Compressive Blind Source Separation

Yiyue Wu, Yuejie Chi and Robert Calderbank
Department of Electrical Engineering
Princeton University, Princeton, NJ 08544, USA
Email: {yiyuewu,ychi,calderbk}@princeton.edu

***Abstract*—The central goal of compressive sensing is to reconstruct a signal that is sparse or compressible in some basis using very few measurements. However reconstruction is often not the ultimate goal and it is of considerable interest to be able to deduce attributes of the signal from the measurements without explicitly reconstructing the full signal. This paper solves the blind source separation problem not in the high dimensional data domain, but in the low dimensional measurement domain. It develops a Bayesian inference framework that integrates hidden Markov models for sources with compressive measurement. Posterior probabilities are calculated using a Markov Chain Monte Carlo (MCMC) algorithm. Simulation results are provided for one-dimensional signals and for two-dimensional images, where hidden Markov tree models of the wavelet coefficients are considered. The integrated Bayesian framework is shown to outperform standard approaches where the mixtures are separated in the data domain.**

I. INTRODUCTION

Compressive sensing (CS) is a new approach to data acquisition first proposed by Candés, Romberg and Tao [1] and Donoho [2]. It has the potential to greatly reduce the number of samples required by the Shannon/Nyquist sampling theorem for signals that are sparse or compressable with respect to some basis. This is a very attractive feature for applications where the cost of data acquisition is very high.

Compressive sensing is a natural fit to Blind Source Separation (BSS) where the aim is to separate a mixture of sources with little knowledge of the source signals or the mixing process. The source signals are assumed to be independent of or uncorrelated with each other. Many important problems in speech recognition, network anomaly detection, and medical signal processing can be viewed as BSS problems. The standard approach is Independent Component Analysis (ICA) which lacks resilience to noise and fails to take advantage of source correlation. A more recent approach [3] to BSS of images employs Bayesian methods in combination with hidden Markov Tree Models (HMT) in the wavelet domain.

This paper transfers the Bayesian framework for source separation to the compressive measurement domain. We consider the blind source separation problem directly from the compressed mixtures obtained from compressive sensing measurements, instead of recovering the mixtures in the data domain at first. We propose a unified Bayesian inference framework for the problem and the hidden Markov tree models of the wavelet coefficients are considered. We also propose a Markov chain Monte Carlo algorithm in order to calculate the posterior probability and recover the separated sources in the data domain. We provide simulations for both one-dimensional signals and two-dimensional images. Our proposed algorithm provides effective recoveries of original signals and experimental results show that our unified Bayesian inference method outperforms the separate procedure, which firstly recovers the mixtures in the data domain and secondly separates sources in the data domain.

The following of this paper is organized as follows: Section II formulates the problem after a brief review of compressive sensing and blind source separation. Section III describes the integrated Bayesian framework, and in particular MCMC inference using the Gibbs sampler. Section IV presents simulation results. This paper concludes in Section V.

II. COMPRESSIVE SENSING AND BLIND SOURCE SEPARATION

A. Background on Compressive Sensing

Let $\mathbf{s} \in \mathbb{R}^N$ be a signal and let the matrix $\Psi = [\psi_1, \psi_2, \ldots, \psi_N]$ be a basis in $\mathbb{R}^{N\times N}$ such that $\mathbf{s}$ can be expressed as $\mathbf{s} = \Psi\theta$. We say that $\mathbf{s}$ is K-sparse if $||\theta||_{\ell_0} = K$. The signal $\mathbf{s}$ is not measured directly; rather we measure linear projections $\mathbf{x} = \Pi\mathbf{s} = \Pi\Psi\theta = \Phi\theta$ using an $M \times N$ matrix Π, where the number of projections $M \ll N$. The matrix $\Phi = \Pi\Psi$ is rank deficient so we are starting from a heavily underdetermined linear system. When the signal $\mathbf{s}$ is sparse, we are interested in finding the sparsest solution to the underdetermined problem, i.e.

$$\hat{\theta} = \arg\min_{\theta} ||\theta||_{\ell_0} \text{ subject to } \mathbf{x} = \Phi\theta. \tag{1}$$

This problem is NP-hard and computationally intractable. Surprisingly, it is shown that If however the matrix Φ acts as an isometry on K-sparse vectors (this is the Restricted Isometry Property or RIP introduced in [4]) then we can use $M = \mathcal{O}(K\log(N/K))$ measurements and perform the ℓ_1 minimization by linear programming

$$\hat{\theta} = \arg\min_{\theta} ||\theta||_{\ell_1} \text{ subject to } \mathbf{x} = \Phi\theta. \tag{2}$$

This can be accomplished by Basis Pursuit (BP) [5], and given RIP, the solution to (2) coincides with the solution to (1). Other efficient algorithms are also proved to give equivalent solutions to equation (1) with high probability such as orthogonal matching pursuit (OMP) [6] and LASSO [7]. Recently bayesian recovery algorithms based on probabilistic inference are also proposed in [8]. These algorithms are especially powerful when considering the tree-structure inherited in the wavelet coefficients.

B. Background on Blind Source Separation

In blind source separation, the task is to recover T unobserved sources $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_T]$ with $\mathbf{s}_i \in \mathbb{R}^N$ via observations of L linear combinations of these sources, possibly corrupted by input of noise. We denote the obervations by $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_L]$ where $\mathbf{x}_i \in \mathbb{R}^N$. The model can be represented by the following equation

$$\mathbf{X} = \mathbf{S}\mathbf{A} + \epsilon \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{T\times L}$ is the mixing matrix, and ϵ is the corresponding noise.

Let $\Psi = [\psi_1, \psi_2, \ldots, \psi_N]$ be a wavelet basis for $\mathbb{R}^N$. We apply the wavelet transform at both sides of equation (3). In the wavelet domain, BSS becomes

$$\tilde{\mathbf{\Theta}} = \mathbf{\Theta}\mathbf{A} + \Theta_\epsilon \tag{4}$$

where $\mathbf{X} = \Psi\tilde{\Theta} = \Psi[\tilde{\theta}_1, \cdots, \tilde{\theta}_T]$, $\mathbf{S} = \Psi\mathbf{\Theta} = \Psi[\theta_1, \cdots, \theta_T]$, $\epsilon = \Psi\Theta_\epsilon$.

C. Problem Formulation: Compressive Blind Source Separation

In this paper, we assume the mixtures of signals are observed via compressed measurements $\mathbf{V}$, possibly with noise $\mathbf{N}$, i.e.

$$\mathbf{V} = \Pi\mathbf{X} + \mathbf{N} = \Phi\mathbf{\Theta}\mathbf{A} + \mathbf{N}, \tag{5}$$

where $\mathbf{V} \in \mathbb{R}^{M\times L}$ and the k^{th} column of the additive noise $\mathbf{N}$ follows the Gaussian distribution $\mathcal{N}(0, (\alpha_k^{\mathbf{N}})^{-1}\mathbf{I})$. We are interested in recovering the source signals in $\mathbf{S}$ directly from the compressed measurements $\mathbf{V}$, in particular, deriving maximum a posterior probabilistic inference. The key is to infer the mixing process and wavelet coefficients jointly from the compressed measurements.

III. Compressive Blind Source Separation

With the problem formulation in equation (5), our proposed method is to maximize the following posterior:

$$\begin{aligned} p(\mathbf{\Theta}, \mathbf{A}, \mathbf{N}|\mathbf{V}, \Phi) \propto \quad & p(\mathbf{V}|\Phi, \mathbf{\Theta}, \mathbf{A}, \alpha_{\mathbf{N}})\pi(\mathbf{N}|\alpha_{\mathbf{N}}) \\ & \cdot\pi(\mathbf{A}|\alpha_{\mathbf{A}})\pi(\mathbf{\Theta}|\alpha_{\mathbf{\Theta}}) \end{aligned} \tag{6}$$

and we denote $\alpha = [\alpha_{\mathbf{N}}, \alpha_{\mathbf{A}}, \alpha_{\mathbf{\Theta}}]$ as the set of the hyperparameters.

A. Wavelet Hidden Markov Tree Model

Let all source signals be decomposed into S scales in wavelet demain and each source fits into a different HMT model. The hidden Markov tree (HMT) [9] is a graphical model where the inner dependencies between the wavelet transform coefficients are explored explicitly. Fig. 1 illustrates a hidden Markov tree where each parent state has two children states. Each white node represents an unobserved hidden state and each black node represents an observed wavelet coefficient. The hidden Markov tree model captures the following properties.

- **Persistence:** If the parent node has a large/small coefficient, it is of higher probability that the child node also has a large/small coefficient, and vice versa. If a particular wavelet coefficient is large/small, then adjacent coefficients are also very likely to be large/small. In our context, we only take into account the dependency between the parent and child nodes.
- **Mixed Gaussian Model:** We denote $c^{s,i}$ as the i^{th} hidden state at wavelet decomposition scale s; it has two possible state values, $c^{s,i} \in \{S, L\}$ and its associated wavelet coefficient $\theta^{s,i}$ is subject to the following distribution,

$$\theta^{s,i} \sim (1 - \pi^{s,i})\delta_0 + \pi^{s,i}\mathcal{N}(0, (\alpha^s)^{-1}) \tag{7}$$

where $\mathcal{N}(\cdot)$ denotes Gaussian distribution and α^s is the precision parameter (variance $\sigma^2 = (\alpha^s)^{-1}$) on the wavelet decomposition scale s.

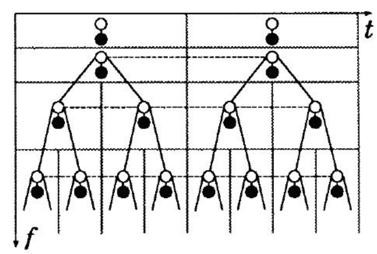

Fig. 1. Hidden Markov Tree as a Graphical Model

One way of training and maximizing likelihood estimation in the HMT is via the EM algorithm in HMT [9], [10]. Instead, we will use Gibbs sampling Markov chain Monte Carlo inference in this paper.

B. Prior Distributions

1) Noise Variance Prior Distribution: As shown in equation (5), the additive noise $\mathbf{N}$ follows Gaussian distribution with zero mean and precision $\alpha_{\mathbf{N}} = \{\alpha_k^{\mathbf{N}}\}_{k=1}^N$. The noise precision $\alpha_k^{\mathbf{N}}$ is assigned as Gamma distribution

$$\alpha_k^{\mathbf{N}} \sim \text{Gamma}(a_0, b_0). \tag{8}$$

The measured linear projections $\mathbf{V}$ follows conditional Gaussian as

$$\mathbf{V}_k|\Phi, \mathbf{\Theta}, \mathbf{A}, \alpha_k^{\mathbf{N}} \sim \mathcal{N}(\Phi\mathbf{\Theta}\mathbf{A}_k, (\alpha_k^{\mathbf{N}})^{-1}\mathbf{I}) \tag{9}$$

where $\mathbf{V}_k$ is the k^{th} column of $\mathbf{V}$ and $\mathbf{A}_k$ is the k^{th} column of $\mathbf{A}$.

2) Mixing Matrix Prior Distribution: The prior distribution of mixing matrix $\mathbf{A}$ is indeed determined by the mixing process. We consider that each entry of the mixing matrix $\mathbf{A}$ follows Gaussian distribution as $a_{i,j} \sim \mathcal{N}(\mu_{i,j}, \alpha_{i,j}^{-1})$. The hyperparameter set hence is $\alpha_{\mathbf{A}} = \{\mu_{i,j}, \alpha_{i,j}\}_{1\le i\le T, 1\le j\le L}$.

3) Prior Distributions of Wavelet Coefficients: According to the hidden Markov tree model, the prior distributions of wavelet coefficients are assigned as mixed Gaussian with precision having gamma priors.

$$\theta^{s,i} \sim (1-\pi^{s,i})\delta_0 + \pi^{s,i}\mathcal{N}(0,(\alpha^s)^{-1}),$$
$$\text{with } \pi^{s,i} = \begin{cases} \pi^r, & \text{if } s=1, \\ \pi^{s0}, & \text{if } 2 \le s \le S, \theta^{p(s,i)} = 0, \\ \pi^{s1}, & \text{if } 2 \le s \le S, \theta^{p(s,i)} \neq 0, \end{cases}$$
$$\alpha^s \sim \text{Gamma}(c_0,d_0),\ \pi^r \sim \text{Beta}(e_0^r, f_0^r),$$
$$\pi^{s0} \sim \text{Beta}(e_0^{s0}, f_0^{s0}),\ \pi^{s1} \sim \text{Beta}(e_0^{s1}, f_0^{s1}), \tag{10}$$

where $\theta^{p(s,i)}$ denotes the coefficient for the parent node of node $\theta^{s,i}$. If the parent coefficient is zero, the children coefficients are more likely to be zero due to the persistence property, therefore we assign a relatively small π^{s0}, otherwise we assign a relative large π^{s1}. For the root node, no preference is given to π^r. Each source is assigned a different set of hyperparameters. The overall parameter set concerning the wavelet coefficients is denoted by $\alpha_{\mathbf{\Theta}}$.

C. Markov Chain Monte Carlo Inference

In this subsection, we illustrate posterior computations through Markov chain Monte Carlo inference by Gibbs sampling. The Gibbs sampler approximates true posterior distributions by sequentially sampling from conditional distributions of the rest parameters and unknown variables. For the posterior specified in equation (6), the Gibbs sampler samples from the following conditional distributions at iteration t,

$$\begin{aligned} &\theta_k^{s,i}(t) \sim p(\theta_k^{s,i}|\mathbf{V},\Phi,\mathbf{A}(t-1),\alpha_k^{\mathbf{N}}(t-1), \\ &\qquad\qquad \alpha_k^s(t-1),\pi_k^{s,i}(t-1)), \\ &\mathbf{A}(t) \sim p(\mathbf{A}|\mathbf{V},\Phi,\mathbf{A}(t-1),\Theta(t-1)), \\ &\alpha_k^{\mathbf{N}}(t) \sim p(\alpha_k^{\mathbf{N}}|\mathbf{V}_k,\Phi,\mathbf{A}_k(t-1),\Theta(t-1)), \\ &\alpha_k^s(t) \sim p(\alpha_k^s|\theta_k^{s,i}(t-1)), \\ &\pi_k^{s,i}(t) \sim p(\pi_k^{s,i}|\theta_k^{s,i}(t-1)), \end{aligned} \tag{11}$$

where $\{\theta_k^{s,i}(t)\}$ is the set of wavelet coefficients associated with the k^{th} source in the t^{th} iteration. Note that, for each source, the wavelet coefficients are different and they need to be inferred separately in a similar fashion. The posterior distributions are specified as follows. Most of them follow are fairly easy to derive since they all come from the prior and the likelihood distribution are conjugate.

1) Conditional Distributions of Wavelet Coefficients: The conditional distribution of the wavelet coefficient $\theta_k^{s,i}$ for the k^{th} source is specified as

$$p(\theta_k^{s,i}|-) = (1-\tilde{\pi}_k^{s,i})\delta_0 + \tilde{\pi}_k^{s,i}\mathcal{N}(\tilde{\mu}_k^{s,i},(\tilde{\alpha}_k^s)^{-1}). \tag{12}$$

We assume $\theta_k^{s,i}$ is the j^{th} entry of the N-dimensional vector θ_k, denoted by $\theta_k^j = \theta_k^{s,i}$, then

$$\begin{aligned} \tilde{\alpha}_k^s &= \alpha_k^s + \alpha_k^{\mathbf{N}}\phi_j^T\phi_j, \\ \tilde{\mu}_k^{s,i} &= (\tilde{\alpha}_k^s)^{-1}\alpha_k^{\mathbf{N}}\phi_j^T\tilde{\mathbf{V}}_k^j, \\ \frac{\tilde{\pi}_k^{s,i}}{1-\tilde{\pi}_k^{s,i}} &= \frac{\pi_k^{s,i}}{1-\pi_k^{s,i}}\frac{\mathcal{N}(0|(\alpha_k^s)^{-1})}{\mathcal{N}(\tilde{\mu}_k^{s,i}|(\tilde{\alpha}_k^s)^{-1})}, \end{aligned}$$

where ϕ_j is the j^{th} column of Φ and $\tilde{\mathbf{V}}_k^j = \mathbf{V}_k - \sum_{p\neq j}\phi_p\theta_k^p$.

2) Conditional Distribution of Coefficients hyperparameters: The conditional distribution of hyperparameters for wavelet coefficients are specified as

- $p(\alpha_k^s|-) =$ $\text{Gamma}(c_0 + \frac{1}{2}\sum_{i=1}^{M_s}\mathbf{1}_{\{\theta_k^{s,i}\neq 0\}}, d_0 + \frac{1}{2}\sum_{i=1}^{M_s}(\theta_k^{s,i})^2)$,
- $p(\pi_k^r|-) =$ $\text{Beta}(e_0^r + \frac{1}{2}\sum_{i=1}^{M_s}\mathbf{1}_{\{\theta_k^{s,i}\neq 0\}}, f_0^r + \frac{1}{2}\sum_{i=1}^{M_s}\mathbf{1}_{\{\theta_k^{s,i}=0\}})$, $s=1$,
- $p(\pi_k^{s0}|-) = \text{Beta}(e_0^{s0} + \frac{1}{2}\sum_{i=1}^{M_s}\mathbf{1}_{\{\theta_k^{s,i}\neq 0,\ \theta_{p(s,i)}(k)=0\}}$, $f_0^{s0} + \frac{1}{2}\sum_{i=1}^{M_s}\mathbf{1}_{\{\theta_k^{s,i}=0,\ \theta_k^{p(s,i)}=0\}})$, $s\neq 1$,
- $p(\pi_k^{s1}|-) = \text{Beta}(e_0^{s1} + \frac{1}{2}\sum_{i=1}^{M_s}\mathbf{1}_{\{\theta_k^{s,i}\neq 0,\ \theta_{p(s,i)}(k)\neq 0\}}$, $f_0^{s1} + \frac{1}{2}\sum_{i=1}^{M_s}\mathbf{1}_{\{\theta_k^{s,i}=0,\ \theta_k^{p(s,i)}\neq 0\}})$, $s\neq 1$.

3) Conditional Distribution of Noise Variance: The conditional distribution of the noise precision $\alpha_k^{\mathbf{N}}$ is

$$p(\alpha_k^{\mathbf{N}}|-) = \text{Gamma}(a_0 + \tfrac{M}{2}, b_0 + \tfrac{1}{2}(\mathbf{V}_k - \Phi\mathbf{\Theta}\mathbf{A}_k)^T(\mathbf{V}_k - \Phi\mathbf{\Theta}\mathbf{A}_k)). \tag{13}$$

4) Conditional Distribution of Mixing Matrix: The conditional distribution of $a_{i,j}$ is

$$p(a_{i,j}|-) = \mathcal{N}(\tilde{\mu}_{i,j}, \tilde{\alpha}_{i,j}^{-1}) \tag{14}$$

and

$$\tilde{\mu}_{i,j} = \frac{\mu_{i,j}\alpha_{i,j} + \mathbf{P}_{i,j}\mathbf{Q}_{i,j}\alpha_j^{\mathbf{N}}}{\alpha_{i,j} + \mathbf{Q}_{i,j}\alpha_j^{\mathbf{N}}}, \tag{15}$$
$$\tilde{\alpha}_{i,j} = \alpha_{i,j} + \mathbf{Q}_{i,j}\alpha_j^{\mathbf{N}}, \tag{16}$$

where $\mathbf{P} = \left((\Phi\mathbf{\Theta})^H(\Phi\mathbf{\Theta})\right)^{-1}(\Phi\mathbf{\Theta})^H\mathbf{V}$ and $\mathbf{Q} = \left((\Phi\mathbf{\Theta})^H(\Phi\mathbf{\Theta})\right)^{-1}(\Phi\mathbf{\Theta})^H\mathbf{I}_{M\times L}$.

IV. Simulation Results

In formulation (5), we note that, there is an underlying algorithm. That is separate procedure of compressive sensing and wavelet-based blind source separation, i.e.

- Step 1: Inference of $\tilde{\mathbf{\Theta}}$ from linear projection $\mathbf{V}$ (see [8]).
- Step 2: Inference of $\mathbf{\Theta}$ from $\tilde{\mathbf{\Theta}}$ (see [3]).

We call this the decoupled method and note that it is generic. Indeed there is no other generic methods to handle with compressive blind source separation.We now present simulations that compare our integrated approach with the decoupled approach.

In one round of MCMC inference, we set burn-in process as 400 iterations and the collection period as 100 iterations.

A. One-dimensional Signals

Here, the task is to separate two one-dimensional sparse signals $\mathbf{s}_1$ and $\mathbf{s}_2$. The two sources and mixing matrix are generated randomly. In the wavelet domain, their coefficient vectors are θ_1 and θ_2 respectively with dimensions as 512. There are two mixtures and 200 measurements are taken from each mixture. Therefore, the linear projection $\mathbf{V}$ has size 200×2 and the compression rate is $\frac{200}{512}$.

In Fig. 2, 'Sep-Recovered Signals' denotes signals recovered by the separate method; 'Bayes-Recovered Signals' denotes signals recovered by our proposed method. It shows that

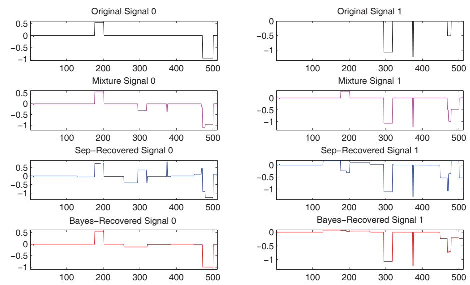

Fig. 2. Bayesian compressive blind separation of one dimensional signals

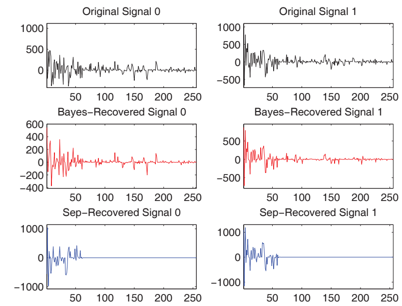

Fig. 5. Comparisons of the first 256 recovered wavelet coefficients

our proposed method provides better recoveries. The wavelet coefficients are provided in Fig. 3.

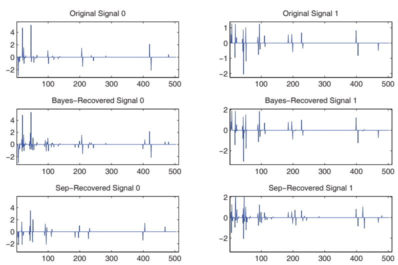

Fig. 3. Comparisons of recovered wavelet coefficients

B. Two-dimensional Images

We consider blind separation of two 32×32 images. Each coefficient vector in the wavelet domain has 1024 entries. Two mixtures are generated randomly and we take 500 measurements from each mixture. Therefore, the compression rate is $\frac{500}{1024}$.

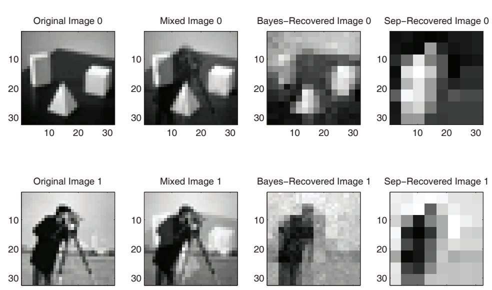

Fig. 4. Bayesian compressive blind separation of two images

Fig. 4 shows that our proposed method provides better recoveries. The first 256 wavelet coefficients are provided in Fig. 5, which also verifies better recoveries by our proposed method.

We note that, both methods do not give very clear recoveries; one reason is that the original images has very low definition and the second reason is that these two low-definition images are not sparse enough. From Fig. 5, we notice that these images are not sparse but *nearly sparse* in wavelet domain (most entries are close to zero but not equal to zero); yet, our algorithm mainly aims separating and recovering sparse signals. It would be interesting to tune our algorithms to handle these nearly sparse images. This will be our future work.

V. Conclusion

In this paper, we have considered the blind source separation problem directly from the compressed mixtures obtained from compressive sensing measurements. We have proposed a bayesian inference which separates and recovers original sparse signals directly from the compressed mixtures. We have also shown that our proposed algorithm provides effective recoveries and outperforms its peer, especially for one-dimensional signals. As for the future work, we will focus on improving our proposed method in separation and recovery of nearly sparse signals by exploring their special structures.

References

[1] E. J. Candés, “Compressive sampling,” vol. 3, 2006, pp. 1433–1452.
[2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[3] M. M. Ichir and A. Mohammad-Djafari, “Hidden markov models for wavelet-based blind source separation,” *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1887–1899, 2006.
[4] E. J. Candés and T. Tao, “Decoding by linear programming,” *IEEE Trans. Info. Theory*, vol. 51, pp. 4203– 4215, 2005.
[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1999.
[6] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Info. Theory*, vol. 53, pp. 4655–4666, 2007.
[7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics (with discussion)*, vol. 32, pp. 407–499, 2004.
[8] L. He and L. Carin, “Exploiting structure in wavelet-based bayesian compressive sensing,” *IEEE Trans. Signal Processing*, vol. 57, no. 9, pp. 3488–3497, 2009.
[9] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, “Wavelet-based statistical signal processing using hidden markov models,” *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 882–902, 1998.
[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *J. Roy. Stat. Soc. B*, vol. 39, pp. 1–17, 1977.