

BLIND ITERATIVE RECEIVER FOR MULTIUSER MIMO SYSTEMS

Ejaz Khan, Dirk Slock

Mobile Communications Department
Institut Eurécom, 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 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

Eurécom's research is partially supported by its industrial partners: Ascom, Cégétel, France Télécom, Hitachi, IBM France, Motorola, Swisscom, Texas Instruments, and Thomson CSF

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

$$x_t = H d_t + n_t \quad (1)$$

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

$$H = [H_1 H_2 \cdots H_K] \quad (2)$$

where H_i is as follows

$$H_i = \begin{pmatrix} \alpha_{1,1}^i & \alpha_{2,1}^i & \cdots & \alpha_{N,1}^i \\ \alpha_{1,2}^i & \alpha_{2,2}^i & \cdots & \alpha_{N,2}^i \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1,M}^i & \alpha_{2,M}^i & \cdots & \alpha_{N,M}^i \end{pmatrix} \quad (3)$$

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 θ are parameters of the model and D are the data. Suppose that this optimization problem would be simplified by the knowledge of the additional variable χ , known as missing or hidden data. The set $D_c = D \cup \chi$ is referred to as the complete data set (in the same context D is referred to as incomplete data set). Correspondingly, the loglikelihood function $l_c(\theta; D_c)$ is referred to as complete data likelihood. χ is chosen such that the function $l_c(\theta; D_c)$ would be easily maximized if χ were known. However, since χ is not observable, l_c is a random variable and cannot be maximized directly. Thus, the EM algorithm relies on integrating over the distribution of χ , with the auxiliary function $Q(\theta, \hat{\theta}) = E_\chi[l_c(\theta; D_c | D, \hat{\theta})]$, which is the expected value of the complete data likelihood, given the observed data D and the parameter $\hat{\theta}$ computed at the previous iteration. Intuitively, computing Q corresponds to filling the missing data using the knowledge of the observed data and previous parameters. The auxiliary function is deterministic and can be maximized. An EM algorithm iterates the following two steps, for $k=1,2,\dots$, until local or global maximum of the likelihood is found.

Expectation: Compute

$$Q(\theta; \theta^{(k)}) = E_\chi[l_c(\theta; D_c | D, \theta^{(k)})] \quad (4)$$

Maximization: Update the parameters as

$$\theta^{(k+1)} = \arg_{\max_\theta} Q(\theta; \theta^{(k)}), \quad (5)$$

In some cases, it is difficult to analytically maximize $Q(\theta; \theta^{(k)})$, as required by the M-step of the above algorithm, and we are only able to compute a new value $\theta^{(k+1)}$ that produces an increase of Q at each iteration. In this case we have so called generalized EM (GEM) algorithm.

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 $h_j^i \in N(0, R_{h_j h_j})$. In the first approach we assume that the interfering symbols as Gaussian i.i.d random

variables with known variance σ_r^2 . Given T snapshot, i.e., $\{x_t\}_1^T$, 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 . Without loss of generality, we will detect user 1 first. The pdf of the complete data set is given by

$$f(x, d_r, H_r; H_1, d_1) = f(x|H, d) f(d_r; H_1, d_1) f(H_r; H_1, d_1) \quad (6)$$

d_1 vector is composed of user 1 transmitted data at all time instants, $f(x|H, d)$, $f(d_r; H_1, d_1)$ and $f(H_r; H_1, d_1)$ are given by

$$f(x|H, d) = K_1 \exp\left(\frac{-1}{\sigma^2} (x - Hd)^H (x - Hd)\right) \quad (7)$$

where K_1 is constant not depending on parameters to be estimated, $(\cdot)^H$ is the Hermitian transpose and

$$f(d_r; H_1, d_1) = K_2 \exp\left(\frac{-1}{2\sigma_r^2} d_r^T d_r\right) \quad (8)$$

where K_2 is another constant. In the above equation we have assumed without loss of generality that the prior mean for the interfering users' symbols is zero and the variance σ_r^2 of the symbols is known.

$$f(H_r; H_1, d_1) = K_3 \exp\left(-\sum_{i=2}^{TK} h_i^H R_{h_i h_i}^{-1} h_i\right) \quad (9)$$

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 .

$$Q(H_1, d_1; H_1^{(k)}, d_1^{(k)}) = E\{\log f(x, d_r, H_r; H_1, d_1 | x; H_1^{(k)}, d_1^{(k)})\} \quad (10)$$

where $(\cdot)^{(k)}$ 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

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

d_{ri} are the symbols transmitted by interfering users (with N transmit antennas each) at time instant i , x_i is the received signal at instant i , d_{1i} is the transmitted data vector of user 1 at instant i , H_1 is the channel matrix for user 1 and H_r is

the channel matrix for the interfering users. $H = [H_1|H_r]$. The above equation can be further written as

$$Q(H_1, d_1; H_1^{(k)}, d_1^{(k)}) = \sum_{i=1}^T E\{(x_i - D_{1i}h_1 - D_{ri}h_r)^H (x_i - D_{1i}h_1 - D_{ri}h_r) | x; H_1^{(k)}, d_1^{(k)}\} \quad (12)$$

where $D_{1i} = d_{1i} \otimes I$ and $h_1 = \text{vec}(H_1)$ and I is identity matrix. Similarly, we can define $D_{ri} = d_{ri} \otimes I$ and $h_r = \text{vec}(H_r)$. We have used the property that $\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$, \otimes is Kronecker product. Differentiating the E-step equation with respect to h_1 yields

$$h_1 = \sum_{i=1}^T (D_{1i}^T D_{1i})^{-1} \sum_{i=1}^T (D_{1i}^T x_i - D_{1i}^T \hat{H}_r \hat{d}_r) \quad (13)$$

where

$$\hat{d}_{ri} = E\{d_{ri} | x; H, d_1^{(k)}\} \quad (14)$$

and \hat{H}_r is

$$\hat{H}_r = E\{H_r | x; H_1^{(k)}, d_1^{(k)}\} \quad (15)$$

Now the problem is to derive the expressions for \hat{d}_{ri} 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

$$f(x; H, d_1) = K_3 \exp(-x^H R_{xx}^{-1} x) \quad (16)$$

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

$$R_{xx} = H_1 d_1 d_1^H H_1^H + H_r d_r d_r^H H_r^H + \sigma^2 I \quad (17)$$

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_{ri} 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)

$$f(d_{ri} | x; H, d_1) = f(x_i | H, d_i) f(d_{ri}) / f(x_i; H, d_{1i}) \quad (18)$$

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

$$f(d_{ri} | x; H, d_1) = \frac{K_1 K_2}{K_3} \exp\left(-\frac{1}{\sigma^2} (x_i - H d_i)^H (x_i - H d_i) - \frac{1}{2\sigma_r^2} d_{ri}^T d_{ri} + x^H R_{xx}^{-1} x\right) \quad (19)$$

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

$$\hat{d}_{ri} = \frac{R_{dd}}{\sigma^2} (H_r^H x_i - H_r^H H_1 d_{1i}) \quad (20)$$

where

$$R_{dd}^{-1} = \frac{1}{\sigma^2} H_r^H H_r + \frac{I}{2\sigma_r^2}, \quad (21)$$

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

$$\hat{h}_r = \text{vec}(\hat{H}_r) = R_{hh} \sum_{i=1}^T \frac{1}{\sigma^2} (D_{ri}^T x_i - D_{ri}^T H_1 d_{1i}) \quad (22)$$

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

$$R_{hh}^{-1} = \frac{1}{\sigma^2} \sum_{i=1}^T D_{ri}^T D_{ri} + R_{h_r h_r}^{-1} \quad (23)$$

and $D_{ri} = d_{ri} \otimes I$.

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_{1i} 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}_{ri} 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_{ri} 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_{ri} is given by

$$\hat{d}_{ri} = E\{d_{ri} | x; H, d_1^{(k)}\} = \sum_{d_{ri}} d_{ri} f(d_{ri} | x; H, d_1^{(k)}) \quad (24)$$

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., $Q(d_{ri}) = \prod_j Q_j(d_{ri}^j)$. We now choose a distribution which is close to the true distribution, i.e., $f(d_{ri}|x; H, d_1)$. The parameter of the distribution is chosen so as to minimize Kullback-Leibler (KL) distance, i.e.,

$$KL(Q||f(d_{ri}|x; H, d_1)) = \sum_{d_{ri}} Q(d_{ri}) \ln \frac{Q(d_{ri})}{f(d_{ri}|x; H, d_1)} \quad (25)$$

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

$$f(d_{ri}|x; H, d_1) = \frac{f(x_i|H, d_i)}{\sum_{d_{ri}} f(x_i|H, d_i)} = \frac{\exp(-H(d_i))}{Z} \quad (26)$$

where Z is independent of d_{ri} , $f(x_i|H, d_i)$ 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

$$H(d_i) = \frac{1}{\sigma^2} (-x_i^H H d_i - d_i^T H^H x_i + d_i^T H^H H d_i) \quad (27)$$

The above equation has the form

$$H(d_i) = \sum_{j,n} d_{ri}^j J_{jn} d_{ri}^n - 2 \sum_j d_{ri}^j \theta_j + C \quad (28)$$

where C is a term independent of d_{ri} , $J_{jn} = \frac{1}{\sigma^2} (H^H H)_{j,n}$ and $\theta_j = \text{real}(\frac{1}{\sigma^2} (H^H x_i)_j)$, $(\cdot)_j$ is the j^{th} element of the vector $H^H x_i$. The KL distance between Q and $f(d_{ri}|x; H, d_1)$ can be written as

$$KL(Q||f(d_{ri}|x; H, d_1)) = \ln Z + V[Q] - S[Q] \quad (29)$$

where

$$S[Q] = - \sum_{d_{ri}} Q(d_{ri}) \ln Q(d_{ri}) \quad (30)$$

is the entropy and

$$V[Q] = \sum_{d_{ri}} Q(d_{ri}) H(d_i) \quad (31)$$

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

$$Q_j(d_{ri}^j; m_j) = \frac{1 + d_{ri}^j m_j}{2} \quad (32)$$

where m_j is the variational parameter which corresponds to the mean, i.e., $m_j = E\{d_{ri}^j\}$. The entropy can be written as

$$S[Q] = - \sum_j \frac{1 + m_j}{2} \ln \frac{1 + m_j}{2} + \frac{1 - m_j}{2} \ln \frac{1 - m_j}{2} \quad (33)$$

and similarly variational energy can be written as

$$V[Q] = \sum_{j,n} J_{jn} m_j m_n - 2 \sum_j m_j \theta_j \quad (34)$$

In order to evaluate m_j we have to minimize the variational free energy, i.e.,

$$F[Q] = V[Q] - S[Q] \quad (35)$$

Differentiating this equation with respect to m_j 's gives nonlinear fixed point equations, i.e.,

$$m_j = \tanh(- \sum_n J_{jn} m_n + \beta_j), j = 1, 2, \dots, (K-1)N \quad (36)$$

In the matrix form we can write the above equation as

$$\mathbf{m} = \mathbf{tanh}(-\mathbf{J}\mathbf{m} + \beta) \quad (37)$$

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 $f(d_{ri}|x; H, d_1)$ 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 MFT (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.

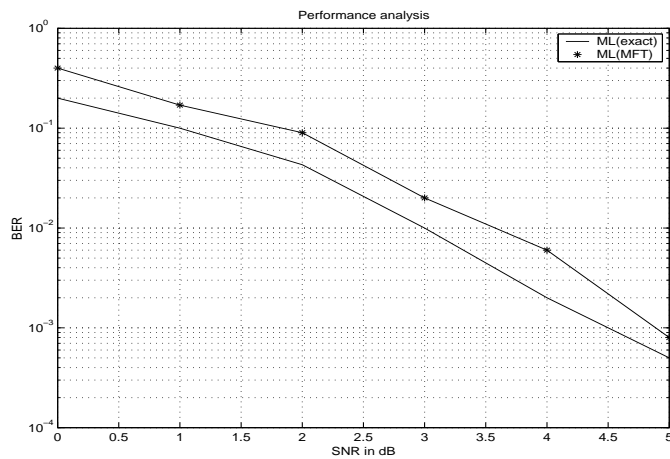


Fig. 1. Av. BER of $K=4$ $N=2$, $M=2$ vs SNR (dB).

7. REFERENCES

- [1] M. Opper and D. Saad, eds, "Advanced mean field methods: Theory and practice". MIT press, 2001.
- [2] B. Lu and X. Wang, "Iterative receivers for multiuser Space-time coding systems", IEEE JSAC vol. 18, No. 11, November 2000.
- [3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," IEEE JSAC vol. 17, No. 3, March 1999.
- [4] E.de. Carvalho and D.T.M. Slock, "Maximum likelihood FIR multi-channel estimation with Gaussian prior for the symbols," Proc. ICASSP, Munich, Germany, April 1997.
- [5] G. McLachlan and T. Krishnan, "The EM algorithm and extensions", John Wiley ad sons Inc. 1997.
- [6] V. Tarokh, A. Naguib, N. Seshadri, and A. R. calderbank, "Combined array processing and space-time coding", IEEE Tran. Information Theory, vol. 45, No. 4, May 1999.
- [7] G. Parisi, "Statistical field theory", Addison Wesley, Redwood, CA, 1988.