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 answer the following questions for a given channel:**

i. What is the optimum training sequence size in a training based method ?

ii. How do the training based, blind (considering Constant modulus algorithm, CMA, specifically) and semiblind methods compare with each other ?

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

1. INTRODUCTION

Wireless channels are necessary for ubiquitous connectivity. However, due to time varying multipath fading, broadcast nature and limited power and bandwidth, it is important to optimize the wireless resources. Therefore, multiple transmit and receive antennas, adaptive power control, modulation and coding are employed to increase the transmission rate and reduce the bit error rate [14]. An important component in the success of these adaptive techniques 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 ([13]) 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 [13] 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. To-wards 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. 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.

Consider a frame involving N channel uses with N_t training symbols. In a training based method, the channel is estimated from these symbols and then the information symbols are decoded. Using the probability of error provided by this method one can compute the capacity C of the composite channel per channel use. Then the overall channel capacity per channel use in the frame will be $(N - N_t)C/N$. In a blind algorithm (say CMA) $N_t = 0$ but using the general statistics of the arrival process one estimates the channel (or may directly obtain an equalizer) as the information symbols arrive. After all the N symbols of a frame arrive, we obtain an equalizer. We use this equalizer to estimate the transmitted information symbols of the frame. The resulting probability of error can be used to obtain the capacity of the composite channel C_b, which will also be the overall channel capacity per channel use in this case. Comparing $(N - N_t)C/N$ with C_b provides a reasonable comparison. One can compare these capacities with the capacity obtained by a semi-blind method also.

This problem for the training based methods has also been studied in [11] and [12]. They obtain a lower bound on the channel capacity and find the optimal training sequence length (and also placement in case of [11]). We not only study the problem of optimal training sequence length, but also compare it to blind and semi blind methods. Also, we actually obtain the channel capacity and compare the different methods.

The paper is organized as follows. Section **2** describes our model and the approach to be followed for comparing different methods. Section **3** considers the training based channel estimation and section **4** studies the blind algorithm (CMA) while section **5** combines the two approaches to obtain a semiblind algorithm. Section **6** compares the capacity of the three algorithms using few examples and Section **7** concludes the paper.

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

2. THE MODEL AND OUR APPROACH

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 symbols are chosen from a finite alphabet $\mathcal{S} = \{s_1, s_2, \cdots, s_L\}$. At time k, vector $\bar{A}(k) \varepsilon \mathcal{S}^m$ is transmitted from the m transmit antennas. We represent the elements of $\mathcal{S}^m$ by $\mathcal{S}^m = \{\bar{S}_i; 1 \leq i \leq L^m\}$ i.e each $\bar{S}_i$ is a m length complex vector formed from elements of $\mathcal{S}$. The channel gain matrix 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

$$\bar{Y}(k) = H\bar{A}(k) + \bar{N}(k) \;\; \forall 1 \leq k \leq N \tag{1}$$

where $\bar{N}(k)$ is an iid sequence of complex Gaussian vectors with mean 0 and co-variance $\sigma_n^2 I$ (denoted by $C\mathcal{N}(0, \sigma_n^2 I)$). We also assume that $H \sim C\mathcal{N}(\mu_H, \sigma_h^2 I)$: 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. (1) also implies that there is inter-channel interference but no inter-symbol interference (ISI). This results from a flat fading channel.

The channel statistics is available at the transmitter and the receiver but the actual channel gain matrix H is not known to the receiver and the transmitter. The receiver tries to estimate 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 H (say via the maximum likelihood estimation (MLE)). 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. Further we expect that by combining some of the blind methods with the training based 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 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 $(N - N_t)/N \; C$. 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) ([13]) has been one of the most used and successful algorithms, we will consider CMA. We assume perfect carrier phase estimation to resolve the phase ambiguity left after CMA algorithm. We will obtain the channel capacity of the composite channel corresponding to this system using the results in [1]. 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) [2] (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.

Define $h = vect(H), \mu_h = vect(\mu_H)$ and $\hat{h} = vect(\hat{H})$, where $vect(H)$ denotes the complex matrix H in real vector form by concatenating the real parts of all columns one after the other and then concatenating the imaginary parts. If $\bar{A}$ is a complex vector then $vect(\bar{A})$ once again represents the corresponding real vector and we use A to represent this vector. That is $A = vect(\bar{A})$. This notation is used throughout the paper, i.e if a letter e represents real vector of size $2n$ then $\bar{e}$ will represent the corresponding complex vector of size n.

Let $\bar{Y}_{TS} \stackrel{\triangle}{=} A_{TS}\bar{h} + \bar{N}_{TS}$ represent the nN_t length received data corresponding to all mN_t training symbols. Here 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 N_{TS}.

The MMSE(MLE) channel estimator ([2]) is given by $\bar{\mu}_h + \sigma_h^2 A_{TS}^H(\sigma_h^2 A_{TS}A_{TS}^H + \sigma_n^2 I)^{-1}(\bar{Y}_{TS} - A_{TS}\bar{\mu}_h)$ and form real vector $\hat{h}$ from the above estimator. $(h, \hat{h})$ are jointly Gaussian with mean (μ_h, μ_h). Here, as explained before, $\bar{\mu}_h$ is a complex vector such that $\mu_h = vect(\bar{\mu}_h)$. We use ^{H}, ^{HT} to denote hermition, complex conjugate respectively.

Given $\hat{h}$ and hence $\hat{H}$, the MMSE equalizer is given by, $E(\hat{H}) = \sqrt{\frac{m}{E_A}} \left(\hat{H}^H\hat{H} + \frac{m}{E_A}\sigma_n^2 I\right)^{-1} \hat{H}^H$, where E_A is the energy per symbol.

We first compute the transition probabilities (with $\hat{\bar{A}}$ representing output of the decoder corresponding to the input vector $\bar{A}$), $\{P(\hat{\bar{A}} = \bar{S}_j/\bar{A} = \bar{S}_i, H, \hat{H}); \bar{S}_i, \bar{S}_j \in \mathcal{S}^m\}$ of the composite channel given $H, \hat{H}$ and then average over all values of $h, \hat{h}$ to obtain the overall transition probabilities of the composite channel, $\{P(\widehat{\bar{A}}/\bar{A})\}$. It is easy to see that the transition probabilities given $H, \hat{H}$ are same for all the symbols in the frame as the channel is quasi stationary and the same equalizer 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 $B_i \triangleq \{Y \in R^{2n} : \cap_{l \neq i} \|\bar{S}_l - E(\hat{H})\bar{Y}\|^2 \geq \|\bar{S}_i - E(\hat{H})\bar{Y}\|^2\}$.

Then transition probabilities are given (with $\bar{Y} = H\bar{S}_j + \bar{N}$, $\bar{N} \sim C\mathcal{N}(0, \sigma_n^2 I)$ and $\mathcal{E}$ representing expectation) by,

$P(\widehat{\bar{A}} = \bar{S}_i / \bar{A} = \bar{S}_j; H, \hat{H}) = Prob(Y \in B_i)$, and

$P(\widehat{\bar{A}} = \bar{S}_i / \bar{A} = \bar{S}_j) = \mathcal{E}_{h,\hat{h}} P(\widehat{\bar{A}} = \bar{S}_i / \bar{A} = \bar{S}_j; H, \hat{H})$.

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

Since $P(\widehat{\bar{A}}/\bar{A})$ is independent of the input pmf $P(\bar{A})$, the mutual information $I(\widehat{\bar{A}}, \bar{A})$ is a concave function of $P(\bar{A})$ ([5], p. 31) and hence optimization over the convex set of probability mass functions results in a global maximum.

4. BLIND CMA EQUALIZER

The CMA Equalizer for single user MIMO flat fading channel with same source alphabet for all transmit antennae can be written as ([3]),

$$E_{CMA} = \arg\min_{E=(\bar{e}_1,\ldots,\bar{e}_m)} \Sigma_{l=1}^m \mathcal{E}\left(|\bar{e}_l \bar{Y}(k)^{HT}|^2 - R_2^2\right)^2$$

or equivalently (terms in the summation are positive)

$$\bar{e}_{cma_l} = \arg\min_{\bar{e}_l} \mathcal{E}\left(|\bar{e}_l \bar{Y}(k)^{HT}|^2 - R_2^2\right)^2, \; l = 1, 2, \ldots m,$$

where $\bar{e}_l$ represents the l^{th} row and $R_2 = \mathcal{E}|\bar{A}|^4 / \mathcal{E}|\bar{A}|^2$.

To obtain the above optimum, the corresponding m update equations are ($1 \leq l \leq m$) , ($e_l = vect(\bar{e}_l)$, contains all the real parts first and then the imaginary parts, similar to the definition for a column vector.)

$$e_l(k+1) = e_l(k) + \mu H_{CMA}\left(e_l(k), vect(H\bar{A}(k)), N(k)\right) \tag{2}$$

where with $Y = Z + N$, and vector $\check{Y} \stackrel{def}{=} vect(i\bar{Y})$ ($i = \sqrt{-1}$), $H_{CMA}(e, Z, N) \triangleq -((eY)^2 + (e\check{Y})^2 - R_2^2)\left((eY)Y^T + (e\check{Y})\check{Y}^T\right)$

A close look at (2) shows that all m sub cost functions are same and the different equalizers should be initialized appropriately to extract the desired source symbols. In this work, we choose the initial condition E_0^* (which will be used in all frames) such that the channel capacity is optimized. In [3] a new joint CMA algorithm is proposed that ensures that the MIMO CMA separates all the sources successfully irrespective of the initial conditions. In future we will analyze this algorithm using our analysis.

In the next subsection, we show how analytically we can obtain the value of CMA equalizer 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 $P(\bar{A})$, must be independent from symbol to symbol (One symbol means the entire $m \times 1$ complex vector transmitted at the same time instant). Further, as the transmitter and receiver is completely unaware of the channel state, capacity achieving input distribution will be iid (independent and identically distributed).

Each one of the m update equations in (2) is similar to the CMA update equation for SISO with ISI. Therefore it is easy to that all the proofs in [1] 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(2) can be approximated by the trajectory of the ODE,

$$\dot{e}_l(t) = \hat{H}_{CMA}(e_l(t)) \stackrel{\triangle}{=} \mathcal{E}_Z[\mathcal{E}_N(H_{CMA}(e_l, Z, N))] \tag{3}$$

where $\bar{Z} \stackrel{\triangle}{=} H\bar{A}$. The approximation can be made accurate with high probability by taking μ 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 (3) numerically and obtain the equalizer co-efficients $E(T)$ at time $T = \mu N$ (μ is the step size) which approximates the CMA equalizer after N channel uses. These co-efficients are used for ML decoding of the entire block. We will show that the transitional probabilities of the approximate composite channel will be a continuous function of pmf $P(\bar{A})$ and the common initial equalizer setting E_0. Also it is easy to see that, $E(T)$ can be computed at the transmitter and the receiver once the original channel statistics are known.

Given a value of H, with $\bar{Z} := H\bar{A}$, $\check{Z} := vect(i\bar{Z})$, $\mathcal{I} := \sigma_n^{-2}\mathcal{E}_N(N\check{N}^T)$ (which is permutation of identity matrix), the equation (3) becomes [1],

$$\begin{aligned}
\dot{e}_l(t) &= \\
&-\mathcal{E}_{\bar{A}}[(e_l(t)Z)^3 Z^T] + R_2^2\mathcal{E}_{\bar{A}}(ZZ^T)e_l(t) + R_2^2\sigma_n^2 e_l(t) \\
&-3\sigma_n^2(e_l(t)\mathcal{E}_{\bar{A}}(ZZ^T)e_l(t)^T)e_l(t) - 3\sigma_n^4 e_l(t)^{(3)} \\
&-3\sigma_n^2\|e_l(t)\|^2 e_l(t)\mathcal{E}_{\bar{A}}(ZZ^T) - 3\sigma_n^4\|e_l(t)\|^2 e_l(t) \\
&-\mathcal{E}_{\bar{A}}[(e_l(t)\check{Z})^3 \check{Z}^T] + R_2^2\mathcal{E}_{\bar{A}}(\check{Z}\check{Z}^T)e_l(t) + R_2^2\sigma_n^2 e_l(t) \\
&-3\sigma_n^2(e_l(t)\mathcal{E}_{\bar{A}}(\check{Z}\check{Z}^T)e_l(t)^T)e_l(t) - 3\sigma_n^4 e_l(t)^{(3)} \\
&-3\sigma_n^2\|e_l(t)\|^2 e_l(t)\mathcal{E}_{\bar{A}}(\check{Z}\check{Z}^T) - 3\sigma_n^4\|e_l(t)\|^2 e_l(t) \\
&-\mathcal{E}_{\bar{A}}[(e_l(t)Z)^2(e_l(t)\check{Z})\check{Z}^T] - \sigma_n^2\|e_l(t)\|^2 e_l(t)\mathcal{E}_{\bar{A}}(\check{Z}\check{Z}^T) \\
&-2\sigma_n^2\mathcal{E}_{\bar{A}}(e_l(t)Ze_l(t)\check{Z})e_l(t)\mathcal{I} - \sigma_n^2 e_l(t)\mathcal{E}_{\bar{A}}(e_l(t)Z)^2 \\
&-2\sigma_n^2 e_l(t)\mathcal{I}e_l(t)^T e_l(t)\mathcal{E}_{\bar{A}}\left(Z\check{Z}^T\right) - 2\sigma_n^4\|e_l(t)\|^2 e_l(t) \\
&-3\sigma_n^4(e_l(t))^{(2.*1)} - 3\sigma_n^4(e_l(t))^{(1.*2)} \\
&-\mathcal{E}_{\bar{A}}[(e_l(t)\check{Z})^2(e_l(t)Z)Z^T] - \sigma_n^2\|e_l(t)\|^2 e_l(t)\mathcal{E}_{\bar{A}}(ZZ^T) \\
&-2\sigma_n^2\mathcal{E}_{\bar{A}}(e_l(t)\check{Z}e_l(t)Z)e_l(t)\mathcal{I} - \sigma_n^2 e_l(t)\mathcal{E}_{\bar{A}}(e_l(t)\check{Z})^2 \\
&-2\sigma_n^2 e_l(t)\mathcal{I}e_l(t)^T e_l(t)\mathcal{E}_{\bar{A}}\left(\check{Z}Z^T\right) - 2\sigma_n^4\|e_l(t)\|^2 e_l(t) \\
&-3\sigma_n^4(e_l(t))^{(2.*1)} - 3\sigma_n^4(e_l(t))^{(1.*2)}
\end{aligned} \tag{4}$$

Here $R_{\bar{A}}$ is the source covariance matrix, $e_l(t)^{(3)}$ is the vector formed by taking cube of the individual terms, $\|.\|$ represents the norm of the vector. $(e_l(t))^{(1.*2)}$ is the vector formed by taking square of the individual terms in $\check{e}_l(t)$ and then multiplying term by term with vector $e_l(t)$. $(e_l(t))^{(2.*1)}$ is defined in a similar way.

It is clear to see that $E(T)$ is a function of H, E_0 and $P(\bar{A})$. Define $E(E_0, H, P(\bar{A})) \stackrel{def}{=} \begin{bmatrix} Re(E(T)) & -Im(E(T)) \\ Im(E(T)) & Re(E(T)) \end{bmatrix}$, where $Re(), Im()$ represent the real and imaginary parts respectively. Next we prove a few properties which are useful in numerically computing the channel capacity for this system.

Lemma 1 *$E(E_0, H, P(\bar{A}))$ is a continuously differentiable function of E_0, H and $P(\bar{A})$.*

Proof : Please refer to the Appendix.

Given $H, P(\bar{A})$ and E_0, the transitional probabilities of the approximate composite channel obtained by solving the ODE are, (with $\bar{Y} \stackrel{def}{=} H\bar{S}_j + \bar{N}$)

$$P(\widehat{\bar{A}}{=}\bar{S}_i/\bar{A}{=}\bar{S}_j;E_0, P(\bar{A}), H){=}P_{rob}(E(E_0,H,P(\bar{A}))Y{\in}B_i) \quad (5)$$

where $B_i \stackrel{\triangle}{=} \{X \in R^{2m} : \cap_{l\neq i}\|\bar{S}_l - \bar{X}\|^2 \geq \|\bar{S}_i - \bar{X}\|^2\}$. For any given $(P(\bar{A}), E_0)$, the transition probabilities are, $P(\widehat{\bar{A}}{=}\bar{S}_i/\bar{A}{=}\bar{S}_j;E_0, P(\bar{A})){=}\mathcal{E}_h P(\widehat{\bar{A}}{=}\bar{S}_i/\bar{A}{=}\bar{S}_j;E_0,P(\bar{A}),H)$. The following two lemmas prove the continuity of these overall transition probabilities.

Lemma 2 *For any given $(E_0, P(\bar{A}))$, $E(E_0, H, P(\bar{A}))$ is full rank for almost all H.*

Proof : Please refer to the Appendix.

Lemma 3 *$P(\widehat{\bar{A}} = \bar{S}_i/\bar{A} = \bar{S}_j; E_0, P(\bar{A}))$ is a continuous function of E_0 and $P(\bar{A})$.*

Proof: Please refer to the Appendix.

From Lemma 3, one also obtains that the conditional mutual information $I(\widehat{\bar{A}}; \bar{A}/E_0, P(\bar{A}))$ between $\bar{A}$ and $\widehat{\bar{A}}$ is a continuous function of $(E_0, P(\bar{A}))$. Also, $\mathcal{P}(\bar{A})$, the set of probability mass functions on $\bar{A}$ is compact. (Note that $\bar{A} \in \mathcal{S}^m$, but we denote the set of probability mass functions by $\mathcal{P}(\bar{A})$).

Thus Capacity $C(E_0) := \sup_{P(\bar{A})\in\mathcal{P}(\bar{A})} I(\widehat{\bar{A}}; \bar{A}/E_0)$ of the approximate channel for a given E_0, can be achieved. Note that the approximate composite channel for a given E_0, is a discrete memoryless channel as in the case of the training based equalizer.

Lemma 4 *$C(E_0)$ is a continuous function of E_0.*

Proof : Please refer to the Appendix.

When the receiver and the transmitter have the knowledge of channel statistics, one can compute ${E_0}^*$, where $E_0^* := \arg\max_{E_0 \in C^{m\times n}} C(E_0)$, if it exists. Even if it does not, one can choose E_0^* such that $C(E_0^*)$ is as close to $\sup_{E_0} C(E_0)$ as required. Therefore, approximate Capacity of the channel with the CMA equalizer per channel use is equal to

$$C_{CMA} \approx C(E_0^*) = \sup_{P(\bar{A})} I(\widehat{\bar{A}}; \bar{A}/E_0^*)$$

and hence the capacity for the whole frame is $= NC_{CMA}$.

5. SEMI-BLIND CMA ALGORITHM

In this section we consider a system where the training data of length N_t is placed in the beginning of the frame (it is not really required, we assume it for the convenience). After the training data we use CMA algorithm to further improve the system performance. We use MMSE equalizer of the training based channel estimator $\hat{H}$ 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 E_0 is given by the training based channel estimator. Now $T = \mu(N - N_t)$.

As seen in section **3** $(h, \hat{h})$ are jointly Gaussian with mean (μ_h, μ_h). The initializer for the CMA, $E_0(\hat{H}) = \sqrt{\frac{m}{E_A}} \left(\hat{H}^H\hat{H} + \frac{m}{E_A}\sigma_n^2 I\right)^{-1} \hat{H}^H$ is a continuously differentiable function of $\hat{H}$.

By Lemma 1 $E(E_0(\hat{H}), H, P(\bar{A}))$ (The matrix now corresponds to time $T = \mu(N - N_t)$) is continuously differentiable in $H, P(\bar{A})$ and $\hat{E}_0(\hat{H})$ and hence also in $\hat{H}$.

Then by a small modification of the proof in Lemma 2 we can show that for a given $P(\bar{A})$, $E(E_0(\hat{H}), H, P(\bar{A}))$ is full rank for almost all H and $\hat{H}$.

Once again, if the transmitter and receiver have the information about the statistics of the channel, they can know the averaged conditional probabilities and hence, the channel capacity can be computed as in blind CMA.

Defining the conditional probabilities as in the blind case, and following the same steps as in Lemma 3 one can see that,

$$P(\widehat{\bar{A}} = \bar{S}_j/\bar{A} = \bar{S}_i; P(\bar{A}))$$
$$:= \mathcal{E}_{h,\hat{h}} P(\widehat{\bar{A}} = \bar{S}_j/\bar{A} = \bar{S}_i; P(\bar{A}), H, \hat{H})$$

is a continuous function of $P(\bar{A})$. Thus the mutual information $I(\widehat{\bar{A}}; \bar{A}/P(\bar{A}))$ is a continuous function of $P(\bar{A})$. Therefore by compactness of $\mathcal{P}(\bar{A})$, $\sup_{P(\bar{A})\in\mathcal{P}(\bar{A})} \left(I(\widehat{\bar{A}}; \bar{A}/P(\bar{A}))\right)$ is achieved at some $P(\bar{A})^*$. Thus the approximate capacity of the channel with the semiblind equalizer is

$$C_{SB} \approx \frac{N-N_t}{N} I(\widehat{\bar{A}}; \bar{A}/P(\bar{A})^*).$$

Having obtained the capacity of the channel with training based, blind and semiblind methods, one can compare them for any MIMO wireless channel. Then one can obtain the optimal scheme (say a semiblind channel with a given length N_t of the training sequence). In the next section we carry out this comparison for a few example cases.

The usual optimization techniques will provide only a local optimum for the blind and semi blind methods. But comparing the local maxima of blind/semiblind algorithm with the global optimum of the training based method would indicate that blind/semiblind method are better at least by the amount shown in simulations.

6. SIMULATIONS

Simulations have been carried out over real Gaussian channels with BPSK modulations. We consider 2×2 MIMO channels for simulations. We fixed the frame length at 64 symbols. We follow the systematic approach explained above for calculating the capacity of the composite channel for all the equalizers. We normalized the channel gain to one for both the receive antennas and fixed noise variance to 1. In Figure 1 we have plotted capacity of the training, semi-blind and blind equalizers versus transmitted power, which in this case becomes received SNR. We varied the mean μ_h and variance σ_h^2 of the channel during our experiments.

When the channel variance is small and the mean is large (first sub figure in Figure 1), it is seen that there is an improvement of up to 0.7 db (approx 16% improvement in TX power) in semiblind algorithms. But as the mean approaches zero (the third sub figure in Figure 1 is with mean 0), it is seen that the improvement diminishes. Thus semiblind algorithms make good improvement in the presence of line of sight (LOS) ray and may not be useful for mean 0 channels.

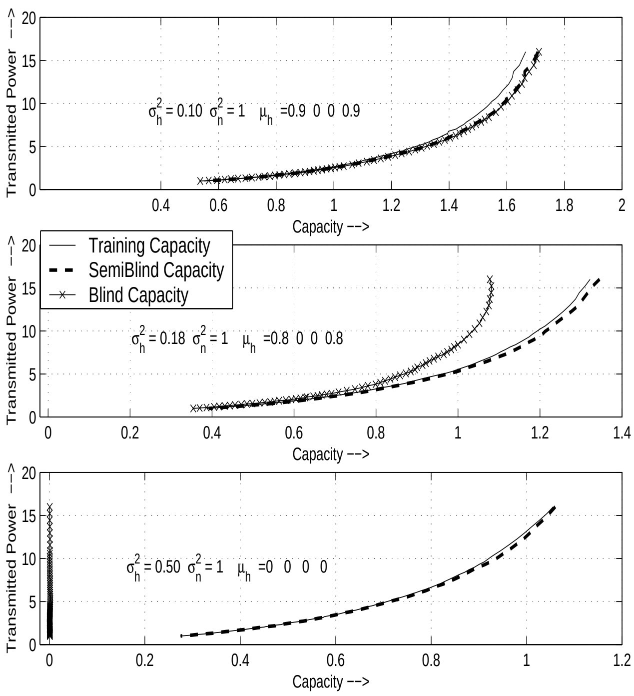

Fig. 1. Capacity versus transmitted power

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, we can estimate the 'optimal' number of training sequences (see Table 1 for some examples). It is observed that usually less than 10 symbols are sufficient for achieving the maximum capacity. It is interesting to note that the same procedure can also be used for choosing the optimal training sequences (for a given N_t). This is possible mainly because the input alphabet is finite and hence a finite number of comparisons will do the job. This might be tedious and we have not conducted experiments in this regard.

Table 1 also shows comparison of various equalizers with respect to noise variance. A substantial improvement in performance is observed for semiblind in comparison with any other equalizers in most of the cases.

Table 1. Capacity of the various equalizers for $m, n = 2$, $\mu_h = [0.85\ \ 0.01\ \ 0.01\ \ 0.85]$ and $\sigma_h^2 = 0.1387$ and $E_A = 16.0$

σ_n^2	Training (Cap, Opt N_T)	Semi Blind (Cap, Opt N_T)	Blind Capacity
3.0	(1.1117, 5)	(1.1593, 2)	1.1088
4.0	(0.9975, 2)	(1.0428, 2)	1.0127
5.0	(0.9130, 2)	(0.9458, 1)	0.9180
6.0	(0.8379, 2)	(0.8698, 2)	0.8473
6.5	(0.7945, 1)	(0.8392, 1)	0.8196
7.0	(0.7813, 1)	(0.8144, 1)	0.7914
10.0	(0.6451, 1)	(0.6710, 1)	0.6379

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 16% improvement in transmit power). But under mean 0 conditions the improvement is negligible. 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] 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.

[2] S. M. Kay, "Fundamentals of Statistical Signal Processing ESTIMATION THEORY", Prentice Hall, New Jersey, 1993.

[3] Z. Ding, Y. Li, "Blind Equalization and Identification", Marcel Dekker Inc, NewYork 2000.

[4] J. Jacod, P. Protter, "Probability Essentials", Springer-Verlag Berlin Heidelberg 2000.

[5] T. Cover, J. A. Thomas, "Elements of Information Theory", John Wiley & Sons, 1991.

[6] P. Billingsley, "Probability and Measure", Wiley NewYork, 1991.

[7] E. A.Coddington, N. Levinson, "Theory of Ordinary Differential Equations" Tata McGRAW-HILL, New Delhi, 1991.

[8] R. K. Sundaram, "A First Course in Optimization Theory", Cambridge University Press, Cambridge, 1996.

[9] M. Spivak, "Differential Geometry", Vol.1, Publish or Perish Inc, Berkeley, 1979.

[10] L. Dingjun, T. Libang "Qualitative Theory of Dynamical Systems", World Scientific, 1993.

[11] 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.

[12] 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.

[13] G.B.Giannakis, Y.Hua, P.Stoica, L.Tong, "Signal Processing Advances in Wireless and Mobile Communicaitons", Trends in Channel estimation and equalization. Vol.1, Prentice Hall, Upper Saddle River, NJ 2000.

[14] E.G.Larsson and P.Stoica, "Space-Time Block Coding for wireless Communications", Cambridge University Press, 2003.

APPENDIX

Proof of Lemma1 : It suffices to show the result for $vect(E(T)_l)$, the l^{th} row $\forall l$. Since the result is independent of the number of row, l, we omit l for ease of notations.

From ODE (4), we observe that $\hat{H}_{CMA}(e; P(\bar{A}), H)$ is a continuously differentiable function of e, $P(\bar{A})$, H. It then satisfies uniform Lipschitz condition with respect to e, $P(\bar{A})$, H in any compact domain. The required C^1 property, now follows for local solution from theorem 7.5 of ([7], p. 30) and hence for the global solution. □

Proof of Lemma2 : Fix $E_0, P(\bar{A})$. Assume the covariance matrix of h is full rank. From Lemma1, $g(H) \triangleq E(E_0, H, P(\bar{A}))$, mapping domain $C^{n\times m}$ to $R^{2m\times 2n}$, is C^1.

Let $N = \{E \in R^{2m\times 2n} : rank(E) < 2m\}$. Then $N = \cup_{i=0}^{2m-1} N_i$ where, $N_i = \{E \in R^{2m\times 2n}$: First i rows are linearly independent and $i + 1^{th}$ row is linearly dependent on the first i rows $\}$ for $i \neq 1$ and N_0 contains all matrices whose first row is all zero vector.

Consider the collection of all possible $(i + 1) \times (i + 1)$ sub matrices formed from the first $(i + 1)$ rows and define some order for the collection. With respect to each such k^{th} sub matrix, define $N_{ik} = \{E \in R^{2m\times 2n} : k^{th} (i+1) \times (i+1)\text{submatrix rank } = i\}$.

Clearly, $N_i \subset \cup_k N_{ik}$ and hence $N \subset \cup_{i=0}^{n-1} \cup_k N_{ik}$. Therefore, it suffices to show that $g^{-1}(N_{ik})$ has measure 0 for all i and k.

Using the same order as used above, define M_{ik} and a function f_{ik} with domain M_{ik} by,

$M_{ik} = \{E \in R^{2m\times 2n} : k^{th} (i + 1) \times (i + 1)\text{submatrix rank} \geq i\}$
$f_{ik}(E) = det(\text{of } k^{th} (i + 1) \times (i + 1)\text{sub matrix } of E)\ \forall E \in M_{ik}$.

M_{ik} is a open subset of $R^{2m\times 2n}$ and hence a C^∞ manifold of dimension $4nm$, N_{ik} equals $f_{ik}^{-1}(\{0\})$ and f_{ik} has constant rank 1 on M_{ik}. (For definition of rank of a function refer to p. 52 [9]).

By Proposition 12 of (p. 65 [9]) N_{ik} is an $4nm-1$ dimensional submanifold of M_{ik} and hence of $R^{2m\times 2n}$. From Theorem 1.8 of (p. 54 of [10]) $g^{-1}(N_{ik})$ is empty or has dimension $4nm - 1$ which is strictly less than the dimension of $C^{n\times m}$.

Now the lemma follows from Lemma1.3 ([10], p.47) by using the inclusion map from $g^{-1}(N_{ik})$ to $C^{n\times m}$ and the fact that the distribution of channel h is absolutely continuous with respect to the Lebesgue measure.

If the covariance matrix is not full rank, the lemma follows by restricting domain of the map g to the hyperplane on which the probability mass of the channel h is concentrated. □

Proof of Lemma3: Let $(E_{0n}, P(\bar{A})_n) \to (E_0, P(\bar{A}))$. Denote $E(E_{0n}, H, P(\bar{A})_n)$ and $E(E_0, H, P(\bar{A}))$ by $E_n(H), E(H)$ respectively. From Lemma1 $E_n(H) \to E(H)$ for all H. Also from Lemma2 $E(H)$ is invertible for all most all H.

Let $Z_{n,j}(H)$ denote a random variable with the distribution equal to the conditional distribution of $E_n(H)Y$, given $\bar{A} = \bar{S}_j$ was the transmitted vector. Here Y is the real vector corresponding to the complex channel output vector with input $\bar{A} = \bar{S}_j$. Then
$Z_{n,j}(H) \sim \mathcal{N}\left(\left(E_n(H)vect(H\bar{S}_j)\right),\ \sigma_n^2\left(E_n(H)E_n(H)^T\right)\right)$.
Let Z_j be defined in a similar way for $E(H)Y$.

Then from (5), $P(\hat{\bar{A}} = \bar{S}_i/\bar{A} = \bar{S}_j; E_{0n}, P(\bar{A})_n, H)$ equals $Prob(Z_{n,j}(H) \in B_i)$.

Since for almost all H, $E(H)$ is full rank, $E(H)E(H)^T$ is invertible and hence $Z_j(H)$ is absolutely continuous with respect to Lebesgue measure.

$E(H)$ is full rank for almost all H implies that $E_n(H)$ is full rank for all $n > N$ for some $N > 0$ for almost all H . Thus $Z_{n,j}(H)$ has density for all $n > N$ and for almost all H. For all such H, the density of $Z_{n,j}(H)$ converges pointwise to that of $Z_j(H)$ and hence by Scheffe's theorem ([6]), for all i, j,
$Prob(Z_{n,j}(H) \in B_i) \to Prob(Z_j(H) \in B_i)$. Therefore, for almost all H, $P(\hat{\bar{A}} = \bar{S}_i/\bar{A} = \bar{S}_j; E_{0n}, P(\bar{A})_n, H)$

$$\to P(\hat{\bar{A}} = \bar{S}_i/\bar{A} = \bar{S}_j; E_0, P(\bar{A}), H)$$

And, by bounded convergence theorem,
$P(\hat{\bar{A}} = \bar{S}_i/\bar{A} = \bar{S}_j; E_{0n}, P(\bar{A})_n)$

$$\to P(\hat{\bar{A}} = \bar{S}_i/\bar{A} = \bar{S}_j; E_0, P(\bar{A}))$$ □

Proof of Lemma4: $I(\hat{\bar{A}}; \bar{A}/E_0, P(\bar{A}))$ is continuous function of E_0 and $P(\bar{A})$. For every E_0, constraint set $D(E_0) = \mathcal{P}(\bar{A})$ is compact. Thus correspondence $E_0 \mapsto D(E_0)$ is compact and constant and hence continuous. Thus the required continuity follows from Maximum Theorem ([8] p. 235) □