Estimating Jakes' Doppler Power Spectrum Parameters Using the Whittle Approximation

Aleksandar Dogandžić and *Benhong Zhang*

ECpE Department, Iowa State University

3119 Coover Hall, Ames, IA 50011

Phone: (515) 294-0500, Fax: (515) 294-8432

email: {`ald, zhangbh`}`@iastate.edu`

Abstract

We derive methods for asymptotic maximum likelihood (ML) estimation of Jakes' Doppler power spectrum parameters from complex noisy estimates of the fading channel. We consider both single-input single-output (SISO) and smart-antenna scenarios and utilize the Whittle approximation to the likelihood to estimate the Doppler spread, noise variance, and channel covariance parameters. Asymptotic Cramér-Rao bounds for the unknown parameters are derived. We also discuss the initialization of the proposed methods and their generalization to the Ricean-fading scenario. Numerical simulations demonstrate the performance of the proposed methods.

Index Terms

Cramér-Rao bound, Doppler-spread (velocity) estimation, fading, Jakes' (Clarke's) spectrum, signal-to-noise ratio estimation, Whittle approximation.

I. INTRODUCTION

DOPPLER spread and signal-to-noise ratio (SNR) are important parameters for assessing the quality and rate of change of wireless communication channels [1]–[10]. The Doppler spread determines the rate of channel variation and fading type[1], and can be used for adaptive modulation, coding, and interleaving, channel tracker step-size selection (at the receiver), and for network control algorithms, such as handoff and channel allocation in cellular systems [1]–[10]. Similarly, the SNR information is instrumental for adaptive modulation, handoff, channel access, and power control [1]–[7]. In smart-antenna systems, modeling spatial fading correlations, analyzing their effects on capacity and error-probability performance, and the use of fading correlations in the design of noncoherent ML space-time receivers and transmit precoding schemes have recently attracted considerable attention, see e.g. [12]–[17] and references therein.

Most existing methods for estimating statistical properties of fading channels are based on signal-amplitude or power measurements and do not explicitly account for noise effects, see e.g. [1, Ch. 12], [2], [4]–[6], [18], and [19]. Also, the Doppler spread and signal strength were estimated separately [1], [2]. In [20], an approximate average maximum likelihood method was proposed for estimating the Doppler spread from noisy channel estimates under a single-input single-output (SISO) Rayleigh fading scenario. In [9], an exact ML estimator of the Doppler spread was derived for this scenario, assuming that the SNR is known and multiple independent data slots are available. In

---

[1]See e.g. [11, Ch. 5.5.2] for slow- and fast-fading channel characterization.

[17], we derived methods for estimating the mean and covariance parameters of multi-input multi-output (MIMO) *block-fading* channels. In this paper (see also [21]), we develop asymptotic ML methods for the *joint* estimation of the Doppler spread, noise variance, and channel covariance parameters from *complex noisy* channel estimates [containing *both* the in-phase and quadrature-phase (I/Q) components of the fading channel] under Jakes' SISO and single-input multi-output (SIMO) smart-antenna scenarios.

In Section II, we utilize the Whittle approximation to develop asymptotic ML methods for Jakes' Doppler power spectrum estimation in SISO systems and derive asymptotic Cramér-Rao bounds (CRBs) for the unknown parameters (Section II-A). A generalization of the proposed method to the Ricean fading scenario is derived in Section II-B. The smart-antenna scenario is considered in Section III, where we present iterative algorithms for asymptotic ML estimation for unstructured and independent fading (Sections III-A and III-B) and corresponding asymptotic CRBs (Section III-C). In Section IV, we evaluate the accuracy of the proposed methods via numerical simulations. The asymptotic ML estimates of the Doppler spread are compared with the sample-covariance-based and approximate ML methods in [6], [10], and [20] and their multivariate extensions. Concluding remarks are given in Section V.

II. Estimating Jakes' Power Spectrum Parameters in SISO Systems

Assume that we have obtained N noisy channel estimates $y(1), y(2), \ldots, y(N)$ from a Rayleigh fading channel with Jakes' Doppler power spectrum[2] [1], [22], [23]. For example, if we transmit an unmodulated carrier, then the real and imaginary parts of $y(t)$ are the I/Q components of the received baseband signal at time $t \in \{1, 2, \ldots, N\}$. We assume that the channel estimates are corrupted by additive white circularly symmetric complex Gaussian noise with an unknown variance σ^2 and that the noise is independent from the fading process. The real and imaginary parts of the fading process are assumed to be independent, which follows from [23, App. A] and [1, Ch. 2.1.1]. Define the indicator function:

$$i_A(f) = \left\{ \begin{array}{ll} 1, & f \in A, \\ 0, & \text{otherwise} \end{array} \right. . \tag{2.1}$$

Then, the noisy power spectral density (PSD) of $y(t)$ can be written as[3]

$$P_{yy}(f; \boldsymbol{\theta}) = \sigma^2 \cdot g(f; \boldsymbol{\rho}) = \sigma^2 \cdot \left[\frac{s \cdot i_{[0, f_{\mathrm{D}})}(f)}{(f_{\mathrm{D}}^2 - f^2)^{1/2}} + \frac{s \cdot i_{(1 - f_{\mathrm{D}}, 1]}(f)}{[f_{\mathrm{D}}^2 - (1 - f)^2]^{1/2}} + 1 \right], \tag{2.2}$$

where $f \in [0, 1]$ is the discrete-time frequency,

$$\boldsymbol{\theta} = [\sigma^2, \boldsymbol{\rho}^T]^T \tag{2.3}$$

is the vector of unknown parameters,

$$\boldsymbol{\rho} = [f_{\mathrm{D}}, s]^T, \tag{2.4}$$

---

[2]This model implies isotropic scattering, i.e. the multipath components are assumed to arrive at the receiver array uniformly from all directions, see [1], [22], [23].

[3]We define the PSD of a stationary zero-mean random process $y(t)$ as $P_{yy}(f) = \sum_{n=-\infty}^{\infty} \mathrm{E}\left[y(t)y(t + n)^*\right] \cdot \exp(-j2\pi f n)$, where "*" denotes complex conjugation.

and "T" denotes a transpose. Here,

- $f_\mathrm{D} \in (0, 1/2]$ is the unknown maximum Doppler frequency (corresponding to the Doppler spread of $2f_\mathrm{D}$);
- s is the unknown signal-to-noise ratio, defined as the ratio between the scattering power of the fading channel and the noise variance σ^2.

The maximum Doppler frequency is proportional to the speed of the mobile v: $f_\mathrm{D} = v f_\mathrm{c}/c$, where f_c is the carrier frequency and c is the speed of light, see e.g. [1], [11], [22], [23]. The first two terms in (2.2) model the Jakes' spectrum [1], [22], [23], whereas the third term is attributed to additive white Gaussian noise. For large N, we can apply the Whittle approximation to the log-likelihood of the measurements $\boldsymbol{y} = [y(1), y(2), \ldots, y(N)]^T$ (see [24, Chs. 7.9 and 15.9] and [25]):

$$
\begin{aligned}
l(\boldsymbol{y}; \boldsymbol{\theta}) &= -N \ln \pi - \sum_{k=0}^{N-1} \left\{ \ln[P_{yy}(f_k; \boldsymbol{\theta})] + \frac{C_{y,N}(f_k)}{P_{yy}(f_k; \boldsymbol{\theta})} \right\} \tag{2.5a} \\
&= -N \ln \pi - \sum_{k=0}^{N-1} \left\{ \ln\left[\sigma^2 \cdot g(f_k; \boldsymbol{\rho})\right] + \frac{C_{y,N}(f_k)}{\sigma^2 \cdot g(f_k; \boldsymbol{\rho})} \right\}, \tag{2.5b}
\end{aligned}
$$

where

$$
f_k = k/N, \quad k = 0, 1, \ldots, N-1 \tag{2.6}
$$

and

$$
\begin{aligned}
C_{y,N}(f) &= \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-j2\pi ft} \right|^2 = |y_\mathrm{DTFT}(f)|^2, \\
y_\mathrm{DTFT}(f) &= \frac{1}{\sqrt{N}} \sum_{t=1}^{N} y(t) e^{-j2\pi ft} \tag{2.7}
\end{aligned}
$$

are the the periodogram and normalized discrete-time Fourier transform (DTFT) of $y(t)$, $t = 1, 2, \ldots, N$. Then, $y_\mathrm{DTFT}(f_k)$, $k = 0, 1, \ldots, N-1$ form the normalized discrete Fourier transform (DFT) of $y(t)$, $t = 1, 2, \ldots, N$, which can be computed efficiently using the fast Fourier transform (FFT) if N is a power of two. Here, (2.5b) follows by substituting (2.2) into (2.5a). In [24, Ch. 7.9], the estimation of unknown parameters by maximizing the Whittle log-likelihood is referred to as *asymptotic ML estimation*. For stationary processes, it is typically more convenient to parametrize the PSD rather than the autocorrelation function, which makes the Whittle approximation very appealing. Here, the Jakes' PSD is a closed-form expression of the unknown parameters, whereas the corresponding covariance matrix of the observations $\boldsymbol{y}$ is not analytically tractable:

$$
\mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^H\right] = \sigma^2 \cdot [s\,\boldsymbol{J}(f_\mathrm{D}) + \boldsymbol{I}_N], \tag{2.8}
$$

where the (p, q) element of the matrix $\boldsymbol{J}(f_\mathrm{D})$ is

$$
[J(f_\mathrm{D})]_{p,q} = \pi \cdot J_0(2\pi f_\mathrm{D}(p - q)), \tag{2.9}
$$

"^{H}" denotes the Hermitian (conjugate) transpose, $\boldsymbol{I}_N$ the identity matrix of size N, and $J_0(\cdot)$ the zeroth-order Bessel function of the first kind [1], [22], [23].

We now compute the asymptotic (Whittle) ML estimate of $\boldsymbol{\theta}$ by maximizing (2.5). For fixed $\boldsymbol{\rho}$, there exists a closed-form expression for the asymptotic ML estimate of σ^2:

$$\widehat{\sigma}^2(\boldsymbol{\rho}) = \frac{1}{N} \sum_{k=0}^{N-1} \frac{C_{y,N}(f_k)}{g(f_k; \boldsymbol{\rho})}, \tag{2.10}$$

see also [26] and [27]. Substituting (2.10) into the Whittle log-likelihood function (2.5b) and neglecting constant terms yields the concentrated likelihood function:

$$l_{\mathrm{c}}(\boldsymbol{y}; \boldsymbol{\rho}) = l(\boldsymbol{y}, [\widehat{\sigma}^2(\boldsymbol{\rho}), \boldsymbol{\rho}^T]^T) = -N \ln \Big[\sum_{k=0}^{N-1} \frac{C_{y,N}(f_k)}{g(f_k; \boldsymbol{\rho})} \Big] - \sum_{k=0}^{N-1} \ln \big[g(f_k; \boldsymbol{\rho}) \big] \tag{2.11}$$

to be maximized with respect to $\boldsymbol{\rho}$.

Let us now introduce the following notation: define

$$z_k(f_{\mathrm{D}}) = \frac{i_{[0,f_{\mathrm{D}})}(f_k)}{(f_{\mathrm{D}}^2 - f_k^2)^{1/4}} + \frac{i_{(1-f_{\mathrm{D}},1]}(f_k)}{[f_{\mathrm{D}}^2 - (1 - f_k)^2]^{1/4}}. \tag{2.12}$$

Observe that, since $f_{\mathrm{D}} \in (0, 1/2]$, the following identity holds:

$$z_k^n(f_{\mathrm{D}}) = \frac{i_{[0,f_{\mathrm{D}})}(f_k)}{(f_{\mathrm{D}}^2 - f_k^2)^{n/4}} + \frac{i_{(1-f_{\mathrm{D}},1]}(f_k)}{[f_{\mathrm{D}}^2 - (1 - f_k)^2]^{n/4}}, \tag{2.13}$$

implying that

$$g(f_k, \boldsymbol{\rho}) = s \cdot z_k^2(f_{\mathrm{D}}) + 1, \quad k = 0, 1, \ldots, N - 1. \tag{2.14}$$

Approximate ML Estimator of f_{D}**:** In Appendix A, we derive the following *approximate ML estimator:*

$$\widehat{f}_{\mathrm{D}}^{\mathrm{app}} = \arg \max_{f_{\mathrm{D}}} \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot C_{y,N}(f_k), \tag{2.15}$$

which is closely related to the *approximate average ML* method in [20]. Interestingly, $\widehat{f}_{\mathrm{D}}^{\mathrm{app}}$ may outperform the asymptotic (Whittle) ML estimator of f_{D}, see Section IV.

***Initialization*:** The algorithms for maximizing (2.11) and (2.15) can be initialized using the following simple estimator of f_{D} (see [10, eq. (10)]):

$$\widehat{f}_{\mathrm{D}}^{\mathrm{HS}} = \frac{1}{2\pi} \cdot \sqrt{\frac{2 \sum_{t=1}^{N-1} |y(t) - y(t+1)|^2}{\sum_{t=1}^{N-1} |y(t)|^2}}, \tag{2.16}$$

which can be viewed as an extension of the *sample-covariance-based estimator* by Holtzman and Sampath [6] to the complex (I/Q) measurement scenario. The above estimator performs well if the number of observations N is small, but does not have good asymptotic properties, see Section IV. We also propose the following initial estimate of s:

$$s^{(0)} = \frac{(1/N^{(0)}) \cdot \sum_{k=0, \, z_k(f_{\mathrm{D}}^{(0)}) \neq 0}^{N-1} C_{y,N}(f_k) / z_k^2(f_{\mathrm{D}}^{(0)})}{[1/(N - N^{(0)})] \cdot \sum_{k=0, \, z_k(f_{\mathrm{D}}^{(0)}) = 0}^{N-1} C_{y,N}(f_k)}, \tag{2.17}$$

where $f_{\mathrm{D}}^{(0)}$ can be chosen as, for example, $\widehat{f}_{\mathrm{D}}^{\mathrm{HS}}$ or $\widehat{f}_{\mathrm{D}}^{\mathrm{app}}$, and $N^{(0)}$ is the number of terms in the summation in the numerator of (2.17), equal to the number of indices k for which $z_k(f^{(0)})$ is nonzero.

A. Asymptotic Cramér-Rao Bound

We derive asymptotic CRB expressions for the unknown parameters and simplify them in the case where the SNR s and number of samples N are large.

The asymptotic CRB for the unknown parameters $\boldsymbol{\theta}$ is the inverse of the asymptotic Fisher information matrix (FIM) $\mathcal{I}(\boldsymbol{\theta})$:

$$\mathbf{CRB}(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})^{-1}, \tag{2.18}$$

where $\mathcal{I}(\boldsymbol{\theta})$ easily follows from the Whittle approximation of the log-likelihood in (2.5a):

$$\mathcal{I}(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \cdot \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}^T}, \tag{2.19}$$

see also [24, Ch. 15.9]. We now specialize (2.18) and (2.19) to the Jakes' PSD model in (2.2), which leads to

$$\frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial \sigma^2} = \frac{1}{\sigma^2}, \tag{2.20a}$$

$$\frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial s} = \frac{\partial \ln[g(f; \boldsymbol{\rho})]}{\partial s} = \frac{1}{g(f_k; \boldsymbol{\rho})} \cdot z_k^2(f_{\mathrm{D}}), \tag{2.20b}$$

$$\frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial f_{\mathrm{D}}} = \frac{\partial \ln[g(f_k; \boldsymbol{\rho})]}{\partial f_{\mathrm{D}}} = -\frac{s\, f_{\mathrm{D}}}{g(f_k; \boldsymbol{\rho})} \cdot z_k^6(f_{\mathrm{D}}) \tag{2.20c}$$

for $k = 0, 1, \ldots, N-1$, where we used the following identity to derive (2.20c):

$$\frac{\partial z_k^2(f_{\mathrm{D}})}{\partial f_{\mathrm{D}}} = -f_{\mathrm{D}} \cdot z_k^6(f_{\mathrm{D}}). \tag{2.21}$$

Consequently,

$$\mathcal{I}_{1,1}(\boldsymbol{\theta}) = \mathcal{I}_{1,1}(\sigma^2) = \sum_{k=0}^{N-1} \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial \sigma^2} \cdot \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial \sigma^2} = \frac{N}{(\sigma^2)^2}, \tag{2.22a}$$

$$\mathcal{I}_{1,2}(\boldsymbol{\theta}) = \mathcal{I}_{2,1}(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial \sigma^2} \cdot \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial f_{\mathrm{D}}} = -\frac{s\, f_{\mathrm{D}}}{\sigma^2} \cdot \sum_{k=0}^{N-1} \frac{z_k^6(f_{\mathrm{D}})}{g(f_k; \boldsymbol{\rho})}, \tag{2.22b}$$

$$\mathcal{I}_{1,3}(\boldsymbol{\theta}) = \mathcal{I}_{3,1}(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial \sigma^2} \cdot \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial s} = \frac{1}{\sigma^2} \cdot \sum_{k=0}^{N-1} \frac{z_k^2(f_{\mathrm{D}})}{g(f_k; \boldsymbol{\rho})}, \tag{2.22c}$$

$$\mathcal{I}_{2,2}(\boldsymbol{\theta}) = \mathcal{I}_{2,2}(\boldsymbol{\rho}) = \sum_{k=0}^{N-1} \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial f_{\mathrm{D}}} \cdot \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial f_{\mathrm{D}}} = (s\, f_{\mathrm{D}})^2 \cdot \sum_{k=0}^{N-1} \frac{z_k^{12}(f_{\mathrm{D}})}{g^2(f_k; \boldsymbol{\rho})}, \tag{2.22d}$$

$$\mathcal{I}_{2,3}(\boldsymbol{\theta}) = \mathcal{I}_{3,2}(\boldsymbol{\rho}) = \sum_{k=0}^{N-1} \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial f_{\mathrm{D}}} \cdot \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial s} = -s\, f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} \frac{z_k^8(f_{\mathrm{D}})}{g^2(f_k; \boldsymbol{\rho})}, \tag{2.22e}$$

$$\mathcal{I}_{3,3}(\boldsymbol{\theta}) = \mathcal{I}_{3,3}(\boldsymbol{\rho}) = \sum_{k=0}^{N-1} \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial s} \cdot \frac{\partial \ln[P_{yy}(f_k; \boldsymbol{\theta})]}{\partial s} = \sum_{k=0}^{N-1} \frac{z_k^4(f_{\mathrm{D}})}{g^2(f_k; \boldsymbol{\rho})}. \tag{2.22f}$$

We can partition $\mathcal{I}(\boldsymbol{\theta})$ as follows:

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{I}_{\sigma^2,\sigma^2}(\sigma^2) & \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})^T \\ \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta}) & \mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho}) \end{bmatrix}, \tag{2.23a}$$

where

$$\mathcal{I}_{\sigma^2,\sigma^2}(\sigma^2) = \mathcal{I}_{1,1}(\sigma^2), \tag{2.23b}$$

$$\mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta}) = \left[\begin{array}{c} \mathcal{I}_{1,2}(\boldsymbol{\theta}) \\ \mathcal{I}_{1,3}(\boldsymbol{\theta}) \end{array} \right], \tag{2.23c}$$

$$\mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\theta}) = \left[\begin{array}{cc} \mathcal{I}_{2,2}(\boldsymbol{\rho}) & \mathcal{I}_{2,3}(\boldsymbol{\rho}) \\ \mathcal{I}_{2,3}(\boldsymbol{\rho}) & \mathcal{I}_{3,3}(\boldsymbol{\rho}). \end{array} \right]. \tag{2.23d}$$

In the sequel, we use the same block partitioning of the CRB as for the above FIM. Using (2.22a) and the formula for the inverse of a partitioned matrix in e.g. [28, Theorem 8.5.11] yields

$$\mathbf{CRB}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho}) = \left[\mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho}) - \frac{\sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta}) \cdot \sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})^T}{N} \right]^{-1} \tag{2.24a}$$

$$\mathrm{CRB}_{\sigma^2\sigma^2}(\boldsymbol{\theta}) = \frac{\sigma^4}{N - \sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})^T \cdot \mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho})^{-1} \cdot \sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})}. \tag{2.24b}$$

Observe that $\sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})$ does not depend on σ^2. Hence, $\mathbf{CRB}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho})$ is a function of $\boldsymbol{\rho}$ only, implying that the asymptotic accuracy of (efficiently) estimating the maximum Doppler frequency f_{D} and SNR parameter s depends on the unknown parameters only through f_{D} and s. Interestingly (2.24a) further implies that $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ is a function of s only through $g(f_k; \boldsymbol{\rho})$ and a function of f_{D} through $z_k(f_{\mathrm{D}})$, $k = 0, 1, \ldots, N-1$. Also, $\mathrm{CRB}_{\sigma^2\sigma^2}(\boldsymbol{\theta})$ depends on σ^2 only through σ^4 in the numerator of (2.24b); it is a function of s through $g(f_k; \boldsymbol{\rho})$ and a function of f_{D} through $z_k(f_{\mathrm{D}})$, $k = 0, 1, \ldots, N-1$.

Since the Jakes' PSD in (2.2) is infinite at $f = f_{\mathrm{D}}$ and $f = 1 - f_{\mathrm{D}}$, the regularity conditions for the information inequality are not satisfied, see [29, Ch. 3.4.2]. To avoid this problem, we assume that the frequencies f_k, $k = 0, 1, \ldots, N-1$ in the DFT grid do not coincide with f_{D} or $1 - f_{\mathrm{D}}$. Note that $\mathrm{CRB}_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho})$ strongly depends on the minimum distance between f_{D} and the DFT grid $f_0, f_1, \ldots, f_{N-1}$. For a fixed f_{D}, a small variation in the number of samples N may significantly change this distance, which causes oscillatory behavior of $\mathrm{CRB}_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho})$ as a function of N, see Fig. 1 in Section IV. For a large fixed N (i.e. dense DFT grid), a small variation of f_{D} causes a significant change in the minimum distance, resulting in oscillations in $\mathrm{CRB}_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho})$ as a function of f_{D}, see Fig. 3 in Section IV. (The exact $\mathrm{CRB}_{f_{\mathrm{D}},f_{\mathrm{D}}}$ in [10, Fig. 2], computed under a different measurement model, also fluctuates as a function of f_{D}.)

$\mathrm{CRB}_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho})$ *and* $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ *for Large* s *and* N: For large s and N, the following approximations hold:

$$\frac{1}{N} \sum_{k=0}^{N-1} \frac{z_k^2(f_{\mathrm{D}})}{g(f_k, \boldsymbol{\rho})} \approx \int_{-f_{\mathrm{D}}}^{f_{\mathrm{D}}} \frac{df}{s + \sqrt{f_{\mathrm{D}}^2 - f^2}} \approx \frac{2 f_{\mathrm{D}}}{s}, \tag{2.25a}$$

$$\frac{1}{N} \sum_{k=0}^{N-1} \frac{z_k^4(f_{\mathrm{D}})}{g^2(f_k, \boldsymbol{\rho})} \approx \int_{-f_{\mathrm{D}}}^{f_{\mathrm{D}}} \frac{df}{(s + \sqrt{f_{\mathrm{D}}^2 - f^2})^2} \approx \frac{2 f_{\mathrm{D}}}{s^2}. \tag{2.25b}$$

Using (2.27), we simplify (2.24a):

$$\mathbf{CRB}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho}) = \left[\begin{array}{cc} \kappa_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho}) & \kappa_{f_{\mathrm{D}},s}(\boldsymbol{\rho}) \\ \kappa_{f_{\mathrm{D}},s}(\boldsymbol{\rho}) & \kappa_{s,s}(\boldsymbol{\rho}) \end{array} \right]^{-1}, \tag{2.26}$$

where

$$\kappa_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho}) \approx f_{\mathrm{D}}^2 \cdot \sum_{k=0}^{N-1} \left\{ z_k^4(f_{\mathrm{D}}) - \left[(1/N) \cdot \sum_{l=0}^{N-1} z_l^4(f_{\mathrm{D}}) \right] \right\}^2, \tag{2.27a}$$

$$\kappa_{f_{\mathrm{D}},s}(\boldsymbol{\rho}) \approx -\frac{f_{\mathrm{D}}(1-2f_{\mathrm{D}})}{s} \cdot \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}), \tag{2.27b}$$

$$\kappa_{s,s}(\boldsymbol{\rho}) \approx \frac{2f_{\mathrm{D}}N}{s^2} \cdot (1-2f_{\mathrm{D}}). \tag{2.27c}$$

Then, the CRBs for f_{D} and s can be approximated as shown in (2.28):

$$\mathrm{CRB}_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho}) \approx \frac{2N}{f_{\mathrm{D}} \cdot \left(2f_{\mathrm{D}}N \cdot \sum_{k=0}^{N-1} \left\{ z_k^4(f_{\mathrm{D}}) - \left[(1/N) \cdot \sum_{l=0}^{N-1} z_l^4(f_{\mathrm{D}}) \right] \right\}^2 - (1-2f_{\mathrm{D}}) \cdot \left[\sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \right]^2 \right)}, \tag{2.28a}$$

$$\mathrm{CRB}_{s,s}(\boldsymbol{\rho}) \approx \frac{s^2 \cdot \sum_{k=0}^{N-1} \left\{ z_k^4(f_{\mathrm{D}}) - \left[(1/N) \cdot \sum_{l=0}^{N-1} z_l^4(f_{\mathrm{D}}) \right] \right\}^2}{(1-2f_{\mathrm{D}}) \cdot \left(2f_{\mathrm{D}}N \cdot \sum_{k=0}^{N-1} \left\{ z_k^4(f_{\mathrm{D}}) - \left[(1/N) \cdot \sum_{l=0}^{N-1} z_l^4(f_{\mathrm{D}}) \right] \right\}^2 - (1-2f_{\mathrm{D}}) \cdot \left[\sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \right]^2 \right)}$$

$$\approx \frac{s^2}{2N \cdot f_{\mathrm{D}}(1-2f_{\mathrm{D}})}. \tag{2.28b}$$

Clearly, as the number of samples N grows, the approximate $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ decreases proportionally to $1/N$, see also Fig. 1 (right) in Section IV. Furthermore, the approximate $\mathrm{CRB}_{f_{\mathrm{D}},f_{\mathrm{D}}}(\boldsymbol{\rho})$ does not depend on s [see also Fig. 2 (left) in Section IV] and $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ is proportional to s^2, which is also confirmed in Fig. 2 (right).

Note that σ^2 is not identifiable when $f_{\mathrm{D}} = 0.5$, implying that the SNR parameter s is *not identifiable* as well. (Recall that s is defined as the ratio between the scattering power of the fading channel and the noise variance σ^2.) Consequently, $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ goes to infinity as f_{D} approaches 0.5, see (2.28b) and Fig. 4 in Section IV. However, s is identifiable when σ^2 is *known,* see also Section IV. This scenario is of practical interest, since σ^2 may be estimated from noise-only data. Inverting the approximate $\mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho})$ [where the approximate formulas (2.25) were used to compute its elements] yields the approximate CRB for s when σ^2 is known

$$\mathrm{CRB}_{s,s}(\boldsymbol{\rho} \,|\, \mathrm{known}\ \sigma^2) \approx s^2/(2Nf_{\mathrm{D}}), \tag{2.29}$$

which decreases proportionally to $1/f_{\mathrm{D}}$ as f_{D} increases, see also Fig. 4.

B. Extension to Ricean Fading

In the Ricean-fading scenario, $\boldsymbol{y}$ has nonzero mean and (noisy) Jakes' covariance, described by

$$\mathrm{E}\left[\boldsymbol{y}\right] = x \cdot [\varphi(1; f_{\mathrm{LOS}}), \varphi(2; f_{\mathrm{LOS}}), \ldots, \varphi(N; f_{\mathrm{LOS}})]^T, \tag{2.30a}$$

$$\mathrm{cov}[\boldsymbol{y}] = \mathrm{E}\left\{ (\boldsymbol{y} - \mathrm{E}\left[\boldsymbol{y}\right])(\boldsymbol{y} - \mathrm{E}\left[\boldsymbol{y}\right])^H \right\} = \sigma^2 \cdot [s\,\boldsymbol{J}(f_{\mathrm{D}}) + \boldsymbol{I}_N], \tag{2.30b}$$

where

$$\varphi(t; f_{\mathrm{LOS}}) = \exp(j2\pi f_{\mathrm{LOS}}t), \quad t = 1, 2, \ldots, N \tag{2.31}$$

and

- x is the (unknown) complex amplitude of the line-of-sight component;
- $f_{\text{LOS}} \in [0, f_{\text{D}}) \cup [1 - f_{\text{D}}, 1)$ is the Doppler shift due to the line-of-sight component;
- s is the SNR of the scattering (diffuse) channel component.

(Note that the line-of-sight Doppler shift can be written as $f_{\text{LOS}} = f_{\text{D}} \cdot \cos \vartheta$ where ϑ is the angle between the line-of-sight and mobile velocity vectors.)

The vector of unknown parameters is now[4]

$$\boldsymbol{\theta} = [\sigma^2, \boldsymbol{\rho}^T, \text{Re}\{x\}, \text{Im}\{x\}, f_{\text{LOS}}]^T \tag{2.32}$$

and the Whittle approximation to the log-likelihood (for the measurements $\boldsymbol{y}$) becomes:

$$l(\boldsymbol{y}; \boldsymbol{\theta}) = -N \ln \pi - \sum_{k=0}^{N-1} \left\{ \ln \left[\sigma^2 \cdot g(f_k; \boldsymbol{\rho}) \right] + \frac{|y_{\text{DTFT}}(f_k) - x \cdot \varphi_{\text{DTFT}}(f_k; f_{\text{LOS}})|^2}{\sigma^2 \cdot g(f_k; \boldsymbol{\rho})} \right\}, \tag{2.33}$$

where $\varphi_{\text{DTFT}}(f; f_{\text{LOS}})$ is the normalized DTFT of the sequence $\varphi(t; f_{\text{LOS}})$, $t = 1, 2, \ldots, N$:

$$\varphi_{\text{DTFT}}(f; f_{\text{LOS}}) = \frac{1}{\sqrt{N}} \cdot \exp[-j2\pi(f - f_{\text{LOS}})] \cdot \frac{1 - \exp[-j2\pi(f - f_{\text{LOS}})N]}{1 - \exp[-j2\pi(f - f_{\text{LOS}})]}. \tag{2.34}$$

For fixed $\boldsymbol{\rho}$ and f_{LOS}, there exist closed-form expressions for the asymptotic ML estimates of x and σ^2 that maximize (2.33):

$$\widehat{x}(\boldsymbol{\rho}, f_{\text{LOS}}) = \frac{\sum_{k=0}^{N-1} [\varphi_{\text{DTFT}}(f_k; f_{\text{LOS}})]^* \cdot y_{\text{DTFT}}(f_k) / g(f_k; \boldsymbol{\rho})}{\sum_{k=0}^{N-1} |\varphi_{\text{DTFT}}(f_k; f_{\text{LOS}})|^2 / g(f_k; \boldsymbol{\rho})}, \tag{2.35a}$$

$$\widehat{\sigma}^2(\boldsymbol{\rho}, f_{\text{LOS}}) = \frac{1}{N} \cdot \sum_{k=0}^{N-1} \frac{|y_{\text{DTFT}}(f_k) - \widehat{x}(\boldsymbol{\rho}, f_{\text{LOS}}) \cdot \varphi_{\text{DTFT}}(f_k; f_{\text{LOS}})|^2}{g(f_k; \boldsymbol{\rho})}$$

$$= \frac{1}{N} \cdot \sum_{k=0}^{N-1} \left(\frac{|y_{\text{DTFT}}(f_k)|^2}{g(f_k; \boldsymbol{\rho})} \right) - \frac{1}{N} \cdot \frac{|\sum_{k=0}^{N-1} [\varphi_{\text{DTFT}}(f_k; f_{\text{LOS}})]^* \cdot y_{\text{DTFT}}(f_k) / g(f_k; \boldsymbol{\rho})|^2}{\sum_{k=0}^{N-1} |\varphi_{\text{DTFT}}(f_k; f_{\text{LOS}})|^2 / g(f_k; \boldsymbol{\rho})}. \tag{2.35b}$$

Substituting (2.35) into the Whittle log-likelihood function (2.33) and neglecting constant terms yields the concentrated likelihood function

$$\begin{aligned} l_{\text{c}}(\boldsymbol{y}; \boldsymbol{\rho}, f_{\text{LOS}}) &= l\left(\boldsymbol{y}, [\widehat{\sigma}^2(\boldsymbol{\rho}, f_{\text{LOS}}), \boldsymbol{\rho}^T, \text{Re}\{\widehat{x}(\boldsymbol{\rho}, f_{\text{LOS}})\}, \text{Im}\{\widehat{x}(\boldsymbol{\rho}, f_{\text{LOS}})\}, f_{\text{LOS}}]^T \right) \\ &= -N \ln \left[\widehat{\sigma}^2(\boldsymbol{\rho}, f_{\text{LOS}}) \right] - \sum_{k=0}^{N-1} \ln \left[g(f_k; \boldsymbol{\rho}) \right] \end{aligned} \tag{2.36}$$

to be maximized with respect to $\boldsymbol{\rho}$ and f_{LOS}.

Initialization: The maximization of (2.36) can be initialized using the sample-covariance-based estimate $\widehat{f}_{\text{D}}^{\text{HS}}$ of the maximum Doppler frequency f_{D} [see (2.16)], nonlinear least-squares estimate of the line-of-sight Doppler shift f_{LOS}:

$$\widehat{f}_{\text{LOS}}^{\text{LS}} = \arg \max_{f_{\text{LOS}}} \left| \sum_{t=1}^{N} y(t) \cdot \exp(-j2\pi f_{\text{LOS}} t) \right| = \arg \max_{f_{\text{LOS}}} |y_{\text{DTFT}}(f_{\text{LOS}})|, \tag{2.37}$$

---

[4]See (2.4) for the definition of $\boldsymbol{\rho}$.

and an estimator of the scattering SNR s similar to (2.17):

$$s^{(0)} = \frac{(1/N^{(0)}) \cdot \sum_{k=0,\, z_k(\widehat{f}_{\mathrm{D}}^{\mathrm{HS}}) \neq 0}^{N-1} |y_{\mathrm{DTFT}}(f_k) - N^{-1/2} \cdot y_{\mathrm{DTFT}}(\widehat{f}_{\mathrm{LOS}}^{\mathrm{LS}}) \cdot \varphi_{\mathrm{DTFT}}(f_k; \widehat{f}_{\mathrm{LOS}}^{\mathrm{LS}})|^2 \,/\, z_k^2(\widehat{f}_{\mathrm{D}}^{\mathrm{HS}})}{[1/(N-N^{(0)})] \cdot \sum_{k=0,\, z_k(\widehat{f}_{\mathrm{D}}^{\mathrm{HS}}) = 0}^{N-1} |y_{\mathrm{DTFT}}(f_k) - N^{-1/2} \cdot y_{\mathrm{DTFT}}(\widehat{f}_{\mathrm{LOS}}^{\mathrm{LS}}) \cdot \varphi_{\mathrm{DTFT}}(f_k; \widehat{f}_{\mathrm{LOS}}^{\mathrm{LS}})|^2}, \tag{2.38}$$

where $N^{(0)}$ is the number of terms in the summation in the numerator of (2.38).

III. ESTIMATING JAKES' POWER SPECTRUM PARAMETERS IN SMART-ANTENNA SYSTEMS

Consider now a SIMO smart-antenna Rayleigh fading channel with n_{R} receiver antennas. Denote by $\boldsymbol{y}(t)$ an $n_{\mathrm{R}} \times 1$ vector of the complex fading channel estimates at time $t \in \{1, 2, \ldots, N\}$. We assume that $\boldsymbol{y}(t)$ are corrupted by spatially and temporally white circularly symmetric complex Gaussian noise with unknown variance σ^2 and that the noise is independent of the fading process. If the fading-channel components at all antennas share the same Doppler spread and the real and imaginary parts of the fading process are independent (see [23, App. A]), then the noisy $n_{\mathrm{R}} \times n_{\mathrm{R}}$ Jakes' cross-spectral matrix (CSM)[5] of $\boldsymbol{y}(t)$ can be written as

$$\boldsymbol{P_{yy}}(f; \boldsymbol{\theta}) = \sigma^2 \cdot \boldsymbol{G}(f; \boldsymbol{\rho}) = \sigma^2 \cdot \left[\frac{\boldsymbol{S} \cdot i_{[0, f_{\mathrm{D}}]}(f)}{(f_{\mathrm{D}}^2 - f^2)^{1/2}} + \frac{\boldsymbol{S} \cdot i_{(1-f_{\mathrm{D}}, 1]}(f)}{[f_{\mathrm{D}}^2 - (1-f)^2]^{1/2}} + \boldsymbol{I}_{n_{\mathrm{R}}} \right], \tag{3.1}$$

where $\boldsymbol{S}$ is the $n_{\mathrm{R}} \times n_{\mathrm{R}}$ (normalized) spatial fading covariance matrix and $\boldsymbol{\theta} = [\sigma^2, \boldsymbol{\rho}^T]^T$ with

$$\boldsymbol{\rho} = [f_{\mathrm{D}}, \ \boldsymbol{s}^T]^T. \tag{3.2}$$

Here, $\boldsymbol{s}$ describes a parametrization of the fading covariance matrix $\boldsymbol{S}$. [An extension of the above model to the MIMO scenario is straightforward. In the MIMO case, $\boldsymbol{y}(t)$ are $n_{\mathrm{R}} n_{\mathrm{T}} \times 1$ vectors of the estimated MIMO channel coefficients, and $\boldsymbol{P_{yy}}(f; \boldsymbol{\theta})$ and $\boldsymbol{S}$ are $n_{\mathrm{R}} n_{\mathrm{T}} \times n_{\mathrm{R}} n_{\mathrm{T}}$ matrices. For simplicity, we focus on the SIMO scenario in the following discussion.] We consider two models for $\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{s})$:

(i) **unstructured:** $\boldsymbol{s} = [\mathrm{Re}\{\mathrm{vech}(\boldsymbol{S})\}^T, \mathrm{Im}\{\underline{\mathrm{vech}}(\boldsymbol{S})\}^T]^T$ (the correlation structure of the fading channel is completely unknown),

(ii) **diagonal (independent fading):** $\boldsymbol{S} = \mathrm{diag}\{s_1, s_2, \ldots, s_{n_{\mathrm{R}}}\}$ and $\boldsymbol{s} = [s_1, s_2, \ldots, s_{n_{\mathrm{R}}}]^T$ (the fading-channel coefficients are independent with non-equal variances),

where the vech and $\underline{\mathrm{vech}}$ operators create a single column vector by stacking elements below the main diagonal columnwise; vech includes the main diagonal, whereas $\underline{\mathrm{vech}}$ omits it. Note that $\boldsymbol{s}$ is a valid parametrization *only* if $\boldsymbol{S}$ is a positive semidefinite Hermitian matrix. For notational simplicity, we do not explicitly specify the dependence of $\boldsymbol{S}$ on $\boldsymbol{s}$ in the following discussion. Note that the covariance matrix of the observations $\boldsymbol{y} = [\boldsymbol{y}(1)^T, \boldsymbol{y}(2)^T, \ldots, \boldsymbol{y}(N)^T]^T$ that corresponds to the CSM in (3.1) is not analytically tractable:

$$\mathrm{E}\left[\boldsymbol{y}\boldsymbol{y}^H\right] = \sigma^2 \cdot [\boldsymbol{J}(f_{\mathrm{D}}) \otimes \boldsymbol{S} + \boldsymbol{I}_{n_{\mathrm{R}} N}]. \tag{3.3}$$

---

[5]We define the CSM of a stationary zero-mean multivariate random process $\boldsymbol{y}(t)$ as $\boldsymbol{P_{yy}}(f) = \sum_{n=-\infty}^{\infty} \mathrm{E}\left[\boldsymbol{y}(t)\boldsymbol{y}(t+n)^H\right] \cdot \exp(-j2\pi f n)$.

Here, $\boldsymbol{J}(f_\mathrm{D})$ was defined in (2.9) and $\otimes$ denotes the Kronecker product. The multivariate Whittle approximation to the log-likelihood can be derived along the lines of [30] (see also [31, Ch. 13.7]):

$$
\begin{aligned}
l(\boldsymbol{y};\boldsymbol{\theta}) &= -Nn_\mathrm{R}\ln\pi - \sum_{k=0}^{N-1}\left\{\ln|\boldsymbol{P_{yy}}(f_k;\boldsymbol{\theta})| + \mathrm{tr}[\boldsymbol{C_{y},N}(f_k)\,\boldsymbol{P_{yy}}(f_k;\boldsymbol{\theta})^{-1}]\right\} \qquad\text{(3.4a)}\\
&= -Nn_\mathrm{R}\ln\pi - \sum_{k=0}^{N-1}\left\{\ln|\sigma^2\boldsymbol{G}(f_k,\boldsymbol{\rho})| + \mathrm{tr}[\boldsymbol{C_{y},N}(f_k)\,\boldsymbol{G}(f_k,\boldsymbol{\rho})^{-1}]/\sigma^2\right\}, \qquad\text{(3.4b)}
\end{aligned}
$$

where

$$
\boldsymbol{G}(f_k,\boldsymbol{\rho}) = z_k^2(f_\mathrm{D})\cdot\boldsymbol{S} + \boldsymbol{I}_{n_\mathrm{R}}, \qquad k = 0,1,\ldots,N-1, \tag{3.5}
$$

see (3.1). Here $|\cdot|$ denotes the determinant and $\boldsymbol{C_{y},N}(f)$ is the $n_\mathrm{R}\times n_\mathrm{R}$ periodogram matrix of the received data:

$$
\boldsymbol{C_{y},N}(f) = \frac{1}{N}\left(\sum_{t=1}^{N}\boldsymbol{y}(t)e^{-j2\pi ft}\right)\cdot\left(\sum_{t=1}^{N}\boldsymbol{y}(t)e^{-j2\pi ft}\right)^{H} = \boldsymbol{y}_\mathrm{DTFT}(f)\cdot\boldsymbol{y}_\mathrm{DTFT}(f)^{H}, \tag{3.6}
$$

where

$$
\boldsymbol{y}_\mathrm{DTFT}(f) = \frac{1}{\sqrt{N}}\cdot\sum_{t=1}^{N}\boldsymbol{y}(t)e^{-j2\pi ft} \tag{3.7}
$$

is the normalized DTFT of $\boldsymbol{y}(t)$, $t = 1,2,\ldots,N$. Hence, $\boldsymbol{y}_\mathrm{DTFT}(f_k)$, $k = 0,1,\ldots,N-1$ form the normalized discrete DFT of $\boldsymbol{y}(t)$, $t = 1,2,\ldots,N$. For fixed $\boldsymbol{\rho}$, there exists a closed-form expression for the asymptotic ML estimate of σ^2 which maximizes (3.4):

$$
\widehat{\sigma}^2(\boldsymbol{\rho}) = \frac{1}{n_\mathrm{R}N}\sum_{k=0}^{N-1}\mathrm{tr}\{\boldsymbol{C_{y},N}(f_k)\,\boldsymbol{G}(f_k;\boldsymbol{\rho})^{-1}\}. \tag{3.8}
$$

Substituting (3.8) into (3.4b) and neglecting constant terms yields the concentrated likelihood function:

$$
l_\mathrm{c}(\boldsymbol{y};\boldsymbol{\rho}) = -n_\mathrm{R}N\cdot\ln\left\{\sum_{k=0}^{N-1}\mathrm{tr}[\boldsymbol{C_{y},N}(f_k)\,\boldsymbol{G}(f_k;\boldsymbol{\rho})^{-1}]\right\} - \sum_{k=0}^{N-1}\ln\left|\boldsymbol{G}(f_k;\boldsymbol{\rho})\right| \tag{3.9}
$$

to be maximized with respect to $\boldsymbol{\rho}$.

In the following, we derive algorithms for computing the asymptotic ML estimates of the unknown parameters under the unstructured and independent fading scenarios.

A. Asymptotic ML Estimation for Unstructured Fading

We compute the asymptotic ML estimates of the unknown parameters for unstructured fading. We first present a parameter-expanded expectation-maximization (PX-EM) algorithm for computing the asymptotic ML estimates $\widehat{\sigma}^2(f_\mathrm{D})$ and $\widehat{\boldsymbol{S}}(f_\mathrm{D})$ of σ^2 and $\boldsymbol{S}$ when f_D is *known* and then propose its extension to the case where f_D is *unknown* (in addition to $\boldsymbol{S}$ and σ^2).

Known f_D**:** In Appendix B, we apply the parameter-expansion approach in [32] to derive the following PX-EM algorithm for estimating $\boldsymbol{s}$ and σ^2 when f_D is *known:* iterate between

Step 1:

$$\boldsymbol{W}_k^{(i)}(f_{\mathrm{D}}) = [(\sigma^2)^{(i)}(f_{\mathrm{D}}) \cdot \boldsymbol{I}_{n_{\mathrm{R}}} + z_k^2(f_{\mathrm{D}}) \cdot \boldsymbol{\Psi}^{(i)}(f_{\mathrm{D}})]^{-1}, \tag{3.10a}$$

$$\boldsymbol{h}_{\mathrm{a},k}^{(i)}(f_{\mathrm{D}}) = z_k(f_{\mathrm{D}}) \cdot \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}(f_{\mathrm{D}}) \, [\boldsymbol{A}^{(i)}(f_{\mathrm{D}})]^H \, \boldsymbol{W}_k^{(i)}(f_{\mathrm{D}}) \, \boldsymbol{y}_{\mathrm{DTFT}}(f_k), \tag{3.10b}$$

$$\boldsymbol{Q}_k^{(i)}(f_{\mathrm{D}}) = \boldsymbol{h}_{\mathrm{a},k}^{(i)}(f_{\mathrm{D}}) \cdot [\boldsymbol{h}_{\mathrm{a},k}^{(i)}(f_{\mathrm{D}})]^H - z_k^2(f_{\mathrm{D}}) \cdot \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}(f_{\mathrm{D}}) \, [\boldsymbol{A}^{(i)}(f_{\mathrm{D}})]^H \, \boldsymbol{W}_k^{(i)}(f_{\mathrm{D}}) \, \boldsymbol{A}^{(i)}(f_{\mathrm{D}}) \, \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}(f_{\mathrm{D}}), \tag{3.10c}$$

for $k = 1, 2, \ldots, N-1$ and

Step 2:

$$\boldsymbol{A}^{(i+1)}(f_{\mathrm{D}}) = \left\{ \sum_{k=0}^{N-1} z_k(f_{\mathrm{D}}) \cdot \boldsymbol{y}_{\mathrm{DTFT}}(f_k) \cdot [\boldsymbol{h}_{\mathrm{a},k}^{(i)}(f_{\mathrm{D}})]^H \right\} \cdot \left\{ \Big[\sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \Big] \cdot \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}(f_{\mathrm{D}}) + \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \boldsymbol{Q}_k^{(i)}(f_{\mathrm{D}}) \right\}^{-1}, \tag{3.11a}$$

$$\boldsymbol{\Psi}_{\mathrm{a}}^{(i+1)}(f_{\mathrm{D}}) = \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}(f_{\mathrm{D}}) + \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{Q}_k^{(i)}(f_{\mathrm{D}}), \tag{3.11b}$$

$$\boldsymbol{\Psi}^{(i+1)}(f_{\mathrm{D}}) = \boldsymbol{A}^{(i+1)}(f_{\mathrm{D}}) \, \boldsymbol{\Psi}_{\mathrm{a}}^{(i+1)}(f_{\mathrm{D}}) \, [\boldsymbol{A}^{(i+1)}(f_{\mathrm{D}})]^H, \tag{3.11c}$$

$$(\sigma^2)^{(i+1)}(f_{\mathrm{D}}) = \frac{1}{n_{\mathrm{R}} N} \sum_{k=0}^{N-1} \Big\{ [\boldsymbol{y}_{\mathrm{DTFT}}(f_k) - z_k(f_{\mathrm{D}}) \cdot \boldsymbol{A}^{(i)}(f_{\mathrm{D}}) \, \boldsymbol{h}_{\mathrm{a},k}^{(i)}(f_{\mathrm{D}})]^H \cdot [\boldsymbol{y}_{\mathrm{DTFT}}(f_k) - z_k(f_{\mathrm{D}}) \cdot \boldsymbol{A}^{(i)}(f_{\mathrm{D}}) \, \boldsymbol{h}_{\mathrm{a},k}^{(i)}(f_{\mathrm{D}})]$$

$$+ (\sigma^2)^{(i)}(f_{\mathrm{D}}) \cdot \mathrm{tr} \left[\boldsymbol{I}_{n_{\mathrm{R}}} - (\sigma^2)^{(i)}(f_{\mathrm{D}}) \cdot \boldsymbol{W}_k^{(i)}(f_{\mathrm{D}}) \right] \Big\}. \tag{3.11d}$$

The above iteration is performed until $\boldsymbol{\Psi}^{(i)}(f_{\mathrm{D}})$ and $(\sigma^2)^{(i)}(f_{\mathrm{D}})$ converge; denote by $\widehat{\boldsymbol{\Psi}}(f_{\mathrm{D}}) = \boldsymbol{\Psi}^{(\infty)}(f_{\mathrm{D}})$ and $\widehat{\sigma}^2(f_{\mathrm{D}}) = (\sigma^2)^{(\infty)}(f_{\mathrm{D}})$ the estimates of $\boldsymbol{\Psi}$ and σ^2 obtained upon convergence. Then, the asymptotic ML estimate of $\boldsymbol{S}$ is computed as

$$\widehat{\boldsymbol{S}}(f_{\mathrm{D}}) = \frac{\widehat{\boldsymbol{\Psi}}(f_{\mathrm{D}})}{\widehat{\sigma}^2(f_{\mathrm{D}})}, \tag{3.12}$$

see also (B.5) in Appendix B. The PX-EM algorithm shares the same monotonic convergence properties as the "classical" expectation-maximization (EM) algorithm, see [32, Theorem 1]. It also outperforms the EM algorithm in the global rate of convergence (see [32, Theorem 2]), where the performance improvement is particularly significant in the low-SNR scenarios, i.e. when σ^2 is large compared with the entries of $\boldsymbol{\Psi} = \sigma^2 \boldsymbol{S}$.

The iteration (3.10)–(3.11) can be initialized as follows:

$$\boldsymbol{A}^{(0)} = \boldsymbol{I}_{n_{\mathrm{R}}}, \tag{3.13a}$$

$$\boldsymbol{\Psi}^{(0)}(f_{\mathrm{D}}) = \boldsymbol{\Psi}_{\mathrm{a}}^{(0)}(f_{\mathrm{D}}) = \frac{1}{N^{(0)}} \cdot \sum_{\substack{k=0, \\ z_k(f_{\mathrm{D}}) \neq 0}}^{N-1} \frac{\boldsymbol{y}_{\mathrm{DTFT}}(f_k) \, \boldsymbol{y}_{\mathrm{DTFT}}(f_k)^H}{z_k^2(f_{\mathrm{D}})}, \tag{3.13b}$$

$$(\sigma^2)^{(0)} = \frac{1}{n_{\mathrm{R}}(N - N^{(0)})} \cdot \sum_{\substack{k=0, \\ z_k(f_{\mathrm{D}}) = 0}}^{N-1} \boldsymbol{y}_{\mathrm{DTFT}}(f_k)^H \boldsymbol{y}_{\mathrm{DTFT}}(f_k), \tag{3.13c}$$

where $N^{(0)}$ is the number of terms in the summation in (3.13b).

Unknown f_{D}**:** If f_{D} is unknown (in addition to $\boldsymbol{S}$ and σ^2), we propose the following *alternating-projection* algorithm for computing the asymptotic ML estimates of $f_{\mathrm{D}}, \boldsymbol{S}$, and σ^2:

Step 1: Fix $f_{\mathrm{D}} = \widehat{f}_{\mathrm{D}}$ and compute $\widehat{\boldsymbol{S}}(f_{\mathrm{D}})$ using (3.10)–(3.11) and (3.12);

Step 2: Fix $\boldsymbol{S} = \widehat{\boldsymbol{S}}(f_{\mathrm{D}})$ [see (3.12)] and find $\widehat{f}_{\mathrm{D}}$ that maximizes $l_{\mathrm{c}}(\boldsymbol{y}; [f_{\mathrm{D}}, \boldsymbol{s}^T]^T)$ in (3.9),

which increases the likelihood function (of the unknown parameters f_{D}, $\boldsymbol{S}$, and σ^2) at each iteration cycle.

The above iteration can be initialized using the following simple estimator of f_{D}:

$$\widehat{f}_{\mathrm{D}}^{\mathrm{HS}} = \frac{1}{2\pi} \cdot \sqrt{\frac{2 \sum_{t=1}^{N-1} \sum_{p=1}^{n_{\mathrm{R}}} |[\boldsymbol{y}(t)]_p - [\boldsymbol{y}(t+1)]_p|^2}{\sum_{t=1}^{N-1} \sum_{p=1}^{n_{\mathrm{R}}} |[\boldsymbol{y}(t)]_p|^2}}, \tag{3.14}$$

which generalizes the sample-covariance-based method in (2.16) to the SIMO scenario.

B. Asymptotic ML Estimation for Independent Fading

We compute the asymptotic ML estimates of the unknown parameters $\boldsymbol{\theta}$ for independent fading where $\boldsymbol{S} = \mathrm{diag}(s_1, s_2, \ldots, s_{n_{\mathrm{R}}})$. As in Section III-A, we first propose a method for estimating $\boldsymbol{S}$ and σ^2 for *known* f_{D} and then generalize it for an unknown f_{D}.

Known f_{D}**:** The PX-EM algorithm for estimating $\boldsymbol{s} = [s_1, s_2, \ldots, s_{n_{\mathrm{R}}}]^T$ and σ^2 when f_{D} is *known* easily follows (see Appendix B): iterate between

Step 1:

$$w_{k,p}^{(i)}(f_{\mathrm{D}}) = [(\sigma^2)^{(i)}(f_{\mathrm{D}}) + z_k^2(f_{\mathrm{D}}) \cdot \psi^{(i)}(f_{\mathrm{D}})]^{-1}, \tag{3.15a}$$

$$h_{\mathrm{a},k,p}^{(i)}(f_{\mathrm{D}}) = z_k(f_{\mathrm{D}}) \cdot \psi_{\mathrm{a},p}^{(i)}(f_{\mathrm{D}}) \cdot [a_p^{(i)}(f_{\mathrm{D}})]^* \cdot w_{k,p}^{(i)}(f_{\mathrm{D}}) \cdot [\boldsymbol{y}_{\mathrm{DTFT}}(f_k)]_p, \tag{3.15b}$$

$$q_{k,p}^{(i)}(f_{\mathrm{D}}) = |h_{\mathrm{a},k,p}^{(i)}(f_{\mathrm{D}})|^2 - z_k^2(f_{\mathrm{D}}) \cdot [\psi_{\mathrm{a}}^{(i)}(f_{\mathrm{D}})]^2 \cdot |a_p^{(i)}(f_{\mathrm{D}})|^2 \cdot w_{k,p}^{(i)}(f_{\mathrm{D}}), \tag{3.15c}$$

for $k = 1, 2, \ldots, N-1$, $p = 1, 2, \ldots, n_{\mathrm{R}}$ and

Step 2:

$$a_p^{(i+1)}(f_{\mathrm{D}}) = \frac{\sum_{k=0}^{N-1} z_k(f_{\mathrm{D}}) \cdot [\boldsymbol{y}_{\mathrm{DTFT}}(f_k)]_p \cdot [h_{\mathrm{a},k,p}^{(i)}(f_{\mathrm{D}})]^*}{\left[\sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}})\right] \cdot \psi_{\mathrm{a},p}^{(i)}(f_{\mathrm{D}}) + \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot q_{k,p}^{(i)}(f_{\mathrm{D}})}, \tag{3.16a}$$

$$\psi_{\mathrm{a},p}^{(i+1)}(f_{\mathrm{D}}) = \psi_{\mathrm{a},p}^{(i)}(f_{\mathrm{D}}) + \frac{1}{N} \sum_{k=0}^{N-1} q_{k,p}^{(i)}(f_{\mathrm{D}}), \tag{3.16b}$$

$$\psi_p^{(i+1)}(f_{\mathrm{D}}) = |a_p^{(i)}(f_{\mathrm{D}})|^2 \cdot \psi_{\mathrm{a},p}^{(i)}(f_{\mathrm{D}}), \tag{3.16c}$$

for $p = 1, 2, \ldots, n_{\mathrm{R}}$ and

$$(\sigma^2)^{(i+1)}(f_{\mathrm{D}}) = \frac{1}{n_{\mathrm{R}} N} \sum_{k=0}^{N-1} \sum_{p=1}^{n_{\mathrm{R}}} \left\{ \left| [\boldsymbol{y}_{\mathrm{DTFT}}(f_k)]_p - z_k(f_{\mathrm{D}}) \cdot a_p^{(i)}(f_{\mathrm{D}}) \cdot h_{\mathrm{a},k,p}^{(i)}(f_{\mathrm{D}}) \right|^2 + (\sigma^2)^{(i)}(f_{\mathrm{D}}) \cdot [1 - (\sigma^2)^{(i)}(f_{\mathrm{D}}) \cdot w_{k,p}^{(i)}(f_{\mathrm{D}})] \right\}. \tag{3.16d}$$

Here "*" denotes complex conjugation and $[\boldsymbol{y}_{\mathrm{DTFT}}(f_k)]_p$ the pth element of $\boldsymbol{y}_{\mathrm{DTFT}}(f_k)$. The above iteration is performed until $\psi_p^{(i)}(f_{\mathrm{D}})$, $p = 1, 2, \ldots, n_{\mathrm{R}}$ and $(\sigma^2)^{(i)}(f_{\mathrm{D}})$ converge to $\widehat{\psi}_p(f_{\mathrm{D}}) = \psi_p^{(\infty)}(f_{\mathrm{D}})$, $p = 1, 2, \ldots, n_{\mathrm{R}}$ and $\widehat{\sigma}^2(f_{\mathrm{D}}) = (\sigma^2)^{(\infty)}(f_{\mathrm{D}})$. Then, the asymptotic ML estimate of $\boldsymbol{S}$ is computed as

$$\widehat{\boldsymbol{S}}(f_{\mathrm{D}}) = \mathrm{diag}\left\{ \frac{\widehat{\psi}_1(f_{\mathrm{D}})}{\widehat{\sigma}^2(f_{\mathrm{D}})}, \frac{\widehat{\psi}_2(f_{\mathrm{D}})}{\widehat{\sigma}^2(f_{\mathrm{D}})}, \ldots, \frac{\widehat{\psi}_{n_{\mathrm{R}}}(f_{\mathrm{D}})}{\widehat{\sigma}^2(f_{\mathrm{D}})} \right\}, \tag{3.17}$$

see also (B.18) in Appendix B. In analogy with (3.13), we can initialize the iteration (3.15)–(3.16) using

$$a_p^{(0)} = 1, \tag{3.18a}$$

$$\psi_p^{(0)}(f_{\mathrm{D}}) = \psi_{\mathrm{a},p}^{(0)}(f_{\mathrm{D}}) = \frac{1}{N^{(0)}} \cdot \sum_{\substack{k=0, \\ z_k(f_{\mathrm{D}}) \neq 0}}^{N-1} \frac{|[\boldsymbol{y}_{\mathrm{DTFT}}(f_k)]_p|^2}{z_k^2(f_{\mathrm{D}})}, \tag{3.18b}$$

$$(\sigma^2)^{(0)} = \frac{1}{n_{\mathrm{R}}(N - N^{(0)})} \cdot \sum_{\substack{k=0, \\ z_k(f_{\mathrm{D}}) = 0}}^{N-1} \boldsymbol{y}_{\mathrm{DTFT}}(f_k)^H \boldsymbol{y}_{\mathrm{DTFT}}(f_k), \tag{3.18c}$$

for $p = 1, 2, \ldots, n_{\mathrm{R}}$, where $N^{(0)}$ is the number of terms in the summation in (3.18b).

Unknown f_{D}**:** If f_{D} is unknown, we can estimate it using (3.14) when N is small. For large N, (3.14) can be used to initialize the *alternating-projection* algorithm for asymptotic ML estimation of $f_{\mathrm{D}}, \boldsymbol{S}$, and σ^2:

Step 1: Fix $f_{\mathrm{D}} = \widehat{f}_{\mathrm{D}}$ and compute $\widehat{\boldsymbol{S}}(f_{\mathrm{D}})$ using (3.15)–(3.16) and (3.17),

Step 2: Fix $\boldsymbol{S} = \widehat{\boldsymbol{S}}(f_{\mathrm{D}})$ [see (3.17)] and find $\widehat{f}_{\mathrm{D}}$ that maximizes $l_c(\boldsymbol{y}; [f_{\mathrm{D}}, \boldsymbol{s}^T]^T)$ in (3.9),

which is similar to the asymptotic ML algorithm for unstructured $\boldsymbol{S}$ and unknown f_{D} in Section III-A.

C. Asymptotic Cramér-Rao Bound

The asymptotic CRB for the unknown parameters $\boldsymbol{\theta}$ is the inverse of the asymptotic FIM $\mathcal{I}(\boldsymbol{\theta})$, see (2.18). The (p,q) element of the asymptotic FIM for $p = 1, 2 \ldots, \dim(\boldsymbol{\theta})$, $q = 1, 2 \ldots, \dim(\boldsymbol{\theta})$ is computed as

$$\mathcal{I}_{p,q}(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} \mathrm{tr}\left[\boldsymbol{P}_{\boldsymbol{yy}}(f_k; \boldsymbol{\theta})^{-1} \cdot \frac{\partial \boldsymbol{P}_{\boldsymbol{yy}}(f_k; \boldsymbol{\theta})}{\partial \theta_p} \cdot \boldsymbol{P}_{\boldsymbol{yy}}(f_k; \boldsymbol{\theta})^{-1} \cdot \frac{\partial \boldsymbol{P}_{\boldsymbol{yy}}(f_k; \boldsymbol{\theta})}{\partial \theta_q} \right], \tag{3.19}$$

which follows by adapting the results of [30] to the complex data model. For $p = 1$ and the Jakes' spectrum model in (3.1), the above expression simplifies to (3.20):

$$\mathcal{I}_{1,q}(\boldsymbol{\theta}) = [\mathcal{I}]_{\sigma^2, \theta_q} = \sum_{k=0}^{N-1} \mathrm{tr}\left\{ [\sigma^2 \boldsymbol{G}(f_k, \boldsymbol{\rho})]^{-1} \frac{\partial[\sigma^2 \boldsymbol{G}(f_k, \boldsymbol{\rho})]}{\partial \sigma^2} [\sigma^2 \boldsymbol{G}(f_k, \boldsymbol{\rho})]^{-1} \frac{\partial[\sigma^2 \boldsymbol{G}(f_k, \boldsymbol{\theta})]}{\partial \theta_q} \right\}$$

$$= \frac{1}{\sigma^4} \sum_{k=0}^{N-1} \mathrm{tr}\left\{ [\boldsymbol{G}(f_k, \boldsymbol{\rho})]^{-1} \frac{\partial[\sigma^2 \boldsymbol{G}(f_k, \boldsymbol{\rho})]}{\partial \theta_q} \right\} = \begin{cases} N n_{\mathrm{R}}/\sigma^4, & q = 1 \\ -(f_{\mathrm{D}}/\sigma^2) \cdot \sum_{k=0}^{N-1} z_k^6(f_{\mathrm{D}}) \cdot \mathrm{tr}\left[\boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \boldsymbol{S} \right], & q = 2 \\ (1/\sigma^2) \cdot \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \mathrm{tr}\left[\boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \cdot \partial \boldsymbol{S}/\partial s_{q-2} \right], & q > 2 \end{cases}. \tag{3.20}$$

Similarly, for $p = 2$ and $q \geq 2$, we obtain (3.21):

$$\mathcal{I}_{2,q}(\boldsymbol{\theta}) = [\mathcal{I}]_{f_{\mathrm{D}}, \theta_q} = \sum_{k=0}^{N-1} \mathrm{tr}\left\{ \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \frac{\partial \boldsymbol{G}(f_k, \boldsymbol{\rho})}{\partial f_{\mathrm{D}}} \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \frac{\partial \boldsymbol{G}(f_k, \boldsymbol{\theta})}{\partial \theta_q} \right\}$$

$$= -f_{\mathrm{D}} \sum_{k=0}^{N-1} z_k^6(f_{\mathrm{D}}) \mathrm{tr}\left\{ \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \boldsymbol{S} \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \frac{\partial \boldsymbol{G}(f_k, \boldsymbol{\theta})}{\partial \theta_q} \right\}$$

$$= \begin{cases} (f_{\mathrm{D}})^2 \cdot \sum_{k=0}^{N-1} z_k^{12}(f_{\mathrm{D}}) \cdot \mathrm{tr}\left\{ \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \boldsymbol{S} \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \boldsymbol{S} \right\}, & q = 2 \\ -f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} z_k^8(f_{\mathrm{D}}) \cdot \mathrm{tr}\left[\boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \boldsymbol{S} \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \cdot \partial \boldsymbol{S}/\partial s_{q-2} \right], & q > 2 \end{cases}. \tag{3.21}$$

For $p, q \geq 3$, we have (3.22):

$$
\begin{aligned}
\mathcal{I}_{p,q}(\boldsymbol{\theta}) &= [\mathcal{I}]_{s_{p-2},s_{q-2}} = \sum_{k=0}^{N-1} \operatorname{tr} \left\{ \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \frac{\partial \boldsymbol{G}(f_k, \boldsymbol{\rho})}{\partial s_{p-2}} \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \frac{\partial \boldsymbol{G}(f_k, \boldsymbol{\theta})}{\partial s_{q-2}} \right\} \\
&= \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \operatorname{tr} \left\{ \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \frac{\partial \boldsymbol{S}}{\partial s_{p-2}} \boldsymbol{G}(f_k, \boldsymbol{\rho})^{-1} \frac{\partial \boldsymbol{S}}{\partial s_{q-2}} \right\}.
\end{aligned}
\tag{3.22}
$$

To derive (3.20)–(3.22), we used (3.5) and the identity (2.21). See Appendix C for further simplifications of these expressions under the unstructured and independent fading scenarios.

Using the same block partitioning as in (2.23a), we have $\mathcal{I}_{\sigma^2,\sigma^2}(\sigma^2) = \mathcal{I}_{1,1}(\boldsymbol{\theta}) = Nn_{\mathrm{R}}/\sigma^4$, and $\mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})$ and $\mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\theta})$ can be constructed using (3.20)–(3.22). Then, the formula for the inverse of a partitioned matrix [28, Theorem 8.5.11] yields

$$
\mathbf{CRB}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho}) = \left[\mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho}) - \frac{\sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta}) \cdot \sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})^T}{Nn_{\mathrm{R}}} \right]^{-1},
\tag{3.23a}
$$

$$
\mathrm{CRB}_{\sigma^2 \sigma^2}(\boldsymbol{\theta}) = \frac{\sigma^4}{Nn_{\mathrm{R}} - \sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})^T \cdot \mathcal{I}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho})^{-1} \cdot \sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})}.
\tag{3.23b}
$$

As in the SISO case, $\sigma^2 \mathcal{I}_{\boldsymbol{\rho},\sigma^2}(\boldsymbol{\theta})$ and $\mathbf{CRB}_{\boldsymbol{\rho},\boldsymbol{\rho}}(\boldsymbol{\rho})$ are functions of $\boldsymbol{\rho}$ only, implying that the asymptotic accuracy of (efficiently) estimating the maximum Doppler frequency f_{D} and normalized spatial fading covariance parameters $\boldsymbol{s}$ depends on the unknown parameters only through f_{D} and $\boldsymbol{s}$. Furthermore, (3.23a) implies that $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ is a function of f_{D} through $z_k(f_{\mathrm{D}})$, $k = 0, 1, \ldots, N-1$. Also, $\mathrm{CRB}_{\sigma^2 \sigma^2}(\boldsymbol{\theta})$ depends on σ^2 only through σ^4 in the numerator of (3.23b) and is a function of f_{D} through $z_k(f_{\mathrm{D}})$, $k = 0, 1, \ldots, N-1$.

As in Section II-A, we assume that the sampling frequencies f_k, $k = 0, 1, \ldots, N-1$ do not coincide with f_{D} or $1 - f_{\mathrm{D}}$ since the Jakes' CSM in (3.1) is infinite at $f = f_{\mathrm{D}}$ and $f = 1 - f_{\mathrm{D}}$.

IV. NUMERICAL EXAMPLES

We evaluate the performance of the proposed methods using numerical simulations. Our performance metric is the mean-square error (MSE) of an estimator, calculated using 400 independent trials. In all the examples, we have chosen unit noise variance: $\sigma^2 = 1$. The noise variance is assumed to be unknown, unless specified otherwise (see Fig. 4).

SISO Rayleigh-fading Scenario: In the first set of simulations, we examine the MSE performances of

- the asymptotic (Whittle) ML estimators of the unknown parameters of interest (f_{D} and s) and
- the approximate ML and sample-covariance-based estimators of f_{D} in (2.15) and (2.16), respectively.

Simulated data was generated using the Jakes' correlation model in (2.8); in particular, we simulated the measurement vector $\boldsymbol{y}$ by premultiplying a white unit-variance complex Gaussian vector by a square root of the Jakes' covariance matrix in (2.8). In this scenario, we also generated simulated data using the sum of complex exponentials (as in e.g. [2, eq. (3)] or [20, eq. (2)]), corrupted by additive white circularly symmetric complex Gaussian noise. The obtained MSE results were almost identical to the results reported here.
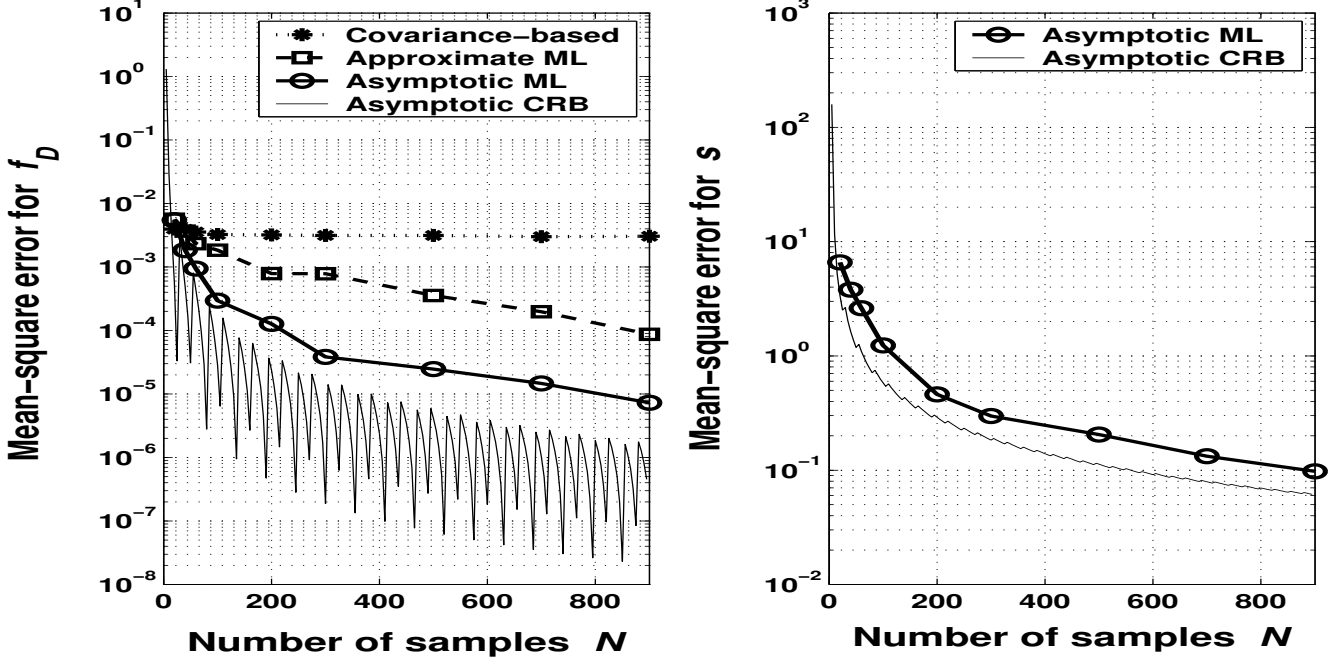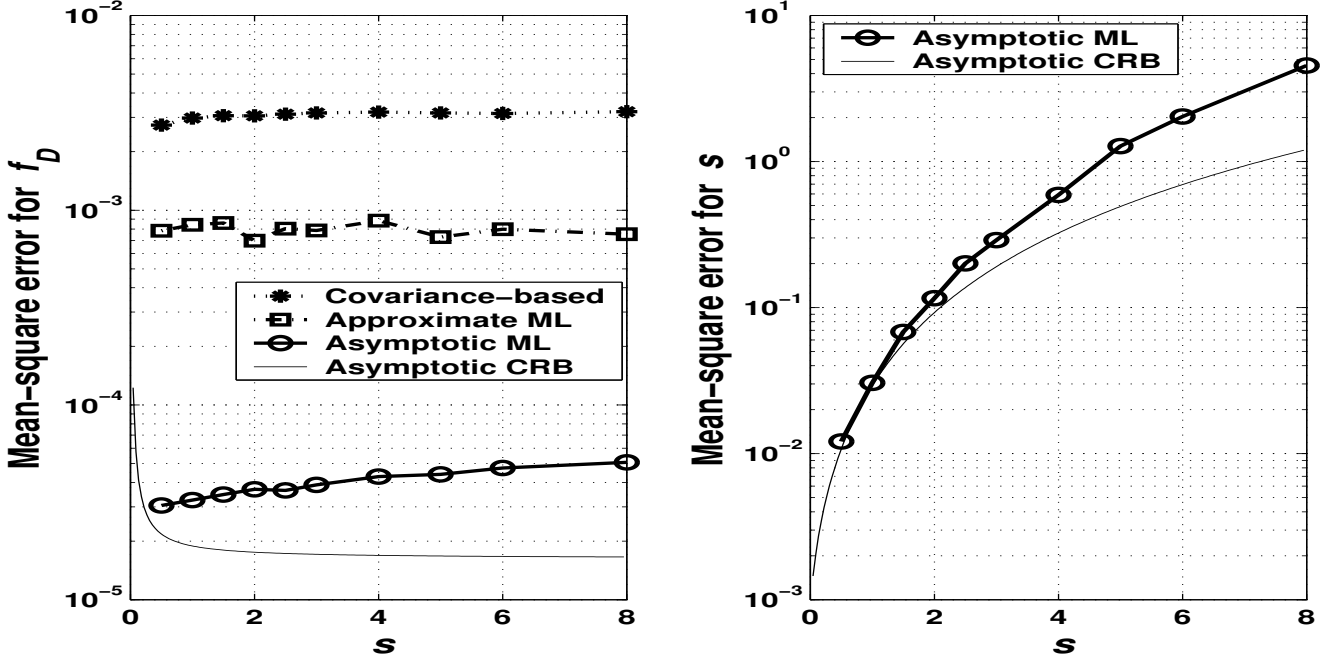
Fig. 1. Mean-square errors and asymptotic Cramér-Rao bounds for the asymptotic ML, approximate ML, and sample-covariance-based estimators of f_D (left) and asymptotic ML estimator of s (right) as functions of N, for $f_D = 0.365$ and $s = 3$.

In Fig. 1, we show the MSEs (and corresponding asymptotic CRBs) for the above estimators as functions of the number of samples N. The maximum Doppler frequency and SNR were fixed and set to $f_D = 0.365$ and $s = 3$. In this scenario, the asymptotic ML estimator of f_D clearly outperforms the approximate ML and sample-covariance-based methods, see Fig. 1 (left). Here, the asymptotic ML estimator of f_D achieves *excellent* performance for $N < 100$ samples, compared with 500 samples needed for the approximate ML method. We consider the accuracy of estimating the maximum Doppler frequency to be excellent if it is approximately within 6% of the true value, which corresponds to $\mathrm{MSE}_{\mathrm{excellent}} = 0.06^2 \cdot f_D^2 \approx 5 \cdot 10^{-4}$. This choice is consistent with the best performance in [19, Fig. 5], see also [1, Ch. 12].) Note that the sample-covariance-based method fails to reach $\mathrm{MSE}_{\mathrm{excellent}}$. The MSE for the asymptotic ML estimate of s in Fig. 1 (right) is close to the corresponding $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ for all values of N and is approximately proportional to $1/N$, as predicted in Section II-A.

In Fig. 2, we present the MSEs of the above estimators as functions of s, for fixed $f_D = 0.365$ and $N = 300$. When $s > 1$, $\mathrm{CRB}_{f_D,f_D}(\boldsymbol{\rho})$ is approximately independent of s whereas $\mathrm{CRB}_{s,s}(\boldsymbol{\rho})$ increases with s proportionally to s^2, see also Section II-A. Clearly, f_D is not identifiable when $s = 0$, which explains the sharp increase in $\mathrm{CRB}_{f_D,f_D}(\boldsymbol{\rho})$ as s decreases toward zero.

Figures 3 and 4 show the MSEs and asymptotic CRBs for f_D and s (respectively) as functions of f_D, for $s = 3$ and $N = 300$. In this scenario, the approximate ML estimator (2.15) outperforms the asymptotic ML method for $f_D < 0.25$, see Fig. 3. However, the approximate ML method performs poorly when f_D is large, in contrast to the asymptotic ML estimator which is (approximately) insensitive to the choice of f_D. Motivated by the discussion on

Fig. 2.   Mean-square errors and asymptotic Cramér-Rao bounds for the asymptotic ML, approximate ML, and sample-covariance-based estimators of f_D (left) and asymptotic ML estimator of s (right) as functions of s, for $f_D = 0.365$ and $N = 300$.

the identifiability of s in Section II-A, we now study the performances of the asymptotic ML estimators of s for both *unknown* and *known* noise level σ^2, see Fig. 4. For unknown σ^2, the estimation of s deteriorates as f_D approaches 0.5, as predicted by the approximate CRB results in Section II-A. In contrast, for known σ^2, the estimation of s improves as f_D increases[6]. The MSEs for the asymptotic ML estimates of s are close to the corresponding CRBs under both scenarios.

To compute the asymptotic ML estimates of f_D and s and approximate ML estimates of f_D, we utilized the *Nelder-Mead simplex method*[7] [33, Ch. 10.4], which converged in 24 iterations (on average).

SISO Ricean-fading Scenario: We now analyze the performances of the Rayleigh-fading based estimators of f_D and s and the Ricean-fading based asymptotic ML estimators of these parameters (Section II-B) under the Ricean fading scenario. Simulated data was generated using the model in (2.30). Figs. 5 and 6 show the MSEs of the above estimators as functions of the Ricean K factor $K = |x|^2/(\pi\sigma^2 s)$, defined as the ratio of the powers of the line-of-sight and scattering (diffuse) channel components [1, Ch. 2.1.2.2], [11, Ch. 5.6.2]. Here, we set the scattering SNR to $s = 3$ and consider two choices of the maximum Doppler frequency: $f_D = 0.265$ (Fig. 5) and $f_D = 0.365$ (Fig. 6). For $K = 0$ (Rayleigh fading), the approximate ML method for estimating f_D outperforms other methods when $f_D = 0.265$; however, it is outperformed by both the Rayleigh- and Ricean-fading based asymptotic ML estimators when $f_D = 0.365$, which is also consistent with the results in Fig. 3. For large K, the approximate ML and Rayleigh-

---

[6]For known σ^2, the asymptotic ML estimates of s and f_D are obtained by maximizing (2.5b).

[7]The simplex method was implemented using MATLAB's `fminsearch` function and initialized using (2.16) and (2.17).
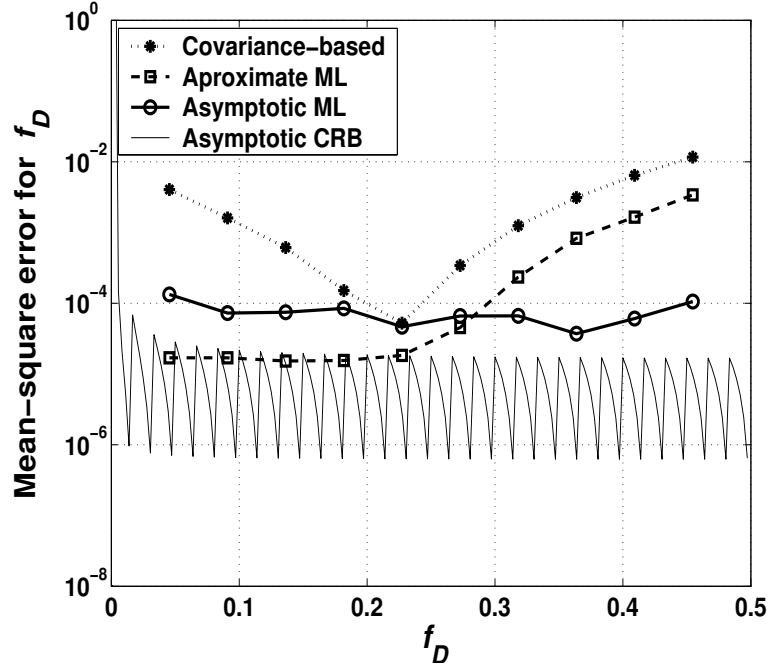
Fig. 3. Mean-square errors and asymptotic Cramér-Rao bounds for the asymptotic ML, approximate ML, and sample-covariance-based estimators of f_{D} as functions of f_{D}, for $s = 3$ and $N = 300$.
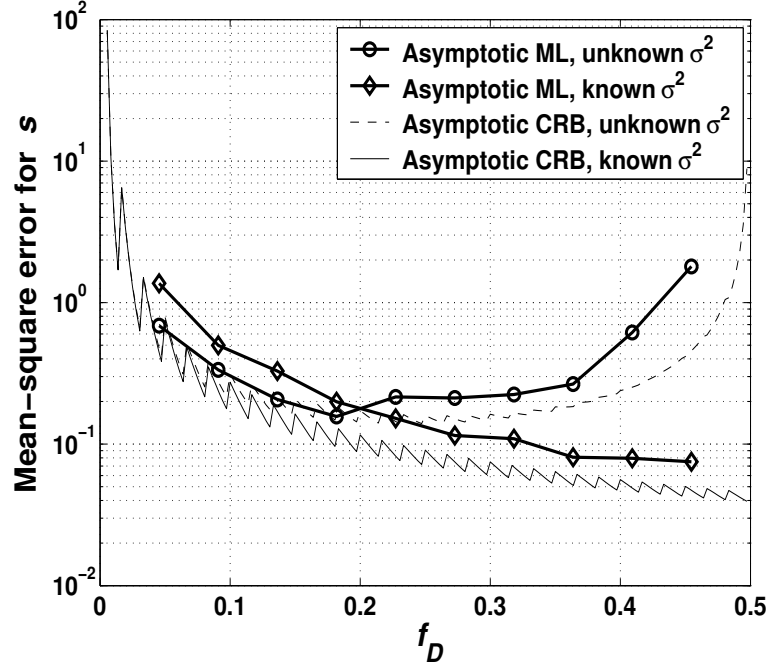

Fig. 4. Mean-square errors and asymptotic Cramér-Rao bounds for the asymptotic ML estimator of s (unknown σ^2) and asymptotic ML estimator of s for known σ^2 as functions of f_{D}, for $s = 3$ and $N = 300$.
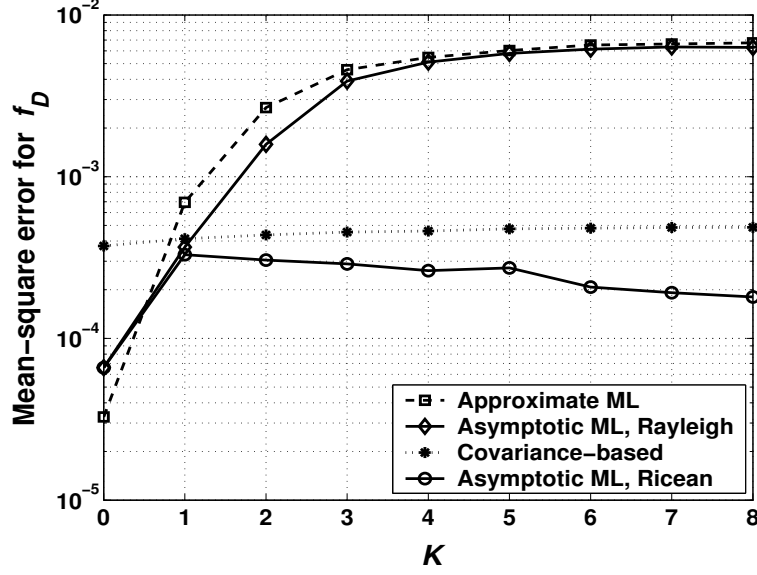
Fig. 5. Mean-square errors and asymptotic Cramér-Rao bounds for the Rayleigh- and Ricean-fading based estimators of f_D as functions of K, for $s = 3$ and $f_\mathrm{D} = 0.265$.
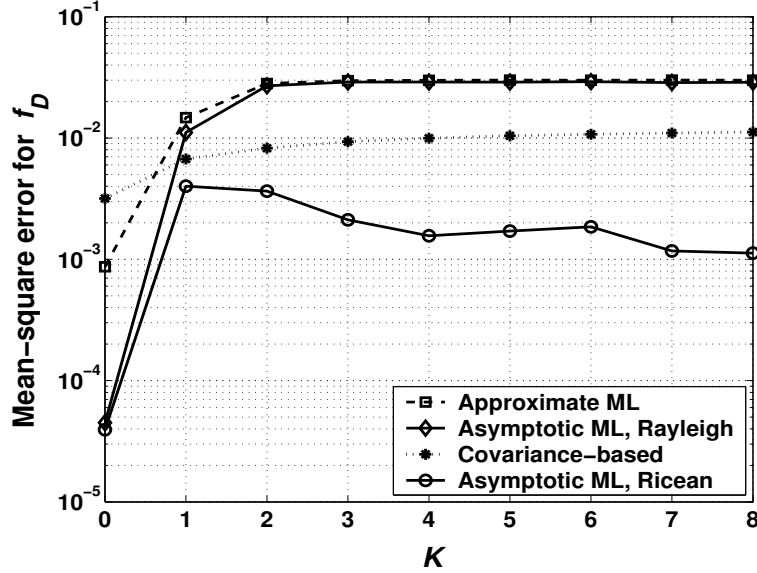
Fig. 6. Mean-square errors and asymptotic Cramér-Rao bounds for the Rayleigh- and Ricean-fading based estimators of f_D as functions of K, for $s = 3$ and $f_\mathrm{D} = 0.365$.

fading based asymptotic ML methods perform poorly. Interestingly, the sample-covariance-based estimator (2.16) is quite robust to the presence of the line-of-sight component. As expected, the best overall performance is achieved by the Ricean-fading based asymptotic ML method.

SIMO Rayleigh-fading Scenario: Consider the SIMO Rayleigh-fading scenario in Section III with the maximum Doppler frequency and number of receiver antennas set to $f_\mathrm{D} = 0.265$ and $n_\mathrm{R} = 2$. Simulated data was generated using the Jakes' correlation model for SIMO channels in (3.3); in particular, we simulated the measurement vector

y by premultiplying a white unit-variance complex Gaussian vector by a square root of the Jakes' covariance matrix in (3.3). We first examine the asymptotic ML method for correlated fading, where the normalized spatial fading covariance matrix was chosen as follows:

$$\boldsymbol{S} = \begin{bmatrix} s_{11} & s_{12} \\ s_{12}^* & s_{22} \end{bmatrix} = \begin{bmatrix} 0.33 & 0.17 - j0.1 \\ 0.17 + j0.1 & 0.5 \end{bmatrix}. \tag{4.1}$$

In Figs. 7–9, we show the MSEs (and corresponding asymptotic CRBs) for the asymptotic ML estimates of the unknown parameters f_{D} and $\boldsymbol{S}$, under the correlated fading scenario (see Section III-A), as functions of the number of samples N. Fig. 7 also compares the MSE performance of the asymptotic ML estimator of f_{D} with the sample-covariance-based estimator in (3.14) and approximate ML method:

$$\widehat{f}_{\mathrm{D}}^{\,\mathrm{app}} = \arg\max_{f_{\mathrm{D}}} \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \mathrm{tr}[\boldsymbol{C}_{\boldsymbol{y},N}(f_k)], \tag{4.2}$$

which generalizes (2.15) to the SIMO scenario, see also Appendix A. As expected, the asymptotic ML estimator outperforms the approximate ML and sample-covariance-based methods for large N. In this example, the sample-covariance-based estimator outperforms the asymptotic and approximate ML methods when N is small (less than 100). Simplicity and good performance for small numbers of observations are important when fast computation of f_{D} is needed (e.g. in adaptive modulation schemes, see [2]); then the sample-covariance-based estimator (3.14) may be the method of choice. As in the SISO case, the MSEs for the asymptotic ML estimates of (the elements of) $\boldsymbol{S}$ are close to the corresponding asymptotic CRBs for all values of N and are approximately proportional to $1/N$, see Figs. 8 and 9. The PX-EM algorithm for estimating $\boldsymbol{S}$ and σ^2 in (3.10)–(3.11) converged in 40 iterations (on average). The estimation of f_{D} was performed using the Nelder-Mead simplex method, which converged in less than 20 iterations.

We now consider the independent fading scenario with $\boldsymbol{S} = \mathrm{diag}\{s_1, s_2\} = \mathrm{diag}\{0.33, 0.5\}$. In Figs. 10 and 11, we show the MSEs (and corresponding asymptotic CRBs) for the asymptotic ML estimates of the unknown parameters f_{D} and s_1, s_2, (respectively) under the independent fading scenario (see Section III-B), as functions of the number of samples N. Fig. 10 also shows the MSE performances of the sample-covariance-based method in (3.14) and asymptotic ML estimator (4.2). The asymptotic ML estimator outperforms the approximate ML and sample-covariance-based methods for large N, whereas the sample-covariance-based method outperforms the asymptotic and approximate ML methods when N is less than 100. As in the SISO and unstructured SIMO fading scenarios, the MSEs for the asymptotic ML estimates of s_1 and s_2 are close to the corresponding asymptotic CRBs for all values of N and are approximately proportional to $1/N$, see Fig. 11.

The PX-EM algorithm for estimating $\boldsymbol{S}$ and σ^2 [see (3.15)–(3.16)] converged in less than 25 iterations. The asymptotic ML estimation of f_{D} was performed using the Nelder-Mead simplex method, which converged in less than 20 iterations.
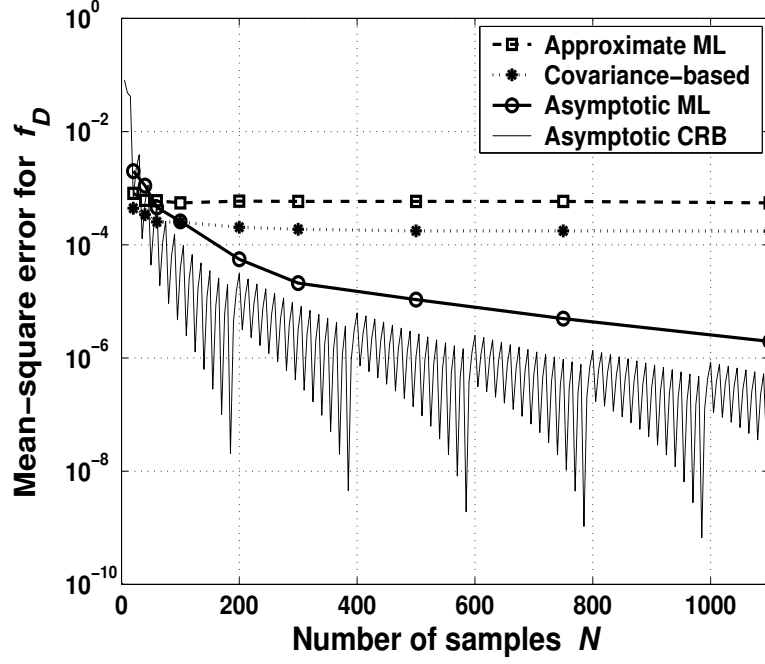
Fig. 7. Mean-square errors and asymptotic Cramér-Rao bounds for the asymptotic ML, sample-covariance-based, and approximate ML estimators of f_D in a SIMO system with correlated fading, as functions of N.
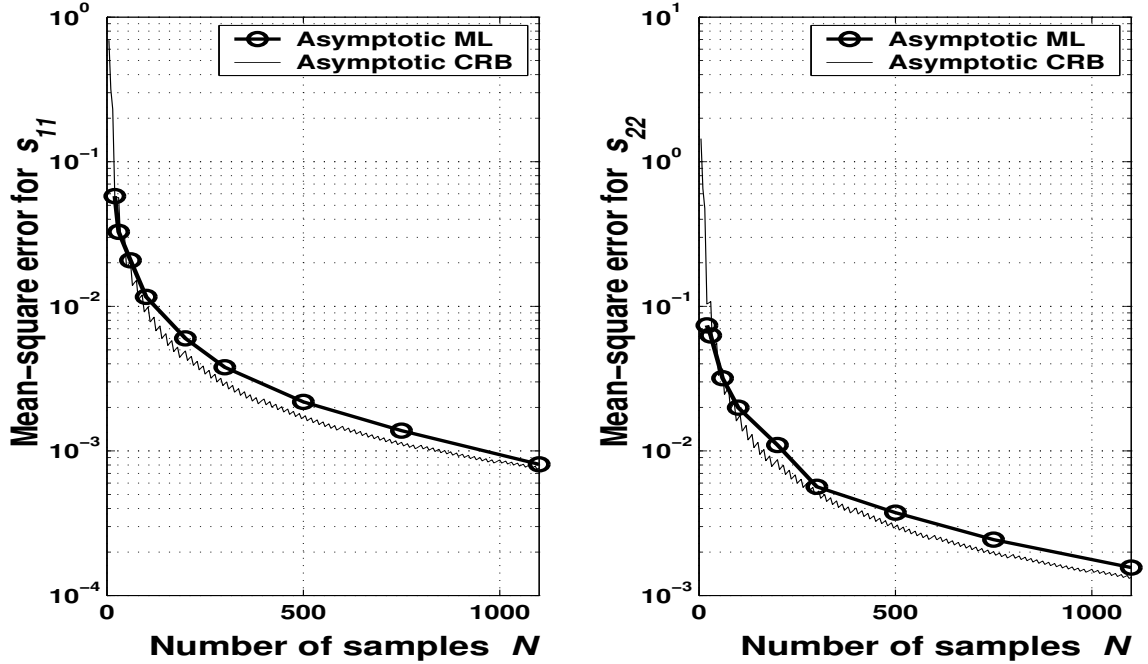

Fig. 8. Mean-square errors and corresponding asymptotic Cramér-Rao bounds for the asymptotic ML estimates of s_{11} and s_{22} in a SIMO system with correlated fading, as functions of N.
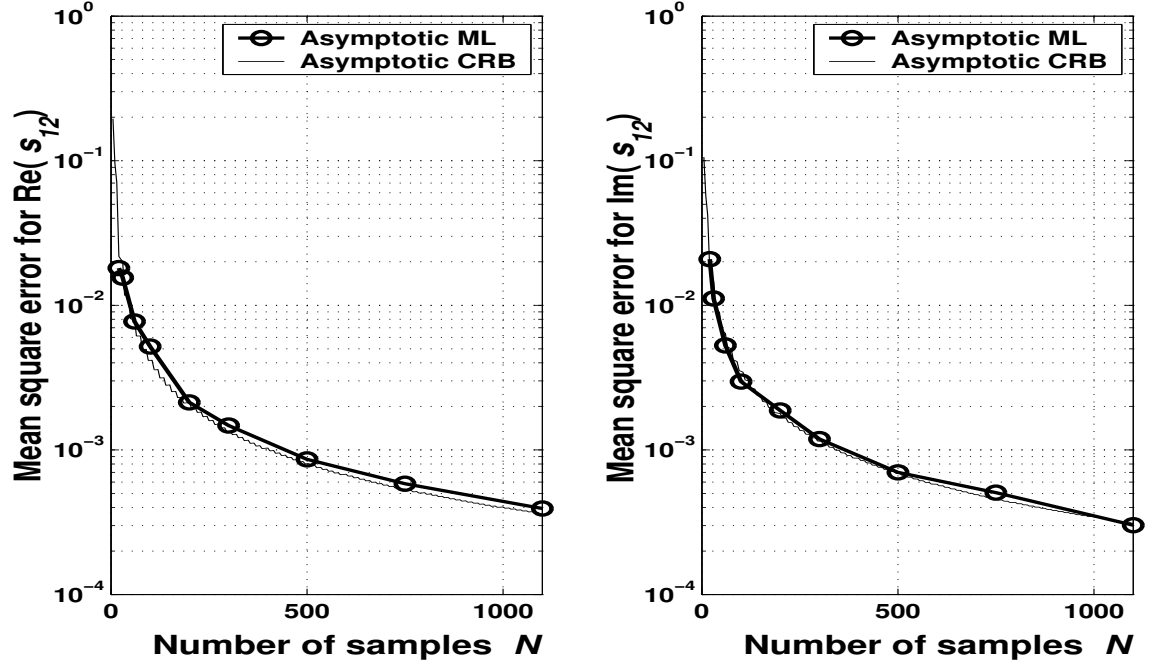
Fig. 9. Mean-square errors and corresponding asymptotic Cramér-Rao bounds for the asymptotic ML estimates of $\mathrm{Re}\{s_{12}\}$ and $\mathrm{Im}\{s_{12}\}$ in a SIMO system with correlated fading, as functions of N.
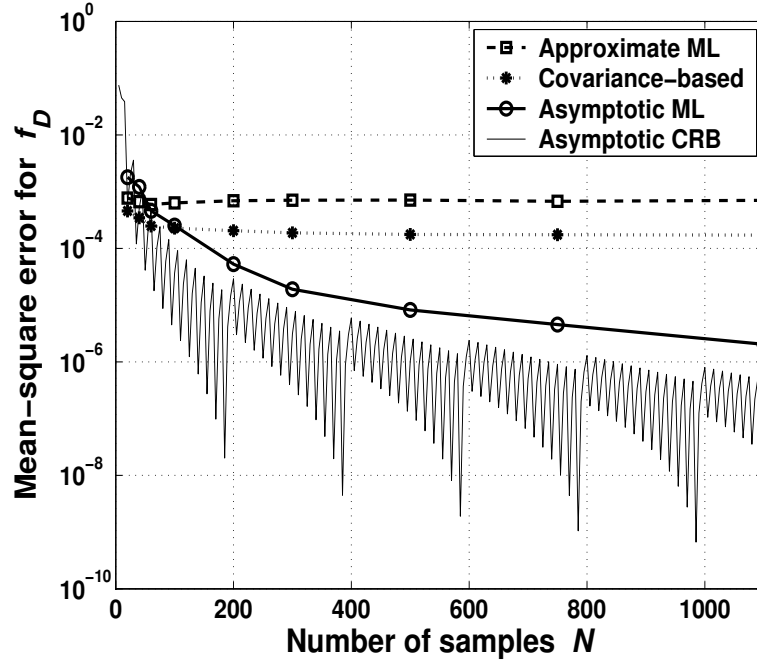

Fig. 10. Mean-square errors and asymptotic Cramér-Rao bounds for the asymptotic ML and sample-covariance-based estimators of f_{D} in a SIMO system with independent fading, as functions of N.
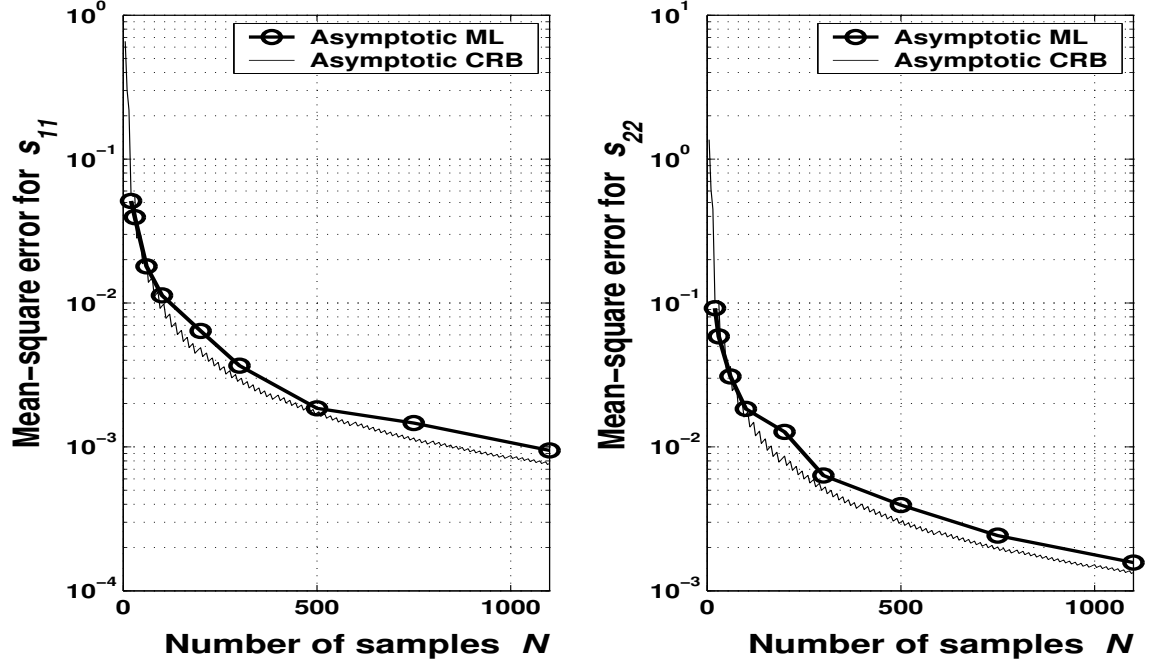
Fig. 11. Mean-square errors and corresponding asymptotic Cramér-Rao bounds for the asymptotic ML estimates of s_1 and s_2 in a SIMO system with independent fading, as functions of N.

V. Concluding Remarks

We derived asymptotic ML methods for estimating the Doppler-spread, noise variance, and channel covariance parameters from fading-channel estimates containing both the in-phase and quadrature-phase components under SISO and smart-antenna scenarios. Asymptotic Cramér-Rao bounds were derived for the unknown parameters. We also generalized the sample-covariance-based and approximate ML methods for estimating the Doppler spread in [6], [10], and [20] to the smart-antenna SIMO scenario. The performance of the proposed methods was evaluated for all parameters of interest under various simulation scenarios. We compared several Doppler-spread estimators and discussed their relative merits (For example, we observed in Section IV that the approximate ML and sample-covariance-based methods for Doppler-spread estimation may outperform the asymptotic ML estimator in some scenarios, e.g. when the Doppler spread or the number of samples are small.) In general, the asymptotic ML method shows excellent performance for large data records. We also discussed identifiability of the signal-to-noise ratio parameter and showed how its estimation may be improved (and the identifiability problem resolved) when the noise level is known.

Further research will include:

- examining the performance of the proposed methods in realistic *non-uniform* angle-of-arrival (AoA) and *impulsive-noise* environments,

- developing methods that account for non-uniform AoA distributions (along the lines of [2], [10], and [34]) and *man-made* and *atmospheric* impulsive noise (along the lines of [35], [36]),

- efficiently estimating the AoA and noise-distribution parameters,

- extending the SIMO asymptotic ML estimators in Section III to the Ricean fading scenario, and

- computing *exact* Cramér-Rao bounds for the Rayleigh and Ricean fading models in Sections II and III.

APPENDIX A.. APPROXIMATE ML ESTIMATION OF THE DOPPLER SPREAD

We derive the approximate ML estimator (2.15) of the Doppler-spread parameter f_D in the SISO scenario, discuss its relationship with the approximate average ML method in [20], and extend it to the smart-antenna SIMO scenario. **SISO Scenario:** We start with the following basis-function representation of the measurements $y(t)$ (see e.g. [37, eq. (2.6) and Fig. 1] and references therein):

$$y(t) = \sum_{l=1}^{L} x_l \cdot \exp[j2\pi f_\mathrm{D} \cos(\pi l/L)t] + e(t), \quad t = 1, 2, \ldots, N, \tag{A.1}$$

which is a linear combination of $L \leq N$ complex exponentials at frequencies $f_\mathrm{D} \cos(\pi l/L)$, $l = 1, 2, \ldots, L$, weighted by the (unknown) complex amplitudes x_l, $l = 1, 2, \ldots, L$, and corrupted by additive white circularly symmetric complex Gaussian noise $e(t)$ with unknown variance σ^2. Equation (A.1) approximates a multipath fading channel with uniformly distributed scatterers around the mobile (see also a similar model in [20, eq. (2)]). For a fixed f_D and under the model (A.1), the ML estimates of x_l, $l = 1, 2, \ldots, L$ and σ^2 easily follow (see e.g. [37]):

$$[\widehat{x}_1(f_\mathrm{D}), \widehat{x}_2(f_\mathrm{D}), \ldots, \widehat{x}_L(f_\mathrm{D})] = [y(1), y(2), \ldots, y(N)] \, \mathbf{\Phi}(f_\mathrm{D})^H \cdot [\mathbf{\Phi}(f_\mathrm{D})\mathbf{\Phi}(f_\mathrm{D})^H]^{-1}, \tag{A.2a}$$

$$\widehat{\sigma}^2(f_\mathrm{D}) = (1/N) \cdot [y(1), y(2), \ldots, y(N)] \cdot \left\{ \boldsymbol{I}_N - \mathbf{\Phi}(f_\mathrm{D})^H \, [\mathbf{\Phi}(f_\mathrm{D})\mathbf{\Phi}(f_\mathrm{D})^H]^{-1}\mathbf{\Phi}(f_\mathrm{D}) \right\} \cdot [y(1), y(2), \ldots, y(N)]^H, \tag{A.2b}$$

where

$$\mathbf{\Phi}(f_\mathrm{D}) = \begin{bmatrix} \exp[j2\pi f_\mathrm{D} \cos(\pi/L)] & \exp[j2\pi f_\mathrm{D} \cos(\pi/L) \cdot 2] & \cdots & \exp[j2\pi f_\mathrm{D} \cos(\pi/L) \cdot N] \\ \exp[j2\pi f_\mathrm{D} \cos(\pi 2/L)] & \exp[j2\pi f_\mathrm{D} \cos(\pi 2/L) \cdot 2] & \cdots & \exp[j2\pi f_\mathrm{D} \cos(\pi 2/L) \cdot N] \\ \vdots & \vdots & \vdots & \vdots \\ \exp[j2\pi f_\mathrm{D} \cos(\pi(L-1)/L)] & \exp[j2\pi f_\mathrm{D} \cos(\pi(L-1)/L) \cdot 2] & \cdots & \exp[j2\pi f_\mathrm{D} \cos(\pi(L-1)/L) \cdot N] \\ \exp(-j2\pi f_\mathrm{D}) & \exp(-j2\pi f_\mathrm{D} \cdot 2) & \cdots & \exp(-j2\pi f_\mathrm{D} \cdot N) \end{bmatrix}.$$

Substituting (A.2) into the likelihood function for the above measurement model yields the concentrated likelihood function (see e.g. [37, eq. (5.5)]):

$$l_\mathrm{c}^{\mathrm{app}}(f_\mathrm{D}) = \frac{(1/N) \cdot \sum_{t=1}^{N} |y(t)|^2}{\widehat{\sigma}^2(f_\mathrm{D})} \tag{A.3}$$

to be maximized with respect to f_D. For large N, we can use the following approximation (see [31, p. 157]):

$$\mathbf{\Phi}(f_\mathrm{D})\mathbf{\Phi}(f_\mathrm{D})^H \approx N\boldsymbol{I}_L \tag{A.4}$$

and a monotonic transformation $(l_\mathrm{c})^{\mathrm{app})\prime}(f_\mathrm{D}) = [1 - 1/l_\mathrm{c}^{\mathrm{app}}(f_\mathrm{D})] \cdot (1/N) \sum_{t=1}^{N} |y(t)|^2$ to obtain a simpler form of the concentrated likelihood

$$(l_\mathrm{c}^{\mathrm{app}})\prime(f_\mathrm{D}) = \sum_{l=1}^{L} C_{y,N}(f_\mathrm{D} \cos(\pi l/L)). \tag{A.5}$$

The above expression is equivalent to the approximate average log-likelihood function for f_{D} in [20, eq. (16)], where it was derived using a different measurement model. Note that (A.5) cannot be computed efficiently because of non-uniform sampling of the periodogram at frequencies that depend on the unknown parameter f_{D}. For large L, (A.5) is proportional to

$$l_{\mathrm{c}}^{\mathrm{app}}(f_{\mathrm{D}}) = \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot C_{y,N}(f_k) \tag{A.6}$$

and (2.15) follows. Note that (A.6) is different from Equation (17) in [20]. In [20], the authors mistakenly suggest maximizing $\sum_{k=0}^{N-1} f_{\mathrm{D}} \cdot z_k^2(f_{\mathrm{D}}) \cdot C_{y,N}(f_k)$ with respect to f_{D}, see [20, eqs. (3) and (17)].

SIMO Scenario: The above derivation is easily extended to the SIMO scenario in Section III, yielding (4.2).

APPENDIX B.. PX-EM ALGORITHM DERIVATION

We derive PX-EM algorithms for estimating S and σ^2 assuming that f_{D} is known.

First, note that the Whittle log-likelihood in (3.4) can be rewritten as

$$l(\boldsymbol{y}; \boldsymbol{\theta}) = - \sum_{k=0}^{N-1} \left\{ \ln |\pi \boldsymbol{P_{yy}}(f_k; \boldsymbol{\theta})| + \boldsymbol{y}_{\mathrm{DTFT}}(f_k)^H \boldsymbol{P_{yy}}(f_k; \boldsymbol{\theta})^{-1} \boldsymbol{y}_{\mathrm{DTFT}}(f_k) \right\}, \tag{B.1}$$

which would be the exact log-likelihood in the scenario where $\boldsymbol{y}_{\mathrm{DTFT}}(f_k)$, $k = 0, 1, \ldots, N-1$ are independent zero-mean circularly symmetric complex Gaussian random vectors with covariances $\boldsymbol{P_{yy}}(f_k; \boldsymbol{\theta})$. In the following, PX-EM algorithms are derived for maximizing (B.1) with respect to S and σ^2, assuming that f_{D} is known. We consider both the unstructured and independent fading scenarios, see the following discussion.

A. Unstructured Fading

We now derive the PX-EM algorithm for estimating S and σ^2 for known f_{D} and the unstructured fading scenario. Consider the following expanded measurement model:

$$\boldsymbol{y}_{\mathrm{DTFT}}(f_k) = z_k(f_{\mathrm{D}}) \cdot \boldsymbol{A} \boldsymbol{h}_{\mathrm{a},k} + \boldsymbol{e}_k \tag{B.2a}$$

$$= z_k(f_{\mathrm{D}}) \cdot (\boldsymbol{h}_{\mathrm{a},k}^T \otimes \boldsymbol{I}_{n_{\mathrm{R}}}) \cdot \mathrm{vec}(\boldsymbol{A}) + \boldsymbol{e}_k \tag{B.2b}$$

for $k = 0, 1, \ldots, N-1$, where $\boldsymbol{A}$ is the matrix of the auxiliary parameters, $\boldsymbol{h}_{\mathrm{a},k}$ are independent, identically distributed (i.i.d) zero-mean complex Gaussian random vectors with covariance $\boldsymbol{\Psi}_{\mathrm{a}} = \mathrm{E}\,[\boldsymbol{h}_{\mathrm{a},k}\boldsymbol{h}_{\mathrm{a},k}^H]$, and $\boldsymbol{e}_k$ is additive zero-mean white complex Gaussian noise with covariance $\mathrm{E}\,[\boldsymbol{e}_k\boldsymbol{e}_k^H] = \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}$. Here, the vec operator stacks the columns of a matrix one below another into a single column vector and (B.2b) follows from (B.2a) by applying the following identity:

$$\mathrm{vec}(\boldsymbol{PQ}) = (\boldsymbol{Q}^T \otimes \boldsymbol{I}) \, \mathrm{vec}(\boldsymbol{P}) \tag{B.3}$$

which holds for arbitrary conforming matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ and an identity matrix $\boldsymbol{I}$ of appropriate dimensions, see [28, eq. (2.11) at p. 342]. We assume that $\boldsymbol{h}_{\mathrm{a},k}$ and $\boldsymbol{e}_k$ are independent, implying that $\mathrm{E}\,[\boldsymbol{h}_{\mathrm{a},k_1}\boldsymbol{e}_{k_2}^H] = \boldsymbol{0}$ for $k_1, k_2 \in$

$\{0, 1, \ldots, N-1\}$. Clearly, the covariances of $\boldsymbol{y}_{\mathrm{DTFT}}(f_k)$ can be written as

$$\boldsymbol{P_{yy}}(f_k; \boldsymbol{\theta}) = z_k^2(f_{\mathrm{D}}) \cdot \boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}, \tag{B.4}$$

implying that

$$\boldsymbol{S} = \boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H / \sigma^2. \tag{B.5}$$

We wish to find the ML estimates of the following parameters of the expanded model: σ^2, $\boldsymbol{A}$, and $\boldsymbol{\Psi}_{\mathrm{a}}$; then the ML estimate of $\boldsymbol{S}$ easily follows by using (B.5). To compute these ML estimates, we derive the EM algorithm (see e.g. [38] and [39]) for the model (B.2) by treating $\boldsymbol{h}_{\mathrm{a},k}$, $k = 0, 1, \ldots, N-1$ as the *unobserved* (or missing) data. Then, the complete-data log-likelihood function is

$$L_{\mathrm{c}}(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}}; \sigma^2, \boldsymbol{A}, \boldsymbol{\Psi}_{\mathrm{a}}) = -2Nn_{\mathrm{R}}\ln\pi - Nn_{\mathrm{R}}\ln\sigma^2 - N\ln|\boldsymbol{\Psi}_{\mathrm{a}}|$$
$$- \sum_{k=0}^{N-1} \{[\boldsymbol{y}_{\mathrm{DTFT}}(f_k) - z_k(f_{\mathrm{D}}) \cdot \boldsymbol{A}\boldsymbol{h}_{\mathrm{a},k}]^H [\boldsymbol{y}_{\mathrm{DTFT}}(f_k) - z_k(f_{\mathrm{D}}) \cdot \boldsymbol{A}\boldsymbol{h}_{\mathrm{a},k}]/\sigma^2\} - \mathrm{tr}\left(\boldsymbol{\Psi}_{\mathrm{a}}^{-1} \sum_{k=0}^{N-1} \boldsymbol{h}_{\mathrm{a},k}\boldsymbol{h}_{\mathrm{a},k}^H\right) \tag{B.6}$$

and the complete-data sufficient statistics are

$$t_0(\boldsymbol{y}) = \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{y}_{\mathrm{DTFT}}(f_k)^H \boldsymbol{y}_{\mathrm{DTFT}}(f_k), \tag{B.7a}$$

$$\boldsymbol{T}_1(\boldsymbol{h}_{\mathrm{a}}) = \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{h}_{\mathrm{a},k}\boldsymbol{h}_{\mathrm{a},k}^H, \tag{B.7b}$$

$$\boldsymbol{T}_2(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}}) = \frac{1}{N} \sum_{k=0}^{N-1} z_k(f_{\mathrm{D}}) \cdot \boldsymbol{y}_{\mathrm{DTFT}}(f_k) \boldsymbol{h}_{\mathrm{a},k}^H, \tag{B.7c}$$

$$\boldsymbol{T}_3(\boldsymbol{h}_{\mathrm{a}}) = \frac{1}{N} \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \boldsymbol{h}_{\mathrm{a},k} \boldsymbol{h}_{\mathrm{a},k}^H, \tag{B.7d}$$

where $\boldsymbol{h}_{\mathrm{a}} = [\boldsymbol{h}_{\mathrm{a},0}^T, \boldsymbol{h}_{\mathrm{a},1}^T, \ldots, \boldsymbol{h}_{\mathrm{a},N-1}^T]^T$. The complete-data log-likelihood (B.6) is easily maximized with respect to $\boldsymbol{A}, \boldsymbol{\Psi}_{\mathrm{a}}$, and σ^2, yielding the following estimates:

$$\widehat{\boldsymbol{A}} = \left[\sum_{k=0}^{N-1} z_k(f_{\mathrm{D}}) \cdot \boldsymbol{y}_{\mathrm{DTFT}}(f_k) \cdot \boldsymbol{h}_{\mathrm{a},k}^H\right] \cdot \left[\sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \boldsymbol{h}_{\mathrm{a},k}\boldsymbol{h}_{\mathrm{a},k}^H\right]^{-1} = \boldsymbol{T}_2(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}}) \cdot \boldsymbol{T}_3(\boldsymbol{h}_{\mathrm{a}})^{-1}, \tag{B.8a}$$

$$\widehat{\boldsymbol{\Psi}}_{\mathrm{a}} = \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{h}_{\mathrm{a},k}\boldsymbol{h}_{\mathrm{a},k}^H = \boldsymbol{T}_1(\boldsymbol{h}_{\mathrm{a}}), \tag{B.8b}$$

and, for a given $\boldsymbol{A}$,

$$\widehat{\sigma}^2 = \frac{1}{n_{\mathrm{R}}N} \cdot \sum_{k=0}^{N-1} [\boldsymbol{y}_{\mathrm{DTFT}}(f_k) - z_k(f_{\mathrm{D}}) \cdot \boldsymbol{A} \cdot \boldsymbol{h}_{\mathrm{a},k}^{(i)}]^H \cdot [\boldsymbol{y}_{\mathrm{DTFT}}(f_k) - z_k(f_{\mathrm{D}}) \cdot \boldsymbol{A} \cdot \boldsymbol{h}_{\mathrm{a},k}^{(i)}]$$
$$= \frac{1}{n_{\mathrm{R}}} \cdot \{t_0(\boldsymbol{y}) - \mathrm{tr}[\boldsymbol{A}^H \boldsymbol{T}_2(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}})] - \mathrm{tr}[\boldsymbol{T}_2(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}})^H \boldsymbol{A}] + \mathrm{tr}[\boldsymbol{A}^H \boldsymbol{A}\boldsymbol{T}_3(\boldsymbol{h}_{\mathrm{a}})]\}. \tag{B.8c}$$

The equation (B.8a) follows as a least-squares solution to the linear regression problem in (B.2b):

$$\mathrm{vec}(\widehat{\boldsymbol{A}}) = \left\{\left[\sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \boldsymbol{h}_{\mathrm{a},k}^* \boldsymbol{h}_{\mathrm{a},k}^T\right] \otimes \boldsymbol{I}_{n_{\mathrm{R}}}\right\}^{-1} \cdot \sum_{k=0}^{N-1} z_k(f_{\mathrm{D}}) \cdot (\boldsymbol{h}_{\mathrm{a},k}^* \otimes \boldsymbol{I}_{n_{\mathrm{R}}}) \cdot \boldsymbol{y}_{\mathrm{DTFT}}(f_k), \tag{B.9}$$

where we used the Kronecker product theorem [28, Lemma 16.1.2] stating that for arbitrary conforming matrices $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{R}$, and $\boldsymbol{S}$, $(\boldsymbol{P} \otimes \boldsymbol{Q}) \cdot (\boldsymbol{R} \otimes \boldsymbol{S}) = (\boldsymbol{P}\boldsymbol{R}) \otimes (\boldsymbol{Q}\boldsymbol{S})$. Then, (B.8a) follows by applying (B.3) to (B.9).

The complete-data likelihood belongs to an exponential family of distributions, i.e. the log-likelihood (B.6) is linear in the *natural sufficient statistics* (B.7); see e.g. [29, ch. 1.6.2] for the definition of the multiparameter exponential family and natural sufficient statistics. Then, it easily follows from (B.6) that the expectation (E) step reduces to computing the conditional expectations of the complete-data natural sufficient statistics [in (B.7)] given the observed data $\boldsymbol{y}$. Observe that, for the measurement model in (B.2), $\boldsymbol{y}_{\text{DTFT}}(f_k)$ and $\boldsymbol{h}_{\text{a},k}$ are jointly complex Gaussian with the following mean and covariance:

$$\text{E} \left\{ \begin{bmatrix} \boldsymbol{y}_{\text{DTFT}}(f_k) \\ \boldsymbol{h}_{\text{a},k} \end{bmatrix} \right\} = \boldsymbol{0}, \tag{B.10a}$$

$$\text{cov} \left\{ \begin{bmatrix} \boldsymbol{y}_{\text{DTFT}}(f_k) \\ \boldsymbol{h}_{\text{a},k} \end{bmatrix} \right\} = \begin{bmatrix} z_k^2(f_{\text{D}}) \cdot \boldsymbol{A}\boldsymbol{\Psi}_{\text{a}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_{n_{\text{R}}} & z_k(f_{\text{D}}) \cdot \boldsymbol{A}\boldsymbol{\Psi}_{\text{a}} \\ z_k(f_{\text{D}}) \cdot \boldsymbol{\Psi}_{\text{a}}\boldsymbol{A}^H & \boldsymbol{\Psi}_{\text{a}} \end{bmatrix}, \tag{B.10b}$$

and then it easily follows from [24, result 7. at pp. 508–509] that $\boldsymbol{h}_{\text{a},k}$ conditioned on $\boldsymbol{y}_{\text{DTFT}}(f_k)$ is a complex Gaussian vector with the mean and covariance equal to

$$\text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right] = z_k(f_{\text{D}}) \cdot \boldsymbol{\Psi}_{\text{a}}\boldsymbol{A}^H \boldsymbol{W}_k \, \boldsymbol{y}_{\text{DTFT}}(f_k), \tag{B.11a}$$

$$\text{cov}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right] = \boldsymbol{\Psi}_{\text{a}} - z_k^2(f_{\text{D}}) \cdot \boldsymbol{\Psi}_{\text{a}}\boldsymbol{A}^H \boldsymbol{W}_k \boldsymbol{A}\boldsymbol{\Psi}_{\text{a}}, \tag{B.11b}$$

where $\boldsymbol{W}_k = [z_k^2(f_{\text{D}}) \cdot \boldsymbol{A}\boldsymbol{\Psi}_{\text{a}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_{n_{\text{R}}}]^{-1}$. Define also

$$\boldsymbol{Q}_k = \text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right] \cdot \text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right]^H - z_k^2(f_{\text{D}}) \cdot \boldsymbol{\Psi}_{\text{a}}\boldsymbol{A}^H \boldsymbol{W}_k \boldsymbol{A}\boldsymbol{\Psi}_{\text{a}}. \tag{B.12}$$

We now use the above expressions to find the conditional expectations of the complete-data natural sufficient statistics (B.7):

$$\text{E}\left[t_0(\boldsymbol{y})\,|\,\boldsymbol{y}; \sigma^2, \boldsymbol{A}, \boldsymbol{\Psi}_{\text{a}}\right] = t_0(\boldsymbol{y}), \tag{B.13a}$$

$$\text{E}\left[\boldsymbol{T}_1(\boldsymbol{h}_{\text{a}})\,|\,\boldsymbol{y}; \sigma^2, \boldsymbol{A}, \boldsymbol{\Psi}_{\text{a}}\right] = \frac{1}{N}\sum_{k=0}^{N-1}\left\{\text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right]\cdot\text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right]^H + \text{cov}(\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k))\right\}$$

$$= \boldsymbol{\Psi}_{\text{a}} + \frac{1}{N}\sum_{k=0}^{N-1}\boldsymbol{Q}_k, \tag{B.13b}$$

$$\text{E}\left[\boldsymbol{T}_2(\boldsymbol{y},\boldsymbol{h}_{\text{a}})\,|\,\boldsymbol{y}; \sigma^2, \boldsymbol{A}, \boldsymbol{\Psi}_{\text{a}}\right] = \frac{1}{N}\sum_{k=0}^{N-1} z_k(f_{\text{D}}) \cdot \boldsymbol{y}_{\text{DTFT}}(f_k) \cdot \text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right]^H, \tag{B.13c}$$

$$\text{E}\left[\boldsymbol{T}_3(\boldsymbol{h}_{\text{a}})\,|\,\boldsymbol{y}; \sigma^2, \boldsymbol{A}, \boldsymbol{\Psi}_{\text{a}}\right] = \frac{1}{N}\sum_{k=0}^{N-1} z_k^2(f_{\text{D}})\cdot\left\{\text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right]\cdot\text{E}\left[\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k)\right]^H + \text{cov}(\boldsymbol{h}_{\text{a},k}|\boldsymbol{y}_{\text{DTFT}}(f_k))\right\}$$

$$= \frac{1}{N}\Big[\sum_{k=0}^{N-1} z_k^2(f_{\text{D}})\Big]\cdot\boldsymbol{\Psi}_{\text{a}} + \frac{1}{N}\sum_{k=0}^{N-1} z_k^2(f_{\text{D}})\cdot\boldsymbol{Q}_k, \tag{B.13d}$$

where we emphasize the dependence of the above conditional expectations on the parameters σ^2, $\boldsymbol{A}$, and $\boldsymbol{\Psi}_{\text{a}}$. Now, the maximization (M) step follows by replacing the complete-data sufficient statistics (B.7) that occur in the complete-data

ML estimates (B.8) with their conditional expectations computed in (B.13):

$$\boldsymbol{A}^{(i+1)} = \mathrm{E}\left[\boldsymbol{T}_2(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}}) \,|\, \boldsymbol{y};\, (\sigma^2)^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}\right] \cdot \mathrm{E}\left[\boldsymbol{T}_3(\boldsymbol{h}_{\mathrm{a}}) \,|\, \boldsymbol{y};\, (\sigma^2)^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}\right]^{-1}, \tag{B.14a}$$

$$\boldsymbol{\Psi}_{\mathrm{a}}^{(i+1)} = \mathrm{E}\left[\boldsymbol{T}_1(\boldsymbol{h}_{\mathrm{a}}) \,|\, \boldsymbol{y};\, (\sigma^2)^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}\right], \tag{B.14b}$$

$$\begin{aligned}
(\sigma^2)^{(i+1)} = {} & \frac{1}{n_{\mathrm{R}}} \cdot \Big(t_0(\boldsymbol{y}) - \mathrm{tr}\{(\boldsymbol{A}^{(i)})^H \cdot \mathrm{E}\left[\boldsymbol{T}_2(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}}) \,|\, \boldsymbol{y};\, (\sigma^2)^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}\right]\} \\
& - \mathrm{tr}\{\mathrm{E}\left[\boldsymbol{T}_2(\boldsymbol{y}, \boldsymbol{h}_{\mathrm{a}}) \,|\, \boldsymbol{y};\, (\sigma^2)^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}\right]^H \cdot \boldsymbol{A}^{(i)}\} \\
& + \mathrm{tr}\{(\boldsymbol{A}^{(i)})^H \boldsymbol{A}^{(i)} \cdot \mathrm{E}\left[\boldsymbol{T}_3(\boldsymbol{h}_{\mathrm{a}}) \,|\, \boldsymbol{y};\, (\sigma^2)^{(i)}, \boldsymbol{A}^{(i)}, \boldsymbol{\Psi}_{\mathrm{a}}^{(i)}\right]\} \Big),
\end{aligned} \tag{B.14c}$$

and the equations (3.11) easily follow. To derive (3.11d), we used the following identity [see also (B.11)]:

$$\begin{aligned}
z_k^2(f_{\mathrm{D}}) \,\mathrm{tr}\{\boldsymbol{A}^H \boldsymbol{A} \cdot \mathrm{cov}[\boldsymbol{h}_{\mathrm{a},k}|\boldsymbol{y}_{\mathrm{DTFT}}]\} = {} & \mathrm{tr}\Big\{ z_k^2(f_{\mathrm{D}})\boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H - z_k^2(f_{\mathrm{D}})\boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H \\
& \cdot [z_k^2(f_{\mathrm{D}})\boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}]^{-1} \cdot [z_k^2(f_{\mathrm{D}})\boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}} - \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}]\Big\} \\
= {} & \mathrm{tr}\Big(z_k^2(f_{\mathrm{D}})\boldsymbol{A}^H \boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A} \cdot \{\boldsymbol{I}_{n_{\mathrm{R}}} - \boldsymbol{I}_{n_{\mathrm{R}}} + \sigma^2 \cdot [z_k^2(f_{\mathrm{D}})\boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}]^{-1}\}\Big) \\
= {} & \sigma^2 \cdot \mathrm{tr}\Big\{ [z_k^2(f_{\mathrm{D}})\boldsymbol{A}^H \boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A} + \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}} - \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}] \cdot [z_k^2(f_{\mathrm{D}})\boldsymbol{A}\boldsymbol{\Psi}_{\mathrm{a}}\boldsymbol{A}^H + \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}]^{-1}\Big\} = \sigma^2 \cdot \mathrm{tr}\{\boldsymbol{I}_{n_{\mathrm{R}}} - \sigma^2 \cdot \boldsymbol{W}_k\}. \tag{B.15}
\end{aligned}$$

B. Independent Fading

The PX-EM algorithm for estimating $\boldsymbol{s}$ and σ^2 for known f_{D} and the independent fading scenario is easily derived from the following expanded measurement model:

$$[\boldsymbol{y}_{\mathrm{DTFT}}(f_k)]_p = z_k(f_{\mathrm{D}}) \cdot a_p \cdot h_{\mathrm{a},k,p} + [\boldsymbol{e}_k]_p, \tag{B.16}$$

for $p = 1, 2, \ldots, n_{\mathrm{R}}$, $k = 0, 1, \ldots, N - 1$, where a_p are the auxiliary parameters, $h_{\mathrm{a},k,p}$ are independent zero-mean complex Gaussian random variables with variances ψ_p, and $\boldsymbol{e}_k$ is additive zero-mean white complex Gaussian noise with covariance $\mathrm{E}\left[\boldsymbol{e}_k \boldsymbol{e}_k^H\right] = \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}$, independent from $h_{\mathrm{a},k,p}$. Under the above assumptions, the covariance matrices of $\boldsymbol{y}_{\mathrm{DTFT}}(f_k)$ can be written as

$$\boldsymbol{P}_{\boldsymbol{yy}}(f_k; \boldsymbol{\theta}) = z_k^2(f_{\mathrm{D}}) \cdot \mathrm{diag}\left\{|a_1|^2 \psi_{\mathrm{a},1}, |a_2|^2 \psi_{\mathrm{a},2}, \ldots, |a_{n_{\mathrm{R}}}|^2 \psi_{\mathrm{a},n_{\mathrm{R}}}\right\} + \sigma^2 \boldsymbol{I}_{n_{\mathrm{R}}}, \tag{B.17}$$

implying that

$$\boldsymbol{s} = \mathrm{diag}\{|a_1|^2 \psi_{\mathrm{a},1}/\sigma^2, |a_2|^2 \psi_{\mathrm{a},2}/\sigma^2, \ldots, |a_{n_{\mathrm{R}}}|^2 \psi_{\mathrm{a},n_{\mathrm{R}}}/\sigma^2\}. \tag{B.18}$$

The PX-EM algorithm for independent fading in Section III-B follows from the above measurement model by using arguments similar to those in Appendix B-A where the PX-EM algorithm was derived for the unstructured fading scenario.

APPENDIX C.. ASYMPTOTIC FISHER INFORMATION MATRIX

We utilize (3.20)–(3.22) to compute the elements of $\mathcal{I}(\boldsymbol{\theta})$ that correspond to the normalized fading covariance parameters $\boldsymbol{s}$ in unstructured and independent fading scenarios (see Appendices C-A and C-B below). To simplify the notation, we omit the dependences of $\mathcal{I}(\boldsymbol{\theta})$ and $\boldsymbol{G}(f; \boldsymbol{\rho})$ on $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, respectively.

A. FIM for Unstructured S

Denote by $S_{p,q}$ the (p,q) element of S, where $p, q = 1, 2, \ldots, n_{\mathrm{R}}$. First, we use (3.20) to compute

$$[\mathcal{I}]_{\sigma^2, S_{p,p}} = \frac{1}{\sigma^2} \cdot \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \mathrm{tr}\left[G(f_k)^{-1} \cdot \frac{\partial S}{\partial S_{p,p}}\right] = \frac{1}{\sigma^2} \cdot \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \left[G(f_k)^{-1}\right]_{p,p} \tag{C.1}$$

and, for $p > q$,

$$[\mathcal{I}]_{\sigma^2, \mathrm{Re}\{S_{p,q}\}} = \frac{1}{\sigma^2} \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \, \mathrm{tr}\left[G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Re}\{S_{p,q}\}}\right] = \frac{2}{\sigma^2} \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \mathrm{Re}\{\left[G(f_k)^{-1}\right]_{p,q}\}, \tag{C.2a}$$

$$[\mathcal{I}]_{\sigma^2, \mathrm{Im}\{S_{p,q}\}} = \frac{1}{\sigma^2} \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \, \mathrm{tr}\left[G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Im}\{S_{p,q}\}}\right] = \frac{2}{\sigma^2} \sum_{k=0}^{N-1} z_k^2(f_{\mathrm{D}}) \cdot \mathrm{Im}\{\left[G(f_k)^{-1}\right]_{p,q}\}. \tag{C.2b}$$

Similarly, (3.21) implies

$$[\mathcal{I}]_{f_{\mathrm{D}}, S_{p,p}} = -f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} z_k^8(f_{\mathrm{D}}) \cdot \mathrm{tr}\left[G(f_k)^{-1} S G(f_k)^{-1} \cdot \frac{\partial S}{\partial S_{p,p}}\right] = -f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} z_k^8(f_{\mathrm{D}}) \cdot \left[G(f_k)^{-1} S G(f_k)^{-1}\right]_{p,p} \tag{C.3}$$

and, for $p > q$,

$$[\mathcal{I}]_{f_{\mathrm{D}}, \mathrm{Re}\{S_{p,q}\}} = -f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} z_k^8(f_{\mathrm{D}}) \cdot \mathrm{tr}\left[G(f_k)^{-1} S G(f_k)^{-1} \cdot \frac{\partial S}{\partial \mathrm{Re}\{S_{p,q}\}}\right]$$

$$= -2f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} z_k^8(f_{\mathrm{D}}) \cdot \mathrm{Re}\left\{[G(f_k)^{-1} S G(f_k)^{-1}]_{p,q}\right\} \tag{C.4a}$$

$$[\mathcal{I}]_{f_{\mathrm{D}}, \mathrm{Im}\{S_{p,q}\}} = -f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} z_k^8(f_{\mathrm{D}}) \cdot \mathrm{tr}\left[G(f_k)^{-1} S G(f_k)^{-1} \cdot \frac{\partial S}{\partial \mathrm{Im}\{S_{p,q}\}}\right]$$

$$= -2f_{\mathrm{D}} \cdot \sum_{k=0}^{N-1} z_k^8(f_{\mathrm{D}}) \cdot \mathrm{Im}\left\{[G(f_k)^{-1} S G(f_k)^{-1}]_{p,q}\right\}. \tag{C.4b}$$

Finally, for $p_1 > q_1$ and $p_2 > q_2$, (3.22) simplifies to

$$\mathcal{I}_{\mathrm{Re}\{S_{p_1,q_1}\}, \mathrm{Re}\{S_{p_2,q_2}\}} = \mathcal{I}_{\mathrm{Re}\{S_{p_2,q_2}\}, \mathrm{Re}\{S_{p_1,q_1}\}} = \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \, \mathrm{tr}\left\{G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Re}\{S_{p_1,q_1}\}} \cdot G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Re}\{S_{p_2,q_2}\}}\right\}$$

$$= 2 \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \, \mathrm{Re}\left\{[G(f_k)^{-1}]_{q_2,p_1} \cdot [G(f_k)^{-1}]_{q_1,p_2} + [G(f_k)^{-1}]_{q_2,q_1} \cdot [G(f_k)^{-1}]_{p_1,p_2}\right\}, \tag{C.5a}$$

$$\mathcal{I}_{\mathrm{Re}\{S_{p_1,q_1}\}, \mathrm{Im}\{S_{p_2,q_2}\}} = \mathcal{I}_{\mathrm{Im}\{p_2,q_2\}, \mathrm{Re}\{S_{p_1,q_1}\}} = \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \, \mathrm{tr}\left\{G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Re}\{S_{p_1,q_1}\}} \cdot G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Im}\{p_2, q_2\}}\right\}$$

$$= -2 \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \, \mathrm{Im}\left\{[G(f_k)^{-1}]_{q_2,p_1} \cdot [G(f_k)^{-1}]_{q_1,p_2} + [G(f_k)^{-1}]_{q_2,q_1} \cdot [G(f_k)^{-1}]_{p_1,p_2}\right\}, \tag{C.5b}$$

$$\mathcal{I}_{\mathrm{Im}\{S_{p_1,q_1}\}, \mathrm{Im}\{S_{p_2,q_2}\}} = \mathcal{I}_{\mathrm{Im}\{S_{p_2,q_2}\}, \mathrm{Im}\{S_{p_1,q_1}\}} = \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \, \mathrm{tr}\left\{G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Im}\{S_{p_1,q_1}\}} \cdot G(f_k)^{-1} \frac{\partial S}{\partial \mathrm{Im}\{S_{p_2,q_2}\}}\right\}$$

$$= 2 \sum_{k=0}^{N-1} z_k^4(f_{\mathrm{D}}) \, \mathrm{Re}\left\{-[G(f_k)^{-1}]_{q_2,p_1} \cdot [G(f_k)^{-1}]_{q_1,p_2} + [G(f_k)^{-1}]_{q_2,q_1} \cdot [G(f_k)^{-1}]_{p_1,p_2}\right\}, \tag{C.5c}$$

and for $p_1 = q_1$ and $p_2 > q_2$,

$$
\begin{aligned}
\mathcal{I}_{S_{p_1,p_1},\mathrm{Re}\{S_{p_2,q_2}\}} = \mathcal{I}_{\mathrm{Re}\{S_{p_2,q_2}\},S_{p_1,p_1}} &= \sum_{k=0}^{N-1} z_k^4(f_\mathrm{D}) \,\mathrm{tr}\left\{ \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial S_{p_1,p_1}} \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial \mathrm{Re}\{S_{p_2,q_2}\}} \right\} \\
&= 2\sum_{k=0}^{N-1} z_k^4(f_\mathrm{D}) \,\mathrm{Re}\left\{ [\boldsymbol{G}(f_k)^{-1}]_{q_2,p_1} \cdot [\boldsymbol{G}(f_k)^{-1}]_{p_1,p_2} \right\},
\end{aligned}
\tag{C.6a}
$$

$$
\begin{aligned}
\mathcal{I}_{S_{p_1,p_1},\mathrm{Im}\{S_{p_2,q_2}\}} = \mathcal{I}_{\mathrm{Im}\{S_{p_2,q_2}\},S_{p_1,p_1}} &= \sum_{k=0}^{N-1} z_k^4(f_\mathrm{D}) \,\mathrm{tr}\left\{ \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial S_{p_1,p_1}} \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial \mathrm{Im}\{S_{p_2,q_2}\}} \right\} \\
&= -2\sum_{k=0}^{N-1} z_k^4(f_\mathrm{D}) \cdot \mathrm{Im}\left\{ [\boldsymbol{G}(f_k)^{-1}]_{q_2,p_1} \cdot [\boldsymbol{G}(f_k)^{-1}]_{p_1,p_2} \right\},
\end{aligned}
\tag{C.6b}
$$

and for $p_1 = q_1$ and $p_2 = q_2$,

$$
\mathcal{I}_{S_{p_1,p_1},S_{p_2,p_2}} = \mathcal{I}_{S_{p_2,p_2},S_{p_1,p_1}} = \sum_{k=0}^{N-1} z_k^4(f_\mathrm{D}) \,\mathrm{tr}\left\{ \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial S_{p_1,p_1}} \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial S_{p_2,p_2}} \right\} = \sum_{k=0}^{N-1} z_k^4(f_\mathrm{D}) \cdot \left| [\boldsymbol{G}(f_k)^{-1}]_{p_1,p_2} \right|^2.
\tag{C.7}
$$

B. FIM for Independent Fading

For independent fading (i.e. $\boldsymbol{S} = \mathrm{diag}\{s_1, s_2, \ldots, s_{n_\mathrm{R}}\}$ and $\boldsymbol{s} = [s_1, s_2, \ldots, s_{n_\mathrm{R}}]^T$), the elements of $\mathcal{I}(\boldsymbol{\theta})$ related to $\boldsymbol{s}$ can be computed as follows:

$$
\begin{aligned}
[\mathcal{I}]_{\sigma^2,s_p} &= \frac{1}{\sigma^2} \cdot \sum_{k=0}^{N-1} z_k^2(f_\mathrm{D}) \cdot \mathrm{tr}\left[\boldsymbol{G}(f_k)^{-1} \cdot \frac{\partial \boldsymbol{S}}{\partial s_p} \right] = \frac{1}{\sigma^2} \cdot \sum_{k=0}^{N-1} \frac{z_k^2(f_\mathrm{D})}{1 + z_k^2(f_\mathrm{D})s_p},
\end{aligned}
\tag{C.8a}
$$

$$
\begin{aligned}
[\mathcal{I}]_{f_\mathrm{D},s_p} &= -f_\mathrm{D} \cdot \sum_{k=0}^{N-1} z_k^8(f_\mathrm{D}) \cdot \mathrm{tr}\left[\boldsymbol{G}(f_k)^{-1}\boldsymbol{S}\boldsymbol{G}(f_k)^{-1} \cdot \frac{\partial \boldsymbol{S}}{\partial s_p} \right] \\
&= -f_\mathrm{D} \cdot \sum_{k=0}^{N-1} z_k^8(f_\mathrm{D}) \cdot \frac{s_p}{(1 + z_k^2(f_\mathrm{D})s_p)^2}
\end{aligned}
\tag{C.8b}
$$

for $p = 1, 2, \ldots, n_\mathrm{R}$, and

$$
\mathcal{I}_{s_{p_1},s_{p_2}} = \sum_{k=0}^{N-1} z_k^4(f_\mathrm{D}) \,\mathrm{tr}\left\{ \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial s_{p_1}} \boldsymbol{G}(f_k)^{-1}\frac{\partial \boldsymbol{S}}{\partial s_{p_2}} \right\} = \begin{cases} 0, & p_1 \neq p_2 \\ \sum_{k=0}^{N-1} z_k^4(f_\mathrm{D})/[1 + z_k^2(f_\mathrm{D})s_{p_1}]^2, & p_1 = p_2 \end{cases}
\tag{C.9}
$$

for $p_1, p_2 \in \{1, 2, \ldots, n_\mathrm{R}\}$.

REFERENCES

[1] G.L. Stüber, *Principles of Mobile Communication,* 2nd. ed., Norwell, MA: Kluwer, 2001.

[2] C. Tepedelenlioğlu, A. Abdi, G. B. Giannakis, and M. Kaveh, "Estimation of Doppler spread and signal strength in mobile communications with applications to handoff and adaptive transmission," *Wirel. Commun. Mobile Comput.*, vol. 1, pp. 221–242, Apr.-June 2001.

[3] A. Duel-Hallen, S.Q. Hu, and H. Hallen, "Long-range prediction of fading signals: Enabling adapting transmission for mobile radio channels," *IEEE Signal Processing Mag.,* vol. 17, pp. 62–75, May 2000.

[4] M.D. Austin and G.L. Stüber, "Velocity adaptive handoff algorithms for microcellular systems," *IEEE Trans. Veh. Technol.,* vol. 43, pp. 549–561, Aug. 1994.

[5] A.J. Goldsmith, L.J. Greenstein, and G.J. Foschini, "Error statistics of real-time power measurements in cellular channels with multipath and shadowing," *IEEE Trans. Veh. Technol.,* vol. 43, pp. 439–446, Aug. 1994.

[6] J.M. Holtzman and A. Sampath, "Adaptive averaging methodology for handoffs in cellular systems," *IEEE Trans. Veh. Technol.,* vol. 44, pp. 59–66, Feb. 1995.

[7] A.J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[8] M.J. Chu and W.E. Stark, "Effect of mobile velocity on communications in fading channels," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 202–210, Jan. 2000.

[9] L. Krasny, H. Arslan, D. Koilpillai, and S. Chennakeshu, "Doppler spread estimation in mobile radio systems," *IEEE Commun. Letters*, vol. 5, pp. 197–199, May 2001.

[10] C. Tepedelenlioğlu, "Performance analysis of velocity (Doppler) estimators in mobile communications," in *Proc. Int. Conf. Acoust., Speech, Signal Processing,* Orlando, FL, May 2002, vol. 3, pp. III-2201–III-2204.

[11] T.S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2002.

[12] D.-S. Shiu, G.J. Foschini, M.J. Gans, and J.M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.

[13] C.-N. Chuah, D.N.C. Tse, J.M. Kahn, and R.A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.

[14] A. Dogandžić, "Chernoff bounds on pairwise error probabilities of space-time codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1327–1336, May 2003.

[15] *IEEE Journal on Selected Areas in Communications*, Special Issue on MIMO Systems and Applications, Apr. and June 2003.

[16] *IEEE Transactions on Information Theory*, Special Issue on Space-time Transmission, Reception, Coding, and Signal Processing, Oct. 2003.

[17] A. Dogandžić and J. Jin, "Estimating statistical properties of MIMO fading channels," in revision for *IEEE Trans. Signal Processing*.

[18] D. Wong and D.C. Cox, "Estimating local mean signal power level in a Rayleigh fading environment," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 956–959, May 1999.

[19] R. Narasimhan and D.C. Cox, "Speed estimation in wireless systems using wavelets," *IEEE Trans. Commun.*, vol. 47, pp. 1357–1364, Sept. 1999.

[20] H. Hansen, S. Affes, and P. Mermelstein, "A Rayleigh Doppler frequency estimator derived from maximum likelihood theory," *Proc. 2nd IEEE Workshop Signal Process. Adv. Wireless Commun.,* Annapolis, MD, May 1999, pp. 382–386.

[21] A. Dogandžić and B. Zhang, "Maximum likelihood estimation of Jakes' Doppler power spectrum parameters for SIMO channels using the Whittle approximation," in *Proc. 2003 IEEE Workshop Stat. Signal Process.,* St. Louis, MO, Sept. 2003, pp. 85–88.

[22] W.C. Jakes, Ed., *Microwave Mobile Communications*, New York: Wiley, 1974.

[23] R.H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.,* vol. 47, pp. 957–1000, Jul.-Aug. 1968.

[24] S.M. Kay, *Fundamentals of Statistical Signal Processing — Estimation Theory*, Englewood Cliffs, NJ: Prentice Hall, 1993.

[25] P. Whittle, *Hypothesis Testing in Time Series Analysis*, Uppsala, Sweden: Almquist and Wiksel, 1951.

[26] P. Whittle, "Some recent contributions to the theory of stationary processes," Appendix 2 in *A Study in the Analysis of Stationary Time Series,* H. Wold, 2nd ed., Stockholm, Sweden: Almquist and Wiksel, 1954.

[27] M.J. Levin, "Power spectrum parameter estimation," *IEEE Trans. Inform. Theory*, vol. 11, pp. 100–107, Jan. 1965.

[28] D.A. Harville, *Matrix Algebra From a Statistician's Perspective,* New York:Springer-Verlag, 1997.

[29] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics,* 2nd ed., Upper Saddle River, NJ: Prentice Hall, 2000.

[30] P. Whittle, "The analysis of multiple stationary time series," *J. R. Stat. Soc., Ser. B,* vol. 15, pp. 125–139, 1953.

[31] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory,* Englewood Cliffs, NJ: Prentice Hall, 1998, pt. II.

[32] C.H. Liu, D.B. Rubin, and Y.N. Wu, "Parameter expansion to accelerate EM: The PX-EM algorithm," *Biometrika,* vol. 85, pp. 755-770, Dec. 1998.

[33] W.H. Press, S.A. Teukolsky, W.T. Vettering, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing,* 2nd ed., Cambridge, UK: Cambridge Univ. Press, 1992.

[34] A. Abdi, J.A. Barger, and M. Kaveh, "A parametric model for the distribution of the angle of arrival and the associated correlation function and power spectrum at the mobile station," *IEEE Trans. Veh. Technol.,* vol. 51, pp. 425-434, May 2002.

[35] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Trans. Electromagnetic Compatibility,* vol. 19, pp. 106–127, 1977.

[36] H.M. Hall, "A new model for impulsive phenomena: Application to atmospheric-noise communication channels," Stanford Univ., Stanford, CA, Tech. Rep. 3412-8, Aug. 1966.

[37] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," *IEEE Trans. Signal Processing,* vol. 50, pp. 457–474, March 2002.

[38] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc., Ser. B,* vol. 39, pp. 1–38, Jul. 1977.

[39] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, New York: Wiley, 1997.