ITERATIVE BLIND DEMODULATION OF SYNCHRONOUS CDMA

Ejaz Khan and Dirk.T.M Slock

Mobile Communications Department
Institut Euré com,
B.P. 193, 06904 Sophia–Antipolis CEDEX, France
Tel: +33 4 93 00 26 76, Fax: +33 4 93 00 26 27
E-mail: firstname.lastname@eurecom.fr

Abstract - In this paper iterative blind estimation of the complex amplitudes of the users is considered. A Gaussian mixture model formulation of the problem is introduced and Expectation Maximization (EM) algorithm for estimation of parameters for Gaussian mixture observation model is used. Simulation results compare the performance of the proposed algorithm with the Cramer-Rao bound.

Keywords - Maximum Likelihood, expectation maximization, Gaussian mixture, CDMA, hidden variable, convergence

I. INTRODUCTION

Code Division Multiple Access (CDMA) is one of the most common multiple access techniques for wireless communication systems. In CDMA all users use entire frequency band and are separated at the receiver by each user's quasi-orthogonal spreading codes to reduce inter-user interference. In recent years, various kinds of receivers have been proposed for the CDMA system. In this paper we consider the problem of estimating the received amplitudes of the users knowing only their spreading codes. Talwar, et al [5] proposed iterative least square with enumeration (ILSE) which solves the problem by estimating the channel by short training sequence or from previous estimates and find the data sequence over all possible data in the Finite Alphabet (FA). They also proposed iterative least square with projection (ILSP) which also initially estimates the channel with the same method as for ILSE and treats the problem as continuous optimization problem and projects the results onto closest discrete alphabet. In [4,9], the authors consider projection of the received signal on the signal subspace of the received signal autocorrelation matrix and after apply the method of parameter estimation for Gaussian mixture. Their proposed algorithm is faced with two problems

1) eigenvalue decomposition of the received signal autocorrelation matrix (a computationally complex operation), an other algorithm must be used for signal subspace tracking and also signal subspace mismatch can deteriorate estimation of the parameters
2) The most important one is that by projecting the data vector onto signal subspace or any other matrix of lower dimension, it can be imagined (and this is borne out by experience with EM like techniques), that the result would not converge to true means of the Gaussians. This can be explained by the following reason: Let M be the number of Gaussians and P be the dimensionality of the data. If the dimension is decreased from P to q, the average distance between any two means decreases as $\sqrt{q/P}$, and the probability that the means are separated by less than 2σ increases. The criterion for the separation of two Gaussian distributions in one dimension is that the distance between two means is greater than twice the standard deviation (2σ). Furthermore, Gaussians that are poorly separated in the original dimension will tend to become even more poorly separated as the dimensionality is decreased. Thus, it is very important that the Gaussians remain well separated after projection onto lower dimensional space. If they are not, it will be difficult for the EM algorithm to recognize overlapping components as distinct Gaussian distributions resulting in a total failure of the EM algorithm.

In our approach we consider directly the output of the channel (the received signal) as the mixture of a known number of Gaussian and estimate its parameters thus avoiding the above mentioned problems and by keeping spreading factor not too high, the computational complextiy is kept moderate (comparable to the case in which projection is done) . Direct Maximum Likelihood (ML)estimation of parameters is complex and therefore we use expectation maximization (EM) algorithm to find the parameters of our model. Mixture models, in particular mixtures of Gaussian, have been a popular tool for density estimation, clustering and unsupervised learning with wide range of applications. Mixture models are one of the most useful tools for handling incomplete data, in particular hidden variables. For Gaussian mixtures the hidden variable indicate for each data point the index of the Gaussian that generated it. The EM technique is used to iteratively update the maximum likelihood estimate of the parameters of the mixture which are used to obtain amplitudes of the users.

Eurecom's research is partially supported by its industrial partners: Ascom, Cé gé tel, France Té lé com, Hitachi, IBM France, Motorola, Swisscom, Texas Instruments, and Thomson CSF

The rest of the paper is organized as follows: The signal model for the problem is described in section 2. Section 3 is devoted to the development of the EM based algorithm for solving the problem. In section 4,5 and 6 convergence rate, simulations, performance is analyzed, and conclusion is drawn, respectively.

II. Signal model

We consider DS CDMA with K-users with processing gain P. The output of the channel is chip matched filtered and sampled at the chip rate. The system is assumed to be synchronous. In a single data interval we have P-dimensional vector $\mathbf{x}$, given by

$$\mathbf{x} = \mathbf{SAb} + \mathbf{n} \tag{1}$$

where $\mathbf{S}$ is $P \times K$ matrix whose columns are K users normalized spreading sequences:

$$\mathbf{S} = [s_1|s_2|.......s_K] \tag{2}$$

$\mathbf{A}$ = diag$(A_1, A_2, ...A_K)$, k users' received amplitudes, $\mathbf{b} = [b_1, b_2, \ldots, b_K]$ contains the symbols transmitted by the users and $\mathbf{n}$ is P dimensional Gaussian random vector for noise with covariance matrix given by $\sigma^2 I$, where I is identity matrix.

We assume that symbols of different users are independent i.e. $E[b_k b_l] = 1$ for $k = l$ and 0 otherwise.

We can write equation (1) as

$$\mathbf{x} = \mathbf{Hb} + \mathbf{n}, \tag{3}$$

where $\mathbf{H} = \mathbf{SA}$ is $(P \times K)$ dimensional matrix.

Given model of equation (3) our goal is to estimate $\mathbf{A}$(i.e. users signal amplitudes) from multiple independent observations of $\mathbf{x}$.

III. EM framework for Maximum Likelihood estimation

First of all, we briefly describe EM algorithm. EM algorithm 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 $l(\theta; D) = logL(\theta; D)$, 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 refered to as the complete data set (in the same context D is refered to as incomplete data set). Correspondingly, the loglikelihood function $l_c(\theta; D_c)$ is refered 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 obsevable, 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,...., until local or global maximum of the likelihood is found.

Expectation: Compute

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

Maximization: Update the parameters as

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

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.

We consider the BPSK case in which the transmitted data takes on two possible values $\{-1, +1\}$ with all symbol vectors being equally likely.

In ML estimation problem we have density function $P(\mathbf{x}|\theta)$ that is governed by the set of parameters θ (e.g. P might be set of Gaussians and θ could be the means and covariances). The data is of size N, supposedly drawn from this distribution, i.e $X = [\mathbf{x}_1, ...\mathbf{x}_N]$. That is, we assume that these data vectors are independent identically distributed (i.i.d) with distribution P. Therefore the resulting density for the samples is

$$p(X|\theta) = \prod_{t=1}^{N} P(\mathbf{x}_t|\theta) = L(\theta|X).$$

This function $L(\theta|X)$ is called the likelihood of the parameters given the data, or just the likelihood function. In the ML problem, our goal is to find θ that maximizes L. That is, we wish to find θ^* where

$$\theta^* = \arg\max_\theta L(\theta|X). \tag{6}$$

Assuming that the channel output i.e. $\mathbf{x}$ can be approximated by Gaussian distributions i.e. $P(\mathbf{x}/\theta)$ can be modeled as P-dimensional mixture of Gaussians. We can write

$$P(\mathbf{x}|\theta) = \sum_{j=1}^{M} \alpha_j P(\mathbf{x}|\mathbf{m}_j, \Sigma_j), \tag{7}$$

where $M = 2^K$ and

$$P(\mathbf{x}|\mathbf{m}_j, \Sigma_j) = \tag{8}$$

$$\frac{1}{(2\pi)^{(P/2)}|\Sigma_j|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_j)^T \Sigma_j^{-1}(\mathbf{x} - \mathbf{m}_j)\right), \tag{9}$$

$\alpha_j \geq 0$, and $\sum_{j=1}^{M} \alpha_j = 1$. The parameter vector θ consists of mixing proportions α_j, the means vectors $\mathbf{m}_j$, and the covariance matrices Σ_j. Given M and given N independent, i.i.d samples $\{x_t\}_1^N$, we obtain the following likelihood

$$l(\theta) = \sum_{t=1}^{N} \log \sum_{j=1}^{M} \alpha_j P(\mathbf{x}|\mathbf{m}_j, \Sigma_j) \tag{10}$$

which is difficult to optimize because it contains logarithm of a sum. If we consider X as incomplete, since we do not know which index j, within the mixture probability density function output has originated. The complete ”data set” in this case is $[\mathbf{x}_1, \ldots, \mathbf{x}_N, i_1, \ldots, i_N]$, where i_n denotes the component of the pdf from which $\mathbf{x}_n$ is drawn. Using complete data set we can optimize our problem using EM algorithm (see, e.g [1],[2])

The update for means is given by the following equation.

$$\mathbf{m}_j^{(k+1)} = \frac{\sum_{t=1}^{N} h_j^{(k)}(t)\mathbf{x}_t}{\sum_{t=1}^{N} h_j^{(k)}(t)}, \tag{11}$$

where the posteriori probabilities $h_j^{(k)}(t)$ is defined as follows:

$$h_j^{(k)}(t) = \frac{\alpha_j^{(k)} P\left(\mathbf{x}_t|\mathbf{m}_j^{(k)}, \Sigma_j^{(k)}\right)}{\sum_{i=1}^{M} \alpha^{(k)} P\left(\mathbf{x}_t|\mathbf{m}_j^{(k)}, \Sigma_j^{(k)}\right)}. \tag{12}$$

The mixing proportions (α_j) and Covariance matrices in our case are constant and is given by 2^{-K} and $\sigma^2 I$ respectively.

The algorithm works as follows, first posteriori probabilities are calculated using initial estimates of means. The posteriori probabilities tells us the likelihood that a point belongs to each of the separate component densities with respect to the current parameters. These posteriori estimates are used to find the update means of the mixture. These two steps are repeated until convergence. The convergence of the EM algorithm to a solution and the number of iterations depends on the tolerance, the initial parameters, the data set, etc. After convergence of the algorithm the estimate of $\mathbf{H}$ is given by

$$\hat{\mathbf{H}} = \sum_{j=1}^{M} \mathbf{m}_j \mathbf{b}_j^T \left(\sum_{j=1}^{M} \mathbf{b}_j \mathbf{b}_j^T \right)^{-1}. \tag{13}$$

From an $\hat{\mathbf{H}}$, estimate of the matrix $\mathbf{H}$, it is straightforward to obtain estimates of the users' signal amplitudes by following equation

$$\hat{A} = (S^T S)^{-1} S^T \hat{H} \tag{14}$$

Data sequence can be estimated by employing method of least square.

IV. Convergence rate of EM algorithm for Gaussian mixtures

Theorem: For Gaussian mixtures the convergence rate r of EM algorithm for means and hence for the channel is bounded by

$$r = \frac{\left\| m_j^{(k+1)} - m_j^* \right\|}{\left\| m_j^{(k)} - m_j^* \right\|} \leq \left\| I + P_{m_j}^* h_{mj}^* \right\| = \tag{15}$$

$$||A|| = \frac{\left\| H^{(k+1)} - H^* \right\|}{\left\| H^{(k)} - H^* \right\|} \tag{16}$$

where m_j, H are means and channel coefficients respectively and $*$ denotes the converged point. I and h_{mj}^* denotes the Identity matrix and Hessian of the likelihood function at m_j^* and $P_{m_j} = \frac{I}{\sum_{t=1}^{N} h_j^{(k)}(t)}$. Higher the values of $||A||$, slower will be the convergence.

Proof: Xu and Jordan (1996) showed that each iteration the following relationship holds between the gradient of the loglikelihood and the EM update step:

$$m_j^{(k+1)} - m_j^{(k)} = P_{m_j}^{(k)} \frac{\partial l}{\partial m_j}|_{m_j = m_j^{(k)}} \tag{17}$$

From the above equation, using Taylor expansion around the convergent point m_j^* for large k and noting that $P_{m_j}^{(k)} \frac{\partial l}{\partial m_j}|_{m_j = m_j^*} = 0$, we have

$$m_j^{(k+1)} = m_j^{(k)} + P_{m_j}^* h_{mj}^* (m_j^{(k)} - m_j^*) \tag{18}$$

which can be written as

$$m_j^{(k+1)} - m_j^* = m_j^{(k)} - m_j^* + P_{m_j}^* h_{mj}^* (m_j^{(k)} - m_j^*) \tag{19}$$

$$m_j^{(k+1)} - m_j^* = (I + P_{m_j}^* h_{mj}^*)(m_j^{(k)} - m_j^*) \tag{20}$$

which after using Schwarz inequality follows the result. It has been proved by Ma et al. that asymtotic convergence rate of EM for Gaussian mixtures locally around the true solution m_j^* is $O(e^{0.5-\epsilon}(m_j^*))$ where $\epsilon > 0$ is an arbitrary small number, O(x) means that it is higher order infinitismal as $x \to 0$, and $e(m_j^*)$ is a measure of overlap of Gaussians in the mixture. In other words, large sample local convergence rate for the EM algorithm tends to be asymtotically superlinear when $e(m_j^*)$ tends to zero.

V. Simulations

The performance of the proposed method was evaluated as a function of SNR (signal to noise ratio) based on Monte carlo simulations. The method was tested for 500 Monte Carlo trials per SNR point across range of SNR's. In each trial, the amplitude estimation error was recorded. The data block of 32 symbols were used in all simulations. The spreading gain was 32. The proposed method worked quite well for the two and three users case (due to the fact that there were only four and eight mixture of Gaussians respectively). In figure 1, the performance is compared with the approximate Cramer Rao bound which is not as tight as Cramer-Rao bound (CRB) and the difference between the simulation and CRB can be explained by the fact that the initial parameter values for the EM algorithm were given as random numbers i.e initial values were not confined to be in the vacinity the true value of the parameter. This was done in order to show results for EM in more realistic way (because in reality it is very difficult to know a priori good starting points for an algorithm). Figure 2 shows comparison of the estimation error for three and four users. Beyond three users the estimation error increased quite substantially (as is clear from figure 2). This effect can be explained from the fact that as the number of users increases it is more probable for the EM algorithm to converge at false means of the mixture of Gaussians (if random initialization is done as in our case). Therefore very good initialization is needed when number of users grows large.

VI. Conclusion

In this paper we presented a Gaussian mixture formulation of the problem to blindly estimate the users amplitudes for the synchronous CDMA system. We proposed EM based algorithm to estimate the parameters of the mixture. Theoretical convergence rate for the means in the Gaussian mixture case was also presented. Simulations result shows usefulness of the method. The estimation error is compared with approximate Cramer-Rao lower bound.

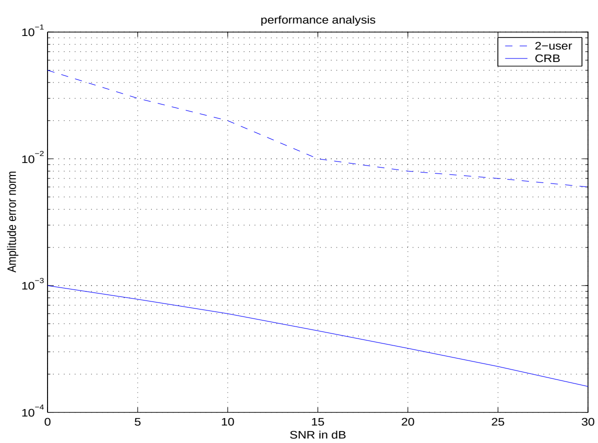

Fig. 1. Amplitude estimation error.

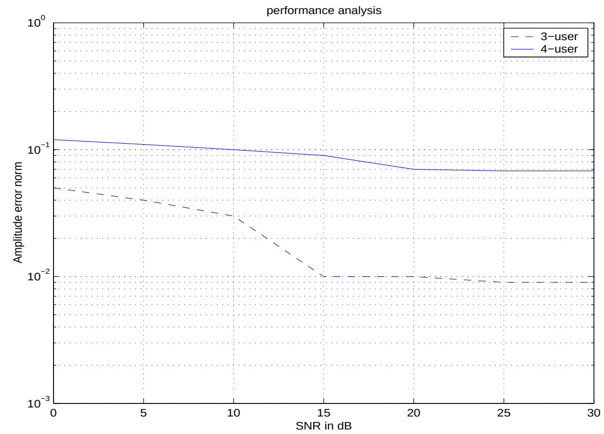

Fig. 2. Amplitude estimation error.

References

[1] J .A .Bilmes, "A gentle tutorial of EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov model," tech. rep., U.C.Berkely, April 1998.

[2] A .P .Dempster, N .M .Laird, and D .B .Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J.Royal Statistical Society B*, no. 39, pp. 1–38, 1977.

[3] G. McLachlan and T. Krishnan, *The EM algorithm and extension*. John Wiley and sons Inc, 1997.

[4] Y .Yao and H .V .Poor, "EM based blind demodulation of synchronous cdma," IEEE 6^{th} Int. symp. on spread spectrum Tech. and Application, 2000.

[5] S .Talwar, M . Viberg, and A .Paulraj,"Blind separation of synchronous co-channel digital signals using an antenna array-Part-1: algorithms," IEEE trans. signal Processing, 44,pp. 1184-1197, 1996

[6] L.Xu and M. I. Jordan,"On convergence properties of the EM algorithm for gaussian mixtures". Neural Computation,8, 129-151, 1996.

[7] J. Ma, L. Xu and M. I. Jordan,"Asymtotic convergence rate of the EM algorithm for gaussian mixtures". Neural computation,12, 2881-2907, 2000.

[8] S. Dasgupta,"Learning mixture of Gaussians". IEEE symposium on foundation of computer science, 1999.

[9] Y .Yao and H .V .Poor, "Eavesdropping in the synchronous CDMA channel: An EM based approach" IEEE transaction on signal processing, vol.49, No.8, August 2001.