SUPPORT RECOVERY FOR SOURCE LOCALIZATION BASED ON OVERCOMPLETE SIGNAL REPRESENTATION

Gongguo Tang and Arye Nehorai

Electrical and Systems Engineering
Washington University in St. Louis
St. Louis, MO 63130-1127

ABSTRACT

We analyze the performance of a direction-of-arrival (DOA) estimation scheme based on overcomplete signal representation in this paper. We formulate the problem as a support recovery problem with joint sparsity constraints and analyze it in a hypothesis testing framework. We derive both upper and lower bounds on the probability of error by using Chernoff bound and Fano's inequality, respectively. The lower bound implies that the minimal number of samples necessary for accurate DOA estimation is proportional to the logarithm of the discretization level for arbitrary isotropic sensor arrays. We apply the upper bound to study the effect of noise. For uniform linear array (ULA) with only one source, the upper bound exponent indicates that the optimal overcomplete representation is achieved by uniform partition of the wave number space instead of the DOA space.

***Index Terms*—** Direction-of-arrival estimation, overcomplete representation, support recovery, Chernoff bound, Fano's inequality

1. INTRODUCTION

Recent reformulation of source localization by overcomplete signal representation demonstrates the supper-resolution power of sparse signal processing algorithms. In [1], the authors transform the process of source localization using sensor arrays into estimating the spectrum of a sparse signal by discretizing the parameter manifold. This sparse signal reconstruction based method exhibits super-resolution in DOA estimation compared with traditional techniques such as beamforming, Capon, and MUSIC [2]. Since the basic model employed in [1] applies to several other important problems in signal processing (see [3] and references therein), the principle of overcomplete signal representation by discretization is readily applicable to those cases.

We reformulate the basic problem of this sparse signal reconstruction based method as one of support recovery with joint sparsity constraints. The DOA estimation for sensor arrays aims at estimating the parameter $\tilde{\theta}$ in the following model:

$$\boldsymbol{y}(t) = A(\tilde{\theta})\tilde{\boldsymbol{x}}(t) + \boldsymbol{w}(t), t = 1, 2, \ldots, T, \tag{1}$$

where $\boldsymbol{y}(t), \tilde{\boldsymbol{x}}(t), \boldsymbol{w}(t)$ are sensor measurements, source signal amplitudes and additive noises, respectively. The steering matrix $A(\tilde{\theta}) = [\ a(\tilde{\theta}_1)\cdots a(\tilde{\theta}_K)\ ]$, where $\tilde{\theta} = [\tilde{\theta}_1 \cdots \tilde{\theta}_K]'$ is the vector of source DOAs, and $a(\cdot)$ is the transfer vector function [3]. To estimate $\tilde{\theta}$, D. Malioutov *et al.* [1] discretize the parameter manifold and assume that the true DOAs $\{\tilde{\theta}_1, \ldots, \tilde{\theta}_K\}$ are among the grid points $\{\theta_1, \ldots, \theta_N\}$. Then model (1) can be rewritten as

$$\boldsymbol{y}(t) = A(\theta)\boldsymbol{x}(t) + \boldsymbol{w}(t), t = 1, 2, \ldots, T, \tag{2}$$

where $\theta = [\theta_1 \cdots \theta_N]'$, $\boldsymbol{x}_n(t) = \tilde{\boldsymbol{x}}_k(t)$ when $\tilde{\theta_k} = \theta_n$ and zero otherwise. Then estimation of DOAs is equivalent to the recovery of the common support for $\{\boldsymbol{x}(t)\}_{t=1}^T$. Therefore, model (2)'s ability of exact support recovery is key to the effectiveness of the method. However, the performance measures on sparse signal recovery of most existing work, e.g. [4], [5], are concerned with bounds on various norms of the difference between the true signals and their estimates for a particular recovery algorithm. The performance of traditional DOA estimation algorithms based on Maximum Likelihood Estimation is usually measured by the Cramér-Rao bound [3], which is also not fitting here because of the discretization procedure. With joint sparsity constraints, a natural measure of performance would be model (2)'s potential for correctly identifying the true common support, and an algorithm's ability to achieve this potential. Therefore, we propose using the probability of error as a performance measure for support recovery and derive bounds that reveal the intrinsic ability of system (2) in signal recovery.

The paper is organized as follows. In Section II, we introduce the mathematical model and relevant assumptions. Section III is devoted to derivation of upper bounds, its application to ULA, and analysis of noise effect. In Section IV, an information theoretic lower bound is given by using Fano's inequality, and a necessary condition is derived guaranteeing a varnishing probability of error. The paper is concluded in Section V.

This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, ONR Grants N000140810849 and N000140910496.

2. MODELS AND ASSUMPTIONS

We introduce the mathematical model studied in this paper. Suppose $\{\boldsymbol{x}(t) \in \mathbb{C}^N\}_{t=1}^T$ are jointly sparse signals with a common support S; that is, only a few components of $\boldsymbol{x}(t)$ are non-zero and the index sets corresponding to these non-zero components are the same for all $t = 1, \ldots, T$. The common support $S = \text{supp}(\boldsymbol{x}(t))$ has known size K.

For measurement model (2), we assume that the vectors $\{\boldsymbol{x}^S(t)\}_{t=1}^T$, formed by the non-zero components of $\boldsymbol{x}(t)$, follow *i.i.d.* complex Gaussian distribution $\mathbb{C}\mathcal{N}(0, I_K)$. Suppose we have M sensors and the columns of steering matrix $A(\theta) \in \mathbb{C}^{M\times N}$ are specified by samples of a transfer vector function $a(\cdot)$. The additive noises $\boldsymbol{w}(t) \in \mathbb{C}^M$ follow *i.i.d.* $\mathbb{C}\mathcal{N}\left(0, \sigma^2 \mathrm{I}_M\right)$. Note that assuming unit variance for signals loses no generality since only the ratio of signal variance to noise variance appears in all subsequent analysis.

We start our analysis for general transfer vector function $a(\cdot)$. We assume through out the paper that every $M \times M$ submatrix of $A(\theta)$ is non-singular, a property referred as non-degenerateness. We henceforth consider a set of fixed grid points and denote $A(\theta)$ by A for simplicity.

Clearly the support S of $\boldsymbol{x}(t)$ reveals the DOAs for source signals. Model (2)'s performance in DOA estimation is determined by the maximal achievable ability in recovering the common support S from the noisy sensor measurements. Therefore, we consider two hypothesis testing problems and analyze the associated probability of error. The first one is a binary support recovery problem of deciding between two different supports. The second problem is a multiple support recovery problem that we choose one among all $\binom{N}{K}$ possible candidate supports.

3. UPPER BOUND ON PROBABILITY OF ERROR

In this section, we first review the derivation of the Chernoff bound [6] on the probability of error for our support recovery problems. Detailed proofs for theorems can be found in [7,8]. We then analyze the effect of noise and derive bound exponent for ULA with one source.

3.1. Upper bound for general steering matrix

Under model (2) and the assumptions, the observation $\boldsymbol{Y} = \left[\boldsymbol{y}(1) \quad \cdots \quad \boldsymbol{y}(T)\right]$ follows a complex matrix variate Gaussian distribution $\boldsymbol{Y}|S \sim \mathbb{C}\mathcal{N}_{M,T}(0, \Sigma_S \otimes \mathrm{I}_T)$ with $\Sigma_S = A_S A_S^\dagger + \sigma^2 \mathrm{I}_M$, when the true support is S [9]. Here A_S denotes the submatrix of A formed by its columns indicated by S and I_M is the identity matrix of dimension M.

The binary support recovery problem of deciding the true support between two candidate supports S_i and S_j is equivalent to a linear Gaussian binary hypothesis testing problem with equal prior probability:

$$\begin{cases} \mathrm{H}_0 : \boldsymbol{Y} \sim \mathbb{C}\mathcal{N}_{M,T}(0, \Sigma_{S_i} \otimes \mathrm{I}_T) \\ \mathrm{H}_1 : \boldsymbol{Y} \sim \mathbb{C}\mathcal{N}_{M,T}(0, \Sigma_{S_j} \otimes \mathrm{I}_T) \end{cases}. \tag{3}$$

For notation convenience, we henceforth denote Σ_{S_i} by Σ_i.

The optimal decision rule with minimal probability of error for (3) is given by the likelihood ratio test [6]. The application of Chernoff bound yields an upper bound on the probability of error P_{err}:

$$P_{\text{err}} \leq \frac{1}{2} e^{\mu(\frac{1}{2})} = \frac{1}{2} \exp\left[-T \log \left| \frac{H_{i,j}^{1/2} + H_{i,j}^{-1/2}}{2} \right| \right] \tag{4}$$

where $\mu(\cdot)$ is the log-moment-generating function of the likelihood ratio, and $H_{i,j} = \boldsymbol{\Sigma}_i^{1/2} \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\Sigma}_i^{1/2}$.

Note that the tightest bound for P_{err} is obtained by minimizing $\mu(s)$ *w.r.t.* s. However, due to the difficulty of finding the minimizer, we have taken $s = \frac{1}{2}$ in the Chernoff bound (4), which turns out to be a good bound in most cases.

The numbers of eigenvalues of $H_{i,j}$ that are equal to 1, less than 1 and greater than 1, as well as the values of the later, play a significant role. Thus, we present the following two propositions on the eigenvalue structure of $H_{i,j}$ [7]:

Proposition 1 *Suppose* $k_{\mathrm{i}} = |S_i \cap S_j|, k_0 = |S_i \backslash S_j|, k_1 = |S_j \backslash S_i|$ *and* $M \geqslant k_0 + k_1$, *where we use* $|\cdot|$ *to denote the size of a finite set and* S_0/S_1 *is the set of elements that are in* S_0 *but not in* S_1*; then* k_0 *eigenvalues of* $H_{i,j}$ *are greater than* 1, k_1 *less than* 1, *and* $M - (k_0 + k_1)$ *equal to* 1.

Proposition 2 *The sorted eigenvalues of* $H_{i,j}$ *that are greater than one are lower bounded by the corresponding eigenvalues of* $\mathrm{I}_{k_0} + RR^\dagger/\sigma^2$, *where* R *is the* $k_0 \times k_0$ *submatrix at the lower-right corner of the upper triangle matrix in the QR decomposition of* $\left[A_{S_j \backslash S_i} \quad A_{S_j S_i} \quad A_{S_i \backslash S_j}\right]$*; and upper bounded by the corresponding eigenvalues of* $\mathrm{I}_{k_0} + A_{S_i \backslash S_j} A_{S_i \backslash S_j}^\dagger / \sigma^2$.

For binary support recovery with $|S_i| = |S_j| = K$ and $k_0 = k_1 \triangleq k_{\mathrm{d}}$, we have:

Proposition 3 *The probability of error for the binary support recovery problem is bounded by*

$$P_{\text{err}} \quad \leq \quad \frac{1}{2} \left[\bar{\lambda}_{S_i,S_j} \bar{\lambda}_{S_j,S_i}/16\right]^{-k_{\mathrm{d}}T/2}, \tag{5}$$

where $\bar{\lambda}_{S_i,S_j}$ *is the geometric mean of the eigenvalues of* $H_{i,j} = \Sigma_i^{1/2} \Sigma_j^{-1} \Sigma_i^{1/2}$ *that are greater than one.*

Now we turn to analyze the probability of error for the multiple support recovery problem. Under the equal prior probability assumption, an application of the union bound gives an upper bound on the probability of error associated with the optimal MAP estimator [6]. For its proof, please refer to [7].

Theorem 1 *If* $\bar{\lambda} \triangleq \min_{i \neq j}\{\bar{\lambda}_{S_i,S_j}\} > 4\left[K(N-K)\right]^{\frac{1}{T}}$, *then the probability of error for the full support recovery problem is bounded by*

$$P_{\text{err}} \leq \frac{1}{2} \frac{K(N-K)\left(\bar{\lambda}/4\right)^{-T}}{1 - K(N-K)\left(\bar{\lambda}/4\right)^{-T}}. \tag{6}$$

We note that increasing the number of temporal samples can force the probability of error to 0 exponentially fast as long as $\bar{\lambda}$ exceeds the threshold $4[K(N-K)]^{\frac{1}{T}}$. The final bound (6) is of the same order as the probability of error when $k_{\rm d} = 1$. The probability of error $P_{\rm err}$ is dominated by the probability of error in cases for which the estimated support differs by only one index from the true support, which are the most difficult cases for the decision rule to make a choice.

3.2. Effect of noise

We now present a corollary on the effect of noise:

Corollary 1 *For binary and multiple support recovery problems with support size K, suppose $M \geq 2K$; then there exists constant $c > 0$ that depends only on the steering matrix A such that*

$$1 + \frac{c}{\sigma^2} \leq \bar{\lambda} \leq 1 + \frac{M}{\sigma^2}. \quad (7)$$

This corollary implies from (5) *and* (6) *that*

$$\lim_{\sigma^2 \to 0} P_{\rm err} = 0 \quad (8)$$

and the speed of convergence is approximately $(\sigma^2)^{k_{\rm d}T}$ and $(\sigma^2)^T$ for the binary and multiple cases, respectively.

Proof: According to Proposition 2, for any fixed S_i, S_j, the eigenvalues of $H_{i,j} = \Sigma_i^{1/2}\Sigma_j^{-1}\Sigma_i^{1/2}$ that are greater than 1 are lower bounded by those of $\mathrm{I}_{k_{\rm d}} + RR^\dagger/\sigma^2$; hence we have inequality

$$\bar{\lambda}_{S_i,S_j} \geq \left|\mathrm{I}_{k_{\rm d}} + RR^\dagger/\sigma^2\right|^{1/k_{\rm d}} \geq 1 + \frac{1}{\sigma^2}\left(\prod_{l=1}^{k_{\rm d}} r_{ll}^2\right)^{1/k_d}, \quad (9)$$

where r_{ll} is the lth diagonal element of R. Since A is non-degenerate, $\begin{bmatrix} A_{S_j \backslash S_i} & A_{S_j S_i} & A_{S_i \backslash S_j} \end{bmatrix}$ is of full rank and $r_{ll}^2 > 0, 0 \leq l \leq k_{\rm d}$ for all S_i, S_j. Defining c as the minimal value of $\left(\prod_{l=1}^{k_{\rm d}} r_{ll}^2\right)^{1/k_{\rm d}}$'s over all possible support pairs S_i, S_j, we then have $c_1 > 0$ and

$$\bar{\lambda} \geq 1 + \frac{c}{\sigma^2}. \quad (10)$$

On the other hand, the upper bound on the eigenvalues of $H_{i,j}$ in Proposition 2 yields

$$\bar{\lambda} \leq 1 + \frac{M}{\sigma^2}. \quad (11)$$

All other statements in the theorem follows immediately from (5) and (6). ■

Corollary 1 suggests that in the limiting case that there is no noise, $M \geq 2K$ is sufficient to recover a $K-$sparse signal. Our results also shows that the optimal decision rule, which is unfortunately inefficient in computation, is robust to noise.

3.3. Uniform linear array

In this subsection, we bound $\bar{\lambda}$ for ULAs with only one source. It is hard to obtain the exact expression of $\bar{\lambda}$ or even reasonable approximations for general K.

Suppose that M sensors are located uniformly along the x-axis with the x-coordinate for the mth sensor $(m - \frac{M+1}{2})d$, where d is the distance between adjacent sensors. Then the transfer vector function is

$$a(\phi) = e^{-i\omega_c(m-\frac{M+1}{2})d\cos\phi}, \quad (12)$$

where ω_c is the carrier frequency for the narrow band source signals.

In the case of only one source, the QR decomposition in Proposition 2 becomes trivial and we obtain for $S_i = \{i\}$ and $S_j = \{j\}$:

$$R^2 = M\left[1 - \left(a(\theta_i)^\dagger a(\theta_j)/M\right)^2\right]. \quad (13)$$

Therefore, further algebraic manipulation yields

$$\bar{\lambda} \geq 1 + \frac{1}{\sigma^2}M\left[1 - \max_{i\neq j}\frac{\sin^2\left(M\omega_c d\Delta_{ij}/2\right)}{M^2\sin^2(\omega_c d\Delta_{ij}/2)}\right] \quad (14)$$

with $\Delta_{ij} = \cos(\theta_i) - \cos(\theta_j)$.

This lower bound on $\bar{\lambda}$ has several implications. Since Δ_{ij} can get close to 2 when the range of DOA is $(0, \pi)$, in order to avoid ambiguity we impose the constraint $\omega_c d \leq \pi$, which is equivalent to that d is less than half wavelength. We also note that the maximization in (14) is achieved by adjacent grid points. To make the lower bound as large as possible, we should choose grid points $\{\theta_n\}_{n=1}^N$ such that Δ_{ij} is constant for adjacent grid points. Hence, it is a better strategy to discretize uniformly the wave number space than the DOA space.

4. AN INFORMATION THEORETICAL LOWER BOUND ON PROBABILITY OF ERROR

In this section, we derive an information theoretical lower bound on the probability of error for *any* decision rule in the multiple support recovery problem. Suppose that we have a random vector $\boldsymbol{Y}$ with L possible densities $f_0, \ldots, f_{L-1}$. Then by Fano's inequality [10], [11], the probability of error for *any* decision rule to identify the true density is lower bounded by

$$P_{\rm err} \geq 1 - (\beta + \log 2)/\log L, \quad (15)$$

where $\beta = \frac{1}{L^2}\sum_{i,j} D_{KL}(f_i||f_j)$ is the average of the Kullback-Leibler divergence between all pairs of densities.

Since in the multiple support recovery problem, all the distributions involved are matrix variate Gaussian distributions with mean 0 and different variance, a direct computation gives the expression of $\beta = \frac{T}{2L^2}\sum_{i,j}\left[\mathrm{tr}(H_{i,j}) - M\right]$, where $H_{i,j} = \Sigma_i^{1/2}\Sigma_j^{-1}\Sigma_i^{1/2}$.

Invoking Proposition 1 and the second part of Proposition 2, we derive an upper bound on the average Kullback-Leibler divergence

$$\beta \leq \frac{T}{2L^2\sigma^2}\sum_{i=0}^{L-1}\sum_{k_{\rm d}=1}^{K}\binom{K}{K-k_{\rm d}}\binom{N-K}{k_{\rm d}} \times \sum_{n\in S_i\backslash S_j}\|a(\theta_n)\|^2. \quad (16)$$

Under the narrowband signal assumption and narrowband array assumption, the transfer vector function for isotropic sensor arrays always has norm $\sqrt{M}$ [2]. Hence, we obtain

$$\beta \leq MTK(1 - K/N)/(2\sigma^2). \tag{17}$$

We conclude with the following theorem:

Theorem 2 *Under the narrowband signal assumption and narrowband array assumption, suppose we have an isotropic sensor array and $M \geq 2K$, then for the multiple support recovery problem, the probability of error for* ***any*** *decision rule is lower bounded by*

$$P_{\text{err}} \geq 1 - \frac{MTK\,(1 - K/N)}{2\sigma^2 \log \binom{N}{K}} + o\,(1)\,. \tag{18}$$

An immediate consequence is a necessary condition on the number samples for a vanishing probability of error:

Corollary 2 *With the same setup as in Theorem 2, in order to let the probability of error $P_{\text{err}} < \varepsilon$ with $\varepsilon > 0$ for* ***any*** *decision rule, the number of measurements must satisfy the following:*

$$MT \geq (1 - \varepsilon)\, 2\sigma^2 \log(N/K) + o(1). \tag{19}$$

Note that Theorem 2 is most strong when the number of sources is small, which is usually the case in the practice of radar and sonar. Our result shows that the number of samples is lower bounded by $\log N$ in this case. Note that N is the number of intervals we use to divide the whole range of DOA; hence, it is a measure of resolution. Therefore, the number of samples needs to increase only in the logarithm of N, which is very desirable. The symmetric roles played by M and T are also desirable since M is the number of sensors and is expensive to increase. As a consequence, we simply increase the number of samples to achieve a desired probability of error as long as M is greater than $2K$.

5. CONCLUSION

We analyzed the performance for overcomplete signal representation based DOA estimation method. We reformulated the DOA estimation problem as one of support recovery problem and analyzed it in a hypothesis testing framework. We derived bounds on the probability of error for the support recovery problem. We applied these bounds to study the effect of noise, to find the best discretization method, and to determine minimal number of samples needed for a pre-specified recovery accuracy. We found that discretizing the wave number space is a better strategy than discretizing the DOA space. The minimal number of samples necessary for accurate support recovery is proportional to the logarithm of the number of partitions in the parameter space.

6. REFERENCES

[1] D. 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.

[2] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[3] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramér-Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.

[4] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[5] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[6] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley-Interscience, New York, 2001.

[7] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," to appear on *IEEE Trans. Inf. Theory*.

[8] G. Tang and A. Nehorai, "Performance analysis for support recovery with joint sparsity constraints," in *Proc. 47th Allerton Conf. Commun. Control, and Comput.*, Monticello, IL, Sept. 2009.

[9] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Chapman & Hall/CRC, Boca Raton, FL, 1999.

[10] Y. G. Yatracos, "A lower bound on the error in nonparametric regression type problems," *Ann. Statist.*, vol. 16, no. 3, pp. 1180–1187, 1988.

[11] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," in *IEEE Int. Symp. Information Theory (ISIT 2007)*, Nice, France, June 2007, pp. 961–965.