Semi-blind SIMO Flat-Fading Channel Estimation in Unknown Spatially Correlated Noise Using the EM Algorithm

Aleksandar Dogandžić,[†] *Wei Mo*, and *Zhengdao Wang*
ECpE Department
Iowa State University
3119 Coover Hall, Ames, IA 50010
{ald,mowei,zhengdao}@iastate.edu
Fax: (515) 294-8432
Phone: (515) 294-0500 (A. Dogandžić), (515) 294-2307 (W. Mo), (515) 294-8362 (Z. Wang)

Abstract

We present a maximum likelihood (ML) method for semi-blind estimation of single-input multi-output (SIMO) flat-fading channels in spatially correlated noise having unknown covariance. An expectation-maximization (EM) algorithm is utilized to compute the ML estimates of the channel and spatial noise covariance. We derive the Cramér-Rao bound (CRB) matrix for the unknown parameters and present a symbol detector that utilizes the EM channel estimates. Numerical simulations demonstrate the performance of the proposed method.

I. Introduction

Expectation-maximization (EM) and related algorithms (see [1]–[3]) have been applied to carrier phase recovery [4], demodulation for unknown carrier phase [5], timing estimation [6], and channel estimation [7]–[9] in single-input single-output (SISO) communication systems, and, more recently, to symbol detection [10]–[12] and channel estimation [13]–[15] in *smart antenna* systems. In this paper (see also [17]), we treat the unknown symbols as the *unobserved* (or missing) data and propose an EM algorithm for semi-blind maximum likelihood (ML) estimation of *both* the channel and spatial noise covariance in a single-input multi-output (SIMO) smart antenna scenario. This is unlike previous work in [7]–[9] and [13]–[15], where EM algorithms were applied to channel estimation in white noise.

The signal and noise models are introduced in Section II. In Section III, we derive the EM algorithm for estimating the unknown channel and noise parameters and in Section IV, we compute the Cramér-Rao bound (CRB) matrix for these parameters. The EM channel estimates are incorporated into the receiver design in Section V. In Section VI, we give some numerical examples, and conclude the paper in Section VII.

II. Measurement Model

Consider a single-input multi-output (SIMO) flat-fading channel with equiprobable constant-modulus symbols. Denote by $\boldsymbol{y}(t)$ an $n_{\mathrm{R}} \times 1$ data vector (snapshot) received by an array of n_{R} antennas at time t and assume that we have collected N snapshots from a coherent interval containing the unknown symbols. Under a single-user slow flat-fading scenario, $\boldsymbol{y}(t)$ can be modeled as

$$\boldsymbol{y}(t) = \boldsymbol{h} \cdot u(t) + \boldsymbol{e}(t), \quad t = 1, 2, \ldots, N, \tag{2.1}$$

[†]Corresponding author.

where

- $\boldsymbol{h}$ is an unknown $n_{\mathrm{R}} \times 1$ channel response vector;
- $u(t)$ is an unknown symbol received by the array at time t;
- $\boldsymbol{e}(t)$ is temporally white and circularly symmetric zero-mean complex Gaussian noise vector with unknown positive definite spatial covariance matrix Σ.

The channel $\boldsymbol{h}$ and noise covariance matrix Σ are assumed to be constant for $t \in \{1, 2, \ldots, N\}$. The spatially correlated noise model accounts for co-channel interference (CCI) and receiver noise[1]. We further assume that the symbols $u(t)$ belong to an M-ary constant-modulus constellation

$$\mathcal{U} = \{u_1, u_2, \ldots, u_M\}, \tag{2.2a}$$

where

$$|u_m| = 1, \quad m = 1, 2, \ldots, M. \tag{2.2b}$$

[The constant-modulus assumption can be relaxed, see Appendix A.] We model $u(t), \ t = 1, 2, \ldots, N$ as independent, identically distributed (i.i.d.) random variables with probability mass function

$$p(u(t)) = \frac{1}{M}\, i(u(t)), \tag{2.3}$$

where

$$i(u) = \begin{cases} 1, & u \in \mathcal{U}, \\ 0, & \text{otherwise} \end{cases}. \tag{2.4}$$

Our goal is to estimate the unknown channel and noise parameters $\boldsymbol{h}$ and Σ. To allow unique estimation of the channel $\boldsymbol{h}$ (e.g. to resolve the phase ambiguity), we further assume that N_{T} *known* (training) symbols

$$u_{\mathrm{T}}(\tau) \in \mathcal{U}, \quad \tau = 1, 2, \ldots, N_{\mathrm{T}} \tag{2.5}$$

are embedded in the transmission scheme and denote the corresponding snapshots received by the array as $\boldsymbol{y}_{\mathrm{T}}(\tau), \ \tau = 1, 2, \ldots, N_{\mathrm{T}}$. Then, the measurement model (2.1) holds for the training symbols as well, with $\boldsymbol{y}(t)$ and $u(t)$ replaced by $\boldsymbol{y}_{\mathrm{T}}(\tau)$ and $u_{\mathrm{T}}(\tau)$, respectively.

In the following, we present an EM algorithm for computing the ML estimates of $\boldsymbol{h}$ and Σ under the above measurement model.

III. ML Estimation

The EM algorithm is a general iterative method for computing ML estimates in the scenarios where ML estimation cannot be easily performed by directly maximizing the likelihood function of observed measurements. Each EM iteration consists of maximizing the expected complete-data log-likelihood function, where the expectation is computed with respect to the conditional distribution of the unobserved data given the observed measurements. A good choice of unobserved data allows easy maximization of the expected

[1]This noise model has been used in numerous recent publications to account for unstructured interference, see e.g. [18] and references therein.

complete-data log-likelihood. The algorithm converges monotonically to a local or the global maximum of the observed-data likelihood function, see e.g. [3, Ch. 3]. Here, the unknown symbols $u(t),\ t = 1, 2, \ldots, N$ are modeled as the unobserved (or missing) data. Given $u(t)$, the corresponding *observed* snapshot $\boldsymbol{y}(t)$ is distributed as a complex multivariate Gaussian vector with probability density function (pdf):

$$f(\boldsymbol{y}(t)|u(t), \boldsymbol{h}, \Sigma) = \frac{1}{|\pi\Sigma|} \cdot \exp\Big\{ - [\boldsymbol{y}(t) - \boldsymbol{h}u(t)]^H \Sigma^{-1} [\boldsymbol{y}(t) - \boldsymbol{h}u(t)]\Big\}, \tag{3.1}$$

where "^{H}" denotes the Hermitian (conjugate) transpose. The above expression also holds for the training data, with $\boldsymbol{y}(t)$ and $u(t)$ replaced by $\boldsymbol{y}_{\mathrm{T}}(\tau)$ and $u_{\mathrm{T}}(\tau)$. The joint distribution and mass function of $\boldsymbol{y}(t)$, $u(t)$ (for $t = 1, 2, \ldots, N$), and $\boldsymbol{y}_{\mathrm{T}}(\tau)$ (for $\tau = 1, 2, \ldots, N_{\mathrm{T}}$) can be written as

$$\prod_{t=1}^{N} p(u(t))\ f(\boldsymbol{y}(t)|u(t), \boldsymbol{h}, \Sigma) \cdot \prod_{\tau=1}^{N_{\mathrm{T}}} f(\boldsymbol{y}_{\mathrm{T}}(\tau)|u_{\mathrm{T}}(\tau), \boldsymbol{h}, \Sigma), \tag{3.2}$$

which is also known as the *complete-data likelihood function*. The *observed-data likelihood function* to be maximized is then

$$\begin{aligned} &\Big[\sum_{u(1)\in\mathcal{U}} \sum_{u(2)\in\mathcal{U}} \cdots \sum_{u(N)\in\mathcal{U}} \prod_{t=1}^{N} p(u(t))\ f(\boldsymbol{y}(t)|u(t), \boldsymbol{h}, \Sigma)\Big] \cdot \prod_{\tau=1}^{N_{\mathrm{T}}} f(\boldsymbol{y}_{\mathrm{T}}(\tau)|u_{\mathrm{T}}(\tau), \boldsymbol{h}, \Sigma) \\ &= \prod_{t=1}^{N} \Big[\sum_{m=1}^{M} \frac{1}{M} \cdot f(\boldsymbol{y}(t)|u_m, \boldsymbol{h}, \Sigma)\Big] \cdot \prod_{\tau=1}^{N_{\mathrm{T}}} f(\boldsymbol{y}_{\mathrm{T}}(\tau)|u_{\mathrm{T}}(\tau), \boldsymbol{h}, \Sigma). \end{aligned} \tag{3.3}$$

In Appendix A, we derive the EM algorithm for maximizing (3.3): iterate between

Step 1:

$$\boldsymbol{h}^{(k+1)} = \frac{1}{N + N_{\mathrm{T}}} \Big\{ \sum_{t=1}^{N} \Big[\boldsymbol{y}(t) \sum_{m=1}^{M} u_m^* \cdot \rho_m^{(k)}(t)\Big] + \sum_{\tau=1}^{N_{\mathrm{T}}} \boldsymbol{y}_{\mathrm{T}}(\tau) u_{\mathrm{T}}(\tau)^* \Big\}, \tag{3.4a}$$

where

$$\rho_m^{(k)}(t) = \frac{\exp\{-[\boldsymbol{y}(t) - \boldsymbol{h}^{(k)}u_m]^H (\Sigma^{(k)})^{-1} [\boldsymbol{y}(t) - \boldsymbol{h}^{(k)}u_m]\}}{\sum_{n=1}^{M} \exp\{-[\boldsymbol{y}(t) - \boldsymbol{h}^{(k)}u_n]^H (\Sigma^{(k)})^{-1} [\boldsymbol{y}(t) - \boldsymbol{h}^{(k)}u_n]\}}, \quad \text{and} \tag{3.4b}$$

Step 2:

$$\Sigma^{(k+1)} = R_{\boldsymbol{yy}} - \boldsymbol{h}^{(k+1)}(\boldsymbol{h}^{(k+1)})^H. \tag{3.5}$$

Here

$$R_{\boldsymbol{yy}} = \frac{1}{N + N_{\mathrm{T}}} \Big[\sum_{t=1}^{N} \boldsymbol{y}(t)\boldsymbol{y}(t)^H + \sum_{\tau=1}^{N_{\mathrm{T}}} \boldsymbol{y}_{\mathrm{T}}(\tau)\boldsymbol{y}_{\mathrm{T}}(\tau)^H \Big] \tag{3.6}$$

is the sample correlation matrix of the observed data and "*" denotes complex conjugation. Note that the terms in the summation over t in (3.4a) can be computed in parallel. To ensure that the estimates of the spatial noise covariance matrix in (3.5) are positive definite with probability one, the following condition should be satisfied:

$$N + N_{\mathrm{T}} \geq n_{\mathrm{R}} + 1, \tag{3.7}$$

see also the discussion in Appendix A. Expression (3.4b) can be further simplified by canceling out terms in the numerator and denominator:

$$\rho_m^{(k)}(t) = \frac{\exp[2\,\mathrm{Re}\{\boldsymbol{y}(t)^H (\Sigma^{(k)})^{-1} \boldsymbol{h}^{(k)} u_m\}]}{\sum_{n=1}^{M} \exp[2\,\mathrm{Re}\{\boldsymbol{y}(t)^H (\Sigma^{(k)})^{-1} \boldsymbol{h}^{(k)} u_n\}]}, \tag{3.8}$$

where we have used the constant-modulus property of the transmitted symbols. In the $(k+1)$st iteration, Step 2 requires computing $(\Sigma^{(k)})^{-1}$, which can be done using the matrix inversion lemma in e.g. [20, Cor. 18.2.10]:

$$(\Sigma^{(k)})^{-1} = R_{\boldsymbol{yy}}^{-1} + \frac{R_{\boldsymbol{yy}}^{-1}\boldsymbol{h}^{(k)}(\boldsymbol{h}^{(k)})^H R_{\boldsymbol{yy}}^{-1}}{1-(\boldsymbol{h}^{(k)})^H R_{\boldsymbol{yy}}^{-1}\boldsymbol{h}^{(k)}}, \tag{3.9a}$$

where $R_{\boldsymbol{yy}}^{-1}$ needs to be evaluated only once, before the iteration starts. Then, $(\Sigma^{(k)})^{-1}\boldsymbol{h}^{(k)}$ simplifies to:

$$(\Sigma^{(k)})^{-1}\boldsymbol{h}^{(k)} = \frac{R_{\boldsymbol{yy}}^{-1}\boldsymbol{h}^{(k)}}{1-(\boldsymbol{h}^{(k)})^H R_{\boldsymbol{yy}}^{-1}\boldsymbol{h}^{(k)}}. \tag{3.9b}$$

In the following, we discuss phase correction of the EM channel estimates.

A. Phase Correction

We describe a method for correcting the phases of the channel estimates in the EM iteration. Observe that the first product term in (3.3) is due to the unknown symbols, whereas the second term

$$\prod_{\tau=1}^{N_{\mathrm{T}}} f(\boldsymbol{y}_{\mathrm{T}}(\tau)|u_{\mathrm{T}}(\tau), \boldsymbol{h}, \Sigma) \tag{3.10}$$

is due to the training symbols, and is equal to the likelihood function for the case where *only* the training data are available. For i.i.d. constant-modulus symbols considered here [see (2.3)], the first term in (3.3) has M equal maxima (due to the phase ambiguity), which could cause the above EM iteration to converge to a local maximum of the likelihood function. We correct the phase of the EM channel estimates $\boldsymbol{h}^{(k)}$ to ensure that (3.10) is maximized. For example, for a QPSK constellation, we find which of the following four vectors: $\boldsymbol{h}^{(k)}$, $\boldsymbol{h}^{(k)}\exp(j\pi/2)$, $\boldsymbol{h}^{(k)}\exp(-j\pi/2)$, and $\boldsymbol{h}^{(k)}\exp(j\pi)$ maximizes the training-data likelihood function in (3.10) and update $\boldsymbol{h}^{(k)}$ accordingly. This test is computationally very efficient and may not need to be performed at every step of the EM iteration.

IV. Cramér-Rao Bound

We derive the CRB matrix for the unknown parameters under the measurement model in Section II. First, define the vector of the unknown channel and noise parameters $\boldsymbol{\zeta} = [\boldsymbol{\eta}^T, \boldsymbol{\psi}^T]^T$, where $\boldsymbol{\eta} = [\mathrm{Re}(\boldsymbol{h})^T, \mathrm{Im}(\boldsymbol{h})^T]^T$ and $\boldsymbol{\psi} = [\mathrm{Re}\{\mathrm{vech}(\Sigma)\}^T, \mathrm{Im}\{\underline{\mathrm{vech}}(\Sigma)\}^T]^T$. (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.) Define also the vector of the observed data

$$\boldsymbol{v} = [\boldsymbol{y}(1)^T, \boldsymbol{y}(2)^T, \ldots, \boldsymbol{y}(N)^T, \boldsymbol{y}_{\mathrm{T}}(1)^T, \boldsymbol{y}_{\mathrm{T}}(2)^T, \ldots, \boldsymbol{y}_{\mathrm{T}}(N_{\mathrm{T}})^T]^T \tag{4.1a}$$

and the vector of the unobserved data

$$\boldsymbol{u} = [u(1), u(2), \ldots, u(N)]^T. \tag{4.1b}$$

Then, the CRB matrix for the unknown parameters $\boldsymbol{\zeta}$ is computed as (see [3, Ch. 3.8.1]):

$$\mathrm{CRB}(\boldsymbol{\zeta}) = \left\{\mathrm{E}_{\boldsymbol{v}}[\boldsymbol{s}(\boldsymbol{v};\boldsymbol{\zeta})\boldsymbol{s}(\boldsymbol{v};\boldsymbol{\zeta})^T]\right\}^{-1}, \tag{4.2}$$

where the expectation is performed with respect to the distribution of $\boldsymbol{v}$ and $\boldsymbol{s}(\boldsymbol{v};\boldsymbol{\zeta})$ is the observed-data score vector. The observed-data score vector can be computed as (see [3, eq. (3.42)]):

$$\boldsymbol{s}(\boldsymbol{v};\boldsymbol{\zeta}) = \mathrm{E}_{\boldsymbol{u}|\boldsymbol{v}}[\boldsymbol{s}_{\mathrm{c}}(\boldsymbol{v},\boldsymbol{u};\boldsymbol{\zeta})|\boldsymbol{v}], \tag{4.3}$$

where $\boldsymbol{s}_{\mathrm{c}}(\boldsymbol{v},\boldsymbol{u};\boldsymbol{\zeta})$ is the complete-data score vector, obtained by differentiating the complete-data log-likelihood function [i.e. the logarithm of (3.2)] with respect to $\boldsymbol{\zeta}$. Computing the expectations in (4.2) and (4.3) is discussed in Appendix B, where the expression for $\boldsymbol{s}_{\mathrm{c}}(\boldsymbol{v},\boldsymbol{u};\boldsymbol{\zeta})$ is also given.

V. Detection

We now utilize the channel and noise estimates proposed in Section III to detect the unknown transmitted symbols $u(t)$. We apply the (estimated) maximum *a posteriori* (MAP) detector:

$$\begin{aligned}
\widehat{u}(t) &= \arg\max_{u(t)\in\mathcal{U}} \frac{\exp\{-[\boldsymbol{y}(t)-\widehat{\boldsymbol{h}}u(t)]^H\widehat{\Sigma}^{-1}[\boldsymbol{y}(t)-\widehat{\boldsymbol{h}}u(t)]\}}{\sum_{n=1}^{M}\exp\{-[\boldsymbol{y}(t)-\widehat{\boldsymbol{h}}u_n]^H\widehat{\Sigma}^{-1}[\boldsymbol{y}(t)-\widehat{\boldsymbol{h}}u_n]\}} \\
&= \arg\min_{u(t)\in\mathcal{U}} [\boldsymbol{y}(t)-\widehat{\boldsymbol{h}}u(t)]^H\widehat{\Sigma}^{-1}[\boldsymbol{y}(t)-\widehat{\boldsymbol{h}}u(t)] && \text{(5.1a)} \\
&= \arg\max_{u(t)\in\mathcal{U}} \mathrm{Re}\{\boldsymbol{y}(t)^H R_{\boldsymbol{yy}}^{-1}\widehat{\boldsymbol{h}}\cdot u(t)\}, && \text{(5.1b)}
\end{aligned}$$

where $\widehat{\boldsymbol{h}} = \boldsymbol{h}^{(\infty)}$ and $\widehat{\Sigma} = \Sigma^{(\infty)}$ are the ML estimates obtained from the EM iteration (3.4)–(3.5) upon convergence. To derive (5.1b), we have used the identity (3.9b) and the constant-modulus property of the transmitted symbols. Interestingly, the detector in (5.1b) is a function of the channel estimate $\widehat{\boldsymbol{h}}$ *only*, through the $R_{\boldsymbol{yy}}^{-1}\widehat{\boldsymbol{h}}$ term. Note that the above detection problem is equivalent to finding $m \in \{1,2,\ldots,M\}$ that maximizes $\rho_m^{(\infty)}(t)$ in (3.8) [see also (3.4b)]. The detector (5.1) and EM algorithm (3.4)–(3.5) can be easily modified to account for unequal prior probabilities of the transmitted symbols.

VI. Simulation Results

We evaluate the performance of the proposed estimation and detection algorithms using numerical simulations. We consider an array of $n_{\mathrm{R}} = 5$ receiver antennas. Our performance metrics are the mean-square error (MSE) and symbol error rate (SER), averaged over 5000 random channel realizations generated using an i.i.d. Rayleigh fading model with unit-variance channel coefficients. The transmitted symbols were generated from an uncoded QPSK modulated constellation (i.e. $M = 4$) with normalized energy. We added a three-symbol training sequence ($N_{\mathrm{T}} = 3$), which was utilized to obtain the initial channel estimate $\boldsymbol{h}^{(0)}$, computed using least squares. The initial estimate of the noise covariance matrix was chosen as $\Sigma^{(0)} = R_{\boldsymbol{yy}}$. The signal was corrupted by additive complex Gaussian noise with spatial noise covariance matrix Σ whose (p,q)th element is

$$\Sigma_{p,q} = \sigma^2\cdot 0.9^{|p-q|}\cdot\exp[j(\pi/2)(p-q)], \tag{6.1}$$

which is the noise covariance model used in [22] (see also references therein). The bit signal-to-noise ratio (SNR) per receiver antenna was defined as

$$\mathrm{SNR} = 10\log_{10}\left[\frac{1}{\sigma^2\cdot\log_2(M)}\right] = 10\log_{10}\left(\frac{1}{2\sigma^2}\right) \quad \text{(dB)}. \tag{6.2}$$

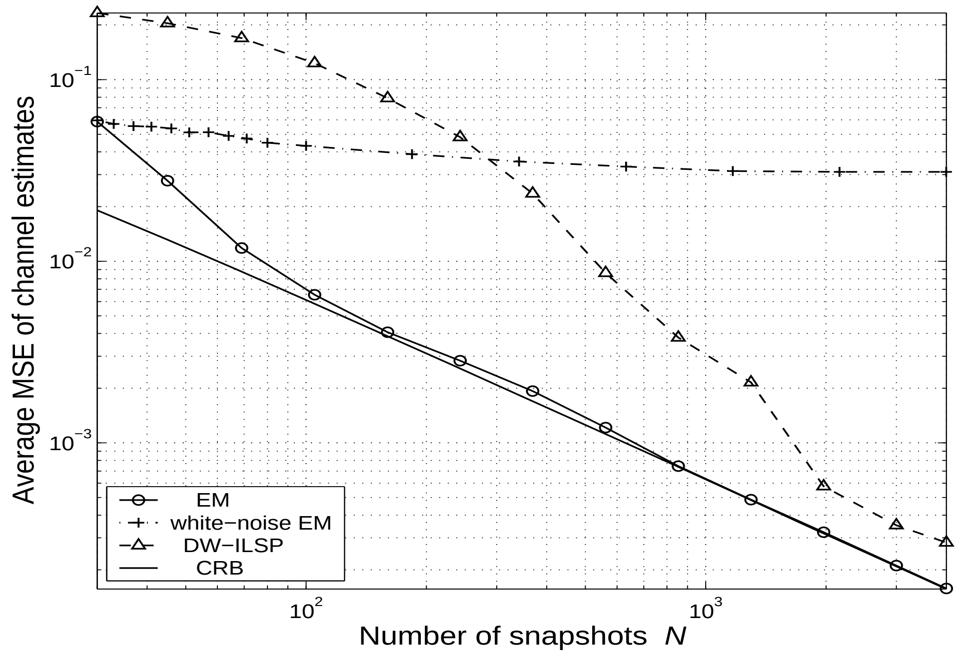


Fig. 1. Average mean-square errors and corresponding Cramér-Rao bounds for the channel estimates obtained using the proposed EM algorithm, DW-ILSP method, and an EM algorithm for spatially white noise, as functions of N for $N_T = 3$ and $\mathrm{SNR} = -1$ dB.

In the cases where the EM algorithm did not converge within 40 iterations, it was restarted using a randomly selected initial value for the channel coefficients[2]. [We implemented the same restart procedure in all algorithms whose performance is analyzed in this section.] We also applied the phase correction technique in Section III-A at every step of the EM iteration.

In the first set of simulations, the bit SNR was set to -1 dB. In Figs. 1 and 2, we show the average MSEs (and corresponding average CRBs) for the ML estimates of the channel coefficients[3] and selected elements of the spatial noise covariance matrix Σ (obtained using the proposed EM algorithm) as functions of the block length N. Fig. 1 also compares the MSE performance of the proposed EM algorithm with

- the decoupled weighted iterative least squares with projection (DW-ILSP) method in [23] and
- an EM algorithm that assumes spatially white noise.

For low SNR (-1 dB), few training symbols ($N_{\mathrm{T}} = 3$) and short block lengths, the proposed EM algorithm clearly outperforms the DW-ILSP method. In this scenario, the proposed method attains the CRB for $N = 100$ symbols, compared with more than 4000 symbols needed for the DW-ILSP method. [Note also that in fading channels the block length N is limited by the coherence time of the channel.] The average numbers of iterations needed for the EM, white-noise EM, and DW-ILSP algorithms to converge were $9, 9,$ and 6, respectively. For $N = 100$, restart was needed in less than 0.1% of the total number of trials. A single EM iteration has higher computational complexity than a DW-ILSP iteration for the same N, and the complexity

[2]Note that fast convergence of the EM algorithm or utilizing the above restart method do not guarantee convergence to the global maximum of the observed-data likelihood function. Hence, our simulation results represent upper bounds on the performance achievable by the exact ML method.

[3]Here, averaging is performed over *both* the channel coefficients for different antennas (i.e. elements of $\widehat{\boldsymbol{h}}$) and random channel and training-data realizations.

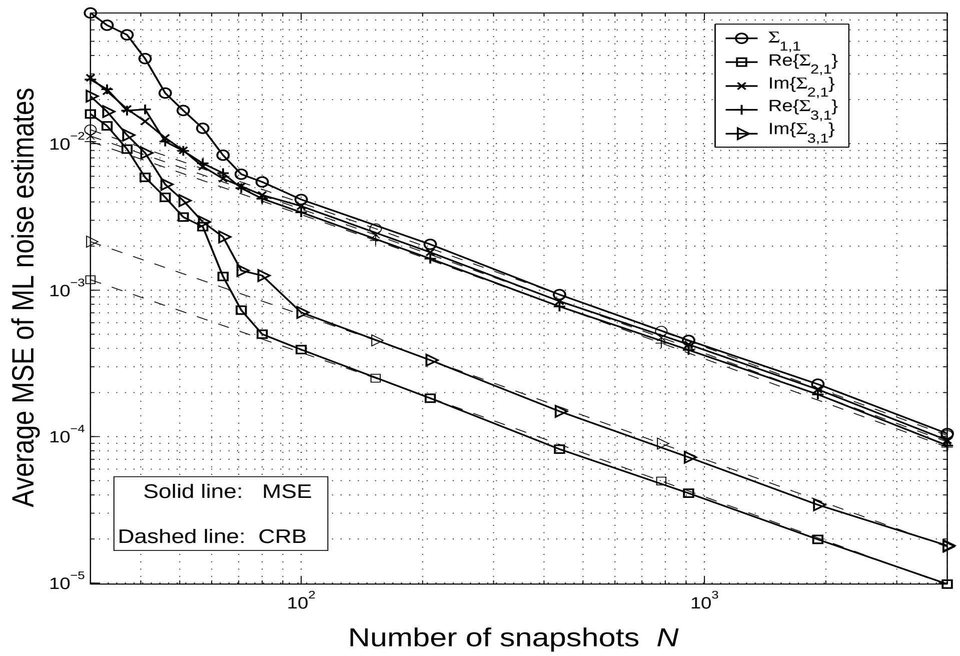


Fig. 2. Average mean-square errors and corresponding Cramér-Rao bound for the ML estimates of $\Sigma_{1,1}$, $\mathrm{Re}\{\Sigma_{2,1}\}$, $\mathrm{Im}\{\Sigma_{2,1}\}$, $\mathrm{Re}\{\Sigma_{3,1}\}$, $\mathrm{Im}\{\Sigma_{3,1}\}$ obtained using the proposed EM algorithm, as functions of N for $N_T = 3$ and $\mathrm{SNR} = -1$ dB.

of both iterations increases linearly with N. However, the proposed EM algorithm typically needs a smaller N to attain the same MSE. To demonstrate the importance of incorporating the spatial color of the noise in channel estimation, we also show the performance of an EM algorithm that assumes spatially white noise in the scenario where the noise is *colored* [with covariance (6.1)]. The EM algorithm for spatially white noise follows from (3.4)–(3.5) by substituting $\Sigma^{(k)} = (\widehat{\sigma}^2)^{(k)} I_{n_{\mathrm{R}}}$ into Step 1 in (3.4) and applying the following Step 2: $(\widehat{\sigma}^2)^{(k+1)} = \mathrm{tr}(\Sigma^{(k+1)})/n_{\mathrm{R}}$, where $\Sigma^{(k+1)}$ was defined in (3.5), and $I_{n_{\mathrm{R}}}$ denotes the identity matrix of size n_{R}. For low SNR (-1 dB) and few training symbols ($N_{\mathrm{T}} = 3$), the white-noise EM algorithm breaks down.

In Fig. 2, we show the average MSEs for the ML estimates of $\Sigma_{1,1}$, $\mathrm{Re}\{\Sigma_{2,1}\}$, $\mathrm{Im}\{\Sigma_{2,1}\}$, $\mathrm{Re}\{\Sigma_{3,1}\}$, and $\mathrm{Im}\{\Sigma_{3,1}\}$ (obtained using the proposed EM algorithm) and the corresponding CRBs as functions of N.

In Fig. 3, the average MSEs for the channel estimates obtained by the proposed EM algorithm for spatially correlated noise, DW-ILSP method, and EM algorithm for spatially white noise are shown as functions of the bit SNR per receiver antenna for block lengths $N = 50, 100,$ and 150. When the average MSE is 0.03 and $N = 100$, the EM algorithm has an advantage of about 9 dB over the DW-ILSP algorithm; this advantage further grows as N decreases. An intuitive explanation for this performance improvement is that the EM algorithm exploits additional information provided by the prior distribution of the unknown symbols in (2.3). Note also that the number of parameters in the random-symbol measurement model in Section II equals $n_{\mathrm{R}}^2 + 2n_{\mathrm{R}}$, and, therefore, is independent of N. This is in contrast with the DW-ILSP and other deterministic ML methods (e.g. [24], see also [25]) where the number of parameters grows with N. For low SNRs, the white-noise EM algorithm performs poorly, see also Fig. 1. However, for high SNRs and small

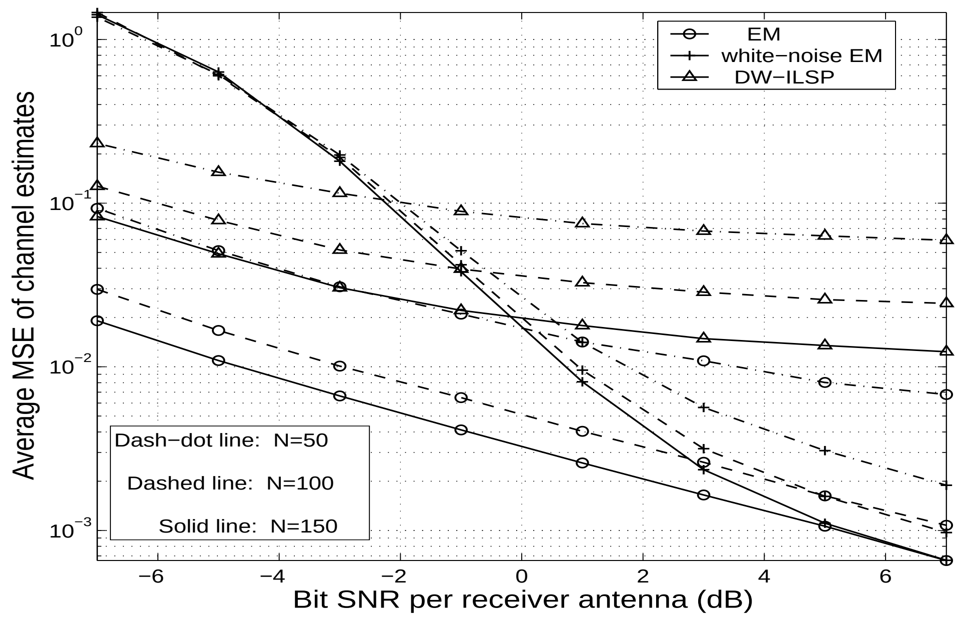


Fig. 3. Average mean-square errors for the channel estimates obtained using the proposed EM algorithm, DW-ILSP method, and EM algorithm for spatially white noise, as functions of the bit SNR per receiver antenna for block lengths $N = 50, 100$, and 150.

block lengths, it outperforms the EM algorithm for spatially correlated noise. Hence, in this scenario, the fact that the white-noise EM algorithm estimates a small number of parameters ($2n_{\mathrm{R}} + 1$) becomes more important than accounting for spatial noise covariance (which, in addition, is poorly estimated due to the small block length).

In Fig. 4, we compare symbol error rates of the detector (5.1) which uses the ML estimates of $\boldsymbol{h}$ and Σ [obtained from the EM iteration (3.4)–(3.5)] with

- the DW-ILSP detector in [23] and
- a white-noise detector

$$\arg \max_{u(t)\in\mathcal{U}} \mathrm{Re}\{\boldsymbol{y}(t)^H \widehat{\boldsymbol{h}}_{\text{white EM}} \cdot u(t)\}, \tag{6.3}$$

where $\widehat{\boldsymbol{h}}_{\text{white EM}}$ is computed using the EM algorithm for spatially white noise.

The symbol error rates are shown as functions of the bit SNR per receiver antenna for block lengths $N = 50, 100$, and 150. For the given range of SNRs and block lengths, the proposed detector significantly outperforms the DW-ILSP detector. As expected, the white-noise detector performs poorly for low SNRs due to poor channel estimates provided by the white-noise EM algorithm. Similarly, for high SNRs and small block lengths it outperforms the detector in (5.1) due to the fact that the white-noise EM algorithm outperforms the EM algorithm for spatially correlated noise in this scenario. The performance of the detector (5.1) improves significantly as the block length increases due to the improved channel estimation. In contrast, the performance of the white-noise detector is insensitive to the choice of the block length (for the block lengths considered in Fig. 4), which can be explained by the fact that the white-noise EM algorithm estimates a small number of parameters (and thus requires a relatively small data size).

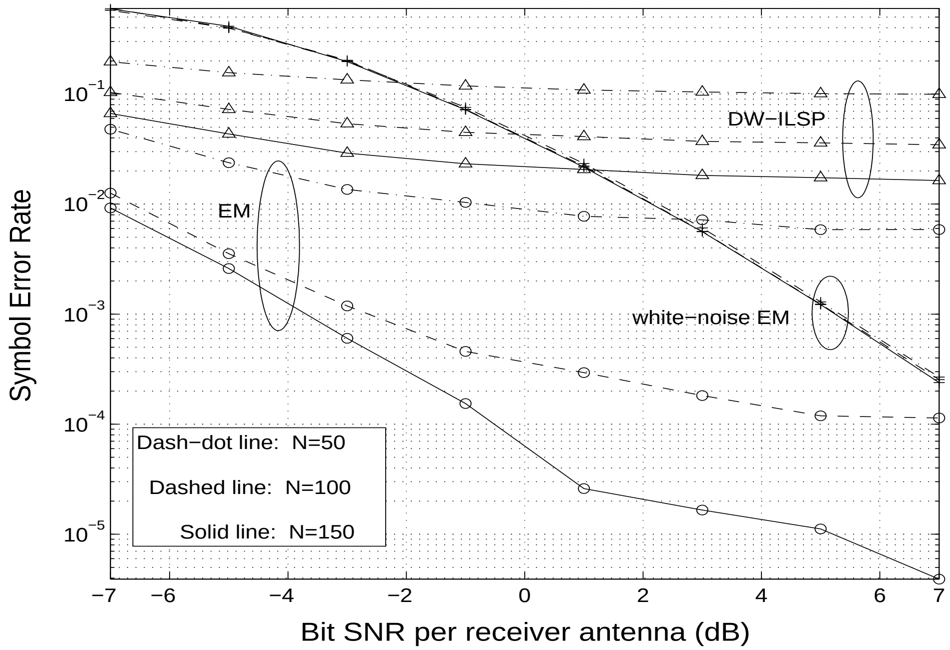


Fig. 4. Symbol error rates of the EM-based and DW-ILSP detectors, as functions of the bit SNR per receiver antenna for block lengths $N = 50, 100$, and 150.

VII. Concluding Remarks

We developed an expectation-maximization algorithm for semi-blind estimation of single-input multi-output fading channels in spatially correlated noise having unknown covariance. We also derived a method for phase correction of the EM channel estimates and computed the Cramér-Rao bounds for the unknown parameters. The proposed channel and noise estimators were incorporated into the receiver design. We presented numerical simulations that demonstrated the performance of the proposed methods, and compared them with the existing techniques.

Further research will include extending the proposed methods to the multi-input multi-output (MIMO) scenario and reducing their computational complexity. For coded transmission, we will develop iterative schemes that combine EM channel estimation with decoding under the correlated noise scenario, generalizing the white-noise based methods in e.g. [26] and [27].

Appendix A. EM Algorithm Derivation

We relax the constant-modulus assumption (2.2b) and first derive the EM algorithm for the general case where the symbols $u(t),\ t = 1, 2, \ldots, N,\ u_{\mathrm{T}}(\tau),\ \tau = 1, 2, \ldots, N_{\mathrm{T}}$ belong to an arbitrary constellation. This algorithm is then simplified to the constant-modulus scenario in Sections II and III.

By taking the logarithm of (3.2) and neglecting terms that do not depend on the parameters $\boldsymbol{h}$ and Σ, we obtain the complete-data log-likelihood function

$$\mathcal{L}(\boldsymbol{h}, \Sigma) = -(N + N_{\mathrm{T}}) \cdot \big\{ \ln |\Sigma| + \mathrm{tr}[\Sigma^{-1} \cdot (R_{\boldsymbol{yy}} - \boldsymbol{r}_{\boldsymbol{yu}} \boldsymbol{h}^H - \boldsymbol{h} \boldsymbol{r}_{\boldsymbol{yu}}^H + r_{\boldsymbol{uu}} \boldsymbol{h} \boldsymbol{h}^H)] \big\}, \tag{A.1}$$

where $|\cdot|$ denotes the determinant and

$$R_{\boldsymbol{yy}} = \frac{1}{N+N_{\mathrm{T}}}\Big[\sum_{t=1}^{N}\boldsymbol{y}(t)\boldsymbol{y}(t)^H + \sum_{\tau=1}^{N_{\mathrm{T}}}\boldsymbol{y}_{\mathrm{T}}(\tau)\boldsymbol{y}_{\mathrm{T}}(\tau)^H\Big], \tag{A.2a}$$

$$\boldsymbol{r}_{\boldsymbol{yu}} = \frac{1}{N+N_{\mathrm{T}}}\Big[\sum_{t=1}^{N}\boldsymbol{y}(t)u(t)^* + \sum_{\tau=1}^{N_{\mathrm{T}}}\boldsymbol{y}_{\mathrm{T}}(\tau)u_{\mathrm{T}}(\tau)^*\Big], \tag{A.2b}$$

$$r_{\boldsymbol{uu}} = \frac{1}{N+N_{\mathrm{T}}}\Big[\sum_{t=1}^{N}|u(t)|^2 + \sum_{\tau=1}^{N_{\mathrm{T}}}|u_{\mathrm{T}}(\tau)|^2\Big], \tag{A.2c}$$

are the *natural complete-data sufficient statistics* for estimating $\boldsymbol{h}$ and Σ, see e.g. [16]. At the kth iteration, the E step computes the conditional expectation of the complete-data log-likelihood given the observed data $\boldsymbol{v}$ [see (4.1a)] at the current parameter estimates $\boldsymbol{h}^{(k)}$ and $\Sigma^{(k)}$:

$$Q(\boldsymbol{h},\Sigma;\boldsymbol{h}^{(k)},\Sigma^{(k)}) = -(N+N_{\mathrm{T}})\cdot\big\{\ln|\Sigma| + \mathrm{tr}[\Sigma^{-1}\cdot(R_{\boldsymbol{yy}} - \boldsymbol{r}_{\boldsymbol{yu}}^{(k)}\boldsymbol{h}^H - \boldsymbol{h}(\boldsymbol{r}_{\boldsymbol{yu}}^{(k)})^H + r_{\boldsymbol{uu}}^{(k)}\boldsymbol{h}\boldsymbol{h}^H)]\big\}, \tag{A.3}$$

where $\boldsymbol{r}_{\boldsymbol{yu}}^{(k)} = \mathrm{E}_{\,u|v}[\boldsymbol{r}_{\boldsymbol{yu}}|\boldsymbol{v};\boldsymbol{h}^{(k)},\Sigma^{(k)}]$ and $r_{\boldsymbol{uu}}^{(k)} = \mathrm{E}_{\,u|v}[r_{\boldsymbol{uu}}|\boldsymbol{v};\boldsymbol{h}^{(k)},\Sigma^{(k)}]$. The above expression is obtained from (A.1) by replacing $\boldsymbol{r}_{\boldsymbol{yu}}$ and $r_{\boldsymbol{uu}}$ with their conditional expectations $\boldsymbol{r}_{\boldsymbol{yu}}^{(k)}$ and $r_{\boldsymbol{uu}}^{(k)}$. The M step maximizes the above Q function with respect to $\boldsymbol{h}$ and Σ to produce

$$\boldsymbol{h}^{(k+1)},\Sigma^{(k+1)} = \arg\max_{\boldsymbol{h},\Sigma} Q(\boldsymbol{h},\Sigma;\boldsymbol{h}^{(k)},\Sigma^{(k)}). \tag{A.4}$$

The maximization of $\mathcal{L}(\boldsymbol{h},\Sigma)$ in (A.1) with respect to $\boldsymbol{h}$ and Σ has well-known solutions given by $\boldsymbol{r}_{\boldsymbol{yu}}/r_{\boldsymbol{uu}}$ and $R_{\boldsymbol{yy}} - \boldsymbol{r}_{\boldsymbol{yu}}\boldsymbol{r}_{\boldsymbol{yu}}^H/r_{\boldsymbol{uu}}$ (respectively), provided that $R_{\boldsymbol{yy}} - \boldsymbol{r}_{\boldsymbol{yu}}\boldsymbol{r}_{\boldsymbol{yu}}^H/r_{\boldsymbol{uu}}$ is a positive definite matrix, see e.g. [18] and [19, Th. 10.1.1]. (These expressions follow from the multivariate analysis of variance (MANOVA) model in multivariate statistical analysis, see [18] and [19].) Hence, the M step is obtained by replacing $\boldsymbol{r}_{\boldsymbol{yu}}$ and $r_{\boldsymbol{uu}}$ in $\boldsymbol{r}_{\boldsymbol{yu}}/r_{\boldsymbol{uu}}$ and $R_{\boldsymbol{yy}} - \boldsymbol{r}_{\boldsymbol{yu}}\boldsymbol{r}_{\boldsymbol{yu}}^H/r_{\boldsymbol{uu}}$ with their conditional expectations and the EM iteration follows:

Step 1:

$$\boldsymbol{h}^{(k+1)} = \frac{1}{N+N_{\mathrm{T}}}\Big\{\sum_{t=1}^{N}\Big[\boldsymbol{y}(t)\sum_{m=1}^{M}u_m^*\cdot\rho_m^{(k)}(t)\Big] + \sum_{\tau=1}^{N_{\mathrm{T}}}\boldsymbol{y}_{\mathrm{T}}(\tau)u_{\mathrm{T}}(\tau)^*\Big\}\Big/ r_{\boldsymbol{uu}}^{(k)}, \tag{A.5a}$$

where

$$r_{\boldsymbol{uu}}^{(k)} = \frac{1}{N+N_{\mathrm{T}}}\Big\{\sum_{t=1}^{N}\sum_{m=1}^{M}[|u_m|^2\cdot\rho_m^{(k)}(t)] + \sum_{\tau=1}^{N_{\mathrm{T}}}|u_{\mathrm{T}}(\tau)|^2\Big\}, \tag{A.5b}$$

Step 2:

$$\Sigma^{(k+1)} = R_{\boldsymbol{yy}} - r_{\boldsymbol{uu}}^{(k)}\cdot\boldsymbol{h}^{(k+1)}(\boldsymbol{h}^{(k+1)})^H, \tag{A.6}$$

where $\rho_m^{(k)}(t)$ is computed using (3.4b). Note that (A.5) and (A.6) each incorporate both E and M steps. The condition (3.7) is needed to ensure that $\Sigma^{(k+1)}$ is a positive definite matrix with probability one, which follows using arguments similar to those in [19, Th. 3.1.4], see also [18, eq. (4)] and [19, Th. 10.1.1].

In the constant-modulus scenario (2.2b), we have that $r_{\boldsymbol{uu}}^{(k)} \equiv 1$ for all k. Hence, setting $r_{\boldsymbol{uu}}^{(k)} = 1$ in (A.5)–(A.6) yields the EM iteration (3.4)–(3.5).

Appendix B. Cramér-Rao Bound

We present the expression for the complete-data score vector $\boldsymbol{s}_{\mathrm{c}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta})$ under the measurement model (2.1)–(2.3) and discuss evaluating the expectations in (4.2) and (4.3), needed for computing the CRB matrix. The complete-data score vector $\boldsymbol{s}_{\mathrm{c}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta})$ for this measurement model is obtained by differentiating the complete-data log-likelihood function (A.1) with respect to $\boldsymbol{\zeta}$ (see [21, App. 15C]) and setting $r_{\boldsymbol{uu}} = 1$:

$$\boldsymbol{s}_{\mathrm{c}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta}) = [\mathrm{Re}\{\boldsymbol{s}_{\mathrm{c},\boldsymbol{h}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta})\}^T, \mathrm{Im}\{\boldsymbol{s}_{\mathrm{c},\boldsymbol{h}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta})\}^T, \boldsymbol{s}_{\mathrm{c},\boldsymbol{\psi}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta})^T]^T, \tag{B.1}$$

where

$$\begin{aligned}
\boldsymbol{s}_{\mathrm{c},\boldsymbol{h}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta}) &= 2 \cdot \Sigma^{-1} \cdot \Big\{ \sum_{t=1}^{N} [\boldsymbol{y}(t) u(t)^* - \boldsymbol{h}] + \sum_{\tau=1}^{N_{\mathrm{T}}} [\boldsymbol{y}_{\mathrm{T}}(\tau) u_{\mathrm{T}}(\tau)^* - \boldsymbol{h}] \Big\} \\
&= 2 \cdot (N + N_{\mathrm{T}}) \cdot \Sigma^{-1} \cdot (\boldsymbol{r}_{\boldsymbol{yu}} - \boldsymbol{h}), \qquad \text{(B.2a)} \\
[\boldsymbol{s}_{\mathrm{c},\boldsymbol{\psi}}(\boldsymbol{v}, \boldsymbol{u}; \boldsymbol{\zeta})]_i &= -(N + N_{\mathrm{T}}) \cdot \mathrm{tr}\Big(\Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Big) + (N + N_{\mathrm{T}}) \cdot \boldsymbol{h}^H \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \boldsymbol{h} \\
&\quad + \sum_{t=1}^{N} \boldsymbol{y}(t)^H \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \boldsymbol{y}(t) + \sum_{\tau=1}^{N_{\mathrm{T}}} \boldsymbol{y}_{\mathrm{T}}(\tau)^H \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \boldsymbol{y}_{\mathrm{T}}(\tau) \\
&\quad - \boldsymbol{h}^H \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \cdot \sum_{t=1}^{N} [\boldsymbol{y}(t) u(t)^*] - \sum_{t=1}^{N} [\boldsymbol{y}(t)^H u(t)] \cdot \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \boldsymbol{h} \\
&\quad - \boldsymbol{h}^H \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \cdot \sum_{\tau=1}^{N_{\mathrm{T}}} [\boldsymbol{y}_{\mathrm{T}}(\tau) u_{\mathrm{T}}(\tau)^*] - \sum_{\tau=1}^{N_{\mathrm{T}}} [\boldsymbol{y}_{\mathrm{T}}(\tau)^H u_{\mathrm{T}}(\tau)] \cdot \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Sigma^{-1} \boldsymbol{h} \\
&= (N + N_{\mathrm{T}}) \cdot \Big\{ - \mathrm{tr}\Big(\Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Big) + \mathrm{tr}\Big(\Sigma^{-1} \boldsymbol{h}\boldsymbol{h}^H \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Big) + \mathrm{tr}\Big(\Sigma^{-1} R_{\boldsymbol{yy}} \Sigma^{-1} \frac{\partial \Sigma}{\partial \psi_i} \Big) \\
&\quad - \mathrm{tr}\Big[(\Sigma^{-1} \boldsymbol{r}_{\boldsymbol{yu}} \boldsymbol{h}^H \Sigma^{-1} + \Sigma^{-1} \boldsymbol{h} \boldsymbol{r}_{\boldsymbol{yu}}^H \Sigma^{-1}) \cdot \frac{\partial \Sigma}{\partial \psi_i} \Big] \Big\}, \quad i = 1, 2, \ldots, n_{\mathrm{R}}^2. \qquad \text{(B.2b)}
\end{aligned}$$

To compute (B.2b), the following identities can be utilized:

$$\mathrm{tr}\Big(A \cdot \frac{\partial \Sigma}{\partial \Sigma_{p,p}} \Big) = A_{p,p}, \quad p = 1, 2, \ldots, n_{\mathrm{R}}, \tag{B.3a}$$

and

$$\mathrm{tr}\Big(A \cdot \frac{\partial \Sigma}{\partial \,\mathrm{Re}\{\Sigma\}_{p,q}} \Big) = 2\,\mathrm{Re}\{A_{p,q}\}, \tag{B.3b}$$

$$\mathrm{tr}\Big(A \cdot \frac{\partial \Sigma}{\partial \,\mathrm{Im}\{\Sigma\}_{p,q}} \Big) = 2\,\mathrm{Im}\{A_{p,q}\}, \quad 1 \le q < p \le n_{\mathrm{R}}, \tag{B.3c}$$

where A is an arbitrary $n_{\mathrm{R}} \times n_{\mathrm{R}}$ Hermitian matrix. It follows from (B.2) that computing the observed-data score vector $\boldsymbol{s}(\boldsymbol{v}; \boldsymbol{\zeta})$ in (4.3) reduces to replacing $\boldsymbol{r}_{\boldsymbol{yu}}$ in (B.2) with its conditional expectation given $\boldsymbol{v}$:

$$\mathrm{E}_{\boldsymbol{u}|\boldsymbol{v}}[\boldsymbol{r}_{\boldsymbol{yu}}|\boldsymbol{v}] = \frac{1}{N + N_{\mathrm{T}}} \Big\{ \sum_{t=1}^{N} \Big[\boldsymbol{y}(t) \sum_{m=1}^{M} u_m^* \cdot \rho_m(t) \Big] + \sum_{\tau=1}^{N_{\mathrm{T}}} \boldsymbol{y}_{\mathrm{T}}(\tau) u_{\mathrm{T}}(\tau)^* \Big\}, \tag{B.4a}$$

where

$$\rho_m(t) = \frac{\exp\{-[\boldsymbol{y}(t) - \boldsymbol{h} u_m]^H \Sigma^{-1} [\boldsymbol{y}(t) - \boldsymbol{h} u_m]\}}{\sum_{n=1}^{M} \exp\{-[\boldsymbol{y}(t) - \boldsymbol{h} u_n]^H \Sigma^{-1} [\boldsymbol{y}(t) - \boldsymbol{h} u_n]\}}. \tag{B.4b}$$

Finally, the CRB matrix is computed using (4.2), which requires multidimensional integration to evaluate the expectation with respect to the distribution of $\boldsymbol{\upsilon}$; this can be performed using Monte Carlo integration, i.e. by averaging $\boldsymbol{s}(\boldsymbol{\upsilon};\boldsymbol{\zeta})\boldsymbol{s}(\boldsymbol{\upsilon};\boldsymbol{\zeta})^T$ over many realizations of $\boldsymbol{\upsilon}$.

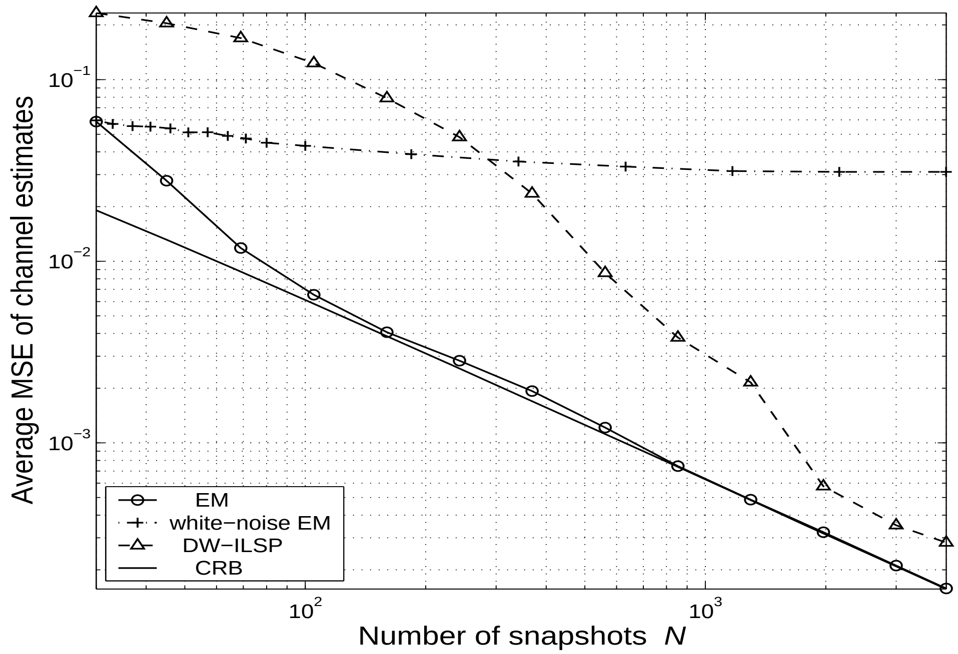
Acknowledgment

The authors are grateful to the anonymous reviewers for their helpful comments.

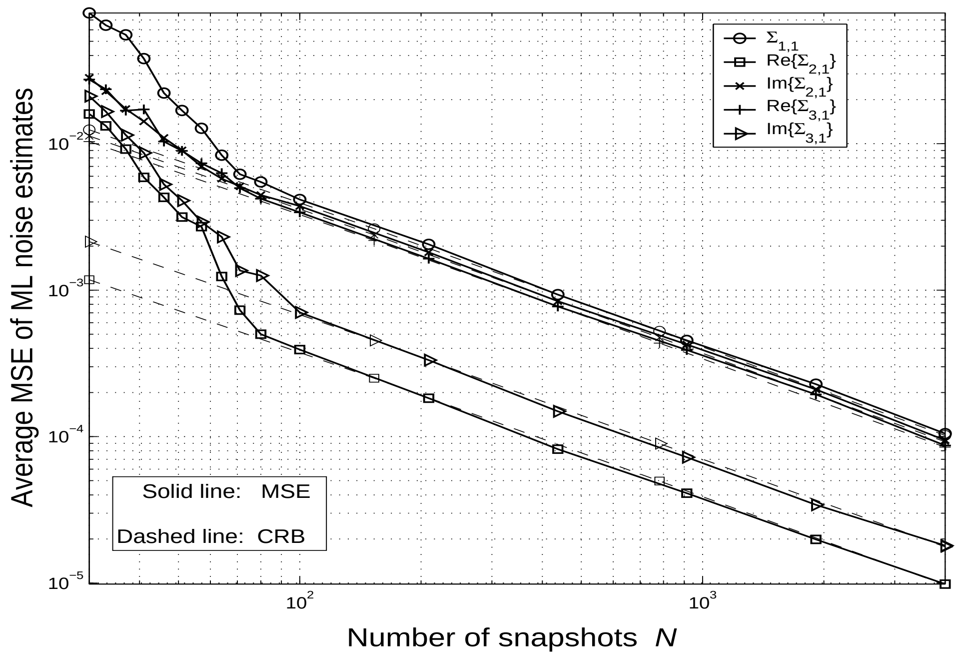
References

[1] 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, July 1977.
[2] T.K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Mag.*, vol. 13, pp. 47–60, Nov. 1996.
[3] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, New York: Wiley, 1997.
[4] M.J. Nissila, S. Pasupathy, and A. Mammela, "An EM approach to carrier phase recovery in AWGN channel," in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, June 2001, pp. 2199–2203.
[5] J. Qin, "Demodulation of binary PSK signals without explicit carrier synchronization," in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 498–501.
[6] C.N. Georghiades and D.L. Snyder, "The expectation-maximization algorithm for symbol unsynchronized sequence detection," *IEEE Trans. Commun.*, vol. 39, pp. 54–61, Jan. 1991.
[7] G.K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, pp. 2406–2413, July 1994.
[8] M. Shao and C.L. Nikias, "An ML/MMSE estimation approach to blind equalization," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Adelaide, Australia, Apr. 1994, pp. IV/569 - IV/572.
[9] H.A. Cirpan and M.K. Tsatsanis, "Stochastic maximum likelihood methods for semi-blind channel estimation," *IEEE Signal Proc. Letters*, vol. 5, pp. 21–24, Jan. 1998.
[10] Y. Li, C.N. Georghiades, and G. Huang, "Iterative maximum-likelihood sequence estimation for space-time coded systems," *IEEE Trans. Commun.*, vol. 49, pp. 948–951, June 2001.
[11] B. Lu, X. Wang and Y. Li, "Iterative receivers for space-time block-coded OFDM systems in dispersive fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 213–225, Apr. 2002.
[12] C. Cozzo and B.L. Hughes, "Joint channel estimation and data symbol detection in space-time communications," in *Proc. IEEE Int. Conf. Commun.*, New Orleans, LA, June 2000, pp. 287–291.
[13] C.H. Aldana and J. Cioffi, "Channel tracking for multiple input, single output systems using EM algorithm," in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, June 2001, pp. 586–590.
[14] J. Míguez and L. Castedo, "Semiblind space-time decoding in wireless communications: A maximum likelihood approach," *Signal Processing*, vol. 82, pp. 1–18, Jan. 2002.
[15] H.A. Çirpan, E. Panayirci, and E. Çekli, "Maximum likelihood blind channel estimation for space-time coding systems," *EURASIP J. Appl. Signal Processing*, vol. 5, pp. 497–506, May 2002.
[16] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd ed., Upper Saddle River, NJ:Prentice Hall, 2000.
[17] A. Dogandžić, W. Mo, and Z. Wang, "Maximum likelihood semi-blind channel and noise estimation using the EM algorithm," in *Proc. 37th Annu. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 2003, pp. WA3.1–WA3.6.
[18] A. Dogandžić and A. Nehorai, "Generalized multivariate analysis of variance: A unified framework for signal processing in correlated noise," to appear in *IEEE Signal Processing Mag.*, vol. 20, no. 4, Sept. 2003.
[19] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, New York: Wiley, 1982.
[20] D.A. Harville, *Matrix Algebra From a Statistician's Perspective*, New York: Springer-Verlag, 1997.
[21] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993, pt. I.
[22] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 996–1004, Apr. 1997.
[23] A. Ranheim, "A decoupled approach to adaptive signal separation using an antenna array," *IEEE Trans. Veh. Tech.*, vol. 48, pp. 676–682, May 1999.
[24] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 1000–1011, Feb.–Apr. 1994.
[25] A. Dogandžić and A. Nehorai, "Finite-length MIMO adaptive equalization using canonical correlation analysis," *IEEE Trans. Signal Processing*, vol. 50, pp. 984–989, Apr. 2002.
[26] Z. Baranski, A.M. Haimovich, and J. Garcia-Frias, "EM-based iterative receiver for space-time coded modulation with noise variance estimation," *Proc. Globecom Conf.*, Taipei, Taiwan, R.O.C., Nov. 2002, pp. 355–359.
[27] R.R. Lopes and J.R. Barry, "Exploiting error-control coding in blind channel estimation," *Proc. Globecom Conf.*, San Antonio, TX, Nov. 2001, pp. 1317–1321.

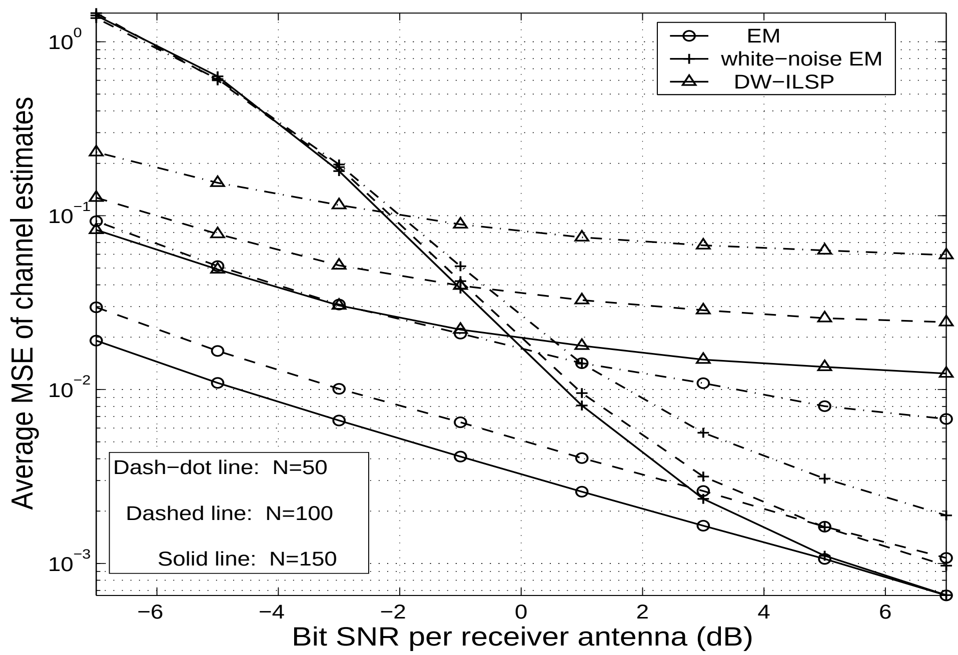
List of Figures

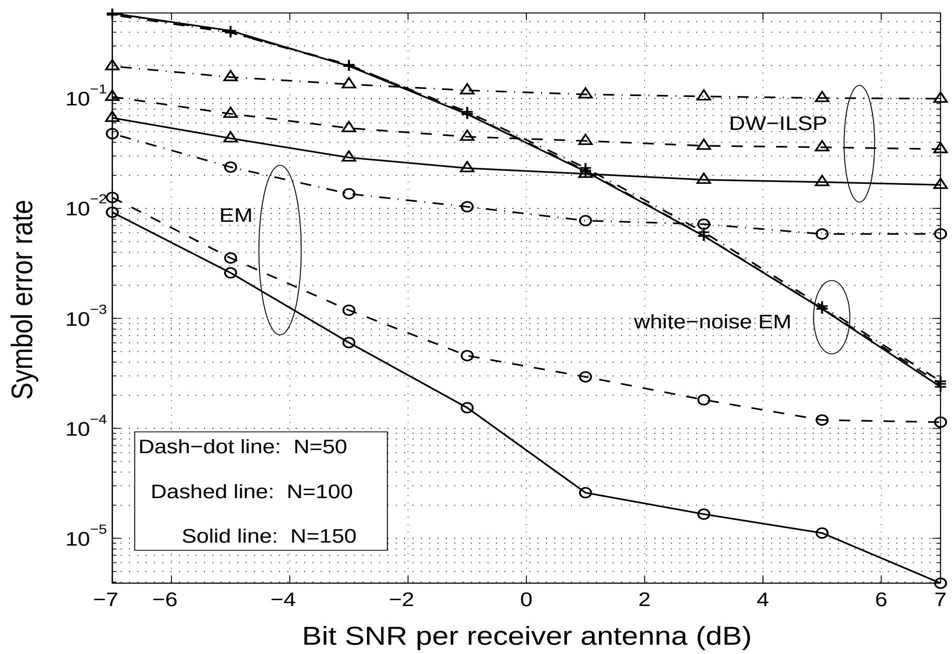


Average MSE of channel estimates
Number of snapshots N
10^{-1}
10^{-2}
10^{-3}
10^{2}
10^{3}
EM
white−noise EM
DW−ILSP
CRB



Average MSE of ML noise estimates
Number of snapshots N
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{2}
10^{3}
$\Sigma_{1,1}$
$\mathrm{Re}\{\Sigma_{2,1}\}$
$\mathrm{Im}\{\Sigma_{2,1}\}$
$\mathrm{Re}\{\Sigma_{3,1}\}$
$\mathrm{Im}\{\Sigma_{3,1}\}$
Solid line: MSE
Dashed line: CRB


EM
white−noise EM
DW−ILSP
Dash−dot line: N=50
Dashed line: N=100
Solid line: N=150
Average MSE of channel estimates
Bit SNR per receiver antenna (dB)
10^{0}
10^{-1}
10^{-2}
10^{-3}
−6
−4
−2
0
2
4
6



Symbol error rate
Bit SNR per receiver antenna (dB)
DW−ILSP
EM
white−noise EM
Dash−dot line: N=50
Dashed line: N=100
Solid line: N=150