AN ENTROPY BASED APPROACH FOR DIRECTION-OF-ARRIVAL ESTIMATION IN GAUSSIAN AND NON-GAUSSIAN CORRELATED NOISE

Yifeng Zhou[1], *Henry Leung*[2], *Patrick C. Yip*[3] *and Q. Jin*[4]

[1]Telexis Corporation, 2427 Holly Lane, Ottawa, Ontario, Canada K1V 7P2
[2]Surface Radar Section, Defence Research Establishment Ottawa, Ottawa, Ontario, Canada K1A OK2
[3]Communications Research Lab, McMaster University, Hamilton, Ontario, Canada L8S 4K1
[4]MITEL Corporation, 350 Legget Dr., P.O. Box 13089, Kanata, Ontario, Canada K2K 1X3

ABSTRACT

In this paper, we present an entropy based approach for DOA estimation in Gaussian and non-Gaussian environments. The DOA estimates are obtained by minimizing an entropy measure of the array data in the noise subspace. We show that the entropy approach leads to the MAP algorithm under the Gaussian assumption. Under the non-Gaussian assumption, we apply the varimax norm as an information measure. An intuitive consistency analysis is also performed. Computer simulations are used to demonstrate the effectiveness of the proposed approach.

1. INTRODUCTIONS

For many sensor systems used in radar, sonar and communications systems, one of the most important problems is to estimate the direction-of-arrival (DOA) of source signals. Many techniques have been developed in recently years including the subspace based algorithms [1] and the nonlinear optimization based approaches [2]. They rely on the assumption that the noise covariance be exactly known. However, this assumption is not true and has always been violated in practice.

Several approaches have been proposed to overcome the difficulty of the unknown noise covariance problem. The MDL (Minimum Description Length) method proposed by Wax [3] is based on the Rissanen's MDL principle [4] and can solve the detection and estimation problem simultaneously. The MAP (Maximum A Posteriori) algorithm [5] by Wong *et al.* is a direct application of the Bayesian inference theory. In the MAP algorithm, a noninformative *a priori* probability density function is assigned to the unknown noise covariance and an integration process is used to represent the ignorance of the unknown noise environment. Geometrically, the MDL method minimizes the volume occupied by the array data in both the signal and the noise subspace while the MAP algorithm only minimizes the projected volume of the array data in the noise subspace. These approaches require no parametric modeling of the noise and thus show robust performance. However, since these techniques are developed exclusively under the Gaussian assumption, their performances deteriorate even when the underlying signal and noise distribution deviates slightly from Gaussian.

In this paper, we present an entropy based DOA estimation approach for Gaussian and non-Gaussian correlated noise environments. The estimation problem is considered from an informational viewpoint. We formulate the entropy approach as the one that minimizes the entropy of the array data in the noise subspace. For Gaussian processes, we use the Shannon entropy and show that the entropy approach leads to the MAP method. We use the varimax norm as an entropy measure for the more general case of non-Gaussian processes and carry out the consistency analysis. The entropy based approach makes no specific assumptions about the distribution of the array data and will have wider applications in practice. Finally, computer simulations are provided to demonstrate the effectiveness of the proposed approach.

2. ARRAY SIGNAL MODEL

Consider an array of M omni-directional sensors. Assume that there are $K(K < M)$ source signals in the far-field of the array. Using analytic representation, we can write the array model as

$$\underline{x}(t) = A(\Theta_0)\underline{s}(t) + \underline{n}(t), \tag{1}$$

where

$$\begin{aligned} \underline{s}(t) &= [s_1(t), s_2(t), \ldots, s_K(t)]^T \\ \underline{x}(t) &= [x_1(t), x_2(t), \ldots, x_M(t)]^T \\ \underline{n}(t) &= [n_1(t), n_2(t), \ldots, n_M(t)]^T, \end{aligned} \tag{2}$$

are the array signal, data and noise vector, respectively, and $A(\Theta_0)$ is defined as the array composite steering matrix determined by the array geometry and the DOA parameters. We assume that $A(\Theta_0)$ has full rank and the array ensures unique estimation solution,

$$A(\Theta_1)T_1 = A(\Theta_2)T_2 \Longrightarrow \Theta_1 = \Theta_2, \tag{3}$$

where T_1 and T_2 are any pair of full rank matrices. Assume that $\{\underline{s}(t)\}$ and $\{\underline{n}(t)\}$ are both *i.i.d.* processes with zero mean and covariance matrices R_s and R_n, respectively. We also assume that the $\{\underline{s}(t)\}$ and $\{\underline{n}(t)\}$ are statistically independent.

3. ENTROPY APPROACH FORMULATION

Define the signal subspace $\mathcal{H}_s$ as the K dimensional subspace spanned by the columns of $A(\Theta)$, where Θ is the assumed DOA parameter, and the noise subspace $\mathcal{H}_n$ as the orthogonal complementary subspace of $\mathcal{H}_s$. We decompose the array data into its components in the noise subspace. Let $U(\Theta)$ denote a set of orthonormal basis of the noise subspace, assuming that Θ is given. The noise subspace components are given by the projection of the array data onto the noise subspace as

$$\begin{aligned} \underline{y}(t) &= U^H(\Theta)\underline{x}(t) \\ &= = U^H(\Theta)A(\Theta_0)\underline{s}(t) + U^H(\Theta)\underline{n}(t). \end{aligned} \quad (4)$$

We formulate the entropy approach for DOA estimation as the one that minimizes the entropy of the array data in the noise subspace

$$\hat{\Theta} = \arg\min_{\Theta} h[\underline{y}(t)], \quad (5)$$

where $h[\cdot]$ denotes an entropy measure.

3.1. Gaussian assumption

Assume that the signal and the noise processes follow the Gaussian distribution. It follows that the noise subspace component $\underline{y}(t)$ is also Gaussian distributed. We use the Shannon entropy as an information measure. The Shannon entropy of a random variable x with a probability density function $f(x)$ is defined by

$$h_s(x) = -\int_S f(x)\log f(x)dx, \quad (6)$$

where S is the support set of x. For a multivariate Gaussian process, the entropy is proportional to the determinant of its covariance [6]. Then, criterion (5) can be written as

$$\hat{\Theta} = \arg\min_{\Theta} det\{U^H(\Theta)R_x U(\Theta)\}. \quad (7)$$

Criterion (7) can be interpreted as minimizing the geometric volume of the array data covariance in the noise subspace. In practical applications, since the exact array data covariance is usually not available, we replace R_x by its maximum likelihood estimate

$$\hat{R}_x = \frac{1}{N}\sum_{t=1}^{N} \underline{x}(t)\underline{x}^H(t), \quad (8)$$

where N is the number of array data. The criterion of the entropy approach under the Gaussian assumption becomes

$$\hat{\Theta} = \arg\min_{\Theta} det\{U^H(\Theta)\hat{R}_x U(\Theta)\}. \quad (9)$$

which is identical to the MAP algorithm [3] based on the Bayesian inference theory.

Since, by the ergodic theorem, the distribution of the array data tends to its theoretical distribution as N increases and the maximum likelihood estimate $\hat{R}_x$ converges to R_x asymptotically with probability one [7], the entropy approach criterion (9) approaches its theoretical version (7). Consider the covariance of the array data in the noise subspace

$$R_{xn} = R_{sn} + R_{nn}, \quad (10)$$

where $R_{xn} = U^H(\Theta)R_x U(\Theta)$, $R_{nn} = U^H(\Theta)R_n U(\Theta)$ and $R_{sn} = U^H(\Theta)A(\Theta_0)R_s A(\Theta_0)U(\Theta)$. Taking the determinant of (10), we have the following inequality as

$$det(R_{xn})^{\frac{1}{M-K}} \geq det(R_{sn})^{\frac{1}{M-K}} + det(R_{nn})^{\frac{1}{M-K}}. \quad (11)$$

where the equality holds only when either R_{ns} or R_{un} is a zero matrix, or $R_{sn} = R_{nn}$. Since R_{nn} is assumed to be unknown, we use a noninformative *a priori* probability density function so that it reflects our ignorance of the noise environments [12]. The general rule of obtaining the noninformative *a priori* probability density function of a set of parameters is to take the one that is proportional to the square root of the determinant of the information matrix. In [3], the noninformative *a priori* probability density function of R_{nn} under the Gaussian assumption has been derived as

$$p(R_{un}^{-1} \mid \Theta) \propto \{det(R_{un}^{-1})\}^{-(M-K)}. \quad (12)$$

To ignore the effects of R_{nn}, we multiply both sides of (11) by $p(R_{nn}^{-1})$ and integrate them over R_{nn}^{-1}

$$det(R_{xn})^{\frac{1}{M-K}} \geq det(R_{sn})^{\frac{1}{M-K}} + \text{const.}, \quad (13)$$

where we have used

$$\int_{-\infty}^{\infty} p(R_{nn}^{-1})dR_{nn}^{-1} = 1. \quad (14)$$

Inequality (13) implies that $det(R_{ux})$ attains minimum if and only if R_{sn} is zero. This is possible only when $U^H(\Theta)A(\Theta_0) = 0$. Then, it follows from the uniqueness condition (3) that the source DOA parameters estimated by minimizing the determinant of array covariance in the assumed noise subspace converge to Θ_0 as N increases.

We examine the case when the sensor noise is known to be spatially white with a common variance, i.e., $R_n = \sigma_n^2 I$, where I denotes the identity matrix and σ_n^2 is the common noise variance. Assuming that R_{sn} is positive, we have

$$\{det(R_{ux})\}^{1/(M-K)} = \min_{G} \frac{1}{M-K} tr\{R_{ux}G\}, \quad (15)$$

where G is restrained to be positive and $det(G) = 1$. Using the basic inequality relationships, we can obtain

$$\min_{\Theta} det\{U^H(\Theta)\hat{R}_x U(\Theta)\} = \min_{\Theta} \frac{1}{M-K} tr\{P_{A(\Theta)}^{\perp} R_x\}, \quad (16)$$

where $P_{A(\Theta}^{\perp}$ is the orthogonal projector onto the null space of $A^H(\Theta)$. When we replace R_x with empirical estimate $\hat{R}_x$, it becomes clear that the entropy approach reduces to the deterministic maximum likelihood (DML) method introduced by Ziskind and Wax [8].

3.2. Non-Gaussian assumption

Assume that the signal and the noise process are both non-Gaussian distributed. It is known that a Gaussian process is characterized by its second-order statistics while non-Gaussian processes contain valuable statistical information in their higher order moments. For a general class of non-Gaussian processes, it is difficult to calculate the entropies when their exact distributions are not provided.

The varimax norm has been defined by data analysts [9][10] in trying to find a simple representation of set of orthogonal vectors. The varimax norm measures the simplicity of a signal. Maximizing the varimax norm has the effects of simplifying the appearance or the entropy of a signal. The minimum entropy deconvolution (MED) [11] by Wiggins is one successful application of the varimax norm to the blind deconvolution problem. The computation of the varimax norm approach depends only on the empirical distribution of the data and requires no statistical assumptions. The varimax norm of a sequence $\{x(t), t = 1, 2, \ldots, N\}$ is defined as

$$V(x) = \frac{\sum_{t=1}^{N} \mid x(t) \mid^4}{\{\sum_{t=1}^{N} \mid x(t) \mid^2\}^2}. \tag{17}$$

When using the varimax norm as an information measure, the entropy approach criterion (5) becomes

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{m=1}^{M-K} V(y_m(t)), \tag{18}$$

where $y_m(t)$ denotes the m^{th} component of $\underline{y}(t)$ and

$$V(y_m(t)) = \frac{\sum_{t=1}^{N} \mid y_m(t) \mid^4}{\{\sum_{t=1}^{N} \mid y_m(t) \mid^2\}^2}, \tag{19}$$

denotes the varimax norm of $y_m(t)$ in the temporal domain. The entropy approach can be interpreted as the one that minimizes the simplicity of the appearance, or the entropy, of the noise subspace components of the array data.

The consistency of the entropy approach is difficult to be rigorously proven. However, we will provide some useful intuitive justifications as follows. Since the varimax norm $V(\underline{y})$ is related to the sampled Kurtosis of $\{y_m(t)\}$ and, as N increases, approaches its theoretical Kurtosis by the ergodic theorem, we can study the theoretical criterion

$$V_t(\underline{y}) = \sum_{m=1}^{M-K} V_k[y_m(t)] = \sum_{m=1}^{M-K} \frac{C_4[y_m(t)]}{C_2^2[y_m(t)]}, \tag{20}$$

instead, where $V_k[\cdot]$ denotes the Kurtosis and

$$C_i[y_m(t)] = E\{\mid y_m(t) \mid^i\}, \;\; i = 2, 4. \tag{21}$$

Let $\mathcal{X}$ be a set of random variables with finite variances which is closed under linear combinations, that is, $\sum a_i X_i + c_i$ is in $\mathcal{X}$ when a_i and c_i are constants. We use the concepts of equivalency and partial order introduced by Donoho [12]. Two random variables X and Y are regarded as equivalent, written $X \doteq Y$, if for some constants c and $a \neq 0$, $aX + c$ has the same distribution as Y. The partial order $X \cdot \geq Y$ means that for constants $\{a_i\}$ with $\sum a_i^2 < \infty$,

$$Y \doteq \sum a_i X_i, \tag{22}$$

where X_i are independent copies of X. For X in $\mathcal{X}$ and Z Gaussian,

$$X \cdot \geq \sum a_i X_i \cdot \geq Z; \tag{23}$$

this order is strict unless either (a) $\mathcal{X}$ is Gaussian or (b) $\mathcal{X}$ is not Gaussian, but the linear combination is trivial (no two a_i's are nonzero) [12].

Relation (11) implies that the linear combinations of independent random variables are more nearly Gaussian than any individual component of the combination. Thus, we can interpret the partial order $X \cdot \geq Y$ as Y being more Gaussian than X. Expressing $y_m(t)$ as

$$y_m(t) = \sum_{i=1}^{K} \alpha_i s_i(t) + \sum_{i=1}^{M} \beta_i n_i(t), \tag{24}$$

where α_i and β_i are determined by $U^H(\Theta)A(\Theta_0)$ and $U(\Theta)$, respectively, we have

$$\sum_{i=1}^{M} \beta_i n_i(t) \cdot \geq y_m(t), \tag{25}$$

where the equality holds only when $\{\alpha_i\}$ are all zeros, and

$$\sum_{i=1}^{M} \alpha_i n_i(t) \cdot > y_m(t), \tag{26}$$

where the inequality is strict since $\{\beta_i\}$ are elements of $U(\Theta)$ which cannot be all zeros. In [12], it has been shown that the Kurtosis agrees with the partial order, i.e., for two random variables X and Y,

$$X \cdot \geq Y \;\text{ implies }\; V_k(X) \geq V_k(Y). \tag{27}$$

It follows that the partial order (25) can be transformed into the following inequality

$$V_k(y_m(t)) \leq V_k(\sum_{i=1}^{M} \beta_i n_i(t)). \tag{28}$$

$V_k(y_m(t))$ reaches its maximum when the equality holds. According to (25), $V_k(y_m(t))$ is maximum only when $\{\alpha_i\}$ are all zeros. When maximizing the Kurtosis for $m = 1, 2, \ldots, M-K$, we can write the necessary condition as

$$U^H(\Theta)A(\Theta_0) = 0, \tag{29}$$

which is valid if and only if $\Theta = \Theta_0$. Thus, the consistency of the entropy approach is readily established.

4. NUMERICAL EXAMPLES

An equispaced linear array of $M = 6$ sensors is simulated with half the source wavelength spacing. Two uncorrelated unit power narrow-band source signals are assumed to be in the far-field of the array at electrical angles $\phi_i = \pi \sin\theta_i = \pm\frac{1}{8}(2\pi/M)$, $i = 1, 2$, to the normal of the array. The source

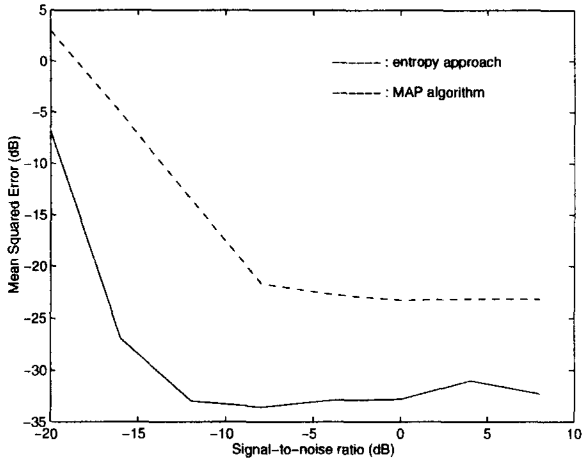

Figure 1. The MSE of the estimates by the entropy approach and the MAP algorithm.

and the noise process are assumed to follow the distribution of a mixture model

$$p[\underline{n}(t)] = (1-\epsilon)\mathcal{N}(R_m, 0) + \epsilon Q, \tag{30}$$

where Q denotes the uniform distribution in $[-1,1]$ and $\mathcal{N}$ represents the Gaussian distribution. R_m is the covariance with its mn^{th} element given by

$$r_{mn} = \rho^{|m-n|} \exp(j\phi_p(m-n)), \tag{31}$$

where $\rho = 0.9$ and $\phi_p = \pi/8$. We choose the mixture coefficient as $\epsilon = 0.6$. Fig. 1 shows the mean squares error (MSE) of the estimated DOA's versus SNR for the MAP algorithm and the entropy approach. For each simulation, 1000 samples are used and each test is repeated 100 times to obtain the average results. It can be seen that the entropy approach outperforms the MAP algorithm when the sensor noise is not Gaussian.

5. CONCLUSIONS

The entropy based DOA estimation approach has been discussed in this paper. This approach is applicable to Gaussian and non-Gaussian correlated noise environments. It has the advantage of making no specific assumptions about the distribution of the array data and will have wider applications in practice. Rigorous analysis of its consistency under the non-Gaussian assumption still needs further investigations.

REFERENCES

[1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation", IEEE Tran. AP, Vol. AP-34, pp. 276-280, 1986

[2] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting", IEEE Trans. ASSP, Vol. ASSP-39, pp 1110-1121, 1991

[3] M. Wax, "Detection and location of multiple sources in noise with unknown covariance", IEEE Trans. SP, Vol. SP-40, pp. 245-249, 1992

[4] J. Rissanen, "Modeling by the shortest data description", *Automatica*, Vol. 14, pp. 465-471, 1978

[5] K. M. Wong, J. P Reilly, Q. Wu and S. Qiao, "Estimation of the direction of arrival of signals in unknown correlated noise, Part I : The MAP approach and its implementation", IEEE Trans. SP, Vol. SP-40, pp. 2007-2017, 1992

[6] T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, Inc., 1991

[7] T. W. Anderson, *Introduction to multivariate statistical analysis*, New York : John Wiley & Sons, Inc., 1984

[8] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection", IEEE Trans. ASSP, Vol. ASSP-37, pp 1190-1196, 1989

[9] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis", Psychometrika, Vol. 23, pp. 187-200, 1958

[10] W. W. Cooley and P. R. Richards, *Multivariate analysis*, Wiley : New York, 1971

[11] R. A. Wiggins, "Minimum entropy deconvolution", Geoexploration, Vol. 16, pp. 21-35, 1978

[12] D. Donoho, "On minimum entropy deconvolution", in *Applied time series analysis II*, Academic Press, pp. 565-608, 1981