BLIND ITERATIVE RECEIVER FOR MULTIUSER MIMO SYSTEMS

Ejaz Khan, Dirk Slock

Mobile Communications Department
Institut Eure\'ecom, B.P. 193, 06904 Sophia–Antipolis CEDEX, France
email: Ejaz.Khan@eurecom.fr

ABSTRACT
The information capacity of wireless communication systems may be increased dramatically by employing multiple transmit and receive antennas. In this chapter, we consider multiuser wireless communication system, employing multiple transmit and receive antennas. We estimate jointly channel and symbols user-wise by Maximum Likelihood approach (ML) approach. Two models are considered for the symbols of the interferers, corresponding to Gaussian and discrete priors. In the latter case, in which the finite alphabet gets exploited for the MAI symbols, a simplification for the posterior MAI symbol probabilities is introduced based on Mean Field Theory.

1. INTRODUCTION
Multiple Input Multiple Output (MIMO) system has gained much interest recently [6]. Deploying multiple antennas at both, the base station and the remote stations increase capacity of the wireless channels. The gain in capacity is because of diversity, spatial multiplexing, interference rejection and array gain. In order to fully exploit the advantages of an antenna array, one must know the channel that will distort the signal as well as interfering noise. In this paper, we consider the problem of estimation of channel-symbols user-wise (i.e., considering other users as interferers). We use two approaches for the parameter estimation. In the Gaussian prior case [4], only the Multiple Access Interference (MAI) are modeled as stationary (white) sequences. We use ML formulation that gets implemented via Expectation Maximization (EM) algorithm to alternate between channel and the User of Interest (UoI) symbols estimates. Alternatively, we consider exploiting the finite alphabet for the MAI symbols, leading to significant MAI reduction capability. To simplify and to reduce the complexity of the resulting EM algorithm, we consider the introduction of Mean Field methods for the approximation of the posterior MAI symbol probabilities. The paper is organized as follows: In section 2, we define the communication model. Section 3 is devoted to the general principle of the EM algorithm for Maximum Likelihood (ML) estimation. Section 4 describes user-wise channel-symbol estimation with Gaussian prior on MAI. In section 5, we describe user-wise channel-symbol estimation procedure using discrete prior on MAI symbols. Conclusions are drawn in section 6.

2. COMMUNICATION MODEL
We model a wireless communication system with K users. Each user is equipped with N transmit antennas. The base station has M receive antennas. We assume flat fading between each transmit-receive pair. We denote $\alpha_{m,n}$ as complex fading gain from the $n^{th}$ transmitter antenna to the $m^{th}$ receive antenna, where $\alpha_{m,n} \sim N_c(0, 1)$ is assumed to be zero mean circularly symmetric complex Gaussian random variable with unit variance. This is equivalent to the assumption that signals transmitted from different antennas undergo independent Rayleigh fades. It is also assumed that the fading gains remain constant over the entire signal frame, but they may vary from one frame to another. The received discrete time signal at instant $t$ can be written as

$$x_t = H d_t + n_t$$

(1)

Where $d_t = [d_{1,t}^T, d_{2,t}^T, \ldots, d_{K,t}^T]^T$, is the symbol vector. $x_t = [x_{t,1}, x_{t,2}, \ldots, x_{t,M}]^T$, is the received signal, $n_t = [n_{t,1}, n_{t,2}, \ldots, n_{t,M}]^T$ and $d_{it} = [d_{i,t,1}, d_{i,t,2}, \ldots, d_{i,t,M}]^T$, where $d_{it} \in \{-1, 1\}$, $n(t)$ is white Gaussian noise. $H$ is the transpose operator. Channel matrix $H$ is given by

$$H = [H_1 H_2 \ldots H_K]$$

(2)

where $H_i$ is as follows

$$H_i = \begin{pmatrix}
\alpha_{1,i} & \alpha_{2,i} & \ldots & \alpha_{N,i} \\
\alpha_{1,2} & \alpha_{2,2} & \ldots & \alpha_{N,2} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{1,M} & \alpha_{2,M} & \ldots & \alpha_{N,M}
\end{pmatrix}$$

(3)
3. EM FRAMEWORK FOR MAXIMUM LIKELIHOOD ESTIMATION

First of all, we briefly describe EM algorithm. EM algorithm [5] is an iterative approach to Maximum Likelihood Estimation (MLE), originally formalized in (Demster, Laird and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood \( l(\theta; D) = \log L(\theta; D) \), where \( \theta \) are parameters of the model and \( D \) are the data. Suppose that this optimization problem would be simplified

\[ f(x, d_r, H_r; H_1, d_1) = f(x|H, d)f(d_r; H_1, d_1)f(H_r; H_1, d_1) \]  

\[ d_1 \text{ vector is composed of user 1 transmitted data at all time instants. } f(x|H, d), f(d_r; H_1, d_1) \text{ and } f(H_r; H_1, d_1) \text{ are given by} \]

\[ f(x|H, d) = K_1 \exp\left(\frac{-1}{\sigma^2}(x - H.d)H^T(x - H.d)\right) \]  

\[ f(d_r; H_1, d_1) = K_2 \exp\left(\frac{-1}{2\sigma_r^2}d_r^T d_r\right) \]  

\[ f(H_r; H_1, d_1) = K_3 \exp\left(-\sum_{i=2}^{T} h_i^H R_{h_i h_i} h_i\right) \]

Having the above equations we are now ready to evaluate the E-step of the algorithm. Since we are conditioning on the received data, we take expectations with respect to \( d_r \) (interfering users’ symbols) and their channel, \( H_r \).

\[ Q(H_1, d_1; H_{11}^{(k)}, d_{11}^{(k)}) = E\{\log f(x, d_r, H_r; H_1, d_1|x; H_{11}^{(k)}, d_{11}^{(k)})\} \]

where \((.,)^{(k)}\) is the iteration index and \( E \) is the expectation operator.

Evaluating the expectations and dropping the terms that do not depend on the parameters the above equation can be written as

\[ Q(H_1, d_1; H_{11}^{(k)}, d_{11}^{(k)}) = \sum_{i=1}^{T} E\{(x_i - H_1 d_{i1} - H_r d_{ri})H^T(x_i - H_1 d_{i1} - H_r d_{ri})|x; H_{11}^{(k)}, d_{11}^{(k)}\} \]

\[ d_{ri} \text{ are the symbols transmitted by interfering users (with } N \text{ transmit antennas each) at time instant } i, x_i \text{ is the received signal at instant } i, d_{i1} \text{ is the transmitted data vector of user 1 at instant } i, H_1 \text{ is the channel matrix for user 1 and } H_r \text{ is the group of the interfering users’ information bits transmitted at all time instants and } H_r \text{ is their channel matrix and } x \text{ is composed of the received vector from time instant 1 to time instant } T . \]

Without loss of generality, we will detect user 1 first. The pdf of the complete data set is given by

4. USER-WISE CHANNEL-SYMBOLS ESTIMATION WITH GAUSSIAN MAI PRIOR

The received signal is given by equation (1). Each user channel is modeled as Gaussian vector which might be correlated in space, i.e., between antennas, but assumed independent between users. The channel vector for user \( i \) can be written as \( h_i^T \in N(0, R_{h_i h_i}) \). In the first approach we assume that the interfering symbols as Gaussian i.i.d random variables with known variance \( \sigma_r^2 \). Given \( T \) snapshot, i.e., \( \{x_i\}_T \), we are now ready to define the complete data set. The complete data set is chosen as \( \{x, d_r, H_r\} \), where \( d_r \) is the group of the interfering users’ information bits transmitted at all time instants and \( H_r \) is their channel matrix and \( x \) is composed of the received vector from time instant 1 to time instant \( T \).
the channel matrix for the interfering users. $H = [H_1 H_2]$. The above equation can be further written as

$$Q(H_1, d_1; H_1^{(k)}, d_1^{(k)}) = \sum_{i=1}^{T} E\{x_i - D_i d_1 - D_i h_i | x; H_1^{(k)} d_1^{(k)}\}$$

(12)

where $D_i = d_i \otimes I$ and $h_i = vec(H_1)$ and $I$ is identity matrix. Similarly, we can define $D_i = d_i \otimes I$ and $h_r = vec(H_r)$. We have used the property that $vec(ABC) = (C^T \otimes A)vec(B)$, $\otimes$ is Kronecker product. Differentiating the E-step equation with respect to $h_1$ yields

$$h_1 = \frac{1}{T} \sum_{i=1}^{T} (D_i^T D_i)^{-1} \sum_{i=1}^{T} (D_i^T x_i - D_i^T h_1 \hat{d}_r)$$

(13)

where

$$\hat{d}_r = E\{d_r | x; H, d_1^{(k)}\}$$

(14)

and $\hat{H}_r$ is

$$\hat{H}_r = E\{H_r | x; H_1^{(k)}, d_1^{(k)}\}$$

(15)

Now the problem is to derive the expressions for $\hat{d}_r$ and $\hat{H}_r$, i.e., the conditional mean of the interfering users bit and the conditional mean of the interfering users channel. In order to accomplish this, we first write the pdf for the observed data

$$f(x; H, d_1) = K_3 \exp(-x^H R^{-1}_{xx} x)$$

(16)

where $K_3$ is another constant and $R_{xx}$ is given by

$$R_{xx} = H_1 d_1 H_1^H + H_r d_r H_r^H + \sigma^2 I$$

(17)

where $d_1$ and $d_r$ are the data vector composed of transmitted symbols at all time instants of user 1 and the rest of the users respectively. In deriving the above equation we used the fact that $E\{d_r\} = 0$. From now we will omit the EM iteration index, i.e., $k$. The conditional pdf of $d_r$ as a function of known pdfs is follows (using the fact that transmitted symbols at instant $i$ results in received vector at the same instant)

$$f(d_r; x; H, d_1) = f(x_i | H, d_i) f(d_r | x; H, d_1)$$

(18)

where $d_i$ is the vector of symbols of all users at instant $i$, $x_i$ is the received vector at instant $i$, $d_1$ are the interfering users data vector transmitted at instant $i$ and $H$ is the channel matrix. Substituting the corresponding expressions and rearranging gives

$$f(d_r; x; H, d_1) = \frac{K_1 K_2}{\sigma^2} \exp\left(-\frac{1}{\sigma^2} (x_i - H d_i)^H (x_i - H d_i) - \frac{1}{\sigma^2} d_r^T d_r + x^H R^{-1}_{xx} x\right)$$

(19)

Since the conditional pdf of $\hat{d}_r$ will be Gaussian, it is easy to show that

$$\hat{d}_r = \frac{R_{dd}}{\sigma^2} (H^H x_i - H^H H_1 d_1)$$

(20)

where

$$R_{dd}^{-1} = \frac{1}{\sigma^2} H^H H + \frac{I}{2\sigma^2}$$

(21)

where $I$ is identity matrix. Similarly the expression for $\hat{H}_r$ is as follows

$$\hat{h}_r = vec(\hat{H}_r) = R_{hh} \sum_{i=1}^{T} \frac{1}{\sigma^2} (D_i^T x_i - D_i^T H_1 d_1)$$

(22)

where $R_{hh}^{-1}$ is given by

$$R_{hh}^{-1} = \frac{1}{\sigma^2} \sum_{i=1}^{T} D_i^T D_i + R_{h,h}^{-1}$$

(23)

and $D_i = d_i \otimes I$.

The algorithm detects user-wise channel- symbols. First, user 1 channel-symbols are estimated from the above procedure. Then the contribution of that user is subtracted from the received signal to get more clean signal. Then user second is detected. The same procedure is repeated for the other users. After convergence of the EM algorithm, solution of $d_i$ from equation (12) is projected on finite alphabet to get symbols estimate. The same process is done for the other users too. The overall algorithm works as follows: first we initialize $H_1$ and $d_1$, 2) We evaluate $\hat{d}_r$ from equation (20) and $\hat{h}_r$ from equation (22). These values are plugged in equation (12) and equation (13) to get the channel-symbol update. These steps are repeated until convergence.

5. USER-WISE SYMBOL ESTIMATION USING DISCRETE MAI PRIOR

The steps for deriving the algorithm are essentially the same except that the conditional mean of $d_r$ will be different than previously discussed, i.e., Gaussian random variable for the priors, which will result in different channel-symbols estimates. The conditional mean for $d_r$ is given by

$$\hat{d}_r = E\{d_r | x; H, d_1^{(k)}\} = \sum_{d_r} d_r f(d_r | x; H, d_1^{(k)})$$

(24)

From now for the sake of simplicity we will omit the EM iteration index, i.e., $k$. In order to calculate the conditional mean we have to evaluate the above expression, which is summation of all interfering users’ symbols at instant $i$ multiplied by their corresponding pdfs, which is computationally very expensive. Mean Field (MF) methods [1,7], provide tractable approximations for the computation of high
dimensional sums and integrals in the probabilistic models. By neglecting certain dependencies between the random variables, a closed set of equations for the expected values of these variables are derived which often can be solved in a time that grows polynomially in the number of variables [1, chapter 2]. The MF approximation is obtained by taking the approximating family of probability distribution by all product distribution, i.e., \( Q(d_{ri}) = \Pi_j Q_j(d_{ri}^{j}) \). We now choose a distribution which is close to the true distribution, i.e., \( f(d_{ri} \mid x ; H, d_{i}) \). The parameter of the distribution is chosen so as to minimize Kullback-Leibler (KL) distance, i.e.,

\[
KL(Q \| f(d_{ri} \mid x ; H, d_{i})) = \sum_{d_{ri}} Q(d_{ri}) \ln \frac{Q(d_{ri})}{f(d_{ri} \mid x ; H, d_{i})}
\]

where \( Q(d_{ri}) = \prod_{j=1}^{N} Q_j(d_{ri}^{j}) \) and \( d_{ri}^{j} \in \{ -1, 1 \} \).

\[
f(d_{ri} \mid x ; H, d_{i}) = \frac{f(x_{i} \mid H, d_{i})}{\sum_{d_{ri}} f(x_{i} \mid H, d_{i})} = \frac{\exp(-H(d_{i}))}{Z}
\]

where \( Z \) is independent of \( d_{ri} \), \( f(x_{i} \mid H, d_{i}) \) has Gaussian distribution and \( d_{i} \) is the vector of symbols of all users at instant \( i \). After some simplification \( H(d_{i}) \) can be written as

\[
H(d_{i}) = \frac{1}{\sigma^2} ( -x_{i}^{T} H d_{i} - d_{i}^{T} H^{T} x_{i} + d_{i}^{T} H H d_{i} )
\]

The above equation has the form

\[
H(d_{i}) = \sum_{j,n} d_{ri}^{j} J_{jn} d_{n}^{j} + C
\]

where \( C \) is a term independent of \( d_{ri} \), \( J_{jn} = \frac{1}{N} \operatorname{tr}(H H)_{j,n} \) and \( \theta_{j} = \text{real}(\frac{1}{N} (H H)_{j,j}) \), \( (\cdot , \cdot) \) is the \( j^{th} \) element of the vector \( H H x_{i} \). The KL distance between \( Q \) and \( f(d_{ri} \mid x ; H, d_{i}) \) can be written as

\[
KL(Q \| f(d_{ri} \mid x ; H, d_{i})) = \ln Z + V[Q] - S[Q]
\]

where

\[
S[Q] = -\sum_{d_{ri}} Q(d_{ri}) \ln Q(d_{ri})
\]

is the entropy and

\[
V[Q] = \sum_{d_{ri}} Q(d_{ri}) H(d_{i})
\]

is the variational energy. The most general form of probability distribution for our problem is

\[
Q_{j}(d_{ri}^{j} ; m_{j}) = \frac{1 + \phi_{ri}^{j} m_{j}}{2}
\]

where \( m_{j} \) is the variational parameter which corresponds to the mean, i.e., \( m_{j} = E\{ \phi_{ri}^{j} \} \). The entropy can be written as

\[
S[Q] = -\sum_{j} \frac{1 + \phi_{ri}^{j} m_{j}}{2} \ln \frac{1 + \phi_{ri}^{j} m_{j}}{2} + \frac{1 - \phi_{ri}^{j} m_{j}}{2} \ln \frac{1 - \phi_{ri}^{j} m_{j}}{2}
\]

and similarly variational energy can be written as

\[
V[Q] = \sum_{j,n} J_{jn} m_{j} m_{n} - 2 \sum_{j} m_{j} \theta_{j}
\]

In order to evaluate \( m_{j} \) we have to minimize the variational free energy, i.e.,

\[
F[Q] = V[Q] - S[Q]
\]

Differentiating this equation with respect to \( m_{j} \)’s gives non-linear fixed point equations, i.e.,

\[
m_{j} = \text{tanh}(\frac{1}{2} \sum_{n} J_{jn} m_{n} + \beta_{j}) , j = 1, 2 \cdots (K - 1) N
\]

In the matrix form we can write the above equation as

\[
m = \text{tanh}(\frac{1}{2} J m + \beta)
\]

where \( \beta_{j} = 2 \theta_{j} \). The huge computational task (complexity grows exponentially with the number of interfering users times the transmitted symbols per user) of exact averages over \( f(d_{ri} \mid x ; H, d_{i}) \) has been replaced by solving the above set of \( (K -1)N \) nonlinear equations, which often can be done in time that grows only polynomially with number of interfering users times their transmitted bits. As the above equation is nonlinear there may be local minima or saddle points. In order to avoid it, the solution must be compared by their value of variational free energy \( F[Q] \).

6. CONCLUSIONS

In this paper, we derived two receivers for user-wise joint channel-symbols estimate. In the first approach, the Gaussian prior on the interfering users’ symbols is assumed and EM algorithm is used for user-wise channel symbols estimation. In the second proposed receiver a discrete prior is assumed on the interfering users’ bits. In the later case, the complexity of computing the posteriori probabilities grows exponentially in the number of interfering users times the symbols per user. We derived low complexity method to circumvent this problem. The exact posteriori probabilities are replaced by the approximate separable distributions. The distributions are calculated by MPT (variational approach). Simulation results, using estimated channel, shows very close performance in terms of BER to the exact ML (i.e., when the channel is exactly known) approach.
Fig. 1. Av. BER of K=4 N=2, M=2 vs $SNR\,(dB)$.

7. REFERENCES


