An Iterative Receiver for Distributed Multi-Input Multi-Output (MIMO) Flat-Fading Channels

The-Hanh Pham  
National University of Singapore,  
Singapore 119260  
Email: elepth@nus.edu.sg

Arumugam Nallanathan  
Division of Engineering,  
King’s College of London  
Email: arumugam.nallanathan@kcl.ac.uk

Ying-Chang Liang  
Institute for Infocomm Research (I2R),  
1 Fusionopolis Way, #21-01 Connexis  
Singapore 138632  
Email: ycliang@i2r.a-star.edu.sg

Abstract—In this paper, we consider the problem of joint channel and frequency offset estimation, data detection in a flat-fading multi-input multi-output (MIMO) system. We assume that each pair of transmit and receive antennas has a different frequency offset value. We present a computationally efficient iterative algorithm based on the space-alternating generalized expectation-maximization (SAGE) and the expectation conditional maximization (ECM) algorithms.

I. INTRODUCTION

Multi-antenna transmission over multi-input multi-output (MIMO) channels has been proved to be effective in combating multipath fading, as well as increasing the channel capacity [1, 2]. In conventional MIMO systems, the transmit/receive antennas are collocated; hence, it is usually assumed that there is only one frequency offset within the system [3, 4]. Recently, there is an increasing interest in the research of distributed MIMO systems [5, 6] where each of the transmit antennas is utilized by one user and the receive antennas are distributed in various locations in order to compensate the large scale fading (shadowing) effect. One typical distributed MIMO scenario is the cellular systems where several cell edge users communicate with several base stations. In this case, each transmit/receive antenna is equipped with its own oscillator, thus different transmit-receive pair may have different frequency offset.

In practice, coherent detection requires information of channel and frequency offset, thus channel and frequency offset estimation has become a critical task. The transmitted signals from the transmitter consists of two parts. One is the known training signals and the other is the data signals. This results in two approaches to obtain the necessary information for detection process at the receiver. One approach is only use the received signals over the transmission of the known training signals. This approach for distributed MIMO systems is investigated in [7–9] for channel and frequency offset estimation and [10] for frequency offset estimation only. In spite of their simplicity, the accuracy of those methods may not be guaranteed in case of insufficient number of pilots and/or the very high noise environment. The other approach is to exploit not only the information provided by the training signals but also the detected signals to improve the quality of the interested parameters [11, 12].

To the best of authors’ knowledge, this approach has not been investigated in distributed MIMO systems.

In this paper, we propose an iterative receiver to estimate the channel coefficients and frequency offsets as well as to detect the transmitted signals in a flat-fading MIMO channel.

This approach is based on the Space-Alternating Generalized Expectation Maximization (SAGE) algorithm [13] and the Expectation Conditional Maximization (ECM) algorithm [14]. We assume that each pair of transmit-receive antenna has a distinct frequency offset value. Our proposed algorithm decouples the multi-dimensional optimization problem into many one-dimensional optimization problems. Simulations show that the proposed method improve the bit-error-rate (BER) performance compared with the one use only information given by the training symbols.

The organization of the rest of this paper is as follows. Section II presents the system model. The proposed iterative receiver is given in Section III. Section IV provides the simulation results. Finally, conclusions and necessary derivations are given the Section V and the Appendix, respectively.

Notations: The capital bold letters denote matrices and the small bold letters denote row/column vectors; transpose, conjugate and Hermitian conjugate of a vector/matrix are denoted by \( \cdot^T \), \( \cdot^* \) and \( \cdot^H \), respectively; \( \Re\{\cdot\} \), \( \Im\{\cdot\} \), \( |\cdot| \) and \( \cdot^\dagger \) denote the real part, imaginary part, absolute value and conjugate of a complex number \( \cdot \), respectively; \( \odot \) denotes the element-wise product of two vectors/matrices. \( I_N \) denotes the identity matrix of size \( N \). \( \{\cdot\} \) is a diagonal matrix constructed from vector \( \cdot \). \( \text{sign}\{\cdot\} \) is the sign function.

II. SYSTEM MODEL

We consider a distributed MIMO system with \( N_t \) transmit and \( N_r \) receive antennas working under a flat-fading environment. The channel coefficient and the frequency offset from the \( t \)-th transmit antenna to the \( k \)-th receive antenna are denoted by \( h_{k,t} \) and \( w_{k,t} \), respectively. We assume that \( h_{k,t} \)’s and \( w_{k,t} \)’s are unchanged over a transmission of \( T \) symbols. Furthermore, we assume that the \( h_{k,t} \)’s are statistically independent and complex-value Gaussian random variables with zero-mean. If the transmitted symbol at time \( t \) from the \( l \)-th transmit antenna is denoted by \( s_l(t) \), the received signal at the \( k \)-th receive antenna, \( k = 1, 2, \ldots, N_r \), at that time is written as

\[
y_k(t) = \sum_{l=1}^{N_t} h_{k,l} e^{jw_{k,l} t} s_l(t) + n_k(t), \quad t = 1, 2, \ldots, T, \tag{1}
\]

where noise sequence at the \( k \)-th receive antenna, \( \{n_k(t)\}_{t=1}^T \), is a sequence of zero-mean i.i.d. complex value Gaussian random variables with variance of \( \sigma_n^2 \). We further assume that the noise at \( N_r \) receive antennas are statistically independent. The transmitted symbol at time \( t \) from the \( l \)-th transmit antenna belongs.
to an M-ary constant modulus constellation with \( |s_l(t)|^2 = \sigma_n^2 \), \( t = 1, 2, \ldots, T \).

We define \( h_t = [h_{1,t} \ h_{2,t} \ \cdots \ h_{N_r,t}]^T \) and \( w_t = [w_{1,t} \ w_{2,t} \ \cdots \ w_{N_r,t}]^T \) as the vector containing the channel coefficients and vector containing frequency offsets from the \( l^{th} \) transmit antenna to all \( N_r \) receive antennas, respectively. The signals from the \( l^{th} \) transmit antenna is denoted as \( s_l = [s_l(1) \ s_l(2) \ \cdots \ s_l(T)]^T \). If we collect all receive signals from \( N_r \) receive antennas at time \( t \) to form \( y(t) = [y_1(t) \ y_2(t) \ \cdots \ y_{N_r}(t)]^T \), this vector can be written as

\[
y(t) = \sum_{l=1}^{N_r} (h_t \circ e_t(t)) s_l(t) + n(t), \quad t = 1, 2, \ldots, T,
\]

where \( n(t) = [n_1(t) \ \cdots \ n_{N_r}(t)]^T \), and \( e_t(t) = [e^{jw_{1,t}} \ e^{jw_{2,t}} \ \cdots \ e^{jw_{N_r,t}}]^T \).

Our objective here is to use \( \{y(t)\}_{T=1}^T \) to estimate \( \theta = [\theta_1^T \ \cdots \ \theta_l^T \ \cdots \ \theta_{N_r}^T] \) where \( \theta_l = [w_{1,t} \ h_{1,t} \ s_l^T]^T \) is the parameters of the \( l^{th} \) transmit antenna.

III. PROPOSED ITERATIVE RECEIVER

The Maximum-likelihood (ML) estimate of \( \theta \) is obtained by the following maximization equation

\[
\hat{\theta}_{ML} = \arg \max_{\theta} \Lambda(\theta),
\]

where \( \Lambda(\theta) \) is the log-likelihood function (after dropping terms that are not related to \( \theta \)). It is given by

\[
\Lambda(\theta) = \sum_{t=1}^{T} \left( \|y(t) - \sum_{l=1}^{N_r} (h_t \circ e_t(t)) s_l(t)\|^2 \right).
\]

Unfortunately, this operation requires an exhaustive search over the multi-dimensional space spanned by \( \theta \) which is prohibitively complex for practical implementation.

To avoid this obstacle, we propose and iterative receiver which is based on the SAGE and ECM algorithms. More explicitly, SAGE-based algorithm decouples the complex maximization in (3) into \( N_t \) smaller maximization problems. Each of these smaller problems is used to estimate the \( \theta_l, l = 1, 2, \ldots, N_t \). For the parameter \( \theta_l \), we determine the contribution from the \( l^{th} \) transmit antenna in the \( \{y(t)\}_{T=1}^T \) as

\[
y_l(t) = (h_t \circ e_t(t)) s_l(t) + n_l(t), \quad t = 1, 2, \ldots, T,
\]

where \( n_l(t), t = 1, 2, \ldots, T, \) is a zero-mean complex Gaussian random vector and \( \sum_{l=1}^{N_t} n_l(t) = n(t) \) for \( t = 1, 2, \ldots, T \). However, the exact value of \( \{y_l(t)\}_{T=1}^T \) is hard to calculate exactly. Its value is estimated along side with our algorithm's iteration.

The iteration, our algorithm is composed of \textit{iterations}; one iteration is the process of estimating all \( \theta_l \)'s. Furthermore, one iteration consists of \( N_t \) \textit{sub-iterations}; the \( l^{th} \) sub-iteration updates the parameter \( \theta_l \) of the \( l^{th} \) transmit antenna.

Suppose that after the \( (m - 1)^{th} \) iteration, we have the estimates \( \hat{\theta}_l^{[m]} = [\hat{w}_{1,l}^{[m]} \ \hat{h}_{1,l}^{[m]} \ \hat{s}_l^{[m]}]^T \), for \( l = 1, 2, \ldots, N_t \). At the \( l^{th} \) sub-iteration of the \( m^{th} \) iteration, we calculate the estimate of the contribution of the \( l^{th} \) transmit antenna as

\[
\hat{y}_l^{[m]}(t) = y(t) - \sum_{n=1 \atop n \neq l}^{N_r} \hat{x}_n^{[m]}(t), \quad t = 1, 2, \ldots, T,
\]

where

\[
\hat{x}_l^{[m]}(t) = (\hat{h}_l^{[m]} \circ \hat{e}_l^{[m]}(t)) \hat{s}_l^{[m]}(t), \quad t = 1, 2, \ldots, T.
\]

and

\[
\hat{e}_l^{[m]}(t) = [e^{j\hat{w}_{1,l}^{[m]}} e^{j\hat{w}_{2,l}^{[m]}} \ \cdots \ e^{j\hat{w}_{N_r,l}^{[m]}}]^T.
\]

Substituting (2) into (6) yields

\[
y_l^{[m]}(t) = (h_l \circ e_l(t)) s_l(t) + \hat{n}_l^{[m]}(t), \quad t = 1, 2, \ldots, T,
\]

where

\[
\hat{n}_l^{[m]}(t) = n(t) + \sum_{n=1 \atop n \neq l}^{N_r} (x_n(t) - \hat{x}_n^{[m]}(t)), \quad t = 1, 2, \ldots, T.
\]

and \( x_n(t) = (h_n \circ e_n(t)) s_n(t) \). We further assume that the estimates \( \hat{w}_{n,l}^{[m]} \) for \( n \neq l \) are sufficiently accurate so that we can approximate \( \hat{x}_n^{[m]}(t) \) by \( x_n(t) \). Hence, \( \hat{n}_l^{[m]}(t) \) in (9) reduces to \( n(t) \) for \( t = 1, 2, \ldots, T \).

The set \( \{\hat{y}_n^{[m]}(t)\}_{T=1}^T \) is the incomplete data space for the parameter \( \theta_l \). Of the elements of \( \theta_l \), we consider \( h_l \) as the missing data and we define \( \{\hat{y}_n^{[m]}(t)\}_{T=1}^T, h_l \) as the complete data space for the remaining variables \( \beta_l \triangleq [w_{1,l} \ s_l]^T \).

In order to update the parameters \( \theta_l \) at the \( l^{th} \) sub-iteration, we have two step as follows.

A. E-step

In this step, we determine

\[
Q(\beta_l | \hat{\beta}_l^{[m]}) = E \left[ \log f(\{\hat{y}_n^{[m]}(t)\}_{T=1}^T | h_l, w_l, s_l) \right] \{\hat{y}_n^{[m]}(t)\}_{T=1}^T, \hat{\beta}_l^{[m]} \},
\]

where \( \hat{\beta}_l^{[m]} = [\hat{w}_{n,l}^{[m]} \ \hat{s}_l^{[m]}]^T \). We have

\[
f(\{\hat{y}_n^{[m]}(t)\}_{T=1}^T | h_l, w_l, s_l) =
\]

\[
f(\{\hat{y}_n^{[m]}(t)\}_{T=1}^T | h_l, w_l, s_l) f(h_l | w_l, s_l)
\]

where

\[
f(\{\hat{y}_n^{[m]}(t)\}_{T=1}^T | h_l, w_l, s_l) = \frac{1}{(\pi \sigma_n^2)^T \pi} \exp \left\{ - \frac{1}{\sigma_n^2} \sum_{t=1}^{T} \|y_n^{[m]}(t) - (h_l \circ e_l(t)) s_l(t)\|^2 \right\}
\]

and

\[
f(h_l | w_l, s_l) = \frac{1}{\pi|K_{h_l}|} \exp \left\{ -h_l^T K_{h_l}^{-1} h_l \right\},
\]

due to the independence of \( h_l \) with \( w_l \) and \( s_l, K_{h_l} = E \{h_l h_l^T\} \).

After substituting (12) and (13) into (10) and dropping some terms that do not relate to \( \beta_l \), we have

\[
Q(\beta_l | \hat{\beta}_l^{[m]}) = -E \left[ \sum_{t=1}^{T} \|\hat{y}_n^{[m]}(t) - (h_l \circ e_l(t)) s_l(t)\|^2 \right] \{\hat{y}_n^{[m]}(t)\}_{T=1}^T, \hat{\beta}_l^{[m]} \}.
\]
The derivation of (14) is given in (A-7) of the Appendix with \( \beta_l \) replaced by \( \hat{\beta}_l^{[m]} \). From (A-7), after dropping some terms that do not relate to \( \beta_l \), we obtain

\[
Q(\beta_l | \hat{\beta}_l^{[m]}) = -\sum_{t=1}^{T}||\hat{y}_l^{[m]}(t) - (\hat{h}_l^{[m+1]} \odot e_l(t))s_l(t)||^2, \tag{15}
\]

where

\[
\hat{h}_l^{[m+1]} = \left[ \hat{h}_l^{[m+1]} \; \hat{h}_{l+1}^{[m+1]} \; \cdots \; \hat{h}_{N_r}^{[m+1]} \right]^T
= \frac{1}{\sigma_n^2} \left( K_{h_l} + \frac{T \sigma_n^2}{\sigma_n^2} I_{N_r} \right)^{-1}
\sum_{t=1}^{T} (\hat{z}_l^{[m]}(t))^*(\hat{e}_l^{[m]}(t))^* \odot \hat{y}_l^{[m]}(t). \tag{16}
\]

Equation (16) comes from (A-4) in the Appendix where \( \beta_l \) is replaced by \( \hat{\beta}_l^{[m]} \).

**B. M-step**

The M-step is to find the updated value of \( \beta_l, \hat{\beta}_{l}^{[m+1]} \), as follows

\[
\hat{\beta}_{l}^{[m+1]} = \arg \max_{\beta_l} Q(\beta_l | \hat{\beta}_{l}^{[m]}). \tag{17}
\]

In (17), the maximization is still a complicated problem where the updating process of parameters is taken place simultaneously for \( w_l \) and \( s_l \). Hence, we resort to the ECM algorithm to overcome this difficulty. The ECM algorithm minimizes (17) in two steps. In the first step, (17) is minimized with respect to one of \( (w_l, s_l) \) while the other is kept at their most updated values. We denote \( \hat{\beta}_{l}^{[m+c/2]} \) as the estimate of \( \beta_l \) at \( c \)-th step of the ECM algorithm in the \( t \)-th sub-iteration, \( c = 1, 2, \ldots \).

1) Step 1: In this step, we determine the updated value of \( w_l \) while \( s_l \) is fixed at \( \hat{\beta}_{l}^{[m]} \), i.e., we determine \( \hat{\beta}_{l}^{[m+1/2]} = [\hat{w}_l^{[m+1]} \; \hat{s}_l^{[m]}]^T \) where

\[
\hat{w}_l^{[m+1]} = \arg \max_{w_l} Q(\beta_l | \hat{\beta}_{l}^{[m]}), \tag{18}
\]

\[
= \arg \min_{w_l} \sum_{t=1}^{T} ||\hat{y}_l^{[m]}(t) - (\hat{h}_l^{[m+1]} \odot e_l(t))\hat{s}_l^{[m]}(t)||^2
= \arg \max_{w_l} \sum_{t=1}^{T} \sum_{k=1}^{N_r} T \Re \{(\hat{y}_l^{[m]}(t))^*(\hat{h}_l^{[m+1]} \odot e_l(t))(e^{j\omega_{k,l} t})\}, \tag{19}
\]

In (18), \( \hat{y}_l^{[m]}(t) \) is the \( k \)-th element of \( \hat{y}_l^{[m]}(t) \); \( \hat{h}_l^{[m+1]} \) is the \( k \)-th element of \( \hat{h}_l^{[m+1]} \). We can see from this equation that we can further decouple the updating process of whole vector \( w_l \) into \( N_r \) smaller updating processes. More explicitly, the updated value of \( w_{k,l} \) for \( k = 1, 2, \ldots, N_r \) is calculated as

\[
\hat{w}_{k,l}^{[m+1]} = \arg \min_{w_{k,l}} \sum_{t=1}^{T} T \Re \{(\hat{y}_l^{[m]}(t))^*(\hat{h}_l^{[m+1]} \odot e_l(t))(e^{j\omega_{k,l} t})\}. \tag{20}
\]

To overcome the nonlinearity of (19), we resort to the Taylor's series expansion of \( e^{j\omega_{k,l} t} \) around \( \hat{w}_{k,l}^{[m]} \) to the second-order term

\[
e^{j\omega_{k,l} t} \approx e^{j\hat{w}_{k,l}^{[m]} t} + (w_{k,l} - \hat{w}_{k,l}^{[m]} )(\hat{y}_l^{[m]}(t)) e^{j\omega_{k,l} t} + \frac{1}{2}(w_{k,l} - \hat{w}_{k,l}^{[m]} )^2 (\hat{y}_l^{[m]}(t))^2 e^{j\omega_{k,l} t}. \tag{20}
\]

Therefore, (19) can be written, after dropping some terms that do not relate to \( w_{k,l} \) as in (20) on the top of next page.

Differentiating the function inside \( \{ \cdot \} \) of (21) and equating the result to 0, we obtain the updated value \( \hat{w}_{k,l}^{[m+1]} \) as

\[
\hat{w}_{k,l}^{[m+1]} = \hat{w}_{k,l}^{[m]} - \frac{1}{2} \sum_{t=1}^{T} T \Re \{(\hat{y}_l^{[m]}(t))^*(\hat{h}_l^{[m+1]} \odot e_l^{[m+1]}(t))(e^{j\omega_{k,l} t})\} \tag{22}
\]

2) Step 2: In this step we calculate \( \hat{\beta}_{l}^{[m+1]} = [\hat{w}_l^{[m+1]} \; \hat{s}_l^{[m+1]}]^T \) where

\[
\hat{s}_l^{[m+1]} = \arg \max_{s_l} Q(\beta_l | \hat{\beta}_{l}^{[m]}), \tag{22}
\]

\[
= \arg \min_{s_l} \sum_{t=1}^{T} ||\hat{y}_l^{[m]}(t) - (\hat{h}_l^{[m+1]} \odot e_l^{[m+1]}(t))s_l(t)||^2
= \arg \max_{s_l} \sum_{t=1}^{T} \sum_{k=1}^{N_r} T \Re \{(\hat{y}_l^{[m]}(t))^*(\hat{h}_l^{[m+1]} \odot e_l^{[m+1]})(e^{j\omega_{k,l} t})s_l(t)\}. \tag{23}
\]

where \( e_l^{[m+1]}(t) = [e^{j\omega_{1,l} t} \; e^{j\omega_{2,l} t} \; \cdots \; e^{j\omega_{N_r,l} t}]^T \).

From (23), it is easy to observe that we can determine the updated value of each element of \( s_l \) separately. Specially, the updated value of element \( s_l(t), \; t = 1, 2, \ldots, T \), is determined as

\[
\hat{s}_l^{[m+1]}(t) = \arg \max_{s_l(t)} \sum_{k=1}^{N_r} T \Re \{(\hat{y}_l^{[m]}(t))^*(\hat{h}_l^{[m+1]} \odot e_l^{[m+1]})(e^{j\omega_{k,l} t})s_l(t)\}. \tag{24}
\]

If BPSK symbols are used at the transmit antennas, then (24) reduces to

\[
\hat{s}_l^{[m+1]}(t) = \begin{cases} \sum_{k=1}^{N_r} \Re \{(\hat{y}_l^{[m]}(t))^*(\hat{h}_l^{[m+1]} \odot e_l^{[m+1]})(e^{j\omega_{k,l} t})\} \end{cases}. \tag{25}
\]

After finishing the \( l \)-th sub-iteration where \( \theta_l \) is updated, the \( \chi_l^{[m]}(t), \; t = 1, 2, \ldots, T, \) is updated as

\[
\chi_l^{[m]}(t) = (\hat{h}_l^{[m+1]} \odot e_l^{[m+1]})(\hat{y}_l^{[m+1]}(t)), \; t = 1, 2, \ldots, T. \tag{26}
\]

The next sub-iteration is performed to estimate the next parameter \( \theta_{l+1} \).

**IV. Simulation Results**

In this section we present some simulation results for our proposed algorithm in a distributed MIMO system. The system has \( N_t = N_r = 2 \) antennas. The signals form all transmitted antennas are grouped into frames of length of \( F \). We assume that the channel is static over a frame and randomly changes from frame to frame and \( K_{h_l} = I_{N_r} \). In a single frame, the first \( L \)
symbols at the beginning are used for training purpose and the
T = F − L remaining symbols are used for data transmission.
We consider the case that L training symbols constitutes 10%
of the frame length. Over the training period, the method in[9]is used to estimate the interested parameters. These values are used as the initialization values for our algorithm. We use
BPSK for all simulations.
In Fig. 2, we present the BER performance versus the SNR
values for the case of L = 64 symbols and the frequency
offsets, in each frame, are realizations of uniformly distributed
random variables over the interval of 2π(−0.1, 0.1). The ini-
tialization curve is the performance obtained from the coherent
detection using the channel information from training symbols
only. We see that our algorithm provides better performance.
However, the improvement becomes smaller with the increase
of the iteration number. After 3 iteration, the gap between
the obtained performance and the initialization is around 4dB.
To help in comparison, in this figure, we also present the
ideal performance, i.e., the coherent detection using perfect
channel coefficients and frequency offsets. The gap between
the two performance and this ideal performance is around 5dB.

To see deeply into the BER performance versus the iteration
number, Fig. 3 provides, for different SNR values, the BER
performance as a function of iterations. It is observed that after 3
iterations, the improvement is marginal. Comparing among SNR
values, we see that BER drops faster for bigger SNR values.
Note that in this figure, the points belonging to 0 iteration are
from the initialization curve in Fig. 2.

V. CONCLUSIONS

In this paper, the problem of estimation of the channel co-
eficients and frequency offsets as well as data detection for MIMO
distributed systems are investigated. Unlike the conventional
MIMO systems, here, we assume the presence of multiple
frequency offsets in the systems. Our algorithm is based on the
SAGE and the ECM algorithms. The BER performance with
respect to some system parameters are obtained by simulations.
It is shown that our algorithm can provide better performance
compared with the approach that uses only the information
given by training period to detect the signal.

APPENDIX

In this appendix we prove (15) and (16). We repeat the model
of (8) as follows
\[ \hat{y}_i^{[m]}(t) = (h_i \odot e_i(t))s_i(t) + n(t), \quad t = 1, 2, \ldots, T, \quad (A-1) \]
where we have used the assumption that \( \hat{x}_n^{[m]}(t) \) is close enough
to \( x_n(t) \) for \( n \neq l \) and for \( t = 1, 2, \ldots, T \).

Equation (A-1) can be written in another form as
\[ \hat{y}_i^{[m]}(t) = \text{diag}\{s_i(t)e_i(t)\}h_i + n(t) = S_i(t)h_i + n(t), t = 1, 2, \ldots, T \quad (A-2) \]
where \( S_i(t) = \text{diag}\{s_i(t)e_i(t)\} \). Note that \( S_i^H(t)S_i(t) = \sigma_n^2I_n \), and \( \beta_i \) is embedded in the form of \( \{S_i(t)\}_{t=1}^{T} \).
If we stack all \( y_i^{[m]}(t) \), \( t = 1, 2, \ldots, T \) to construct the vector
\[ y_{[m]} = \left[ \left( y_i^{[m]}(1) \right)^T \left( y_i^{[m]}(2) \right)^T \cdots \left( y_i^{[m]}(T) \right)^T \right] \]

\[ \text{it can be written as} \]
\[ y_{[m]} = Sh_i + n, \quad (A-3) \]
where \( S = \left[ S_i^T(1) \ S_i^T(2) \cdots S_i^T(T) \right]^T \), and \( n = \left[ n^T(1) \ n^T(2) \cdots n^T(T) \right]^T \) is a zero-mean complex Gaussian
random vector with covariance of \( K_n = \sigma_n^2I_{Nn} \).
The conditional mean (and covariance) of \( h_i \) given \( y_{[m]} \) and \( S \)
are
\[ \hat{h}_i = E\{h_i|y_{[m]}, S\} = E\{h_i|y_{[m]}, \beta_i\} = (K_{h_i} + S^H\sigma_n^{-1}K_n^{-1}S)^{-1}S^H\sigma_n^{-1}y_{[m]} \]
\[ = \left( \frac{1}{\sigma_n^2} \left[ \sum_{t=1}^{T} (S_i(t))^H S_i(t) \right] \right)^{-1} \frac{1}{\sigma_n^2} \sum_{t=1}^{T} (S_i(t))^H y_i^{[m]}(t) \]
\[ = \frac{1}{\sigma_n^2} \left( K_{h_i} + \frac{T\sigma_n^2}{\sigma_n^2}I_{N_n} \right)^{-1} \sum_{t=1}^{T} (s_i^T(t)e_i^*(t)) y_i^{[m]}(t) \], \quad (A-4) \]
and
\[ K_{h_i} = E\{(h_i - \hat{h}_i)(h_i - \hat{h}_i)^H|y_{[m]}, S\} = (K_{h_i} + S^H\sigma_n^{-1}K_n^{-1}S)^{-1} \quad (A-5) \]

Equations (A-4) and (A-5) are well-known conditional expectation formulas in the Gaussian case using Matrix Inversion
Lemma [15].

Because of the linearity of (A-3) then \( \hat{y}_i^{[m]} - Sh_i \) given \( y_{[m]} \) and \( S \) is also conditionally Gaussian random vector with mean \( (y_{[m]} - Sh_i) \) and covariance matrix of \( S^H \sigma_n^2 \). We note that
\[ E\{(y_{[m]} - Sh_i)(y_{[m]} - Sh_i)^H|y_{[m]}, S\} = S^H \sigma_n^2 + (y_{[m]} - Sh_i)(y_{[m]} - Sh_i)^H \]. \quad (A-6) \]
We have
\[ -E\{\sum_{t=1}^{T}\|\hat{y}_t[m](t) - (h_t \odot e_i(t))s_i(t)\|_2^2\}_{t=1}, \beta_t} \]
\[ = -E\{\|\hat{y}_t[m] - S h_t H^T(y_t[m] - S h_t)\|_2^2\} \]
\[ = -E\{\text{tr}\{\hat{y}_t[m] - S h_t H^T(y_t[m] - S h_t)\} \}
\[ = -\text{tr}\{S h_t H^T\} - \text{tr}\{\text{tr}\{\hat{y}_t[m] - S h_t H^T(y_t[m] - S h_t)\}\} \]
\[ = -T \sigma^2 \text{tr}\{K_h\} - \sum_{t=1}^{T}\|\hat{y}_t[m](t) - (h_t \odot e_i(t))s_i(t)\|_2^2. \]  

(A-7)

REFERENCES


