Weighted Block Sparse Bayesian Learning for Basis Selection

Ahmed Al Hilli[1,2] and Athina Petropulu[1]

[1]Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ 08854, USA
[2]Al-Furat Al-Awsat Technical University, Engineering Technical College of Al-Najaf, Najaf, Iraq

***Abstract*— Block Sparse Bayesian Learning (BSBL) methods estimate a block sparse vector by maximizing the posterior distribution and using sparsity-inducing priors. In BSBL works, all hyperparameters priors are assumed to follow the same distribution with the same parameters. In this paper, we propose to assign different parameters to each hyperparameter, giving more importance to some hyperparameters over others. The importance weights are obtained by leveraging a low resolution estimate of the underlying sparse vector, for example, an estimate obtained via a method that does not encourage sparsity. We refer to the proposed approach as Weighted Block Sparse Bayesian Learning (WBSBL). Simulation results show that, as compared to BSBL, WBSBL achieves substantial improvement in terms of probability of detection and probability of false alarm in the low signal to noise ratio regime. Also, WBSBL's performance degrades slower than that of BSBL as the number of active blocks increases.**

I. Introduction

Sparse signal recovery problems arise in many contexts, including biomedical imaging [1]–[5], and radar [6]–[10]. In such problems, we need to estimate a vector with the minimum number of active entries that satisfies certain constrains. Mathematically, this corresponds to finding the least ℓ_0-norm solution. However, since this is an NP-hard problem [11], a lower complexity ℓ_1-norm minimization problem is solved instead. Conditions under which the ℓ_0 and ℓ_1-norm minimization problems are strictly equivalent include the Restricted Isometry Property (RIP) [12], the Null Space Property (NSP) [13], the Mutual Coherence [14], and the Range Space Property [15]. Weighted approaches have also been proposed for sparse signal recovery. In [16], a reweighted ℓ_1-norm approach has been proposed for sparsity enhancement of the recovered vector. Also, a weighted ℓ_1-norm approach has been proposed in [2] for the cases in which the dictionary matrix exhibits high coherence. Probabilistic approaches for sparse signal recovery have also been proposed, where a Bayesian posterior is maximized, using sparsity inducing priors. In Sparse Bayesian Learning (SBL) [17], [18], Gaussian priors with distinct variances for each entry are used. The variances are estimated by maximizing the marginal likelihood function. A weighted version of SBL (WSBL) was proposed in [19] and shown to improve the performance of SBL under low SNR scenarios. Bayesian approaches have a global minimum, which, unlike ℓ_1-norm minimzation based approaches, is the sparsest solution in noise free scenarios [18].

Block sparse signals constitute an interesting class of signals in which groups of entries are active simultaneously. In block sparse signal recovery problems, we seek a solution with the smallest number of active groups that best describe the observations vector. Since this is a complex problem, a relaxation is proposed in [20], where we seek the smallest sum of group energies. Conditions for equivalence between the original problem and the relaxed one include the Generalized RIP condition [20], the Null Space Characterization [21], the Block Mutual Coherence [22], and the Generalized Range Space Property [23]. Bayesian approaches have also been proposed for group sparse probems by generalizing SBL [24], [24]. As in SBL, the Bayesian approach global minimum is the sparsest solution in noise free scenarios [24], which is not the case, in general, for approaches that solve for the smallest sum of group energies.

Motivated by the good performance of WSBL as compared to SBL, in this paper, we propose a weighted approach to recover block sparse signals. The weights are estimated using a low resolution estimate of the underlying signal. The proposed approach shows robustness in low SNR scenarios, and its performance degrades slower than that of BSBL as the number of active blocks increases.

The paper is organized as follows. Section II provides some background on BSBL as proposed in [24] and [25]. Section III introduces the proposed WBSBL approach, Section IV simulation results, while Section VII provides concluding remarks.

II. Overview of BSBL

Consider the following linear system

$$\mathbf{y} = \mathbf{G}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{G} \in R^{M\times N}$ with $M \ll N$ is the dictionary, or sensing matrix, $\mathbf{y} \in R^{M\times 1}$ is the observation vector, $\mathbf{n} \in R^{M\times 1}$ is the noise vector, and $\mathbf{x} \in R^{N\times 1}$ is a block sparse vector to be estimated. Assume each block $\mathbf{x}_i \in R^{d_i\times 1}$ in $\mathbf{x}$ follows a parametrized multivariate Gaussian distribution, i.e.,

$$p(\mathbf{x}_i; g_i, \mathbf{B}_i) \sim \mathcal{N}(0, g_i\mathbf{B}_i), \tag{2}$$

where g_i is a non-negative parameter that controls the block sparsity of $\mathbf{x}$ (i.e., $g_i > 0$ for active blocks, and $g_i = 0$ for

The work was supported by NSF under Grant ECCS-1408437

non-active blocks), and $\mathbf{B}_i$ is a positive definite matrix which describes the correlation between the block entries. Assuming independence between the blocks, $p(\mathbf{x})$ can be written as $p(\mathbf{x}) \sim \mathcal{N}(0, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0 = \text{diag}\{g_1\mathbf{B}_1, ..., g_m\mathbf{B}_m\}$. Assuming white Gaussian noise, i.e., $n \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, the posterior of $\mathbf{x}$ is [25]

$$p(\mathbf{x}; \mathbf{y}, \sigma^2, \{g_i, \mathbf{B}_i\}_{i=1}^m) = \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{\Sigma}_x), \tag{3}$$

where

$$\boldsymbol{\mu}_x = \mathbf{\Sigma}_0\mathbf{G}^T(\sigma^2\mathbf{I} + \mathbf{G}\mathbf{\Sigma}_0\mathbf{G}^T)^{-1}\mathbf{y}, \tag{4}$$

and

$$\mathbf{\Sigma}_x = \left(\mathbf{\Sigma}_0^{-1} + \sigma^{-2}\mathbf{G}^T\mathbf{G}\right)^{-1}. \tag{5}$$

Given the parameters σ^2 and $\{g_i, \mathbf{B}_i\}_{i=1}^m$, the Maximum a Posteriori (MAP) estimate of $\mathbf{x}$ is

$$\hat{\mathbf{x}} = \boldsymbol{\mu}_x. \tag{6}$$

A type II maximum likelihood procedure can be used to estimate the parameters σ^2 and $\{g_i, \mathbf{B}_i\}_{i=1}^m$ [17], which is equivalent to minimizing the following cost function [25]:

$$\begin{aligned} L(\sigma^2, \{g_i, \mathbf{B}_i\}_{i=1}^m) = \log|\sigma^2\mathbf{I} + \mathbf{G}\mathbf{\Sigma}_0\mathbf{G}^T| + \\ \mathbf{y}^T(\sigma^2\mathbf{I} + \mathbf{G}\mathbf{\Sigma}_0\mathbf{G}^T)^{-1}\mathbf{y}. \end{aligned} \tag{7}$$

Differentiating L w.r.t. g_i, σ^2, and $\mathbf{B}_i$, and equating to zero we get

$$g_i = \frac{1}{d_i}\text{Tr}[\mathbf{B}_i^{-1}(\mathbf{\Sigma}_x^i + \boldsymbol{\mu}_x^i(\boldsymbol{\mu}_x^i)^T)], \;\; i = 1, 2, ..., m, \tag{8}$$

$$\sigma^2 = \frac{\|\mathbf{y} - \mathbf{G}\boldsymbol{\mu}_x\|_2 + \text{Tr}[\mathbf{\Sigma}_x\mathbf{G}^T\mathbf{G}]}{M}, \tag{9}$$

and

$$\mathbf{B}_i = \frac{1}{m}\sum_{i=1}^{m}\frac{\mathbf{\Sigma}_x^i + \boldsymbol{\mu}_x^i(\boldsymbol{\mu}_x^i)^T}{g_i}, \tag{10}$$

respectively, where $\boldsymbol{\mu}_x^i$ is the i^{th} block in $\boldsymbol{\mu}_x$, $\mathbf{\Sigma}_x^i$ is the corresponding i^{th} principal diagonal block in $\mathbf{\Sigma}_x$, and d_i is the length of the i^{th} block. Note that in BSBL, most of g_is tend to be zero, thus resulting in a block sparse estimate. BSBL is a recursive approach; in each iteration, the parameters $\{g_i, \mathbf{B}_i\}_{i=1}^m$ and σ^2 are estimated, and the g_is that are below a small threshold (around zero) are excluded in the next iteration. Given the parameters $\{g_i, \mathbf{B}_i\}_{i=1}^m$ and σ^2, $\boldsymbol{\mu}_x$ and $\mathbf{\Sigma}_0$ are calculated using (4) and (5), respectively. The algorithm stops when $\boldsymbol{\mu}_x$ converges.

III. The Proposed Approach

In the the proposed approach, we follow the BSBL idea, except that we consider $\alpha_i = \frac{1}{g_i}$ as a random variable. Since we know that α_i should be positive, we model α_i as a Gamma distribution with parameters a_i and b_i, i.e.,

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{m}\text{Gamma}(\alpha_i; a_i, b_i), \tag{11}$$

where $\text{Gamma}(\alpha, a, b) = \Gamma(a)^{-1}b^a\alpha^{a-1}e^{-b\alpha}$, and $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ is the Gamma function. Using a Type II maximum likelihood procedure as in BSBL, the cost function to be minimized, after dropping the irrelevant terms, can be written as

$$\begin{aligned} L(\sigma^2, \{g_i, \mathbf{B}_i\}_{i=1}^m) = \log|\sigma^2\mathbf{I} + \mathbf{G}\mathbf{\Sigma}_0\mathbf{G}^T| + \\ \mathbf{y}^T(\sigma^2\mathbf{I} + \mathbf{G}\mathbf{\Sigma}_0\mathbf{G}^T)^{-1} + 2\sum_{i=1}^{m}\frac{b_i}{g_i} + 2\sum_{i=1}^{m}a_i\log(g_i). \end{aligned} \tag{12}$$

Differentiating w.r.t. g_i, σ^2, and $\mathbf{B}_i$, we get

$$g_i = \frac{\text{Tr}[\mathbf{B}_i^{-1}(\mathbf{\Sigma}_x^i + \boldsymbol{\mu}_x^i(\boldsymbol{\mu}_x^i)^T)] + 2b_i}{d_i + 2a_i} \tag{13}$$

and σ^2 and $\mathbf{B}_i$ are as described in (9) and (10), respectively. Note that the update rule of the weighted approach in (13) has the parameters a_i and b_i. One can use these parameters to give importance to some g_is. The relative importance can be determined by some rough estimate of the underlying sparse vector.

Now, suppose we have a weight vector $\mathbf{w}$, which contains large values corresponding to active $\mathbf{x}_i$ blocks, and low values corresponding to non active blocks in $\mathbf{x}$. Let us assign $a_i = \frac{1}{w_i}$ and $b_i = w_i$. Assuming that $w_i \neq 0$, the final update rule for g_i can be written as

$$g_i = \frac{\text{Tr}[\mathbf{B}_i^{-1}(\mathbf{\Sigma}_x^i + \boldsymbol{\mu}_x^i(\boldsymbol{\mu}_x^i)^T)^T] + 2w_i}{d_i + \frac{2}{w_i}}. \tag{14}$$

We call the above recursive approach Weighted Block Sparse Bayesian Learning (WBSBL). In each iteration of WBSBL, the parameters $\{g_i, \mathbf{B}_i\}_{i=1}^m$ and σ^2 are estimated via (13), (10) and (9), respectively. The α_is that are larger from a predefined threshold are excluded from the next iteration. Given the parameters $\{g_i, \mathbf{B}_i\}_{i=1}^m$ and σ^2, $\boldsymbol{\mu}_x$ and $\mathbf{\Sigma}_0$ are calculated using (4) and (5), respectively. The algorithm stops when $\boldsymbol{\mu}_x$ converges, or some other criterion is satisfied.

The weight vector used in this approach can be any rough or blurred estimate of the underlying sparse vector, and all the entries should be made non-zero to avoid losing potentially important components in the recovered vector [26]. The threshold

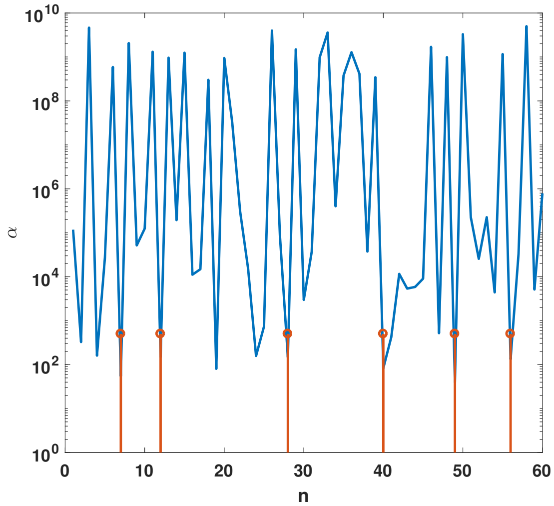


Fig. 1. The values of α in log scale after convergence for the non-weighted approach. The red lines show the indices of the true active blocks.

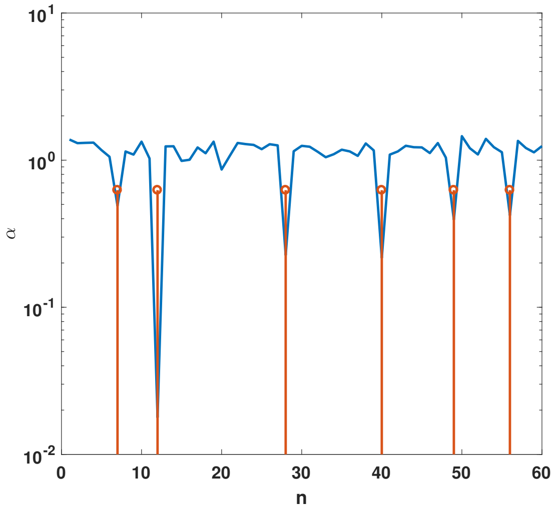


Fig. 2. The values of α in log scale after convergence for the weighted approach. The red lines show the indices of the true active blocks.

that is used to exclude small g_is depends on the weights. One can see from (14) that after convergence of WBSBL, the values of α_is are bounded between 0 and $\frac{d_i+2/w_i}{2w_i}$, and the threshold should belong to this interval.

In the following, we show through an example, how the hyperparameters are distributed after convergence for both BSBL and WBSBL. The dictionary matrix $\mathbf{A}$ of size 60×120 is constructed by choosing its entries to follow Gaussian distribution with zero mean and unit variance. The number of active blocks is set to 6, each with block size of 2. The non-zero entries of the block sparse vector follow Gaussian distribution of mean 5 and variance 0.25. The weights that are used in this example is MUltiple SIgnal Classification (MUSIC) estimated based on 100 snapshots. The SNR is set to be 15 dB. Fig. (1) shows the values of $\alpha_i = \frac{1}{g_i}$s of BSBL after convergence. It is obvious that the values of α_is that are associated with active blocks, along with other non-active blocks, have small values, and are considered in the final estimation. Also, one can observe large variance among the values of the non-active blocks of α_is; this makes it difficult to choose a threshold to distinguish between active and non-active blocks. Fig. (2) shows the values of α_is in log scale after convergence for WBSBL. One can see that the active and non-active blocks have been completely separated, and only the true active blocks have small values; those blocks will be considered in the final estimate. Also, in WBSBL, there is an upper limit on the values of α_i that correspond to non-active blocks with low variance among these α_is. This behavior makes choosing the threshold easier than in BSBL.

IV. Simulation Results

In this section, we provide simulation results for the proposed approach, and compare the performance of WBSBL and BSBL. Monte Carlo simulations with 1000 trials were performed. In each trial, a matrix $\mathbf{A}$ of size 64×120 with Gaussian distributed entries with zero mean and unit variance was constructed. k blocks of size d were randomly selected as active blocks, and the value of the entries in the selected blocks were set to follow the Gaussian distribution with mean 5 and standard deviation 0.25. White Gaussian noise was added to $\mathbf{Ax}$ at various SNR levels. The Receiver Operating Characteristics (ROC) graph was used to compare the performance of WBSBL to BSBL. For the cases with more than one active blocks, successful detection was claimed if all the active blocks were detected. The weights were constructed based on the MUSIC estimate [27], constructed using 100 snapshots. The threshold for BSBL was set to 0.001, while for WBSBL, the threshold was set to $0.75\frac{2w_{\min}}{d_i+2/w_{\min}}$, where $w_{\min}$ is the smallest non-zero entry in the weighting vector, i.e., MUSIC.

Fig. 3 shows the ROC curves of MUSIC (green curve), BSBL (blue curve) and WBSBL (red curve) for SNR=10 and 5, and for $k = 3$ active blocks of size 2. One can see that WBSBL improves significantly upon the low resolution estimate used for constructing the weights. Also, one can see that both BSBL and WBSBL have comparative performance in the case of high SNR scenarios. Fig. 4 considers the same scenario but at SNR = 4, 2, and 0. One can see that the performance of BSBL drops dramatically under this low SNR, while WBSBL remains robust. The performance of BSBL, MUSIC, and WBSBL as function of the number of active blocks k, with block size of $d = 2$ is shown in Fig. 5. One can see that WBSBL degrades slower than BSBL as the number of active blocks increases. In summary, WBSBL shows improved performance as compared to BSBL in cases of low SNR regimes, and with different number of active blocks.

V. Conclusions

Weighted Block Sparse Bayesian Learning approach has been proposed, which assigns distinct variance priors to each block in the block sparse vector, giving some hyperparameters some importance over the others. The importance of a specific parameter is obtained based on rough estimate of the underlying block sparse vector. Simulation results have shown significant improvement in terms of probability of detection and probability of false alarm, especially at low SNR scenarios, as compared to BSBL. WBSBL degrades slower as the number of active block increased, as compared to BSBL.

References

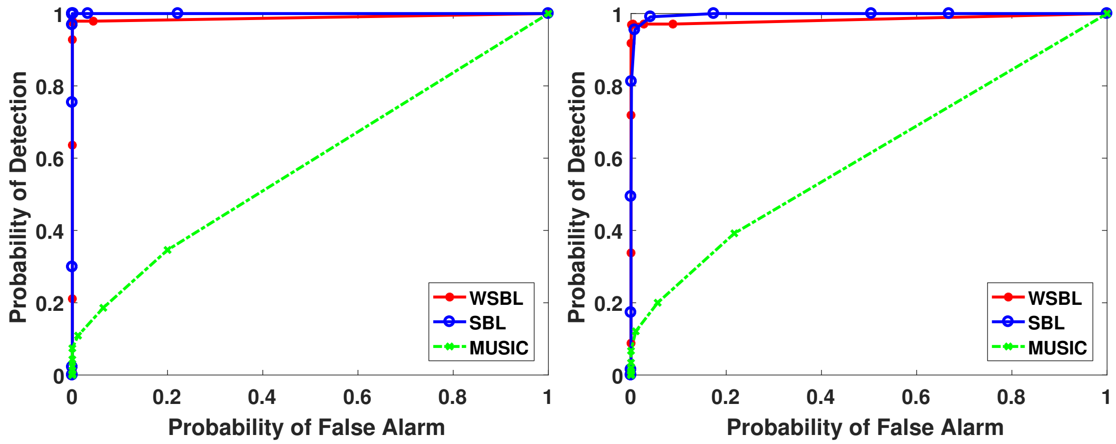
[1] S. C. Wu and A. L. Swindlehurst, "Matching pursuit and source deflation for sparse EEG/MEG dipole moment estimation," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2280–2288, 2013.

[2] A. Al Hilli, L. Najafizadeh, and A. Petropulu, "EEG sparse source localization via range space rotation," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*. IEEE, 2015, pp. 265–268.

[3] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing mri," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.

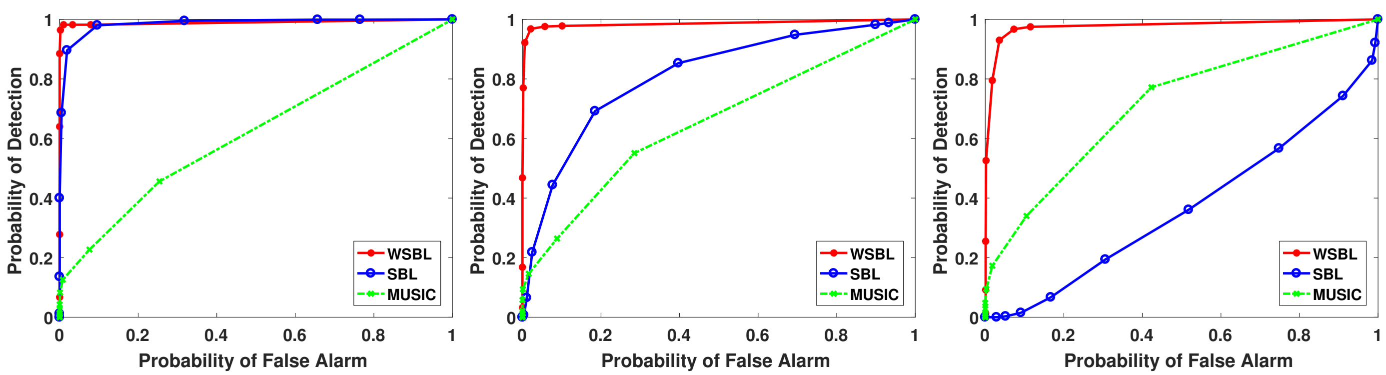
[4] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE/SP 14th Workshop Stat. Signal Process*, 2007, pp. 181–184.

[5] J. P. Haldar, D. Hernando, and Z.-P. Liang, "Compressed-sensing mri with random encoding," *IEEE transactions on Medical Imaging*, vol. 30, no. 4, pp. 893–903, 2011.



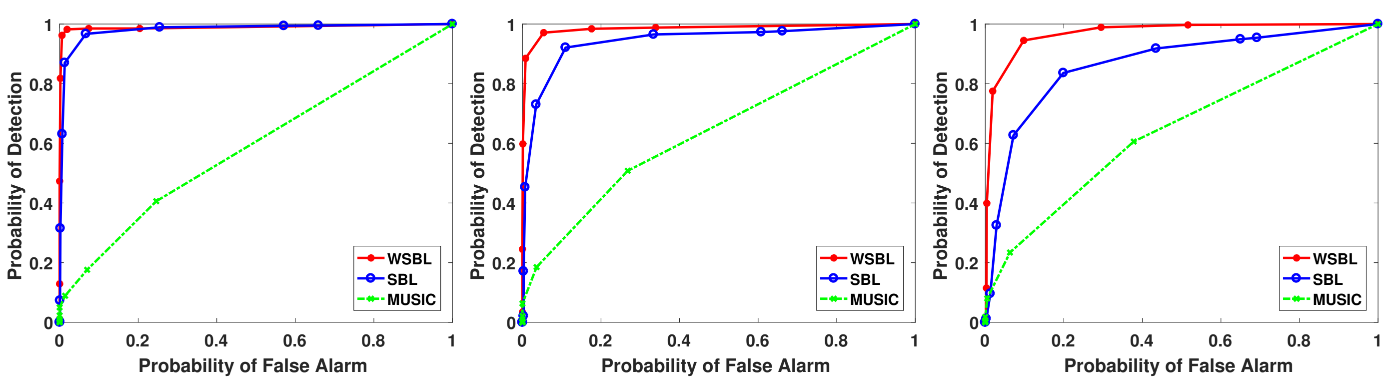
(a) (b)

Fig. 3. ROC curves of BSBL and WBSBL for k=3 sources, block size of 2, and a) SNR=10 dB ,b) SNR=5 dB.



(a) (b) (c)

Fig. 4. ROC curves for k=3 sources, block size of 2, and a) SNR=4 dB ,b) SNR=2 dB , c) SNR=0 dB.



(a) (b) (c)

Fig. 5. ROC curves for SNR=5, block size of 2, and a) $k = 4$, b) $k = 5$ dB, c) $k = 6$.

[6] Y. Yu, A. P. Petropulu, and H. V. Poor, “MIMO radar using compressive sampling,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 146–163, 2010.

[7] A. Al Hilli, L. Najafizadeh, and A. Petropulu, “Sparse target scene reconstruction for SAR using range space rotation,” in *2016 IEEE Radar Conference (RadarConf)*. IEEE, 2016, pp. 1–5.

[8] M. Herman and T. Strohmer, “Compressed sensing radar,” in *2008 IEEE Radar Conference*. IEEE, 2008, pp. 1–6.

[9] J. H. Ender, “On compressive sensing applied to radar,” *Signal Processing*, vol. 90, no. 5, pp. 1402–1414, 2010.

[10] M. T. Alonso, P. López-Dekker, and J. J. Mallorquí, “A novel strategy for radar imaging based on compressive sensing,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4285–4295, 2010.

[11] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.

[12] E. J. Candes and T. Tao, “Decoding by linear programming,” *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, 2005.

[13] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best k-term approximation,” *Journal of the American mathematical society*, vol. 22, no. 1, pp. 211–231, 2009.

[14] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[15] Y.-B. Zhao, “Rsp-based analysis for sparsest and least-norm solutions to underdetermined linear systems,” *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5777–5788, 2013.

[16] E. J. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted 1 minimization,” *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[17] M. E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[18] D. P. Wipf and B. D. Rao, “Sparse bayesian learning for basis selection,” *Signal Processing, IEEE Transactions on*, vol. 52, no. 8, pp. 2153–2164, 2004.

[19] A. Al Hilli, L. Najafizadeh, and A. Petropulu, “Weighted sparse bayesian learning (WSBL) for basis selection in linear underdetermined systems,” in *Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), 2016 4th International Workshop on*. IEEE, 2016, pp. 115–119.

[20] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5302–5316, 2009.

[21] M. Stojnic, F. Parvaresh, and B. Hassibi, “On the reconstruction of block-sparse signals with an optimal number of measurements,” *Signal Processing, IEEE Transactions on*, vol. 57, no. 8, pp. 3075–3085, 2009.

[22] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3042–3054, 2010.

[23] A. Al Hilli, L. Najafizadeh, and A. Petropulu, “Generalized range space property for group sparsity of linear underdetermined systems,” in *2016 Annual Conference on Information Science and Systems (CISS)*. IEEE, 2016, pp. 568–571.

[24] Z. Zhang and B. D. Rao, “Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.

[25] ——, “Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation,” *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2009–2015, 2013.

[26] I. F. Gorodnitsky and B. D. Rao, “Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm,” *IEEE Transactions on signal processing*, vol. 45, no. 3, pp. 600–616, 1997.

[27] J. C. Mosher, P. S. Lewis, and R. M. Leahy, “Multiple dipole modeling and localization from spatio-temporal MEG data,” *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 6, pp. 541–557, 1992.