

Article

Sparse Bayesian Learning for DOA Estimation with Mutual Coupling

Jisheng Dai ^{1,2,*}, Nan Hu ³, Weichao Xu ⁴ and Chunqi Chang ³

¹ School of Electrical and Information Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, China

² National Mobile Communications Research Laboratory, Southeast University, 2 Sipailou Road, Nanjing 210096, China

³ School of Electronic and Information Engineering, Soochow University, 178 East Ganjiang Road, Suzhou 215006, China; E-Mails: hunan@suda.edu.cn (N.H.); cqchang@suda.edu.cn (C.C.)

⁴ Department of Automatic Control, Guangdong University of Technology, 100 Huanxi Road, Guangzhou 510006, China; E-Mail: wxu@gdut.edu.cn

* Author to whom correspondence should be addressed; E-Mail: jsdai@ujs.edu.cn; Tel./Fax: +86-511-8882-2666.

Academic Editor: Vittorio M. N. Passaro

Received: 23 August 2015 / Accepted: 10 October 2015 / Published: 16 October 2015

Abstract: Sparse Bayesian learning (SBL) has given renewed interest to the problem of direction-of-arrival (DOA) estimation. It is generally assumed that the measurement matrix in SBL is precisely known. Unfortunately, this assumption may be invalid in practice due to the imperfect manifold caused by unknown or misspecified mutual coupling. This paper describes a modified SBL method for joint estimation of DOAs and mutual coupling coefficients with uniform linear arrays (ULAs). Unlike the existing method that only uses stationary priors, our new approach utilizes a hierarchical form of the Student t prior to enforce the sparsity of the unknown signal more heavily. We also provide a distinct Bayesian inference for the expectation-maximization (EM) algorithm, which can update the mutual coupling coefficients more efficiently. Another difference is that our method uses an additional singular value decomposition (SVD) to reduce the computational complexity of the signal reconstruction process and the sensitivity to the measurement noise.

Keywords: Sparse Bayesian Learning (SBL); Direction-of-Arrival (DOA); Uniform Linear Array (ULA); mutual coupling

1. Introduction

The problem of estimating the direction-of-arrival (DOA) of multiple narrow-band sources has received considerable attention in many fields, e.g., radar, sonar, radio astronomy and mobile communications [1]. Many high-resolution DOA estimation algorithms have been proposed in the last few decades. Their excellent performance relies crucially on perfect knowledge of the array manifold. In practice, however, the array manifold usually suffers from imperfections, such as unknown mutual coupling between antenna elements, imperfectly-known sensor positions and orientations and gain-phase imbalances. Without array manifold calibration, the performance of DOA estimation may degrade substantially. Hence, it is necessary to calibrate imperfections prior to carrying out DOA estimation.

A large number of calibration methods has been proposed to deal with the imperfect manifold caused by unknown or misspecified mutual coupling [2–11]. For example, some iterative mutual coupling auto-calibration methods for uniform linear arrays (ULAs) and uniform circular arrays (UCAs) were proposed in [2,3]. However, the high computational complexity required by iterations may be time consuming, and the convergence is not theoretical guaranteed, thus resulting in algorithmic instability. Recently, by taking advantage of the special structure of the mutual coupling matrix for ULAs, Ye and coauthors [4,5] demonstrated that DOAs can be accurately estimated without compensating for mutual coupling if a group of auxiliary sensors are added into ULAs. Dai and coauthors further extended the result to the spatial smoothing method [6], real-valued method [7] and l_1 -norm regularized method [12]. These methods are referred to as the auxiliary methods. Neither a calibration source nor iteration is required in the auxiliary methods. However, because of only applying a middle subarray, they reduce the working array aperture, which limits the applicability of these methods to cases with either few DOAs or many array elements. Some alternatives that can calibrate the imperfect manifold with the whole array in the presence of mutual coupling are proposed in [8–11].

Recently, the emerging technique of compressive sensing (CS) has given renewed interest to the problem of DOA estimation [12–18]. These methods exhibit many advantages, e.g., improved robustness to noise, limited number of snapshots and correlation of signals. Sparse Bayesian learning (SBL) is a popular and important technique for the sparse signal recovery in CS [19–21], which formulates the signal recovery problem from a Bayesian perspective, while the sparsity information is exploited by assuming a sparse prior for the signal of interest. It is worth noting that l_1 -norm regularized optimization is deemed to be a special example of SBL, if a maximum *a posteriori* (MAP) optimal estimate is adopted with a Laplace signal prior. Theoretical and empirical results show that SBL methods can achieve enhanced performance over l_1 -norm regularized optimization [20,21]. It is generally assumed that the measurement matrix in SBL is precisely known. Unfortunately, this assumption is invalid if perturbations on the measurement matrix (e.g., when the array manifold suffers from the aforementioned imperfections) are considered. Additive perturbations [22,23] and multiplicative perturbations [16] have been addressed in the recent literature.

In this paper, we will propose a modified Bayesian method for joint estimation of DOAs and the mutual coupling coefficients with ULAs, where the optimization problem is formulated in a SBL framework with a two-stage hierarchical prior, and then, we adopt an expectation-maximization (EM) algorithm that treats the signals of interest and the mutual coupling coefficients as hidden variables and

parameters, respectively. To the best of our knowledge, there are very limited SBL methods that can resolve the DOAs with mutual coupling. This problem was first addressed in [24], where a unified SBL framework was proposed to address the DOA estimation problem with typical perturbations of mutual coupling, gain-phase uncertainty and sensor location error. The main differences between our method and the method in [24] are:

- Our SBL model constitutes a two-stage hierarchical form, which results in the Student t prior in a hierarchical manner; while the model in [24] has only stationary priors. The advantage of the Student t prior is that it enforces the sparsity constraint more heavily [21,25].
- Our method provides a distinct Bayesian inference for the EM algorithm, which can update the mutual coupling coefficients more efficiently.
- Our method uses an additional singular value decomposition (SVD) to reduce the computational complexity of the signal reconstruction process and the sensitivity to the measurement noise.

These differences guarantee that our method can give a better joint estimation performance of DOAs and the mutual coupling coefficients.

2. Data Model from the Bayesian Viewpoint

2.1. DOA Estimation Model

Consider K narrow-band far-field sources impinging on an M -element ULA, where the distance between adjacent sensors is d . The K signals, $s_1(t), s_2(t), \dots, s_K(t)$, arrive at the array from distinct directions, $\theta_1, \theta_2, \dots, \theta_K$, with respect to the normal line of the array. The $M \times 1$ array output vector $\mathbf{y}(t)$ is then given by:

$$\mathbf{y}(t) = \mathbf{C}\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \quad t \in \{t_1, t_2, \dots, t_T\} \quad (1)$$

where $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_M(t)]^T$, $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_K(t)]^T$, $\mathbf{A} = [\boldsymbol{\alpha}(\theta_1), \boldsymbol{\alpha}(\theta_2), \dots, \boldsymbol{\alpha}(\theta_K)]$, $\boldsymbol{\alpha}(\theta_k) = [1, e^{j\phi(\theta_k)}, \dots, e^{j(M-1)\phi(\theta_k)}]^T$, $\phi(\theta_k) = (-2\pi d/\lambda) \sin(\theta_k)$ and $\mathbf{n}(t) = [n_1(t), n_2(t), \dots, n_M(t)]^T$ is an unknown noise vector. The matrix $\mathbf{C} \in \mathbb{C}^{M \times M}$ is the mutual coupling matrix (MCM) for ULAs. Many theoretical or experimental studies [26,27] have demonstrated that a banded symmetric Toeplitz MCM with m ($\leq \frac{M-1}{2}$) coefficients can provide a good approximation of real-world situations, *i.e.*, $\mathbf{C} = \text{toeplitz}(\mathbf{c})$, where $\mathbf{c} = [c_0, c_1, c_2, \dots, c_m]$ with $0 < |c_1|, |c_2|, \dots, |c_m| < c_0 = 1$ and $\text{toeplitz}(\mathbf{c})$ denotes a symmetric Toeplitz matrix constructed by the vector \mathbf{c} .

Denoting $\mathbf{Y} = [\mathbf{y}(t_1), \mathbf{y}(t_2), \dots, \mathbf{y}(t_T)]$, $\mathbf{N} = [\mathbf{n}(t_1), \mathbf{n}(t_2), \dots, \mathbf{n}(t_T)]$ and $\mathbf{S} = [\mathbf{s}(t_1), \mathbf{s}(t_2), \dots, \mathbf{s}(t_T)]$, we have:

$$\mathbf{Y} = \mathbf{C}\mathbf{A}\mathbf{S} + \mathbf{N}. \quad (2)$$

Following the convention in [14], we use the singular value decomposition (SVD) to reduce the computational complex of the signal reconstruction process and the sensitivity to the measurement noise. Let the SVD of \mathbf{Y} be written in the form of:

$$\mathbf{Y} = \mathbf{U}_s \boldsymbol{\Lambda}_s \mathbf{V}_s^H + \mathbf{U}_n \boldsymbol{\Lambda}_n \mathbf{V}_n^H \quad (3)$$

where $\mathbf{U}_s \in \mathbb{C}^{M \times \hat{K}}$ and $\mathbf{V}_s \in \mathbb{C}^{T \times \hat{K}}$ are unitary matrices whose columns are the singular vectors corresponding to the \hat{K} largest singular values, while the columns of $\mathbf{U}_n \in \mathbb{C}^{M \times (M - \hat{K})}$ and $\mathbf{V}_n \in \mathbb{C}^{T \times (M - \hat{K})}$ are the singular vectors corresponding to the rest $M - \hat{K}$ singular values. Note that the value of \hat{K} is generally determined by the singular values; especially if the number of sources K is exactly known, \hat{K} is set to K .

Using Equation (2) and defining $\hat{\mathbf{Y}} = \mathbf{Y}\mathbf{V}_s$, we obtain:

$$\hat{\mathbf{Y}} = \mathbf{C}\mathbf{A}\hat{\mathbf{S}} + \hat{\mathbf{N}} \quad (4)$$

where $\hat{\mathbf{S}} = \mathbf{S}\mathbf{V}_s$ and $\hat{\mathbf{N}} = \mathbf{N}\mathbf{V}_s$. In order to cast the problem of DOA estimation with unknown mutual coupling as a sparse representation problem, we let Ω denote the set of possible locations and let $\hat{\theta}$ be a generic location parameter. Furthermore, let $\{\hat{\theta}_i\}_{i=1}^{K_{\hat{\theta}}}$ denote a grid that covers Ω . If the grid is fine enough, such that the true DOAs lie on (or, practically, close to) the grid, we can use the following model for $\hat{\mathbf{Y}}$:

$$\hat{\mathbf{Y}} = \mathbf{C}\mathbf{A}_{\hat{\theta}}\hat{\mathbf{S}}_{\hat{\theta}} + \hat{\mathbf{N}} \quad (5)$$

where $\mathbf{A}_{\hat{\theta}} \triangleq [\boldsymbol{\alpha}(\hat{\theta}_1), \boldsymbol{\alpha}(\hat{\theta}_2), \dots, \boldsymbol{\alpha}(\hat{\theta}_{K_{\hat{\theta}}})]$ and $\hat{\mathbf{S}}_{\hat{\theta}}$ is a $K_{\hat{\theta}} \times \hat{K}$ complex matrix whose i -th row corresponds to the signal impinging on the array from a possible source at $\hat{\theta}_i$. It is easy to verify that the i -th row is nonzero and equals the k -th row of $\hat{\mathbf{S}}$ if signal k comes from $\hat{\theta}_i$ for some k and zero otherwise. As a result, the aim of this paper is to find a row-sparse $\hat{\mathbf{S}}_{\hat{\theta}}$ (in other words, with a few nonzero rows) and a Toeplitz matrix \mathbf{C} that minimize the following objective function:

$$\|\hat{\mathbf{Y}} - \mathbf{C}\mathbf{A}_{\hat{\theta}}\hat{\mathbf{S}}_{\hat{\theta}}\|_2 \quad (6)$$

where $\|\cdot\|_2$ stands for the Frobenius norm. Finding the sparse solution to the above problem through the l_1 -norm regularized optimization is intractable, as it is a non-convex optimization problem with unknown \mathbf{C} , which cannot be solved in polynomial time. In the next section, we will propose an SBL method for the DOA estimation in the presence of unknown coupling. To this end, we have to preliminarily model the noisy and sparse signals as in [19].

2.2. Noise Model

Firstly, we address the noise model that is commonly used in SBL. Assume elements in the noise vector \mathbf{n} defined in Equation (1) are independent and each has a complex Gaussian distribution with zero mean and a common variance σ^2 . Since the orthogonal invariance property of the Gaussian random matrix makes the distribution impervious to multiplication by orthogonal matrices, each element in $\hat{\mathbf{N}}$ is approximately i.i.d. complex Gaussian with the same mean and variance [15,28], i.e., $p(\hat{N}_{i,j}) = \mathcal{CN}(\hat{N}_{i,j}|0, \beta^{-1})$, where $\beta = \sigma^{-2}$ denotes the noise precision and $\hat{N}_{i,j}$ is the (i, j) -th element in $\hat{\mathbf{N}}$. Then, we have:

$$p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) = \prod_{k=1}^{\hat{K}} \mathcal{CN}(\hat{\mathbf{y}}_k|\mathbf{C}\mathbf{A}_{\hat{\theta}}\hat{\mathbf{S}}_k, \beta^{-1}\mathbf{I}) \quad (7)$$

where \hat{y}_k and \hat{s}_k denote the k -th columns of $\hat{\mathbf{Y}}$ and $\hat{\mathbf{S}}_{\hat{\theta}}$, respectively. Usually, the noise variance σ^2 is unknown, so is the noise precision β . Hence, we model β as a Gamma hyperprior:

$$p(\beta) = \Gamma(\beta; a, b) \quad (8)$$

where we set $a, b \rightarrow 0$ as in [19,21] so as to obtain a broad hyperprior. The reason why we choose the Gamma hyperprior is that it is a conjugate prior (in Bayesian probability theory, a prior $p(\theta)$ is said to be conjugate to $p(x|\theta)$ if the posterior distribution $p(\theta|x)$ has the same functional form as the prior) of the Gaussian distribution.

2.3. Sparse Signal Model

A widely-used sparseness prior for $\hat{\mathbf{S}}_{\hat{\theta}}$ is the Laplace distribution; however, it is not readily accomplished with such a prior in SBL, because the Laplace prior is not conjugate to the Gaussian likelihood, and hence, it is unlikely to perform the associated Bayesian inference in closed form [21]. A typical SBL treatment of $\hat{\mathbf{S}}_{\hat{\theta}}$, proposed in RVM [19], begins by assigning a non-stationary Gaussian prior distribution with a distinct inverse variance δ_i for each row of $\hat{\mathbf{S}}_{\hat{\theta}}$. Letting $\boldsymbol{\delta} = [\delta_1, \delta_2, \dots, \delta_{K_{\hat{\theta}}}]^T$ and $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$, we have:

$$p(\hat{\mathbf{S}}_{\hat{\theta}}|\boldsymbol{\delta}) = \prod_{k=1}^{K_{\hat{\theta}}} \mathcal{CN}(\hat{s}_k|\mathbf{0}, \boldsymbol{\Delta}^{-1}) \quad (9)$$

In order to make the Bayesian inference convenient and to obtain a two-stage hierarchical prior that favors most rows of $\hat{\mathbf{S}}_{\hat{\theta}}$ being zeros, the hyper-parameter δ_i 's are further modeled as independent Gamma distributions [19,21], *i.e.*,

$$p(\boldsymbol{\delta}) = \prod_{i=1}^{K_{\hat{\theta}}} \Gamma(\delta_i; c, d) \quad (10)$$

where $c \rightarrow 1$ and $d \rightarrow 0$. In this case, the integral $\int_0^\infty \mathcal{CN}(\hat{s}_k|\mathbf{0}, \boldsymbol{\Delta}^{-1}) \cdot p(\boldsymbol{\delta}; c, d)d\boldsymbol{\delta}$ corresponds to the Student t distribution [19], which can be evaluated analytically. Alternatively, a two-stage hierarchical method that results in Laplace priors was addressed in [21].

3. The Proposed Sparse Bayesian Learning Method

The associated learning problem becomes the search for the hyperparameters β and $\boldsymbol{\delta}$, as well as the parameter \mathbf{C} . As $p(\hat{\mathbf{S}}_{\hat{\theta}}, \beta, \boldsymbol{\delta}|\hat{\mathbf{Y}}; \mathbf{C})$ cannot be explicitly calculated, a Type-II ML [19] (or evidence maximization) procedure is exploited to perform the Bayesian inference. In other words, the hyperparameters (β and $\boldsymbol{\delta}$) and parameter (\mathbf{C}) are estimated by maximizing $p(\hat{\mathbf{Y}}|\beta, \boldsymbol{\delta}; \mathbf{C})$ or its logarithm $\mathcal{L}(\beta, \boldsymbol{\delta}; \mathbf{C}) \triangleq \ln p(\hat{\mathbf{Y}}|\beta, \boldsymbol{\delta}; \mathbf{C})$. However, explicitly finding the values of β , $\boldsymbol{\delta}$ and \mathbf{C} that maximize $\mathcal{L}(\beta, \boldsymbol{\delta}; \mathbf{C})$ is intractable, and here, we adopt an EM algorithm that treats $\hat{\mathbf{S}}_{\hat{\theta}}$ as a hidden variable. The principle behind the EM algorithm is to instead repeatedly construct a lower bound on $\mathcal{L}(\beta, \boldsymbol{\delta}; \mathbf{C})$ (E-step) and then to optimize that lower-bound (M-step). Specifically,

- E-step: Compute:

$$p(\hat{\mathbf{S}}_{\hat{\theta}}|\hat{\mathbf{Y}}, \beta, \boldsymbol{\delta}; \mathbf{C}) \quad (11)$$

and evaluate:

$$\left\langle \ln p(\hat{\mathbf{Y}}, \hat{\mathbf{S}}_{\hat{\theta}}, \beta, \boldsymbol{\delta}; \mathbf{C}) \right\rangle_{E^{(t)}} \triangleq E_{p(\hat{\mathbf{S}}_{\hat{\theta}}|\hat{\mathbf{Y}}, \beta^{(t)}, \boldsymbol{\delta}^{(t)}; \mathbf{C}^{(t)})} \left[\ln p(\hat{\mathbf{Y}}, \hat{\mathbf{S}}_{\hat{\theta}}, \beta, \boldsymbol{\delta}; \mathbf{C}) \right] \quad (12)$$

where $(\cdot)^{(t)}$ denotes the estimated value in the t -th iteration.

- M-step: Find the hyperparameter updates and the parameter updates:

$$\{\beta^{(t+1)}, \boldsymbol{\delta}^{(t+1)}, \mathbf{C}^{(t+1)}\} = \arg \max_{\beta, \boldsymbol{\delta}, \mathbf{C}} \left\langle \ln p(\hat{\mathbf{Y}}, \hat{\mathbf{S}}_{\hat{\theta}}, \beta, \boldsymbol{\delta}; \mathbf{C}) \right\rangle_{E^{(t)}}. \quad (13)$$

In the rest of this section, we will discuss the two steps in detail.

3.1. E-Step

Using the Bayes rule, we can verify that the posterior distribution of $\hat{\mathbf{S}}_{\hat{\theta}}$ in Equation (11) is also a complex Gaussian [19]:

$$p(\hat{\mathbf{S}}_{\hat{\theta}}|\hat{\mathbf{Y}}, \beta, \boldsymbol{\delta}; \mathbf{C}) = \frac{p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) \cdot p(\hat{\mathbf{S}}_{\hat{\theta}}|\boldsymbol{\delta})}{p(\hat{\mathbf{Y}}|\beta, \boldsymbol{\delta}; \mathbf{C})} = \prod_{k=1}^{\hat{K}} \mathcal{CN}(\hat{s}_k|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}) \quad (14)$$

where:

$$\boldsymbol{\mu}_k = \beta \boldsymbol{\Sigma} \mathbf{A}_{\hat{\theta}}^H \mathbf{C}^H \hat{\mathbf{y}}_k, \quad k = 1, 2, \dots, \hat{K} \quad (15)$$

$$\boldsymbol{\Sigma} = (\beta \mathbf{A}_{\hat{\theta}}^H \mathbf{C}^H \mathbf{C} \mathbf{A}_{\hat{\theta}} + \boldsymbol{\Delta})^{-1} = \boldsymbol{\Delta}^{-1} - \boldsymbol{\Delta}^{-1} (\mathbf{C} \mathbf{A}_{\hat{\theta}})^H (\beta^{-1} \mathbf{I} + \mathbf{C} \mathbf{A}_{\hat{\theta}} \boldsymbol{\Delta}^{-1} (\mathbf{C} \mathbf{A}_{\hat{\theta}})^H)^{-1} \mathbf{C} \mathbf{A}_{\hat{\theta}} \boldsymbol{\Delta}^{-1}. \quad (16)$$

On the other hand, by combining the stages of the hierarchical Bayesian model, $p(\hat{\mathbf{Y}}, \hat{\mathbf{S}}_{\hat{\theta}}, \beta, \boldsymbol{\delta}; \mathbf{C})$ can be rewritten as:

$$p(\hat{\mathbf{Y}}, \hat{\mathbf{S}}_{\hat{\theta}}, \beta, \boldsymbol{\delta}; \mathbf{C}) = p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) p(\hat{\mathbf{S}}_{\hat{\theta}}|\boldsymbol{\delta}) p(\beta) p(\boldsymbol{\delta}). \quad (17)$$

Therefore, Equation (12) leads to:

$$\left\langle \ln p(\hat{\mathbf{Y}}, \hat{\mathbf{S}}_{\hat{\theta}}, \beta, \boldsymbol{\delta}; \mathbf{C}) \right\rangle_{E^{(t)}} = \left\langle \ln p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) p(\hat{\mathbf{S}}_{\hat{\theta}}|\boldsymbol{\delta}) p(\beta) p(\boldsymbol{\delta}) \right\rangle_{E^{(t)}}. \quad (18)$$

3.2. M-Step

In this subsection, we will address the derivation of hyperparameter updates $(\beta^{(t+1)})$ and $\boldsymbol{\delta}^{(t+1)}$, as well as the parameter update $(\mathbf{C}^{(t+1)})$.

1. For β , ignoring terms in the logarithm independent thereof, we just have to maximize:

$$\begin{aligned}
 & \left\langle \ln p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C})p(\beta) \right\rangle_{E^{(t)}} \\
 &= \left\langle \ln p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) \right\rangle_{E^{(t)}} + \ln p(\beta) \\
 &= \hat{K}M \ln \beta - \beta \sum_{k=1}^{\hat{K}} \left\langle \|\hat{\mathbf{y}}_k - \mathbf{C}\mathbf{A}_{\hat{\theta}}\hat{\mathbf{s}}_k\|_2^2 \right\rangle_{E^{(t)}} + (a-1) \ln \beta - b\beta + \text{const} \\
 &= \hat{K}M \ln \beta - \beta \sum_{k=1}^{\hat{K}} \|\hat{\mathbf{y}}_k - \mathbf{C}^{(t)}\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)}\|_2^2 \\
 &\quad - \beta \hat{K} \text{tr}(\mathbf{C}^{(t)}\mathbf{A}_{\hat{\theta}}\boldsymbol{\Sigma}^{(t)}(\mathbf{C}^{(t)}\mathbf{A}_{\hat{\theta}})^H) + (a-1) \ln \beta - b\beta + \text{const} \tag{19}
 \end{aligned}$$

which, through differentiation, gives the update for $\beta^{(t+1)}$:

$$\beta^{(t+1)} = \frac{\hat{K}M + (a-1)}{\left(b + \sum_{k=1}^{\hat{K}} \|\hat{\mathbf{y}}_k - \mathbf{C}^{(t)}\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)}\|_2^2 + \hat{K} \text{tr}(\mathbf{C}^{(t)}\mathbf{A}_{\hat{\theta}}\boldsymbol{\Sigma}^{(t)}(\mathbf{C}^{(t)}\mathbf{A}_{\hat{\theta}})^H) \right)}. \tag{20}$$

2. For δ , also ignoring terms in the logarithm independent thereof, we have:

$$\begin{aligned}
 & \left\langle \ln p(\hat{\mathbf{S}}_{\hat{\theta}}|\boldsymbol{\delta})p(\boldsymbol{\delta}) \right\rangle_{E^{(t)}} \\
 &= \left\langle \ln p(\hat{\mathbf{S}}_{\hat{\theta}}|\boldsymbol{\delta}) \right\rangle_{E^{(t)}} + \ln p(\boldsymbol{\delta}) \\
 &= \hat{K} \ln |\boldsymbol{\Delta}| - \sum_{k=1}^{\hat{K}} \left\langle \hat{\mathbf{s}}_k^H \boldsymbol{\Delta} \hat{\mathbf{s}}_k \right\rangle_{E^{(t)}} - d \sum_{i=1}^{K_{\hat{\theta}}} \delta_i + (c-1) \sum_{i=1}^{K_{\hat{\theta}}} \ln \delta_i + \text{const} \\
 &= (\hat{K} + c - 1) \sum_{i=1}^{K_{\hat{\theta}}} \ln \delta_i - d \sum_{i=1}^{K_{\hat{\theta}}} \delta_i - \sum_{k=1}^{\hat{K}} \text{tr} \left(\left(\boldsymbol{\mu}_k^{(t)} (\boldsymbol{\mu}_k^{(t)})^H + \boldsymbol{\Sigma}^{(t)} \right) \boldsymbol{\Delta} \right) + \text{const}. \tag{21}
 \end{aligned}$$

Setting the derivative of Equation (21), with respect to δ , to zero and solving for each δ_i gives the updates:

$$\delta_i^{(t+1)} = \frac{\hat{K} + c - 1}{d + \sum_{k=1}^{\hat{K}} [\boldsymbol{\Xi}_k^{(t)}]_{ii}}, \quad i = 1, 2, \dots, K_{\hat{\theta}} \tag{22}$$

where $\boldsymbol{\Xi}_k^{(t)} \triangleq \boldsymbol{\mu}_k^{(t)} (\boldsymbol{\mu}_k^{(t)})^H + \boldsymbol{\Sigma}^{(t)}$.

3. For \mathbf{C} , its estimate should maximize the following expected value:

$$\left\langle \ln p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) \right\rangle_{E^{(t)}} = -\beta \sum_{k=1}^{\hat{K}} \|\hat{\mathbf{y}}_k - \mathbf{C}\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)}\|_2^2 - \beta \hat{K} \text{tr}(\mathbf{C}\mathbf{A}_{\hat{\theta}}\boldsymbol{\Sigma}^{(t)}(\mathbf{C}\mathbf{A}_{\hat{\theta}})^H) + \text{const}. \tag{23}$$

In order to calculate the derivative of Equation (23), we need the following lemma.

Lemma 1 (see [2]). *Let \mathbf{C} and \mathbf{c} be defined as earlier, then for any vector \mathbf{x} , we have:*

$$\mathbf{C}\mathbf{x} = \mathbf{T}(\mathbf{x})\mathbf{c} \tag{24}$$

where the matrix $\mathbf{T}(\mathbf{x}) \in \mathbb{C}^{M \times (m+1)}$ is the sum of the two $M \times (m+1)$ matrices:

$$[\mathbf{T}_1(\mathbf{x})]_{p,q} = \begin{cases} [\mathbf{x}]_{p+q-1} & p+q \leq M+1 \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

$$[\mathbf{T}_2(\mathbf{x})]_{p,q} = \begin{cases} [\mathbf{x}]_{p-q+1} & p \geq q \geq 2 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

With the assistance of Lemma 1, we are capable of rewriting the terms $\|\hat{\mathbf{y}}_k - \mathbf{C}\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)}\|_2^2$ and $\text{tr}(\mathbf{C}\mathbf{A}_{\hat{\theta}}\boldsymbol{\Sigma}^{(t)}(\mathbf{C}\mathbf{A}_{\hat{\theta}})^H)$ in Equation (23) as:

$$\|\hat{\mathbf{y}}_k - \mathbf{C}\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)}\|_2^2 = \|\hat{\mathbf{y}}_k - \mathbf{T}(\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)})\mathbf{c}\|_2^2 \tag{27}$$

and:

$$\text{tr}(\mathbf{C}\mathbf{A}_{\hat{\theta}}\boldsymbol{\Sigma}^{(t)}(\mathbf{C}\mathbf{A}_{\hat{\theta}})^H) = \|\mathbf{C}\mathbf{D}^{(t)}\|_2^2 = \sum_{i=1}^M \|\mathbf{C}\mathbf{d}_i^{(t)}\|_2^2 = \sum_{i=1}^M \|\mathbf{T}(\mathbf{d}_i^{(t)})\mathbf{c}\|_2^2 \tag{28}$$

respectively, where $\mathbf{D}^{(t)}(\mathbf{D}^{(t)})^H$ stands for a decomposition of $\mathbf{A}_{\hat{\theta}}\boldsymbol{\Sigma}^{(t)}\mathbf{A}_{\hat{\theta}}^H$ and $\mathbf{d}_i^{(t)}$ denotes the i -th column of $\mathbf{D}^{(t)}$. Using Equations (27) and (28), we can calculate the derivative of $\langle \ln p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) \rangle_{p^{(t)}}$, with respect to \mathbf{c} , as:

$$\begin{aligned} & \frac{\partial}{\partial \mathbf{c}} \langle \ln p(\hat{\mathbf{Y}}|\hat{\mathbf{S}}_{\hat{\theta}}, \beta; \mathbf{C}) \rangle_{p^{(t)}} \\ &= \beta \sum_{k=1}^{\hat{K}} \mathbf{T}(\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)})^H (\hat{\mathbf{y}}_k - \mathbf{T}(\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)})\mathbf{c}) - \beta \hat{K} \sum_{i=1}^M \mathbf{T}(\mathbf{d}_i^{(t)})^H \mathbf{T}(\mathbf{d}_i^{(t)})\mathbf{c}. \end{aligned} \tag{29}$$

Setting the derivative to zero gives:

$$\mathbf{c} = \left(\sum_{k=1}^{\hat{K}} \mathbf{T}(\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)})^H \mathbf{T}(\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)}) + \hat{K} \sum_{i=1}^M \mathbf{T}(\mathbf{d}_i^{(t)})^H \mathbf{T}(\mathbf{d}_i^{(t)}) \right)^{-1} \cdot \sum_{k=1}^{\hat{K}} \mathbf{T}(\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)})^H \hat{\mathbf{y}}_k. \tag{30}$$

Remark 1. Note that [9–11] provided an alternative parameterization for the steering vector:

$$\mathbf{C}\boldsymbol{\alpha}(\theta_k) = g(\theta_k, \mathbf{c})\bar{\mathbf{T}}(\theta_k)\boldsymbol{\tau}(\theta_k, \mathbf{c}) \tag{31}$$

where $g(\theta_k, \mathbf{c}) \triangleq \sum_{m'=-m}^m c_{|m'|} e^{jm'\phi(\theta_k)}$, $\bar{\mathbf{T}}(\theta_k) \triangleq \text{blkdiag}\{\bar{\mathbf{T}}_1(\theta_k), \bar{\mathbf{T}}_2(\theta_k), \bar{\mathbf{T}}_3(\theta_k)\}$ is a block diagonal matrix with $\bar{\mathbf{T}}_1(\theta_k) = \text{diag}\{1, e^{j\phi(\theta_k)}, \dots, e^{j(m-1)\phi(\theta_k)}\}$, $\bar{\mathbf{T}}_2(\theta_k) = [e^{jm\phi(\theta_k)}, \dots, e^{j(M-m-1)\phi(\theta_k)}]^T$ and $\bar{\mathbf{T}}_3(\theta_k) = \text{diag}\{e^{j(M-m)\phi(\theta_k)}, \dots, e^{j(M-1)\phi(\theta_k)}\}$ and $\boldsymbol{\tau}(\theta_k, \mathbf{c}) \triangleq [\nu_1, \dots, \nu_m, 1, \eta_1, \dots, \eta_m]$ with $\nu_k = \frac{1}{g(\theta_k, \mathbf{c})} (\sum_{m'=1-k}^m c_{|m'|} e^{jm'\phi(\theta_k)})$ and $\eta_k = \frac{1}{g(\theta_k, \mathbf{c})} (\sum_{m'=-m}^{m-k} c_{|m'|} e^{jm'\phi(\theta_k)})$. However, it is unlikely that Equation (31) can be applied to the M -step for \mathbf{C} . The most difficult aspect for applying the alternative parameterization in Equation (31) is that it is dependent on the geometric progression in $\boldsymbol{\alpha}(\theta_k)$; while the terms $\mathbf{A}_{\hat{\theta}}\boldsymbol{\mu}_k^{(t)}$ and $\mathbf{d}_i^{(t)}$ in Equations (27) and (28) do not have the property of geometric progression.

The EM algorithm proceeds by repeated application of Equations (20), (22) and (30), until a prescribed accuracy is achieved. Once the algorithm is convergent, we are able to obtain the calibrated \mathbf{C} . As far as the estimated δ_i 's are concerned, we observed that many of δ_i tend to infinity in the EM process. Clearly, the rows corresponding to these infinity δ_i 's are zero; while the rows corresponding to small δ_i 's imply the true DOAs.

4. Simulation Results

In this section, we will present several simulation results to illustrate the performance of our proposed method. We will compare the proposed method to the original SBL method in [24], the iterative method in [3] and the auxiliary methods in [5,12].

Simulation 1 addresses the performance comparisons of the joint estimation of DOAs and mutual coupling coefficients. Consider a scenario where a ULA composed of $M = 10$ sensors is used to receive $K = 2$ uncorrelated signals coming from $\theta_1 = -19.7^\circ\text{C}$ and $\theta_2 = 10.1^\circ\text{C}$. Note that the effect of mutual coupling is negligible between two sensors that are far enough away from each other, because the mutual coupling coefficient is inversely proportional to their distance. Hence, it is reasonable to approximate the mutual coupling effect with just a few nonzero coefficients. In the simulation, we assume that the number of mutual coupling coefficients is $m = 2$ with $c_1 = 0.5 - 0.4i$ and $c_2 = 0.3 + 0.1i$. Figure 1 shows the root mean square error (RMSE) of DOA estimation *versus* input SNR computed via 200 Monte Carlo runs, where the number of snapshots is 100. For the ease of comparison, we also include the curve for CRB. As can be seen from the figure, our method outperforms the state-of-the-art methods. This is because: (1) compared to the method in [24], our method utilizes a hierarchical form of the Student t prior to enforce the sparsity of unknown signal more heavily; moreover, our method uses the SVD to reduce the sensitivity to the noise; (2) compared to the methods in [5,12], our method uses the whole array, rather than a subarray, to estimate DOAs and to compensate for the mutual coupling effect.

Figure 2 shows the estimation bias and variance for each DOA. Compared to the l_1 -norm regularized method [12], our method can significantly reduce the DOA estimation bias, as well as the variance. To assess the performance of mutual coupling calibration, Figure 3 shows the relative RMSE of the estimation of the mutual coupling coefficients. This relative RMSE is defined as:

$$\sqrt{\frac{1}{200} \sum_{n=1}^{200} \sum_{i=1}^m \frac{|c_{ni} - c_i|^2}{|c_i|^2}} \times 100\% \quad (32)$$

where c_{ni} is the estimate of c_i at the n -th Monte Carlo run. As shown from the Figure 3, the mutual coupling coefficients can be estimated more accurately in our method, especially for low SNR signals. Hence, if all of the methods embed the mutual coupling coefficients within the same classical eigendecomposition algorithm, such as MUSIC and ESPRIT, to estimate the DOA, our method can achieve the best performance.

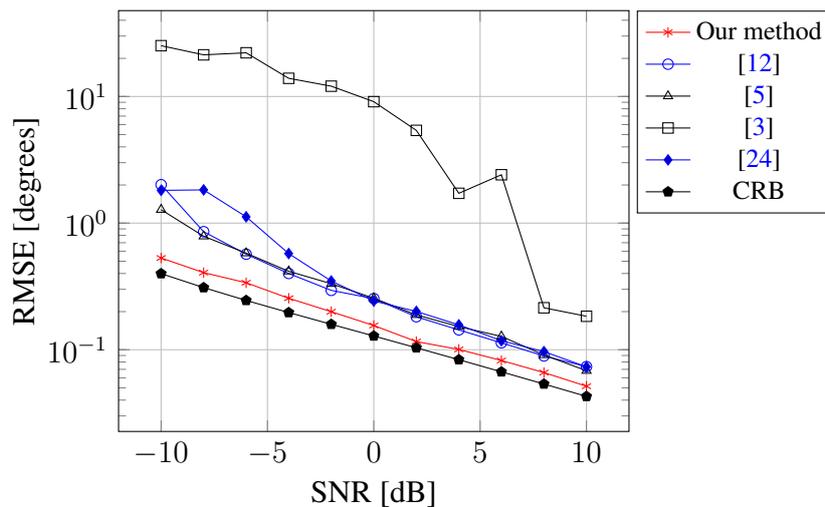


Figure 1. RMSE of the DOA estimate against SNR.

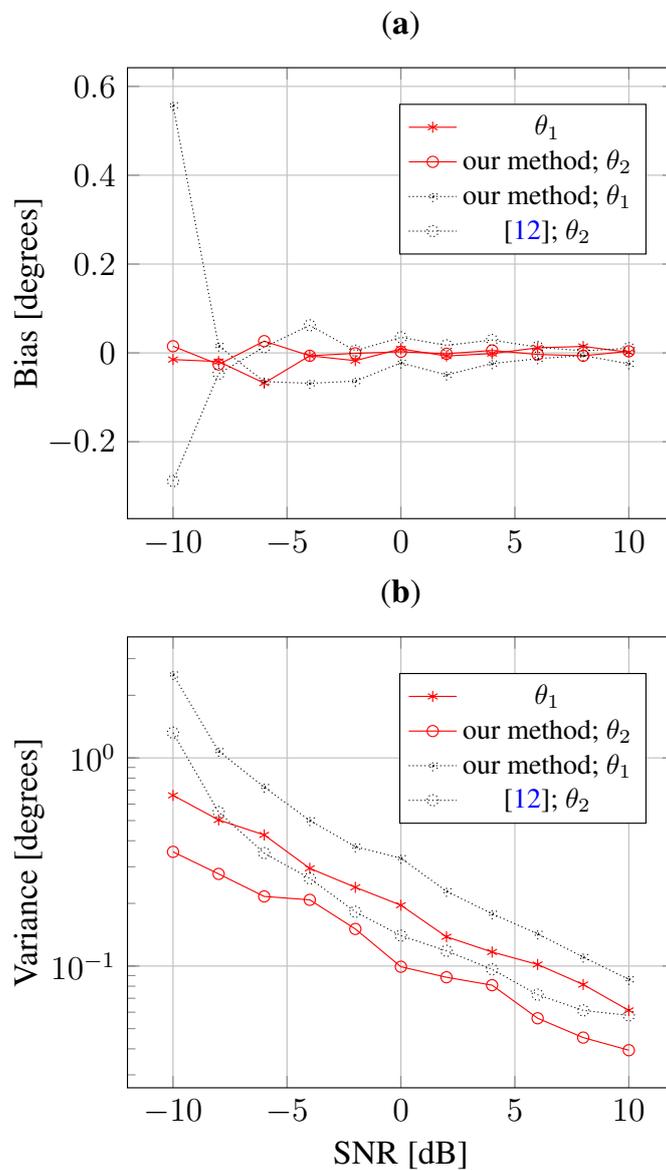


Figure 2. Estimation bias and variance against SNR. (a) Bias; (b) Variance.

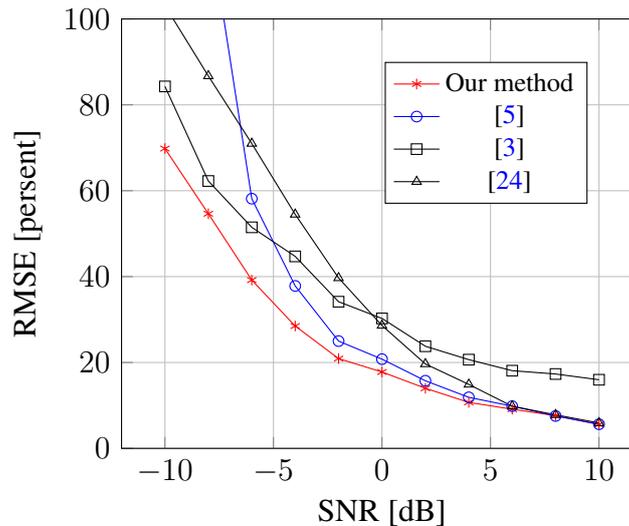


Figure 3. RMSE of mutual coupling coefficients against SNR.

Simulation 2 is used to investigate the capability of resolving closely-spaced sources with limited snapshots ($T = 50$) for several methods. The simulation considers a scenario where a ULA composed of $M = 13$ sensors is used to receive $K = 2$ uncorrelated signals coming from $\theta_1 = -2.5^\circ\text{C}$ and $\theta_2 = 3.5^\circ\text{C}$, and the number of mutual coupling coefficients is $m = 2$ with $c_1 = 0.5 - 0.4i$ and $c_2 = 0.3 + 0.1i$. We say that the two signals are exactly resolved in a given run, if $\max_{k=1,2}\{|\tilde{\theta}_k - \theta_k|\}$ is smaller than $|\theta_1 - \theta_2|/2$, where $\tilde{\theta}_k$ stands for the estimated DOA for the k -th signal. As can be seen from Figure 4, the resolution performance of our method outperforms others. The superior resolution performance of sparse representation is natural, which is consistent with the simulation results in [12].

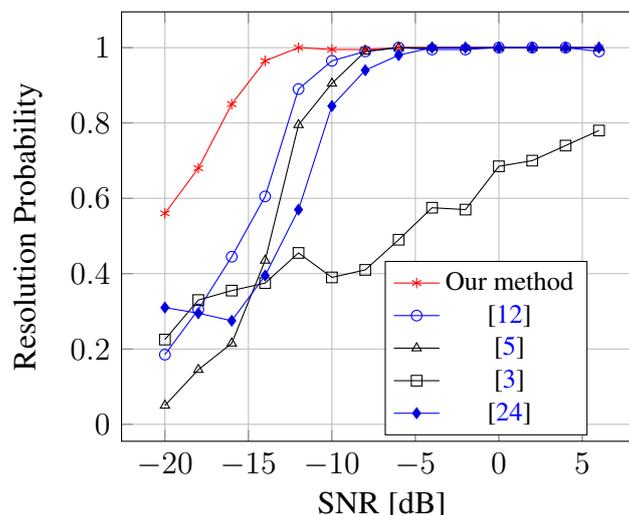


Figure 4. Resolution probability against SNR for closely-spaced sources.

Simulation 3 addresses the “blind angle” phenomenon, which may result in the substantially degraded performance of DOA estimation [4,5]. We consider a scenario where a ULA composed of $M = 12$ sensors is used to receive $K = 4$ uncorrelated signals coming from $\theta_1 = 10^\circ\text{C}$, $\theta_2 = 20^\circ\text{C}$, $\theta_3 = 30^\circ\text{C}$ and $\theta_4 = 40^\circ\text{C}$, and the number of mutual coupling coefficients is $m = 2$ with $c_1 = 0.6 + 0.5i$ and $c_2 = 0.3844 - 0.3476i$. It is easy to verify that the blind angle occurs at $\theta_4 = 40^\circ\text{C}$. The number of

snapshots is 30, and the SNR is 10 dB. Figure 5 illustrates that: (1) our method can estimate all of the true DOAs accurately, including the blind DOA $\theta_4 = 40^\circ$; (2) the l_1 -norm regularized method [12] fails in estimating the blind DOA from $\theta_4 = 40^\circ$, but succeeds in obtaining the DOAs from other directions; (3) the iterative method [3] misses two DOAs $\theta_3 = 30^\circ$ and $\theta_4 = 40^\circ$. Obviously, our method has the best performance of “blind angle” suppression.

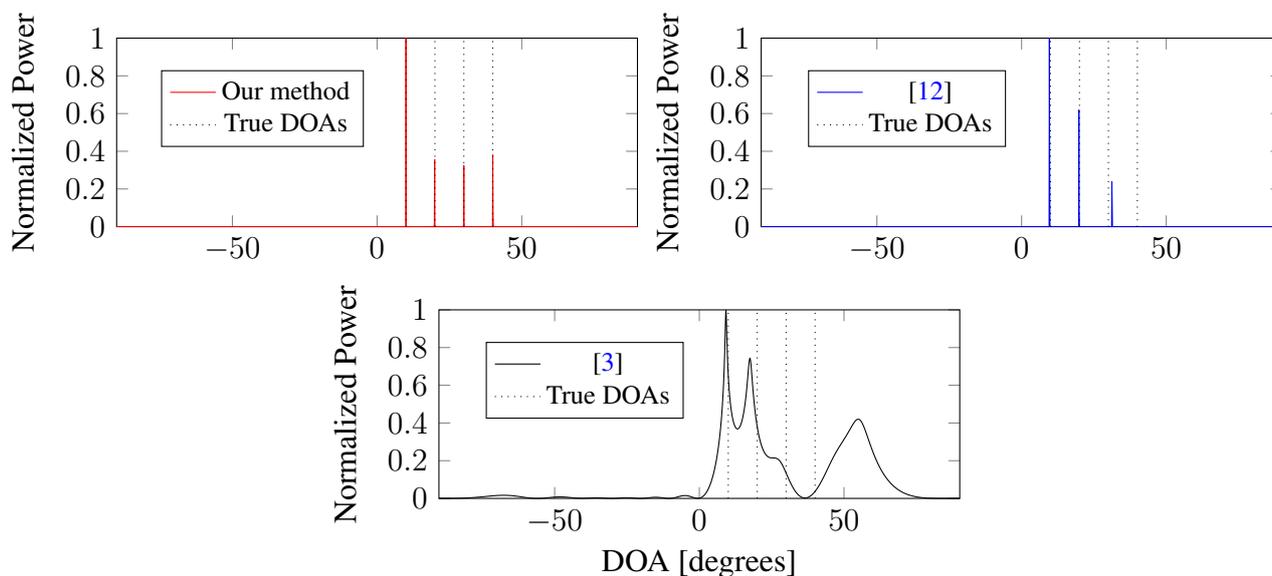


Figure 5. DOA estimation in the case of a “blind angle”.

5. Conclusions

We have proposed a modified SBL method that approaches the problem of DOA estimation with unknown mutual coupling. Unlike the original SBL method in [24] that only uses stationary priors, our new method utilized a hierarchical form of the Student t prior to enforce the sparsity of unknown signal more heavily. To efficiently perform the Bayesian inference, we adopted a refined EM algorithm that treats \hat{S}_θ and C as a hidden variable and a parameter, respectively. It can update the mutual coupling coefficients more efficiently. Another difference is that our method used an additional SVD to reduce the computational complexity of the signal reconstruction process and the sensitivity to the measurement noise. Hence, the proposed method is expected to give a better estimation performance. Simulation results have verified its efficiency.

Acknowledgments

Jisheng Dai’s work is supported in part by National Natural Science Foundation of China under Project 61571211, and in part by the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University under Project 2013D08.

Weichao Xu’s work is supported in part by National Natural Science Foundation of China under Project 61271380, and in part by the Guangdong Natural Science Foundation under Project S2012010009870 and 2014A030313515.

Author Contributions

J.D. and N.H. provided the idea of this work. C.C. conceived of and designed the experiments. J.D. and W.X. performed the experiments and provided all of the figures and data for the paper. C.C. prepared the literature and analyzed the data. J.D. and W.X. wrote the paper. Correspondence and requests for the paper should be addressed to J.D..

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Krim, H.; Viberg, M. Two decades of array signal processing research. *IEEE Signal Proc. Mag.* **1996**, *13*, 67–94.
2. Friedlander, B.; Weiss, A. Direction finding in the presence of mutual coupling. *IEEE Trans. Antennas Propag.* **1991**, *39*, 277–284.
3. Sellone, F.; Serra, A. A novel online mutual coupling compensation algorithm for uniform and linear arrays. *IEEE Trans. Signal Proc.* **2007**, *55*, 560–573.
4. Ye, Z.; Liu, C. On the resiliency of MUSIC direction finding against antenna sensor coupling. *IEEE Trans. Antennas Propag.* **2008**, *56*, 371–380.
5. Ye, Z.; Dai, J.; Xu, X.; Wu, X. DOA estimation for uniform linear array with mutual coupling. *IEEE Trans. Aerosp. Electron. Syst.* **2009**, *45*, 280–288.
6. Dai, J.; Ye, Z. Spatial smoothing for DOA estimation of coherent signals in the presence of unknown mutual coupling. *IET Signal Proc.* **2011**, *5*, 418–425.
7. Dai, J.; Xu, W.; Zhao, D. Real-valued DOA estimation for uniform linear array with unknown mutual coupling. *Signal Proc.* **2012**, *92*, 2056–2065.
8. Wang, W.; Ren, S.; Ding, Y.; Wang, H. An efficient algorithm for direction finding against unknown mutual coupling. *Sensors* **2014**, *14*, 20064–20077.
9. Liao, B.; Zhang, Z.; Chan, S. DOA estimation and tracking of ULAs with mutual coupling. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 891–905.
10. Liao, B.; Chan, S. Adaptive beamforming for uniform linear arrays with unknown mutual coupling. *IEEE Antennas Wirel. Propag. Lett.* **2012**, *11*, 464–467.
11. Liao, B.; Chan, S. A cumulant-based approach for direction finding in the presence of mutual coupling. *Signal Proc.* **2014**, *104*, 197–202.
12. Dai, J.; Zhao, D.; Ji, X. A sparse representation method for DOA estimation with unknown mutual coupling. *IEEE Antennas Wirel. Propag. Lett.* **2012**, *11*, 1210–1213.
13. Hu, N.; Ye, Z.; Xu, X.; Bao, M. DOA estimation for sparse array via sparse signal reconstruction. *IEEE Trans. Aerosp. Electron. Syst.* **2013**, *49*, 760–773.
14. Malioutov, D.; Cetin, M.; Willsky, A. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Proc.* **2005**, *53*, 3010–3022.
15. Dai, J.; Xu, X.; Zhao, D. Direction-of-arrival estimation via real-valued sparse representation. *IEEE Antennas Wirel. Propag. Lett.* **2013**, *12*, 376–379.

16. Zhao, L.; Bi, G.; Wang, L.; Zhang, H. An improved auto-calibration algorithm based on sparse Bayesian learning framework. *IEEE Signal Proc. Lett.* **2013**, *20*, 889–892.
17. Yang, Z.; Xie, L.; Zhang, C. Off-grid direction of arrival estimation using sparse bayesian inference. *IEEE Trans. Signal Proc.* **2013**, *61*, 38–43.
18. Carlin, M.; Rocca, P.; Oliveri, G.; Viani, F.; Massa, A. Directions-of-arrival estimation through Bayesian compressive sensing strategies. *IEEE Trans. Antennas Propag.* **2013**, *61*, 3828–3838.
19. Tipping, M. Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
20. Wipf, D.; Rao, B. Sparse bayesian learning for basis selection. *IEEE Trans. Signal Proc.* **2004**, *52*, 2153–2164.
21. Ji, S.; Xue, Y.; Carin, L. Bayesian compressive sensing. *IEEE Trans. Signal Proc.* **2008**, *56*, 2346–2356.
22. Herman, M.; Strohmer, T. General deviants: An analysis of perturbations in compressed sensing. *IEEE J. Sel. Top. Signal Proc.* **2010**, *4*, 342–349.
23. Yang, Z.; Zhang, C.; Xie, L. Robustly stable signal recovery in compressed sensing with structured matrix perturbation. *IEEE Trans. Signal Proc.* **2012**, *60*, 4658–4671.
24. Liu, Z.; Zhou, Y. A unified framework and sparse bayesian perspective for direction-of-arrival estimation in the presence of array imperfections. *IEEE Trans. Signal Proc.* **2013**, *61*, 3786–3798.
25. Wipf, D.; Palmer, J.; Rao, B. Perspectives on sparse bayesian learning. *Adv. Neural Inf. Proc. Syst.* **2004**, *16*, 1–8.
26. Gupta, L.; Ksienski, A. Effect of mutual coupling in the performance of adaptive arrays. *IEEE Trans. Antennas Propag.* **1983**, *31*, 785–791.
27. Hui, H. A new definition of mutual impedance for application in dipole receiving antenna arrays. *IEEE Antennas Wirel. Propag. Lett.* **2004**, *3*, 364–367.
28. Edelman, A. Random matrix theory. *Acta Numer.* **2005**, *14*, 1–65.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).