Low-Complexity Optimization for Direction-of-Arrival Estimation via Approximate Message Passing

1st Xinyu Zhang
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
zhangxinyu90111@163.com

2nd Kai Huo
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
huokai2001@163.com

3rd Shuanghui Zhang
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
shzhang3@126.com

4th Yongxiang Liu
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
lyx_bible@sina.com

5th Weidong Jiang
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
jwd2232@vip.163.com

6th Xiang Li
College of Electronic Science and Technology
National University of Defense Technology
Changsha, China
lixiang01@vip.sina.com

***Abstract*—Sparsity-inducing techniques have been introduced into direction of arrival (DOA) estimation and achieved a great success in performance. However the computational complexity of the conventional sparsity-inducing techniques is prohibitively high and thus prevents such methods from application. In this paper, we propose a low-complexity DOA estimation algorithm based on approximate message passing (AMP). Derived from the loopy belief propagation, AMP is a fast algorithm to obtain the posterior distribution of the signal. The proposed algorithm combines the AMP with expectation maximization (EM) technique to adaptively learn the hyper-parameters in the Gaussian priori of the signal. Closed-form update rule of signal prior variance is derived using fix-point method, an estimator of sources number and an empirical update rule for noise variance are also derived. Compared with the state-of-the-art algorithms, the proposed algorithm reduces the computational complexity by several orders of magnitude, while obtaining comparable performance of DOA estimation. Numerical simulation demonstrates the advantages of the proposed algorithm.**

***Index Terms*—Direction-of-arrival estimation, approximate massage passing, expectation maximization**

I. Introduction

Direction of arrival (DOA) estimation has been intensively studied for decades. It has found application in many areas such as radar, sonar, navigation and communication etc.. Conventional methods, such as the sub-space based methods [1], [2] and methods based on maximum likelihood paradigm [3], [4] usually require moderate signal to noise ratio (SNR), non-coherent sources assumption and sufficient number of snapshots, which confines their application in practice.

Recently, with the advent of compressed sensing theory [5], [6], the sparsity-inducing techniques have been introduced into DOA estimation and achieved a great success in improving robustness of estimation against noise, limited number of snapshots and correlation of signals. In its most basic form, the targets were assumed to be relatively sparse on the many grids defined by a over-complete dictionary matrix. One class of the sparsity-inducing methods introduced $l_p-\text{norm}, p > 0$ penalty [7], [8] or the atomic norm [9] as an additional constraint on the weight vector that was aimed to recover. Another category of such methods [10]–[12], called sparse Bayesian learning (SBL), approached the problem from Bayesian perspective where a sparse prior was usually assumed. However, the computational cost remained prohibitively expensive, especially when the dimensions of dictionary matrix were large.

The approximate message passing (AMP) algorithm [13], [14] was derived from the loopy belief propagation [15] for bipartite graph. By introducing a message passing term into the iterative thresholding schemes, the AMP algorithm was reported to obtain substantial improvement of the sparsity-undersampling trade-off [13]. In [16], the AMP was extended to incorporate arbitrary distributions on both the input and output of the bipartite graph model, which facilitated the application of the algorithm in practice. Later, the authors in [17] inserted an expectation-maximization step into the iteration, enabling the algorithm to learn the priori. However, their method, named EM-GAMP, was based on the Gaussian-

This research is funded by National Science Foundation of China under contract No. 61025006 and No. 60872134 and China Postdoctoral Science Foundation under grant NO. 2018M633667.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

Bernoulli and Gaussian-Mixture distributions as the priori. Recently, the authors in [18] proposed using the Gaussian scale mixture (GSM) as the priori in EM-GAMP. But their algorithm focuses on the property of the EM-GAMP in general cases, it did not specify the exact update rule for noise variance nor the approach to determine the number of sources.

In this paper, we propose a novel DOA estimation method based on the approximate message passing (AMP). In the proposed method, the priori of the signal is assumed to be zero-mean Gaussian distribution. The DOA estimation is then obtained by estimating the variances in the signal priori. Compared with traditional sparsity-exploiting method such as least absolute shrinkage and selection operator (LASSO), the proposed method does not require the troublesome tuning of the trade-off parameter and was significantly faster without losing accuracy and precision of the estimation. Detailed steps of the algorithm are given and numerical simulation demonstrates the advantages of the proposed method.

II. RELATION TO PRIOR WORK

The work in this paper belongs to the SBL-based optimization approach to DOA estimation. The original AMP algorithm proposed in [13] was derived from loopy belief propagation and exhibited great advantages for its significantly reduced computational complexity. However, it does not have the ability to learn the prior parameters. The work done by Vila and Schniter [17] combined the AMP with expectation maximization (EM) technique to adaptively learn the prior parameters. However, their work assumes the priori of signal to be Bernoulli-Gaussian distribution or Gaussian-mixture distribution. Recently, a novel EM-GAMP algorithm has been proposed for the case where the prior distribution is assumed to be GSM. However, the algorithm did not specify the exact update rule for noise variance nor the approach to determine the number of sources. In this paper, we proposed a novel EM-AMP based DOA estimation method, where we use fix-point method and empirical approach to obtain a simpler update for the prior parameter and give the estimator for the number of sources. Hence, the update rules of our algorithm are different from those in [18].

III. SIGNAL MODEL

Consider a linear array with m elements uniformly spaced by d. K independent far-field sources impinge on the array from different angles $\theta_i, i = 1, 2, \cdots K$. The received signal $\mathbf{y}(t)$ at the tth moment could be written as

$$\mathbf{y}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{A} = [\mathbf{a}(\bar{\theta}_1) \quad \mathbf{a}(\bar{\theta}_2) \quad \cdots \quad \mathbf{a}(\bar{\theta}_n)]$ is the dictionary matrix composed by the array manifold vectors

$$\mathbf{a}(\theta_i) = [1 \quad e^{j\frac{2\pi d\cos(\theta_i)}{\lambda}} \quad \cdots \quad e^{j\frac{2\pi(m-1)d\cos(\theta_i)}{\lambda}}] \tag{2}$$

where $i = 1, 2, \cdots, n$ and λ is the wavelength. $\mathbf{x}(t)$ is a k-sparse vector whose non-zero indexes correspond to the directions of the targets. $\mathbf{n}(t)$ is assumed to be Gaussian white noise whose covariance matrix is $\sigma_0\mathbf{I}_m$ where $\mathbf{I}_m$ is a $m \times m$ unit matrix. For the brevity of statement, we will omit the time index t in the following unless it is required. Naturally we have $k \ll n$, therefore one typical approach is to form the signal recovery problem as

$$\min \|\mathbf{x}\|_1 \quad s.t. \ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_F \leq \varepsilon \tag{3}$$

where $\|\cdot\|_1$ and $\|\cdot\|_F$ represent the l_1-norm and the Frobenius norm respectively, ε is a user-determined parameter which should be optimized according to the noise power. The well-known LASSO problem can be solved by convex programming.

From Bayesian perspective, the above problem is equal to the case where a Laplacian prior of the signal is assumed. In this paper, we assume the prior of the signal $\mathbf{x}$ is complex Gaussian distribution with zero mean and covariance matrix $\mathbf{\Gamma} = \text{diag}(\gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_n)$ where $\gamma_i \geq 0, i = 1, 2, \cdots, n$. Thus, the prior distribution of the signal x_i can be written as

$$x_i \sim \begin{cases} \mathcal{CN}(0, \gamma_i) & \text{if} \gamma_i > 0 \\ \delta(0) & \text{if} \gamma_i = 0 \end{cases} \tag{4}$$

where $\mathcal{CN}(0, \gamma_i)$ represents the complex Gaussian distribution with zero mean and variance γ_i. Here the variance γ_i is considered to be a parameter that needs to be learned from the data and consequently the directions of targets can be obtained accordingly.

IV. APPROXIMATE MESSAGE PASSING

In this section, we introduce the AMP algorithm to solving the DOA estimation problem. The AMP algorithm is derived from loopy belief propagation algorithm designed for bipartite graph model. The loopy belief propagation algorithm is able to produce the posterior distribution $p(\mathbf{x}|\mathbf{y})$ within a few iterations and consequently the minimum mean-squared error (MMSE) estimates of $\mathbf{x}$ and the estimates of $\mathbf{\Gamma}$ can be obtained. The idea of the algorithm is to iteratively decompose the vector valued estimation problem into a sequence of scalar operations at the input and output nodes. Therefore, the algorithm is much faster than the conventional convex programming methods and empirical-Bayesian algorithms based on relevance vector machine.

The steps of the AMP algorithm is listed below where $(\cdot)^*$ represents the conjugate operator, index k means the kth AMP iteration, $\hat{\mathbf{x}} = [\hat{x}_1 \quad \hat{x}_2 \quad \cdots \quad \hat{x}_n]$ is the estimate of $\mathbf{x}$. For brevity of statement, the detailed mathematical derivation of the algorithm is omitted and interested reader can refer to [16] for details. There are two scalar estimation functions g_{in} and g_{out} which depends on the prior distribution $p(\mathbf{x}|\mathbf{\Gamma})$ and the conditional distribution $p(\mathbf{y}|\mathbf{x})$ respectively. Readers are referred to [16] for better understanding of these two functions. In brief, the function g_{in} is the posterior mean of the signal $x_j, j = 1, 2, \cdots, n$ and $\tau_j^r \frac{\partial}{\partial \hat{r}} g_{\text{in}}(k, \hat{r}_j, \gamma_j, \tau_j^r(k))$ is the posterior variance of the signal. Function g_{out} and its derivative are actually quantities that emerged during the derivation of the algorithm.

For our model, the scalar functions g_{in} and g_{out} can be written as

$$\begin{aligned} g_{\text{in}} &= \frac{\hat{r}_j(k)\gamma_j}{\tau_j^r + \gamma_j}, j = 1, 2, \cdots, n \\ g_{\text{out}} &= \frac{y_i - \hat{p}_i(k)}{\sigma_0 + \tau_i^p}, i = 1, 2, \cdots, m \end{aligned} \tag{5}$$

Their derivatives $\frac{\partial}{\partial \hat{r}} g_{\text{in}}$ and $\frac{\partial}{\partial \hat{p}} g_{\text{out}}$ can be written as

$$\begin{aligned} \tfrac{\partial}{\partial \hat{r}} g_{\text{in}} &= \tfrac{\gamma_j}{\tau_j^r(k) + \gamma_j}, j = 1, 2, \cdots, n \\ \tfrac{\partial}{\partial \hat{p}} g_{\text{out}} &= -\tfrac{1}{\sigma_0 + \tau_i^p}, i = 1, 2, \cdots, m \end{aligned} \tag{6}$$

The initial condition for the AMP algorithm should be estimated from the prior distribution of the signal as

$$\begin{aligned} \hat{x}_j(0) &= < x_j >_{p(x_j|\gamma_j)} \\ \tau_j^x(0) &= < |x_j - \hat{x}_j(0)|^2 >_{p(x_j|\gamma_j)} \end{aligned} \tag{7}$$

where $< \cdot >_{p(x_j|\gamma_j)}$ represents the expectation taken over $p(x_j|\gamma_j)$.

After sufficient iterations, the posterior distribution of signal $p(\mathbf{x}|\mathbf{y})$ can be obtained by

$$p(x_j|\mathbf{y}) = \frac{p(x_j)\mathcal{N}(x_j; \hat{r}_j, \tau_j^r)}{\int_{x_j} p(x_j)\mathcal{N}(x_j; \hat{r}_j, \tau_j^r)} \tag{8}$$

The above algorithm can produce the MMSE estimate of the signal $\mathbf{x}$, but the number of iterations is hard to determine. During the iteration of AMP, the parameters $\mathbf{\Gamma}, \sigma_0$ are assumed to be known. The AMP algorithm is used in the E-step in order to obtain the posterior distribution (8). With the posterior distribution, the prior parameters can be updated in the M-step.

Approximate Message Passing

Input: the dictionary matrix $\mathbf{A}$, prior paramters $\mathbf{\Gamma}, \sigma_0$ and received signal $\mathbf{y}$

1) *Initialization:* Set $k = 0$, set $\hat{x}_j(0)$ and $\tau_j^x(0)$ to some initial values and set $\hat{s}_i(-1) = 0$.
2) *Output linear step:* For each $i = 1, 2, \cdots, m$, calculate the following:

$$\begin{aligned} \tau_i^p(k) &= \sum_j |a_{ij}|^2 \tau_j^x(k) \\ \hat{p}_i(k) &= \sum_j a_{ij}\hat{x}_j(k) - \tau_i^p \hat{s}_i(k-1) \\ \hat{z}_i(k) &= \sum_j a_{ij}\hat{x}_j \end{aligned} \tag{9}$$

3) *Output nonlinear step:* For each $i = 1, 2, \cdots, m$, calculate the following:

$$\begin{aligned} \hat{s}_i(k) &= g_{\text{out}}(k, \hat{p}_i(k), y_i, \tau_i^p(k)) \\ \tau_i^s(k) &= -\tfrac{\partial}{\partial \hat{p}} g_{\text{out}}(k, \hat{p}_i(k), y_i, \tau_i^p(k)) \end{aligned} \tag{10}$$

4) *Input linear step:* For each $j = 1, 2, \cdots, n$, calculate the following:

$$\begin{aligned} \tau_j^r(k) &= [\sum_i |a_{ij}|^2 \tau_i^s(k)]^{-1} \\ \hat{r}_j(k) &= \hat{x}_j(k) + \tau_j^r(k) \sum_i (a_{ij})^* \hat{s}_i(k) \end{aligned} \tag{11}$$

5) *Input nonlinear step:* For each $j = 1, 2, \cdots, n$, calculate the following:

$$\begin{aligned} \hat{x}_j(k+1) &= g_{\text{in}}(k, \hat{r}_j(k), \gamma_j, \tau_j^r(k)) \\ \tau_j^x(k+1) &= \tau_j^r \tfrac{\partial}{\partial \hat{r}} g_{\text{in}}(k, \hat{r}_j, \gamma_j, \tau_j^r(k)) \end{aligned} \tag{12}$$

Output: τ_i^p, $\hat{p}_i$, $\hat{z}_i$, $\hat{s}_i$, τ_i^s, $i = 1, 2, \cdots, m, \tau_j^r$, $\hat{r}_j$, $\hat{x}_j$, τ_j^x $j = 1, 2, \cdots, n$

V. DOA Estimation by EM Learning

In this section, we use the EM technique to iteratively learn the prior parameters $\mathbf{q} = [\gamma, \sigma_0]$ where γ is a vector composed by the diagonal elements of the matrix $\mathbf{\Gamma}$. For our model, the update rule of prior parameters at each iteration is

$$\mathbf{q}^{i+1} = \arg\max_{\mathbf{q}} < \ln p(\mathbf{y}, \mathbf{x}; \mathbf{q}) >_{p(\mathbf{x}|\mathbf{y};\mathbf{q}^i)} \tag{13}$$

Here the expectation in the above equation is calculated with respect to the posterior distribution $p(\mathbf{x}|\mathbf{y}; \mathbf{q}^i)$ under the parameter hypothesis $\mathbf{q}^i$ where the superscript $(\cdot)^i$ means the ith EM iteration. Since it is impractical to update all the elements of $\mathbf{q}$ at once, the elements of $\mathbf{q}^i$ need to be updated respectively.

A. EM update for γ

In this subsection, we derive the updates for γ. Since $p(\mathbf{y}|\mathbf{x})$ is irrelevant of γ and the priori $p(\mathbf{x}; \mathbf{q})$ can be decoupled into $\prod_j p(x_j; \gamma_j)$, the update rule for γ_j can be written as

$$\gamma^{i+1} = \arg\max < \ln p(x_j; \gamma_j) >_{p(x_j|\mathbf{y};\mathbf{q}^i)} \tag{14}$$

where $p(x_j|\mathbf{y}; \mathbf{q}^i)$ can be obtained from (8) as

$$p(x_j|\mathbf{y}; \mathbf{q}^i) = \mathcal{CN}(\frac{\hat{r}_j\gamma_j}{\tau_j^r + \gamma_j}, \frac{\tau_j^r\gamma_j}{\tau_j^r + \gamma_j}) \tag{15}$$

Substituting (4) and (15) into (14), one could obtain that

$$\gamma_j^{i+1} = \frac{|\hat{r}_j(t)\gamma_j^i/(\tau_j^r + \gamma_j^i)|^2}{1 - \tau_j^r\gamma_j^i/((\tau_j^r + \gamma_j^i)\gamma_j^i)} \tag{16}$$

It can be noticed from (16) that our update rule for γ is much simpler than the one in []. Furthermore, with (16), one can establish a stopping criteria for the iteration as $\|\gamma^{i+1} - \gamma^i\|_1 \leq \varepsilon$. Here ε is a threshold that determines the number of iterations.

B. Determine K and EM update for σ_0

Another parameter that needs to be learned is the noise power σ_0. The update of σ_0 also influences the optimization of γ and hence the precision of DOA estimation. The conventional EM update for σ_0 tends to under-estimate the noise variance. Thus in this paper, we adopt an empirical approach to estimate σ_0. Before jumping to the estimator of σ_0, we need to determine the number of sources K. Applying the BIC rule [19] to the model, we can obtain the following model order selector

$$K^* = \arg\max_{\hat{K}} \left\{ mT \ln\left(\frac{\operatorname{tr}((\mathbf{I} - \mathbf{P})\mathbf{R})}{m - \hat{K}}\right) - T\hat{K} + \kappa \right\} \tag{17}$$

where $\hat{K}$ is the model order, T is the number of available time samples, $\kappa = \frac{(2\hat{K}T+1)\ln(T)}{2}$, $\operatorname{tr}(\cdot)$ is the trace of a matrix and

$$\mathbf{P} = \mathbf{A}_{\mathcal{M}}(\mathbf{A}_{\mathcal{M}}^H\mathbf{A}_{\mathcal{M}})^{-1}\mathbf{A}_{\mathcal{M}}^H, \mathbf{R} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^H \tag{18}$$

where $(\cdot)^H$ is the conjugate transpose of a matrix, $(\cdot)^{-1}$ is the inverse of a matrix, $\mathbf{Y} = [\mathbf{y}(1) \quad \mathbf{y}(2) \quad \cdots \quad \mathbf{y}(T)]$ is a collection of T received samples, $\mathbf{A}_{\mathcal{M}}$ is a matrix composed of columns from $\mathbf{A}$ corresponding to the set of presumed target directions $\mathcal{M}$. The set $\mathcal{M}$ can be obtained by choosing the $\hat{K}$ directions that correspond to the first $\hat{K}$ largest peaks in γ. Once the number of sources is correctly obtained, $\mathcal{M}$ will be consisted of the estimated target DOAs.

At the optimal point of $\mathbf{q}$, the following condition must be satisfied [20]

$$\mathbf{A}_{\mathcal{M}}^H(\mathbf{R} - \mathbf{\Sigma_Y})\mathbf{A}_{\mathcal{M}} = \mathbf{0} \tag{19}$$

where $\mathbf{\Sigma_Y}$ is the covariance matrix. Then one can derive from (19) the following noise variance update

$$(\sigma^2)^{i+1} = \frac{\text{tr}\left((\mathbf{I} - \mathbf{P})\mathbf{R}\right)}{m - K} \tag{20}$$

C. EM initialization

Although there are no certain constraints about the initial values of $\mathbf{\Gamma}$ and σ_0, prior information about these parameters, should it exists, can be utilized to form preliminary estimates. A good initial estimation of these prior parameters could accelerate the speed of convergence. Since the variances of noise are assumed to be the same across different channels. The following equation can be used to obtain a fair initial estimate of the noise power [18].

$$\hat{\sigma_0} = \frac{\|\mathbf{y}\|_2^2}{(SNR + 1)m} \tag{21}$$

where $\|\cdot\|_2$ is the l_2-norm and SNR is an initial estimate of SNR. If the signal power of different sources are assumed to be the same, the following equation can be used to obtain the initial estimate of the signal power.

$$\gamma_i = \hat{\gamma_0} = \frac{\|\mathbf{y}\|_2^2 - m\hat{\sigma_0}}{\text{tr}(\mathbf{A}^H\mathbf{A})}, \, i = 1, 2, \cdots, n \tag{22}$$

VI. Numerical Simulation

In this section we report the results of a numerical study that investigate the performance of our proposed algorithm under noisy settings. Consider a linear array with 16 elements uniformly spaced by half of wavelength. Two independent sources impinge on the array from $\theta_1 = 20°$ and $\theta_2 = 30°$ respectively. The sparsity-inducing methods sample the $[-45° \quad 45°]$ space with $1°$ interval to form the over-complete dictionary matrix $\mathbf{A}$. The signal-to-noise (SNR) ratio of the two received signals are assumed to be the same. Among the existing methods, the LASSO and Tipping's relevance vector machine (RVM) [21] are used as performance comparison with the proposed algorithm. We used CVX to solve the convex programming posed by LASSO and the priori of the hyper-parameters in the RVM were chosen to be non-informative. In our simulation, we saved the trouble of tuning the weight parameter in LASSO, instead we used an empirical value in each case. In our simulation, the DOA estimation is conducted with only one snapshot. Therefore, we did not use conventional subspace-based methods as comparison.

The reconstructed spatial spectral are plotted in Fig. 1. It can be seen that all the three sparsity recovery algorithms find two peaks at the true location of the targets. However, the proposed algorithm has wider peak width. This is because that the proposed algorithm is an approximate approach to implement the EM iteration in sparse Bayesian learning.

The root-mean-squared error (RMSE) of DOA estimates of different algorithms are compared in Fig. 2 in the case of different SNR. The RMSE of DOA estimates are calculated by 100 Monte Carlo simulations. It can be seen that as the SNR increases, the RMSE of DOA estimates of all these algorithms decrease. The performance of the proposed algorithm is comparable to the LASSO in all cases and slightly better than the RVM. However as it will be seen in the following, the computational complexity of the proposed algorithm is significantly less than the other two methods. Compared with the conventional SBL methods, the proposed algorithm also adopts the EM learning technique, however, the proposed algorithm takes an approximate but much faster way to obtain the posterior distribution of the signal.

In another scenario, the targets are incident from $\theta_1 = 10°$ and $\theta_2 = 20°$ respectively. The rest of condition remains the same as the first scenario. The RMSE of DOA estimates versus SNR is plotted in Fig. 3. It can be seen that the RMSE of the proposed algorithm is similar to the other two algorithms. Compared with Fig. 2, one can find that as the targets approach the boresight of the array, the RMSE of DOA estimation is lower.

To evaluate the computational complexity of different algorithms, we conduct 50 Monte Carlo simulations and record the CPU time spent by each algorithm for DOA estimation. The results are shown in Table I. It can be seen that the proposed algorithm spends significantly less time than the LASSO and the RVM. Thus, the proposed algorithm reduces the computational cost of DOA estimation by several orders of magnitude, which is extremely helpful for application in practical radar system.

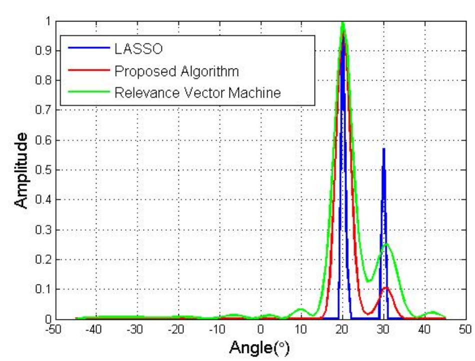

Fig. 1. Reconstructed spectra of different algorithms

VII. Conclusion

In this paper, we propose a novel low-complexity optimization algorithm for DOA estimation. The AMP for complex

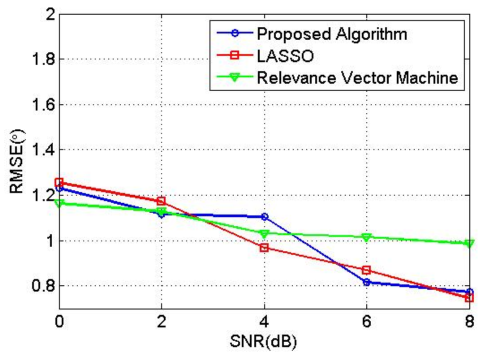

Fig. 2. DOA RMSE versus SNR in scenario 1 ($\theta_1 = 20^\circ$ and $\theta_2 = 30^\circ$)

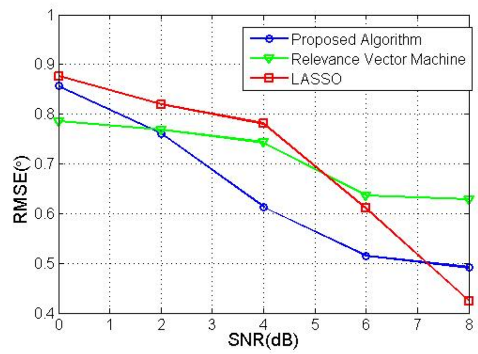

Fig. 3. DOA RMSE versus SNR in scenario 2 ($\theta_1 = 10^\circ$ and $\theta_2 = 20^\circ$)

TABLE I
CPU TIME SPENT BY DIFFERENT ALGORITHMS (S)

SNR(dB)	0	2	4	6	8
Proposed Algorithm	0.0102	0.0090	0.0101	0.0115	0.0149
RVM	12.5559	12.4790	11.9380	11.7604	10.3909
LASSO	0.9793	1.0027	0.9898	0.9924	0.9815

Gaussian priori is developed to obtain the posterior distribution of signal, which is then used to estimate the DOA of target via the EM technique. Closed-form update rules are derived using fix-point method. Numerical simulation shows that the proposed algorithm can reduce the computational complexity by serval orders of magnitude and also achieves comparable performance of DOA estimation. The complexity benefits of the proposed algorithm are particularly attractive for large-scale antenna systems.

REFERENCES

[1] Richard Roy and Thomas Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, pp. 984–995, 1989.

[2] Fenggang Yan, Ming Jin and Xiaolin Qiao, "Low-Complexity DOA Estimation Based on Compressed MUSIC and Its Performance Analysis," IEEE Transactions on Signal Processing, vol. 61, pp. 1915–1930, 2013.

[3] P. Stoica, A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, pp. 720–741, 1989.

[4] P. Stoica, A. Nehorai, "Performance Study of Conditional and Unconditional Direction-of-Arrival Estimation," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 38, pp. 1783–1795, 1990.

[5] David L Donoho, "Compressed Sensing," IEEE Transactions on Information Theory, vol. 52, pp. 1289–1306, 2006.

[6] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht, "Compressed Sensing off the Grid," IEEE Transactions on Information Theory, vol. 59, pp. 7465–7490, 2013.

[7] B. D. Jeffs, "Sparse Inverse Solution Methods for Signal and Image Processing Applications," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 3, pp. 1885–1888, 1998.

[8] 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, pp. 600–616, 1997.

[9] Emmanuel J Candes and Carlos Fernandez-Granda, "Towards a Mathematical Theory of Super-Resolution," Communications on Pure and Applied Mathematics, vol. 67, pp. 906–956, 2014.

[10] David P. Wipf and Bhaskar D. Rao, "An Empirical Bayesian Strategy for Solving the Simultaneous Sparse Approximation Problem," IEEE Transactions on Signal Processing, vol. 55, pp. 3704–3716, 2007.

[11] David P. Wipf and Bhaskar D. Rao, "Latent Variable Bayesian Models for Promoting Sparsity," IEEE Transactions on Information Theory, vol. 57, pp. 6236–6255, 2011.

[12] Xinyu Zhang, Kai Huo, Yongxiang Liu and Xiang Li, "Direction of Arrival Estimation via Joint Sparse Bayesian Learning for Bi-Static Passive Radar," IEEE Access, vol. 7, pp. 72979–72993, 2019.

[13] D. L. Donoho, A. Maleki, A. Montanari, "Message-Passing Algorithms for Compressed Sensing," Proc. Nat. Acad. Sci., vol. 106, pp. 18914–18919, 2009.

[14] M. Bayati and A. Montanari, "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing," IEEE Transactions on Information Theory, vol. 57, pp. 764–785, 2011.

[15] M. J. Wainwright and M. I. Jordan, "Graphical Models, Exponential Families, and Variational Inference," in Foundations and Trends in Machine Learning, vol. I, Hanover, MA: NOW Publishers, 2008.

[16] S. Rangan, "Generalized Approximate Message Passing for Estimation with Random Linear Mixing," arXiv:1010.5141, 2012.

[17] Jeremy Vila and Philip Schniter, "Expectation-Maximization Gaussian-Mixture Approximate Message Passing," IEEE Transactions on Signal Processing, vol. 61, pp. 4658–4672, 2012.

[18] Maher Al-Shoukairi, Philip Schniter and Bhaskar D. Rao, "A gamp-based low complexity sparse bayesian learning algorithm," IEEE Transactions on Signal Processing, vol. 66, pp. 294–307, 2018.

[19] Zhang-Meng Liu, Zhi-Tao Huang, and Yi-Yu Zhou, "An Efficient Maximum Likelihood Method for Direction-of-Arrival Estimation via Sparse Bayesian Learning," IEEE Transactions on Wireless Communications, vol. 11, pp. 3607–3617, 2012.

[20] Peter Gerstoft, Christoph F. Mecklenbrauker, Angeliki Xenaki and Santosh Nannuru, "Multisnapshot Sparse Bayesian Learning for DOA," IEEE Signal Processing Letters, vol. 23, pp. 1469–1473, 2016.

[21] Michael E Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," Journal of Machine Learning Research, vol. 1, pp. 211–244, 2001.