

COMPLEX MULTITASK BAYESIAN COMPRESSIVE SENSING

Qisong Wu[†], Yimin D. Zhang[†], Moeness G. Amin[†], and Braham Himed[‡]

[†] Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA

[‡] RF Technology Branch, Air Force Research Lab (AFRL/RVMD), WPAFB, OH 45433, USA

ABSTRACT

An effective complex multitask Bayesian compressive sensing (CMT-BCS) algorithm is proposed to recover sparse or group sparse complex signals. The existing multitask Bayesian compressive sensing (MT-CS) algorithm is powerful in recovering multiple real-valued sparse solutions. However, a large class of sensing problems deal with complex values. A simple approach, which decomposes a complex value into independent real and imaginary components, does not take into account the group sparsity of these two components and thus yields poor recovery performance. In this paper, we first introduce the CMT-BCS algorithm that jointly treats the real and imaginary components, and then derive a fast and accurate algorithm for the estimation of the prior parameters by solving a surrogate convex function. The proposed CMT-BCS algorithm achieves effective complex sparse signal recovery and outperforms MT-CS and complex group Lasso.

Index Terms— Compressive sensing, Bayesian inference, multitask learning, multiple measurement vector

1. INTRODUCTION

Sparse signal recovery and the associated compressive sensing (CS) problems have attracted significant attention in recent years [1]. CS techniques have the capability of recovering signals from a small number of measurement samples with a high probability, given that the signals are sparse or can be sparsely represented in some domain. They have been widely used in many applications, such as wireless sensing [2], direction-of-arrival (DOA) estimation [3, 4], radar imaging [5, 6, 7], space-time adaptive processing [8, 9], and time-frequency analysis [10, 11, 12].

A typical real-valued CS model with a single measurement vector (SMV) $\mathbf{y} \in \mathcal{R}^N$ is given by

$$\mathbf{y} = \Phi \mathbf{w} + \boldsymbol{\varepsilon}, \quad (1)$$

The work of Q. Wu, Y. D. Zhang, and M. G. Amin was supported in part by a subcontract with Defense Engineering Corporation for research sponsored by the Air Force Research Laboratory under Contract FA8650-12-D-1376 and by a subcontract with Dynetics, Inc. for research sponsored by the Air Force Research Laboratory under Contract FA8650-08-D-1303.

where $\Phi \in \mathcal{R}^{N \times M}$, $N \ll M$, is a known dictionary matrix, and $\boldsymbol{\varepsilon} \in \mathcal{R}^N$ is an unknown noise vector. The task is to estimate the weight vector $\mathbf{w} \in \mathcal{R}^M$. Any K columns of Φ are assumed to be linearly independent so that the unique representation property condition is satisfied to guarantee the sparse signal recovery with a high probability [13]. In addition, to ensure a unique global solution, the dimension of the measurement vector \mathbf{y} should meet the condition $N > O(K \log M/K)$, where K is defined as the sparsity, i.e., the number of non-zero entries in \mathbf{w} . By exploiting the fact that \mathbf{w} is sparse, one may estimate \mathbf{w} accurately by solving the ill-posed problem via an l_1 -regularized formulation,

$$\min \|\mathbf{w}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \Phi \mathbf{w}\|_2^2 \leq \epsilon. \quad (2)$$

The basic framework in (2) has been the starting point for several recent CS algorithms, such as the orthogonal matching pursuit (OMP) [14] and the least absolute shrinkage and selection operator (Lasso) [15].

In many applications, a sequence of measurements in $\{\mathbf{y}_l\}_{l=1,L}$ with $L > 1$ are taken from slow-varying sources or similar scenes. It is referred to as a multiple measurement vector (MMV) model. The mapping from $\{\mathbf{y}_l\}_{l=1,L} \rightarrow \{\mathbf{w}_l\}_{l=1,L}$ can be reformulated as a multitask CS problem. Numerous algorithms, such as the group OMP [16] and group Lasso [17], have been developed to solve such MMV problems by extending from the SMV algorithm counterparts. Bayesian based CS algorithms, which form a different CS class, have recently received great attention because they generally achieve much better reconstruction performance over those obtained by the Lasso and OMP based approaches [18]. The Bayesian CS [19] and its extended multitask CS (MT-CS) [20] algorithms, which are based on the relevance vector machine (RVM), constitute an important family in this class to recover sparse signals in the SMV and MMV models, respectively.

The above Bayesian algorithms have originally been designed to recover real-valued sparse solutions as encountered in image and video processing. However, a large class of sensing data deals with complex-valued sparse weights and observations. A simple approach to deal with such complex problems is to decompose a complex value into independent real and imaginary components [3]. However, such an approach does not utilize the fact that the real and imaginary

components are merely the projection of the same complex value into two orthogonal axes and thus share the same sparsity pattern. As such, it unnecessarily expands the dimension of sparsity entries and thus results in a degraded recovery performance.

In this paper, an effective complex multitask Bayesian compressive sensing (CMT-BCS) algorithm is proposed to recover complex signals in the MMV model, by exploiting the fact that the real and imaginary components of a complex value are likely to share the same sparsity pattern. We first introduce the complex multitask model, and then derive a fast and accurate algorithm for the estimation of the prior parameters based on the majorization-minimization principle to achieve automatic feature determination. Simulation studies are performed that verify that the proposed algorithm outperforms the MT-CS and complex group Lasso (CG-Lasso) for complex sparse signal recovery.

Notations: We use lower-case (upper-case) bold characters to denote vectors (matrices). $(\cdot)^T$ and $(\cdot)^H$, respectively, denote the transpose and conjugate transpose of a matrix or vector. $\text{diag}(\mathbf{x})$ denotes a diagonal matrix that uses the elements of \mathbf{x} as its diagonal elements. $\|\cdot\|$ denotes the Euclidean l_2 norm of a vector, whereas $\|\cdot\|_1$ denotes the l_1 norm. $P_r(\cdot)$ denotes the probability density function (pdf), and $\mathcal{N}(x|a, b)$ denotes that random variable x follows a Gaussian distribution with mean a and variance b . $\text{Re}(x)$ and $\text{Im}(x)$ denote the real and imaginary parts of complex element x , respectively.

2. COMPLEX MULTITASK FRAMEWORK BASED ON BAYESIAN REGRESSION FORMULATION

Assume L sets of complex CS measurements that are statistically correlated as described below. By extending the real-valued SVM model in (1) to a complex-value MMV model, we represent the complex measurements \mathbf{y}_l as

$$\mathbf{y}_l = \Phi_l \mathbf{w}_l + \varepsilon_l, \quad l \in [1, \dots, L], \quad (3)$$

where each complex measurement $\mathbf{y}_l \in \mathcal{C}^{N_l}$ employs its own projection matrix $\Phi_l \in \mathcal{R}^{N_l \times M}$. This generalizes the formulation in (1), wherein a common Φ is employed across all the L tasks. The weight vector \mathbf{w}_l that characterizes task l is assumed to be drawn from a product of the following zero-mean Gaussian distributions:

$$\mathbf{w}_{lm} \sim \mathcal{N}(\mathbf{w}_{lm} | \mathbf{0}, \alpha_m \mathbf{I}_2), \quad m \in [1, \dots, M], \quad (4)$$

where \mathbf{w}_{lm} is a vector that includes the real-part coefficient w_{lmR} and imagery-part coefficient w_{lmI} of the m th weight, α_m is the variance of Gaussian pdf, and \mathbf{I}_2 denotes the 2×2 identity matrix. It is important to note that the parameter $\alpha = \{\alpha_m\}_{m=1, M}$ are shared by all the L groups. It is in this sense that the L tasks are statistically correlated. The m th weight vector tends to be zero with probability 1 across the L groups when α_m is set to zero [18, 20].

To deal with complex problems in the real-valued MT-CS context, the real and imaginary components have been treated as two separate variables in the literature [3]. As described earlier, such an approach does not utilize the fact that, because the real and imaginary components are merely the projection of the same complex coefficients into two orthogonal axes, their non-zero entries usually appear at the same positions. In the proposed approach, on the other hand, the shared parameters α_m are jointly estimated for both real and imagery components. As a result, it clearly improves the sparsity of the estimated weight vectors and yields improved sparse signal recovery.

To promote sparsity over the weight vector \mathbf{w}_l , a Gamma prior is placed on α_m^{-1} , i.e.,

$$\alpha_m^{-1} \sim \text{Gamma}(\alpha_m^{-1} | a, b), \quad (5)$$

where $\text{Gamma}(x^{-1} | a, b) = \Gamma(a)^{-1} b^a x^{-(a+1)} e^{-\frac{b}{x}}$, and $\Gamma(\cdot)$ denotes the Gamma function.

It has been demonstrated in [18] that an appropriate choice of the parameters a and b encourages a sparse representation for the coefficients in \mathbf{w}_l . We set $a = b = 0$ as a default choice which avoids a subjective choice of a and b and leads to simplifications of computation.

Without loss of generality, a Gaussian prior is placed on the additive noise due to the existence of measurement noise, $\varepsilon_{lm} \sim \mathcal{N}(\varepsilon_{lm} | \mathbf{0}, \beta_0 \mathbf{I}_2)$, where ε_{lm} is a vector including the real-part and imaginary-part coefficients of the complex noise and β_0 is the noise variance. Similar to α , β_0^{-1} is placed on the Gamma prior with parameters c and d . We also let $c = d = 0$ as a default choice [18].

Assuming that the parameters α and β_0 are known (their estimation is discussed in the next section), the posterior density function for $\mathbf{w}_l = [w_{l1R}, \dots, w_{lMR}, w_{l1I}, \dots, w_{lMI}]^T$ can be evaluated analytically based on the Bayes' rule as,

$$\Pr(\mathbf{w}_l | \mathbf{y}_l, \Phi, \alpha, \beta_0) = \mathcal{N}(\mathbf{w}_l | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l),$$

where

$$\boldsymbol{\mu}_l = \beta_0^{-1} \boldsymbol{\Sigma}_l \boldsymbol{\Psi}_l^T \mathbf{y}_l, \quad (6)$$

$$\boldsymbol{\Sigma}_l = \left[\beta_0^{-1} \boldsymbol{\Psi}_l^T \boldsymbol{\Psi}_l + \mathbf{A}^{-1} \right]^{-1}, \quad (7)$$

$$\boldsymbol{\Psi}_l = \begin{bmatrix} \text{Re}(\Phi_l) & -\text{Im}(\Phi_l) \\ \text{Im}(\Phi_l) & \text{Re}(\Phi_l) \end{bmatrix}, \quad (8)$$

$$\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_M, \alpha_{M+1}, \dots, \alpha_{2M}), \quad (9)$$

$$\alpha_l = \alpha_{l+M}, \quad l \in [1, \dots, M], \quad (10)$$

with $\mathbf{y}_l = [\text{Re}(\mathbf{y}_l), \text{Im}(\mathbf{y}_l)]^T$. Comparing to the complex MMV model used in [3], which handles complex-value problem by dividing a complex weight into independent real and imaginary components, the proposed approach reduces the sparsity by a factor of two by imposing a reasonable restriction on shared parameter α in the proposed algorithm.

3. BAYESIAN PARAMETER ESTIMATION

With known α and β_0 , the mean and covariance of each scattering coefficients can be derived from (6) and (7). The associated *learning* problem, in the context of the RVM, thus becomes the search for the parameters α and β_0 . The empirical Bayesian estimate for α and β_0 is determined by maximizing the marginal likelihood, or equivalently, its logarithm

$$\{\alpha, \beta_0\} = \arg \max_{\alpha, \beta_0} \mathcal{L}(\alpha, \beta_0), \quad (11)$$

where

$$\begin{aligned} \mathcal{L}(\alpha, \beta_0) &= \sum_{l=1}^L \log P_r(\mathbf{w}_l | \alpha, \beta_0) \\ &= -\frac{1}{2} \sum_{l=1}^L [2N_l \log(2\pi) + \log |\mathbf{C}_l| + \mathbf{y}_l^T \mathbf{C}_l^{-1} \mathbf{y}_l], \end{aligned} \quad (12)$$

and $\mathbf{C}_l = \beta_0 \mathbf{I} + \Psi_l \mathbf{A} \Psi_l^T \in R^{2N_l \times 2N_l}$, $N_l \ll M$. A type-II maximum likelihood (ML) approximation [21] employs the point estimates for β_0 and α to maximize (12), which can be implemented via the expectation maximization (EM) algorithm to yield

$$\alpha_m^{(\text{new})} = \frac{1}{L} \sum_{l=1}^L \{\mu_{l,m}^2 + \mu_{l,M+m}^2 + \Sigma_{l,mm} + \Sigma_{l,(M+m)(M+m)}\}, \quad (13)$$

$$\beta_0^{(\text{new})} = \frac{\sum_{l=1}^L \left\{ \text{Tr}[\Sigma_l \Psi_l \Psi_l^T] + \|\mathbf{y}_l - \Psi_l \boldsymbol{\mu}_l\|_2^2 \right\}}{2ML}. \quad (14)$$

Note that $\alpha^{(\text{new})}$ and $\beta_0^{(\text{new})}$ are functions of $\{\boldsymbol{\mu}_l\}_{l=1,L}$ and $\{\Sigma_l\}_{l=1,L}$, while $\{\boldsymbol{\mu}_l\}_{l=1,L}$ and $\{\Sigma_l\}_{l=1,L}$ are functions of α and β_0 . This suggests an iterative algorithm, which iterates (6), (7) and (13), (14), until a convergence criterion is satisfied.

While benefitting from the general convergence properties of the EM algorithm, we observe this updated rule of α to be very slow on large scale problems. When evaluating (7) one must invert matrices of size $2M \times 2M$, which is an $\mathcal{O}(M^3)$ operations for each of the L groups of data. It motivates the development of fast algorithms. Consider the fact that the marginal likelihood is not a convex function with respect to α , an *approximate* expression of α was developed in [20]. In this paper, we derive a fast and accurate learning rule of α by solving a surrogate convex function based on the majorization-minimization method [22, 23], in which the supporting hyperplane of the second term $\log |\mathbf{C}_l|$ in (12) is

used. Let α^* be a given point in the α -space. Then,

$$\begin{aligned} \mathcal{L}(\alpha) &\triangleq \log |\beta_0 \mathbf{I} + \Psi_l \mathbf{A} \Psi_l^T| \\ &\leq \log |\mathbf{C}_l^*| + \sum_{m=1}^M \text{Tr}[(\mathbf{C}_l^*)^{-1} \Psi_{lm}^T \Psi_{lm} \\ &\quad + (\mathbf{C}_l^*)^{-1} \Psi_{l,m+M}^T \Psi_{l,m+M}] (\alpha_m - \alpha_m^*) \triangleq \tilde{\mathcal{L}}(\alpha), \end{aligned} \quad (15)$$

where $\mathbf{C}_l^* = \beta_0 \mathbf{I} + \Psi_l \mathbf{A}^* \Psi_l^T$, $\mathbf{A}^* = \text{diag}(\alpha_1^*, \dots, \alpha_M^*, \alpha_1^*, \dots, \alpha_M^*)$, and $\Psi_{lm}, \Psi_{l,m+M}$, being the m th and $(m+M)$ th columns in Ψ_l . Denote the corresponding real and imaginary atoms of the m th weight vector $\mathbf{w}_{lm} = [\mathbf{w}_{lmR}, \mathbf{w}_{lmI}]^T$, respectively. The function $\tilde{\mathcal{L}}(\alpha)$ in (15) is convex over α , and when $\alpha = \alpha^*$, we have $\mathcal{L}(\alpha^*) = \tilde{\mathcal{L}}(\alpha^*)$ [23]. Further, for any α_{\min} which minimizes $\tilde{\mathcal{L}}(\alpha)$, we have the following relationship: $\mathcal{L}(\alpha_{\min}) \leq \tilde{\mathcal{L}}(\alpha_{\min}) \leq \tilde{\mathcal{L}}(\alpha^*)$ [23]. Further, based on the identity property we have a surrogate function for the third term in (12),

$$\begin{aligned} \mathbf{y}_l^T [\beta_0 \mathbf{I} + \Psi_l \mathbf{A} \Psi_l^T]^{-1} \mathbf{y}_l &\triangleq \min_{\mathbf{w}_l} \beta_0^{-1} \|\mathbf{y}_l - \Psi_l \mathbf{w}_l\|_2^2 \\ &\quad + \mathbf{w}_l^T \mathbf{A}^{-1} \mathbf{w}_l. \end{aligned} \quad (16)$$

Substituting (15) and (16) into (12), we can define a new function as

$$\begin{aligned} \mathcal{G}(\alpha, \{\mathbf{w}_l\}_{l=1,L}) &\triangleq \sum_{l=1}^L \beta_0^{-1} \|\mathbf{y}_l - \Psi_l \mathbf{w}_l\|_2^2 + \mathbf{w}_l^T \mathbf{A}^{-1} \mathbf{w}_l \\ &\quad + \log |\mathbf{C}_l^*| + \sum_{m=1}^M \text{Tr}[(\mathbf{C}_l^*)^{-1} \Psi_{lm}^T \Psi_{lm} \\ &\quad + (\mathbf{C}_l^*)^{-1} \Psi_{l,m+M}^T \Psi_{l,m+M}] (\alpha_m - \alpha_m^*). \end{aligned} \quad (17)$$

It is important to note that $\mathcal{G}(\alpha, \{\mathbf{w}_l\}_{l=1,L})$ is convex in both $\{\mathbf{w}_l\}_{l=1,L}$ and α . It can be easily shown that the solution α of $\tilde{\mathcal{L}}(\alpha)$ in (12) is the solution $(\alpha, \{\mathbf{w}_l\}_{l=1,L})$ of $\mathcal{G}(\alpha, \{\mathbf{w}_l\}_{l=1,L})$ in (15). Thus, $\mathcal{G}(\alpha, \{\mathbf{w}_l\}_{l=1,L})$ is our final cost function. Setting the derivative of \mathcal{G} with respect to α_m to zero, we obtain

$$\alpha_m^{(\text{new})} = \sqrt{\frac{\sum_{l=1}^L (\mathbf{w}_{lm}^T \mathbf{w}_{lm} + \mathbf{w}_{l,m+M}^T \mathbf{w}_{l,m+M})}{\sum_{l=1}^L \text{Tr}[(\mathbf{C}_l^*)^{-1} (\Psi_{lm}^T \Psi_{lm} + \Psi_{l,m+M}^T \Psi_{l,m+M})]}}. \quad (18)$$

Due to this learning rule for α , the proposed fast algorithm requires far fewer iterations than the one in the EM algorithm (13). We acquire the accurate α expression by solving a surrogate convex function.

4. EXPERIMENTS

In the following experiments, the normalized mean square error (NMSE) $\|\hat{\mathbf{w}} - \mathbf{w}_{\text{gen}}\|_2^2 / \|\mathbf{w}_{\text{gen}}\|_2^2$ is used as the performance index, where $\hat{\mathbf{w}}$ is the estimate of the true signal \mathbf{w}_{gen} .

We consider that an exact reconstruction is achieved when $\text{NMSE} < 10^{-5}$.

4.1. Recovery profiles comparison

In the first experiment, we consider $L=2$ signals of length $M = 256$, each containing 20 unit-norm complex spikes with random locations. All spikes are assumed to have random real and imaginary components of $\pm 1/\sqrt{2}$. The projection matrix Φ_l is created from an independent and identically distributed complex Gaussian distribution. Complex zero-mean Gaussian noise with standard deviation of $\eta = 0.005$ is added to data \mathbf{y}_l . In the experiment, we adopt $N_l = 45$ for $l = [1, 2]$.

In Fig. 1, the left half shows the real component, whereas the right half shows the imaginary component. Only the first signal is shown due to space limitation. The recovered results are compared among the proposed CMT-BCS, unpaired MT-BCS (which represents the result of MT-BCS when the real and imaginary components are treated separately and the parameters α are estimated based on the majorization-minimization method), MT-CS (which represents the result of the algorithm in [20], where the real and imaginary components are treated separately and the parameters α are provided by the approximate expression in [20]), complex group Lasso (in which the real and imaginary components are treated jointly). The proposed CMT-BCS algorithm exactly reconstructs the original complex-value signals. However, both the unpaired MT-BCS and MT-CS algorithms show poor recovery results, because they ignored the identical sparsity pattern between the real and imaginary components. It is also evident that the proposed algorithm outperforms the CG-Lasso even though they both treat the real and imaginary components jointly because the Bayesian sparse learning based on RVM provides a tighter approximation to l_0 -norm sparsity measure than the l_1 -norm used in Lasso [20].

4.2. Recovery performance versus data length

To study how the data length of the observed data affects the recovery performance, in the second experiment we use the same parameters as those in the first experiment except the vector length. In the following, the experiment in each dimension was repeated for 100 trials. The average NMSE performance is compared in Fig. 2. It is evident that the CMT-BCS algorithm is able to exactly recover the signal when the data length is higher than 40, whereas the requisite lengths are 56 and 60, respectively, for the unpaired MT-BCS and MT-CS. For the CG-Lasso algorithm, the recovery errors has a floor of around 10^{-2} . It is evident that the proposed algorithm outperforms the other algorithms compared herein.

5. CONCLUSION

Motivated by the applications that deal with complex data, as commonly encountered in radio frequency sensing problems, we proposed a complex Bayesian compressive sensing

algorithm to effectively recover sparse weights described by the multiple measurement vector model. An accurate expression of the prior parameters is derived by solving a surrogate convex function. We have shown that the proposed algorithm outperforms existing methods, including the simple extension of the original multitask Bayesian compressive sensing algorithm to complex problems and the complex group Lasso.

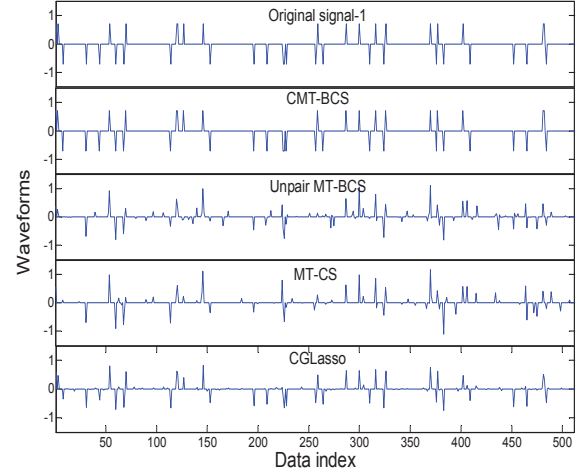


Fig. 1. Comparison of the recovered results for signal 1.

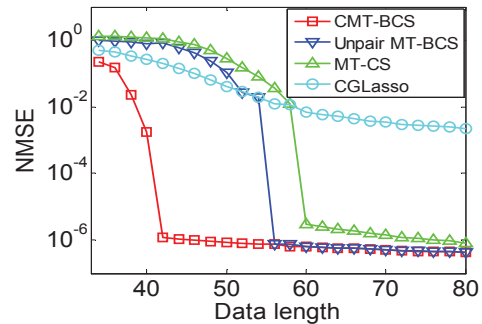


Fig. 2. Recovery performance versus the number of dimensions.

6. REFERENCES

- [1] M. Elad, *Sparse and Redundant Representations*. Springer, 2010.
- [2] Q. Wu, M. Xing, C. Qiu, B. Liu, Z. Bao and T. Yeo, "Motion parameter estimation in the SAR System with low PRF sampling," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 450–454, 2010.
- [3] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, "Directions-of-arrival estimation through bayesian compressive sensing strategies," *IEEE Trans. Antennas Propagat.*, vol. 61, no. 7, pp. 3828–3838, 2013.
- [4] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays, in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013.

- [5] A. Gurbuz, J. McClellan, and W. Scott, "A compressive sensing data acquisition and imaging method for stepped frequency GPRs," *IEEE Trans. Signal Proc.*, vol. 57, no. 7, pp. 2640–2650, 2009.
- [6] M. Amin and F. Ahmad, "Compressive sensing for through-the wall radar imaging," *J. Electronic Imaging*, vol. 22, no. 3, 2013.
- [7] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Multi-static passive radar SAR imaging based on Bayesian compressive sensing," in *Proc. SPIE Compressive Sensing Conf.*, Baltimore, MD, May 2014.
- [8] I. W. Selesnick, S. U. Pillai, K. Y. Li, and B. Himed, "Single-sample STAP using sparsity," in *Proc. U.S./Australia Joint Workshop on Defense Applications of Signal Processing*, Lihue, HI, 2009.
- [9] Y. D. Zhang and B. Himed, "Space-time adaptive processing in bistatic passive radar exploiting complex bayesian learning," in *Proc. IEEE Radar Conf.*, Cincinnati, OH, May, 2014.
- [10] Y. D. Zhang and M. G. Amin, "Compressive sensing in non-stationary array processing using bilinear transforms," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Hoboken, NJ, June 2012.
- [11] Y. D. Zhang, M. G. Amin, and B. Himed, "Reduced interference time-frequency representations and sparse reconstruction of undersampled data," in *Proc. European Signal Proc. Conf.*, Marrakech, Morocco, Sept. 2013.
- [12] M. G. Amin and Y. D. Zhang, "Time-frequency signature reconstruction from random observations using multiple measurement vectors," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014.
- [13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [14] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [15] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Royal Statist. Soc.*, vol. 58, no. 1, pp. 267–288, 1996.
- [16] J. Huang and T. Zhang, "The benefit of group sparsity," *Ann. Statist.*, vol. 38, no. 4, pp. 1978–2004, 2010.
- [17] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statist. Soc. Serie B*, vol. 68, no. 1, pp. 49–67, 2006.
- [18] M. E. Tipping, "Sparse Bayesian shrinkage and selection learning and the relevance vector machine," *J. Machine Learning Research*, vol. 1, no. 9, pp. 211–244, 2001.
- [19] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Proc.*, vol. 56, no. 6, pp. 2346–C2356, 2008.
- [20] S. Ji, D. Dunson, and L. Carin, "Multi-task compressive sampling," *IEEE Trans. Signal Proc.*, vol. 57, no. 1, pp. 92–106, 2009.
- [21] D. J. C. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992.
- [22] Z. Zhang and B. D. Rao, "Extension of SBL algorithm for the recovery of block sparse signals with inter-block correlation," *IEEE Trans. Signal Proc.*, vol. 92, no. 7, pp. 1580–1590, 2012.
- [23] P. Stoica and P. Babu, "Spice and likes: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Proc.*, vol. 61, no. 8, pp. 2009–2015, 2013.