MAXIMUM LIKELIHOOD JOINT ANGLE AND DELAY ESTIMATION IN UNKNOWN NOISE FIELDS

Adel Belouchrani

Electrical Engineering Department,
Ecole Nationale Polytechnique,
PO Box 182, El HArrach 16200, Algiers, Algeria

*Saïd Aouada**

School of Electrical and Computer Engineering,
Curtin University of Technology,
GPO Box U1987, Perth WA 6845, Australia

ABSTRACT

We address the problem of Joint Angle and Delay Estimation using a sensor array in an unknown additive noise field. We propose a stochastic Maximum Likelihood (ML) estimator. The algorithm which is a 2-D extension of the Approximate Maximum Likelihood (AML) is applied to the multiple channel sample model and exploits the shift invariance of the data. The model allows the estimation of more parameter pairs than sensors and robustness of the algorithm makes it possible to use blind techniques to estimate the channel. Basic performances of the ML estimator are assessed through simulations and are compared with other high resolution methods. Comparisons are made against the stochastic Cramér-Rao Bound (CRB) which is derived in the Appendix.

1. INTRODUCTION

Parametric Joint Angle and Delay Estimation (JADE) has received considerable attention in the literature [2, 3]. Among the proposed methods, many applied 2-D MUSIC, ESPRIT [2, 3] and Maximum Likelihood (ML) [4]. Only a few of these methods allow for the simultaneous estimation of more paths than sensors. In addition, most available methods are based on the common simplifying assumption of a Gaussian additive noise.

SI-JADE [2] and JADE-ESPRIT [3] are two powerful methods exploiting the shift invariance in the space-time steering matrix and they are based on a model which exploits the stationarity of the channel parameters over long time intervals. The fading however is varying, thus multiple channel samples are collected. The model derived suggests a two step estimation where channel samples are estimated first using a specially designed training sequence. Blind channel estimation techniques are possible, however, they result in an unknown estimation noise distribution which dramatically limits the parameter estimation step [3].

In what follows, we propose an alternative method based on stochastic ML to deal with unspecified noise distributions. As it will be shown, the ML estimator is a 2-D extension of the Approximate Maximum Likelihood (AML) estimator [5] for the JADE model derived in [2]. The main advantage of the algorithm is that it allows the estimation of more parameter pairs than sensors in a blind scheme.

2. DATA MODEL

Consider the case of a single user transmitting a digital signal in a specular multipath environment. P, the number of multipaths is assumed known. The channel is assumed to be fading but stationary over short time intervals. Each path is parameterized by a Direction Of Arrival (DOA) θ_p, a Time Difference Of Arrival (TDOA) τ_p (measured in terms of normalized symbol periods $T = 1$) and a complex fading β_p. The fading coefficients can vary between time slots but not within symbol periods [3]. At the receiver end, an M element calibrated ULA array is used. $\mathbf{a}(\theta_p) = [1, e^{j\theta_p}, \ldots, e^{j\theta_p(M-1)}]^T$ is the functional form of the array response to a path from direction θ_p, with $(.)^T$ denoting matrix transpose. The channel is assumed to be nonzero over the interval $[0, L)$ with $L = L_g + \tau_{\max}$.

*This work was in part supported by the Australian Telecommunications Cooperative Research Centre (AT-CRC).

In matrix form, the received signals corresponding to the k-th time slot can be written as follows

$$\mathbf{X}^{(k)} = \mathbf{H}^{(k)}\mathbf{S}^{(k)} + \mathbf{N}^{(k)}, \quad k = 1, \ldots, K \tag{1}$$

where $\mathbf{X}^{(k)}$ is a $M \times N$ matrix of received samples over interval k, $\mathbf{H}^{(k)}$ is the $M \times L$ matrix of channel samples, $\mathbf{S}^{(k)}$ is a $L \times N$ Toeplitz matrix of unknown data symbols, $\mathbf{N}^{(k)}$ is the matrix of additive noise samples and N is the number of collected snapshots.

After applying the Fourier transform to the channel samples $\mathbf{H}^{(k)}$ and dividing out the known Fourier transform of the samples pulse shape function $\mathbf{g}(\tau_p) = [g(-\tau_p), \ldots, g(L-1-\tau_p)]^T$, where $g(t)$ is the known modulation pulse shaping function, it is straightforward to observe that the channel satisfies the following factorization [2]

$$\mathbf{H}^{(k)} = \mathbf{A}(\boldsymbol{\theta})\text{diag}\{\boldsymbol{\beta}(k)\}\mathbf{F}^T(\boldsymbol{\tau}) \tag{2}$$

where

$$\boldsymbol{\beta}(k) = [\beta_1(k), \ldots, \beta_P(k)] \tag{3}$$

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_P)] \tag{4}$$

$$\mathbf{F}(\boldsymbol{\tau}) = [\mathbf{f}(\tau_1), \ldots, \mathbf{f}(\tau_P)] \tag{5}$$

with

$$\mathbf{f}(\tau_p) = [1, e^{j2\pi\tau_p/L}, \ldots, e^{j2\pi\tau_p(L-1)/L}]^T \tag{6}$$

is the L-dimensional vector of transformed samples of $\mathbf{g}(\tau_p)$, $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_P]^T$ and $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_P]^T$.

Applying the vec$\{.\}$ operator to $\mathbf{H}^{(k)}$, we obtain the following

$$\begin{aligned}\mathbf{y}(k) = \text{vec}\left\{\mathbf{H}^{(k)}\right\} &= [\mathbf{G}(\boldsymbol{\tau}) \circ \mathbf{A}(\boldsymbol{\theta})]\boldsymbol{\beta}(k) \\ &= \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\boldsymbol{\beta}(k)\end{aligned} \tag{7}$$

where $\circ$ stands for Khatri-Rao product, i.e. a column-wise Kronecker product. The $ML \times P$ space-time steering matrix $\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})$ is assumed to be invariant over the observation interval.

3. PARAMETER ESTIMATION

3.1. Blind Channel Estimation

Matrix $\mathbf{H}^{(k)}$ can be obtained from a blind channel estimation [2, 6]. The noisy channel estimates are

$$\tilde{\mathbf{H}}^{(k)} = \mathbf{H}^{(k)} + \mathbf{V}^{(k)}; \tag{8}$$

where $\mathbf{V}^{(k)}$ is the channel estimation noise.

Collecting the data over K time slots leads to the following matrix form

$$\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(K)] = \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{B} + \boldsymbol{\mathcal{V}}. \tag{9}$$

The noise $\boldsymbol{\mathcal{V}}$ and the signal fading $\mathbf{B}$ are assumed to be uncorrelated. The data covariance matrix is therefore given by

$$\mathbf{R} = \mathsf{E}\left\{\mathbf{y}(k)\mathbf{y}^H(k)\right\} = \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{R_B}\mathbf{U}^H(\boldsymbol{\theta}, \boldsymbol{\tau}) + \mathbf{Q} \tag{10}$$

where $\mathbf{Q} = \mathsf{E}\left\{\mathbf{v}(k)\mathbf{v}^H(k)\right\}$ is the unknown covariance matrix of the channel estimation noise, $\mathbf{R_B} = \mathsf{E}\left\{\boldsymbol{\beta}(k)\boldsymbol{\beta}^H(k)\right\}$, $\mathbf{v}(k) = \text{vec}\left\{\mathbf{V}^{(k)}\right\}$ and $(.)^H$ denotes Hermitian transpose.

The received signal waveforms are assumed to be a random zero-mean Gaussian process [7, 8], satisfying the following

$$\mathbf{y}(k) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{11}$$

3.2. Noise Modeling

We consider the $ML \times ML$-dimensional noise covariance matrix $\mathbf{Q}$ as completely unknown. In what follows, we impose a model on the background noise. The noise is modeled by a linear combination of $(ML)^2$ known base matrices and $(ML)^2$ unknown real parameters. Thus, the spatial covariance matrix of the noise is assumed to have the following structure

$$\mathbf{Q} = \sum_{d=1}^{(ML)^2} q_d \boldsymbol{\Phi}_d \tag{12}$$

where $\mathbf{q} = [q_1, q_2, \ldots, q_{(ML)^2}]^T$ is the vector of unknown real noise parameters and the base matrices $\boldsymbol{\Phi}_d$ are known.

Taking into account the Hermitian structure of matrix $\mathbf{Q}$, it is straightforward to identify the unknown parameters $\{q_d\}$ as follows: $d_1 = ML$ diagonal elements, $d_2 = ML(ML-1)/2$ off-diagonal real parts and $d_3 = ML(ML-1)/2$ off-diagonal imaginary parts of $\mathbf{Q}$. Thus, the corresponding base matrices $\boldsymbol{\Phi}_d$ have accordingly entries 1 and $\pm j$ (see [5]).

3.3. Maximum Likelihood Estimation

Under the above assumptions, the joint density of the data is given by [9]

$$f_{\boldsymbol{\eta}}(\mathbf{y}) = (2\pi)^{-\frac{K}{2}} \det\left\{\mathbf{R}^{-\frac{K}{2}}(\boldsymbol{\eta})\right\} \exp\left\{-\frac{1}{2}\sum_{k=1}^{K}\mathbf{y}^H(k)\mathbf{R}^{-1}(\boldsymbol{\eta})\mathbf{y}(k)\right\} =$$
$$= (2\pi)^{-\frac{K}{2}} \det\left\{\mathbf{R}^{-\frac{L}{2}}(\boldsymbol{\eta})\right\} \exp\left\{\frac{-K}{2}\text{trace}\left[\mathbf{R}^{-1}(\boldsymbol{\eta})\hat{\mathbf{R}}\right]\right\} \tag{13}$$

where $\boldsymbol{\eta} = [\boldsymbol{\theta}^T, \boldsymbol{\tau}^T, \boldsymbol{\beta}^T(1), \ldots, \boldsymbol{\beta}^T(K), \mathbf{q}^T]^T$ is the vector of unknown parameters and $\hat{\mathbf{R}}$ is the sample covariance matrix of the data, defined as

$$\hat{\mathbf{R}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{y}(k)\mathbf{y}^H(k) \tag{14}$$

Omitting the constant terms $\frac{K}{2}$ and $\ln(2\pi)$, the negative LL function of the observed data becomes [9, 5]

$$\mathcal{L}(\boldsymbol{\eta}) = \ln\left(\det\left\{\mathbf{R}(\boldsymbol{\eta})\right\}\right) + \text{trace}\left\{\mathbf{R}^{-1}(\boldsymbol{\eta})\hat{\mathbf{R}}\right\} \tag{15}$$

The AML algorithm [5] is based on the observation that

$$\begin{aligned} \text{vec}\{\mathbf{R}\} &= \text{vec}\left\{\mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{R_B}\mathbf{U}^H(\boldsymbol{\theta}, \boldsymbol{\tau})\right\} + \text{vec}\{\mathbf{Q}\} \\ &= [\mathbf{U}^*(\boldsymbol{\theta}, \boldsymbol{\tau}) \otimes \mathbf{U}(\boldsymbol{\theta}, \boldsymbol{\tau})]\,\text{vec}\{\mathbf{R_B}\} + \boldsymbol{\Omega}\mathbf{q} \\ &= \boldsymbol{\mathcal{U}}(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{r_B} + \boldsymbol{\Omega}\mathbf{q} \\ &= \begin{bmatrix} \boldsymbol{\mathcal{U}}(\boldsymbol{\theta}, \boldsymbol{\tau}) & \boldsymbol{\Omega} \end{bmatrix} \begin{bmatrix} \mathbf{r_B} \\ \mathbf{q} \end{bmatrix} \\ &= \mathbf{M}(\boldsymbol{\theta}, \boldsymbol{\tau})\mathbf{z} \end{aligned} \tag{16}$$

where $(.)^*$ stands for complex conjugate, $\otimes$ denotes matrix Kronecker product and $\boldsymbol{\Omega} = \left[\text{vec}\left\{\boldsymbol{\Phi}_1\right\}, \ldots, \text{vec}\left\{\Phi_{(ML)^2}\right\}\right]$.

From (16), a consistent estimate $\hat{\mathbf{z}}$ can be obtained as [5]

$$\hat{\mathbf{z}} = \left[\mathbf{U}^H\left(\hat{\mathbf{R}}^{-T} \otimes \hat{\mathbf{R}}^{-1}\right)\mathbf{U}\right]^{-1}\mathbf{U}^H\left(\hat{\mathbf{R}}^{-T} \otimes \hat{\mathbf{R}}^{-1}\right)\hat{\mathbf{r}} \tag{17}$$

where the dependence of $\mathbf{U}$ on $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ was dropped for convenience. Substituting (17) back into (15) leads to the modified cost function

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \ln\left(\det\left\{\bar{\mathbf{R}}(\boldsymbol{\theta}, \boldsymbol{\tau})\right\}\right) + \text{trace}\left\{\bar{\mathbf{R}}^{-1}(\boldsymbol{\theta}, \boldsymbol{\tau})\hat{\mathbf{R}}\right\} \tag{18}$$

where

$$\bar{\mathbf{R}}(\boldsymbol{\theta}, \boldsymbol{\tau}) = \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\tau}, \hat{\mathbf{z}}(\boldsymbol{\theta}, \boldsymbol{\tau})) \tag{19}$$

Finally, the 2D-AML algorithm reduces to the joint minimization of (18) with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. In order to achieve convergence, the algorithm needs to be well initialized. One way to find an initial estimate is, first to assume that the noise is uniform Gaussian, i.e., $\mathbf{Q} = \sigma^2\mathbf{I}$ and then use any 2-D high resolution estimator to get the initial estimate. Use the initial estimate to start either a simultaneous search [5] over $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$, or through a numerical iterative concentration of the LL function, where one parameter is assumed known to obtain the other over each step [8].

4. CRAMÉR-RAO BOUND

The unconditional CRB corresponding to the assumptions on our model is as follows

$$\mathbf{CRB}_{\alpha\alpha} = \frac{1}{K}\left\{2\text{Re}\left[\left(\mathbf{R_B}\breve{\mathbf{U}}^H\breve{\mathbf{R}}^{-1}\breve{\mathbf{U}}\mathbf{R_B}\right)^T \odot \left(\breve{\mathbf{D}}^H\mathbf{P}_{\breve{\mathbf{U}}}^{\perp}\breve{\mathbf{D}}\right)\right] - \mathfrak{M}\mathfrak{T}\mathfrak{M}^T\right\}^{-1} \tag{20}$$

where $\odot$ stands for Schur-Hadamard product, and

$$\mathfrak{M} = 2\text{Re}\left\{\boldsymbol{\mathcal{Q}}^T\left[\left(\breve{\mathbf{D}}^H\mathbf{P}_{\breve{\mathbf{U}}}^{\perp}\right) \otimes \left(\mathbf{R_B}{}^T\breve{\mathbf{U}}^T\breve{\mathbf{R}}^{-T}\right)\right]\boldsymbol{\mathcal{P}}^*\right\} \tag{21}$$

$$\mathfrak{T} = 2\text{Re}\left\{\boldsymbol{\mathcal{P}}^H\left(\breve{\mathbf{R}}^{-T} \otimes \mathbf{P}_{\breve{\mathbf{U}}}^{\perp}\right)\boldsymbol{\mathcal{P}}\right\} - \boldsymbol{\mathcal{P}}^H\left(\mathbf{P}_{\breve{\mathbf{U}}}^{\perp T} \otimes \mathbf{P}_{\breve{\mathbf{U}}}^{\perp}\right)\boldsymbol{\mathcal{P}} \tag{22}$$

$$\mathbf{D} = \left[\left.\frac{d\mathbf{u}(\alpha)}{d\alpha}\right|_{\alpha=\alpha_1}, \ldots, \left.\frac{d\mathbf{u}(\alpha)}{d\alpha}\right|_{\alpha=\alpha_{2P}}\right] \tag{23}$$

$$\boldsymbol{\mathcal{Q}} = \left[\text{vec}\left\{\mathbf{e}_1\mathbf{e}_1^T\right\}, \ldots, \text{vec}\left\{\mathbf{e}_p\mathbf{e}_p^T\right\}\right] \tag{24}$$

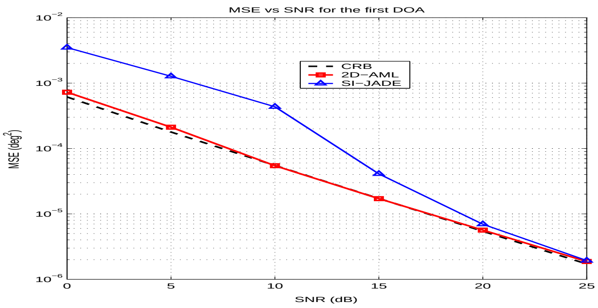
$$\boldsymbol{\mathcal{P}} = \left[\text{vec}\left\{\breve{\boldsymbol{\Phi}}_1\right\}, \ldots, \text{vec}\left\{\breve{\boldsymbol{\Phi}}_{(ML)^2}\right\}\right] \tag{25}$$

$$\breve{\boldsymbol{\Phi}}_d = \mathbf{Q}^{-1/2}\boldsymbol{\Phi}_d \tag{26}$$

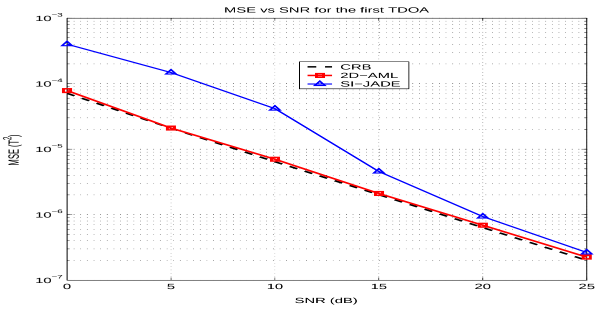
$$\breve{\mathbf{U}} = \mathbf{Q}^{-1/2}\mathbf{U} \tag{27}$$

$$\breve{\mathbf{D}} = \mathbf{Q}^{-1/2}\mathbf{D} \tag{28}$$

$$\breve{\mathbf{R}} = \mathbf{Q}^{-1/2}\mathbf{R}\mathbf{Q}^{-1/2} \tag{29}$$

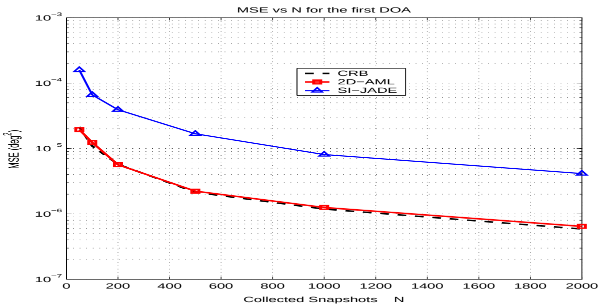


(a) 1^{st} Angle.

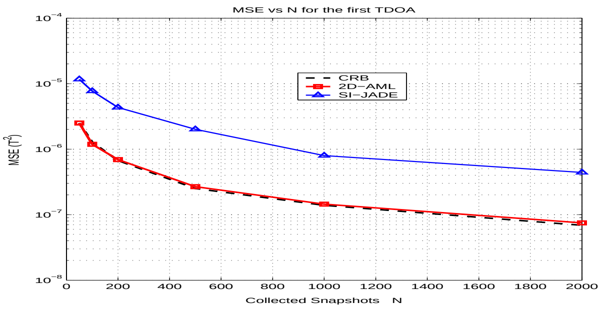


(b) 1^{st} Delay.

Fig. 1. Comparison between 2D-AML and SI-JADE vs SNR.



(a) 1^{st} Angle.



(b) 1^{st} Delay.

Fig. 2. Comparison between 2D-AML and SI-JADE vs N.

with $\mathbf{P}^{\perp}_{\check{\mathbf{U}}}$ being the orthogonal projector onto the column space of $\check{\mathbf{U}}$ and $\mathbf{e}_p$ is a vector whose p-th entry is the only nonzero element as it equals 1. Details relative to the derivation of $\mathbf{CRB}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ are provided in the Appendix.

5. SIMULATIONS

We illustrate the global performance of the 2D-AML estimator for different values of SNR and number of collected snapshots. We compare the 2D-AML algorithm to SI-JADE [2]. In the first simulation $P = 3$ paths are considered with $\boldsymbol{\theta} = [-7°, 0°, 12°]^T$ and $\boldsymbol{\tau} = [3, 0, 8]^T$ and an array of $M = 2$ sensors is used. The channel length is assumed to be $L = 32$ symbol periods. The number of collected snapshots is set to $N = 100$. The channel is estimated over $K = 20$ slots. The results of the simulation are averaged after 500 Monte Carlo runs. In the 2D-AML, an initial estimate is obtained using 2D-MUSIC with $\mathbf{Q} = \mathbf{I}$. SI-JADE employs a joint diagonalization method referred to as the *Q method* [2]. The results are very similar for the three paths, therefore only those corresponding to the first path are shown. As expected, SI-JADE provides erroneous estimates of the parameters as it undergoes severe mis-modeling due to the non white additive noise. On the other hand, 2D-AML approaches the CRB and exploits the favorable data model.For the above settings, convergence of 2D-AML is achieved after a relatively small number of iterations.

In the second simulation, the same settings are applied for different vales of N. The Signal to Noise Ratio (SNR) is fixed at 10dB. SI-JADE again suffers a severe mis-modeling effect as it expects a uniform Gaussian noise whereas 2D-AML approaches the CRB for relatively low N.

6. CONCLUSION

An alternative ML estimation method is presented for the JADE problem. The algorithm is an extension to 2-D of AML and it exploits the shift invariance inherent in the data model. This shift invariance allows the estimation of more paths than sensors. The noise covariance matrix is assumed unknown and its Hermitian structure is exploited. Robustness of the algorithm makes it possible to estimate the channel samples in a blind scheme.

7. APPENDIX

Considering the stochastic case, we define the vector of unknown parameters as follows

$$\boldsymbol{\eta} = [\boldsymbol{\alpha}^T, \mathbf{r_B}^T, \mathbf{q}^T]^T \tag{30}$$

where $\boldsymbol{\alpha} = [\boldsymbol{\theta}^T, \boldsymbol{\tau}^T]^T$, is the vector of the parameters of interest and $\mathbf{r_B} = [\text{Re}(\mathbf{r_B}^T), \text{Im}(\mathbf{r_B}^T)]^T$ is the L^2-dimensional vector defined in (16), of the real parameters of the covariance matrix $\mathbf{R_B}$ of the fading coefficients. The elements of the Fisher Information Matrix (FIM) are given by [7, 8]

$$[\boldsymbol{\mathcal{F}}]_{i,j} = K\text{trace}\left\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_i}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \eta_j}\right\} \tag{31}$$

In the following, we summarize the derivation of all the blocks of the FIM. Note that

$$\frac{\partial \mathbf{R}}{\partial \alpha_i} = \left[\frac{\partial \mathbf{R}}{\partial \theta_i}, \frac{\partial \mathbf{R}}{\partial \tau_i}\right] \tag{32}$$

with

$$\frac{\partial \mathbf{R}}{\partial \theta_i} = \mathbf{D}_\theta \mathbf{e}_i \mathbf{e}_i^T \mathbf{R_B} \mathbf{U}^H + \mathbf{U} \mathbf{R_B} \mathbf{e}_i \mathbf{e}_i^T \mathbf{D}_\theta^H \tag{33}$$

$$\frac{\partial \mathbf{R}}{\partial \tau_i} = \mathbf{D}_\tau \mathbf{e}_i \mathbf{e}_i^T \mathbf{R_B} \mathbf{U}^H + \mathbf{U} \mathbf{R_B} \mathbf{e}_i \mathbf{e}_i^T \mathbf{D}_\tau^H \tag{34}$$

and

$$\mathbf{D}_\theta = \mathbf{F} \circ \left[\frac{d\mathbf{a}(\theta)}{d\theta} \bigg|_{\theta=\theta_1}, \ldots, \frac{d\mathbf{a}(\theta)}{d\theta} \bigg|_{\theta=\theta_P} \right] \quad (35)$$

$$\mathbf{D}_\tau = \left[\frac{d\mathbf{f}(\tau)}{d\tau} \bigg|_{\tau=\tau_1}, \ldots, \frac{d\mathbf{f}(\tau)}{d\tau} \bigg|_{\tau=\tau_P} \right] \circ \mathbf{A} \quad (36)$$

We define $\mathbf{D} = [\mathbf{D}_\theta, \mathbf{D}_\tau]$. Using results of [7, 8, 10] and applying properties of the vec $\{.\}$ and trace $\{.\}$ operators several times, it is straightforward to show that

$$\begin{aligned} \boldsymbol{\mathcal{F}}_{\boldsymbol{\alpha\alpha}} &= K\boldsymbol{\mathcal{E}}^T \left[\mathbf{D}^T \otimes \left(\mathbf{R}_\mathbf{B} \mathbf{U}^H \right) + \left(\mathbf{R}_\mathbf{B}{}^T \mathbf{U}^T \right) \otimes \mathbf{D}^H \right] \\ &\quad \cdot \left[(\mathbf{R}^*)^{-1} \otimes \mathbf{R}^{-1} \right] \\ &\quad \cdot \left[\mathbf{D}^T \otimes \left(\mathbf{R}_\mathbf{B} \mathbf{U}^H \right) + \left(\mathbf{R}_\mathbf{B}{}^T \mathbf{U}^T \right) \otimes \mathbf{D}^H \right]^H \boldsymbol{\mathcal{E}} \quad (37) \end{aligned}$$

Let us introduce the matrix

$$\begin{aligned} [\boldsymbol{\mathcal{F}}_{\mathbf{qq}}]_{i,j} &= K\mathrm{trace}\left\{ \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial q_i} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial q_j} \right\} \\ &= K\mathrm{trace}\left\{ \mathbf{R}^{-1/2} \frac{\partial \mathbf{R}}{\partial q_i} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial q_j} \mathbf{R}^{-1/2} \right\} \\ &= K\mathrm{trace}\left\{ \mathbf{R}^{-1/2} \frac{\partial \mathbf{Q}}{\partial q_i} \mathbf{R}^{-1} \frac{\partial \mathbf{Q}}{\partial q_j} \mathbf{R}^{-1/2} \right\} \\ &= K\mathrm{trace}\left\{ \mathbf{R}^{-1/2} \mathbf{\Phi}_i \mathbf{R}^{-1} \mathbf{\Phi}_j \mathbf{R}^{-1/2} \right\} \quad (38) \end{aligned}$$

Again, using results of [7, 8, 10] and applying properties of the vec $\{.\}$ and trace $\{.\}$ operators, we obtain the following

$$\boldsymbol{\mathcal{F}}_{\mathbf{qq}} = K\boldsymbol{\mathcal{P}}^T \left[(\mathbf{R}^*)^{-1} \otimes \mathbf{R}^{-1} \right] \boldsymbol{\mathcal{P}} \quad (39)$$

where $\boldsymbol{\mathcal{P}}$ was defined in Section 4.

Note that $\mathbf{R}_\mathbf{B}$ has a Hermitian structure and is completely unknown. This suggests that $\mathbf{R}_\mathbf{B}$ can be written as a linear combination of base matrices, in the same way as $\mathbf{Q}$, i.e,

$$\mathbf{R}_\mathbf{B} = \sum_{d=1}^{P^2} \mathbf{r}_{\mathbf{B}_d} \mathbf{\Phi}_d \quad (40)$$

Thus, similarly to the derivation of $[\boldsymbol{\mathcal{F}}_{\mathbf{qq}}]_{i,j}$, we define the following

$$\begin{aligned} [\boldsymbol{\mathcal{F}}_{\mathbf{r_B r_B}}]_{i,j} &= K\mathrm{trace}\left\{ \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \mathbf{r_B}_i} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \mathbf{r_B}_j} \right\} \\ &= K\mathrm{trace}\left\{ \mathbf{R}^{-1/2} \frac{\partial \mathbf{R}}{\partial \mathbf{r_B}_i} \mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial \mathbf{r_B}_j} \mathbf{R}^{-1/2} \right\} \\ &= K\mathrm{trace}\left\{ \breve{\mathbf{U}} \frac{\partial \mathbf{R_B}}{\partial \mathbf{r_B}_i} \breve{\mathbf{U}}^H \breve{\mathbf{U}} \frac{\partial \mathbf{R_B}}{\partial \mathbf{r_B}_j} \breve{\mathbf{U}}^H \right\} \\ &= K\mathrm{vec}^T \left\{ \left[\breve{\mathbf{U}} \frac{\partial \mathbf{R_B}}{\partial \mathbf{r_B}_i} \breve{\mathbf{U}}^H \right]^T \right\} \mathrm{vec}\left\{ \breve{\mathbf{U}} \frac{\partial \mathbf{R_B}}{\partial \mathbf{r_B}_j} \breve{\mathbf{U}}^H \right\} \\ &= K\mathrm{vec}^T \left\{ \left[\breve{\mathbf{U}} \mathbf{\Phi}_i \breve{\mathbf{U}}^H \right]^T \right\} \mathrm{vec}\left\{ \breve{\mathbf{U}} \mathbf{\Phi}_j \breve{\mathbf{U}}^H \right\} \quad (41) \end{aligned}$$

Similarly to (39), we obtain

$$\boldsymbol{\mathcal{F}}_{\mathbf{r_B r_B}} = K\boldsymbol{\mathcal{P}}_P^T \left[\mathbf{U}^T \otimes \mathbf{U}^H \right] \left[(\mathbf{R}^*)^{-1} \otimes \mathbf{R}^{-1} \right] \left[\mathbf{U}^* \otimes \mathbf{U} \right] \boldsymbol{\mathcal{P}}_P \quad (42)$$

where $\boldsymbol{\mathcal{P}}_P = \left[\mathrm{vec}\left\{ \breve{\mathbf{\Phi}}_1 \right\}, \ldots, \mathrm{vec}\left\{ \breve{\mathbf{\Phi}}_{P^2} \right\} \right]$.

From the above derivations, the cross terms are obtained in a straightforward way (see[10] for details). Applying the partitioned matrix inversion formula [8], we obtain

$$\begin{aligned} \mathbf{CRB}_{\boldsymbol{\alpha\alpha}}^{-1} &= \boldsymbol{\mathcal{F}}_{\boldsymbol{\alpha\alpha}} + \boldsymbol{\mathcal{F}}_{\boldsymbol{\alpha}\mathbf{r_B}} \left(\boldsymbol{\mathcal{F}}_{\mathbf{r_B r_B}} - \boldsymbol{\mathcal{F}}_{\mathbf{r_B q}} \boldsymbol{\mathcal{F}}_{\mathbf{qq}}^{-1} \boldsymbol{\mathcal{F}}_{\mathbf{q r_B}} \right)^{-1} \\ &\quad \cdot \left(\boldsymbol{\mathcal{F}}_{\mathbf{r_B q}} \boldsymbol{\mathcal{F}}_{\mathbf{qq}}^{-1} \boldsymbol{\mathcal{F}}_{\mathbf{r_B}\boldsymbol{\alpha}} \right) \\ &\quad + \boldsymbol{\mathcal{F}}_{\boldsymbol{\alpha}\mathbf{q}} \left(\boldsymbol{\mathcal{F}}_{\mathbf{qq}} - \boldsymbol{\mathcal{F}}_{\mathbf{q r_B}} \boldsymbol{\mathcal{F}}_{\mathbf{r_B r_B}}^{-1} \boldsymbol{\mathcal{F}}_{\mathbf{r_B q}} \right)^{-1} \\ &\quad \cdot \left(\boldsymbol{\mathcal{F}}_{\mathbf{q r_B}} \boldsymbol{\mathcal{F}}_{\mathbf{r_B r_B}}^{-1} \boldsymbol{\mathcal{F}}_{\mathbf{q}\boldsymbol{\alpha}} \right) \quad (43) \end{aligned}$$

Finally, applying results of [10] directly to (43) leads to the closed form expression (20).

8. REFERENCES

[1] Z. Lei and T.J. Lim, “Estimation of Directions of Arrival of Multipath Signals in CDMA Systems”, *IEEE Transactions on Communications*, VOL.48, NO.6, pp.1022-1028, June 2000.

[2] A. van der Veen, M.C. Vanderveen and A. Paulraj,”Joint Angle and Delay Estimation using Shift-Invariance Techniques”, *IEEE Transactions on Signal Processing*, VOL. 45, NO. 3, March 1997.

[3] M.C. Vanderveen, A. van der Veen and A. Paulraj, “Estimation of Multipath Parameters in Wireless Communications”, *IEEE Transactions on Signal Processing*, VOL. 46, NO. 3, pp.682-690, March 1998.

[4] M. Wax and A. Leshem, “Joint Estimation of Time Delays and Directions of Arrival of Multipath Reflections of a Known Signal”,*Proceedings of the International Conference on Acoustics, Speech and Signal Processing, ICASSP’1996*, pp. 2622-2625, May 1996.

[5] B. Göransson and B. Ottersten, “Direction Estimation in Partially Unknown Noise Fields”, *IEEE Transaction on Signal Processing*, VOL.47. NO.9, pp. 2375-2385, September 1999.

[6] K. Abed-Meraim and Y. Hua, “Blind Identification of Multi-Output Systems Using Minimum Noise Subspace”, *IEEE Transactions on Signal Processing*, VOL.45, NO.1, January 1995.

[7] P. Stoica and A. Nehorai, “MUSIC, Maximum Likelihood, and the Cramer-Rao Bound”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, VOL. 37, NO 5, pp.720-741, May 1989.

[8] M. Pesavento and A.B. Gershman, “Maximum-Likelihood Direction-Of-Arrival Estimation in the presence of Unknown Nonuniform Noise”, *IEEE Transactions on Signal Processing*, VOL. 49, NO.7, pp. 1310-1324, July 2001.

[9] L.L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison Wesley, 1990.

[10] A.B. Gershman, P. Stoica, M. Pesavento and E.G.Larsson, “Stochastic Cramér-Rao Bound for Direction Estimation in Unknown Noise Fields”, *IEE Proceedings on Radar and Sonar Navig.* VOL. 149, NO. 1, pp.2-8, February 2002.