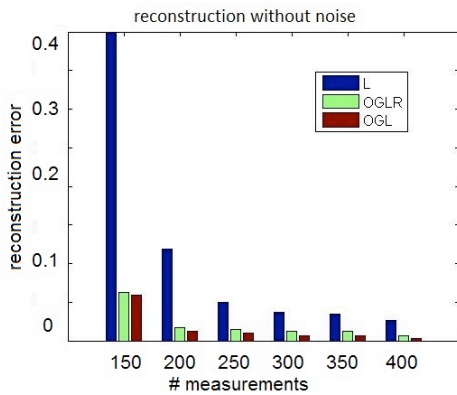

Fig. 7. Effect of varying the number of measurements taken. Note how the overlap lasso needs far fewer measurements than the lasso to achieve low errors. As the number of measurements is increased beyond 300, both methods stop improving significantly.

4.2. Varying the number of measurements

The group penalties also reduce the number of compressed sensing measurements needed to reconstruct images. To study this effect, we varied the number of rows of the matrix. The inputs used were random piecewise constant signals of length 1024, with at most 5 jumps assigned uniformly at random (the 1-D equivalent of Fig. 3), and the associated groups were determined using the binary tree structure of the 1-D DWT, with a group corresponding to a single parent-child pair (an edge of the tree as in Fig. 1(b)). It was observed that we need *far* fewer measurements for robust recovery of signals when this implicit group structure is assumed, as opposed to that needed for the conventional lasso (see Fig. 7).

5. CONCLUSIONS

The proposed group penalties match the sparsity patterns of wavelet coefficients in natural images better than simple coordinate-wise penalties such as the ℓ_1 (lasso) penalty. Like the ℓ_1 penalty, linear inverse problems with group penalties are convex optimizations that can be solved efficiently and exactly. Traditional Markov tree models for coefficients do not lead to convex optimizations. The non-separable nature of the group lasso (when groups are overlapping) was addressed by devising a new optimization criterion that can be solved by standard methods based on separable penalties. The experiments demonstrate the performance gains of the group penalty methods compared to the lasso.

6. REFERENCES

[1] La C. and Do M.N. Tree based orthogonal matching pursuit algorithm for signal reconstruction. *IEEE International Conference on Image Processing, Atlanta, GA.*, pages 1277 – 1280, Oct 2006.

[2] Huang J., Zhang T., and Metaxas D. Learning with structured sparsity. *Technical report, arXiv:0903.3002. Preprint available at http://arxiv.org/pdf/0903.3002v2*, May 2009.

[3] Jacob L., Obozinski G., and Vert J.P. Group lasso with overlap and graph lasso. *Proceedings of the 26th International Conference on machine Learning*, 2009.

[4] Yuan M. and Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the royal statistical society. Series B*, 68:49–67, 2006.

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

[6] Duarte M.F., Wakin M.B., and Baraniuk R.G. Wavelet-domain compressive signal reconstruction using a hidden markov tree model. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 5137–5140, Mar 2008.

[7] Duarte M.F., Cevher V., and Baraniuk R.G. Model-based compressive sensing for signal ensembles. *Allerton*, 2009.

[8] Crouse M.S., Nowak R.D., and Baraniuk R.G. Wavelet based statistical signal processing using hidden markov models. *Transactions on Signal Processing*, 46(4):886–902, 1998.

[9] Schniter P. Turbo reconstruction of structured sparse signals. *Proc. Conference on Information Sciences and Systems*, Mar 2010.

[10] Jenatton R., Mairal J., Obozinski G., , and Bach F. Proximal methods for hierarchical sparse coding. *Technical report, arXiv:1009.3139. submitted*, 2010.

[11] Jenatton R., Audibert J.Y., and Bach F. Structured variable selection with sparsity inducing norms. *Technical report, arXiv:0904.3523. Preprint available at http://arxiv.org/pdf/0904.3523v3*, Sep 2009.

[12] Baraniuk R.G., Cevher V., M.F. Duarte, and Hegde C. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 2010.

[13] Wright S.J., Nowak R.D., and Figueiredo M.A.T. Sparse reconstruction by separable approximation. *Transactions on Signal Processing*, 57:2479–2493, 2009.

[14] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, pages 267–288, 1996.