# Pattern-Coupled Sparse Bayesian Learning for Recovery of Block-Sparse Signals

Yanning Shen\*, Huiping Duan\*, Jun Fang\*, and Hongbin Li<sup>‡</sup>

\*National Key Laboratory on Communications, University of Electronic Science and Technology of China Chengdu 611731, China Emails: 201121260110@std.uestc.edu.cn, {HuipingDuan,JunFang}@uestc.edu.cn <sup>‡</sup>Department of Electrical and Computer Engineering, Stevens Institute of Technology Hoboken, NJ 07030 USA Email: Hongbin.Li@stevens.edu

Abstract-In this paper, we develop a new sparse Bayesian learning method for recovery of block-sparse signals with unknown cluster patterns. A pattern-coupled hierarchical Gaussian prior model is introduced to characterize the statistical dependencies among coefficients, where a set of hyperparameters are employed to control the sparsity of signal coefficients. Unlike the conventional sparse Bayesian learning framework in which each individual hyperparameter is associated independently with each coefficient, in this paper, the prior for each coefficient not only involves its own hyperparameter, but also the hyperparameters of its immediate neighbors. In doing this way, the sparsity patterns of neighboring coefficients are related to each other and the hierarchical model has the potential to encourage structuredsparse solutions. The hyperparameters, along with the sparse signal, are learned by maximizing their posterior probability via an expectation-maximization (EM) algorithm.

Index Terms—Sparse Bayesian learning, pattern-coupled hierarchical model, block-sparse signal recovery.

# I. INTRODUCTION

Compressive sensing is a recently emerged technique of data acquisition through exploiting the inherent sparsity of signals of interest. In practice, sparse signals usually have additional structures that can be exploited to enhance the recovery performance. For example, the atomic decomposition of multiband signals [1] or audio signals [2] usually results in a blocksparse structure in which the nonzero coefficients occur in clusters. In addition, a discrete wavelet transform of an image naturally yields a tree structure of the wavelet coefficients [3]. A number of algorithms, e.g., block-OMP [4], mixed  $\ell_2/\ell_1$  norm-minimization [5], group LASSO [6], and modelbased CoSaMP [7] were proposed for recovery of blocksparse signals. These algorithms, albeit effective, require the knowledge of the block structure (such as locations and sizes of blocks) of sparse signals a priori. In practice, however, the prior information about the block structure of sparse signals is often unavailable. To address this difficulty, a hierarchical Bayesian "spike-and-slab" prior model is introduced in [3], [8] to encourage the sparseness and promote the cluster patterns simultaneously. Nevertheless, for both works [3], [8], the posterior distribution cannot be derived analytically, and a Markov chain Monte Carlo (MCMC) sampling method has to

This work was supported in part by the National Science Foundation of China under Grant 61201274, and the Fundamental Research Funds for the Central Universities (China) under Grant ZYGX2012J025.

be employed for Bayesian inference. In [9], [10], a graphical prior, also referred to as the "Boltzmann machine", was used to model the statistical dependencies between atoms. Specifically, the Boltzmann machine is employed as a prior on the support of a sparse representation. However, the maximum a posterior (MAP) estimator with such a prior involves an exhaustive search over all possible sparsity patterns. To overcome the intractability of the combinatorial search, a greedy method [9] and a variational mean-field approximation method [10] were proposed to approximate the MAP. Recently, a sparse Bayesian learning method was proposed in [11] to address the sparse signal recovery problem when the block structure is unknown, where the components of the signal are partitioned into a number of overlapping blocks and each block is assigned a Gaussian prior. An expanded model is then used to convert the overlapping structure into a block diagonal structure so that the conventional block sparse Bayesian learning algorithm can be readily applied.

In this paper, we develop a new Bayesian method for block-sparse signal recovery when the block-sparse patterns are entirely unknown. To model the block-sparse patterns, we propose a coupled hierarchical Gaussian framework. Such a prior encourages clustered patterns and suppresses "isolated coefficients" whose pattern is different from that of its neighboring coefficients. An expectation-maximization (EM) algorithm is developed to estimate the block-sparse signal.

## II. HIERARCHICAL PRIOR MODEL

We consider the problem of recovering a block-sparse signal  $x \in \mathbb{R}^n$  from noise-corrupted measurements

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w} \tag{1}$$

where  $A \in \mathbb{R}^{m \times n}$  (m < n) is the measurement matrix, and w is the additive multivariate Gaussian noise with zero mean and covariance matrix  $\sigma^2 I$ . The signal x has a block-sparse structure but the exact block pattern such as the location and size of each block is unavailable to us.

In the conventional sparse Bayesian learning framework, to encourage the sparsity of the estimated signal, x is assigned a Gaussian prior distribution

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{i=1}^{n} p(x_i|\alpha_i)$$
(2)

where  $p(x_i|\alpha_i) = \mathcal{N}(x_i|0, \alpha_i^{-1})$ , and  $\alpha \triangleq \{\alpha_i\}$  are nonnegative hyperparameters controlling the sparsity of the signal x. Clearly, when  $\alpha_i$  approaches infinity, the corresponding coefficient  $x_i$  becomes zero. We see that in the above conventional hierarchical Bayesian model, each hyperparameter is associated independently with each coefficient. The prior model assumes independence among coefficients and has no potential to encourage clustered sparse solutions.

To exploit the statistical dependencies among coefficients, we propose a new coupled hierarchical Bayesian model in which the prior for each coefficient not only involves its own hyperparameter, but also the hyperparameters of its immediate neighbors. Specifically, a prior over x is given by

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{i=1}^{n} p(x_i|\alpha_i, \alpha_{i+1}, \alpha_{i-1})$$
(3)

where

$$p(x_{i}|\alpha_{i}, \alpha_{i+1}, \alpha_{i-1}) = \mathcal{N}(x_{i}|0, (\alpha_{i} + \beta\alpha_{i+1} + \beta\alpha_{i-1})^{-1})$$
(4)

and we assume  $\alpha_0 = 0$  and  $\alpha_{n+1} = 0$  for the end points  $x_1$ and  $x_n$ ,  $0 \leq \beta \leq 1$  is a parameter indicating the relevance between the coefficient  $x_i$  and its neighboring coefficients  $\{x_{i+1}, x_{i-1}\}$ . When  $\beta = 0$ , the prior distribution (4) reduces to the prior for the conventional sparse Bayesian learning, while when  $\beta > 0$ , the sparsity of  $x_i$  not only depends on the hyperparameter  $\alpha_i$ , but also on the neighboring hyperparameters  $\{\alpha_{i+1}, \alpha_{i-1}\}$ . Hence the sparsity patterns of neighboring coefficients are related to each other. The model naturally has the tendency to suppress isolated non-zero coefficients and encourage structured-sparse solutions.

We use Gamma distributions as hyperpriors over the hyperparameters  $\{\alpha_i\}$ , i.e.

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{n} \operatorname{Gamma}(\alpha_{i}|a,b) = \prod_{i=1}^{n} \Gamma(a)^{-1} b^{a} \alpha^{a} e^{-b\alpha}$$
(5)

where  $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$  is the Gamma function. In the conventional sparse Bayesian framework, to make the Gamma prior non-informative, very small values, e.g.  $10^{-4}$ , are assigned to the two parameters a and b. Nevertheless, in this paper, we use a more favorable prior which sets a larger a (say, a = 0.5) in order to achieve the desired "pruning" effect for our proposed hierarchical Bayesian model.

# III. PROPOSED BAYESIAN INFERENCE ALGORITHM

We now proceed to develop a sparse Bayesian learning method for block-sparse signal recovery. Based on the above hierarchical model, the posterior distribution of x can be computed as

$$p(\boldsymbol{x}|\boldsymbol{\alpha},\boldsymbol{y}) \propto p(\boldsymbol{x}|\boldsymbol{\alpha})p(\boldsymbol{y}|\boldsymbol{x})$$
(6)

where  $\boldsymbol{\alpha} \triangleq \{\alpha_i\}, p(\boldsymbol{x}|\boldsymbol{\alpha})$  is given by (3), and

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi\sigma^2})^m} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2}{2\sigma^2}\right)$$
(7)

It can be readily verified that the posterior  $p(\boldsymbol{x}|\boldsymbol{\alpha}, \boldsymbol{y})$  follows a Gaussian distribution with its mean and covariance given respectively by

$$\boldsymbol{\mu} = \sigma^{-2} \boldsymbol{\Phi} \boldsymbol{A}^T \boldsymbol{y}$$
  
$$\boldsymbol{\Phi} = (\sigma^{-2} \boldsymbol{A}^T \boldsymbol{A} + \boldsymbol{D})^{-1}$$
(8)

where D is a diagonal matrix with its *i*th diagonal element equal to  $(\alpha_i + \beta \alpha_{i+1} + \beta \alpha_{i-1})$ , i.e.

$$\boldsymbol{D} \triangleq \operatorname{diag}(\alpha_1 + \beta \alpha_2 + \beta \alpha_0, \dots, \alpha_n + \beta \alpha_{n-1} + \beta \alpha_{n+1})$$
(9)

Given a set of estimated hyperparameters  $\{\alpha_i\}$ , the maximum a posterior (MAP) estimate of x is the mean of its posterior distribution, i.e.

$$\hat{\boldsymbol{x}}_{\text{MAP}} = \boldsymbol{\mu} = (\boldsymbol{A}^T \boldsymbol{A} + \sigma^2 \boldsymbol{D})^{-1} \boldsymbol{A}^T \boldsymbol{y}$$
(10)

Our problem therefore reduces to estimating the set of hyperparameters  $\{\alpha_i\}$ . With hyperpriors placed over  $\alpha_i$ , learning the hyperparameters becomes a search for their posterior mode, i.e. maximization of the posterior probability  $p(\alpha|\mathbf{y})$ . A strategy to maximize the posterior probability is to exploit the expectation-maximization (EM) formulation which treats the signal  $\mathbf{x}$  as the hidden variables and maximizes the expected value of the complete log-posterior of  $\alpha$ , i.e.  $E_{\mathbf{x}|\mathbf{y},\alpha}[\log p(\alpha|\mathbf{x})]$ , where the operator  $E_{\mathbf{x}|\mathbf{y},\alpha}[\cdot]$  denotes the expectation with respect to the distribution  $p(\mathbf{x}|\mathbf{y},\alpha)$ . Specifically, the EM algorithm produces a sequence of estimates  $\alpha^{(t)}$ ,  $t = 1, 2, 3, \ldots$ , by applying two alternating steps, namely, the E-step and the M-step.

**E-Step:** Given the current estimates of the hyperparameters  $\alpha^{(t)}$  and the observed data y, the E-step requires computing the expected value (with respect to the missing variables x) of the complete log-posterior of  $\alpha$ , which is also referred to as the Q-function; we have

$$Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)}) = E_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)}}[\log p(\boldsymbol{\alpha}|\boldsymbol{x})]$$
  
=  $\int p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)})\log p(\boldsymbol{\alpha}|\boldsymbol{x})d\boldsymbol{x}$   
=  $\int p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)})\log[p(\boldsymbol{\alpha})p(\boldsymbol{x}|\boldsymbol{\alpha})]d\boldsymbol{x} + c$  (11)

where c is a constant independent of  $\alpha$ . Ignoring the term independent of  $\alpha$ , and recalling (3), the Q-function can be re-expressed as

$$Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)}) = \log p(\boldsymbol{\alpha}) + \frac{1}{2} \sum_{i=1}^{n} \left( \log(\alpha_{i} + \beta \alpha_{i+1} + \beta \alpha_{i-1}) - (\alpha_{i} + \beta \alpha_{i+1} + \beta \alpha_{i-1}) \int p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)}) x_{i}^{2} d\boldsymbol{x} \right)$$
(12)

Since the posterior  $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)})$  is a multivariate Gaussian distribution with its mean and covariance matrix given by (8), we have

$$\int p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)}) x_i^2 d\boldsymbol{x} = E_{\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)}} \left[ x_i^2 \right] = \hat{\mu}_i^2 + \hat{\phi}_{i,i} \quad (13)$$

where  $\hat{\mu}_i$  denotes the *i*th entry of  $\hat{\mu}$ ,  $\hat{\phi}_{i,i}$  denotes the *i*th diagonal element of the covariance matrix  $\hat{\Phi}$ ,  $\hat{\mu}$  and  $\hat{\Phi}$  are computed according to (8), with  $\alpha$  replaced by the current estimate  $\alpha^{(t)}$ . With the specified prior (5), the Q-function can eventually be written as

$$Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)})$$

$$=\sum_{i=1}^{n} \left( a \log \alpha_{i} - b\alpha_{i} + \frac{1}{2} \log(\alpha_{i} + \beta \alpha_{i+1} + \beta \alpha_{i-1}) - \frac{1}{2} (\alpha_{i} + \beta \alpha_{i+1} + \beta \alpha_{i-1}) (\hat{\mu}_{i}^{2} + \hat{\phi}_{i,i}) \right)$$
(14)

**M-Step:** In the M-step of the EM algorithm, a new estimate of  $\alpha$  is obtained by maximizing the Q-function, i.e.

$$\boldsymbol{\alpha}^{(t+1)} = \arg \max_{\boldsymbol{\alpha}} Q(\boldsymbol{\alpha} | \boldsymbol{\alpha}^{(t)})$$
(15)

We see that the hyperparameters in the Q-function (14) are entangled with each other due to the logarithm term  $\log(\alpha_i + \beta\alpha_{i+1} + \beta\alpha_{i-1})$ . In this case, an analytical solution to the optimization (15) is difficult to obtain. Gradient descend methods can certainly be used to search for the optimal solution. Nevertheless, we consider an alternative strategy which aims at finding a simple, analytical sub-optimal solution of (15). Such an analytical sub-optimal solution can be obtained by examining the optimality condition of (15). Suppose  $\alpha^*$  is the optimal solution of (15), then the first derivative of the Qfunction with respect to  $\alpha$  equals to zero at the optimal point, i.e.

$$\frac{\partial Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)})}{\partial \boldsymbol{\alpha}}\Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}*} = \mathbf{0}$$
(16)

To examine this optimality condition more thoroughly, we compute the first derivative of the Q-function with respect to each individual hyperparameter:

$$\frac{\partial Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)})}{\partial \alpha_{i}} = \frac{a}{\alpha_{i}} - b - \frac{1}{2}\omega_{i} + \frac{1}{2}(\nu_{i} + \beta\nu_{i+1} + \beta\nu_{i-1})$$
$$\forall i = 1, \dots, n$$
(17)

where  $\nu_0 = 0$ ,  $\nu_{n+1} = 0$ , and for i = 1, ..., n, we have  $\omega_i \triangleq (\hat{\mu}_i^2 + \hat{\phi}_{i,i}) + \beta(\hat{\mu}_{i+1}^2 + \hat{\phi}_{i+1,i+1}) + \beta(\hat{\mu}_{i-1}^2 + \hat{\phi}_{i-1,i-1})$ 

$$\nu_i \triangleq \frac{1}{\alpha_i + \beta \alpha_{i+1} + \beta \alpha_{i-1}} \tag{18}$$

Note that in (18),  $\{\hat{\mu}_0, \hat{\phi}_{0,0}, \hat{\mu}_{n+1}, \hat{\phi}_{n+1,n+1}\}$  should all set equal to zero, i.e.  $\hat{\mu}_0 = \hat{\mu}_{n+1} = \hat{\phi}_{0,0} = \hat{\phi}_{n+1,n+1} = 0$ . Recalling the optimality condition, we therefore have

$$\frac{a}{\alpha_i^*} + \frac{1}{2}(\nu_i^* + \beta \nu_{i+1}^* + \beta \nu_{i-1}^*) = b + \frac{1}{2}\omega_i \quad \forall i = 1, \dots, n$$
(20)

where  $\nu_0^* = 0$ ,  $\nu_{n+1}^* = 0$ , and

$$\nu_i^* \triangleq \frac{1}{\alpha_i^* + \beta \alpha_{i+1}^* + \beta \alpha_{i-1}^*} \qquad \forall i = 1, \dots, n$$

Since all hyperparameters  $\{\alpha_i\}$  are non-negative, we have

$$\frac{1}{\alpha_i^*} > \nu_i^* > 0 \qquad \forall i = 1, \dots, n$$
$$\frac{1}{\beta \alpha_{i+1}^*} > \nu_i^* > 0 \qquad \forall i = 1, \dots, n-1$$
$$\frac{1}{\beta \alpha_{i-1}^*} > \nu_i^* > 0 \qquad \forall i = 2, \dots, n$$

Hence the term on the left-hand side of (20) is lower and upper bounded respectively by

$$\frac{a+c_0}{\alpha_i^*} \ge \frac{a}{\alpha_i^*} + \frac{1}{2} (\nu_i^* + \beta \nu_{i+1}^* + \beta \nu_{i-1}^*) > \frac{a}{\alpha_i^*}$$
(21)

where  $c_0 = 1.5$  for i = 2, ..., n-1, and  $c_0 = 1$  for  $i = \{1, n\}$ . Combining (20)–(21), we arrive at

$$\alpha_i^* \in \left[\frac{a}{0.5\omega_i + b}, \frac{a + c_0}{0.5\omega_i + b}\right] \quad \forall i = 1, \dots, n$$
 (22)

With a = 0.5, and  $b = 10^{-4}$ , a sub-optimal solution to (15) can be obtained as

$$\hat{\alpha}_i = \frac{\kappa}{0.5\omega_i + 10^{-4}} \quad \forall i = 1, \dots, n$$
(23)

for some  $\kappa$  within the range  $0.5 + c_0 \ge \kappa \ge 0.5$ . Notice that the update rule (23) resembles that of the conventional sparse Bayesian learning work [12], [13] except that the parameter  $\omega_i$  is equal to  $\hat{\mu}_i^2 + \hat{\phi}_{i,i}$  for the conventional sparse Bayesian learning method, while for our case,  $\omega_i$  is a summation of  $\hat{\mu}_i^2 + \hat{\phi}_{j,j}$  for j = i - 1, i, i + 1.

For clarity, we now summarize the EM algorithm as follows.

- 1) At iteration t (t = 0, 1, ...): Given a set of hyperparameters  $\boldsymbol{\alpha}^{(t)} = \{\alpha_i^{(t)}\}$ , compute the mean  $\hat{\boldsymbol{\mu}}$  and covariance matrix  $\hat{\boldsymbol{\Phi}}$  of the posterior distribution  $p(\boldsymbol{x}|\boldsymbol{\alpha}^{(t)}, \boldsymbol{y})$  according to (8), and compute the MAP estimate  $\hat{\boldsymbol{x}}^{(t)}$  according to (10).
- 2) Update the hyperparameters  $\alpha^{(t+1)}$  according to (23), where  $\omega_i$  is given by (18).
- 3) Continue the above iteration until  $\|\hat{x}^{(t+1)} \hat{x}^{(t)}\|_2 \le \epsilon$ , where  $\epsilon$  is a prescribed tolerance value.

Although the above algorithm employs a sub-optimal solution (23) to update the hyperparameters in the M-step, numerical results show that the sub-optimal update rule is quite effective and presents similar recovery performance as using a gradient-based search method. This is because the sub-optimal solution (23) provides a reasonable estimate of the optimal solution when the parameter a is reasonably large, say, a = 0.5.

#### **IV. SIMULATION RESULTS**

In our simulations, we generate the block-sparse signal in a similar way to [11]. Suppose the *n*-dimensional sparse signal contains K nonzero coefficients which are partitioned into L blocks with random sizes and random locations. Also, the nonzero coefficients of the sparse signal x and the measurement matrix  $A \in \mathbb{R}^{m \times n}$  are randomly generated with each entry independently drawn from a normal distribution.



Fig. 1. The original image 'Lena' and the reconstructed images using respective algorithms.

We examine the recovery performance of our proposed algorithm, also referred to as the pattern-coupled sparse Bayesian learning algorithm (PC-SBL), under different choices of  $\beta^1$ . As indicated earlier in our paper,  $\beta$  ( $0 \le \beta \le 1$ ) is a parameter quantifying the dependencies among neighboring coefficients. Fig. 2 depicts the success rates vs. the ratio m/nfor different choices of  $\beta$ , where we set n = 100, K = 25, and L = 4. Results are averaged over 1000 independent runs, with the measurement matrix and the sparse signal randomly generated for each run. The performance of the conventional sparse Bayesian learning method (denoted as "SBL") [12] and the basis pursuit method (denoted as "BP") [14], [15] is also included for our comparison. We see that when  $\beta = 0$ , our proposed algorithm performs the same as the SBL. This is an expected result since in the case of  $\beta = 0$ , our proposed algorithm is simplified as the SBL. Nevertheless, when  $\beta > 0$ , our proposed algorithm achieves a significant performance improvement (as compared with the SBL and BP) through exploiting the underlying block-sparse structure, even without knowing the exact locations and sizes of the non-zero blocks. We also observe that our proposed algorithm is not very sensitive to the choice of  $\beta$  as long as  $\beta > 0$ .

We carry out experiments on real world images. As it is well-known, images have clustered sparse structures whose significant coefficients tend to be located together in certain over-complete basis, such as wavelet or discrete cosine transform (DCT) basis. We compare our proposed algorithm with some other recently developed algorithms for block-sparse signal recovery, namely, the expanded block sparse Bayesian learning method (EBSBL) [11], the Boltzman machine-based greedy pursuit algorithm (BM-MAP-OMP) [9], and the cluster-structured MCMC algorithm (CluSS-MCMC) [8]. The block sparse Bayesian learning method (denoted as BSBL) developed in [11] is included as well. In our experi-

<sup>1</sup>Matlab codes are available at http://www.junfang-uestc.net/codes/PC-SBL.rar



Fig. 2. Success rates of the proposed algorithm vs. the ratio m/n for different choices of  $\beta$ .

ments, the image is processed in a columnwise manner: we sample each column of the  $128 \times 128$  image using a randomly generated measurement matrix  $A \in \mathbb{R}^{m \times 128}$ , recover each column from the *m* measurements, and reconstruct the image based on the 128 estimated columns. Fig. 1 show the original image 'Lena' and the reconstructed images using respective algorithms, where we set m = 64. We see that our proposed algorithm presents the finest image quality among all methods.

# V. CONCLUSIONS

We developed a new Bayesian method for recovery of block-sparse signals whose block-sparse structures are entirely unknown. A pattern-coupled hierarchical Gaussian prior model was introduced to characterize both the sparseness of the coefficients and the statistical dependencies between neighboring coefficients of the signal. Numerical results show that our proposed algorithm achieves a significant performance improvement through exploiting the underlying block-sparse structure, and outperforms other existing methods in blocksparse signal recovery.

## References

- M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: compressed sensing for analog signals," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.
   R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals
- [2] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 101–111, Jan. 2003.
- [3] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 57, no. 9, pp. 3488–3497, Sept. 2009.
- [4] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Trans. Information Theory*, vol. 58, no. 6, pp. 3042–3054, June 2010.
- [5] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Information Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [6] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, pp. 49–67, 2006.
  [7] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based
- [7] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Information Theoy*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [8] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, pp. 259–269, 2012.
- [9] T. Peleg, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Trans. Signal Processing*, vol. 60, no. 5, pp. 2286–2303, May 2012.
- [10] A. Drámeau, C. Herzet, and L. Daudet, "Boltzmann machine and meanfield approximation for structured sparse decompositions," *IEEE Trans. Signal Processing*, vol. 60, no. 7, pp. 3425–3438, July 2012.
- [11] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Processing*, vol. 61, no. 8, pp. 2009–2015, Apr. 2013.
- [12] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
- [13] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.
- [14] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [15] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Information Theory*, no. 12, pp. 4203–4215, Dec. 2005.