

# INFORMATION THEORETIC COMPARISON OF TRAINING, BLIND AND SEMI BLIND SIGNAL SEPARATION ALGORITHMS IN MIMO SYSTEMS

Veeraruna Kavitha

Vinod Sharma

Department of Electrical Communication Engg, Indian Institute of Science, Bangalore, India  
email: kavitha@pal.iisc.ernet.in, vinod@ece.iisc.ernet.in

**Abstract** - We consider the problem of comparing source/signal separation algorithms for MIMO wireless channels. The channel state is not known to the transmitter and the receiver. The receiver estimates the channel via a training sequence or uses blind methods to separate the signals. By comparing the 'capacity' of a composite channel, we find the optimal training size in training and semiblind methods. We also conclude that for Rician channels (with strong line of sight, LOS) (semi)blind algorithms outperform training based methods but for Rayleigh fading one should use training based methods only.

**Keywords** : Channel estimation, MIMO channels, Signal separation algorithms, Training sequence, Blind algorithms, CMA.

## 1. INTRODUCTION

An important component in any communication system is efficient, accurate channel estimation and equalization. Due to time varying nature of the channel, for a good channel estimate, one needs to send the training sequence frequently. Therefore, a significant ( $\sim 18\%$  in GSM) fraction of the channel capacity is consumed by the training sequence. The usual blind equalization techniques have also been found to be inadequate ([7]) due to their slow convergence and/or high computational complexity. Therefore, semi-blind algorithms, which use some training sequence along with blind techniques have been proposed (Chapter 7 of [7] and references therein). In this paper, we provide a systematic comparison of the training based, blind and semi-blind algorithms.

In comparing training based methods with blind algorithms one encounters the problem of comparing the loss in BW in training based methods (due to training symbols) with the gain in BER (due to better channel estimation/equalization accuracies) as compared to the blind algorithms. We overcome this problem by comparing these methods via the channel capacity they provide. Towards this goal, we combine the channel, equalizer and the decoder to form a composite channel. (We use the misnomer equalizer for the source separation algorithm throughout the paper for convenience of presentation). The input to this channel will be symbols from a finite alphabet and the output of this channel will be the decoded symbols (either hard or soft decision decoding). Hence this is a discrete channel. The capacity of this composite channel will be a good measure for comparison of the various signal separating algorithms.

This problem for the training based methods has also

been studied in [1] and [8]. They obtain a lower bound on the channel capacity and find the optimal training sequence length (and also placement in case of [1]). We not only study the problem of optimal training sequence length, but also compare it to blind and semi blind methods. Obtaining channel capacity of the composite channel for training based methods, in our model is not difficult. But for the blind and semiblind methods, one needs to know the equalizer value at the end of the frame. At that point the equalizer will often be away from the equilibrium point. Thus, we need the value of the equalizer at a specific time under transience and it depends upon the initial conditions, the value of fading channel and the receiver noise realizations. This is not practically feasible to obtain the capacity of such a composite channel. We circumvent this problem by using the result in [11], where given the initial conditions, the equalizer value at any time (even under transience) can be obtained from the value of the solution of an ODE. Also, this value is independent of the receiver noise realizations. Based on this analysis we find that in LOS conditions even though the (semi)blind algorithm might not have converged, it can perform better than training based methods.

The paper is organized as follows. Section 2 describes our model and the approach. Section 3, 4 and 5 consider training, blind and semiblind algorithms respectively. Section 6 compares the capacity of the three algorithms using few examples and Section 7 concludes the paper. Proofs are given in the Appendix.

## Notations

The following notations are used throughout the paper. All bold capital letters represent complex matrices, capital letters represent real matrices and small letters represent the complex scalars. Bold small letters with bar on the top represent complex column vectors, while the same bold small letters without bar represent their real counter parts given by,  $\mathbf{y} \triangleq \text{Real}(\bar{\mathbf{y}}) = [\bar{y}_{1,r} \ \bar{y}_{2,r} \ \dots \ \bar{y}_{n,r} \ \bar{y}_{1,i} \ \dots \ \bar{y}_{n,i}]^T$  (for an  $n$  dimensional complex vector  $\bar{\mathbf{y}}$ ). Here  $\bar{y}_{k,r} + i\bar{y}_{k,i}$  represents  $k^{\text{th}}$  element of vector  $\bar{\mathbf{y}}$ .  $^H, ^T, ^{HT}$  represent Hermitian, transpose and conjugate respectively. Two more real vectors corresponding to a complex vector are defined as,  $\tilde{\mathbf{y}} = \text{Real}(\bar{\mathbf{y}}^{HT})$  and  $\hat{\mathbf{y}} = \text{Real}(i\bar{\mathbf{y}}^{HT})$ . Note that  $\mathbf{e}^T \bar{\mathbf{y}}$ ,  $\mathbf{e}^T \hat{\mathbf{y}}$  equal

This research is partially supported by DRDO-IISc program on advanced research in Mathematical Engineering.

the real and imaginary parts of the complex inner product  $\bar{\mathbf{e}}^H \bar{\mathbf{y}}$  respectively. For any complex matrix  $\mathbf{H}$ , the corresponding  $\bar{\mathbf{h}}$  is given by  $\text{vect}(\mathbf{H})$ , where  $\text{vect}(\mathbf{H})$  represents the complex matrix  $\mathbf{H}$  in vector form by concatenating elements of all vectors one after the other. For any vector (matrix)  $\bar{\mathbf{y}}(\mathbf{H})$ ,  $\hat{\bar{\mathbf{y}}}(\hat{\mathbf{H}})$  represents its estimate, obtained either by MMSE criterion or minimum distance criterion.

## 2. THE MODEL AND OUR APPROACH

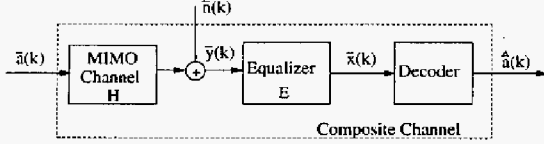


Fig. 1. Block Diagram of the System

Consider a wireless channel with  $m$  transmit antennas and  $n$  receive antennas with  $m \leq n$ . The time axis is divided into frames; each frame consisting of  $N$  channel uses. The transmitted and decoded symbols are from finite alphabets,  $\mathcal{S} = \{s_1, s_2, \dots, s_L\}$  and  $\mathcal{S}_d = \{s_{d_1}, s_{d_2}, \dots, s_{d_{L_d}}\}$  respectively (This includes hard/soft decision decoding). At time  $k$ , vector  $\bar{\mathbf{a}}(k) \in \mathcal{S}^m$  is transmitted from the  $m$ -transmit antennas and the decoded symbol  $\hat{\bar{\mathbf{a}}}(k) \in \mathcal{S}_d^m$ . We represent the elements of  $\mathcal{S}^m$ ,  $\mathcal{S}_d^m$  by  $\mathcal{S}^m = \{\bar{s}_i; 1 \leq i \leq L^m\}$  and  $\mathcal{S}_d^m = \{\bar{s}_{d_i}; 1 \leq i \leq L_d^m\}$  respectively. The channel gain matrix  $\mathbf{H}$  is assumed constant during a frame (quasi static channel). We assume it to vary independently from frame to frame. These are commonly made assumptions. Thus, if we use training based methods, the channel estimate for a frame depends upon the received training sequence during that frame only. Therefore, we will need to consider a single frame in this paper.

The vectors received at the  $n$  receive antennas in any frame are  $\bar{\mathbf{y}}(k) = \mathbf{H}\bar{\mathbf{a}}(k) + \bar{\mathbf{n}}(k) \quad \forall 1 \leq k \leq N$  (1) where  $\bar{\mathbf{n}}(k)$  is an iid sequence of complex Gaussian vectors with mean 0 and co-variance  $\sigma_n^2 I$  (denoted by  $CN(0, \sigma_n^2 I)$ ) and  $\bar{\mathbf{h}} \sim CN(\mu_{\bar{\mathbf{h}}}, \sigma_{\bar{\mathbf{h}}}^2 I)$ . This is a Rayleigh/Rician channel with independent components. Our approach can be easily extended to complex Gaussian channels with any arbitrary covariance matrix as well. This is a flat fading channel. At the receiver one wants to estimate  $\bar{\mathbf{a}}(k)$ . One common way is to use an equalizer at the receiver to nullify the effect of  $\mathbf{H}$  and then detect the transmitted symbol.

We assume that the channel statistics is available at the transmitter and the receiver but the actual channel gain matrix  $\mathbf{H}$  is not known to the receiver and the transmitter. The receiver tries to estimate  $\mathbf{H}$ , or directly obtain an equalizer to estimate/detect the information symbols transmitted. For this the most common method used in wireless channels is to send a known training sequence in the frame. This is used by the receiver to estimate  $\mathbf{H}$  (say via the maximum likelihood estimation (MLE)) and then obtain an equalizer. In the rest of the frame, information symbols are transmitted and are decoded at the receiver using the channel estimate. If a longer training sequence is used, we obtain a better channel estimate

resulting in lower BER. However, we loss channel BW (capacity) because information symbols are sent for a shorter duration. Thus one needs to find the 'optimal' training sequence length for a given channel. Alternatively, one can estimate an equalizer using only the statistics of the received and transmitted signal. These are blind methods and do not require training sequences, but may not be accurate. One can expect that by combining some of the blind methods with the training methods we can obtain the same performance with a shorter training sequence and hence better capacity.

To address the issue of fair comparison of various equalizers (training, blind and semiblind), we form a 'composite' channel, made of the channel, the channel estimator/equalizer and the decoder. It forms a finite input - output alphabet time invariant channel. It would be a time invariant channel as the channel state is not known to the transmitter, and hence the transmitter would experience average behavior in every frame. We will show that one can compute the composite channel's transition matrix and hence it's capacity  $C$  using the statistics of the original channel at transmitter and receiver. Since all the symbols in the frame undergo same fading and also since the equalizer used by them is same, the capacity of the channel per channel use is  $(N - N_t)C/N$ . We need to find  $N_t$  which maximizes  $(N - N_t)C$ .

We will carry out the above program in section 3 for training based methods. Next we will consider the blind channel estimation/equalization algorithms. Since Constant Modulus Algorithm (CMA) ([7]) has been one of the most used and successful algorithms, we will consider CMA. We will obtain the channel capacity of the composite channel corresponding to this system using the results in [11]. This will provide a more systematic comparison of the training based and blind algorithms. Finally we will combine the two methods and obtain a semi-blind algorithm.

## 3. TRAINING BASED CHANNEL ESTIMATION

We first estimate the channel via the Minimum Mean Square Estimator (MMSE) [10] (which is also MLE in this situation) using  $mN_t$  training symbols. A MMSE equalizer is designed using the channel estimate and then used in Maximum Likelihood (ML) decoding (equivalent to minimum distance decoding in Gaussian channels) of the entire frame.

Let  $\bar{\mathbf{y}}_{TS} \hat{=} \mathbf{A}_{TS} \bar{\mathbf{h}} + \bar{\mathbf{n}}_{TS}$  (note  $\bar{\mathbf{h}} = \text{vect}(\mathbf{H})$ ) represent the  $nN_t$  length received data corresponding to all  $mN_t$  training symbols. Here  $\mathbf{A}_{TS}$  is an  $nN_t \times nm$  complex matrix suitably formed from the known training symbols and the noise in the observations at the receiver is  $\bar{\mathbf{n}}_{TS}$ .

The MMSE(MLE) channel estimator ([10]) is given by  $\hat{\bar{\mathbf{h}}} = \mu_{\bar{\mathbf{h}}} + \sigma_{\bar{\mathbf{h}}}^2 \mathbf{A}_{TS}^H (\sigma_{\bar{\mathbf{h}}}^2 \mathbf{A}_{TS} \mathbf{A}_{TS}^H + \sigma_n^2 I)^{-1} (\bar{\mathbf{y}}_{TS} - \mathbf{A}_{TS} \mu_{\bar{\mathbf{h}}})$ .  $(\bar{\mathbf{h}}, \hat{\bar{\mathbf{h}}})$  are jointly Gaussian with mean  $(\mu_{\bar{\mathbf{h}}}, \mu_{\bar{\mathbf{h}}})$ . The corresponding MMSE equalizer is,  $\mathbf{E}(\hat{\mathbf{H}}) = (\hat{\mathbf{H}}^H \hat{\mathbf{H}} + \sigma_n^2 I)^{-1} \hat{\mathbf{H}}^H$ .

We first compute the transition probabilities (with  $\hat{\bar{\mathbf{a}}}$  representing output of the decoder corresponding to the input

vector  $\bar{\mathbf{a}}$ ),  $\{P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_{i_i}, \mathbf{H}, \hat{\mathbf{H}}); \bar{\mathbf{s}}_i \in \mathcal{S}^m, \bar{\mathbf{s}}_{d_j} \in \mathcal{S}_d^{m_i}\}$  of the composite channel given  $\mathbf{H}, \hat{\mathbf{H}}$  and then average over all values of  $\hat{\mathbf{h}}, \hat{\mathbf{h}}$  to obtain the overall transition probabilities of the composite channel,  $\{P(\hat{\mathbf{a}}/\bar{\mathbf{a}})\}$ . It is easy to see that the transition probabilities given  $\mathbf{H}, \hat{\mathbf{H}}$  are same for all the symbols in the frame as the channel is quasi stationary and the same equalizer is being used for the entire frame. The overall transition probabilities can be estimated at the transmitter and receiver once they have channel statistics as explained below.

Let  $B_i \triangleq \{\mathbf{x} \in \mathcal{R}^{2m} : \forall l \neq i, \|\bar{\mathbf{s}}_{d_l} - \bar{\mathbf{x}}\|^2 \geq \|\bar{\mathbf{s}}_{d_i} - \bar{\mathbf{x}}\|^2\}$ . With  $\mathcal{E}$  representing expectation,  $\bar{\mathbf{x}} = \mathbf{E}(\hat{\mathbf{H}})(\mathbf{H}\bar{\mathbf{a}} + \bar{\mathbf{n}})$  and  $\bar{\mathbf{n}} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I})$ , the transition probabilities are given by,  $P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{H}, \hat{\mathbf{H}}) = \text{Prob}(\mathbf{x} \in B_i/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{H}, \hat{\mathbf{H}})$ , and  $P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j) = \mathcal{E}_{\hat{\mathbf{h}}, \hat{\mathbf{h}}} P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{H}, \hat{\mathbf{H}})$ .

The composite channel now becomes a time invariant channel with capacity  $C = \sup_{P(\bar{\mathbf{a}})} I(\hat{\mathbf{a}}, \bar{\mathbf{a}}/P(\bar{\mathbf{a}}))$  where,  $I(\hat{\mathbf{a}}, \bar{\mathbf{a}}/P(\bar{\mathbf{a}}))$  represents the mutual information with input pmf (probability mass function)  $P(\bar{\mathbf{a}})$  and transition probabilities  $\{P(\hat{\mathbf{a}}/\bar{\mathbf{a}})\}$ . The overall capacity per channel use is  $(N - N_t)C/N$ .

Since  $P(\hat{\mathbf{a}}/\bar{\mathbf{a}})$  is independent of the input pmf  $P(\bar{\mathbf{a}})$ , the mutual information  $I(\hat{\mathbf{a}}, \bar{\mathbf{a}})$  is a concave function of  $P(\bar{\mathbf{a}})$  ([5], p. 31) and hence optimization over the convex and compact set of probability mass functions  $\mathcal{P}(\mathcal{S}^m)$  (Note that  $\bar{\mathbf{a}} \in \mathcal{S}^m$ ), results in a global maximum.

#### 4. BLIND CMA EQUALIZER

The CMA cost function for single user MIMO flat fading channel with same source alphabet for all transmit antennae can be written as ([6]),

$\mathbf{E}_{CMA} = \arg \min_{\mathbf{E}=[\bar{\mathbf{e}}_1, \dots, \bar{\mathbf{e}}_m]^T} \sum_{l=1}^m \mathcal{E} (|\bar{\mathbf{e}}_l^H \bar{\mathbf{y}}(k)|^2 - R_2^2)^2$   
or equivalently (terms in the summation are positive)

$\bar{\mathbf{e}}_{cmal} = \arg \min_{\bar{\mathbf{e}}_l} \mathcal{E} (|\bar{\mathbf{e}}_l^H \bar{\mathbf{y}}(k)|^2 - R_2^2)^2, l=1, 2, \dots, m$  (2)  
where  $\bar{\mathbf{e}}_l^T$  represents the  $l^{\text{th}}$  row and  $R_2 = \mathcal{E}|\bar{\mathbf{a}}|^4 / \mathcal{E}|\bar{\mathbf{a}}|^2$ .

To obtain the above optimum, the corresponding  $m$  update equations are ( $1 \leq l \leq m$ ), (note  $\mathbf{e}_l = \text{Real}(\bar{\mathbf{e}}_l)$ .)

$$\mathbf{e}_l(k+1) = \mathbf{e}_l(k) + \mu H_{CMA}(\mathbf{e}_l(k), \text{Real}(\mathbf{H}\bar{\mathbf{a}}(k)), \bar{\mathbf{n}}(k)) \quad (3)$$

where with  $\bar{\mathbf{y}} = \bar{\mathbf{z}} + \bar{\mathbf{n}}$  ( $\bar{\mathbf{y}}$  and  $\bar{\mathbf{y}}$  are defined in notations)  
 $H_{CMA}(\mathbf{e}, \mathbf{z}, \mathbf{n}) \triangleq ((\mathbf{e}^T \bar{\mathbf{y}})^2 + (\mathbf{e}^T \bar{\mathbf{y}})^2 - R_2^2) ((\mathbf{e}^T \bar{\mathbf{y}}) \bar{\mathbf{y}} + (\mathbf{e}^T \bar{\mathbf{y}}) \bar{\mathbf{y}})$ .  
We observe that  $H_{CMA}(0, \mathbf{z}, \mathbf{n}) = 0 \forall \mathbf{z}, \mathbf{n}$ , but 0 is not a minimizer of the cost function in (2). Thus,  $\mathbf{e}_l(0)$  should not be initialized with 0 in (3).

A close look at (3) shows that all  $m$  sub cost functions are same and the different equalizers should be initialized appropriately to extract the desired source symbols. In [6] a new joint CMA algorithm is proposed that ensures that the MIMO CMA separates all the sources successfully irrespective of the initial conditions. In this work, we choose the initial condition  $\mathbf{E}_0^*$  (which will be used in all frames) such that the channel capacity is optimized. This solves initialization problem to a great extent in blind case with the original

CMA itself. The problem is solved to a greater extent in the semiblind algorithm, as here a rough estimate of the training based equalizer forms the initializer.

For real constellations like BPSK, a more suitable CMA cost function would be  $((\mathbf{e}^T \bar{\mathbf{y}})^2 - R_2^2)^2 + (\mathbf{e}^T \bar{\mathbf{y}})^4$  (in comparison to the original cost function (2)). We used this cost function for BPSK over complex channels in computing and comparing the capacity of the three equalizers.

In the next subsection, we show how analytically we can obtain the value of CMA equalizer approximately at any time  $t$  and then proceed with obtaining the channel capacity with that equalizer.

##### 4.1. CMA Equalizer approximated by ODE

When there is no ISI, the capacity achieving input distribution  $P(\bar{\mathbf{a}})$ , must be independent from symbol to symbol. Further, as the transmitter is completely unaware of the channel state, the capacity achieving input distribution will be iid (independent and identically distributed).

Each one of the  $m$  update equations in (3) is similar to the CMA update equation for SISO with ISI. Therefore it is easy to see that all the proofs in [11] for convergence of the CMA trajectory to the solution of an ODE hold (note that the input distribution would be iid). Thus the update equation (3) for any given  $\mathbf{H}$ , can be approximated by the trajectory of the ODE,

$$\dot{\mathbf{e}}_i(t) = \hat{H}_{CMA}(\mathbf{e}_i(t)) \triangleq \mathcal{E}_{\mathbf{z}}[\mathcal{E}_{\mathbf{n}}(H_{CMA}(\mathbf{e}_i(t), \mathbf{z}, \mathbf{n}))] \quad (4)$$

where  $\hat{\mathbf{z}} \triangleq \mathbf{H}\bar{\mathbf{a}}$ . The approximation can be made accurate with high probability by taking  $\mu$  small enough.

We obtain the capacity of the composite channel approximately by obtaining the capacity of the channel using the solution of the ODE as equalizer. We can solve (4) numerically and obtain the equalizer co-efficients  $\mathbf{E}(T)$  at time  $T = \mu N$  which approximates the CMA equalizer after  $N$  channel uses. These co-efficients are used for decoding of the entire block.

Given a value of  $\mathbf{H}$  equation (4) becomes [11],

$$\begin{aligned} \dot{\mathbf{e}}_l(t) = & \mathcal{E}_{\bar{\mathbf{a}}}[f(\bar{\mathbf{z}}, \mathbf{e}_l(t))] + \mathcal{E}_{\bar{\mathbf{a}}}[f(\bar{\mathbf{z}}, \mathbf{e}_l(t))] - 6\sigma_n^4 \mathbf{e}_l(t)^{(3)} \\ & + 2R_2^2 \sigma_n^2 \mathbf{e}_l(t) - 8\sigma_n^4 \|\mathbf{e}_l(t)\|^2 \mathbf{e}_l(t) - 6\sigma_n^4 \mathbf{e}_l(t)^{(2,1)} \\ & - \mathcal{E}_{\bar{\mathbf{a}}}[(\mathbf{e}_l^T(t) \bar{\mathbf{z}})(\mathbf{e}_l^T(t) \bar{\mathbf{z}})(\bar{\mathbf{z}} \bar{\mathbf{z}}^T + \bar{\mathbf{z}} \bar{\mathbf{z}}^T)] \mathbf{e}_l(t) \end{aligned} \quad (5)$$

where,  $f(\mathbf{z}, \mathbf{e}) = R_2^2 \mathbf{e}^T \mathbf{z} \mathbf{z} - (\mathbf{e}^T \mathbf{z})^3 \mathbf{z} - 4\sigma_n^2 ((\mathbf{e}^T \mathbf{z})^2 \mathbf{e} + \|\mathbf{e}\|^2 \mathbf{e}^T \mathbf{z} \mathbf{z})$ . Here  $R_{\bar{\mathbf{a}}}$  is the source covariance matrix,  $\mathbf{e}_l(t)^{(3)}$  is the vector formed by taking cube of the individual terms and  $\|\cdot\|$  represents the norm of the vector. Also  $\mathbf{e}_l(t)^{(2,1)}$  is the vector formed by taking square of the individual terms in  $\dot{\mathbf{e}}_l(t)$  and then multiplying term by term with vector  $\mathbf{e}_l(t)$ .

Define  $\mathbf{E}(\mathbf{E}_0, \mathbf{H}, P(\bar{\mathbf{a}})) \triangleq \begin{bmatrix} \text{Re}(\mathbf{E}(T)) & -\text{Im}(\mathbf{E}(T)) \\ \text{Im}(\mathbf{E}(T)) & \text{Re}(\mathbf{E}(T)) \end{bmatrix}$ , where  $\text{Re}(\mathbf{E}(T))$  and  $\text{Im}(\mathbf{E}(T))$  represent matrices formed by keeping only the real/imaginary part, respectively of, each component of the matrix  $\mathbf{E}(T)$ . Given  $\mathbf{H}, P(\bar{\mathbf{a}})$  and  $\mathbf{E}_0$ , the transitional probabilities of the approximate composite channel obtained by solving the ODE are, ( $\bar{\mathbf{y}} \triangleq \mathbf{H}\bar{\mathbf{s}}_j + \bar{\mathbf{n}}$  and  $B_i$  same in previous section)

$$P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}), \mathbf{H}) = \text{Prob}(E(\mathbf{E}_0, \mathbf{H}, P(\bar{\mathbf{a}})) \mathbf{y} \in B_i / \bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}), \mathbf{H}) \quad (6)$$

Given  $(P(\bar{\mathbf{a}}), \mathbf{E}_0)$ , the overall transition probabilities are,  $P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}})) = \mathcal{E}_{\mathbf{H}} P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}), \mathbf{H})$ .

The proofs of the following lemmas are in the Appendix.

**Lemma 1**  $P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}))$  and hence,  $I(\hat{\mathbf{a}}; \bar{\mathbf{a}}/\mathbf{E}_0, P(\bar{\mathbf{a}}))$ , is a continuous function of  $\mathbf{E}_0$  and  $P(\bar{\mathbf{a}})$ .

By Lemma 1 and by compactness of  $\mathcal{P}(\mathcal{S}^m)$ , capacity  $C(\mathbf{E}_0) \triangleq \sup_{P(\bar{\mathbf{a}}) \in \mathcal{P}(\mathcal{S}^m)} I(\hat{\mathbf{a}}; \bar{\mathbf{a}}/\mathbf{E}_0, P(\bar{\mathbf{a}}))$  of the approximate channel for a given  $\mathbf{E}_0$ , can be achieved.

**Lemma 2**  $C(\mathbf{E}_0)$  is a continuous function of  $\mathbf{E}_0$ .

By Lemma 2, once the receiver and the transmitter have knowledge of the channel statistics, one can compute  $\mathbf{E}_0^*$ , where  $\mathbf{E}_0^* = \arg \max_{\mathbf{E}_0} C(\mathbf{E}_0)$ , if it exists. Even if it does not, one can choose  $\mathbf{E}_0^*$  such that  $C(\mathbf{E}_0^*)$  is as close to  $\sup_{\mathbf{E}_0} C(\mathbf{E}_0)$  as required. Therefore, approximate Capacity of the channel with the CMA equalizer per channel use is equal to  $C_{CMA} \approx C(\mathbf{E}_0^*) = \sup_{P(\bar{\mathbf{a}})} I(\hat{\mathbf{a}}; \bar{\mathbf{a}}/\mathbf{E}_0^*)$ .

## 5. SEMI-BLIND CMA ALGORITHM

In this variant of the semiblind algorithm, we use MMSE equalizer of the training based channel estimator  $\hat{\mathbf{H}}$  obtained in section 3 as the initializer for the CMA algorithm. The equalizer co-efficients obtained from the CMA at the end of the frame are used for decoding of data for the whole frame.

Once again we use the ODE approximation of the CMA trajectory in the capacity analysis. The difference from the blind case, being that the initializer  $\mathbf{E}_0$  is given by the training based channel estimator. Now  $T = \mu(N - N_t)$ . Defining the conditional probabilities as in the blind case, and following the same steps as in Lemma 1 one can show that,

$P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_i; P(\bar{\mathbf{a}})) \triangleq \mathcal{E}_{\mathbf{H}, \hat{\mathbf{H}}} [P(\hat{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_i; P(\bar{\mathbf{a}}), \mathbf{H}, \hat{\mathbf{H}})]$  and mutual information,  $I(\hat{\mathbf{a}}; \bar{\mathbf{a}}/P(\bar{\mathbf{a}}))$ , are continuous in  $P(\bar{\mathbf{a}})$  (We also use the facts that  $\mathbf{E}(\hat{\mathbf{H}}) \neq \mathbf{0}$  with probability one and  $\mathbf{E}(\hat{\mathbf{H}})$  is continuous in  $\hat{\mathbf{H}}$ ). Therefore, the approximate capacity of the channel with the semi-blind equalizer  $C_{SB} \approx \frac{N - N_t}{N} \sup_{P(\bar{\mathbf{a}})} I(\hat{\mathbf{a}}; \bar{\mathbf{a}}/P(\bar{\mathbf{a}}))$  is achieved at some  $P(\bar{\mathbf{a}})$ .

Global maximum is guaranteed only for training algorithms. Comparing this global maximum with that of a local maximum of the blind/semiblind algorithms (as in next section) indicates that the later methods are better by at least the amount obtained.

## 6. EXAMPLES

We illustrate the theory developed with a few examples of practical interest. As a result we also draw some interesting conclusions. We compared the three equalizers over complex Gaussian channels with BPSK modulations. We set  $m=n=2$ ,  $N=64$  and  $S_{d_i}^m = S^m$  (i.e. hard decision decoding). We observed, during the experiments, that the equalizer value at the end of the frame was away from the equilibrium point in both blind/semiblind algorithms. We normalized the channel

gain to one for both the receive antennas. In Figure 2 we set  $\sigma_n^2=1$  and plotted the capacity of the three equalizers versus the transmitted power for varying K-factors (ratio of power in the mean component of the channel to that in the varying component). When K-factor is large, it is seen that there is an improvement of up to 0.88 dB (approx 20% improvement in Tx power) in semiblind algorithms at around 12dB compared to the other two algorithms. But blind and semiblind algorithms stay very close. As the K-factor approaches 0 (Rayleigh channel), it is seen that the improvement in semiblind compared to the training based diminishes but the blind becomes much worse. (Infact, in Rayleigh channel, the capacity of the blind algorithm is almost zero).

We have observed that for training and semiblind equalizers, the capacity increases with the number of training symbols, reaches a maximum and starts decreasing. From this, one can estimate the 'optimal' number of training symbols,  $N_t^*$  (see Table 1 for some examples).

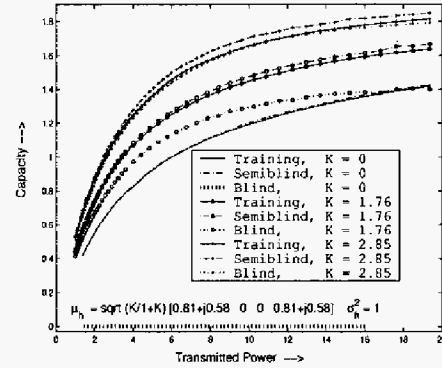


Fig. 2. Capacity versus transmitted power

Table 1.  $\mu_h = \sqrt{K/(1+K)}$  [0.995 0.0995 0.0995 0.995].  $E_A=16.0$

$\sigma_n^2$	K=5			K=1.5		
	Training (C. $N_t^*$ )	Semi (C. $N_t^*$ )	Blind C	Training (C. $N_t^*$ )	Semi (C. $N_t^*$ )	Blind C
1	(1.746, 2)	(1.781, 2)	1.725	(1.535, 5)	(1.540, 3)	0.913
1.04	(1.741, 2)	(1.777, 2)	1.719	(1.525, 5)	(1.535, 3)	0.911
1.21	(1.705, 2)	(1.750, 2)	1.692	(1.483, 5)	(1.594, 5)	0.900
2	(1.540, 2)	(1.596, 2)	1.562	(1.329, 5)	(1.343, 5)	0.844
4	(1.245, 3)	(1.297, 2)	1.287	(1.067, 5)	(1.087, 5)	0.720

Table 1 also shows comparison of various equalizers with respect to noise variance for two different K factors, for a fixed transmit power  $E_A=16$ . These channels also have non zero mean for ICI. For high K systems as noise variance decreases, the difference between the performance of training and blind algorithms decreases. At high noise variances the blind algorithm becomes infact the best. This is because the other algorithms are loosing in capacity because of training symbols. But we have seen that at very low SNR's (below 0dB, not shown in the table) eventually the blind algorithm becomes worse. As commented in section 4, we observed in case of high K, the blind algorithm was successfully separating all the sources (which also explains comparatively good performance of blind algorithms at high K). But for low K

values, we noticed that for some cases, the blind algorithm was not separating the sources resulting in bad performance (K = 1.5 case in Table 1). Also from Table 1, we can see an SNR improvement of 20% (4%) at noise variance 1 for K=5 (K=1.5) in semiblind over the training based method.

## 7. CONCLUSIONS AND FUTURE WORK

We compared blind/semiblind source separation algorithms with training based schemes. The difficulty is in comparing the loss in accuracy of the blind algorithms with that of loss in data rate in training based methods. Information capacity is the most appropriate measure for doing this performance evaluation. Using this capacity analysis, we could see that the semiblind methods perform superior to training as well as blind methods in LOS conditions ( $\approx 20\%$  improvement in transmit power) even when they have not converged to the equilibrium point. But under mean 0 conditions the improvement in semiblind over training based is negligible and the blind methods become completely useless. Thus considering the extra complexity and delays involved one may simply use the training based methods. This method could also be used to obtain the optimum number of training symbols.

We have extended this work to frequency selective channels (ISI channels). Our preliminary results show much more improvement in LOS conditions. One can also try extending this work to continuously varying channels.

## 8. REFERENCES

- [1] S. Adireddy, L. Tong, H. Viswanathan, "Optimal Placement of Training for Frequency-Selective Block-Fading Channels," *IEEE Trans. on Inf. Theory*, Vol. 48, 2338-2353, 2002.
- [2] P. Billingsley, "Probability and Measure", Wiley New York, 1991.
- [3] P. Billingsley, "Convergence of Probability Measures", Wiley New York, 1999.
- [4] E. A.Coddington, N. Levinson, "Theory of Ordinary Differential Equations" Tata McGRAW-HILL, New Delhi, 1991.
- [5] T. Cover, J. A. Thomas, "Elements of Information Theory", John Wiley Sons, 1991.
- [6] Z.Ding, Y. Li, "Blind Equalization and Identification", Marcel Dekker Inc, New York 2000.
- [7] G.B.Giannakis, Y.Hua, P.Stoica; L.Tong, "Signal Processing Advances in Wireless and Mobile Communications", Trends in Channel estimation and equalization. Vol.1, Prentice Hall, Upper Saddle River, NJ 2000.
- [8] B. Hassibi, B.M. Hochwald, "How much training is needed in Multiple-Antenna wireless links? *IEEE Trans. on Inf. Theory*, Vol. 49, 951-963, 2003.
- [9] J. Jacod, P. Protter, "Probability Essentials", Springer-Verlag Berlin Heidelberg 2000.
- [10] S. M. Kay, "Fundamentals of Statistical Signal Processing ESTIMATION THEORY", Prentice Hall, New Jersey, 1993.
- [11] V.Sharma, Naveen Raj V, "Convergence and Performance Analysis of Godard family and Multi Modulus Algorithms for Blind Equalization", to appear in *IEEE Trans. on Signal Proc.*
- [12] R. K. Sundaram, "A First Course in Optimization Theory", Cambridge University Press, Cambridge, 1996.

## APPENDIX

**Lemma 3**  $E(T)$  and hence  $E(\mathbf{E}_0, \mathbf{H}, P(\bar{\mathbf{a}}))$  is a continuously differentiable function of  $\mathbf{E}_0, \mathbf{H}$  and  $P(\bar{\mathbf{a}})$ .

**Proof of Lemma3** : It suffices to show the result for every individual row of  $\mathbf{E}(T)$ , which are obtained from ODE(5). Since the result is independent of the number of row,  $l$ , we omit  $l$  for ease of notation.

From ODE (5), we observe that  $\hat{H}_{CMA}(\mathbf{e}; P(\bar{\mathbf{a}}), \mathbf{H})$  is a continuously differentiable function of  $\mathbf{e}, P(\bar{\mathbf{a}}), \mathbf{H}$ . It then satisfies uniform Lipschitz condition with respect to  $\mathbf{e}, P(\bar{\mathbf{a}}), \mathbf{H}$  in any compact domain. The required  $C^1$  property, now follows for local solution from theorem 7.5 of ([4], p. 30) and hence for the global solution. ■

**Proof of Lemma1**: Let  $(\mathbf{E}_{0n}, P(\bar{\mathbf{a}})_n) \rightarrow (\mathbf{E}_0, P(\bar{\mathbf{a}}))$ . Denote  $E(\mathbf{E}_{0n}, \mathbf{H}, P(\bar{\mathbf{a}})_n)$  and  $E(\mathbf{E}_0, \mathbf{H}, P(\bar{\mathbf{a}}))$  (defined in section4) by  $E_n(\mathbf{H}), E(\mathbf{H})$  respectively. From Lemma3  $E_n(\mathbf{H}) \rightarrow E(\mathbf{H}) \forall \mathbf{H}$ .

Let  $X_{n,j}(\mathbf{H})$  denote a random variable with the distribution equal to the conditional distribution of  $E_n(\mathbf{H})\mathbf{y}$ , given that  $\bar{\mathbf{a}}=\bar{\mathbf{s}}_j$  was the transmitted vector. Here  $\bar{\mathbf{y}}=(\mathbf{H}\bar{\mathbf{a}}+\bar{\mathbf{n}})$ , corresponds to the channel output with input  $\bar{\mathbf{a}}=\bar{\mathbf{s}}_j$ . With  $\mathbf{y}_{s_j} \hat{=} Real(\mathcal{E}[\mathbf{H}\bar{\mathbf{s}}_j])$ ,  $X_{n,j}(\mathbf{H}) \sim \mathcal{N}(E_n(\mathbf{H})\mathbf{y}_{s_j}, \sigma_n^2 E_n(\mathbf{H})E_n(\mathbf{H})^T)$ . Let  $X_j(\mathbf{H})$  be defined in a similar way for  $E(\mathbf{H})\mathbf{y}$ . Then from (6),  $P(\bar{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_{0n}, P(\bar{\mathbf{a}})_n, \mathbf{H})=Prob(X_{n,j}(\mathbf{H}) \in B_i)$ .

Characteristic function of  $X_{n,j}(\mathbf{H})$  converges pointwise to that of  $X_j(\mathbf{H})$ . Thus,  $X_{n,j}(\mathbf{H}) \xrightarrow{w} X_j(\mathbf{H})$ . We will show below that  $Prob(X_j(\mathbf{H}) \in \partial B_i)=0 \forall \mathbf{H}$  ( $\partial B_i$  denotes the boundary of  $B_i$ ). Then by Portmanteau Theorem (Theorem 2.1 in p.16 [3])  $\forall \mathbf{H}$ ,  $P(\bar{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_{0n}, P(\bar{\mathbf{a}})_n, \mathbf{H}) \rightarrow P(\bar{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}), \mathbf{H})$ . Now the lemma follows by bounded convergence theorem i.e.

$$P(\bar{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_{0n}, P(\bar{\mathbf{a}})_n) \rightarrow P(\bar{\mathbf{a}}=\bar{\mathbf{s}}_{d_i}/\bar{\mathbf{a}}=\bar{\mathbf{s}}_j; \mathbf{E}_0, P(\bar{\mathbf{a}}))$$

Next we prove  $Prob(X_j(\mathbf{h}) \in \partial B_i)=0 \forall \mathbf{h}$ . For  $\forall s_{d_j} \in \mathcal{S}_d$ , with  $\bar{x}=x_1 + ix_2 \forall \mathbf{x}=(x_1, x_2) \in \mathbb{R}^2$ , define the two dimensional set  $C_{s_{d_j}}$  by  $C_{s_{d_j}} \hat{=} \{\mathbf{x} \in \mathbb{R}^2 : \|s_{d_j} - \bar{x}\| \leq \|s_{d_l} - \bar{x}\| \forall l \neq j\}$ . Let  $\bar{s}_{d_i, l}$  be the  $l^{th}$  element of  $\bar{s}_{d_i}$ . Clearly  $\prod_{l=1}^{m(N-N_i)} C_{\bar{s}_{d_i, l}} \subset B_i$ . It is proved below using contradiction that,  $B_i = \prod_{l=1}^{m(N-N_i)} C_{\bar{s}_{d_i, l}}$ .

Say  $\mathbf{x} \in B_i$ . Then  $\|\bar{x} - \bar{s}_{d_i}\| \leq \|\bar{x} - \bar{s}_{d_l}\| \forall l \neq i$ . Let  $x_k$  represent the  $k^{th}$  element of vector  $\mathbf{x}$ . Say for any  $k, x_k \in C_{s_{d_l}}$  where  $s_{d_l} \neq \bar{s}_{d_i, k}$ . Let  $\bar{s}_{d_j}$  represent the vector formed from  $\bar{s}_{d_i}$  by only replacing its  $k^{th}$  element with  $s_{d_l}$ . Then  $\mathbf{x} \in B_j$ , which will be a contradiction.

Let  $\partial C_{s_{d_j}}$  represent the boundary of set  $C_{s_{d_j}}$ . Let  $D_{i, l, k} \hat{=} \mathbb{R}^2 \forall k \neq l$  and  $D_{i, l, l} \hat{=} \partial C_{\bar{s}_{d_i, l}}$ . Then,  $\partial B_i \subset \bigcup_{l=1}^{m(N-N_i)} \prod_{k=1}^{m(N-N_i)} D_{i, l, k}$ .

Suffices to show that  $Prob(X_j(\mathbf{H}) \in \prod_{k=1}^{m(N-N_i)} D_{i, l, k})=0 \forall i, l, j, \mathbf{H}$ . With  $X_{j, k}(\mathbf{H})$  representing the  $k^{th}$  coordinate of  $X_j(\mathbf{H})$ , define  $\mathbf{x}_{j, k}(\mathbf{H}) \hat{=} Real(X_{j, k}(\mathbf{H}) + iX_{j, (k+m(N-N_i))}(\mathbf{H}))$ . The above condition is now equivalent to  $Prob(\mathbf{x}_{j, k}(\mathbf{H}) \in \partial C_{s_{d_l}})=0 \forall l, j, k, \mathbf{H}$ . Thus, it suffices to prove  $Prob(\mathbf{x}_{j, k}(\mathbf{H}) \in \partial C_{s_{d_l}, s_{d_l}})=0 \forall i \neq l, j, k, \mathbf{H}$ , where ( $\bar{x}$  is defined for every  $\mathbf{x}$  as in the above para)

$$\partial C_{s_{d_l}, s_{d_l}} \hat{=} \{\mathbf{x} \in \mathbb{R}^2 : \|s_{d_l} - \bar{x}\| = \|s_{l_1} - \bar{x}\|\} = \left\{ \mathbf{x} : \bar{x} = \frac{\|s_{d_l}\|^2 - \|s_{d_l}\|^2}{2(s_{d_l} - s_{d_l})} \right\}$$

The above probability equals 0 when  $\mathbf{x}_{j, k}(\mathbf{H})$ , the 2 dimensional Gaussian random vector, has nonzero variance i.e. when none of the rows of  $E(\mathbf{H})$  equal  $\mathbf{0}$ . For every value of  $\mathbf{h}$ ,  $\nabla \hat{H}_{CMA}(\mathbf{0})$  is a positive definite matrix as seen from equation (5). Hence  $\mathbf{0}$  is a repeller of the ODE(5). As mentioned earlier,  $\mathbf{0}$  is not taken as an initial condition, and therefore, none of the rows of  $E(\mathbf{H})$  equal  $\mathbf{0}$ . ■

**Proof of Lemma2**:  $I(\bar{\mathbf{a}}; \bar{\mathbf{a}}/E_0, P(\bar{\mathbf{a}}))$  is continuous function of  $E_0$  and  $P(\bar{\mathbf{a}})$ . For every  $E_0$ , the constraint set  $D(E_0)=P(S^m)$  is compact. Thus the correspondence  $E_0 \mapsto D(E_0)$  is compact and constant and hence continuous. Therefore, the required continuity follows from Maximum Theorem ([12] p. 235) ■