Progress In Electromagnetics Research C, Vol. 21, 13–31, 2011

JOINT ANGLE-DELAY ESTIMATION BASED ON SMOOTHED MAXIMUM-LIKELIHOOD ALGORITHM

L. Zhang and Y. Zhu

Communication and Security Lab
Shenzhen Graduate School
Peking University, Shenzhen, China

Abstract—In this paper, a novel maximum likelihood algorithm for joint angle and delay estimation is developed to identify the specular components of channel fading for uniform linear array based on the physical propagation channel model. Frequency domain pre-smoothing is applied to the structured frequency transfer matrix before the estimation procedure in order to utilize substantial observations. Iterative Gauss-Newton method is used to solve the multidimensional optimization problem, and a new compact matrix form is presented. Further simplification of the iteration is derived based on the assumption of independent channel parameters. Both simulations and measurement results are investigated for performance analysis. The simulations reveal that the proposed algorithm leads to higher performance with appropriate complexity. Also, a comparison with other algorithms is carried out to validate the accuracy of algorithm by using the power delay profile measured in a real environment, and the results show the proposed algorithm performs well.

1. INTRODUCTION

In mobile communications, source positioning is of interest for emergency service, military navigation and handover schemes in a cellular or the Global Position System (GPS). The accuracy of positioning relies highly on the performance of employed estimation algorithms which involve joint estimation of directions of departure (DOD), directions of arrival (DOA), time/time-difference of arrival (TOA/TDOA) or other physical channel parameters [1–3]. Thus,

Received 24 January 2011, Accepted 21 March 2011, Scheduled 5 April 2011
Corresponding author: Lin Zhang (zhangl06843@szcie.pku.edu.cn).

developing a robust estimation algorithm with high performance is a key issue for source positioning and other radar array applications.

In radio propagation channel, it is common to classify the signals reaching the receiver by specular or diffuse scattering components [4, 5]. The specular components, e.g., line of sight (LOS) and mirror reflection, are often believed to carry most of the path power. They are often modeled by a relatively small number of deterministic paths with small angular spread and delay spread. Diffuse scattering component is usually regarded as noise and neglected in the outdoor scenarios. However, the contribution of all diffuse components together is significant due to the comparative large amount within limited area, e.g., indoor or industry scenario. In this paper, the proposed joint estimation algorithm is designed to solve the dominant specular components of multipath channel for outdoor scenario. Diffuse component estimation, such as [4], is out of the scope of the paper.

In literatures, various high-resolution multidimensional parameter estimation methods have been proposed. ESPRIT introduced in [7] is an attractive estimation algorithm exploiting the rotational invariance of the signal subspace. The algorithm is extended to two-dimensional (2D) case in [8] for joint angle-delay estimation (JADE-ESPRIT). The so called multi-invariance MUSIC (MI-MUSIC) [10] reduces the 2D angle-delay estimation problem to one-dimensional (1D) search process with the exploitation of array invariance as ESPRIT. Maximum likelihood (ML) algorithms have been proposed for estimation of DOA [11]. Based on the knowledge of statistical properties of the transmitted signal, stochastic ML and deterministic ML algorithms, are presented leading to two distinct solutions with the stochastic one providing more accurate estimations for certain bad scenarios, like low signal to noise ratio (SNR) or high correlated channels. Several ML algorithms are extended to multidimensional channel parameter estimation. SAGE [12] allows the complex multidimensional ML based optimization problem reduced to 1D optimization process which can be performed sequentially. In [4], a more complicated ML scheme called RIMAX, has been proposed to estimate not only specular component parameters, e.g., DOA or TOA, but also diffuse component parameters.

In practice, channel sounding is needed to exploit the measured data for physical channel parameter estimation. A switched antenna sounder which contains only one physical transmit and one receive channel is used to implement multiantenna system sounding [14]. With respect to each antenna element, sequential sounding over a number of frequency subchannels, e.g., 385 over 120 MHz bandwidth in [15], is adopted to achieve high time-delay resolution since vector network analyzer can be applied. However, long measurement time interval

is needed to sample each frequency subchannel. With the limitation of channel coherent time, only a few temporal observations can be measured to keep channel statistic stationary [15, 16]. Thus, there are not enough observations to support multidimensional parameter estimation because of the rank deficiency problem. In this paper, a frequency domain smoothing technique which preserves original structure of frequency transfer function is employed to restore matrix rank by increasing number of observations. Similar smoothing scheme has been introduced in time and spatial domain for ESPRIT based joint angle and frequency estimation in [6].

In this paper, based on physical channel model an iterative ML method to estimate both TOA and DOA jointly has been proposed for uniform linear array (ULA) at receive side. Frequency smoothing is employed to the structured frequency transfer matrix as a preprocess to increasing the observation size and restore rank of data matrix. Based on the DOA estimation method [11], a similar Gauss-Newton method with a compact matrix form is extended to the 2D optimization problem. Asymptotic Hessian matrix with block diagonal structure is used to reduce the complexity further at the cost of slight performance degradation. The proposed approach is able to avoid the difficulty of optimizing very high dimensional function due to pre-smoothing. Simulations are conducted to compare the root mean-square error (RMSE) of the smoothed ML estimates with several high resolution algorithms. Power delay profile (PDP) of measured data and PDP estimated using aforementioned algorithms are also involved to validate superior performance of the proposed algorithm in real applications.

The remainder of this paper is organized as follows. In Section 2, the system model is introduced. In Section 3, the algorithm for parameter estimation is developed. In Section 4, both simulation analysis and measured results are presented to verify the performance of proposed algorithm. Throughout the paper, the notations $[\cdot]^T$ and $[\cdot]^H$ denote the transpose and Hermitian transpose, respectively. $\otimes$, $\odot$ and $\diamond$ denote the Kronecker, Hadamard and Khatri-Rao product, respectively. $|\cdot|$, $\mathrm{tr}\{\cdot\}$ and $\mathrm{E}\{\cdot\}$ denote the matrix determinant, trace and expectation, respectively. $\mathbf{I}_q$ denotes the identity matrix of size q. $\mathbf{1}_{p\times q}$ and $\mathbf{0}_{p\times q}$ denote the $p \times q$ matrix with all entries being unities and zeros, respectively. $[\cdot]_{p,q}$ denotes the (p, q)th entry of a matrix and $[\cdot]_q$ denotes qth column (entry) of a matrix (vector).

2. SYSTEM MODEL

In the paper, a wireless system with a symbol sequence $\{x(t)\}$ transmitted over a flat fading channel is considered. The receive signal

is measured using an antenna array with M isotropic elements. The noise free received signal at mth receive antenna has the form

$$y_m(t) = \sum_{l=1}^{L} h_m(t, \tau_l) x(t - \tau_l) \tag{1}$$

where L denotes the number of propagation paths, and τ_l is the excess TOA of lth specular path with first arrival delay τ_1 being zero. Since ULA is only used at receive side, DOD estimation isn't considered. The corresponding narrowband channel impulse response (CIR) referred to [1, 3, 5] is expressed by superposition of the L multipaths

$$h_m(t, \tau) = \sum_{l=1}^{L} \tilde{\gamma}_l e^{-j[2\pi(m-1)\Delta_d \sin(\theta_l(t))+\Phi_l]} \delta(t - \tau_l(t)) \tag{2}$$

where $\tilde{\gamma}_l$ denotes complex magnitude of lth path and Φ_l denotes a uniform distributed random phase term. The relative phase difference between mth element and the reference element is $\phi = -2\pi(m-1)\Delta_d \sin(\theta_l(t))$, where Δ_d denotes antenna element spacing of ULA in wavelength at receiver side and θ_l is the incidence azimuth of the lth path. In a real world MIMO sounding system, channel frequency response (CFR) is more reasonable for physical parameter estimation than CIR because the received data is measured in frequency domain [13, 16]. The equivalent CFR of (2) is

$$h_m(t, f) = \sum_{l=1}^{L} \tilde{\gamma}_l e^{-j2\pi f \tau_l(t)} e^{-j[2\pi(m-1)\Delta_d \sin(\theta_l(t))+\Phi_l]} \tag{3}$$

The radio channel is measured in the frequency domain with a broadband-signal of N equally spaced subchannels around the carrier frequency. Given the channel bandwidth B and the number of subchannels N, the frequency domain element spacing on frequency bandwidth is calculated as $\Delta_f = B/N$. Assume K discrete time observations of (3) have been sampled within the duration of channel coherent time which keeps the channel statistical stationary. The discrete CFR for mth receiver antenna element of the nth frequency subcarrier at kth time instance is expressed as

$$h_m(k, n) = \sum_{l=1}^{L} \tilde{\gamma}_l e^{-j2\pi(n-1)\Delta_f \tau_l} e^{-j[2\pi(m-1)\Delta_d \sin(\theta_l)+\Phi_l]} \tag{4}$$

The time dependency of the channel parameters has been omitted in (4) since the assumption of stationary channel statistic. The desired physical parameters which should be estimated are $\beta = [\theta^T \tau^T]^T \in$

$C^{2L\times 1}$, where $\theta = [\theta_1, \theta_2, \ldots, \theta_L]^T$ and $\tau = [\tau_1, \tau_2, \ldots, \tau_L]^T$ denote multipath DOA and TOA, respectively. It is natural to stack the above channel response samples into a NM elements vector from

$$\tilde{\mathbf{h}}_k = [h_1(k,1), \ldots, h_1(k,N), h_2(k,1), \ldots, h_M(k,1), \ldots, h_M(k,N)]^T \quad (5)$$

After dropping the time index k in $\tilde{\mathbf{h}}_k$ for notation simplicity, the stacked CFR vector is represented in matrix form

$$\tilde{\mathbf{h}} = \tilde{\mathbf{A}}\tilde{\gamma} + \mathbf{n} \quad (6)$$

where $\mathbf{n} \in C^{NM\times 1}$ is referred to as a zero mean Gaussian noise vector and $\tilde{\gamma} = [\tilde{\gamma}_1 e^{-j\Phi_1} \ \ldots \ \tilde{\gamma}_L e^{-j\Phi_L}]^T$ denotes the channel complex gain vector. The structured transfer matrix of the radio channel is

$$\tilde{\mathbf{A}} = \mathbf{A}(\theta) \diamond \tilde{\mathbf{A}}(\tau) = [\mathbf{a}(\theta_1) \otimes \tilde{\mathbf{a}}(\tau_1) \ \ldots \ \mathbf{a}(\theta_L) \otimes \tilde{\mathbf{a}}(\tau_L)] \in C^{NM\times L} \quad (7)$$

Each column of $\tilde{\mathbf{A}}$ involves contribution from an independent path l with specific DOA θ_l and TOA τ_l. $\mathbf{a}(\theta_l) = [1 \ e^{-j2\pi\Delta_d \sin(\theta_l)} \ e^{-j2\pi(M-1)\Delta_d \sin(\theta_l)}]^T$ is the DOA based spatial domain steering vector of lth path. Since it is assumed a flat fading channel, the TOA based frequency domain manifold vector $\tilde{\mathbf{a}}(\tau_l) = [1 \ e^{-j2\pi\Delta_f \tau_l} \ \ldots \ e^{-j2\pi(N-1)\Delta_f \tau_l}]^T$ also preserves a Vandermonde structure as $\mathbf{a}(\theta_l)$ of ULA. Similar channel model can be found in [4, 5] which are developed for MIMO radio channel estimation. Our model can be regarded as a simplified version with reduced apertures in the spatial domain at the transmit side.

3. JOINT ANGLE AND DELAY ESTIMATION

In this section, a data stacking process, referred to as frequency smoothing, to the stacked CFR vector (6) is presented before the ML algorithm is introduced in order to restore the column rank of channel transfer matrix by increasing observations. Then, the joint parameter estimation method is developed under the assumption of small independent observation size.

3.1. Frequency Pre-smoothing

In real channel sounding system, channel sounder has to measure channel statistics at a high speed within the channel coherent time to guarantee the time invariant channel character. Because of large number of subchannels and antenna elements, each subchannel $h_m(k,n)$ can obtain only a few observations within coherent time, e.g., in one measurement campaign [16] of RUSK channel sounder family [17] with 80 MHz bandwidth, each of 944 subchannels takes

12.8 μs for measurement which suggests the time interval between adjacent observation of the same subchannel is at least 0.012 s. As the snapshot rate taking 75 Hz, there is only one observation measured for each subchannel. Thus, if the number of independent observations is small or strong diffuse components exist, estimation performance degrades significantly. A way to improve the performance is to decompose independent observations into overlapped observations by smoothing.

Since frequency subchannels N is often much larger than antenna elements M for channel transfer matrix $\tilde{\mathbf{A}}$ in practice, smoothing procedure in frequency domain is considered with two fold benefits. First, observation number of $\tilde{\mathbf{h}}$ is increased to guarantee feasibility of the optimization problem. Secondly, row of transfer matrix $\tilde{\mathbf{A}}(\tau)$ is reduced from N to P which decreases computational complexity of a high matrix decomposition. Without loss of generality, case for $K = 1$ is used to describe our algorithm, and case for $K > 1$ can be extended straightforwardly. Let the ith selection matrix $\mathbf{J}_i \in C^{PM \times NM}$ be

$$\mathbf{J}_i = \mathbf{I}_M \otimes [\mathbf{0}_{P\times(i-1)}\ \mathbf{I}_P\ \mathbf{0}_{P\times(N-P-i+1)}] \quad i = 1, \ldots, N-P+1 \tag{8}$$

By selecting part of the data vector $\tilde{\mathbf{h}}$ that corresponds to the ith overlapped observation $\mathbf{h}_i = \mathbf{J}_i\tilde{\mathbf{h}}$, the reconstructed channel matrix is

$$\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_{N-P+1}] \in C^{PM\times(N-P+1)} \tag{9}$$

With reduction of frequency bandwidth from $N\Delta_f$ to $P\Delta_f$, the number of observations is effectively increased by a factor of $N-P+1$. The smoothed channel model referred to (6) is expressed as

$$\mathbf{H} = \mathbf{A}\Gamma + \mathbf{N} \tag{10}$$

where each column of $\mathbf{N} \in C^{PM\times(N-P+1)}$ is a Gaussian white noise vector which is shuffled in the same way as $\mathbf{h}_i$ with entries being zero-mean and variance σ^2. Based on the unconditional model assumption (UMA) of [11], complex gain Γ is also assumed a zero-mean Gaussian random process which indicates $E\{\mathbf{h}_i\} = \mathbf{0}_{PM\times 1}$. The new transfer matrix is defined as $\mathbf{A} = \mathbf{J}_1\tilde{\mathbf{A}} \in C^{PM\times L}$ whose column corresponds to the new TOA-based frequency manifold vector $\mathbf{a}(\tau_l) = [1\ e^{-j2\pi\Delta_f\tau_l}\ \ldots\ e^{-j2\pi(P-1)\Delta_f\tau_l}]^T \in C^{P\times 1}$. For simplicity, let $\mathbf{a}_l(\theta,\tau) = \mathbf{a}(\theta_l)\otimes\mathbf{a}(\tau_l)$, and channel transfer matrix (7) is rewritten as

$$\mathbf{A} = \mathbf{A}(\theta)\diamond\mathbf{A}(\tau) = [\mathbf{a}_1(\theta,\tau)\ \ldots\ \mathbf{a}_L(\theta,\tau)] \in C^{PM\times L} \tag{11}$$

Compared with the unsmoothed transfer matrix (7), the only difference is the reduction of frequency bandwidth from $N\Delta_f$ to $P\Delta_f$ which renders the multipath resolution.

3.2. Maximum Likelihood Estimation

Since no LOS component is considered in the paper, the Rayleigh fading is reasonable for the stacked channel frequency response matrix $\mathbf{H}$ in (10) with the covariance matrix of $\mathbf{h}_i \sim \mathcal{N}(\mathbf{0}_{PM\times 1}, \mathbf{R}_H)$ expressed as

$$\mathbf{R}_H = \mathrm{E}\{\mathbf{h}_i \mathbf{h}_i^H\} = \mathbf{A}\mathbf{R}_\Gamma \mathbf{A}^H + \sigma^2 \mathbf{I}_{PM} \tag{12}$$

where both covariance matrix of complex gain $\mathbf{R}_\Gamma = \frac{1}{N-P+1}\mathrm{E}\{\Gamma\Gamma^H\}$ and the noise variance σ^2 are unknown. The joint probability density function (pdf) of $\mathbf{H}$ for $N-P+1$ overlapped observations is expressed by the product of pdf $p(\mathbf{h}_i)$

$$p(\mathbf{H}) = \prod_{i=1}^{N-P+1} p(\mathbf{h}_i) = \frac{1}{|\pi \mathbf{R}_H|^{N-P+1}} e^{-\mathrm{tr}\left\{\mathbf{H}^H \mathbf{R}_H^{-1} \mathbf{H}\right\}} \tag{13}$$

After dropping the constants term and dividing through by $N-P+1$, the negative log-likelihood function of (13) is reduced to

$$\mathcal{L}_0(\theta, \tau, \mathbf{R}_\Gamma, \sigma^2) = \ln |\mathbf{R}_H| + \mathrm{tr}\left\{\mathbf{R}_H^{-1} \hat{\mathbf{R}}_H\right\} \tag{14}$$

where $\hat{\mathbf{R}}_H$ is the sample covariance matrix defined as $\hat{\mathbf{R}}_H = \frac{1}{N-P+1}\mathbf{H}\mathbf{H}^H$. Based on the derivation of [9], ML estimator of σ^2 and $\mathbf{R}_\Gamma$ with parameter θ and τ are calculated by minimizing (14)

$$\begin{aligned} \hat{\sigma}^2(\theta, \tau) &= \frac{1}{PM - L} \mathrm{tr}\left\{\left(\mathbf{I}_{PM} - \mathbf{A}\mathbf{A}^\dagger\right) \hat{\mathbf{R}}_H\right\} \\ &= \frac{1}{PM - L} \mathrm{tr}\left\{\mathbf{P}_A^\perp \hat{\mathbf{R}}_H\right\} \\ \hat{\mathbf{R}}_\Gamma(\theta, \tau) &= \mathbf{A}^\dagger \left(\hat{\mathbf{R}}_H - \hat{\sigma}^2 \mathbf{I}_{PM}\right) \mathbf{A}^{\dagger H} \end{aligned} \tag{15}$$

where $\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ and $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{A}\mathbf{A}^\dagger$ are pseudo-inverse matrix and orthogonal projection matrix of $\mathbf{A}$, respectively. If $\mathbf{R}_\Gamma$ is strictly positive definite, since the ML estimates are consistent, $\hat{\mathbf{R}}_\Gamma$ tends to $\mathbf{R}_\Gamma$ as observations increase to infinite. Thus, $\hat{\mathbf{R}}_\Gamma$ must be positive definite for a valid ML estimator. Then, substituting $\hat{\mathbf{R}}_\Gamma$ and $\hat{\sigma}^2$ back into (14) and after some manipulation, the desired parameters β are estimated by minimizing the objective function (16)

$$\beta = \arg\min_{\theta,\tau} \left\{\ln \left|\mathbf{A}(\theta, \tau) \hat{\mathbf{R}}_\Gamma(\theta, \tau) \mathbf{A}^H(\theta, \tau) + \hat{\sigma}^2(\theta, \tau) \mathbf{I}_{PM}\right|\right\} \tag{16}$$

The parameters (θ, τ) are suppressed for simplicity of notation. Then using $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$ and (15), (16) can be rewritten as

$$\begin{aligned}\mathcal{L} &= \ln\{\hat{\sigma}^{2PM}|\hat{\sigma}^{-2}\hat{\mathbf{R}}_\Gamma \mathbf{A}^H\mathbf{A} + \mathbf{I}_L|\} \\ &= \ln\{\hat{\sigma}^{2PM}|\hat{\sigma}^{-2}\mathbf{A}^\dagger(\hat{\mathbf{R}}_H - \hat{\sigma}^2\mathbf{I}_{PM})\mathbf{A} + \mathbf{I}_L|\} \\ &= \ln\{\hat{\sigma}^{2(PM-L)}|\mathbf{A}^\dagger\hat{\mathbf{R}}_H\mathbf{A}|\} \end{aligned} \tag{17}$$

It is hard to derive a closed form solution for the nonlinear Equation (17). Thus, Gauss-Newton method is utilized to solve the optimization problem. Since computation of exact Hessian matrix for the ML criterion is cumbersome, an alternative way to overcome the difficulties is to use a less complex approximation of Hessian matrix $\mathbf{W}$ which is also guaranteed a positive semidefinite matrix. At the jth iteration

$$\beta_{j+1} = \beta_j - \mu_j \mathbf{W}_j^{-1}\mathbf{g}_j. \tag{18}$$

where $\mu \in (0, 1)$ denotes the preassigned iteration step, and the $2L \times 1$ gradient vector is $\mathbf{g} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = (\frac{\partial \mathcal{L}}{\partial \theta_1} \ \dots \ \frac{\partial \mathcal{L}}{\partial \theta_L} \ \frac{\partial \mathcal{L}}{\partial \tau_1} \ \dots \ \frac{\partial \mathcal{L}}{\partial \tau_L})^T$ which will be $\mathbf{0}$ when it converges to a stationary point. The entries of the gradient vector $\mathbf{g}$ with respect to θ_l and τ_l derived in Appendix A are

$$\frac{\partial \mathcal{L}}{\partial \theta_l} = 2\mathrm{Re}\left\{\mathrm{tr}\left[\left((\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A})^{-1} - \frac{(\mathbf{A}^H\mathbf{A})^{-1}}{\hat{\sigma}^2}\right)\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{P}_A^\perp\frac{\partial \mathbf{A}}{\partial \theta_l}\right]\right\} \tag{19}$$

$$\frac{\partial \mathcal{L}}{\partial \tau_l} = 2\mathrm{Re}\left\{\mathrm{tr}\left[\left((\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A})^{-1} - \frac{(\mathbf{A}^H\mathbf{A})^{-1}}{\hat{\sigma}^2}\right)\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{P}_A^\perp\frac{\partial \mathbf{A}}{\partial \tau_l}\right]\right\} \tag{20}$$

where $\mathrm{Re}\{\cdot\}$ denotes the real part of a complex number. The lth column of partial derivative matrix of $\mathbf{A}$ with respect to θ_l and τ_l are

$$\left[\frac{\partial \mathbf{A}}{\partial \theta_l}\right]_l = \frac{\partial \mathbf{a}_l(\theta, \tau)}{\partial \theta_l} = \mathbf{d}(\theta_l) \quad \text{and} \quad \left[\frac{\partial \mathbf{A}}{\partial \tau_l}\right]_l = \frac{\partial \mathbf{a}_l(\theta, \tau)}{\partial \tau_l} = \mathbf{d}(\tau_l) \tag{21}$$

respectively, and the remaining columns are zeros. Then, an intermediate matrix $\Omega = (\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A})^{-1} - \hat{\sigma}^{-2}(\mathbf{A}^H\mathbf{A})^{-1}$ is employed for brevity. After some manipulation, (19) and (20) are rewritten as

$$\frac{\partial \mathcal{L}}{\partial \theta_l} = 2\mathrm{Re}\left\{\left[\Omega\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{P}_A^\perp\mathbf{d}(\theta_l)\right]_l\right\} \tag{22}$$

$$\frac{\partial \mathcal{L}}{\partial \tau_l} = 2\mathrm{Re}\left\{\left[\Omega\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{P}_A^\perp\mathbf{d}(\tau_l)\right]_l\right\} \tag{23}$$

With two new stacked derivative matrix $\mathbf{D}_\theta = [\mathbf{d}(\theta_1)\,\mathbf{d}(\theta_2)\,\dots\,\mathbf{d}(\theta_L)]$ and $\mathbf{D}_\tau = [\mathbf{d}(\tau_1)\,\mathbf{d}(\tau_2)\,\dots\,\mathbf{d}(\tau_L)]$ defined, the gradient vector with

respect to θ and τ are reduced to

$$\mathbf{g}_\theta = 2\mathrm{Re}\left\{\mathrm{diag}\left((\Omega\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{P}_A^\perp)\mathbf{D}_\theta\right)\right\} \tag{24}$$

$$\mathbf{g}_\tau = 2\mathrm{Re}\left\{\mathrm{diag}\left((\Omega\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{P}_A^\perp)\mathbf{D}_\tau\right)\right\} \tag{25}$$

where diag$(\cdot)$ denotes a vector consisting of the diagonal element of matrix. Finally, let $\mathbf{D} = [\mathbf{D}_\theta\mathbf{D}_\tau]$ and the gradient vector $\mathbf{g} = [\mathbf{g}_\theta^T\mathbf{g}_\tau^T]^T$ in compact matrix form is expressed as

$$\mathbf{g} = 2\mathrm{Re}\left\{\mathrm{diag}\left[\left(\mathbf{1}_{2\times1}\otimes(\Omega\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{P}_A^\perp)\right)\mathbf{D}\right]\right\} \tag{26}$$

In a rough analogy, the asymptotic Hessian matrix of $\mathcal{L}$ is

$$\mathbf{W} = 2\hat{\sigma}^2\mathrm{Re}\left\{\left[\mathbf{1}_{2\times2}\otimes(\Omega\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A}\Omega)^T\right]\odot(\mathbf{D}^H\mathbf{P}_A^\perp\mathbf{D})\right\} \tag{27}$$

The derivation of (27) is included in Appendix B. The iteration is continued until a convergence criterion, such as $||\beta_{j+1}-\beta_j||^2 < \varepsilon$, is satisfied [18], where ε is an arbitrarily small constant.

In order to implement the iteration, a matrix of second derivatives has to be computed and a $2L \times 2L$ matrix is inverted. A simpler approximation which can converge in an adequate rate is found in our work. Since DOA and TOA are assumed two independent parameter vectors, it is reasonable to assume the cross correlation between the two vectors are zeros which is used to simplify (27) to a block diagonal matrix. Then DOA and TOA can be estimated separately

$$\begin{cases} \theta_{j+1} = \theta_j - \mu_j\mathbf{W}_\theta^{-1}\mathbf{g}_\theta \\ \tau_{j+1} = \tau_j - \mu_j\mathbf{W}_\tau^{-1}\mathbf{g}_\tau \end{cases} \tag{28}$$

where $\mathbf{W}_\theta$ and $\mathbf{W}_\tau$ as expressed in (B1) are the $L\times L$ block submatrix of $\mathbf{W}$ associated with θ and τ, respectively.

3.3. Multi-invariance MUSIC

For comparison with the proposed algorithm, MI-MUSIC referred to [10] is introduced in this subsection. First, eigenvalue decomposition is adopted for $\mathbf{R_H}$ in (12), and let $\mathbf{E}_s$ and $\mathbf{E}_n$ be referred to as signal space eigenvector and noise space eigenvector matrixes of $\mathbf{R_H}$, respectively. Then, the MI-MUSIC spatial spectrum function is

$$f_{music}(\tau_l,\theta_l) = \frac{1}{(\mathbf{a}(\theta_l)\otimes\mathbf{a}(\tau_l))^H\,\mathbf{E}_n\mathbf{E}_n^H\,(\mathbf{a}(\theta_l)\otimes\mathbf{a}(\tau_l))}$$

After manipulations as presented in [10], the 2D angle-delay estimation problem reduces to 1D search process which means DOA and TOA can be estimated serially. The score function to estimate the lth angle is

$$\hat{\theta}_l = \arg\min_{\theta_l}\mathbf{a}^H(\tau_l)\left\{(\mathbf{a}(\theta_l)\otimes\mathbf{I}_N)^H\,\Pi_s^\perp\,(\mathbf{a}(\theta_l)\otimes\mathbf{I}_N)\right\}\mathbf{a}(\tau_l)$$

where $\Pi_s^{\perp} = \mathbf{I} - \mathbf{E}_s(\mathbf{E}_s^H\mathbf{E}_s)^{-1}\mathbf{E}_s^H$ denotes the orthogonal projection matrix of $\mathbf{E}_s$. Similarly, the score function to estimate the lth delay is

$$\hat{\tau}_l = \arg\min_{\tau_l} \mathbf{a}^H(\theta_l)\left\{(\mathbf{I}_M \otimes \mathbf{a}(\tau_l))^H \Pi_s^{\perp} (\mathbf{I}_M \otimes \mathbf{a}(\tau_l))\right\}\mathbf{a}(\theta_l)$$

3.4. Cramer-Rao Bounds

The theoretical Cramer-Rao bounds (CRB) for stochastic signal model is also employed for performance comparison. Since a zero-mean complex Gaussian random process with covariance matrix $\mathbf{R_H}$ has been assumed for $\mathbf{H}$ in (10), based on the corresponding derivation of [18, 8.4.1] with (12), the submatrix of CRB with respect to $\omega \in \{\theta, \tau\}$ is

$$\mathbf{C}_{CR}(\omega) = \frac{\sigma^2}{2K}\left\{\mathrm{Re}\left[\left(\mathbf{R}_\Gamma \mathbf{A}^H \mathbf{R}_{\mathbf{H}}^{-\mathbf{1}} \mathbf{A}\mathbf{R}_\Gamma\right) \odot \left(\mathbf{D}_\omega^H \mathbf{P}_A^{\perp} \mathbf{D}_\omega\right)^T\right]\right\}^{-1} \quad (29)$$

which was first used as the lowest bound for angle ML estimator in [11]. In the paper, bounds associated with both angle and delay are taken into account.

4. SIMULATION AND EXPERIMENTAL RESULTS

In order to examine the performance of the proposed ML estimation algorithm, several simulations are conducted for different scenarios. Measurement based power delay spectrum analysis are also involved to evaluate the algorithm's performance.

4.1. Simulation Analysis

For all simulations, stationary channel statistics during the transmission and independent fading from one to another observation are assumed. A four-element ULA is considered with separation $\Delta_d = 1/2$. The carrier frequency is 5.2 GHz with channel bandwidth $B = 20$ MHz corresponding to $N = 80$ frequency subchannels which satisfies the constraint of stationary channel statistics. Two far field and equipower specular path signals are impinging on the antenna array. DOA and TOA are defined as $\theta = \{-20°, 12.5°\}$ and $\tau = \{0.27\,\mu\mathrm{s}, 0.45\,\mu\mathrm{s}\}$, respectively. Pre-smoothing is performed only in frequency domain for TOA refinement. Extension for DOA estimation is straightforward. RMSE is investigated versus the number of independent observations K, SNR and the number of subchannels P. All simulation results are based on 2000 Monte Carlo runs. The behaviors are summarized

in Figures 1–4. The proposed smoothed ML (S-ML) algorithm is labeled with a □ marker, the conventional ML (C-ML) method [11] extended to 2D case without smoothing is labeled △, ESPRIT involving smoothing technique as [6] (S-ESPRIT) is labeled with a + marker, MI-MUSIC [10] is labeled with ∗, and CRB is considered as performance reference.

4.1.1. Different Observation Number

In Figure 1, the RMSE of S-ML is compared with that of S-ESPRIT, C-ML and MI-MUSIC as a function of independent observation size K for SNR = 5 dB. For TOA estimation, S-ML always outperforms S-ESPRIT and C-ML, while MI-MUSIC shows comparable performance except for small K region. For DOA estimation, MI-MUSIC degrades rapidly as K decreasing, while S-ML and C-ML have the same performance since no smoothing is used in spatial domain.

In Figure 2, comparison for SNR = 20 dB is shown. Similar performance behaviors to the SNR = 5 dB case can be observed. Meanwhile, the smaller K is, the better the relative performance is. In both Figures 1 and 2, S-ML achieves CRB associated with TOA asymptotically when K is large. Since no smoothing is performed for DOA estimation, the CRB associated with TOA hasn't been attained.

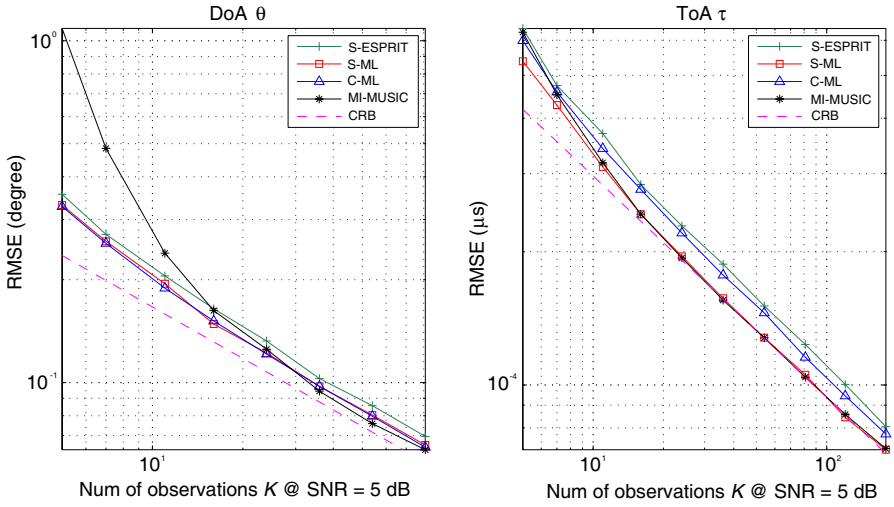

Figure 1. RMSE of DOA and TOA as a function of independent observation number K, $P = 70$ and SNR = 5 dB.

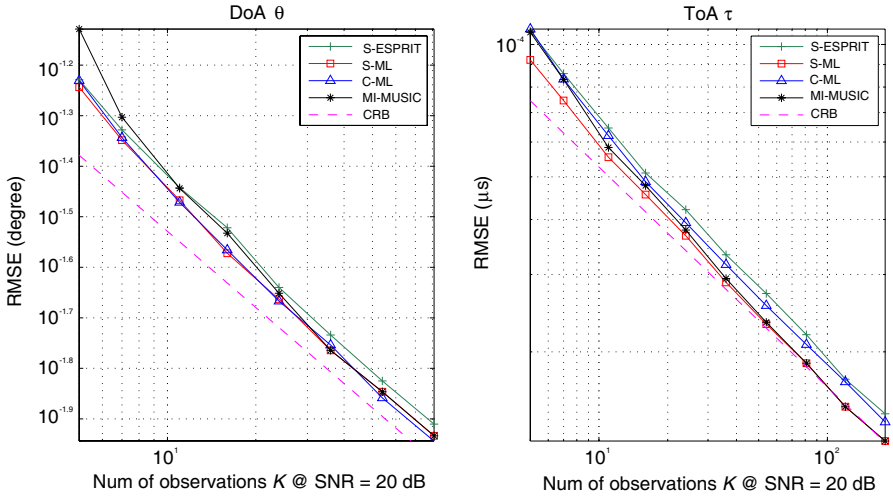

Figure 2. RMSE of DOA and TOA as a function of independent observation number K, $P = 70$ and SNR $= 20\,\mathrm{dB}$.

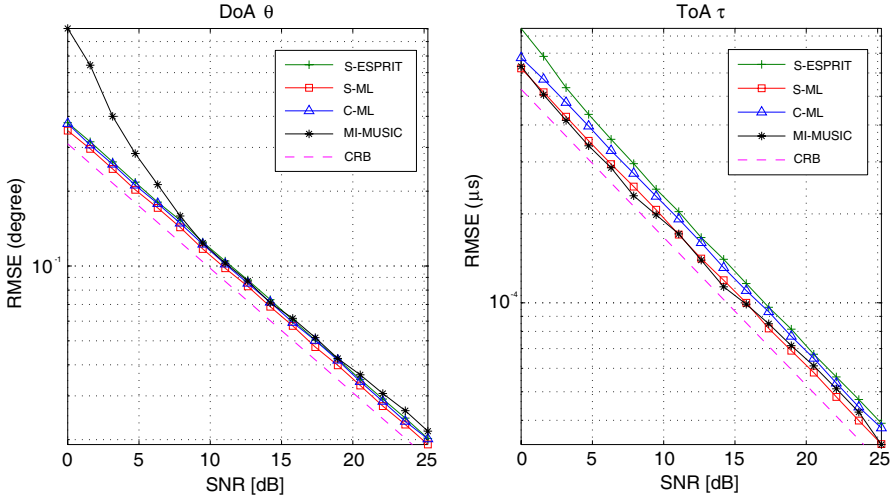

Figure 3. RMSE of DOA and TOA versus SNR, $K = 10$ and $P = 70$.

4.1.2. Different SNR

RMSEs of all the algorithms are compared with varying SNR from 0 to 25 dB at fixed observations $K = 10$ and $P = 70$ in Figure 3. Compared with S-ESPRIT and C-ML, S-ML has lower asymptotic

estimation error for TOA estimation. For the DOA estimation without pre-smoothing, slight improvement of S-ML can be observed, while the performance of MI-MUSIC is worse than the other algorithms in low SNR region. Because small K is assumed, there is still a performance gap to achieve CRB in despite of high SNR.

4.1.3. Different Smoothing Factor

The effect of frequency domain smoothing on TOA estimation is shown in Figure 4. The smoothing factor is expressed as the ratio of P over N. The simulation was designed with a small observation size $K = 10$ for SNR $= 5$ dB and SNR $= 20$ dB, respectively. RMSE of C-ML and MI-MUSIC remain constant over the varying P region because no smoothing is used for both algorithms. For SNR $= 20$ dB as shown in the right subfigure, the performance of S-ESPRIT degrades monotonously as P increases. Meanwhile, S-ML achieves CRB first, e.g., $P = 76$, before degradation, and finally attains the performance of C-ML at $P = N$. Notably, if P is much smaller than N, S-ESPRIT may outperform S-ML which implies that appropriate choice of P is desired to obtain optimal performance of S-ML. Since the subchannel size P has no impact on DOA estimation errors of all the algorithms, no corresponding analysis is taken into account in this subsection.

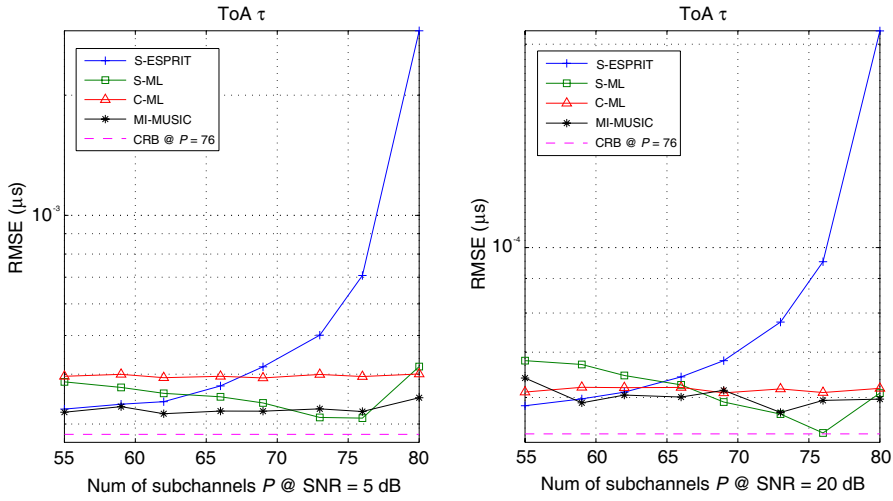

Figure 4. RMSE of TOA versus frequency subchannels P, $K = 10$.

4.2. Measurement Verification

Measured data is free accessible from the RUSK channel sounder [17]. The measurements take place in a suburban environment in Ilmenau [15]. The receiver is fixed with a 8 elements ULA, and the transmitter with an isotropic antenna moves at 10 km/h. NLOS propagation is considered during the measurement. The data including 20 consecutive observations is measured at a carrier frequency of 5.2 GHz with 512 frequency subchannels corresponding to bandwidth $B = 120$ MHz.

Figure 5 depicts measured PDP by averaging over observations and remainder PDP by taking away the estimated specular components from measured data as introduced in [16]. TOAs and DOAs of 4 independent specular paths are estimated by C-ML, S-ESPRIT and S-ML, respectively. Then the estimated CFR are reconstructed by (6) and remainder PDP are calculated. In Figure 5, + marked curve denotes the measured PDP, and the other curves denote remainder PDPs estimated by aforementioned algorithms. The major difference between measured and remainder PDP comes from low delay region, e.g., 0.3 μs and 0.5 μs, where dominant specular components exist. With specular components discarded, lower remainder PDP which decays exponentially [2] corresponding to diffuse components indicates a better estimation performance. At the delay of 0.3 μs, S-ML can get approximately 1 dB improvement than S-ESPRIT. This is a strong indication that the proposed method can achieve superior performance.

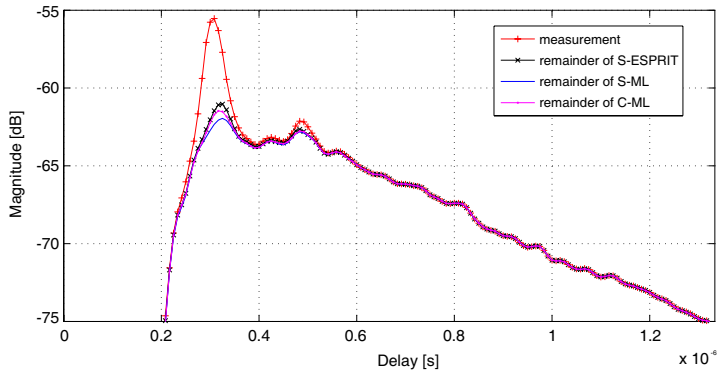

Figure 5. Power delay profile as a function of time delay.

5. SUMMARY

In this paper, a joint angle and delay ML algorithm is proposed for specular component parameters estimation based on physical propagation channel model. The stacked channel frequency response vectors are rearranged by frequency domain smoothing before the estimation to increase observation size. The shift-invariance structures of the 2D channel transfer matrix is preserved after smoothing. Gauss-Newton method is used to solve the ML optimization problem and a compact matrix form is derived. Further simplification is developed to make the iterative convergence faster. Comparisons between the proposed algorithm and several high resolution algorithms are studied which indicates that pre-smoothing can improve the performance. S-ML achieves its optimal performance when the number of frequency subchannels is appropriately chosen. In real applications with sounded channel data, its superior ability to handle slight inaccuracies of other algorithms is also shown by comparing the corresponding PDPs.

ACKNOWLEDGMENT

The work described in this paper was supported by the Shenzhen municipal of China "Shuang Bai Project". The authors thank Dr. S. H. Leung in City University of Hong Kong for the algorithm discussion.

APPENDIX A. DERIVATION OF GRADIENT VECTOR

Let us decompose (17) into two components as

$$\mathcal{L} = \underbrace{(PM - L)\ln\{\hat{\sigma}^2\}}_{\mathcal{L}_1} + \underbrace{\ln\{|\mathbf{A}^\dagger \hat{\mathbf{R}}_H \mathbf{A}|\}}_{\mathcal{L}_2} \tag{A1}$$

In order to compute the derivatives of (A1) with respect to each parameter it will involve the following relations [17, Appendix A.3]

$$\frac{\partial \mathbf{A}^\dagger}{\partial \theta} = (\mathbf{A}^H\mathbf{A})^{-1}\frac{\partial \mathbf{A}^H}{\partial \theta}\mathbf{P}_A^\perp - \mathbf{A}^\dagger\frac{\partial \mathbf{A}}{\partial \theta}\mathbf{A}^\dagger \tag{A2}$$

$$\frac{\partial \mathbf{P}_A^\perp}{\partial \theta} = -2\mathrm{Re}\left\{\mathbf{P}_A^\perp\frac{\partial \mathbf{A}}{\partial \theta}\mathbf{A}^\dagger\right\} \tag{A3}$$

Take derivative of $\mathcal{L}_1$ with respect to ith DOA θ_i as example, and substitution of (15) and (A3) into (A4) results in

$$\frac{\partial \mathcal{L}_1}{\partial \theta_i} = \frac{1}{\hat{\sigma}^2}\mathrm{tr}\left\{\frac{\partial \mathbf{P}_A^\perp}{\partial \theta_i}\hat{\mathbf{R}}_H\right\} = -\frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{\mathrm{tr}\left[\mathbf{A}^\dagger\hat{\mathbf{R}}_H\mathbf{P}_A^\perp\frac{\partial \mathbf{A}}{\partial \theta_i}\right]\right\} \tag{A4}$$

Similarly, plug in (A2) and calculate the derivative of $\mathcal{L}_2$

$$\begin{aligned}\frac{\partial \mathcal{L}_2}{\partial \theta_i} &= \mathrm{tr}\left\{(\mathbf{A}^\dagger \hat{\mathbf{R}}_H \mathbf{A})^{-1} \frac{\partial (\mathbf{A}^\dagger \hat{\mathbf{R}}_H \mathbf{A})}{\partial \theta_i}\right\} \\ &= \mathrm{tr}\left\{(\mathbf{A}^\dagger \hat{\mathbf{R}}_H \mathbf{A})^{-1}\left[(\mathbf{A}^H\mathbf{A})^{-1}\frac{\partial \mathbf{A}^H}{\partial \theta_i}\mathbf{P}_A^\perp \hat{\mathbf{R}}_H \mathbf{A}\right.\right. \\ &\quad \left.\left. -\mathbf{A}^\dagger \frac{\partial \mathbf{A}}{\partial \theta_i}\mathbf{A}^\dagger \hat{\mathbf{R}}_H \mathbf{A} + \mathbf{A}^\dagger \hat{\mathbf{R}}_H \frac{\partial \mathbf{A}}{\partial \theta_i}\right]\right\} \\ &= 2\mathrm{Re}\left\{\mathrm{tr}\left[(\mathbf{A}^H \hat{\mathbf{R}}_H \mathbf{A})^{-1}\mathbf{A}^H \hat{\mathbf{R}}_H \frac{\partial \mathbf{A}}{\partial \theta_i} - \mathbf{A}^\dagger \frac{\partial \mathbf{A}}{\partial \theta_i}\right]\right\} \\ &= 2\mathrm{Re}\left\{\mathrm{tr}\left[(\mathbf{A}^H \hat{\mathbf{R}}_H \mathbf{A})^{-1}\mathbf{A}^H \hat{\mathbf{R}}_H \left(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger\right)\frac{\partial \mathbf{A}}{\partial \theta_i}\right]\right\} \quad \text{(A5)}\end{aligned}$$

Combining (A4) and (A5), the ith entry of gradient vector (19) can be derived, and in a similar way (20) can be obtained.

APPENDIX B. ASYMPTOTIC HESSIAN MATRIX

Based on the definition of asymptotic Hessian matrix:

$$\mathbf{W} = \mathrm{E}\left\{\frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta^T}\right\} = \left[\begin{array}{c|c} \mathbf{W}_\theta & \mathbf{W}_{\theta,\tau} \\ \mathbf{W}_{\tau,\theta} & \mathbf{W}_\tau \end{array}\right] \quad \text{(B1)}$$

Take the second derivative of $\mathcal{L}$ with respect to θ, $[\mathbf{W}_\theta]_{i,j}$, as example. First, differentiate (19) with respect to θ_j

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} &= 2\mathrm{Re}\left\{\mathrm{tr}\left[\frac{\partial (\Omega \mathbf{A}^H \hat{\mathbf{R}}_H \mathbf{P}_A^\perp)}{\partial \theta_j}\frac{\partial \mathbf{A}}{\partial \theta_i} + \Omega \mathbf{A}^H \hat{\mathbf{R}}_H \mathbf{P}_A^\perp \frac{\partial}{\partial \theta_j}\left(\frac{\partial \mathbf{A}}{\partial \theta_i}\right)\right]\right\} \\ &= 2\mathrm{Re}\left\{\mathrm{tr}\left[\left(\frac{\partial \Omega}{\partial \theta_j}\mathbf{A}^H \hat{\mathbf{R}}_H \mathbf{P}_A^\perp + \Omega \frac{\partial (\mathbf{A}^H \hat{\mathbf{R}}_H \mathbf{P}_A^\perp)}{\partial \theta_j}\right)\frac{\partial \mathbf{A}}{\partial \theta_i}\right.\right. \\ &\quad \left.\left. +\Omega \mathbf{A}^H \hat{\mathbf{R}}_H \mathbf{P}_A^\perp \frac{\partial^2 \mathbf{A}}{\partial \theta_i \partial \theta_j}\right]\right\}\end{aligned}$$

Since $\mathbf{R}_H(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)$ spans the null subspace of $\mathbf{A}^H$ [19], $\mathbf{A}^H \mathbf{R}_H \mathbf{P}_A^\perp = 0$ holds. Meanwhile, the fact that $\mathrm{E}\{\hat{\mathbf{R}}_H\} = \mathbf{R}_H$ is used, then (i,j)th

entry of submatrix $\mathbf{W}_\theta$ can be expressed as

$$\begin{aligned}
[\mathbf{W}_\theta]_{i,j} &= \mathrm{E}\left\{\frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j}\right\} \\
&= 2\mathrm{Re}\left\{\mathrm{tr}\left[\mathbf{\Omega}\left(\frac{\partial \mathbf{A}^H}{\partial\theta_j}\hat{\mathbf{R}}_H\mathbf{P}_A^\perp + \mathbf{A}^H\hat{\mathbf{R}}_H\frac{\partial \mathbf{P}_A^\perp}{\partial\theta_j}\right)\frac{\partial \mathbf{A}}{\partial\theta_i}\right]\right\} \\
&= 2\mathrm{Re}\left\{\mathrm{tr}\left[\Omega\left(\frac{\partial \mathbf{A}^H}{\partial\theta_j}\hat{\mathbf{R}}_H\mathbf{P}_A^\perp\right.\right.\right. \\
&\qquad \left.\left.\left. -\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\frac{\partial \mathbf{A}^H}{\partial\theta_j}\mathbf{P}_A^\perp\right)\frac{\partial \mathbf{A}}{\partial\theta_i}\right]\right\} \\
&= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{\mathrm{tr}\left[\mathbf{U}\frac{\partial \mathbf{A}^H}{\partial\theta_j}\mathbf{P}_A^\perp\frac{\partial \mathbf{A}}{\partial\theta_i}\right]\right\} \qquad \text{(B2)}
\end{aligned}$$

where $\mathbf{U}$ denotes an intermediate matrix which is introduced in [11],

$$\begin{aligned}
\mathbf{U} &= \left(\hat{\sigma}^2(\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A})^{-1} - (\mathbf{A}^H\mathbf{A})^{-1}\right)\left(\hat{\sigma}^2\mathbf{I} - (\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A})(\mathbf{A}^H\mathbf{A})^{-1}\right) \\
&= \hat{\sigma}^4\Omega\mathbf{A}^H\hat{\mathbf{R}}_H\mathbf{A}\Omega \in C^{L\times L} \qquad \text{(B3)}
\end{aligned}$$

Therefore, the other submatrices of $\mathbf{W}$ are derived in a similar way

$$[\mathbf{W}_{\tau,\theta}]_{i,j} = \mathrm{E}\left\{\frac{\partial^2 \mathcal{L}}{\partial\theta_i\partial\tau_j}\right\} = \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{\mathrm{tr}\left[\mathbf{U}\frac{\partial \mathbf{A}^H}{\partial\tau_j}\mathbf{P}_A^\perp\frac{\partial \mathbf{A}}{\partial\theta_i}\right]\right\} \qquad \text{(B4)}$$

$$[\mathbf{W}_{\tau}]_{i,j} = \mathrm{E}\left\{\frac{\partial^2 \mathcal{L}}{\partial\tau_i\partial\tau_j}\right\} = \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{\mathrm{tr}\left[\mathbf{U}\frac{\partial \mathbf{A}^H}{\partial\tau_j}\mathbf{P}_A^\perp\frac{\partial \mathbf{A}}{\partial\tau_i}\right]\right\} \qquad \text{(B5)}$$

Then, the substitution of (21) into (B2), (B4) and (B5) yields

$$\begin{aligned}
[\mathbf{W}_\theta]_{i,j} &= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{[\mathbf{U}]_{i,j}\mathbf{d}(\theta_j)^H\mathbf{P}_A^\perp\mathbf{d}(\theta_i)\right\} \\
&= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{[\mathbf{U}]_{i,j}\left[\mathbf{D}_\theta^H\mathbf{P}_A^\perp\mathbf{D}_\theta\right]_{j,i}\right\} \qquad \text{(B6)} \\
[\mathbf{W}_{\tau,\theta}]_{i,j} &= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{[\mathbf{U}]_{i,j}\mathbf{d}(\tau_j)^H\mathbf{P}_A^\perp\mathbf{d}(\theta_i)\right\} \\
&= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{[\mathbf{U}]_{i,j}\left[\mathbf{D}_\tau^H\mathbf{P}_A^\perp\mathbf{D}_\theta\right]_{j,i}\right\} \qquad \text{(B7)} \\
[\mathbf{W}_{\tau}]_{i,j} &= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{[\mathbf{U}]_{i,j}\mathbf{d}(\tau_j)^H\mathbf{P}_A^\perp\mathbf{d}(\tau_i)\right\} \\
&= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{[\mathbf{U}]_{i,j}\left[\mathbf{D}_\tau^H\mathbf{P}_A^\perp\mathbf{D}_\tau\right]_{j,i}\right\} \qquad \text{(B8)}
\end{aligned}$$

The asymptotic Hessain matrix (B1) is reduced to:

$$\begin{aligned}\mathbf{W} &= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{\left[\begin{array}{c|c} \mathbf{U}^T \odot \left(\mathbf{D}_\theta^H \mathbf{P}_A^\perp \mathbf{D}_\theta\right) & \mathbf{U}^T \odot \left(\mathbf{D}_\tau^H \mathbf{P}_A^\perp \mathbf{D}_\theta\right) \\ \mathbf{U}^T \odot \left(\mathbf{D}_\theta^H \mathbf{P}_A^\perp \mathbf{D}_\tau\right) & \mathbf{U}^T \odot \left(\mathbf{D}_\tau^H \mathbf{P}_A^\perp \mathbf{D}_\tau\right) \end{array}\right]\right\} \\ &= \frac{2}{\hat{\sigma}^2}\mathrm{Re}\left\{\left(\mathbf{1}_{2\times 2} \otimes \mathbf{U}^T\right) \odot \left(\mathbf{D}^H \mathbf{P}_A^\perp \mathbf{D}\right)\right\} \end{aligned} \tag{B9}$$

Plug in (B3), and the asymptotic Hessian matrix (27) can be derived.

REFERENCES

1. Steinbauer, M., A. Molisch, and E. Bonek, "The doubledirectional radio channel," *IEEE Antennas and Propagation Magazine*, Vol. 43, No. 4, 51–63, Aug. 2001.
2. Asplund, H., A. A. Glazunov, A. F. Molisch, K. I. Pedersen, and M. Steinbauer, "The COST 259 directional channel modelpart II: Macrocells," *IEEE Trans. on Wireless Communications*, Vol. 5, No. 12, 3434–3450, 2006.
3. Calcev, G., D. Chizhik, B. Goransson, S. Howard, H. Huang, A. Kogiantis, A. Molisch, A. Moustakas, D. Reed, and H. Xu, "A wideband spatial channel model for system-wide simulations," *IEEE Trans. on Vehicle Technology*, Vol. 56, No. 2, 389–403, 2007.
4. Richter, A., "Estimation of radio channel parameters: Models and algorithms," Ph.D. Dissertation, Technische Universität Ilmenau, Ilmenau, Germany, 2005.
5. Salmi, J., A. Richter, and V. Koivunen, "Detection and tracking of MIMO propagation path parameters using state-space approach," *IEEE Trans. on Signal Processing*, Vol. 57, No. 4, 1538–1550, 2009.
6. Lemma, A. N., A.-J. van der Veen, and E. F. Deprettere, "Analysis of joint angle-frequency estimation using ESPRIT," *IEEE Trans. on Signal Processing*, Vol. 51, No. 5, 1264–1283, 2003.
7. Roy, R. and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol. 37, No. 7, 984–995, 1989.
8. Van der Veen, A.-J., M. C. Vanderveen, and A. J. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. on Signal Processing*, Vol. 46, No. 2, 405–418, 1998.
9. Jaffer, A. G., "Maximum likelihood direction finding of stochastic sources: A separable solution," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2893–2896, New York, USA, Apr. 11–14, 1988.

10. Zhang, X., G. Feng, and D. Xu, "Blind direction of angle and time delay estimation algorithm for uniform linear array employing multi-invariance MUSIC," *Progress In Electromagnetics Research Letters*, Vol. 13, 11–20, 2010.
11. Stoica, P. and A. Nehorai, "Performance study of conditional and unconditional direction of arrival estimation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 38, No. 10, 1783–1795, Oct. 1990.
12. Fleury, B. H., M. Tschudin, R. Heddergott, D. Dahlhaus, and K. Ingeman Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE Journal on Selected Areas in Communications*, Vol. 17, No. 3, 434–450, 1999.
13. Landmann, M., "Limitations of experimental channel characterization," Ph.D. Dissertation, Technischen Universität Ilmenau, Ilmenau, Germany, Mar. 2008.
14. Thoma, R. S., D. Hampicke, A. Richter, G. Sommerkorn, A. Schneider, U. Trautwein, and W. Wirnitzer, "Identification of time-variant directional mobile radio channels," *IEEE Trans. on Instrumentation and Measurement*, Vol. 49, No. 2, 357–364, 2000.
15. Hampicke, D., A. Richter, A. Schneider, G. Sommerkorn, and R. S. Thoma, "Statistical analysis of time-variant directional mobile radio channels based on wide-band measurements," *Proc. European Wireless'99*, 273–278, Munich, Germany, Oct. 1999.
16. Schneider, C., et al., "Multi-user MIMO channel reference data for channel modelling and system evaluation from measurements," *Proc. Int. ITG Workshop on Smart Antennas (WSA 2009)*, Berlin, Germany, Feb. 16–19, 2009.
17. www.channelsounder.de.
18. Van Trees, H. L., *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, Wiley-Interscience, New York, 2002.
19. Golub, G. H. and C. F. Van Loan, *Matrix Computations*, 3rd edition, Johns Hopkins University Press, 1996.