Copyright©2005 IEEE. Reprinted from:


This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the Helsinki University of Technology’s products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this material, you agree to all provisions of the copyright laws protecting it.
Joint Angular- and Delay-Domain MIMO Propagation Parameter Estimation Using Approximate ML Method

Cássio B. Ribeiro, Member, IEEE, Andreas Richter, Member, IEEE, and Visa Koivunen, Member, IEEE

Abstract—In this paper, we derive an estimation method that jointly estimates the parameters of the concentrated propagation paths and the distributed scattering component that are frequently observed in multiple-input multiple-output (MIMO) channel sounding measurements. The joint angular-delay domain model leads to a correlation matrix with high dimensionality, which makes direct implementation of a maximum-likelihood (ML) estimator unfeasible. We derive low-complexity methods for computing approximate ML estimates that exploit the structure of the covariance matrices. We propose an iterative two-step procedure that alternates between the estimation of the parameters of the concentrated propagation paths and the parameters of the distributed scattering. For the distributed scattering, the estimator first optimizes the parameters describing their time-delay structure. Then, using the estimated time-delay parameters, the parameters of the angular distributions are optimized. We present simulation results and compare the estimated time-delay and angular distributions to the actual distributions, demonstrating that high-quality estimates are obtained. The large sample performance of the estimator is studied by establishing the Cramér–Rao lower bound (CRLB) and comparing it to the variances of the estimated parameters. The simulations show that the variance of the proposed estimation technique reaches the CRLB for relatively small sample size for most parameters, and no bias is observed.

Index Terms—Channel parameter estimation, maximum-likelihood estimation, multipath channels, radio propagation.

NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Normalized DFT matrix.</td>
</tr>
<tr>
<td>h</td>
<td>Spatial content of the diffuse scattering component (DSC).</td>
</tr>
<tr>
<td>T</td>
<td>Fisher information matrix.</td>
</tr>
<tr>
<td>M_f</td>
<td>Number of frequency samples.</td>
</tr>
<tr>
<td>M_o</td>
<td>(M_o = M_rM_TM_f).</td>
</tr>
<tr>
<td>M_r</td>
<td>Number of RX antennas.</td>
</tr>
<tr>
<td>M_s</td>
<td>Number of observations.</td>
</tr>
<tr>
<td>M_t</td>
<td>Number of TX antennas.</td>
</tr>
<tr>
<td>N_{sp}</td>
<td>Number of specular propagation paths.</td>
</tr>
<tr>
<td>P</td>
<td>Number of clusters.</td>
</tr>
<tr>
<td>s</td>
<td>One observation of the specular component.</td>
</tr>
<tr>
<td>(v(\theta_w))</td>
<td>Sampled version of the PDP in frequency domain.</td>
</tr>
<tr>
<td>(\mathbf{v}_y, \mathbf{v}_w, \mathbf{v}_h)</td>
<td>Eigenvectors of (\mathbf{C}_y, \mathbf{C}_w, \mathbf{C}_h), respectively.</td>
</tr>
<tr>
<td>(w)</td>
<td>One observation of the frequency content of the DSC.</td>
</tr>
<tr>
<td>(y(f))</td>
<td>Received signal in frequency domain.</td>
</tr>
<tr>
<td>(\alpha_t)</td>
<td>Maximum power.</td>
</tr>
<tr>
<td>(\beta_d)</td>
<td>Normalized coherence bandwidth of the DSC.</td>
</tr>
<tr>
<td>(\epsilon_p)</td>
<td>Mixture proportions of the DSC.</td>
</tr>
<tr>
<td>(\varphi_{TJ})</td>
<td>Receive azimuth angle of the (l)th specular path.</td>
</tr>
<tr>
<td>(\gamma_l)</td>
<td>Complex gain of specular component.</td>
</tr>
<tr>
<td>(\kappa_p)</td>
<td>Scattering about the symmetry center.</td>
</tr>
<tr>
<td>(\mathbf{A}_y, \mathbf{A}_w, \mathbf{A}_h)</td>
<td>Eigenvalues of (\mathbf{C}_y, \mathbf{C}_w, \mathbf{C}_h), respectively.</td>
</tr>
<tr>
<td>(\mu_p)</td>
<td>Symmetry center or “mean angle.”</td>
</tr>
<tr>
<td>(\theta_w)</td>
<td>Frequency-domain DSC parameters: (\theta_w = [\alpha_1, \beta_d, \gamma_l]).</td>
</tr>
<tr>
<td>(\Theta_h)</td>
<td>Angular-domain DSC parameters: (\Theta_h = [\mu_p, \kappa_p, \epsilon_p]), (p = 1, \ldots, P).</td>
</tr>
<tr>
<td>(\theta_{h,p})</td>
<td>Angular-domain DSC parameters for cluster (p): (\theta_{h,p} = [\mu_p, \kappa_p, \epsilon_p]).</td>
</tr>
<tr>
<td>(\theta_{wn})</td>
<td>Frequency-domain DSC parameters plus noise variance.</td>
</tr>
<tr>
<td>(\theta_{dn})</td>
<td>DSC parameters plus noise variance ([\theta_{wn}, \theta_{hn}]).</td>
</tr>
</tbody>
</table>

I. INTRODUCTION

The interest in the multidimensional structure of the mobile radio channel is growing rapidly. This is mainly due to the fact that future beyond 3G wireless systems will employ multiantenna transceivers in order to improve spectral efficiency and radio link quality. Consequently, realistic channel models that are verified by real-world measurement campaigns are...
needed especially for transceiver design and network planning purposes. Channel sounding and related propagation parameter estimation are key tasks in creating such channel models. In particular, the double-directional modeling of the radio channel has attracted a lot of interest because it gives a better physical insight into the wave propagation mechanism in real-world radio environments [1], and it has the ability to remove the measurement antenna influence from the channel observation. Moreover, developing and comparing the performance of various multiple-input multiple-output (MIMO) transceiver structures requires such advanced channel models.

In channel parameter estimation from channel sounding measurements, two types of models have mainly been employed so far. A widely used model approximates the observed radio channel with a large number of discrete (deterministic) waves (see, e.g., [2] and [3]). The other approach describes the radio channel as a zero-mean circular complex Gaussian process [4], [5]. The former model is well suited to describe dominant concentrated (specular) propagation paths, i.e., paths with a small delay and angular spread, each contributing significantly to the wave propagation between transmitter (TX) and receiver (RX). The latter model is well suited to describe distributed (diffuse) scattering (DDS). It has been shown in [6] that both propagation mechanisms contribute significantly to the wave propagation. In [3], a data model for channel parameter estimation is proposed, which combines these two models. It is shown that an estimator that accounts for both concentrated propagation paths and distributed scattering outperforms estimators that ignore either of the channel components. However, in [3], the estimator is derived assuming that the contribution of the distributed scattering is an independent and identically distributed (i.i.d.) process in the angular domain at TX and RX. It only accounts for the correlation of the distributed scattering in the time delay and the frequency domain and ignores the correlation in angular domain. So far, no channel parameter estimation results have been published that provide more information about the angular properties of DDS in radio channels. This is mainly due to the fact that no estimator has been developed that estimates also the spatial correlation of the DDS. The model and the algorithm derived in this paper fill this gap.

A maximum-likelihood estimator for scattered sources is proposed in [7], where the angular distribution is assumed to be Gaussian. The angular spread is assumed to be small so that the correlation between the antenna elements can be approximated by a Taylor series expansion. In [8], for small angular spreads, the authors propose a first-order Taylor expansion of the spatial signature of each source. This leads to the generalized array manifold (GAM) model, which can be used in conjunction with well-known subspace-based methods, such as MUSIC. This method provides a parametric model for an instantaneous realization of the fading channel, as shown in [9]. Again, by assuming small angular spreads, the authors in [10] show that it is possible to approximate a scattered source as a combination of two rays symmetrically located around the nominal direction. This approximation allows using computationally efficient algorithms such as ESPRIT and root-MUSIC. The resulting algorithms are called Spread ESPRIT, Spread root-MUSIC, and so on [10]. Other methods stemming from these can be found, e.g., in [11] and [12]. These methods are derived only for narrowband channels.

In this paper, we derive an approximate ML estimator for channel parameter estimation from channel sounding measurements, which also takes the angular distribution of the distributed scattering component into account. It estimates the parameters of the concentrated propagation paths and the distributed scattering component jointly. The power-delay profile of the scattering component is modeled using an exponential distribution, which is typically observed in measurement campaigns [1]. The angular power profile is modeled using a mixture of angular von Mises distributions. The mixture distributions are employed in order to estimate the propagation parameters in scattering environments with multiple clusters of scatterers with high fidelity [13]. The von Mises distribution is described in detail in [14].

We propose a two-step procedure that alternates between the estimation of the parameters of the concentrated propagation paths and the parameters of the distributed scattering. For the distributed scattering component, the time-delay domain parameters modeled using an exponential distribution are estimated first. The angular domain parameters are then improved by fixing the time-delay domain parameters. This way, the correlation of the distributed scattering in the time-delay and frequency domains is fully taken into account. A whitening transformation is applied to the data prior to the estimation of the specular component [15], thus decoupling the problem of estimating the specular and diffuse components. As a result, well-known estimation methods can be used to estimate the parameters of the specular components with low complexity, like the SAGE-based procedure in [2] and [16].

The joint angular-delay model for the distributed scattering leads to a correlation matrix with high dimensionality. This makes the use of a maximum-likelihood (ML) estimator using finite precision arithmetics and finite memory resources unfeasible. In this paper, computationally efficient methods for finding approximate ML estimates are derived. The structure of the covariance matrices is fully exploited in the process.

Moreover, we derive the Cramér–Rao lower bound (CRLB) for the problem. It gives the lower bound on the variance by any unbiased estimator. We show in simulation that the variance of the estimates is close to the bound for a relatively small number of iterations and small number of samples.

This paper is organized as follows: in Section II, we describe the channel model considered in this paper. In Section III, the alternating technique for parameter estimation is described. In Section IV, the Cramér–Rao lower bound is established. Finally, in Section V, we present some simulation results and compare the large sample performance of the estimation technique to the CRLB.

II. SIGNAL MODEL

Assuming a channel sounding arrangement with $M_r$ antennas at the receiver and $M_t$ antennas at the transmitter, the signal at the receiver in frequency-domain is given by

$$ y(f) = s(f) + n_d(f) + n(f) $$

(1)
where \( s(f) \) represents the specular components of the propagation paths, \( n_i(f) \) represents the diffuse scattering component, and \( \tilde{r}(f) \) represents the zero-mean complex Gaussian measurement noise. The sampled data are arranged in vector \( y(f) \) as

\[
y(f) = \text{vec} \left( \begin{bmatrix} y_{11}(f) & \cdots & y_{1M}(f) \\ \vdots & \ddots & \vdots \\ y_{M1}(f) & \cdots & y_{MM}(f) \end{bmatrix} \right)
\]

where the vec operator stacks the columns of a matrix. The specular components are modeled as

\[
s(f) = \sum_{l=1}^{N_{\text{SP}}} s_l(f)
\]

\[
= \sum_{l=1}^{N_{\text{SP}}} \gamma_l b(\varphi_{RL}, \varphi_{TL}) \exp(-j2\pi f \tau_l) u(f)
\]

(2)

where \( u(f) \) is the transmitted signal, \( \gamma_l \) is the complex gain, \( b(\varphi_{RL}, \varphi_{TL}) \) is the array response to receive azimuth angle \( \varphi_{RL} \) and transmit azimuth angle \( \varphi_{TL} \), and \( \tau_l \) is the normalized delay for path \( l = 1, \ldots, N_{\text{SP}} \). The array response is given as a function of the receive and transmit array responses, \( b(\varphi_{RL}) \) and \( b(\varphi_{TL}) \), respectively, as \( b(\varphi_{RL}, \varphi_{TL}) = b(\varphi_{RL}) \otimes b(\varphi_{TL}) \), where \( \otimes \) denotes the Kronecker product.

We assume that the excitation signal \( u(f) \) is a multicarrier spread-spectrum signal (MCSSS) [3], which is designed such that \( |u(f)| \) is constant over the bandwidth of interest. Therefore, a raw estimate of the channel observation can be determined without changing the statistics of the noise \( n(f) \), simply by dividing the received samples by the known excitation function \( u(f) \). Hence, (1) can be rewritten as

\[
y(f) = s(f) + n_d(f) + n(f).
\]

(3)

Let \( M_f \) be the number of observed frequency samples. We then define the \( M_o \times 1 \) vector \( y' \) as

\[
y' = \begin{bmatrix} y(0) \\ \vdots \\ y(M_f-1) \end{bmatrix} = \begin{bmatrix} s(0) + n_d(0) + n(0) \\ \vdots \\ s(M_f-1) + n_d(M_f-1) + n(M_f-1) \end{bmatrix} = \bar{s} + \bar{n}_d + \bar{n}
\]

(4)

where \( M_o = M_r M_t M_f \), and

\[
\bar{s} = [s^T(0) \cdots s^T(M_f-1)]^T \quad \text{and} \quad \bar{n}_d = [n_d^T(0) \cdots n_d^T(M_f-1)]^T \quad \text{and} \quad \bar{n} = [n^T(0) \cdots n^T(M_f-1)]^T.
\]

The channel sounding technique assumed in this paper is based on time-division multiplexing of each transmitter and receiver antenna. Such technique is used, for example, in the PropSound channel sounder used in [16]. This particular structure makes it possible to measure the channel between each transmit and receive antenna pair independently. For further discussion of radio channel sounding techniques, see [17]. Deterministic maximum-likelihood estimation techniques such as the SAGE-based method in [16] represent the received signal as a combination of a large number of discrete waves. Consequently, parameters from each wave must be estimated. Hence, the algorithms often have convergence problems, and the estimates contain artifacts due to local maxima in the likelihood function and high dimensionality of the parameter space.

The following assumptions are employed throughout this article:

a) the process \( n_d(f) \) is zero-mean complex temporally white circular Gaussian;

b) the channel can be treated as constant during the time it takes to measure one realization of the channel;

c) the additive noise \( \bar{n} \) is an i.i.d. zero-mean circular complex Gaussian process with known covariance matrix, \( C_n = \mathbb{E} [\bar{n}\bar{n}^H] = \sigma_n^2 \mathbb{I} \), and independent of \( n_d \) and \( \bar{s} \).

Assumption a) comes from the observation that a very large number of waves, having independent weights, reach the receiver from scattered sources. Therefore, the central limit theorem can be used in this case to show that the received data are Gaussian distributed. The covariance matrix of \( \bar{n}_d(f) \) is obtained by writing the DSC as a sum of a very large number \( (N) \) of specular-like components, each one modeled similarly to (2). Hence

\[
n_d = \frac{1}{\sqrt{N M_t}} \sum_{i=1}^{N} \alpha_i w_i \otimes h_i
\]

(8)

where \( \alpha_i \) is the complex gain, \( P_a = \mathbb{E} [|\alpha_i|^2] \), the vector \( h_i \) of dimension \( M_r M_t \times 1 \) represents the spatial content of the DSC and is a function of the array response, and the vector \( w_i \) of dimension \( M_f \times 1 \) represents the frequency-dependent content of the DSC. Assume also that \( \alpha_i, w_i, \) and \( h_i \) are zero-mean and independent \( \forall i \). The covariance matrix of \( n_d \) is given by

\[
\mathbb{E} [\bar{n}_d \bar{n}_d^H] = \frac{1}{N P_a} \sum_{i=1}^{N} \alpha_i^2 w_i \otimes h_i \mathbb{E} [w_i^H \otimes h_i^H] = \frac{1}{N P_a} \sum_{i=1}^{N} P_a \mathbb{E} [w_i^H \otimes h_i^H] = \mathbb{E} [w w^H] \otimes \mathbb{E} [h h^H]
\]

(9)

Based on the assumptions above, the PDF of the received signal \( y' \) is completely characterized by its mean, \( E[y'] = \bar{s} \), and its \( M_o \times M_o \) covariance matrix

\[
C_y = E[(y' - \bar{s}) (y' - \bar{s})^H] = E[(\bar{n}_d + \bar{n}) (\bar{n}_d + \bar{n})^H]
\]

\[
= E[\bar{w} \bar{w}^H] \otimes E[\bar{h} \bar{h}^H] + E[\bar{n} \bar{n}^H]
\]

\[
= C_w \otimes C_h + \sigma_n^2 \mathbb{I}
\]

(10)

where \( \mathbb{I} \) is the \( M_o \times M_o \) identity matrix, \( C_w = E[\bar{w} \bar{w}^H] \), and \( C_h = E[\bar{h} \bar{h}^H] \). Observe that our model is based on the assumption that the covariance matrix of the DSC can be factorized into
a Kronecker product \((10)\). It is not assumed that the covariance matrix of the complete MIMO channel observation can be factorized into a Kronecker product.

### A. Delay and Frequency-Domain Characterization

For the delay domain, we use the model in [3], which is based on the observation that the power delay profile (PDP) has an exponential decay over time and a base delay that is related to the distance between the transmitter and receiver. The PDF for infinite bandwidth is given by

\[
\psi(\tau) = E[|W(\tau)|^2] = \begin{cases} 
0, & \tau < \tau_d^f \\
\alpha_1/2, & \tau = \tau_d^f \\
\alpha_1 e^{-B_d(\tau-\tau_d^f)}, & \tau > \tau_d^f 
\end{cases} \tag{11}
\]

where \(B_d\) is the coherence bandwidth, \(\alpha_1\) denotes the maximum power, and \(\tau_d^f\) is the base delay.

The Fourier transform of (11), the correlation function of the channel in the frequency domain, is given by

\[
\psi(\Delta f) = \frac{\alpha_1}{\beta_d + j2\pi \Delta f} e^{-j2\pi \Delta f \tau_d^f} \tag{12}
\]

where \(\beta_d = B_d/B_m\) is the normalized coherence bandwidth, and \(B_m\) is the measurement bandwidth. Let us define the sampled version of the correlation function \(v(\theta_w)\), \(\theta_w = \{\alpha_3, \beta_3, \tau_3\}\), in frequency domain as

\[
v(\theta_w) = \frac{\alpha_3}{M_f} \left( e^{-j2\pi \tau_3 / \beta_d + j2\pi \tau_3 M_f} \cdots e^{-j2\pi (M_f-1) \tau_3 / \beta_d + j2\pi (M_f-1) \tau_3 M_f} \right) \tag{13}
\]

where \(\tau_d^f\) is the normalized base delay.

Since the process is stationary, the correlation between components at different frequencies is given by

\[
\Psi(f_1, f_2) = \psi(f_1 - f_2), \quad \tag{14}
\]

Hence, the covariance matrix of the diffuse scattering (assuming the received signal is spatially white) may be modeled as a Toeplitz matrix

\[
C_w = \text{toeplitz}(v(\theta_w)), \quad \text{toeplitz}(a, b^H) \quad \text{as its first column and } b^H \text{ as its first row, with } a_1 = b_1^H.
\]

### B. Angular-Domain Characterization

Using the extended Saleh–Valenzuela (SVA) channel model in [4], we can write \(E[h|h]^H\) as a function of the angular parameters. A similar model is obtained in [5] and [18] following a different approach that is related to the geometry of the distribution of scatterers. Assuming that for the diffuse scattering there is statistical independence between AOAs and AODs, we can use the modeling strategy in [4] to characterize the covariance matrix of the channel in angular domain as

\[
C_h = C_h^R \otimes C_h^T \tag{16}
\]

where \(C_h^R\) and \(C_h^T\) denote the covariance matrix at the receiver and transmitter side, respectively. From a parameter estimation point of view, the extension from the spatially uncorrelated to the correlated case is more complex than extension from single input multiple output (SIMO) to MIMO, as long as the Kronecker structure in (16) holds. Hence, we will limit the discussion to unidirectional estimation, but the results can be naturally extended to the double directional case. We also assume for simplicity that a uniform linear array (ULA) is used at the receiver. Then the correlation at the receiver side is given by

\[
C_{h_{m_1m_2}}(\theta_h) = \int_0^{\pi} \exp(b_{m_1m_2} \cos(\phi)) f(\phi, \theta_h) d\phi \tag{17}
\]

where \(f(\phi, \theta_h)\) is any angular PDF of \(\phi\), characterized by parameters \(\Theta_h\), \(b_{m_1m_2} = j2\pi d_{m_1m_2}\), and \(d_{m_1m_2}\) is the distance between elements \(m_1\) and \(m_2\) in the receive array. An angular PDF must at least satisfy \(f(\phi, \theta_h) = f(\phi + 2\pi k, \theta_h) \forall k \in \mathbb{Z}\), with \(\phi \in [\phi_0, \phi_0 + 2\pi]\), \(\phi_0 \in \mathbb{R}\). Hence, a Gaussian PDF that has an infinite support is not suitable. The von Mises distribution [14] defined in angular domain is more appropriate. It is defined as follows:

\[
f(\phi, \Theta_h) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\phi - \mu)) \tag{18}
\]

where \(\mu\) is the symmetry center or “mean angle” and \(I_0(\cdot)\) is the modified Bessel function of the first kind of order zero. The parameter \(\kappa\) is related to the variance of the von Mises distribution, i.e., how scattered about the symmetry center the data are. It can be chosen between 0 (isotropic scattering) and \(\infty\) (extremely concentrated). Fig. 1 illustrates the von Mises PDF for different values of \(\kappa\).

In channel measurements, multimodal angular PDFs are often found as a result of signals arriving from a number of different clusters. This can be modeled using a mixture of angular PDFs

\[
f(\phi, \Theta_h) = \sum_{p=1}^{P} \epsilon_p f_p(\phi, \theta_{h,p}) \tag{19}
\]

where \(P\) is the number of mixture components, \(\sum_{p=1}^{P} \epsilon_p = 1\), \(\epsilon_p\) are unknown mixture proportions, and \(f_p(\phi)\) is any valid angular PDF. With this definition of the angular PDF, the angular domain parameters are defined as \(\Theta_h = \{\theta_{h,1}, \ldots, \theta_{h,P}\}\), with \(\theta_{h,p} = \{\mu_p, \kappa_p, \epsilon_p\}, p = 1, \ldots, P\).
Using (18), the cross correlation in (17) may be written analytically as [5], [18]

\[
C_{l,m_1m_2}(\Theta_l) = \sum_{p=1}^{P} I_0 \left( \frac{k_p^2 + l_{m_1m_2}^2 + 2k_p b_{m_1m_2} \cos(\mu_I)}{I_0(k_p)} \right),
\]

(20)

III. PARAMETER ESTIMATION

Let us denote by \( Y_{m} \) the \( m \)th observation of \( Y \), \( m = 1, \ldots, M_r \). Assuming \( Y \) is circular complex Gaussian and that the realizations \( Y_{m} \) are i.i.d., we can write the log-likelihood function as

\[
L(Y_1, \ldots, Y_{M_r}) = -M_s \log \pi - M_s \log |C_{y}| - \sum_{m=1}^{M_r} (Y_m - s)^H C_{y}^{-1}(Y_m - s)
\]

(21)

where \( M_s \) is the number of observations. We will also assume that the noise is circular complex white Gaussian with variance \( \sigma_n^2 \).

A method that estimates \( C_w \) and the noise variance is proposed in [3] assuming the input signal to be spatially white, i.e., \( C_h = I \). The method exploits the Toeplitz structure of \( C_w \) for the computation of the ML estimates. This reduces the computational complexity by avoiding the direct computation of determinants and matrix inversions. Unfortunately, it is not possible to directly extend the method for joint estimation of \( C_w \), \( C_h \), and \( \sigma_n^2 \) since \( C_y \) in (10) is not Toeplitz in general. Also, direct optimization of the likelihood function using (10) is not feasible due to the high dimensionality of the matrices involved. In current sounding systems, typical values for \( M_f \) and \( M_r, M_t \) are in the range \( M_f = [100,2000] \), and \( M_r, M_t = [4 \ldots 64] \). But with the rapid development of the channel sounders, these values may grow. This leads to a dimension of \( C_y \) ranging from \( 400 \times 400 \) to \( 128 \times 128 \), or even higher.

We propose an estimation method that reduces the computational complexity by using the following iterative procedure.

1) Optimize for the parameters of the specular components such as azimuth and elevation angle of arrival/departure, time delay, doppler spread etc., using the previously estimated covariance matrix.

2) Remove the contribution of the specular components from data and optimize for the covariance matrix of the diffuse scattering components plus noise variance.

3) Repeat the procedure until convergence or a maximum number of iterations is reached.

Step 2) can be further decomposed into two steps:

2a) Optimize for the frequency-domain parameters and noise variance.

2b) Optimize for the angular-domain parameters, with \( C_w \) as calculated in the previous step.

This iterative procedure is illustrated in Fig. 2.

A key benefit of the proposed method is the separate optimization of specular and diffuse scattering components. This reduces the number of variables for each local optimization. A similar approach has been considered in [3], where the signal was assumed to be spatially white.

The further decomposition of Step 2) into Steps 2a) and 2b) is important due to the high dimensionality of the matrices involved. With this two-step procedure, it is possible to exploit the Toeplitz structure of \( C_w \) for the computation of the ML estimates. Also, the covariance matrix manipulated in Step 2b) is only \( C_h \), which is typically much smaller in dimension than \( C_w \).

A. Specular Component

Most algorithms for parameter estimation of specular components assume that any additive nosilke like components are white. Hence, they cannot be directly applied to the model used in this paper, unless the covariance matrix of \( n(f) + \hat{n}(f) \) is the identity matrix. In order to avoid this problem and allow the use of well-known low-complexity algorithms for parameter estimation of specular components, a prewhitening transform is applied to the data such that its covariance matrix becomes a constant times the identity matrix [15]. We define the prewhitening matrix \( U \), such that

\[
E[U^{-H}Y - U^{-H}s](U^{-H}Y - U^{-H}s)^H = U^{-H} (C_w \otimes C_h + \sigma_n^2 I) U^{-1} = I.
\]

(22)

Therefore, \( U^{-H}Y \) can be used to estimate the parameters of the specular-like propagation paths using any well-known algorithm, such as the SAGE-based procedure in [2] and [16].

The matrix \( U \) is any square root matrix of \( C_y \) such that \( C_y = U H U \). It can be obtained, e.g., by the Cholesky decomposition of \( C_y \), since the presence of additive noise guarantees that \( C_y \) is positive definite. Another possibility to calculate \( U \) is through the eigenvalue decomposition of \( C_y \). This approach proves to be useful in the next sections. Let \( V_y \) and \( \Lambda_y \) denote the eigenvectors and eigenvalues of \( C_y \), i.e.,

\[
C_y = V_y \Lambda_y V_y^H
\]

(23)

where we have used the fact that the eigenvectors of a Hermitian matrix are unitary. Similarly, we can also define \( V_w \), \( \Lambda_w \), \( V_h \), and \( \Lambda_h \), such that \( C_w = V_w \Lambda_w V_w^H \) and \( C_h = V_h \Lambda_h V_h^H \). We can then write

\[
C_y = (V_w \Lambda_w V_w^H) \otimes (V_h \Lambda_h V_h^H) + \sigma_n^2 I = (V_w \otimes V_h) (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I) (V_w^H \otimes V_h^H).
\]

(24)
Comparing (24) with (23), we conclude that
\[
V_y = V_w \otimes V_h
\]
\[
\Lambda_y = \Lambda_w \otimes \Lambda_h + \sigma_n^2 I
\]
(25) (26)

Hence, \( U^{-H} \) can be defined as
\[
U^{-H} = (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I)^{-1/2} (V_w^H \otimes V_h^H).
\]
(27)
The product \( U^{-H} \) can be computed efficiently as
\[
U^{-H} y = \text{diag} \left( (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I)^{-1/2} \right) \circ \text{vec} \left[ (V_w^H \otimes V_h^H) \text{mat}(\gamma_m, M_f, M_t, M_f)^T \right]
\]
(28)
where the mat operator reshapes a vector into a matrix, as follows:
\[
\text{mat} \left( \begin{array}{c}
v_1^m \\
\vdots \\
v_N^m \\
\end{array} \right), M, N \right) = \left[ v_1^m \cdots v_N^m \right]
\]
(29)
the vec operator stacks the columns of a matrix, and \( \circ \) denotes elementwise multiplication, i.e., the Hadamard product.

The benefits of this implementation over the Cholesky decomposition is that the eigenvalues and eigenvectors of \( C_y \) and \( C_h \) can be used to simplify the estimator of the diffuse scattering component. Also, it will be shown in the next section that the computation of \( V_w^H \) and its multiplication by a vector can be done in a very efficient way. The Kronecker product in (28) has to be computed only for the diagonal elements, and the multiplication of a diagonal matrix and a vector has a complexity of \( O(M_0) \).

### B. Frequency-Domain Parameters

In the sequel, we will assume that the specular components have been estimated and removed from the observation, and hence the likelihood function is given by
\[
L(\gamma_1, \ldots, \gamma_{M_s}) = -M_s \log \pi - M_s \log |C_y| - \sum_{m=1}^{M_s} \gamma_m^H C_y^{-1} \gamma_m.
\]
(30)

An estimator for the frequency-domain parameters is derived in [3]. It is assumed that the channel covariance matrix has the structure \( C_{y} = (C_w + \sigma_n^2 I) \otimes I \) i.e., the channel is assumed to be spatially white. This is a special case of the situation expressed in (20), with \( P = 1 \) and \( \kappa = 0 \). Since we do not assume whiteness in the spatial domain, the method in [3] is not directly applicable here.

Using the decomposition of the estimation procedure outlined in Section III already allows some simplification in the optimization. This is due to lower dimensional searches. However, we still have not solved the problem of calculating the determinant and inverse of \( C_y \) at every iteration. These computations can be simplified by writing \( C_y \) as a function of its eigenvalues and eigenvectors, as in Section III-A.

We can exploit the Kronecker structure of the eigenvalues and eigenvectors of \( C_y \) in order to simplify the optimization procedure. This allows us to compute only the eigenvalues and eigenvectors of \( C_w \) and \( C_h \), and then obtain \( V_y \) and \( \Lambda_y \). The logarithm of the determinant of \( C_y \) can now be calculated as
\[
\log |C_y| = \log (|C_w \otimes C_h| (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I) |V_w^H \otimes V_h^H|)
\]
\[
= \log \prod_{j=1}^{M_0} (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I)_{(j,j)}
\]
\[
= \sum_{j=1}^{M_0} \log (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I)_{(j,j)}
\]
(31)
where \( \Lambda_w \) and \( \Lambda_h \) are vectors containing the eigenvalues of \( C_w \) and \( C_h \), respectively. \( 1_{M_0} \) is a \( M_0 \times 1 \) vector whose entries are equal to 1, and \( (\cdot)_{(j,j)} \) denotes the \( j \)th element of \( (\cdot) \). It is clear that the computational complexity of calculating the determinant is reduced. Another important observation is that the exchange of the order in which the log is computed allows for easier implementation with finite precision. This is due to the fact that the eigenvalues \( \Lambda_w \) can have a large spread, since they are an approximation to the PDP in (11).

The computation of \( C_y^{-1} \) can also be simplified using
\[
C_y^{-1} = \left( (V_w \otimes V_h) (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I) (V_w^H \otimes V_h^H) \right)^{-1}
\]
\[
= (V_w \otimes V_h) (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I)^{-1} (V_w^H \otimes V_h^H)
\]
(32)
Further simplifications are possible if we take into account that \( C_h \) is fixed while optimizing for \( C_y \). Let us rewrite the likelihood function in (30) as a function of \( V_w, \Lambda_w, V_h, \Lambda_h, \) and \( \sigma_n^2 \)
\[
L(\gamma_1, \ldots, \gamma_{M_s}) \propto - \sum_{j=1}^{M_f} \log (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I)_{(j,j)} - \frac{1}{M_0} \sum_{m=1}^{M_s} \gamma_m^H (V_w \otimes V_h) \gamma_m
\]
\[
\times (\Lambda_w \otimes \Lambda_h + \sigma_n^2 I)^{-1} (V_w^H \otimes V_h^H) \gamma_m
\]
(33)
Clearly, it is not necessary to calculate \( V_w^H \otimes V_h^H \) and the multiplication of this \( M_f \times M_0 \) matrix by \( \gamma_m \) at every iteration of (30), since \( V_h \) and \( \gamma_m \) are fixed. In order to simplify the problem, we will define the transformed signal
\[
\tilde{\gamma}_m = \left( I_{M_f} \otimes \Lambda_h^{-1/2} V_h \right) \gamma_m
\]
(34)
where it is assumed that \( C_h \) is nonsingular. An efficient form for computation of \( \tilde{\gamma}_m \) is given by
\[
\tilde{\gamma}_m = \text{vec} \left( \Lambda_h^{-1/2} V_h \text{mat}(\tilde{\gamma}_m, M_f, M_f, M_f) \right)
\]
(35)
The covariance matrix of \( \tilde{\gamma}_m \) is given by
\[
C_{\tilde{\gamma}} = \left( I_{M_f} \otimes \Lambda_h^{-1/2} V_h \right) E[\gamma_m^H] \left( I_{M_f} \otimes V_h \Lambda_h^{-1/2} \right)
\]
\[
= \left( I_{M_f} \otimes \Lambda_h^{-1/2} V_h \right) \left( I_{M_f} \otimes V_h \Lambda_h^{-1/2} \right)
\times \left( V_h^H \otimes V_h^H \right) \left( I_{M_f} \otimes V_h \Lambda_h^{-1/2} \right)
\]
\[
= (V_w \otimes I_{M_f, M_t}) (\Lambda_w \otimes I_{M_f, M_t} + \sigma_n^2 I_{M_f} \otimes \Lambda_h^{-1})
\times \left( V_h^H \otimes I_{M_f, M_t} \right)
\]
(36)
The assumption that \( C_h \) is nonsingular implies that \( C_{\tilde{\gamma}} \) is positive definite. Even though the covariance matrix \( C_{\tilde{\gamma}} \) is not
block diagonal, it is clear that the elements of $\mathbf{\tilde{Y}}_m$ can be reorganized such that the resulting covariance matrix is block diagonal. This implies that blocks of $M_f$ elements of $\mathbf{\tilde{Y}}_m$ are uncorrelated. Furthermore, there is no Kronecker product among the eigenvalues of $\mathbf{C}_w$ and $\mathbf{C}_h$, which simplifies numerical implementation. Hence, we will estimate the angular parameters using $\mathbf{\tilde{Y}}_m$ instead of $\mathbf{Y}_m$. Let us define the $M_o \times M_o$ diagonal matrix $\mathbf{\tilde{\Lambda}}$ as
\[
\mathbf{\tilde{\Lambda}} = (\mathbf{\Lambda}_w \otimes \mathbf{I}_{M_f M_t} + \sigma_n^2 M_f \otimes \mathbf{\Lambda}_h^{-1})
\]
and the $M_o \times 1$ vector $\mathbf{\tilde{\lambda}} = \text{diag}(\mathbf{\tilde{\Lambda}})$. Now we can rewrite the likelihood function in (33) as
\[
L(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{M_o}) \propto -\sum_{j=1}^{M_o} \log \tilde{\lambda}_j - \frac{1}{M_o} \sum_{m=1}^{M_o} \mathbf{\tilde{Y}}_m^H (\mathbf{\tilde{\Lambda}}_w \otimes \mathbf{I}_{M_f M_t})^{-1} \mathbf{\tilde{\Lambda}} \mathbf{\tilde{Y}}_m.
\]
Let us define the $M_o \times 1$ vector $\mathbf{\tilde{X}}_m$ as
\[
\mathbf{\tilde{X}}_m = (\mathbf{\tilde{V}}_w^H \otimes \mathbf{I}_{M_f M_t}) \mathbf{\tilde{Y}}_m.
\]
An efficient form for computing $\mathbf{\tilde{X}}_m$ is given by
\[
\mathbf{\tilde{X}}_m = \text{vec} \left( (\mathbf{\tilde{V}}_w^H \otimes \mathbf{I}_{M_f M_t}) \mathbf{\tilde{Y}}_m \right)^T.
\]
Finally, we can write $L(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{M_o})$ in a computationally efficient form as
\[
L(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{M_o}) \propto -\sum_{j=1}^{M_o} \log \tilde{\lambda}_j - \frac{1}{M_o} \sum_{m=1}^{M_o} \mathbf{\tilde{X}}_m^H \mathbf{\tilde{\Lambda}}^{-1} \mathbf{\tilde{X}}_m
\]
where it should be noted that the multiplication between a diagonal matrix and a vector is elementwise multiplication of the main diagonal and the vector. The ML estimates of $\mathbf{\tilde{\theta}}_{w,n} = \{\tilde{\theta}_{w,n,1}, \ldots, \tilde{\theta}_{w,n,M_f}\}$ maximize the likelihood function in (41).

Even further reduction of complexity is possible in the calculation of the eigenvalues and eigenvectors of $\mathbf{R}_w$ since it is a large Toeplitz matrix. Consequently, it can be approximated by a circulant matrix [19], [20]. A circulant matrix can be diagonalized as
\[
\mathbf{R}_w = \mathbf{F} \mathbf{D}_w \mathbf{F}^H
\]
where $\mathbf{F}$ is the unitary discrete Fourier transform (DFT) matrix and $\mathbf{D}_w$ is a diagonal matrix with the eigenvalues of $\mathbf{R}_w$. Hence, $\mathbf{V}_w = \mathbf{F}$, and $\mathbf{A}_w = \mathbf{F}^H \mathbf{R}_w \mathbf{F}$. In case $\mathbf{R}_w$ is a power of 2, the computational complexity of $\mathbf{A}_w$ and $\mathbf{\tilde{Y}}_m$ in (35) can be reduced even further by using fast Fourier transform (FFT). For other values of $M_f$, other algorithms that optimize the computation of the DFT can be used, like the Goertzel algorithm. There is no need to compute the off-diagonal elements of $\mathbf{F}^H \mathbf{R}_w \mathbf{F}$, and we use the following computationally efficient mapping from $\mathbf{v}(\mathbf{\theta}_w)$ to $\mathbf{\lambda}_w$ [3]:
\[
\mathbf{\lambda}_w = T(\mathbf{v}(\mathbf{\theta}_w)) = \frac{1}{\sqrt{M}} \mathbf{F}^H (\mathbf{W}_1 \mathbf{v}(\mathbf{\theta}_w) + \mathbf{W}_2 \mathbf{v}^*(\mathbf{\theta}_w)).
\]
where
\[
\mathbf{W}_1 = \text{diag}([M_f M_f - 1 \cdots 1])
\]
and
\[
\mathbf{W}_2 = \begin{bmatrix}
0 & \cdots & 0 \\
\vdots & \ddots & 1 \\
0 & 2 & 0 \\
M_f - 1 & \cdots & \cdots & 0
\end{bmatrix}.
\]

C. Angular-Domain Parameters

In Section II-B, we modeled the covariance matrix in angular domain, $\mathbf{C}_h$, as a function of a mixture of von Mises distributions. Assuming the number of mixture components in angular domain $P$ is known or reliably estimated, the angular parameters are the parameters of the mixture of von Mises distributions: $\mathbf{\theta}_h = \{\mu_1, \kappa_1, \epsilon_1, \ldots, \mu_P, \kappa_P, \epsilon_P\}$, $p = 1, \ldots, P$, with $\sum_{p=1}^{P} \epsilon_p = 1$. Due to the model in (10), the path loss is already estimated as part of the delay-domain parameters.

For the estimation of the angular domain parameters, we assume that the frequency-domain parameters are fixed. The computational complexity can be simplified as in Section III-B. Assuming $\mathbf{C}_w$ is nonsingular, let us define the transformed signal as
\[
\mathbf{\tilde{Y}}_m = (\mathbf{\Lambda}_w^{-1/2} \mathbf{F}^H \otimes \mathbf{I}_{M_f M_t}) \mathbf{Y}_m.
\]
An efficient form for computation of $\mathbf{\tilde{Y}}_m$ is given by
\[
\mathbf{\tilde{Y}}_m = \text{vec} \left( (\mathbf{\Lambda}_w^{-1/2} \mathbf{F}^H \otimes \mathbf{I}_{M_f M_t}) \mathbf{Y}_m \right)^T.
\]
The covariance matrix of $\mathbf{\tilde{Y}}_m$ is given by
\[
\mathbf{\tilde{C}}_m = (\mathbf{\Lambda}_w^{-1/2} \mathbf{F}^H \otimes \mathbf{I}_{M_f M_t}) \mathbf{E}(\mathbf{\tilde{Y}}_m \mathbf{\tilde{Y}}_m^H) (\mathbf{\Lambda}_w^{-1/2} \mathbf{F} \otimes \mathbf{I}_{M_f M_t})
\]
\[
= (\mathbf{I}_{M_f} \otimes \mathbf{V}_h) \left( \mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{\Lambda}_w^{-1} \otimes \mathbf{I}_{M_f M_t} \right) \mathbf{\Lambda}_w^{-1/2} \mathbf{F} \otimes \mathbf{V}_h^H.
\]
The assumption that $\mathbf{C}_w$ is nonsingular implies that $\mathbf{\tilde{C}}_m$ is positive definite. The covariance matrix $\mathbf{\tilde{C}}_m$ is block diagonal, implying that blocks of $M_f$ elements of $\mathbf{\tilde{Y}}_m$ are uncorrelated. Also, there is no Kronecker product between the eigenvalues of $\mathbf{C}_w$ and $\mathbf{C}_h$. Consequently, we will estimate the angular parameters using $\mathbf{\tilde{Y}}_m$ instead of $\mathbf{\tilde{Y}}_m$. The likelihood function for $\mathbf{\tilde{Y}}_m$ is given by
\[
L(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{M_o}) \propto -\sum_{j=1}^{M_o} \log \tilde{\lambda}_j - \frac{1}{M_o} \sum_{m=1}^{M_o} \mathbf{\tilde{Y}}_m^H \mathbf{\tilde{C}}_m^{-1} \mathbf{\tilde{Y}}_m.
\]
Let us define the $M_o \times M_o$ diagonal matrix $\mathbf{\tilde{\Lambda}}$ as
\[
\mathbf{\tilde{\Lambda}} = (\mathbf{I}_{M_f} \otimes \mathbf{\Lambda}_h + \sigma_n^2 \mathbf{\Lambda}_w^{-1} \otimes \mathbf{I}_{M_f M_t})
\]
and the $M_o \times 1$ vector $\mathbf{\tilde{\lambda}} = \text{diag}(\mathbf{\tilde{\Lambda}})$. Now we can write the likelihood function in (49) as
\[
L(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{M_o}) \propto -\sum_{j=1}^{M_o} \log \tilde{\lambda}_j
\]
\[
- \frac{1}{M_o} \sum_{m=1}^{M_o} \mathbf{\tilde{Y}}_m^H (\mathbf{I}_{M_f} \otimes \mathbf{V}_h) \mathbf{\tilde{\Lambda}}^{-1} \mathbf{\tilde{Y}}_m.
\]
Furthermore, let us define the $M_o \times 1$ vector $\hat{x}_m$ as
\[
\hat{x}_m = (I_{M_f} \otimes V_H^H) \hat{y}_m.
\] (52)

An efficient form for computation of $\hat{x}_m$ is given by
\[
\hat{x}_m = \text{vec} \left( V_H^H \text{mat}(\hat{y}_m, M_r, M_f, M_f) \right).
\] (53)

Finally, we can write $L(\hat{y}_1, \ldots, \hat{y}_{M_s})$ in a computationally efficient form as
\[
L(\hat{y}_1, \ldots, \hat{y}_{M_s}) \propto -\sum_{j=1}^{M_o} \log \hat{\lambda}_j - \frac{1}{M_s} \sum_{m=1}^{M_s} \hat{x}_m^H \hat{x}_m
\] (54)

where it should be noted again that the multiplication between a diagonal matrix and a vector is elementwise multiplication of the main diagonal and the vector. The estimates of $\theta_{\text{hp}p} = \{\nu_{\text{hp}}, \mu_{\text{hp}}, \epsilon_{p}\}$, $p = 1, \ldots, P$, are those values that maximize the cost function in (54).

The optimization of the nonlinear functions in (41) and (54) can be performed using the Levenberg–Marquardt algorithm [21]. It requires the computation of the gradient and the approximation of the Hessian, which can be found in Appendix I. Also, the derivates presented in Appendix I are used for the derivation of the performance bounds in Section IV. Efficient implementations of the gradient and the approximate Hessian can be found in Appendix II.

D. Initialization

In this section, we describe one approach for the initialization of the algorithm. The specular paths are initialized assuming
\[
C_y = I.
\]

Well-known estimators available in the literature can be used to get initial estimates for the parameters of the specular paths [2], [16].

For the diffuse scattering, the frequency-domain parameters are initialized as in [3], and assuming $C_h = I$.

The angular-domain parameters can be initialized using the following procedure.

1) Get initial estimate for $\nu_{\text{hp}}$, $p = 0, \ldots, P - 1$, using, e.g., ESPRIT assuming $P$ sources.
2) Choose initial values for $\mu_{\text{hp}}$, $p = 0, \ldots, P - 1$, uniformly drawn from the interval [0, 50], and $\epsilon_p$, $p = 0, \ldots, P - 1$, randomly in the range [0, 1].
3) Refine the initial estimates using the method in [13], [18], with $\mathcal{Y}$ normalized by $P_y$, that represents the total power in $\mathbf{w}(\tau)$
\[
P_y = \frac{\alpha}{\beta} e^{\alpha \beta \tau_d}.
\]

If an arbitrary 2-D or 3-D antenna element arrangement is used, beamforming or ES-root MUSIC [22] can be applied in Step 1.

E. Summary of the Algorithm

Table I summarizes the reduced complexity algorithm for estimation of frequency- and angular-domain parameters.

We evaluate the computational complexity of the algorithm in terms of real multiplications. We assume the multiplication of two complex numbers corresponds to four real multiplications, and the multiplication of a complex and a real number corresponds to two real multiplications. Fig. 3 shows the computational complexity as a function of the number of frequency samples, $M_f$, assuming 20 evaluations of the cost function during the optimization, and that three cycles of the estimation procedure are used. We assume that only the diffuse components are estimated. The complexity of calculating the eigendecomposition of $C_h$ is of order $O((M_s M_p)^3)$ [23]. The exact number of multiplications depends on the matrix structure, but the order $O((M_s M_p)^3)$ remains. Conditioning of the matrix plays a role as well.

Fig. 3 illustrates the complexities of different solutions. It is clear that the optimized method reduces the complexity by several orders of magnitude compared to the maximum-likelihood estimation, especially if FFTs are used when multiplying by $V_w$.
The effectiveness of this reduction depends on the number of cycles. A reasonable criterion for convergence is to stop the algorithm if after any two iterations the relative change for all parameters is less than a predefined threshold, e.g., $10^{-2}$. For this threshold value, the parameters usually converge in less than five iterations for all versions of the algorithm.

IV. PERFORMANCE BOUNDS

The $i, k$th element of the Fisher information matrix (FIM) for a circular complex white Gaussian variable are given by [15]

$$I_{ik} = M_s 2R \left\{ \left( \frac{\partial h_i}{\partial \theta_k} \right)^T C_y^{-1} \left( \frac{\partial h_i}{\partial \theta_k} \right) + M_s tr \left\{ C_y^{-1} \left( \frac{\partial C_y}{\partial \theta_i} \right) C_y^{-1} \left( \frac{\partial C_y}{\partial \theta_k} \right) \right\} \right\}. \quad (55)$$

Hence, the FIM is block diagonal if $\theta_{sp}$ and $\theta_{dn}$ contain uncorrelated parameters, i.e.,

$$I = \begin{bmatrix} I(\theta_{sp}) & 0 \\ 0 & I(\theta_{dn}) \end{bmatrix} \quad (56)$$

where $I(\theta_{sp})$ is the FIM for the specular components, $\theta_{dn} = \{ \theta_{wn}, \Theta_{h} \}$, and $I(\theta_{dn})$ is the FIM for the diffuse scattering parameters plus noise variance. As a consequence, its inverse is also block diagonal [23]. This means that the parameters of the specular and diffuse components are asymptotically decoupled, and the respective CRLBs can be derived independently. The CRLB for the parameters of the specular components can be found in [2], [3]. The CRLB for the parameters of the diffuse scattering components is given by (55) as

$$I_{ik}(\theta_{dn}) = M_s tr \left\{ C_y^{-1} \left( \frac{\partial C_y}{\partial \theta_i} \right) C_y^{-1} \left( \frac{\partial C_y}{\partial \theta_k} \right) \right\}. \quad (57)$$

Let us define the matrices containing the partial derivatives with respect to all $N_p$ elements of the parameter vector $\theta_{dn}$, as (58) and (59), shown at the bottom of the page. Using (58) and (59), the FIM can be expressed in a compact form as

$$I = M_s D_1^T D_2, \quad (60)$$

The partial derivatives of $C_y$ with respect to the diffuse scattering component parameters are given by

$$\frac{\partial C_y}{\partial \theta_i} = C_w \otimes \frac{\partial C_h}{\partial \theta_i}, \quad \theta_i \in \Theta_h \quad (61)$$

$$\frac{\partial C_y}{\partial \sigma^2} = I \quad (62)$$

The partial derivatives of $C_w$ and $C_h$ with respect to the propagation parameters can be found in Appendix I.

V. SIMULATION RESULTS

In this section, simulation examples are presented in order to illustrate the performance of the described parameter estimation procedure. The receiver is equipped with a ULA having $M_t = 8$ antennas and the transmitter uses $M_t = 1$ antenna. The number of frequency points is $M_f = 128$, and the number of channel realizations is $M_s = 5$. For the frequency-domain parameters, typical values often observed in channel sounding experiments are used: $\sigma_n^2 = 0.1$, $\alpha_1 = 1$, $\beta = 0.07$, and $\tau = 0.1$. The angular-domain parameters are defined as $\phi = \{ 60^\circ, 120^\circ \}$, $\kappa = \{ 10, 50 \}$, $\epsilon = \{ 0.4, 0.6 \}$, corresponding to two clusters in the angular domain. One specular component is assumed to be present, and it is modeled as

$$s(k) = \gamma b(\varphi_R) \exp(-j2\pi k\tau) \quad (64)$$

where $\gamma$ is the complex gain, $b(\varphi_R)$ is the steering vector for receive azimuth angle $\varphi_R$, and $\tau$ is the normalized delay. For the simulation, the values are set as $\gamma = 0.8 e^{j\pi/5}$, $\varphi_R = 80^\circ$, and $\tau = 0.12$.

The received signal is generated as

$$y(k) = s(k) + R^{1/2} n_2(k) + n(k) \quad (65)$$

where $n_2(k)$ is a circular complex white Gaussian process and $R^{1/2}$ is obtained by the Cholesky decomposition of $C_w \otimes C_h$. This implies that the covariance matrix of $R^{1/2} n_2(k)$ is given by $C_w \otimes C_h$. The vector $n(k)$ is a circular complex white Gaussian process representing the measurement noise.

The iterative procedure described in Section III is repeated five times, starting with the estimation of the specular component and proceeding as illustrated in Fig. 2. The model order, i.e., number of clusters and specular paths, is assumed known. The frequency-domain parameters are computed using the approximation of a Toeplitz matrix as a circulant matrix described in Section III-B. The parameters are initialized as described in Section III-D. In Figs. 4 and 5, we compare the power delay profile (PDP) and power angular profile (PAP) obtained using the estimation procedure described in this article with the actual PDP and PAP, respectively. The curves overlap almost perfectly. In the example, $M_s \ll M_t$, i.e., the full sample covariance matrix is rank deficient. Still, the estimator is able to provide high-precision estimates for the time-delay distribution, angular distribution, and specular component. The PAP is compared to the output of the Bartlett beamformer, showing the gain in using the combined procedure to estimate both signal components iteratively. The beamformer is only able to estimate the angle of the specular component, but it does not provide any useful information about the diffuse scattering component.
The specular paths are not estimated individually, while the diffuse component corresponds to the mixture of von Mises distributions. Hence, for the simulated values of $\theta$, we observe that all parameters of the diffuse component are the same as in the previous simulation. The angular spread is shown in degrees using the mapping $X = \psi$. The curves overlap almost perfectly. Also shown is the output of the Bartlett beamformer. The specular component is identified as a sharp peak at $\psi = 80^\circ$, while the diffuse component corresponds to the mixture of von Mises distributions.

The algorithms that approximate the mean angle and angular variance using two separate paths around the mean are called Spread F [10], where F denotes the underlying algorithm to estimate the paths. Table II shows the estimates obtained using the proposed method and Spread ESPRIT [10]. The results are an average over 300 runs. The parameters of the DSC and specular components are the same as in the previous simulation. The angular spread is shown in degrees using the mapping $X = \psi$. The curves overlap almost perfectly. Also shown is the output of the Bartlett beamformer. The specular component is identified as a sharp peak at $\psi = 80^\circ$, while the diffuse component corresponds to the mixture of von Mises distributions.

In Figs. 6–10, we compare the mean-square error (MSE) of the estimates after two cycles with the CRLB as a function of the number of channel realizations, $M_n$. The angular-domain parameters are set to $\phi = \{50^\circ, 100^\circ\}$, $\kappa = \{5, 150\}$, $\epsilon = \{0.4, 0.6\}$, corresponding to two clusters in the angular domain. The frequency-domain parameters remain unchanged. No specular components are present. It can be observed that all parameters converge close to the CRLB for a relatively small sample size. The exception is the relative delay $\tau_n$, which presents a noticeable gap with respect to its CRLB. However, no bias is observed, since the curve is parallel to the CRLB.

In order to verify the robustness of the algorithm, we apply the estimator to data that do not follow exactly the assumptions used for its derivation. We generate the data as a sum of individual propagation paths

$$\mathbf{y}(k) = \sum_{l=1}^{L} \mathbf{s}_l(k) + \mathbf{n}(k)$$

where $L$ is the number of individual paths, and $\mathbf{s}_l(k)$ are defined as in (2). For this simulation, we use $L = 24$, the noise variance is 0.01, and $M_n = 4$. Convergence is achieved after four iterations, assuming the parameters converge after changing by less than $5 \times 10^{-5}$. The specular paths are not estimated individually, but rather the joint angle-delay distribution of diffuse scattering

<table>
<thead>
<tr>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\sigma_{\mu_1}$</th>
<th>$\sigma_{\mu_2}$</th>
<th>$\sigma_\psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$60^\circ$</td>
<td>$120^\circ$</td>
<td>$18.12^\circ$</td>
<td>$8.10^\circ$</td>
<td>$80^\circ$</td>
</tr>
<tr>
<td>$59.98^\circ$</td>
<td>$119.99^\circ$</td>
<td>$18.14^\circ$</td>
<td>$8.07^\circ$</td>
<td>$79.99^\circ$</td>
</tr>
<tr>
<td>$56.66^\circ$</td>
<td>$119.15^\circ$</td>
<td>$9.20^\circ$</td>
<td>$3.09^\circ$</td>
<td>$79.97^\circ$</td>
</tr>
</tbody>
</table>

The curves overlap almost perfectly. The specular component is identified as a sharp peak at $\psi = 80^\circ$, while the diffuse component corresponds to the mixture of von Mises distributions.

**Fig. 4.** Comparison of estimated power delay profile and actual power delay profile. The curves overlap almost perfectly. The specular component is identified as a sharp peak at $\tau_n = 0.12$, while the diffuse component corresponds to the exponential curve.

**Fig. 5.** Comparison of estimated power angular profile and actual power angular profile. The curves overlap almost perfectly. Also shown is the output of the Bartlett beamformer. The specular component is identified as a sharp peak at $\psi = 80^\circ$, while the diffuse component corresponds to the mixture of von Mises distributions.

**Fig. 6.** CRLB of frequency-domain parameters $\{\alpha_1, \beta_\alpha\}$ and noise variance as a function of the number of channel realizations, $M_n$. The MSE after two iterations of the estimation procedure is shown for comparison.

**Table II**

<table>
<thead>
<tr>
<th>Correct values</th>
<th>Proposed method</th>
<th>Spread ESPRIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>$60^\circ$</td>
<td>$59.98^\circ$</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>$120^\circ$</td>
<td>$119.99^\circ$</td>
</tr>
<tr>
<td>$\sigma_{\mu_1}$</td>
<td>$18.12^\circ$</td>
<td>$18.14^\circ$</td>
</tr>
<tr>
<td>$\sigma_{\mu_2}$</td>
<td>$8.10^\circ$</td>
<td>$8.07^\circ$</td>
</tr>
<tr>
<td>$\sigma_\psi$</td>
<td>$80^\circ$</td>
<td>$79.99^\circ$</td>
</tr>
<tr>
<td>$\beta_\alpha$</td>
<td>$56.66^\circ$</td>
<td>$119.15^\circ$</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$9.20^\circ$</td>
<td>$3.09^\circ$</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$79.97^\circ$</td>
<td>$79.97^\circ$</td>
</tr>
</tbody>
</table>
VI. CONCLUSION

In this paper, we derived a joint estimator for the parameters of the concentrated propagation paths and the distributed scattering component that are frequently found in MIMO radio propagation measurements. In particular, we estimate the parameters of the scattering component in both spatial and temporal domains. The power-delay profile of the scattering component is modeled using an exponential distribution, which is typically observed in measurement campaigns. The power angular profile is modeled using a mixture of angular von Mises distributions. The simulation results show that this procedure converges to the estimates of both specular and diffuse components with high fidelity. Convergence is achieved with few iterations.

We derived computationally efficient methods for finding the approximate ML estimates. The structure of the covariance matrices is fully exploited. Complexity studies show that the reduction in the number of real multiplications is approximately three to five orders of magnitude. We also present computationally efficient methods to compute the gradients and Hessians, which are useful for the implementation of the optimization routines.

Furthermore, we derived the CRLB for the problem and showed that the variance of the estimates converges close to the bound for a relatively small number of cycles of the estimation procedure and small number of channel realizations (small sample size). For some parameters the CRLB is not attained, but no error floor indicating bias is observed.
APPENDIX I
GRADIENT AND APPROXIMATE HESSIAN

The derivative of (38) with respect to the $i$th element of $\theta_w = \{\alpha_1, \beta_d, \tau_d\}$ is given by

$$\frac{\partial L(\hat{Y})}{\partial \theta_w}_i = -\text{tr} \left( C^{-1}_g \frac{\partial C_g}{\partial \theta_w}_i \right)$$

$$+ \frac{1}{M_s} \sum_{m=1}^{M_s} \hat{Y}_m^H C^{-1}_g \frac{\partial C_g}{\partial \theta_w}_i C^{-1}_g \hat{Y}_m$$

$$= -\text{tr} \left( C^{-1}_g \left( \frac{\partial C_w}{\partial \theta_w}_i \otimes I \right) \right)$$

$$+ \frac{1}{M_s} \sum_{m=1}^{M_s} \hat{Y}_m^H C^{-1}_g \left( \frac{\partial C_w}{\partial \theta_w}_i \otimes I \right) C^{-1}_g \hat{Y}_m. \quad (67)$$

The derivative of $C_w$ with respect to the frequency-domain parameters is given by

$$\frac{\partial C_w}{\partial \theta_w}_i = \text{toep} \left( \frac{\partial \theta(w)}{\partial \theta_w}_i, \frac{\partial \theta(w)}{\partial \theta_w}_i^H \right). \quad (68)$$

and $(\partial \theta(w)) / (\partial \theta(w)_i)$ is calculated from (13) as (69)–(71), shown at the bottom of the page. From (36), the derivative of $C_w$ with respect to the noise variance $\sigma_n^2$ is given by

$$\frac{\partial C_w}{\partial \sigma_n^2} = I \otimes \Lambda_h^{-1} \quad (70)$$

and hence (67) simplifies to

$$\frac{\partial L(\hat{Y})}{\partial \sigma_n^2} = -\text{tr} \left( C^{-1}_g \frac{\partial C_g}{\partial \sigma_n^2} \right)$$

$$+ \frac{1}{M_s} \sum_{m=1}^{M_s} \hat{Y}_m^H C^{-1}_g \frac{\partial C_g}{\partial \sigma_n^2} C^{-1}_g \hat{Y}_m$$

$$= -\text{tr} \left( C^{-1}_g \left( I \otimes \Lambda_h^{-1} \right) \right)$$

$$+ \frac{1}{M_s} \sum_{m=1}^{M_s} \hat{Y}_m^H C^{-1}_g \left( I \otimes \Lambda_h^{-1} \right) C^{-1}_g \hat{Y}_m.$$ \quad (73)

An approximation of the Hessian is given by the FIM, which gives the expected value of the Hessian. Using the fact that $\hat{Y}$ is a circular complex white Gaussian variable, we obtain the partial derivative with respect to the $i$th and $k$th elements of $\theta_{wn} = \{\alpha_1, \beta_d, \tau_d, \sigma_n^2\}$ as

$$E \left[ \frac{\partial^2 L(\hat{Y})}{\partial (\theta_{wn})_i \partial (\theta_{wn})_k} \right]$$

$$= M_s \text{tr} \left\{ C^{-1}_g \left( \frac{\partial C_y}{\partial (\theta_{wn})_i} \right)^T C^{-1}_g \left( \frac{\partial C_y}{\partial (\theta_{wn})_k} \right) \right\}. \quad (74)$$

Recalling that the trace of the product between two square matrices can be expressed as

$$\text{tr}(AB) = \text{vec}(A^T) \cdot \text{vec}(B) \quad (75)$$

we can write

$$E \left[ \frac{\partial^2 L(\hat{Y})}{\partial (\theta_{wn})_i \partial (\theta_{wn})_k} \right]$$

$$= M_s \text{vec} \left\{ \left( C^{-1}_g \left( \frac{\partial C_y}{\partial (\theta_{wn})_i} \right) \right)^T \right\}$$

$$\times \text{vec} \left\{ C^{-1}_g \left( \frac{\partial C_y}{\partial (\theta_{wn})_k} \right) \right\}. \quad (76)$$

Let us define the matrices $D_{1,wn}$ and $D_{2,wn}$ containing the partial derivatives with respect to all $N_w$ elements of the parameter

$$\frac{\partial \theta(w)}{\partial \alpha_1} = \frac{1}{M_f} \left[ \frac{1}{\beta_d} \cdots e^{-j2\pi(M_f-1)\tau_d} \right]$$

$$\frac{\partial \theta(w)}{\partial \beta_d} = -\frac{\alpha_1}{M_f} \left[ \frac{1}{(\beta_d)^2} \cdots e^{-j2\pi(M_f-1)\tau_d} \right]$$

and

$$\frac{\partial \theta(w)}{\partial \tau_d} = \frac{\alpha_3}{M_f} \left[ 0 \frac{-j2\pi e^{-j2\pi \tau_d}}{\beta_d + j2\pi \frac{M_f-1}{M_f}} \cdots \frac{-j2\pi(M_f-1)e^{-j2\pi(M_f-1)\tau_d}}{\beta_d + j2\pi \frac{M_f-1}{M_f}} \right]. \quad (69)$$
vector $\theta_n$ and the partial derivative with respect to the noise variance $\sigma_n^2$ as (77) and (78), shown at the bottom of the page. Hence, the approximation of the Hessian can be expressed in a compact form as

$$E \left[ \frac{\partial^2 L(\hat{y})}{\partial (\theta_{um}) \partial (\theta_{um})^T} \right] = M_s D_{1,um}^T D_{2,um}. \quad (79)$$

The derivative of (51) with respect to the angular-domain parameters is given by

$$\frac{\partial L(\hat{y})}{\partial (\theta_{h})_i} = -\text{tr} \left( C^{-1}_y \frac{\partial C_y}{\partial (\theta_{h})_i} \right) + \frac{1}{M_a} \sum_{m=1}^{M_a} \hat{y}_m^H C^{-1}_y \frac{\partial C_y}{\partial (\theta_{h})_i} C^{-1}_y \hat{y}_m$$

$$= -\text{tr} \left( C^{-1}_y \left( I \otimes \frac{\partial C_h}{\partial (\theta_{h})_i} \right) \right) + \frac{1}{M_a} \sum_{m=1}^{M_a} \hat{y}_m^H C^{-1}_y \left( I \otimes \frac{\partial C_h}{\partial (\theta_{h})_i} \right) C^{-1}_y \hat{y}_m. \quad (80)$$

The $(m_1,m_2)$th element of the derivative of $C_h$ with respect to the angular-domain parameters for the SVA model in (20) is given by

$$\frac{\partial C_h}{\partial \mu_p} = -\epsilon_p I_0^1(\kappa_p) I_1(\beta_p) \beta_p^{-1} \kappa_p b_{m_1 m_2} \sin \mu_p,$$

$$\frac{\partial C_h}{\partial \kappa_p} = \epsilon_p I_0^{-1}(\kappa_p) I_1(\beta_p) + I_0^{-1}(\kappa_p) I_1(\beta_p) \beta_p^{-1} (\kappa_p + b_{m_1 m_2} \cos \mu_p). \quad (81)$$

The condition $\sum_{p=1}^{P} \epsilon_p = 1$ can be rewritten as $\epsilon_p = 1 - \sum_{p=1}^{P} \epsilon_p$, and then we can calculate the derivatives of $C_y(\theta)$ with respect to the mixture proportions $\epsilon_p, p = 1, \ldots, P - 1$, as

$$\frac{\partial C_h}{\partial \epsilon_p} = I_0^{-1}(\kappa_p) I_1(\beta_p) - I_0^{-1}(\kappa_p) I_0(\beta_p). \quad (83)$$

The derivative of $C_h$ with respect to the angular-domain parameters for the one-ring model can be found in [18].

Similarly to the frequency-domain and noise parameters, an approximation of the Hessian for estimation of the angular-domain parameters is given by

$$E \left[ \frac{\partial^2 L(\hat{y})}{\partial (\theta_{h}) \partial (\theta_{h})^T} \right] = M_s \text{tr} \left\{ C^{-1}_y \left( \frac{\partial C_y}{\partial (\theta_{h})_k} \right) C^{-1}_y \left( \frac{\partial C_y}{\partial (\theta_{h})_k} \right) \right\}$$

$$= M_s \text{vec} \left\{ \left( C^{-1}_y \left( \frac{\partial C_y}{\partial (\theta_{h})_k} \right) \right)^T \right\}$$

$$\times \text{vec} \left\{ C^{-1}_y \left( \frac{\partial C_y}{\partial (\theta_{h})_k} \right) \right\}. \quad (84)$$

Let us define the matrices containing the partial derivatives with respect to all $N_h$ elements of the parameter vector $\theta_h$ as (85) and (86), shown at the bottom of the page. Hence, the approximation of the Hessian can be expressed in a compact form as

$$E \left[ \frac{\partial^2 L(\hat{y})}{\partial (\theta_{h}) \partial (\theta_{h})^T} \right] = M_s D_{1,h}^T D_{2,h}. \quad (87)$$

Efficient formulas for computing (67), (73), (77), (78), (80), (85), and (86) can be found in Appendix II.

### APPENDIX II

**EFFECTIVE IMPLEMENTATIONS OF MATRIX OPERATIONS**

In order to derive an efficient implementation for (67), we define

$$A_1 = -\text{tr} \left( C^{-1}_y \left( \frac{\partial C_w}{\partial (\theta_{w,h})} \otimes I \right) \right)$$

$$= -\text{tr} \left( (V_w \otimes I_{N_M}) \Lambda^{-1} (V_w^H \otimes I_{N_M}) \right)$$

$$\times \left( \frac{\partial C_w}{\partial (\theta_{w,h})} \otimes I \right)$$

$$= -\text{tr} \left( \Lambda^{-1} (V_w^H \frac{\partial C_w}{\partial (\theta_{w,h})} V_w \otimes I) \right) \quad (88)$$

\[
D_{1,um} = \text{vec} \left\{ C^{-1}_y \left( \frac{\partial C_w}{\partial (\theta_{um})_1} \otimes I \right) \right\}^T \quad \cdots \quad \text{vec} \left\{ C^{-1}_y \left( \frac{\partial C_w}{\partial (\theta_{um})_{N_u}} \otimes I \right) \right\}^T \quad \text{vec} \left\{ (C^{-1}_y \left( I \otimes \Lambda^{-1}_h \right)) \otimes I \right\}^T \quad (77)
\[
D_{2,um} = \text{vec} \left\{ C^{-1}_y \left( \frac{\partial C_w}{\partial (\theta_{um})_1} \otimes I \right) \right\} \quad \cdots \quad \text{vec} \left\{ C^{-1}_y \left( \frac{\partial C_w}{\partial (\theta_{um})_{N_u}} \otimes I \right) \right\} \quad \text{vec} \left\{ C^{-1}_y \left( I \otimes \Lambda^{-1}_h \right) \right\}. \quad (78)
\]

\[
D_{1,h} = \text{vec} \left\{ C^{-1}_y \left( I \otimes \frac{\partial C_h}{\partial (\theta_{h})_1} \right) \right\}^T \quad \cdots \quad \text{vec} \left\{ C^{-1}_y \left( I \otimes \frac{\partial C_h}{\partial (\theta_{h})_{N_h}} \right) \right\}^T \quad (85)
\]

\[
D_{2,h} = \text{vec} \left\{ C^{-1}_y \left( I \otimes \frac{\partial C_h}{\partial (\theta_{h})_1} \right) \right\} \quad \cdots \quad \text{vec} \left\{ C^{-1}_y \left( I \otimes \frac{\partial C_h}{\partial (\theta_{h})_{N_h}} \right) \right\}. \quad (86)
\]
where \( (\theta_{w})_i \) is the \( i \)th element of \( \theta_w = \{\alpha_1, \beta_1, \tau_i\} \). It is assumed in Section III-B that the eigenvectors of a Toeplitz matrix are approximated by the DFT matrix, and \( \psi_{w}^{H}((\partial C_w)/\partial (\theta_{w})_i))V_w \) is also a diagonal matrix, given by \( T((\partial \psi_{w}))/((\partial \theta_{w})_i)) \), with \( T(\cdot) \) defined in (43). Since \( \bar{A}^{-1} \) and \( (\psi_{w}^{H}(\partial C_w)/\partial (\theta_{w})_i)V_w \otimes I) \) are diagonal matrices, (88) can be written as

\[
A_1 = -\text{diag}(\bar{A}^{-1})^T \text{vec} \left\{ \begin{bmatrix} T^T \left( \frac{\partial \bar{A}}{\partial (\theta_{w})_i} \right) \\
\vdots \\
T^T \left( \frac{\partial \bar{A}}{\partial (\theta_{w})_i} \right) \end{bmatrix} \right\}. \tag{89}
\]

We also define

\[
B_1 = \frac{1}{M_s} \sum_{m=1}^{M_s} \bar{V}_{m}^{H} \left( C_{g}^{-1} \left( \psi_{w} \otimes I \right) \right) C_{g}^{-1} \bar{V}_{m} = \frac{1}{M_s} \sum_{m=1}^{M_s} \left( \bar{A}^{-1} \bar{V}_{m} \right)^{H} \left( \psi_{w}^{H} \frac{\partial C_{w}}{\partial (\theta_{w})_i} \psi_{w} \otimes I \right) \bar{A}^{-1} \bar{V}_{m} = \frac{1}{M_s} \sum_{m=1}^{M_s} \left( \text{diag}(\bar{A}^{-1}) \bar{V}_{m} \right)^{H} \left( \psi_{w}^{H} \frac{\partial C_{w}}{\partial (\theta_{w})_i} \right) \text{mat}(\text{diag}(\bar{A}^{-1}) \bar{V}_{m}, M_r, M_t, M_f)^{T} \right) \]  \tag{90}

where \( \text{vec} \) denotes elementwise multiplication and was used the definition of \( \bar{V}_m \) in (39) as well as its corresponding efficient implementation (40). The \( \text{mat}(\cdot, M, N) \) reshapes a vector into an \( M \times N \) matrix, as defined in (29).

Hence, (67) can be implemented efficiently as

\[
\frac{\partial L(\bar{Y})}{\partial (\theta_{w})_i} = A_1 + B_1. \tag{91}
\]

Similarly, in order to derive an efficient implementation for (73), we define

\[
A_2 = -\text{tr} \left( C_{g}^{-1} \frac{\partial C_{g}}{\partial \bar{A}} \right) = -\text{tr} \left( (\psi_{w} \otimes I_{M_r, M_t}) \bar{A}^{-1} (\psi_{w}^{H} \otimes I_{M_r, M_t}) (I \otimes \bar{A}^{-1}) \right) = -\text{tr} \left( \bar{A}^{-1} (I \otimes \bar{A}^{-1}) \right) = -\text{diag}(\bar{A}^{-1})^{T} \begin{bmatrix} \bar{A}^{-1} \\
\vdots \\
\bar{A}^{-1} \end{bmatrix} \tag{92}
\]

and

\[
B_2 = \frac{1}{M_s} \sum_{m=1}^{M_s} \bar{V}_{m}^{H} \left( C_{g}^{-1} \right) C_{g}^{-1} \bar{V}_{m} = \frac{1}{M_s} \sum_{m=1}^{M_s} \left( \bar{A}^{-1} \bar{V}_{m} \right)^{H} C_{g}^{-1} \left( I \otimes \bar{A}^{-1} \right) C_{g}^{-1} \bar{V}_{m} = \frac{1}{M_s} \sum_{m=1}^{M_s} \left( \bar{A}^{-1} \bar{V}_{m} \right)^{H} \left( I \otimes \bar{A}^{-1} \right) \bar{A}^{-1} \bar{V}_{m} = \frac{1}{M_s} \sum_{m=1}^{M_s} \left( \text{diag}(\bar{A}^{-1}) \bar{V}_{m} \right)^{H} \left( I \otimes \bar{A}^{-1} \right) \bar{A}^{-1} \bar{V}_{m} = \frac{1}{M_s} \sum_{m=1}^{M_s} \left( \text{diag}(\bar{A}^{-1}) \bar{V}_{m} \right)^{H} \left( I \otimes \bar{A}^{-1} \right) \bar{A}^{-1} \bar{V}_{m} \right) \]  \tag{93}

Hence, (73) can be implemented efficiently as

\[
\frac{\partial L(\bar{Y})}{\partial (\theta_{w})_i} = A_2 + B_2. \tag{94}
\]

Following similar steps as those used in deriving the efficient implementation of (67), it is straightforward to see that (80) can be implemented as

\[
\frac{\partial L(\bar{Y})}{\partial (\theta_{w})_i} = -\text{diag}(\bar{A}^{-1})^{H} \begin{bmatrix} \text{diag}(\psi_{w}^{H} \frac{\partial C_{w}}{\partial (\theta_{w})_i} \psi_{w}) \\
\vdots \\
\text{diag}(\psi_{w}^{H} \frac{\partial C_{w}}{\partial (\theta_{w})_i} \psi_{w}) \end{bmatrix} + \frac{1}{M_s} \sum_{m=1}^{M_s} \left( \text{diag}(\bar{A}^{-1}) \bar{V}_{m} \right)^{H} \left( \psi_{w}^{H} \frac{\partial C_{w}}{\partial (\theta_{w})_i} \psi_{w} \otimes I \right) \bar{V}_{m} \]  \tag{95}

where the definition of \( \bar{V}_m \) in (52) and the corresponding efficient implementation (53) were used.

The definition of matrices \( D_{1, w} \) and \( D_{2, w} \), \( M_{o} \times N_{w} \), in (77) and (78) do not take into account that several of their entries are equal to zero as a consequence of the Kronecker product in (\( \partial C_{w} \))/((\partial (\theta_{w})_i)) \). We define

\[
A_3(i) = \text{vec} \left\{ C_{g}^{-1} \left( \frac{\partial C_{w}}{\partial \bar{A}} \right) \right\} = \text{vec} \left\{ \bar{A}^{-1} \left( \psi_{w}^{H} \frac{\partial C_{w}}{\partial (\theta_{w})_i} \psi_{w} \otimes I \right) \right\}. \tag{96}
\]

We can define the \( M_{o} \times 1 \) vector \( A_3^{'i} \) containing the nonzero elements of \( A_3^{'i} \) as

\[
A_3^{'i} = \text{vec} \left\{ C_{g}^{-1} \left( \frac{\partial C_{w}}{\partial \bar{A}} \right) \right\} = \text{vec} \left\{ \bar{A}^{-1} \left( \psi_{w}^{H} \frac{\partial C_{w}}{\partial (\theta_{w})_i} \psi_{w} \otimes I \right) \right\}. \tag{97}
\]

Similarly, we can define

\[
A_4 = \text{vec} \left\{ C_{g}^{-1} \left( I \otimes \bar{A}^{-1} \right) \right\} = \text{vec} \left\{ \bar{A}^{-1} \left( I \otimes \bar{A}^{-1} \right) \right\}. \tag{98}
\]
We can define the \( M_0 \times 1 \) vector \( \mathbf{A}_4' \) containing the nonzero elements of \( \mathbf{A}_4 \) as

\[
\mathbf{A}_4' = \text{diag}(\mathbf{A}_h^{-1}) \odot \begin{bmatrix}
\text{diag} (\mathbf{A}_h^{-1}) \\
\vdots \\
\text{diag} (\mathbf{A}_h^{-1})
\end{bmatrix}.
\]

(99)

Hence, in order to be implemented efficiently, \( \mathbf{D}_{2\text{,vec}} \) can be redefined as

\[
\mathbf{D}_{2\text{,vec}} = [\mathbf{A}_0'(1) \cdots \mathbf{A}_0'(N_w) \mathbf{A}_4'].
\]

(100)

It is straightforward to see that \( \mathbf{D}_{2\text{,vec}} \) can also be implemented efficiently as (101), shown at the top of the page, where the operator \( \mathcal{T}(\cdot) \) is defined in (43).

Also, following similar derivations, we can implement (85) and (86) as (102) and (103), shown at the top of the page. Hence, the approximation of the Hessian can be expressed in a compact form as

\[
E \left[ \frac{\partial^2 L_\theta(Y)}{\partial \theta_i \partial \theta_j} \right] = M_0 \mathbf{D}_{1,\mathcal{T}_h}^T \mathbf{D}_{2,h}.
\]

(104)
Cássio Ribeiro (S’00–M’05) was born in Brazil in 1977. He received the Electronics Eng. degree from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 2000 and the M.Sc. degree in electrical engineering from COPPE/UFRJ in 2002. He is currently working towards the D.Sc. degree at UFRJ and the Helsinki University of Technology, Helsinki, Finland.

In 2006, he worked as a Researcher at Nokia Institute of Technology, Brazil. Since 2007, he has been working at Nokia Research Center, Helsinki, Finland. His current research interests are in statistical signal processing, parameter estimation methods, and cellular systems.

Mr. Ribeiro coauthored a paper receiving the Best Paper Award in IEEE Personal, Indoor and Mobile Radio Communications (PIMRC) 2005.

Andreas Richter (M’04) was born in Germany in 1969. He received the Dipl.-Ing. (M.Sc.) degree in electrical engineering and the Dr.-Ing. (Ph.D.) degree (summa cum laude) from the Technische Universität Ilmenau, Ilmenau, Germany, in 1995 and 2005, respectively.

From 1995 to 2005, he worked as a Research Assistant at the Electronic Measurement Laboratory of Technische Universität Ilmenau. From July to October 2001, he was a Guest Researcher at the Wireless Laboratories, NTT DoCoMo, Yokosuka, Japan. Since 2004, he has been working as a Senior Researcher in the Statistical Signal Processing Laboratory at Helsinki University of Technology, Finland. His research interests are in the fields of digital communication, sensor-array, and statistical signal processing.

Dr. Richter was coauthor or author of five papers receiving a Best Paper Award (EPMCC’01, ISAP’04, PIMRC’05, EUSIPCO’06, and EuCAP’06). In 2005, he received the Siemens Communications Academic Award. For their work on MIMO channel sounding, he and his former colleagues at Technische Universität Ilmenau received the Thuringien Research Award for Applied Research in 2007.

Visa Koivunen (S’87–M’87–SM’98) received the D.Sc. (Tech.) degree with honors from the Department of Electrical Engineering, University of Oulu, Finland.

From 1992 to 1995, he was a visiting researcher at the University of Pennsylvania, Philadelphia. From 1997 to 1999, he was an Associate Professor at the Signal Processing Laboratory, Tampere University of Technology, Tampere, Finland. Since 1999, he has been a Professor of signal processing at the Department of Electrical and Communications Engineering, Helsinki University of Technology (HUT), Finland. He is one of the Principal Investigators in the SMARAD (Smart and Novel Radios) Center of Excellence in Radio and Communications Engineering nominated by the Academy of Finland. Since 2003, he has been Adjunct Professor at the University of Pennsylvania. During his sabbatical term 2006–2007, he is Nokia Visiting Fellow at Nokia Research Center, Helsinki, and visiting fellow at Princeton University, Princeton, NJ. His research interests include statistical, communications, and sensor array signal processing. He has published more than 200 peer-reviewed papers in international scientific conferences and journals.

Dr. Koivunen coauthored papers receiving the Best Paper Awards in IEEE PIMRC 2005, EUSIPCO 2006, and EuCAP 2006 conferences. He served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He is a member of the editorial board for the Signal Processing journal and the Journal of Wireless Communication and Networking. He is also a member of the IEEE Signal Processing for Communication Technical Committee (SPCOM-TC). He is the General Chair of the IEEE Signal Processing Advances in Wireless Communications (SPAWC) 2007 conference in Helsinki, June 2007.