ADAPTIVE SENSING WITH STRUCTURED SPARSITY

Nikhil Rao   *Gongguo Tang*   *Robert Nowak*

Department of Electrical and Computer Engineering
University of Wisconsin - Madison

ABSTRACT

Adaptive sensing strategies have been proven to outperform traditional (non adaptive) compressed sensing, in terms of the signal to noise ratios that can be handled, and/or the number of measurements needed to accurately recover a signal of interest. Most existing adaptive sensing schemes for sparse signals, while work well in practice, do not take into account potential structure present in the sparsity pattern of the signal. In this paper, we focus on the Markov tree structure inherent in the wavelet coefficients of signals, and propose an adaptive sampling technique to recover the same. We adopt a simple "follow the scent" strategy, and show that it outperforms traditional non adaptive techniques in practice.

***Index Terms*—** Adaptive Algorithm, Compressed Sensing, Probability Propagation, Hidden Markov Models

1. INTRODUCTION

Adaptive sensing methods for sparse signals [1, 2] have been proven to outperform traditional compressed sensing [3] in the sense that in the presence of noise, one can detect much weaker signals, and/or one needs fewer measurements to estimate a given signal. However, in many cases, the sparse signal to be detected has some inherent structure present in the sparsity pattern, which most adaptive sampling methods fail to take into account. Given the vast amounts of literature that has been dedicated to non adaptive recovery of structured sparse signals (see [4, 5] and references therein) and the gains that have been proven to achieve, it is only natural to ask if exploiting structure can help in improving adaptive sampling schemes as well.

In this paper, we aim to partially answer that query. Specifically, we focus on adaptive recovery of wavelet transform coefficients of a signal, that can be modeled as lying on a tree. In fact, statistical dependencies among wavelet transform coefficients can be very accurately modeled using Hidden Markov Trees (HMTs) [6, 7]. Moreover, very efficient methods exist for performing inference on such HMT's [8]. In inverse problems (compressed sensing, tomography, etc), HMT's have been used in loopy belief propagation [9], iterative reweighing schemes [10], greedy methods [11] and that based on Approximate Message Passing [12]. In [13], the authors proposed a grouping scheme inspired by HMTs for the coefficients to recover the signal in a convex manner, by solving a group lasso type algorithm.

1.1. Motivating Applications

Our work finds motivation in the context of dynamic imaging in wavelet encoded MRI [14, 15]. The goal here is to acquire images in a sequential manner. Assuming that one has an estimate of the DWT coefficients of an image at some time t, one needs to efficiently acquire measurements of an image at time $t + \Delta t$. Using the change in the estimate of the coefficients as a guide, we can focus our sensing energy in particular regions of interest, leading to a more efficient method to acquire images in a dynamic fashion (see [16] and references therein).

Digital Micromirror Devices (DMD's) can be used to directly acquire wavelet coefficients of signals. This technique has already been used to sample wavelet coefficients in hyperspectral imaging [17, 18]. Long acquisition times in hyperspectral imaging imply that it cannot be used to capture moving data. One way around the problem is to sequentially obtain compressive measurements. Such schemes (both adaptive and nonadaptive) have been used with varying success in medical imaging, geosensing, and object recognition and tracking. Adaptive imaging schemes have also been studied in fluorescence microscopy [19], wherein concepts similar to hyperspectral imaging are applied.

1.2. Model

Letting $\boldsymbol{\theta} \in \mathbb{R}^p$ be the signal of interest, we propose an algorithm to choose wavelet sensing waveforms $\boldsymbol{w}_l$ so that we obtain measurements of the form:

$$\boldsymbol{y}_l = \langle \boldsymbol{w}_l, \boldsymbol{\theta} \rangle + \boldsymbol{n} \qquad \boldsymbol{n} \sim \mathcal{N}(0, \sigma^2)$$

where $\langle \boldsymbol{w}_l, \boldsymbol{\theta} \rangle$ corresponds to a particular wavelet coefficient. We assume that we can obtain the samples $\boldsymbol{y}_l$ *adaptively*, in that $\boldsymbol{w}_l$ can be chosen in a sequential manner depending on past observations.

We leverage the additional knowledge that the sparsity pattern of the wavelet coefficients of the signal follows a Markov tree structure to design novel adaptive sampling techniques. We introduce a simple "follow the scent" procedure that starts from the root of the tree, predicts the states of the unobserved nodes (and estimates those of the observed ones), and proceeds to sample the unsampled node that has the highest probability of being active.

Although we focus on the wavelet coefficients of signals in this paper, it is important to note that our method will work for any signal whose sparsity pattern follows a Markov tree structure. This includes applications in genomics, imaging, disease propagation, etc. Our method also applies to all previously studied methods wherein the sparsity pattern was forced to lie on a rooted tree.

Past work has proposed sampling schemes that take into account this structure [20, 21, 16, 22]. However, a key difference between our method and other contemporary work lies in the modeling of the coefficients: past work assume that the sparsity pattern of the DWT coefficients is a rooted tree, which as we argue below is not always the case. The authors of [23] attempt to circumvent the problem by including a dictionary learning stage that forces tree structured sparsity patterns, but that adds extra complexity in the algorithm,

The authors wish to acknowledge NSF grant CCF-1218189 and the DARPA KECoM program

and as is the case with all dictionary learning methods, provides no convergence guarantees to a global optimum.

1.3. Non Rootedness of DWT Coefficients

Using an illustrative example, we show here that the DWT coefficients of signals do not always lie in a rooted tree. Consider the standard "blocks" signal, and its Haar DWT coefficients (Fig 1). It is possible that at a certain level j and location k, the sum of the signal values corresponding to the positive part of the Haar basis vector cancels with that of the negative part, leading to a small (or zero) wavelet coefficient. However, as we move to finer scales, the variations in the signal may mimic the variations in the Haar wavelet support, resulting in large values of the corresponding coefficients, as can be seen in Fig. 2.

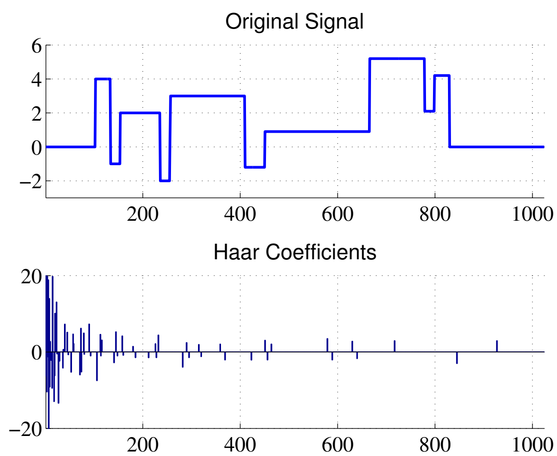

Fig. 1. 'Blocks' and its Haar DWT

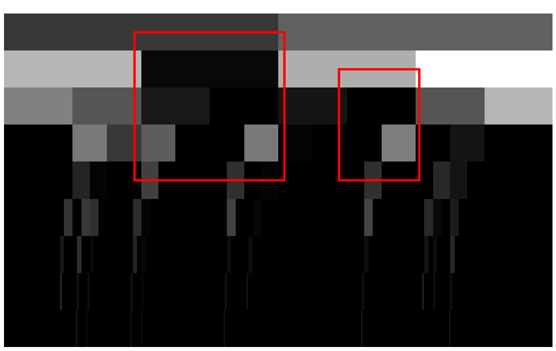

Fig. 2. (Best seen in color) Haar coefficeints in Fig. 1 arranged in a tree. Darker regions correspond to smaller magnitude coefficients. We see that wavelet coefficients can be small (or zero) and still have non zero children, as denoted by the red rectangles.

The rest of the paper is organized as follows: in section 2, we introduce our algorithm to adaptively obtain samples from a hidden markov tree model. In section 3, we perform experiments and report results. We conclude our paper in section 4 and propose future work.

2. ALGORITHM

In this section, we introduce our method to sample DWT coefficients in an adaptive fashion. We first dispense with notations. We denote the wavelet coefficients of a signal by $\boldsymbol{x} \in \mathbb{R}^p$ with, p a power of 2 for simplicity. The coefficients can be arranged as lying on a tree of depth $J = \log_2(p)$. Lowercase letters (j,k) index the node of the wavelet tree at scale $j = 0, 1, \ldots, J-1$ and location $k = 0, 1, \ldots, 2^j - 1$ for the dyadic tree corresponding to the 1D DWT. We focus on the 1D case in this paper to keep explanations simple, but the method can be easily modified to apply to the 2D case. Also, we use the MATLAB notation $(0 : j_0, :)$ to indicate all the nodes from level 0 to j_0 $\forall k = \{0, 1, \ldots, 2^j - 1\}$

Let $\boldsymbol{x}_{(j,k)} = \langle \boldsymbol{w}_{(j,k)}, \boldsymbol{\theta} \rangle$. We assume we have access to point samples of the form $\boldsymbol{y}_{(j,k)} = \boldsymbol{x}_{(j,k)} + \boldsymbol{n}_{(j,k)}$. We model the DWT coefficients $\boldsymbol{x}$ of a signal using a Hidden Markov Tree model [6]. We assume that each (hidden) node in the tree may take one of two states, indicating whether the corresponding DWT coefficient is small (0) or large (1). Letting $\boldsymbol{s}$ be the vector of state variables, we assume the following simple model for the parameters:

$$\begin{aligned} &\mathbb{P}(\boldsymbol{s}_{(j,k)} = 1 | \boldsymbol{s}_{(j-1, \lfloor \frac{k}{2} \rfloor)} = 1) = \gamma \\ &\mathbb{P}(\boldsymbol{s}_{(j,k)} = 0 | \boldsymbol{s}_{(j-1, \lfloor \frac{k}{2} \rfloor)} = 0) = \delta \\ &\mathbb{P}(\boldsymbol{s}_{(0,0)} = 1) = \rho_1 \end{aligned} \tag{1}$$

We model each wavelet coefficient as a 2-stage Gaussian mixture:

$$\begin{aligned} \mathbb{P}(\boldsymbol{x}_{(j,k)} | \boldsymbol{s}_{(j,k)}) = (1 - \boldsymbol{s}_{(j,k)}) &\mathcal{N}(0, \tau_0^2) \\ + \quad \boldsymbol{s}_{(j,k)} &\mathcal{N}(0, \tau_1^2) \end{aligned} \tag{2}$$

Where $\boldsymbol{s}_{(j,k)}$ is a hidden state variable that can be either active or inactive, and $0 \approx \tau_0 \ll \tau_1$. This ensures that when the state is $0(1)$, the corresponding DWT coefficient will be small (large).

To perform inference on the HMT, we make use of the upward-downward (UD) algorithm [6, 8], which we provide here for the sake of completion. We let $L_{(j,k)}(m)$ be the likelihood of node (j,k) being in state m, given observation $\boldsymbol{y}_{(j,k)}$. For the exact formulae for the likelihood, we refer the reader to [24]. We define the transition probabilities:

$$\rho_{(j,k)}(m|n) = \mathbb{P}(\boldsymbol{s}_{(j,k)} = m | \boldsymbol{s}_{(j-1, \lfloor \frac{k}{2} \rfloor)} = n)$$

The quantities $p_{(j,k)}(m)$ (Algorithm 1) are the final posterior (unnormalized) state probabilities of the node (j,k) being in state m.

The UD algorithm computes $\mathbb{P}(\boldsymbol{s}|\boldsymbol{y})$ for the entire HMT and can be seen as Belief Propagation [25] adapted to our specific case. For those nodes in the tree where we have not made any observations, $\mathbb{P}(\boldsymbol{s}|\boldsymbol{y})$ gives a prediction of the state of the node. We define

$$\hat{\boldsymbol{p}}_{(j,k)} = \mathbb{P}(\boldsymbol{s}_{(j,k)} = 1 | \boldsymbol{y}) := \frac{p_{(j,k)}(1)}{p_{(j,k)}(1) + p_{(j,k)}(0)},$$

which gives us the probability that a node is active, given all the observed measurements $\boldsymbol{y}$. When all the nodes in the tree are measured, as is the case for uncompressed non adaptive sensing of signals, $\hat{\boldsymbol{p}}$ gives the exact probability of a node being active. In our case, we obtain only partial measurements, and use $\hat{\boldsymbol{p}}$ to guide our algorithm to make future measurements.

$$q_{(j,k)}(m) = \sum_{i=0}^{1} q_{(j+1,2k)}(i) q_{(j+1,2k+1)}(i) \rho_{(j,k)}(i|m) L_{(j,k)}(i) \tag{3}$$

$$p_{(j,k)}(m) = \sum_{i=0}^{1} \frac{p_{(j-1,\lfloor \frac{k}{2} \rfloor)}(i) \rho_{(j,k)}(m|i) q_{(j+1,2k)}(m) q_{(j+1,2k+1)}(m) L_{(j,k)}(m)}{q_{(j,k)}(i)} \tag{4}$$

Algorithm 1 Upward-Downward Algorithm for Markov Trees

1: **Inputs:** HMT Parameters $\delta, \gamma, \rho_1, \tau_0, \tau_1$
2: **UP STEP**
3: at $j = J - 1$, set
 $q_{(j,k)}(m) = \sum_{i=0}^{1} \rho_{(j,k)}(i|m) L_{(j,k)}(i)$
4: **for** $J - 2 \leq j \leq 1$ **do**
5: set $q_{(j,k)}(m)$ according to (3)
6: **end for**
7: at $j = 0$, set
 $q_{(0,0)}(m) = q_{1,0)}(m) q_{(1,1)}(m) \rho_{(0,0)}(m) L_{(0,0)}(m)$
8: **DOWN STEP**
9: at $j = 0$ set $p_{(0,0)}(m) = q_{(0,0)}(m)$
10: **for** $1 \leq j \leq J - 2$ **do**
11: set $p_{(j,k)}(m)$ according to (4)
12: **end for**
13: at $j = J - 1$ set
 $p_{(j,k)}(m) = \sum_{i=0}^{1} \frac{p_{(j-1,\lfloor \frac{k}{2} \rfloor)}(i) \rho_{(j,k)}(m|i) L_{(j,k)}(m)}{q_{(j,k)}(i)}$

Algorithm 2 Adaptive sensing on Markov Trees

1: **Inputs:** HMT Parameters $\delta, \gamma, \rho_1, \tau_0, \tau_1$, Budget R, initial level j_0, sampled set $S = \{\}$
2: **Initialize** $S = (0 : j_0, :)$, $l = |S|, \boldsymbol{y} = \boldsymbol{x}_S + \boldsymbol{n}$
3: **while** $l \leq R$ **do**
4: Estimate $\hat{\boldsymbol{p}}$ using the **upward-downward** algorithm with observations $\boldsymbol{y}$ and parameters $\delta, \gamma, \rho_1, \tau_0, \tau_1$
5: Find $(\hat{j}, \hat{k}) : \arg\max_{(j,k) \notin S} \hat{\boldsymbol{p}}$
6: In case of a tie, pick a node using any strategy of choice
7: Update samples $\boldsymbol{y} \leftarrow \boldsymbol{y} \cup \boldsymbol{y}_{(\hat{j},\hat{k})}$
8: Update sampled set $S \leftarrow S \cup \{(\hat{j}, \hat{k})\}$
9: $l \leftarrow l + 1$
10: **end while**

2.1. Remarks

In Algorithm 2 the parameter j_0 controls the initial level till which we obtain all the samples. Usually, $j_0 \approx \frac{J}{2}$ [26]. The algorithm then sequentially obtains noisy samples corresponding to certain nodes of the DWT tree, and performs an *exact* inference procedure on the observed nodes, and imputes the probabilities of unobserved nodes being active. We then merely pick the node that has the highest probability of being active. The marginal and conditional probabilities at each scale in the HMT can be computed in parallel, making the procedure highly efficient. Moreover, from (3) and Algorithm 2, it can be seen that $q_{(j,k)} = 1 \ \ \forall (j,k) \notin S$. Hence, we can start the up-step of the UD algorithm only from the leaves of the sampled set S, gaining more efficiency.

To have a more realistic version of the DWT tree model, one can replace $\delta, \gamma, \rho_1, \tau_0, \tau_1$ by scale-dependent quantities. We refrain from doing so in Algorithm 2 to avoid clutter of notations. In the image processing setting, we note that [7] estimated parameters for a "universal" HMT model that we can directly plug-in to our method.

We emphasize that our focus in this paper is the algorithm to sample adaptively on the tree, the implementation of which does not depend on the specific state transition probabilities (δ, γ, ρ_1), or variances (τ_0, τ_1). In the experiments we perform, we learn the Markov transition probabilities on a scale dependent basis, or use the universal HMT parameters [7].

2.2. Analysis

Consider a simplified model, where $\delta = 1$, i.e. we assume that the sparsity pattern forms a rooted tree. In this case, letting k be the sparsity of the signal, we have the following result:

Proposition 2.1 *Consider a k sparse signal, whose sparsity pattern lies on a rooted binary tree ($\delta = 1$). Suppose that the signal to noise ratio per measurement is sufficiently large [1] (5) so as to allow for accurate recovery. Then, sensing $R = 2k + 1$ nodes is sufficient to recover the sparsity pattern of the signal with high probability.*

Although the results are similar to what has been studied previously in the case of rooted trees, our method has the added flexibility of being useful for more general Markovian tree structures. We only provide a sketch of the proof, and omit the details due to space constraints.

Proof Sketch Suppose $j_0 = 0$, meaning we only measure the root node. Then, we obtain $\hat{\boldsymbol{p}}_{(0,0)}$. We will then proceed to measure one of its children, since by the upward-downward algorithm we will have

$$\hat{\boldsymbol{p}}_{(1,k)} \geq \hat{\boldsymbol{p}}_{(j,k)} \qquad \forall j > 1$$

Since the sparsity pattern is rooted, we estimate atleast one of the children to be active. Also, from the upward downward algorithm, we see that, if $j - 1 \in S$ and $j \notin S$,

$$\hat{\boldsymbol{p}}_{(j,k)} = \gamma \hat{\boldsymbol{p}}_{j-1,\lfloor \frac{k}{2} \rfloor}$$

Hence, for some $(j1 - 1, \lfloor \frac{k1}{2} \rfloor)$ and $(j2 - 1, \lfloor \frac{k2}{2} \rfloor) \in S$, and $(j1, k1), (j2, k2) \notin S$

$$\hat{\boldsymbol{p}}_{(j1-1,\lfloor \frac{k1}{2} \rfloor)} > \hat{\boldsymbol{p}}_{(j2-1,\lfloor \frac{k2}{2} \rfloor)} \Rightarrow \hat{\boldsymbol{p}}_{(j1,k1)} > \hat{\boldsymbol{p}}_{(j2,k2)}$$

This means that we will sample children of nodes that have been deemed to be active before sampling children of nodes estimated to be inactive.

Note that, to correctly identify a node $(j,k) \in S$ to be active, we need $\hat{\boldsymbol{p}}_{(j,k)} \geq 0.5$. This would require

$$\boldsymbol{y}_{(j,k)}^2 \geq \frac{2\eta(\tau_1^2 + \sigma^2)(\tau_0^2 + \sigma^2)}{\tau_1^2 - \tau_0^2} \tag{5}$$

where η depends on $\log(\tau_1)$ and γ.

Hence, following this strategy, we will measure all the k active nodes (rooted at $(0,0)$) and its children. The result then follows from arguments similar to that in [20]

[1] We omit the exact values and associated proofs due to space constraints

3. EXPERIMENTS AND RESULTS

We compared our method to the LASSO [27], using [28] to solve the problem. We generated piecewise constant signals, of the form considered in Fig. 1 and 2 of length $p = 1024$, having 20 "pieces" at randomly chosen locations. The magnitude of the pieces was chosen to lie randomly in $[-1, 1]$. We varied the number of measurements from 40 to 280 in steps of 40, and repeated each test 100 times. We learned the (scale dependent) parameters $\gamma(j), \delta(j), \tau_{(0,1)}, \rho_1$ using a test set of size 10000. Fig 3 shows the MSE as we increase the number of measurements taken, for two values of AWGN standard deviation. j_0 was set to 5. For the LASSO, we considered a Gaussian measurement matrix with unit norm rows, and pick a regularization parameter that works best from a fixed grid. Such a clairvoyant scheme is not possible in practice, since we do not have access to the true signal.

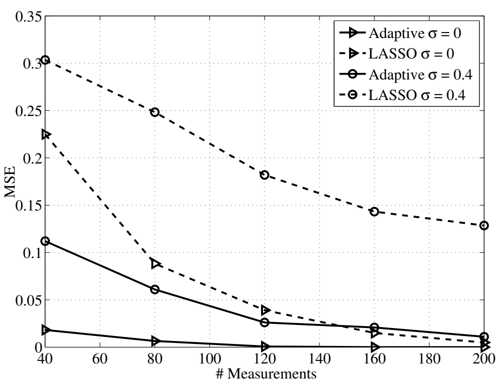

Fig. 3. Comparison to the LASSO. We see that we obtain superior performance to standard compressed sensing, with higher gains when there is noise present in the measurements.

We considerd a noisy $\sigma = 0.02$ version of the normalized 64×64 section of the cameraman image, and fixed our budget to be 2100, 700 for each quad-tree in the 2D-DWT (Fig 4). For the LASSO, we considered an i.i.d. Gaussian measurement matrix of size 2100×4096, and normalized it so that the rows were of unit norm. We see that our method outperforms the LASSO, and does marginally better than Adaptive Compressed Sensing [26] with threshold ($\tau = \sigma\sqrt{2\log(n)}$), and the Tree Structured Bayesian Compressed Sensing (Variational Bayes) [22]. The parameters for the HMT were taken to be the uHMT parameters from [7]. We set $j_0 = 3$.

Finally, we test Proposition 2.1. We consider noiseless measurements of piecewise constant signals of length 4096 with $\delta = 1, \tau_0 = 0.001, \tau_1 = 3$, and vary γ, and hence k, the number of non zeros in the signal. We fix the sampling budget to be $2k + 1$. Fig. 5 plots the Hamming distance between the supports of the estimate and the true signal, averaged over 100 runs. As before, we consider the same number of Gaussian measurements for the LASSO, with unit norm rows, and pick the regularization parameter from a grid.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an adaptive sampling technique that takes into account the parent-child dependencies of the DWT coefficients of a signal. We work with a more realistic model for the coefficients, and showed that our method outperforms standard non-adaptive compressed sensing methods, even with a very simplistic assumption on the HMT.

Although we focus on the wavelet coefficients of signals in this paper, it is important to note that our method will work for any signal whose sparsity pattern can be expressed as lying on a Markov tree. This involves applications in genomics, imaging, dictionary learning, etc. Our method also applies to all previously studied methods wherein the sparsity pattern was forced to lie on a rooted tree.

We propose to analyze our method in more detail in the future. Specifically, we aim to extend proposition 2.1 to general Markov tree models. We also aim to derive SNR bounds for the signal amplitude, for which our method will recover the signal accurately.

(a) LASSO PSNR = 19.83 dB

(b) ACS PSNR = 21.05 dB

(c) TS-BCS PSNR = 22.17 dB

(d) Markov PSNR = 22.73 dB

Fig. 4. Comparison of our method (Markov) to the LASSO for a section of the Cameraman image. We also compare our method to [26] (ACS), and [22] (TS-BCS).

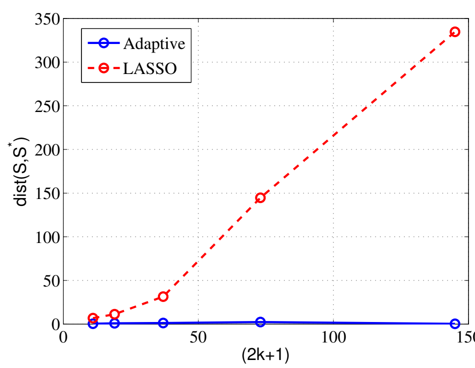

Fig. 5. We see that $2k + 1$ measurements suffice to recover the signal support adaptively, while the LASSO fares far worse. For the lasso to achieve comparable results, we add an extra factor of $\log(n) = \log(4096) \approx 9\times$ measurements. dist(S, S^*) denotes the Hamming distance between the true and recovered supports

5. REFERENCES

[1] M. Malloy and R. Nowak, "Near optimal adaptive sensing and group testing," *Asilomar Conference on Signals, Systems and Computers (to appear)*, 2012.

[2] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," *Asilomar Conference on Signals, Systems and Computers*, 2009.

[3] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, pp. 489–509, 2006.

[4] M. Duarte and Y. Eldar, "Structured compressed sensing: From theory to applications," *preprint arXiv:1106.6224*, 2011.

[5] N. Rao, B. Recht, and R. Nowak, "Universal measurement bounds for structured sparse signal recovery," *Artificial Intelligence and Statistics, Journal of Machine Learning Research*, no. 22, pp. 942–950, 2012.

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

[7] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree structured image modeling using wavelet domain hidden markov models," *Transactions on Image Processing*, March 2000.

[8] B. Frey, "Graphical models for machine learning and digital communications," *MIT Press, Cambridge, Mass.*, 1998.

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

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

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

[12] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," *IEEE transactions on signal processing*, 2011.

[13] N. Rao, R. Nowak, S. Wright, and N. Kingsbury, "Convex approaches to model wavelet sparsity patterns," *IEEE International Conference on Image Processing*, pp. 1917–1920, 2011.

[14] M. Guerquin-Kern, M. Haberlin, K. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, 2011.

[15] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, M. Breakspear *et al.*, "Wavelets and functional magnetic resonance imaging of the human brain," *Neuroimage*, vol. 23, no. Suppl 1, pp. S234–S249, 2004.

[16] L. Panych and F. Jolesz, "A dynamically adaptive imaging algorithm for wavelet-encoded mri," *Magnetic Resonance in Medicine*, vol. 32, pp. 738–748, 1994.

[17] Y. Pfeffer and M. Zibulevsky, "A micro-mirror array based system for compressive sensing of hyperspectral data," Technical report CS-2010-01, Department of Computer Science, Technion-Israel Institute of Technology, Haifa, Israel, Tech. Rep., 2010.

[18] F. MagalhÃŖes, M. Abolbashari, F. AraÃējo, M. Correia, and F. Farahi, "High-resolution hyperspectral single-pixel imaging system based on compressive sensing," *Optical Engineering*, vol. 51, no. 7, pp. 071 406–1, 2012.

[19] V. Studer, J. Bobin, M. Chahid, H. Mousavi, E. Candes, and M. Dahan, "Compressive fluorescence microscopy for biological and hyperspectral imaging," *Proceedings of the National Academy of Sciences*, vol. 109, no. 26, pp. E1679–E1687, 2012.

[20] A. Soni and J. Haupt, "Efficient adaptive compressed sensing using sparse hierarchical learned dictionaries," *Asilomar Conference on Signals, Systems and Computers*, 2011.

[21] A. Averbuch, S. Dekel, and S. Deutsch, "Adaptive compressed image sensing using dictionaries," *SIAM Journal of Imaging Sciences*, vol. 5, no. 1, pp. 57–89, 2012.

[22] L. He, H. Chen, and L. Carin, "Tree-structured compressive sensing with variational bayesian analysis," *Signal Processing Letters, IEEE*, vol. 17, no. 3, pp. 233–236, 2010.

[23] A. Soni and J. Haupt, "Learning sparse representations for adaptive compressive sensing," *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012.

[24] R. Nowak, "Multiscale hidden markov models for bayesian image analysis," *Bayesian Inference in Wavelet Based Models, Springer*, 1999.

[25] J. Pearl, "Fusion, propagation, and structuring in belief networks," *Artificial intelligence*, vol. 29, no. 3, pp. 241–288, 1986.

[26] S. Deutsch, A. Averbush, S. Dekel *et al.*, "Adaptive compressed image sensing based on wavelet modeling and direct sampling," in *SAMPTA'09, International Conference on Sampling Theory and Applications*, 2009.

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

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