DATA-ADAPTIVE REGULARIZATION FOR DOA ESTIMATION USING SPARSE SPECTRUM FITTING

J. Zheng and M. Kaveh

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN USA

ABSTRACT

Regularization parameter selection is critical to the performance of many sparsity-exploiting Direction-Of-Arrival (DOA) estimation algorithms. In this paper, we propose an automatic selector for choosing this parameter in the DOA estimation algorithm presented in [1], which is based on the analysis of its optimality conditions. This selector requires very limited prior information and is computationally efficient. Through simulation examples, the effectiveness and robustness of the selector are illustrated.

Index Terms— Direction-Of-Arrival, Sparse Representation, Regularization Parameter Selection

1. INTRODUCTION

Following the broad appeal and wide-ranging developments in Sparse Signal Recovery (SSR) algorithms [2], a number of sparsity-exploiting Direction-Of-Arrival (DOA) estimation methods have also emerged, e.g. [1] [3][4]. These methods usually use L_1-norm regularization with an associated regularization parameter. The level of regularization through the selection of such a parameter is critical to the effectiveness of the associated estimator. Unfortunately, the range of good values for the parameters are dependent on the true DOAs and signal powers, and need to be carefully tuned for each realization. In this paper, we focus on the development of a computationally-efficient data-adaptive selection mechanism for determining the regularization parameter of a particular DOA estimation algorithm proposed in [1], which is referred to in this paper by Sparse Spectrum Fitting (SpSF) estimator.

Consider an antenna array of M elements and assume L narrow-band far-field signals impinge on the array from directions $\theta_1, ..., \theta_L$ in the presence of additive white Gaussian noise of variance σ^2. The spatial covariance matrix $\mathbf{Q} \in \mathbb{C}^{M\times M}$ for uncorrelated sources is defined by

$$\mathbf{Q} = \sum_{i=1}^{L} s_i a(\theta_i) a^H(\theta_i) + \sigma^2 \mathbf{I}, \tag{1}$$

where $a(\theta) \in \mathbb{C}^{M\times 1}$ is the steering vector at direction θ, "^{H}" is conjugate transpose, $s_i > 0$ is the power of the signal from direction θ_i and $\mathbf{I}$ is the identity matrix. Generally, the DOA estimation methods involve a search over a grid of many candidate directions, which are denoted by $\phi_1, ..., \phi_K$ with $K \gg L$. If the true DOAs belong to this grid, $\theta_i \in \{\phi_1, ..., \phi_K\}$ for $i = 1, ..., L$, (1) can be rewritten as

$$\mathbf{Q} = \sum_{k=1}^{K} p_k^\circ a(\phi_k) a^H(\phi_k) + \sigma^2 \mathbf{I}, \tag{2}$$

where $p_k^\circ = s_i$ if $\exists \theta_i = \phi_k$, otherwise $p_k^\circ = 0$. By defining $a_v(\phi) = vec(a(\phi)a^H(\phi))$ and $\mathbf{A}_K = [a_v(\phi_i), ..., a_v(\phi_K)]$, (2) can be reformulated as:

$$\mathbf{Q}_v = \mathbf{A}_K p^\circ + \sigma^2 \mathbf{I}_v, \tag{3}$$

where $\mathbf{Q}_v = vec(\mathbf{Q})$, $\mathbf{I}_v = vec(\mathbf{I})$, $p^\circ = [p_1^\circ, ..., p_K^\circ]^T$ and $vec(\cdot)$ represents the vectorization operation. In practice, $\mathbf{Q}$, which is unknown, is replaced by the sample covariance matrix $\mathbf{R} = \frac{1}{N}\sum_{t=1}^{N} y(t)y^H(t)$, where $y(t) \in \mathbb{C}^{M\times 1}$ is the array snapshot at time t and N is the number of snapshots. Using an "error term" $\mathbf{E}$ to summarize noise contribution, sample estimates and any other modeling errors in $\mathbf{R}$, we have:

$$\mathbf{R}_v = \mathbf{A}_K p^\circ + \mathbf{E}_v, \tag{4}$$

where $\mathbf{R}_v = vec(\mathbf{R})$ and $\mathbf{E}_v = vec(\mathbf{E})$. By recognizing (4) as a sparse representation, SpSF is formulated as

$$\begin{aligned} p^* = arg\min_{p} \ & \|\mathbf{R}_v - \mathbf{A}_K p\|_2^2 + \beta\|p\|_1 \\ & s.t.\ p_i \geqslant 0, \ \ i = 1, 2, ...K, \end{aligned} \tag{5}$$

where β is a regularization parameter. Note that Equation (5) is an equivalent formulation of SpSF as the one proposed in [1].

In Section 2, the contribution of this paper and its relation to prior works are discussed. Following that, in Section 3, the optimality conditions of SpSF are analyzed and a Monte Carlo process is proposed to efficiently evaluate the probability of perfect support recovery of SpSF. An upper bound on this probability is formulated in Section 4 which, in combination with the Monte Carlo evaluation process, serves as the basis for the automatic β-selector. Simulation results illustrating the effectiveness and the robustness of the selector are presented in Section 5. Finally, the conclusions are made in Section 6.

2. RELATION TO PRIOR WORK

The β-selector proposed in this paper is based on the analysis of the optimality conditions of SpSF, which follows the derivations presented in [5]. Similarly, based on its optimality conditions, the work [6] analyzed the probability of perfectly recovering the support of p° by SpSF and proposed an evaluation process which is similar to the idea of Monte Carlo evaluation process presented in this paper. Further, an explicit formula of the range of βs, which can give perfect support recovery of p°, was presented in [6]. However, such range is usually very small if compared to that for correct DOA estimates. More importantly, the formula relies on the exact knowledge of the true DOAs and, therefore, cannot be used in practical applications. Another choice for selecting β can be Cross Validation (CV). Although CV has been successfully used in SSR, its computational complexity is prohibitive for DOA estimation problems. In this paper, we propose an upper bound on the probability of correct support recovery of SpSF, based on which an automatic β-selector for SpSF is constructed. This selector only requires very limited information and is computationally efficient. Because of limited space, the details of the derivations are omitted in this paper, but can be found in [7].

3. EVALUATION OF THE PROBABILITY OF MSC SOLUTIONS

In this section, we present a Monte Carlo simulation process to efficiently evaluate the probability of perfect support recovery of p° by SpSF, which is obtained through the analysis of the optimality conditions of SpSF. This section starts with some necessary definitions.

The true sparse spatial spectrum is denoted as $p^\circ = [p_1^\circ, ..., p_K^\circ]^T$ where, without loss of generality, only the first L elements are assumed to be non-zero and constitute the subvector $p_{(1)}^\circ = [p_1^\circ, ..., p_L^\circ]^T$. The other elements in p° constitute the subvector $p_{(2)}^\circ$. Similarly, we denote the solutions of SpSF by p^*, and its subvectors by $p_{(1)}^*$ and $p_{(2)}^*$ respectively. **A solution p^* is called Model Selection Consistent (MSC) if $\mathbf{sign}(\mathbf{p_i^*}) = \mathbf{sign}(\mathbf{p_i^\circ})$ for $\mathbf{i = 1, ..., K}$** [5]. **Such solutions provide perfect support recovery of $\mathbf{p^\circ}$ and the probability of the existence of such solutions (in regardless of the value of β) is denoted as $\mathbf{P_{MSC}}$.** Further, for simplicity in derivations, we define $u^* = p^* - p^\circ$ and its corresponding subvectors $u_{(1)}^*$ and $u_{(2)}^*$. By partitioning $\mathbf{A}_K$ into submatrices $\mathbf{A}_{K,(1)}$ and $\mathbf{A}_{K,(2)}$, which consist of the first L and the last $K - L$ columns of $\mathbf{A}_K$ respectively, we can define: for $i, j \in \{1, 2\}$

$$\mathbf{C}_{ij} = \Re[\mathbf{A}_{K,(i)}^H \mathbf{A}_{K,(j)}]$$
$$b_i = \Re[\mathbf{A}_{K,(i)}^H \mathbf{E}_v],$$

where $\Re[\cdot]$ represents the real part and $b = [b_1^T, b_2^T]^T$.

Based on the above definitions, the sufficient and necessary condition for SpSF's MSC solutions is that: $\exists u_{(1)}^*$ satisfying the following element-wise inequalities

$$u_{(1)}^* > -p_{(1)}^\circ \tag{6a}$$
$$\mathbf{C}_{11} u_{(1)}^* - b_1 = -0.5\beta 1_L \tag{6b}$$
$$\mathbf{C}_{21} u_{(1)}^* - b_2 \geqslant -0.5\beta 1_{K-L}. \tag{6c}$$

By the Equation (6b),

$$u_{(1)}^* = \mathbf{C}_{11}^{-1}(b_1 - 0.5\beta 1_L), \tag{7}$$

and the condition can be simplified to: $\exists\beta$ satisfying

$$\beta \mathbf{C}_{11}^{-1} 1_L \leqslant 2\mathbf{C}_{11}^{-1} b_1 + 2p_{(1)}^\circ \tag{8a}$$
$$\beta(\mathbf{C}_{21}\mathbf{C}_{11}^{-1} 1_L - 1_{K-L}) \leqslant 2\mathbf{C}_{21}\mathbf{C}_{11}^{-1} b_1 - 2b_2, \tag{8b}$$

where 1_i, $i \times 1$, is an all-1 vector. In this paper, any β satisfying the inequalities of (8) is also called **MSC** (meaning that at least one of its corresponding solutions is MSC). Since the vectors $\mathbf{C}_{11}^{-1} 1_L$ and $\mathbf{C}_{21}\mathbf{C}_{11}^{-1} 1_L - 1_{K-L}$ contain both negative and positive elements, each of (8a) and (8b) is generally equivalent to two inequalities for β (an upper limit and a lower limit on β). Given a specific realization of $\{b_1, b_2, p_{(1)}^\circ, \theta_1, ..., \theta_L\}$, these four limits and $\beta \geqslant 0$ define a range, S_{MSC}, in which β is MSC. If S_{MSC} is empty, then, for this realization, there is no MSC β and it is impossible for SpSF to find an MSC solution. Therefore, P_{MSC} is equivalent to the probability that S_{MSC} is non-empty.

Although an explicit formula for P_{MSC} can be beneficial, it is very difficult (if not impossible) to obtain for general cases. Thus, we propose to evaluate P_{MSC} through Monte Carlo simulations. Specifically, in N_T independent trials, the signal and the noise samples are randomly generated according to their distribution functions, which are assumed to be known. Then, b_1 b_2 and $p_{(1)}^\circ$ are computed and substituted into (8). Using the values of these variables, S_{MSC} is calculated and the number of trials in which it is non-empty is divided by N_T , providing an estimate of P_{MSC}. N_T is chosen to be sufficiently large so that this estimate becomes accurate enough. Although this evaluation process for P_{MSC} is computationally efficient, it is not useful for practical applications since it requires exact knowledge of the true DOAs and the distribution functions. However, the optimality conditions presented in this section and the idea of the Monte Carlo simulation process serve as the basis for the automatic β-selector proposed in next section.

4. DATA-ADAPTIVE SELECTOR OF β

As a regularized optimization problem, the selection of the regularization parameter β of SpSF is critical to its estimation performance. Generally, the "good" values of β are dependent on the DOAs, signal powers and etc. Thus, automatic

selection of β is highly desired. In this section, we present an upper bound on P_{MSC}, based on which an automatic and computationally efficient β-selector is proposed.

According to its definition, P_{MSC} is always upper bounded by P_θ which is the probability that SpSF can find a solution (either MSC or not) giving correct DOA estimates. Here, a solution p^* gives correct DOA estimates if $min\{p^*_{(1)}\} > max\{p^*_{(2)}\}$, and the corresponding β is called **"proper"**. As in (8), the optimality conditions of such solutions define a range of *proper* β, which is a superset of S_{MSC}, and in order to obtain an upper bound on P_{MSC} we try to enlarge S_{MSC} by using these optimality conditions.

A solution giving correct DOA estimates has to satisfy: $\exists u^*_{(1)}$ and $u^*_{(2)}$ such that

$$2\mathbf{C}_{21}u^*_{(1)} + 2\mathbf{C}_{22}u^*_{(2)} - 2b_2 \geqslant -\beta 1_{K-L}. \tag{9}$$

If p^* is not MSC, $u^*_{(2)} = p^*_{(2)} \geqslant 0$ is non-zero and $\mathbf{C}_{22}u^*_{(2)} \geqslant 0$ since all the elements of $\mathbf{C}_{22}$ are non-negative (easy to prove). Thus, (9) is actually a relaxation of (6c) and, by replacing (6c) with (9), one can obtain an upper bound on P_{MSC}, which unfortunately cannot be evaluated since $u^*_{(2)}$ is unknown. *Therefore, we propose to use a randomly generated non-negative vector to replace the unknown $u^*_{(2)}$ and remove the randomness through the Monte Carlo evaluation process of the upper bound.* Such random vector is chosen to be $(min\{\mathbf{C}_{11}^{-1}b_1 + p^\circ_{(1)}\} - 0.5\beta min\{\mathbf{C}_{11}^{-1}1_L\})e_G$, which is non-negative by (7), and $e_G \in \mathbb{R}^{K-L\times 1}$ is a vector of 0s except for its $r_1^{th}, ..., r_G^{th}$ elements being 1. Here, G is a parameter to be specified by the user and $r_1, ..., r_G$ are the indices randomly generated in each trial of the Monte Carlo process. Using such a random vector to substitute $u^*_{(2)}$ in (9), we obtain

$$2\mathbf{C}_{21}u^*_{(1)} + (2m_{bp} - \beta m_{C1})\mathbf{C}_{22}e_G - 2b_2 \geqslant -\beta 1_{K-L}, \tag{10}$$

with $m_{bp} = min\{\mathbf{C}_{11}^{-1}b_1 + p^\circ_{(1)}\}$ and $m_{C1} = min\{\mathbf{C}_{11}^{-1}1_L\}$. Replacing (10) into (6) gives the inequalities for the upper bound:

$$\beta \geqslant 0 \tag{11a}$$

$$\beta\mathbf{C}_{11}^{-1}1_L \leqslant 2\mathbf{C}_{11}^{-1}b_1 + 2p^\circ_{(1)} \tag{11b}$$

$$\begin{aligned}\beta(\mathbf{C}_{21}\mathbf{C}_{11}^{-1}1_L + m_{C1}\mathbf{C}_{22}e_G - 1_{K-L}) \leqslant \\ 2\mathbf{C}_{21}\mathbf{C}_{11}^{-1}b_1 + 2m_{bp}\mathbf{C}_{22}e_G - 2b_2.\end{aligned} \tag{11c}$$

The range of β defined by (11) is denoted as S_{UP} (for any realization, $S_{MSC} \subseteq S_{UP}$) and, a candidate regularization parameter, β_c, is more likely to be "proper" than the others if it falls in S_{UP} with larger probability. Therefore, P_{UP} is an upper bound on P_{MSC}, where P_{UP} is the probability that S_{UP} is non-empty (when $G = 0$, $P_{UP} = P_{MSC}$). Similar to P_{MSC}, P_{UP} can be evaluated through Monte Carlo evaluation, where $r_1, ..., r_G$ need to be regenerated in each of the trials to remove the randomness of e_G.

Based on the upper bound (11), the automatic β-selector for SpSF is presented in the following. Here, we use $f_i(\zeta_i)$ $i = 1, ..., L$ and $f_n(\zeta_n)$ to denote the distribution functions of sources and noise, respectively, and ζs are the parameters of these distributions (e.g. mean and variance for Gaussian distributions). To utilize the upper bound, the β-selector assumes knowledge of L, G, $f_i, i = 1, ..., L$ and f_n and any other information, including θ_i, $\zeta_i, i = 1, ..., L$ and ζ_n, are replaced by rough estimates obtained by a lower resolution spectrum estimator, such as MVDR [8]. Using $\beta_c,\ c = 1, ..., K_\beta$ to denote all the candidate choices of β, the procedure of the selector is:

1. Use a lower resolution spectrum estimator, such as MVDR, to estimate $\tilde{\theta}_i$, $\tilde{\zeta}_i$ and $\tilde{\zeta}_n$, $i = 1, ..., L$
2. Set $Z_c = 0,\ c = 1, ..., K_\beta$
3. for $t = 1, ..., N_T$
   - Randomly generate vector e_G
   - Generate signal and noise samples by $f_i(\tilde{\zeta}_i)$ and $f_n(\tilde{\zeta}_n)$
   - Calculate S_{UP} from (11)
   - If $\beta_c \in S_{UP}$, $Z_c = Z_c + 1$, $c = 1, ..., K_\beta$
4. Find the maximum among Z_cs and denote its index by c^*
5. Output β_{c^*} for SpSF

Note that, this automatic selector, rather than being fixed, adaptively chooses the regularization parameter β based on actual data-generated covariance. Further, as shown in the next section, this method is not sensitive to the values of L and G, and very rough estimates of θs and ζs are generally effective (combined with SpSF) to achieve satisfactory DOA estimation performance. More importantly, the computational cost of this selector is negligible compared with that of Cross Validation and even the cost of the optimization of SpSF itself.

5. SIMULATION RESULTS

In this section, the effectiveness and robustness of the automatic β-selector are illustrated through simulation examples comparing the DOA estimate Root-Mean-Squared-Error (RMSE) of SpSF (using the selector) with those of MUSIC [9], MVDR [8], L_1-SVD [3] and the Cramer-Rao Lower Bound [10]. We consider a ULA with $M = 8$ elements and $d = 0.5$ normalized inter-element distance. $L = 2$ uncorrelated sources impinge on the array from $\theta_1 = -5^\circ$ and $\theta_2 = 5^\circ$ and $N = 300$ snapshots are available. The grid of candidate directions is chosen to be $\{-80^\circ, -79.9^\circ, ..., 80^\circ\}$. Further, the sources and the noise are both assumed to be zero-mean Gaussian, and the noise power σ^2 is fixed at 1. 1000 independent trials are used to evaluate the RMSEs and $N_T = 1000$ for the β-selector.

In Figure 1, we compare the estimation performance of all the methods using the exact value of L. "SpSF (Manual)" is the RMSE of SpSF with the fixed β exhaustively

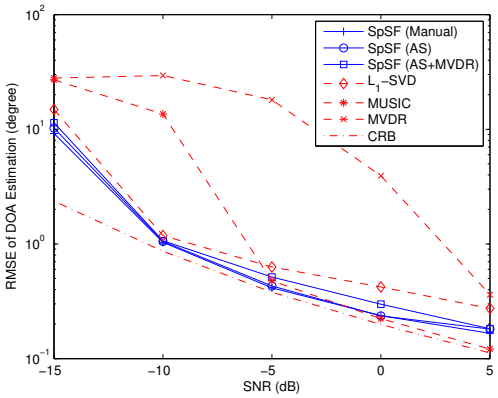

Fig. 1. RMSE of DOA Estimation, $L = 2$ and $G = 4$

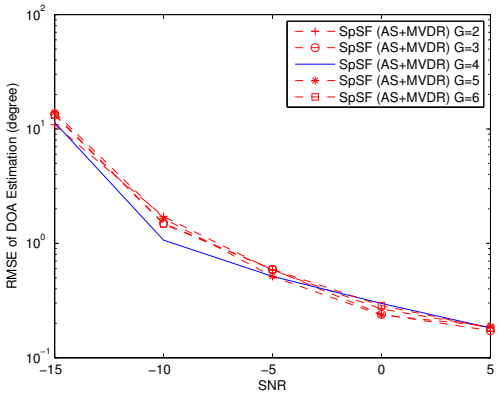

Fig. 2. Robustness with respect to G

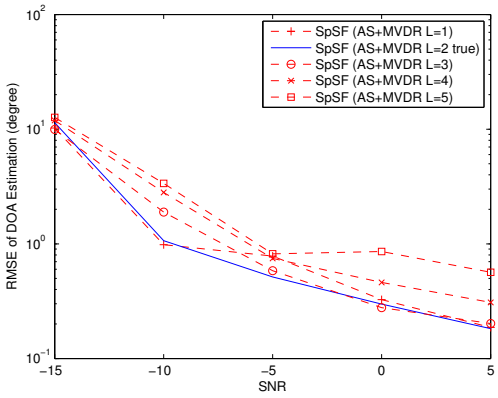

Fig. 3. Robustness with respect to L

searched. "SpSF (AS)" is the RMSE of SpSF combined with the β-selector, using $G = 4$, the true DOAs and the exact distributions. "SpSF (AS+MVDR)" is the same as "SpSF (AS)" except that, **the selector only knows that the sources and the noise are uncorrelated zero-mean Gaussian random variables**. In each trial, it uses the estimated DOAs ($\hat{\theta}_1$ and $\hat{\theta}_2$) by MVDR, the estimated noise power $\hat{\sigma}^2$ by the smallest eigenvalue of $\mathbf{R}$ and the estimated signal powers $\hat{p}_i = p_i^{mvdr} - \hat{\sigma}^2/M$, $i = 1, 2$ where p_i^{mvdr} is the value of the MVDR spectrum at $\hat{\theta}_i$. "MUSIC", "MVDR" and "L_1-SVD" are the RMSEs of the corresponding methods, and the regularization parameter of L_1-SVD is selected in the same way as "SpSF (Manual)". "CRB" stands for the stochastic Cramer-Rao Lower Bound, which uses the knowledge that the sources are uncorrelated. As shown by Figure 1, the βs suggested by the selector using either the true or the initial rough estimates of DOAs and powers can help SpSF achieve almost the same DOA estimation performance as exhaustively searched and fixed regularization parameter ("SpSF (Manual)"). Further, the large error threshold of "SpSF (Manual)", "SpSF (AS)", "SpSF (AS+MVDR)" and L_1-SVD are $5dB$ lower than that of MUSIC.

In Figure 2 and Figure 3, the robustness of the β-selector with respect to the values of G and errors in L is illustrated. In Figure 2, we still assume availability of the exact knowledge of L but use different values of G for "SpSF (AS+MVDR)". In contrast, we keep $G = 4$ in Figure 3 but provide different values of L to the selector. As shown by these figures, SpSF is not sensitive to the values of G and the errors in L. Further, when SNR is relatively large, small perturbations in either G or L cause no significant performance degradation.

6. CONCLUSION

In this paper, based on the analysis of the optimality conditions of SpSF, an upper bound on its probability of correct support recovery and an efficient Monte Carlo evaluation process are presented. Further, by utilizing this upper bound, an automatic selector of the regularization parameter of SpSF is proposed, which requires very limited prior information and is computationally efficient. The effectiveness and robustness (with respect to its parameters) of the selector are illustrated through simulation examples comparing the DOA estimation performance of SpSF with that of MUSIC, L_1-SVD and etc. It is shown that this automatic selector can help SpSF to achieve almost the same performance as exhaustively searched and fixed (ESF) regularization parameters for the cases of uncorrelated sources. Although not presented in this paper, this automatic selector, in combination with SpSF, was also applied to the cases of highly correlated sources in [7] and shown to provide even better DOA estimation performance than using ESF regularization parameters.

7. REFERENCES

[1] J. Zheng, M. Kaveh, and H. Tsuji, "Sparse spectral fitting for direction of arrival and power estimation," in *in Proc. the* 15^{th} *IEEE Workshop on Statistical Signal Process.*, Cardiff, Wales, Sept. 2009.

[2] D.L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization," in *Proc. Natl, Acad. Sci.*, USA, 2003, vol. 100, pp. 2197–2202.

[3] D.M. Malioutov, M. Cetin, and A.S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[4] P. Stoica, P. Babu, and J. Li, "Spice: A sparse covariance-based estimation method for array processing," *IEEE Trans. on Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[5] P. Zhao and B. Yu, "On model selection consistency of lasso," *J. Machine Learning Research*, pp. 2541–2563, Dec. 2006.

[6] J.S. Picard and A.J. Weiss, "Error bounds for convex parameter estimation," *Signal Process.*, vol. 92, no. 5, pp. 1328–1337, Dec. 2011.

[7] J. Zheng and M. Kaveh, "Sparse spatial spectral estimation: A covariance fitting algorithm, performance and regularization," *Submitted to IEEE Trans. on Signal Process.*, Sep. 2012.

[8] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE.*, pp. 1408–1418, Aug. 1969.

[9] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, Rome, NY, Sept. 1979, pp. 243–258.

[10] H.L. Van Trees, *Optimum Array Processing*, John Wiely, 1st edition edition, 2003.