Maximum Likelihood Semi-blind Channel and Noise Estimation using the EM Algorithm

Aleksandar Dogandzić, Wei Mo, and Zhengdao Wang
Department of Electrical and Computer Engr.
Iowa State University, 2215 Coover Hall
Ames, Iowa 50011
e-mail: {ald,mowei,zhengdao}@iastate.edu

Abstract — We present a maximum likelihood (ML) method for semi-blind estimation of single-input multi-output (SIMO) fading channels in spatially correlated noise with unknown covariance, or for estimation of multipath channels in unknown colored noise. An expectation-maximization (EM) algorithm is utilized to compute the ML estimates of the channel and spatial noise covariance. We derive modified and estimated Cramér-Rao bounds for the unknown parameters, and present a symbol detector that utilizes the EM channel estimates. Numerical simulations demonstrate the performance of the proposed methods.

I. INTRODUCTION

Expectation-maximization (EM) and related algorithms (see e.g. [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 in smart antenna systems [10]–[12]. Here, we propose an EM algorithm for semi-blind ML estimation of both the channel and spatial noise covariance matrices in a smart antenna scenario. The proposed algorithm can also be used to estimate multipath channels in unknown colored noise. This is unlike previous work in [7]–[9], where EM algorithms were applied to SISO 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 modified and estimated Cramér-Rao bounds (CRBs) 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 \( y(t) \) an \( m \times 1 \) data vector (snapshot) received by an array of \( m \) antennas at time \( t \) and assume that we have collected \( N \) snapshots. Under a single-user slow flat fading scenario, \( y(t) \) can be modeled as

\[
y(t) = h \cdot u(t) + e(t), \quad t = 1, 2, \ldots, N,
\]

where

- \( h \) is an unknown \( m \times 1 \) channel response vector,
- \( u(t) \) is an unknown symbol received by the array at time \( t \), and
- \( e(t) \) is temporally white and circularly symmetric zero-mean complex Gaussian noise vector with unknown positive definite spatial covariance matrix \( \Sigma \).

The noise term \( e(t) \) models co-channel interference (CCI) and receiver noise. We further assume that the symbols \( u(t) \) belong to an \( M \)-ary constant-modulus constellation \( \{ u_1, u_2, \ldots, u_M \} \), where \( |u_m| = 1 \), \( m = 1, 2, \ldots, M \). The constant-modulus assumption can be relaxed by appropriately modifying the algorithm. We model \( u(t) \), \( t = 1, 2, \ldots, N \) as independent, identically distributed (i.i.d.) random variables with probability mass function

\[
p(u(t)) = \frac{1}{M} i(u(t)),
\]

where

\[
i(u) = \begin{cases} 1, & u \in \{ u_1, u_2, \ldots, u_M \}, \\ 0, & \text{otherwise} \end{cases}
\]

Our goal is to estimate the unknown channel and noise parameters \( h \) and \( \Sigma \). To allow unique estimation of the channel \( h \), assume that \( N_x \) known (training) symbols

\[
u_T(t), \quad t = 1, 2, \ldots, N_T
\]

are embedded in the transmission scheme and denote the corresponding snapshots received by the array as \( y_T(\tau) \), \( \tau = 1, 2, \ldots, N_T \). Then, the measurement model \((1)\) holds for the training symbols as well, with \( y(t) \) and \( u(t) \) replaced by \( y_T(\tau) \) and \( u_T(\tau) \), respectively.

In the following section, we derive an EM algorithm for computing the ML estimates of \( h \) and \( \Sigma \) under the above measurement model.

III. ML Estimation
We treat the unknown symbols \( u(t), t = 1, 2, \ldots, N \) as the unobserved (or missing) data. Given \( u(t) \), the corresponding observed data \( y(t) \) is distributed as a complex multivariate Gaussian vector with probability density function (pdf):

\[
f(y(t)|u(t), h, \Sigma) = \frac{1}{\sqrt{\pi \Sigma}} \cdot \exp \left\{ -[y(t) - hu(t)]^H \Sigma^{-1} [y(t) - hu(t)] \right\}, \tag{5}
\]

where \(^H\) denotes the Hermitian (conjugate) transpose. The above expression also holds for the training data, with \( y(t) \) and \( u(t) \) replaced by \( y_T(\tau) \) and \( u_T(\tau) \). The joint distribution of \( y(t), u(t) \) (for \( t = 1, 2, \ldots, N \)), and \( y_T(\tau) \) (for \( \tau = 1, 2, \ldots, N_T \)) can be written as

\[
\prod_{t=1}^{N} p(u(t)) f(y(t)|u(t), h, \Sigma) \cdot \prod_{\tau=1}^{N_T} f(y_T(\tau)|u_T(\tau), h, \Sigma), \tag{6}
\]

which is also known as the complete-data likelihood function. Then, complete-data sufficient statistics for estimating \( h \) and \( \Sigma \) are

\[
r_{yy} = \frac{1}{N + N_T} \left[ \sum_{t=1}^{N} y(t)u(t)^* + \sum_{\tau=1}^{N_T} y_T(\tau)u_T(\tau)^* \right], \tag{7a}
\]

\[
R_{yy} = \frac{1}{N + N_T} \left[ \sum_{t=1}^{N} y(t)y(t)^H + \sum_{\tau=1}^{N_T} y_T(\tau)y_T(\tau)^H \right], \tag{7b}
\]

where \(^*\) denotes complex conjugation. Note that the complete-data likelihood belongs to an exponential family of distributions, and the EM algorithm is easily derived as follows:

- The expectation (E) step reduces to computing conditional expectations of the complete-data sufficient statistics [in (7)], given the observed data \( y(t), t = 1, \ldots, N \) and \( y_T(\tau), \tau = 1, \ldots, N_T \). [Note that \( R_{yy} \) is constant with respect to this conditional expectation.]

- The maximization (M) step reduces to finding the expressions for the complete-data ML estimates of the unknown parameters \( h \) and \( \Sigma \), and replacing the complete-data sufficient statistics (7) that occur in these expressions with their conditional expectations computed in the E step. Here, the complete-data ML estimates of \( h \) and \( \Sigma \) easily follow as a special case of the multivariate analysis of variance (MANOVA) model, see [13] and [14, Sec. V.A].

Using the above procedure, we derive the EM algorithm shown below:

\[
E \text{ step:} \quad h^{(k+1)} = \frac{1}{N + N_T} \left[ \sum_{t=1}^{N} y(t)u_m^* \exp\left\{ -[y(t) - h^{(k)} u_m]^H (\Sigma^{(k)})^{-1} [y(t) - h^{(k)} u_m] \right\} \right] / \sum_{m=1}^{M} \exp\left\{ -[y(t) - h^{(k)} u_m]^H (\Sigma^{(k)})^{-1} [y(t) - h^{(k)} u_m] \right\}.
\]

\[
M \text{ step:} \quad \Sigma^{(k+1)} = R_{yy} - h^{(k+1)}(h^{(k+1)})^H.
\]

Note that, in the \((k + 1)\)st iteration, the E step requires the computation of \( (\Sigma^{(k+1)})^{-1} \), which can be done using the matrix inversion lemma in e.g. [15, Cor. 18.2.10]:

\[
(\Sigma^{(k+1)})^{-1} = R_{yy}^{-1} + \frac{R_{yy}^{-1} h^{(k+1)}(h^{(k+1)})^H R_{yy}^{-1}}{1 - (h^{(k+1)})^H R_{yy}^{-1} h^{(k+1)}}, \tag{10}
\]

where \( R_{yy}^{-1} \) needs to be evaluated only once, before the iteration starts. Observe also that \( (\Sigma^{(k+1)})^{-1} h^{(k+1)} \) can be simplified as follows:

\[
(\Sigma^{(k+1)})^{-1} h^{(k+1)} = \frac{R_{yy}^{-1} h^{(k+1)} (1 - (h^{(k+1)})^H R_{yy}^{-1} h^{(k+1)})}{1 - (h^{(k+1)})^H R_{yy}^{-1} h^{(k+1)}}. \tag{11}
\]

We summarize the EM algorithm as the following steps.

S1) Compute the sufficient statistics according to (7).

S2) Initialize \( h \) and \( \Sigma \) to be the all-one vector and the identity matrix, respectively.

S3) Iterative between the E-step (8) and the M-step (9), until \( h \) and \( \Sigma \) converge. Use (10) and (11) in (8).

S4) Do symbol detection based on the channel estimation (see Section V).

IV. CRAMÉR-RAO BOUNDS

The exact CRB for the unknown parameters under the measurement model in Section II is difficult to compute. Here, we derive

- the modified CRB (MCRB) [16], which is a lower bound on the exact CRB, and

- an estimate of the exact CRB, obtained by inverting the empirical observed information matrix, see [3, ch. 4.3].

First, define the vector of unknown channel and noise parameters \( \rho = [\eta^T, \psi^T]^T \) where \( \eta = [\text{Re}(h)^T, \text{Im}(h)^T]^T \) and \( \psi = [\text{Re}(\text{vech}(\Sigma))^T, \text{Im}(\text{vech}(\Sigma))^T]^T \). (The \text{vech} and \text{vech} operators create a single column vector by stacking elements below the main diagonal columnwise: \text{vech} includes the main diagonal, whereas \text{vech} omits it.)

**Modified Cramér-Rao Bound:** Assuming the measurement model in Section II, the MCRB for the unknown parameters \( \rho \) is identical to the exact CRB for these parameters when the symbols \( u(t) \) are known, and is equal to:

\[
\text{MCRB}_\rho = \begin{bmatrix} \text{MCRB}_\eta & 0 \\ 0 & \text{MCRB}_\psi \end{bmatrix}, \tag{12}
\]
where

$$\text{MCRB}_s = \frac{1}{2(N + N_\tau)} \begin{bmatrix} \text{Re}\{\Sigma\} & \text{Im}\{\Sigma\} \\ \text{Im}\{\Sigma\} & \text{Re}\{\Sigma\} \end{bmatrix}, \quad (13a)$$

$$\text{MCRB}_e = \frac{1}{N + N_\tau} \mathcal{I}_e^{-1} \quad (13b)$$

and the \((i,k)\)th element of \(\mathcal{I}_e\) is

$$[\mathcal{I}_e]_{i,k} = \text{tr} \left\{ \frac{\partial \Sigma}{\partial \psi_i} \frac{\partial \Sigma}{\partial \psi_k} \right\}, \quad (14)$$

for \(i, k = 1, 2, \ldots, m^2\). Denote by \(\Sigma_{p,q}\) the \((p, q)\) element of \(\Sigma\), for \(p, q = 1, 2, \ldots, m\). Using this notation, we further simplify (14): for \(p_1 > q_1\) and \(p_2 > q_2\), we have

$$[\mathcal{I}_e]\text{Re}\{\Sigma_{p_1,q_1}\}, \text{Re}\{\Sigma_{p_2,q_2}\} = [\mathcal{I}_e]\text{Re}\{\Sigma_{p_2,q_2}\}, \text{Re}\{\Sigma_{p_1,q_1}\}$$

$$= 2 \text{Re} \left\{ [\Sigma^{-1}]_{p_1,q_1} \cdot [\Sigma^{-1}]_{q_1,p_2} + [\Sigma^{-1}]_{q_2,p_1} \cdot [\Sigma^{-1}]_{p_1,q_2} \right\}$$

$$[\mathcal{I}_e]\text{Re}\{\Sigma_{p_1,q_1}\}, \text{Im}\{\Sigma_{p_2,q_2}\} = [\mathcal{I}_e]\text{Im}\{\Sigma_{p_2,q_2}\}, \text{Re}\{\Sigma_{p_1,q_1}\}$$

$$= -2 \text{Im} \left\{ [\Sigma^{-1}]_{p_1,q_1} \cdot [\Sigma^{-1}]_{q_1,p_2} + [\Sigma^{-1}]_{q_2,p_1} \cdot [\Sigma^{-1}]_{p_1,q_2} \right\}$$

$$[\mathcal{I}_e]\text{Im}\{\Sigma_{p_1,q_1}\}, \text{Im}\{\Sigma_{p_2,q_2}\} = [\mathcal{I}_e]\text{Im}\{\Sigma_{p_2,q_2}\}, \text{Im}\{\Sigma_{p_1,q_1}\}$$

$$= 2 \text{Re} \left\{ -[\Sigma^{-1}]_{q_2,p_1} \cdot [\Sigma^{-1}]_{q_1,p_2} + [\Sigma^{-1}]_{q_2,p_1} \cdot [\Sigma^{-1}]_{q_1,p_2} \right\}$$

for \(p_1 = q_1\) and \(p_2 > q_2\),

$$[\mathcal{I}_e]\text{Re}\{\Sigma_{p_1,p_1}\}, \text{Re}\{\Sigma_{p_2,q_2}\} = [\mathcal{I}_e]\text{Re}\{\Sigma_{p_2,q_2}\}, \text{Re}\{\Sigma_{p_1,p_1}\}$$

$$= 2 \text{Re} \left\{ [\Sigma^{-1}]_{q_2,p_1} \cdot [\Sigma^{-1}]_{p_1,q_2} \right\} \quad (16a)$$

$$[\mathcal{I}_e]\text{Re}\{\Sigma_{p_1,q_1}\}, \text{Im}\{\Sigma_{p_2,q_2}\} = [\mathcal{I}_e]\text{Im}\{\Sigma_{p_2,q_2}\}, \text{Re}\{\Sigma_{p_1,q_1}\}$$

$$= -2 \text{Im} \left\{ [\Sigma^{-1}]_{q_2,p_1} \cdot [\Sigma^{-1}]_{p_1,q_2} \right\} \quad (16b)$$

and, for \(p_1 = q_1\) and \(p_2 = q_2\),

$$[\mathcal{I}_e]\text{Re}\{\Sigma_{p_1,p_1}\}, \text{Re}\{\Sigma_{p_2,p_2}\} = [\Sigma^{-1}]_{p_1,p_2}^2 \quad (17)$$

**Estimated Cramér-Rao Bound:** An estimate of the exact CRB for the channel parameters \(\eta\) is given by the observed data \(y(t), t = 1, \ldots, N\) and \(y_\tau(\tau), \tau = 1, \ldots, N_\tau\) as follows:

$$\widehat{\text{CRB}}_\eta = \sum_{t=1}^{N} \left[ \text{Re}\{s(y(t), \hat{\rho})\}^T \text{Im}\{s(y(t), \hat{\rho})\} \right]^T$$

$$+ \sum_{\tau=1}^{N_\tau} \left[ \text{Re}\{s_\tau(y_\tau(\tau), \hat{\rho})\}^T \text{Im}\{s_\tau(y_\tau(\tau), \hat{\rho})\} \right]^T$$

$$\cdot \left[ \text{Re}\{s_\tau(y_\tau(\tau), \hat{\rho})\}^T \text{Im}\{s_\tau(y_\tau(\tau), \hat{\rho})\} \right]^{-1} \quad (18)$$

where \(s(y(t), \hat{\rho})\) and \(s_\tau(y_\tau(\tau), \hat{\rho})\) are defined in (19) and \(\hat{\rho} = \rho^{(\infty)}\) is the ML estimate of \(\rho\) (obtained using the EM algorithm). Here, \(\text{CRB}_\eta\) follows by inverting the empirical observed information matrix, computed using the results of [3, ch. 4.3]. A similar estimated CRB expression can be computed for the noise parameters \(\psi\).

**V. Detection**

We now utilize the channel and noise estimators proposed in Section III to detect the unknown transmitted symbols

$$s(y(t), \hat{\rho}) = 2 \cdot \hat{\Sigma}^{-1} \cdot \left[ \begin{bmatrix} y(t) - \hat{h}_m u_m \end{bmatrix} \right] \quad (21a)$$

$$s_\tau(y_\tau(\tau), \hat{\rho}) = 2 \cdot \hat{\Sigma}^{-1} \cdot \left[ y_\tau(\tau)u_\tau(\tau)^\tau - \hat{h} \right]$$

$$u(t). \text{ We use the maximum a posteriori (MAP) detector:}$$

$$\hat{u}(t) = \arg \max_{u(t) \in \{u_1, u_2, \ldots, u_M\}} \exp\left\{ -[y(t) - \hat{h} u(t)]^T \hat{\Sigma}^{-1}[y(t) - \hat{h} u(t)] \right\} \quad (21a)$$

$$= \arg \min_{u(t) \in \{u_1, u_2, \ldots, u_M\}} \sum_{m=1}^{M} \exp\left\{ -[y(t) - \hat{h}_m u_m]^T \hat{\Sigma}^{-1}[y(t) - \hat{h}_m u_m] \right\}$$

$$= \arg \max_{u(t) \in \{u_1, u_2, \ldots, u_M\}} \text{Re}\{y(t)^T \hat{R}^{-1}_y \hat{h} \cdot u(t)\} \quad (21c)$$

where \(\hat{h}\) and \(\hat{\Sigma}\) are the ML estimates using the EM iteration (8)–(9). To derive (21c), we have used the identities (10) and (11), as well as the constant-modulus property of the transmitted symbols. Interestingly, the detector in (21c) is a function of the channel estimate \(\hat{h}\) only, through the \(\hat{R}^{-1}_y\) term. The expression (21a), maximized with respect to \(u(t)\), is a good indicator of confidence in the decision at time \(t\). Note that the denominator of (21a) is available as a by-product of the EM algorithm, see (8). The detector (21) can be easily modified to account for unequal prior probabilities of the transmitted symbols.

**VI. Simulation Results**

We evaluate the performance of the proposed semi-blind channel and noise estimation using numerical simulations. Our performance metrics are the mean-square error (MSE) and symbol error rate (SER), calculated using 3200 independent trials. We consider an array of \(m = 5\) receiver antennas. 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 \((N_\tau = 3)\), which was used to obtain initial estimates of the channel coefficients (computed using least squares). The signal was corrupted by additive complex Gaussian noise with spatial noise covariance \(\Sigma\) shown in (20) at the bottom of the page. The signal-to-noise ratio (SNR) per receiver antenna was defined as

$$\text{SNR} = 10 \log(1/\sigma^2) \quad (\text{dB}).$$

For the first set of simulations, the SNR per receiver antenna was chosen to be 0 dB. In Figures 1 and 2, we
show the MSEs (and the corresponding MCRBs and estimated CRBs) for the ML estimates of the channel coefficients and selected elements of the spatial noise covariance matrix $\Sigma$, as a function of block length $N$. From equations (13a) and (20), it follows that the MCRBs for the five different channel branches are identical, and are hence represented by only one curve in Figure 1. We also show the estimated CRBs computed by averaging (18) over independent trials (denoted by ECRB in Figure 1). In Figure 2, we show the MSEs for the ML estimates of Re{$\Sigma_{2,1}$}, Im{$\Sigma_{2,1}$}, and $\Sigma_{1,1}$, as well as the corresponding MCRBs.

In Figures 3 and 4, we compare the MSE performances of the EM algorithm and the decoupled weighted iterative least squares with projection (DW-ILSP) method in [17] and [14]. Figure 3 shows the MSEs of the estimated channel coefficients for the EM and DW-ILSP algorithms as functions of the block length $N$. For low SNR (0 dB), few training symbols ($N_T = 3$) and short block length ($N < 200$), the EM algorithm clearly outperforms the DW-ILSP algorithm. In Figure 4, MSEs of the two algorithms are shown as functions of SNR per receiver antenna. When the EM algorithm is used, the estimated MSE of one component of $h$ is 0.04, the EM algorithm has about 4 dB advantage over the DW-ILSP algorithm. 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). Note also that the number of parameters in the random-symbol measurement model in Section II equals $m^2 + 2m$, and is therefore independent of the block size $N$. This is in contrast with DW-ILSP and other deterministic ML methods (e.g. [18]) where the number of parameters grows with $N$.

In Figure 5, we compare symbol error rates of the detector (21) which uses the EM estimates of $h$ and $\Sigma$, and the DW-ILSP detector in [17]. The symbol error rates are shown as functions of the SNR per receiver antenna. When the symbol error rate is $10^{-1}$, the proposed EM algorithm has about 3 dB advantage over the DW-ILSP algorithm.

VII. Concluding Remarks

We derived an expectation-maximization algorithm for semi-blind estimation of single-input multi-output fading channels in spatially correlated noise having unknown covariance. We computed the modified and estimated 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.

The proposed estimation algorithm can also be applied to estimation of multi-input multi-output (MIMO) channels. However, the complexity will increase due to the needed averaging over the information symbols. Further
Figure 2: Mean-square errors for the ML estimates of \( \text{Re}\{\Sigma_{2,1}\}, \text{Im}\{\Sigma_{2,1}\}, \Sigma_{1,1} \) and corresponding MCRBs, as functions of \( N \), for \( N_T = 3 \) and SNR = 0 dB.

Figure 3: Mean-square errors for the estimates of \( h_1, h_2, h_3, h_4 \), and \( h_5 \) obtained using the EM and DW-ILSP algorithms, as functions of \( N \), for \( N_T = 3 \) and SNR = 0 dB.

Figure 4: Mean-square errors for the estimates of \( h_1, h_2, h_3, h_4 \), and \( h_5 \) using the EM and DW-ILSP algorithms, as functions of the SNR per receiver antenna, for \( N = 100 \) and \( N_T = 3 \).

Figure 5: Symbol error rates of the EM-based and DW-ILSP detectors as functions of the SNR per receiver antenna, for \( N = 100 \) and \( N_T = 3 \).
work will include extending the algorithm to the MIMO scenario and reduce the complexity involved. When there is error-control coding in the transmission, we will also develop iterative schemes to combine the proposed channel estimation with decoding.

REFERENCES


