CORRELATED GAUSSIAN DESIGNS FOR COMPRESSIVE IMAGING

Nikhil S. Rao and Robert D. Nowak

Department of Electrical and Computer Engineering
University of Wisconsin- Madison

ABSTRACT

Statistical correlations among wavelet transform coefficients of images are commonly represented using graphical models. But in linear inverse problems like compressed sensing, the sensing matrix linearly mixes up these dependencies, making recovery of the transform coefficients difficult. Past work has involved using greedy methods to recover images in a compressed sensing framework. Recently, message passing and group lasso based methods have been shown to perform at least as well as traditional approaches. Group lasso based methods are especially viable, since they provide the guarantees that come along with solving a convex program. Standard sensing matrices are well-suited to the recovery of unstructured sparse signals, but the sparsity patterns of natural images are highly structured. In this paper, we look to exploit the intra-group dependencies among coefficients to design sensing matrices that are better matched to image structure than conventional compressed sensing matrices. We show that the new sensing matrices based on structural prior knowledge yield considerably better results compared to standard sensing matrices.

Index Terms— wavelet modeling, compressed sensing, sensing matrix design

1. INTRODUCTION

Discrete wavelet transform coefficients can be arranged in a tree, representing the inter scale dependencies among coefficients at similar locations. Graphical models like Hidden Markov Trees (HMTs) [5] are commonly used to represent such statistical dependencies among Discrete Wavelet Transform (DWT) coefficients of images. Graphical models provide superior performance as compared to independent coefficient-wise thresholding methods such as the lasso. 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 [7], iterative re-weighting [4], or variants of the Orthogonal Matching Pursuit [2, 10]. To retain the guarantees that convex optimization methods provide, the authors in [8] propose modeling the coefficients into overlapping groups and use the latent group lasso [6] to efficiently recover images.

In this paper, we propose an extension of the method in [8] which involves modeling the intra-group dependencies among the coefficients, and incorporating this prior knowledge into the design of the sensing matrix in a compressed sensing framework. We aim to show that by redesigning the (standard) i.i.d. Gaussian sensing matrix to reflect the intra group dependencies, significant improvements in image reconstruction are possible.

The rest of the paper is organized as follows: in section 2 we formulate the problem. We then propose our method in section 3. We outline our experiments and present our findings in section 4. We conclude our paper in section 5 and suggest avenues for further research.

2. PROBLEM FORMULATION

Consider the standard compressive sensing framework of images,

$$\begin{aligned} y &= A\theta + \eta \\ &= AW^{-1}x + \eta \end{aligned}$$

where $y \in \mathbb{R}^m$ is a vector of measurements, $A \in \mathbb{R}^{m \times n}$ is the sensing matrix, with $m < n$. W is the DWT matrix, and $\theta \in \mathbb{R}^n$ is the image (vectorized). x is the DWT coefficients of the image, which is known to be (approximately) sparse. $\eta \in \mathbb{R}^m$ is an i.i.d. Gaussian noise vector of zero mean and unit variance.

As in [8], we assume that the wavelet coefficients x can be grouped into parent-child pairs (Fig. 1), and solve the overlapping (latent) group lasso program:

$$\widehat{x}_{Glasso} = \arg\min_{x} \frac{1}{2}\|y - A_{iid}W^{-1}x\|^2 + \lambda_g \Omega(x) \quad (1)$$

where $\Omega(x)$ is the latent group lasso penalty [6]. For a set of groups $G_1, G_2, \ldots, G_M$, the latent group lasso penalty is defined by

$$\Omega(x) = \inf_{\sum_i v_i = x} \sum_{i=1}^{M} \|v_i\|$$

where $v_i \in \mathbb{R}^n$ has support restricted to the indices in group G_i. A_{iid} is the standard gaussian i.i.d. sensing matrix used for compressed sensing. The latent group lasso penalty has the property that the pattern of non zeros (coefficients) recovered by solving (1) can be expressed as a union of groups.

We assume the following signal model for x. Assume that the parent child groups are given by $\mathcal{G} = \{G_i\}_{i=1}^M$. We assume a mixture model, where a group can be active with probability p. Since the signal is group-sparse, p is small ($p << 1$). If a group is active, we assume that the coefficients arise from a distribution with covariance matrix Σ. The matrix Σ encodes the intra group dependencies between variables. Since we are considering only parent-child pairs,

This work was supported by DARPA/ONR grant N66001-10-1-4090.

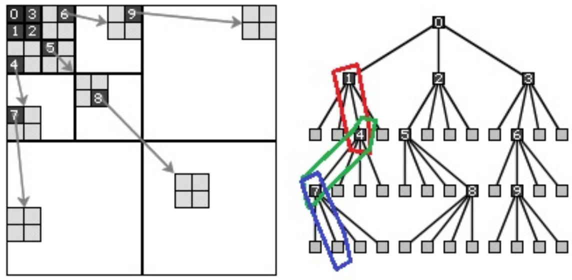

Fig. 1. Quadtree corresponding to the 2-d DWT. The parent-child pairs can be grouped together to reflect inter scale dependencies

$\Sigma \in \mathbb{R}^{2\times 2}$, and the off-diagonal elements will represent the cross-correlation between the parent and child, for each group as shown in Fig 1.

Our goal in this paper is to learn the covariance matrix coresponding to the parent child pair from a training set of images, and use this as prior information to design new sensing matrices that are better matched to the image structure. In doing so, the sensing energy is better aligned to the image structure, and consequently active groups can be reliably identified using fewer measurements than needed with the conventional sensing matrix composed of i.i.d. zero-mean Gaussian variables.

An important point to note is that since the covariance matrix that we aim to learn is only 2×2, we do not need as many samples as one might need to accurately learn the entire $n \times n$ covariance matrix corresponding to the images.

3. MEASUREMENT MATRIX DESIGN

A group sparse signal can be written as a sum of signals whose support is restricted to be the indices corresponding to a single group. For example, the signal s comprising of three groups in Fig. 2 can be decomposed using latent variables [6] into signals s_1, s_2 and s_3

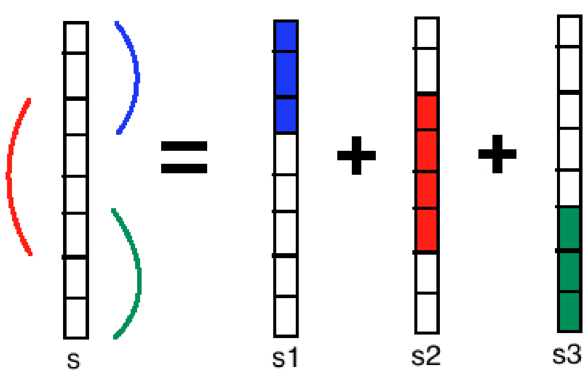

Fig. 2. Decomposition of the group sparse signal in latent group lasso. The curved lines in the LHS represent groups of coefficients. (best seen in color)

The latent groups can be active/inactive independent of each other, and the final sparse signal is the sum (normalized) of the decomposition. It can then be seen that, if the groups G_i have covariance Σ_i, then the final (sum) vector will have a covariance matrix that can be decomposed as in Fig. 3, where C is the matrix after combining the individual covariance matrices Σ_i (shown shaded in Fig 3).

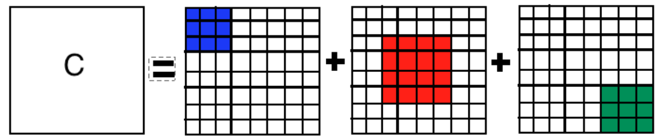

Fig. 3. Decomposition of the covariance matrix of the signal in Fig 2. The shaded parts in the RHS if the figure correspond to covariance matrices of the individual active groups, $\Sigma_i, i = 1, 2, 3$.

Based on this insight, we form measurement vectors (the rows of the measurement matrix): Assume we are given the covariance matrices $\Sigma_i = \Sigma$. We assume all the covariance matrices are the same, for simplicity. We note later in the paper that one can assume the covariance matrices to be different across scales, and still apply our method. Let $a^i \in \mathbb{R}^n$ denote the i-th row of the sensing matrix under construction. We start with $a^i = 0$. Letting a_{G_i} be the sub-vector of a^i indexed by group G_i, we recursively perform the following operation $\forall G_i \in \mathcal{G}$:

$$a_{G_i} = a_{G_i} + v_{G_i} \ , \tag{2}$$

where

$$v_{G_i} \sim \mathcal{N}(0, \Sigma) \ .$$

This procedure yields $a^i \sim \mathcal{N}(0, C)$, where C is the covariance matrix obtained as a result of the composition of the vectors, as in Fig. 3. The measurement matrix is generated by repeating this procedure m times. We then normalize the columns to have unit norm, and call this matrix A_{avg}, distinguishing it from the standard gaussian sensing matrix used for compressed sensing, which we denote by A_{iid}. The columns of A_{iid} are also normalized, so that both A_{avg} and A_{iid} have Frobenius norm n. This puts the two sensing matrices on equal footing, as far as SNR is concerned. Hence, we now solve the group lasso as in (1), but with A_{avg}:

$$\widehat{x}_{CGlasso} = \arg\min_x \ \frac{1}{2}\|y - A_{avg}W^{-1}x\|^2 + \lambda_a\Omega(x) \tag{3}$$

The intuition for this approach arises from work on correlated gaussian designs, as in [9]. In [9], it was shown that gaussian matrices with non *iid* columns also obey the restricted eigenvalue conditions [1] needed for exact recovery in the noiseless setting, and robust recovery in the noisy case. In our case, the sparse vector to be reconstructed itself has correlated entries. To take into account this correlation, we allow the columns in the measurement matrix (called predictor variables in sparse linear regression settings) to be correlated. Hence, selection of any variable will automatically force the selection of a correlated variable, when we use the latent group lasso. We refrain from getting into the theoretical details due to space constraints.

It is well known that to accurately capture a signal, one should measure along the direction of the signal. The matched filter is a typical example that makes use of this principle. By generating sensing vectors from (roughly) the same distribution as the data itself, we ensure that the measurement is correlated with the signal itself, facilitating better recovery. Note that in this case, the “data” corresponds to coefficients from a single group.

4. EXPERIMENTS AND RESULTS

We refer to the method in [8] as Glasso and the method we developed by CGlasso, the ‘C’ indicating the use of the covariance matrix in our design. For details on how the groups are designed, and the Glasso method, we refer the reader to [8]. We solve the latent group lasso problem by the replication strategy elaborated in [3]. We use SpaRSA [11] to solve the optimization problems.

To show the efficacy of our method, we first consider a toy signal, with a known covariance matrix. Note here that the covariance matrix refers to the 2×2 matrix corresponding to the parent child pairs on the DWT tree. We assume a signal consisting of 100 non-overlapping groups of size 10 each, of which 10 are active, and consider 250 measurements. The signals have active groups whose coefficients are generated from a zero mean Gaussian with covariance matrix $U^T U$, where $U \in \mathbb{R}^{10 \times 10}$ is an orthonormal basis for a 10×10 random gaussian matrix. It can be seen from Fig. 4 that CGlasso outperforms Glasso

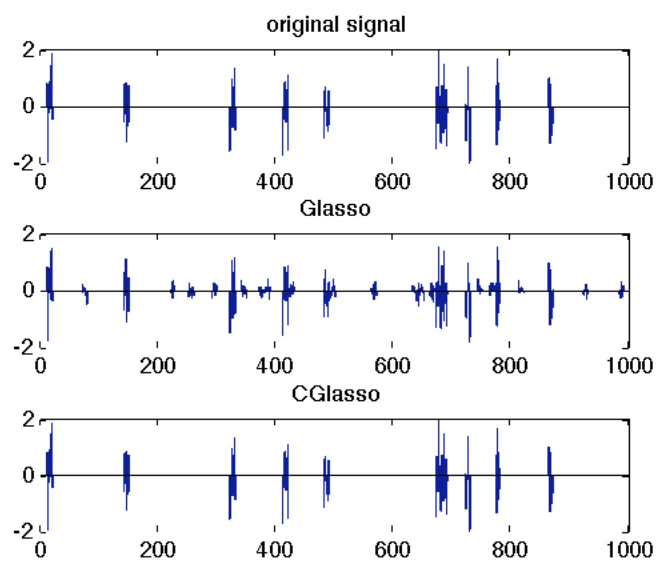

Fig. 4. Comparison of the two methods. It can be seen the CGlasso recovers the signal exactly, while Glasso makes some errors, for the given number of measurements.

We tested our methods for noiseless image recovery using the Microsoft Research Object Class Recognition database[1]. The dataset consists of images categorized into 20 types, with roughly 30 images of each type. We used the first 10 images of each type to generate a training set of 200 images, which was used to learn both the regularization parameters λ_g and λ_a, as well as the covariance matrix Σ. For each image in the training set, we computed its DWT, and isolated the groups that were active. We then computed the unbiased sample covariance matrix (4) from these groups. Letting x^i be the DWT of the i^{th} training image, we used the estimate

$$\widehat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{n} \sum_{k} (x^i_{G_k})^T x^i_{G_k}, \quad (4)$$

where G_k is an active group (non-zero norm) in the DWT of the training set image. To compare our results with other methods tested in [12] (Fig. 6)[2], we compute the Normalized Mean Square Error (NMSE) of our method. The normalized mean square error (in dB) for the true image x is given by $10 \times \log \left(\frac{\|\widehat{x}-x\|^2}{\|x\|^2} \right)$. We resized the images to size 128×128, and obtained 5000 measurements for each image.

We also compared group lasso [8] to our new method, with noisy measurements. Table 1 demonstrates that CGlasso outperforms Glasso under this scenario as well. We again considered the same toy signals as the ones explained with reference to Fig. 4. The signals were of length 1000, and we considered 250 measurements. The results are averaged over 100 tests.

Noise Std. Dev	MSE CGlasso	MSE Glasso
0	$\approx 10^{-31}$	0.0200
0.02	0.0001	0.0225
0.04	0.0003	0.0294
0.06	0.0006	0.0355
0.08	0.0011	0.0497
0.1	0.0019	0.0556

Table 1. Comparison of the two methods under noisy measurements.

Finally, we show that we need fewer measurements for CGlasso to achieve an acceptable recovery as compared to Glasso. Fig. 6 indicates this. We considered toy piecewise constant signals of length 1024, where the locations of the “pieces” were chosen uniformly at random. The magnitude of the blocks in the signals was also uniformly selected to lie between $[-1, +1]$. We used 2500 independent signals to learn λ_a and λ_g and compute $\widehat{\Sigma}$ as in (4). Using these parameters, we tested the two methods over 100 other test signals, and averaged the results per measurement size.

5. CONCLUSIONS AND FUTURE WORK

We proposed a new method to design measurement matrices, based on the inter group relationships between the wavelet coefficients. The new measurement matrix yields results better than traditional group lasso methods, and also outperforms many other existing methods for compressive imaging. Considering that we “adapt” the sensing matrix to the correlations observed within groups of coefficients in the signal, in some sense our method can be called adaptive. We make it clear that we do not update the sensing matrix on the fly, and once the covariance matrix is learned, we fix a sensing matrix. Also, the covariance matrix os of a very low dimension (just 2×2), and hence the amount of data we need to learn the covariance matrix is much less as compared to the data needed to learn the full covariance matrix of the data.

Future work would include theoretically justifying the use of such matrices, as well as integrating the learning of the covariance struc-

[1]http://research.microsoft.com/en-us/projects/ObjectClassRecognition

[2]The authors thank Subhojit Som and Phil Schniter for sharing data for Fig. 5

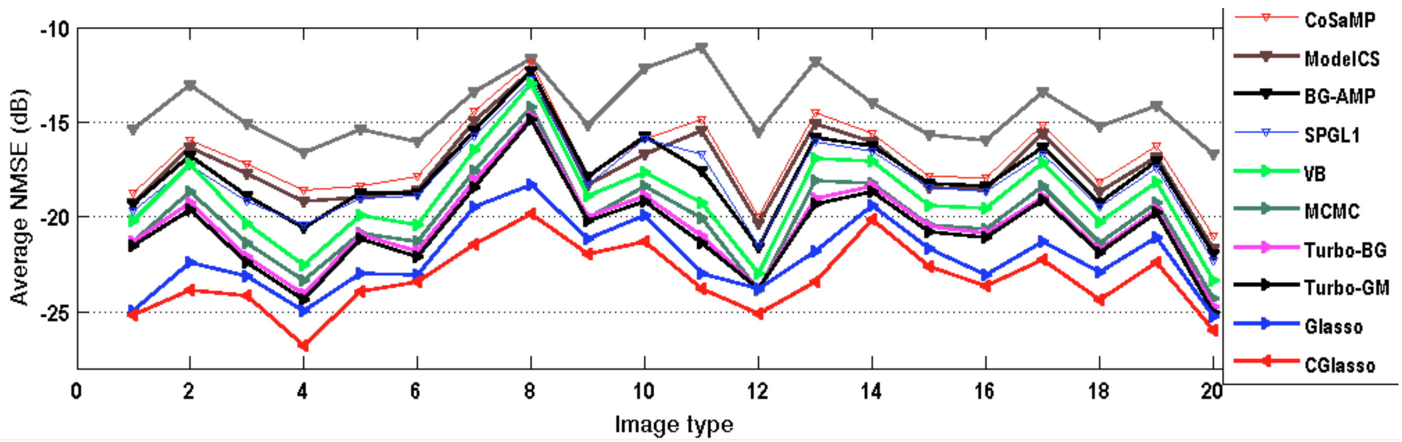

Fig. 5. Comparison of various methods. Note that the group lasso (solid blue curve) outperforms the others on the dataset, and CGlasso (solid red curve) performs marginally better.

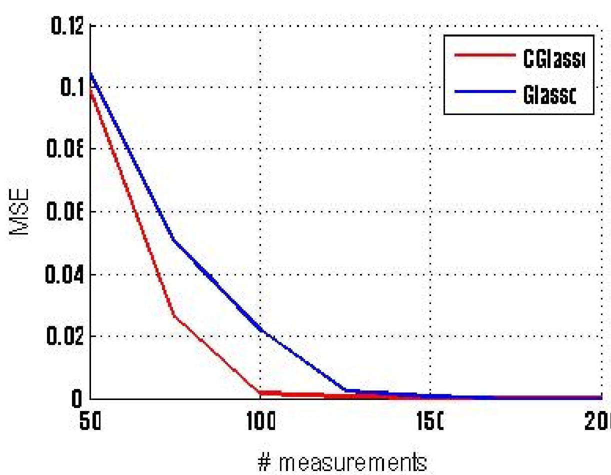

Fig. 6. Number of measurements needed to reconstruct toy signals

ture for the measurement matrix in an online fashion. Also, the covariance matrices may be scale dependent, in which case it would be prudent to learn a different matrix for each scale. Adapting our method to include much higher order dependencies among DWT coefficients could potentially yield better results.

6. REFERENCES

[1] P. Bickel, Y. Ritov, and A. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, 37(4):1705–1732, 2009.

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

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

[5] 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.

[6] G. Obozinski, L. Jacob, and J. P. Vert. Group lasso with overlaps: the latent group lasso approach. *preprint: arXiv:1110.0413v1 [stat.ML]*, 2011.

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

[8] N. Rao, R. Nowak, S. Wright, and N. Kingsbury. Convex approaches to model wavelet sparsity patterns. *International Conference on Image Processing*, 2011.

[9] G. Raskutti, M. Wainwright, and B. Yu. Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research*, 11:2241–2259, 2010.

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

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

[12] S. Som and P. Schniter. Compressive imaging using approximate message passing and a markov-tree prior. *preprint arxiv:1108.2632*, 2011.