Blind Channel Estimation in OFDM Systems by Relying on the Gaussian Assumption of the Input

T. Y. Al-Naffouri[1] and A. A. Quadeer[2]
Electrical Engineering Department
King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia
{naffouri[1], aquadeer[2]}@kfupm.edu.sa

Abstract— *In an OFDM system, the receiver requires an estimate of the channel to recover the transmitted data. Most channel estimation methods rely on some form of training which reduces the useful data rate. Here instead we blindly estimate the channel by maximizing the log likelihood of the channel given the output data. Finding the likelihood function of a linear system can be very difficult. However, in the OFDM case, central limit arguments can be used to argue that the time-domain input is Gaussian. This together with the Gaussian assumption on the noise makes the output data Gaussian. The output likelihood function can then be maximized to obtain the maximum likelihood (ML) estimate of the channel. Unfortunately, the likelihood function is not uni-modal and thus finding the global maxima is challenging. In this paper, we propose two methods to find the global maxima of the ML objective function. One is the blind Genetic algorithm and the other is the semi-blind Steepest descent method. The performance of the proposed algorithms is demonstrated by computer simulations.*

Keywords— *Gaussian assumption on data, Blind channel estimation, Semi-blind channel estimation, Maximum likelihood estimation.*

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) has emerged as a modulation scheme that can achieve high data rates by efficiently combating multipath effects. It does this by dividing the frequency selective fading channel into parallel frequency-flat channels. The additional advantages of simple receiver implementation and high spectral efficiency due to orthogonality contribute towards the increasing interest in OFDM. This is reflected by the many standards that considered and adopted OFDM, including those for digital audio and video broadcasting (DAB and DVB), WIMAX (Worldwide Interoperability for Microwave Access), high speed modems over digital subscriber lines, and local area wireless broadband standards such as the HIPER-LAN/2 and IEEE 802.11a, with data rates of up to 54 Mbps [1]. OFDM is also being considered for fourth-generation (4G) mobile wireless systems [2].

In order to achieve high data rate in OFDM, receivers must estimate the channel efficiently and subsequently the data. Many techniques have been presented by the researchers for channel estimation in OFDM systems. They can be broadly divided into three categories:

1) *Training-based estimation*: It involves sending pilots (symbols which are known to the receiver) with the data symbols so that the channel can be estimated and hence the data at the receiver (see for example [3] - [5]). Use of pilots results in decrease in bandwidth efficiency.
2) *Blind estimation*: The limitations in training based estimation techniques motivated interest in the spectrally efficient blind approach. These techniques use some inherent structure of the communication system which is produced by constraints including finite alphabet constraint [6], [10], cyclic prefix [8], [9], [10], and time and frequency correlation [7], [11]. Blind techniques are generally computationally cumbersome.
3) *Semi-blind estimation*: Semi-blind techniques make use of both pilots and the natural constraints to efficiently estimate the channel. These methods use pilots to obtain an initial channel estimate and improve the estimate by using a variety of a priori information. Thus, in addition to the pilots, semi-blind methods use the cyclic prefix [7], [9], [10], time and frequency correlation [7], [11], gaussian assumption on transmitted data [12], and virtual carriers [13] for channel estimation and subsequent data detection.

In this paper, we perform channel estimation by utilizing the Gaussian assumption on the transmitted data and the cyclic prefix. Specifically, the channel estimate is obtained by maximizing the log likelihood of the channel given the output data. Finding the likelihood function of a linear system can be very difficult. However, in the OFDM case, central limit arguments can be used to argue that the input is Gaussian [12], [14]. Under the assumption that the noise is Gaussian, this makes the output Gaussian and allows us to easily write down the likelihood expression of the output. The likelihood function can then be maximized to obtain the ML estimate of the channel.

A. Paper Organization and Notation

After introducing the notation in Table 1, we give an overview of the OFDM system in Section II. The log likeli-

This work was supported by SABIC Fast Track, an internally funded project from DSR, KFUPM (Project No. SB090008).

Table 1. Notation used in the paper

Variable	**Notation employed**
Scalars	Small-case letters (e.g. x)
Vectors	Small-case boldface letters (e.g. $\boldsymbol{x}$)
Matrices	Upper-case boldface letters (e.g. $\boldsymbol{Q}$)
Vectors in frequency domain	Calligraphic notation (e.g. $\boldsymbol{\mathcal{H}}$)
Individual entries of a vector	$h(l)$
Estimate of a variable	Hat over the variable (e.g. $\hat{\boldsymbol{\mathcal{X}}}$)
Variables as function of time	Time index appears as a subscript (e.g. $\boldsymbol{\mathcal{X}}_i$)
Cyclic prefix	Underlined vector (e.g. $\underline{\boldsymbol{x}}_i$)
Super symbol	Overlined vector (e.g. $\overline{\boldsymbol{x}}_i = [\underline{\boldsymbol{x}}_i^T\ \boldsymbol{x}_i^T]^T = [\underline{\boldsymbol{x}}_i^T\ \tilde{\boldsymbol{x}}_i^T\ \underline{\boldsymbol{x}}_i^T]^T$)

hood function is derived in Section III by using the Gaussian assumption on transmitted data. The channel can be estimated by maximizing the likelihood function. For blind estimation, the likelihood function should have global maxima when it is plotted against the channel taps. Experimental results of log likelihood function plot against channel taps are discussed in Section III-B which show that it is multi-modal. To solve this problem, two methods have been presented in this paper. One is the blind channel estimation method implemented using the Genetic algorithm (Section IV) and the second one is a semi-blind approach using steepest descent algorithm (Section V). The gradient of likelihood function involved in the steepest descent algorithm is also derived in this section. The simulation results are presented in Section VI followed by the conclusion in Section VII.

II. System Overview

In this paper, a simple OFDM system is used which involves transmitting data in symbols $\boldsymbol{\mathcal{X}}_i$ of length N each. Each symbol then undergoes an IDFT operation to produce the time domain symbol $\boldsymbol{x}_i$, i.e.

$$\boldsymbol{x}_i = \sqrt{N}\boldsymbol{Q}\boldsymbol{\mathcal{X}}_i, \tag{1}$$

where $\boldsymbol{Q}$ is an IDFT matrix of size $N \times N$. A cyclic prefix of length L is appended to form the super-symbol $\overline{\boldsymbol{x}}_i$ (Refer to Table 1 for further explanation). We assume an FIR channel of maximum length $L+1$ given by

$$\boldsymbol{h} = \begin{bmatrix} h_0 & h_1 & \cdots & h_L \end{bmatrix} \tag{2}$$

For reasons to be explained shortly, we will focus in this paper on time domain signals. Here, the input/output relationship is given by

$$\begin{bmatrix} \underline{\boldsymbol{y}}_i \\ \boldsymbol{y}_i \end{bmatrix} = \begin{bmatrix} h_L & h_{L-1} & \cdots & 0 \\ 0 & h_L & \cdots & 0 \\ & \ddots & \ddots & \\ 0 & 0 & \cdots & h_0 \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{x}}_{i-1} \\ \boldsymbol{x}_i \\ \tilde{\boldsymbol{x}}_i \\ \underline{\boldsymbol{x}}_i \end{bmatrix} + \begin{bmatrix} \underline{\boldsymbol{n}}_i \\ \boldsymbol{n}_i \end{bmatrix}$$

or in matrix form

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N} \tag{3}$$

where $\boldsymbol{n}$ is the output noise which we take to be white Gaussian. The matrices $\boldsymbol{Y}, \boldsymbol{H}$ and $\boldsymbol{X}$ are of size $(N+L)\times 1$, $(N+L)\times(N+2L)$ and $(N+2L)\times 1$, respectively.

III. Evaluating the Log Likelihood Function

To derive the likelihood function of a the output of a linear system, the input is assumed usually to be Gaussian (otherwise writing down the likelihood function can be very difficult). This is usually not true in a data communication system as the input is generated from a finite alphabet. Fortunately in an OFDM system, the time domain input can be assumed to be Gaussian by central limit theorem arguments [14]. Specifically, from equation (1),we have the element by element relationship

$$\boldsymbol{x}_i(1) = \sqrt{N}q_1\boldsymbol{\mathcal{X}}_i,\ \boldsymbol{x}_i(2) = \sqrt{N}q_2\boldsymbol{\mathcal{X}}_i, ..., \boldsymbol{x}_i(N) = \sqrt{N}q_N\boldsymbol{\mathcal{X}}_i$$

where q_j are the rows of $\boldsymbol{Q}$. In other words, this shows that $\boldsymbol{x}_i(j)$ is a large (weighted) sum of iid random variables. The validity of this assumption is evident from the histogram plot shown in Figure 1 which describes the distribution of the transmitted data $\boldsymbol{x}_i$.

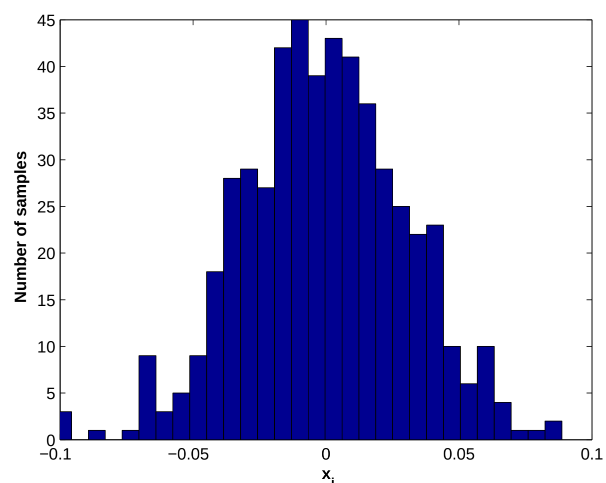

Fig. 1. Number of samples vs transmitted data ($\boldsymbol{x}_i$)

Thus from this and the fact that noise is also Gaussian, we can conclude that output $\boldsymbol{Y}$ will also be Gaussian with pdf

$$\boldsymbol{Y} \sim \boldsymbol{\mathcal{N}}(0, \boldsymbol{\Sigma}_{\mathbf{Y}})$$

where $\boldsymbol{\Sigma}_{\mathbf{Y}}$ is the second order moment of $\boldsymbol{Y}$ which (from equation (3)) is given by

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{Y}} &= E\left[\boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{H}^T\right] + \sigma_n^2\boldsymbol{I} && (4) \\ &= \boldsymbol{H}\boldsymbol{\Sigma}_{\mathbf{X}}\boldsymbol{H}^T + \sigma_n^2\boldsymbol{I} && (5) \end{aligned}$$

where $\mathbf{\Sigma_X}$ is a matrix of size $(N+2L)\times(N+2L)$ given by

$$\begin{aligned}
\mathbf{\Sigma_X} &= \boldsymbol{E}[\boldsymbol{X}\boldsymbol{X}^T] \\
&= \boldsymbol{E}\begin{bmatrix} \boldsymbol{x}_{i-1} \\ \boldsymbol{x}_i \\ \tilde{\boldsymbol{x}}_i \\ \boldsymbol{x}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{i-1}^T & \boldsymbol{x}_i^T & \tilde{\boldsymbol{x}}_i^T & \boldsymbol{x}_i^T \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{I}_L & 0 & 0 & 0 \\ 0 & \boldsymbol{I}_L & 0 & \boldsymbol{I}_L \\ 0 & 0 & \boldsymbol{I}_{N-L} & 0 \\ 0 & \boldsymbol{I}_L & 0 & \boldsymbol{I}_L \end{bmatrix} \qquad (6)
\end{aligned}$$

The pdf of output $\boldsymbol{Y}$ can thus be written as

$$P(\boldsymbol{Y}|\boldsymbol{h}) = \frac{1}{\det(\mathbf{\Sigma_Y})} \exp(-\boldsymbol{Y}^T\mathbf{\Sigma_Y}^{-1}\boldsymbol{Y}) \qquad (7)$$

So, the log likelihood function is given by

$$\boldsymbol{\mathcal{L}}(\boldsymbol{Y}|\boldsymbol{h}) = -\ln\det(\mathbf{\Sigma_Y}) - \boldsymbol{Y}^T\mathbf{\Sigma_Y}^{-1}\boldsymbol{Y} \qquad (8)$$

A. Maximum Likelihood Estimation of the Channel IR

We can use the likelihood function derived above to obtain the ML estimate of the channel $\boldsymbol{h}$ by maximizing it. i.e.

$$\begin{aligned}
\hat{\boldsymbol{h}}_{\mathrm{ML}} &= \max_{\boldsymbol{h}} \boldsymbol{\mathcal{L}} \\
&= \max_{\boldsymbol{h}} -\ln\det(\mathbf{\Sigma_Y}) - \boldsymbol{Y}^T\mathbf{\Sigma_Y}^{-1}\boldsymbol{Y} \qquad (9)
\end{aligned}$$

which depends only on the output data $\boldsymbol{Y}$ and the channel $\boldsymbol{h}$ (through the dependance of $\mathbf{\Sigma_Y}$ on $\boldsymbol{h}$).

This approach for channel estimation using the Gaussian input assumption is quite common in single carrier case, but has not been applied in the OFDM case. There are two disadvantages of applying it in the single carrier case [15], [16]:

- The method assumes that the input is Gaussian which is not the case in a single carrier system.
- Even if input is Gaussian, this method is usually phase blind i.e. it can only be used to identify minimum phase systems.

We avoid both of the problems in the OFDM case as the input is Gaussian by central limit theorem arguments and as the input is cyclostationary (due to the presence of the cyclic prefix)[8].

Unfortunately, as we shall show next, the likelihood function is not uni-modal (it could have local maxima) and so finding the global maxima might be challenging.

B. Plot of Likelihood Function vs Channel Taps

The likelihood function is plotted against the channel taps to investigate whether it has a global maxima. The input data is considered to be Gaussian of length $N = 64$ and a cyclic prefix of length $L = 2$ is used. Channel is considered to be an FIR of length $L+1 = 3$ with first tap fixed at 1 to avoid sign

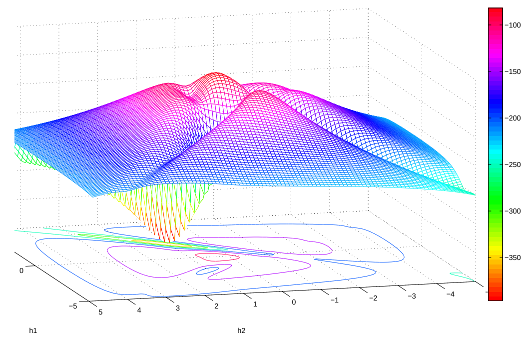

Fig. 2. 3D plot of likelihood function against channel taps with $\sigma_n^2 = 0.1$

ambiguity inherent in all blind techniques[1]. In Figure 2 the log likelihood function is plotted against the remaining two channel taps h_1 and h_2 when $\sigma_n^2 = 0.1$.

This shows that a completely blind approach for channel estimation would be challenging. In what follows, we present two approaches to solve this channel estimation problem.

IV. Blind Estimation using Genetic Algorithm

Genetic algorithm (GA) is an iterative stochastic search algorithm which was introduced by Holland [17] in 1975. GA finds the best solution in a population of candidate solutions (called chromosomes) based on natural selection (survival of the fittest) and evolution. Each chromosome has a fitness value associated with it which in our algorithm is found by evaluating the likelihood function in equation (8). The next generation is produced by using genetic operators like mutation and crossover. As the channel IR is composed of real values, real coded genetic algorithm [18] has been implemented. As the likelihood function is multi-modal (described in Subsection III-B), we have employed GA here for blind channel estimation due to its ability to avoid local maxima/minima. The algorithms used in the main operators in the reproduction process (mutation and crossover) are as follows:

A. Crossover

Crossover is the most important component of a GA. It is a method of combining the features of two parent chromosomes to form two offspring. There are many crossover algorithms present in literature but we selected the *BLX-α* algorithm (with α = 0.5) due to its superior performance in real coded genetic algorithms [18].

B. Mutation

Mutation is a method in which an arbitrary element of a selected chromosome is altered to prevent the premature convergence of GA to suboptimal solutions. GA is able to avoid

[1] A channel with only two effective taps is chosen so that we can plot the likelihood function against them in three dimensions.

local minima/maxima due to this operator. Like crossover, researchers have presented many mutation operators. We have used the *Non-uniform mutation* algorithm as it is very appropriate for real coded genetic algorithms [18].

V. Semi-blind Estimation Using Steepest Descent Algorithm

A semi-blind approach can also be pursued where we use a few pilots to obtain an initial rough estimate of the channel and subsequently improve the channel estimate using the Steepest Descent algorithm i.e.

$$\boldsymbol{h}^T_{(k+1)} = \boldsymbol{h}^T_{(k)} - \mu \, \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{h}_{(k)}} \tag{10}$$

where μ is called step size which is a small scalar value and $\boldsymbol{h}_{(k)}$ represents the estimate of channel $\boldsymbol{h}$ at k^{th} iteration. The algorithm continues to iterate until a maximum number of iterations is reached or until a stopping threshold is crossed. It can be seen from equation (10) that it involves the gradient of the likelihood function with respect to the channel. Thus we need to evaluate this gradient to implement the steepest descent algorithm. We start by representing the channel convolution matrix $\boldsymbol{H}$ in block form.

A. Writing $\boldsymbol{H}$ in Block Form

We can write the convolution matrix in the following block form

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{B} & & & & \\ & \boldsymbol{C} & \boldsymbol{B} & & & \\ & & \ddots & \ddots & & \\ & & & \boldsymbol{C} & \boldsymbol{B} \end{bmatrix} \tag{11}$$

where

$$\boldsymbol{B} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & \cdots & 0 \\ \vdots & & \ddots & \\ h_{L-1} & h_{L-2} & \cdots & h_0 \end{bmatrix} \tag{12}$$

$$\boldsymbol{C} = \begin{bmatrix} h_L & h_{L-1} & \cdots & h_1 \\ 0 & h_L & \cdots & h_2 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & h_L \end{bmatrix} \tag{13}$$

B. Evaluating Second Order Moment of Output $\boldsymbol{Y}$

As the log likelihood function (equation (8)) involves second order moment of output $\boldsymbol{Y}$, we want to evaluate it in terms of channel $\boldsymbol{h}$ or specifically in terms of $\boldsymbol{B}$ and $\boldsymbol{C}$.

The output autocorrelation matrix $\boldsymbol{\Sigma}_{\mathbf{Y}}$ can be decomposed as

$$\boldsymbol{\Sigma}_{\mathbf{Y}} = \boldsymbol{G}\boldsymbol{G}^T + \sigma_n^2 \boldsymbol{I} \tag{14}$$

where

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{B} & 0 & \cdots & 0 \\ 0 & \boldsymbol{C} & \boldsymbol{B} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \boldsymbol{B} \\ 0 & \boldsymbol{B} & 0 & \cdots & \boldsymbol{C} \end{bmatrix}$$

The factor matrix $\boldsymbol{G}$ has the following properties

1) It is a square matrix of size $N + L$.
2) It is full rank if and only if $h_L \neq 0$

Remember that we need to differentiate the likelihood function with respect to the channel IR. Now, the likelihood function is a function of $\boldsymbol{\Sigma}_{\mathbf{Y}}$ which is a function of $\boldsymbol{G}$ (see equation (14)). The matrix $\boldsymbol{G}$ is itself a linear function of the channel IR. Specifically, we can write $\boldsymbol{G}$ as a linear combination of $L+1$ constant matrices $\boldsymbol{F}_0, \boldsymbol{F}_1, ..., \boldsymbol{F}_L$ i.e.

$$\boldsymbol{G} = \sum_{i=0}^{L} h_i \boldsymbol{F}_i \tag{15}$$

The matrix $\boldsymbol{F}_i$ is an indicator matrix, i.e. it indicates the entries of $\boldsymbol{G}$ that depend on $\boldsymbol{h}_i$. We can thus write

$$\boldsymbol{G}^T = \sum_{i=0}^{L} h_i \boldsymbol{F}_i^T \tag{16}$$

or

$$\begin{aligned} \text{vec}(\boldsymbol{G}^T) &= \sum_{i=0}^{L} h_i \, \text{vec}(\boldsymbol{F}_i^T) \\ &= \begin{bmatrix} \text{vec}(\boldsymbol{F}_0^T) & \text{vec}(\boldsymbol{F}_1^T) & \cdots & \text{vec}(\boldsymbol{F}_L^T) \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_L \end{bmatrix} \\ &= \boldsymbol{\mathcal{F}} \boldsymbol{h}^T \end{aligned} \tag{17}$$

where the vec operation transforms a matrix $\boldsymbol{G}$ into a long column vector consisting of the concatenation of the columns of $\boldsymbol{G}$. Thus,

$$\begin{aligned} \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{h}} &\triangleq \frac{\partial \, \text{vec}(\boldsymbol{G}^T)}{\partial \boldsymbol{h}} \\ &= \boldsymbol{\mathcal{F}}^T \end{aligned} \tag{18}$$

We will use this relation in evaluating the gradient of $\boldsymbol{\mathcal{L}}$ w.r.t $\boldsymbol{h}$ in the following subsection.

C. Gradient of likelihood function $\boldsymbol{\mathcal{L}}$ w.r.t channel IR $\boldsymbol{h}$

We would like to find the gradient of $\boldsymbol{\mathcal{L}}$ w.r.t $\boldsymbol{h}$. By the chain rule, we can write

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{h}} = \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{h}} \, \frac{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}{\partial \boldsymbol{G}} \, \frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}} \tag{19}$$

In carrying out the differentiation $\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}$, $\boldsymbol{\Sigma}_{\mathbf{Y}}$ is treated as a general matrix. Thus, despite the fact that $\boldsymbol{\Sigma}_{\mathbf{Y}}$ is symmetric and positive definite, we ignore this fact in obtaining $\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}$. All properties of $\boldsymbol{\Sigma}_{\mathbf{Y}}$ are captured in its relation to $\boldsymbol{G}$ and in the relation of the latter to $\boldsymbol{h}$.

We have already evaluated $\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{h}}$. Lets now evaluate $\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}$. We can show that [19]

$$\begin{aligned}
\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}} &= -\frac{\partial}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}\left(\ln\det(\boldsymbol{\Sigma}_{\mathbf{Y}}) + \boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\boldsymbol{Y}\right) \\
&= -\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-T}) - \frac{\partial}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}\,\mathrm{tr}(\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}) \\
\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}} &= -\mathrm{vec}(\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-T} - \boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-T}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-T}) \qquad (20)
\end{aligned}$$

Similarly in carrying out the differentiation $\frac{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}{\partial \boldsymbol{G}}$, we ignore the sparse structure of $\boldsymbol{G}$. The sparse structure is captured in the relation of $\boldsymbol{G}$ to the channel parameters $\boldsymbol{h}$.

$$\begin{aligned}
\frac{\partial \boldsymbol{\Sigma}_{\mathbf{Y}}}{\partial \boldsymbol{G}} &= \frac{\partial}{\partial \boldsymbol{G}}(\boldsymbol{G}\boldsymbol{G}^T + \sigma_n^2\boldsymbol{I}) \\
&= (\boldsymbol{I}\otimes\boldsymbol{G}^T) + \mathrm{K}_{\mathrm{s,m}}(\boldsymbol{G}^T\otimes\boldsymbol{I}_s) \qquad (21)
\end{aligned}$$

where the second line is obtained by the product rule, and $\otimes$ and $\mathrm{K}_{\mathrm{s,m}}$ stand for Kronecker product and Commutation matrix respectively [19]. Combining the results (20) and (21) yields

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{G}} = -2\ \mathrm{vec}\left[\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1} - \boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\right] \qquad (22)$$

where we used the property that

$$\mathrm{K}_{\mathrm{s,m}}\mathrm{vec}(\boldsymbol{A}^T) = \mathrm{vec}(\boldsymbol{A})$$

So, equation (19) can now be written as

$$\begin{aligned}
\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{h}} &= \frac{\partial \boldsymbol{G}}{\partial \boldsymbol{h}}\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{G}} \\
&= -2\boldsymbol{\mathcal{F}}^T\mathrm{vec}\left[\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1} - \boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\right] \\
\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{h}} &= -2\begin{bmatrix}
\mathrm{tr}(\boldsymbol{F}_0\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1} - \boldsymbol{F}_0\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}) \\
\mathrm{tr}(\boldsymbol{F}_1\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1} - \boldsymbol{F}_1\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}) \\
\vdots \\
\mathrm{tr}(\boldsymbol{F}_L\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1} - \boldsymbol{F}_L\boldsymbol{G}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}\boldsymbol{Y}\boldsymbol{Y}^T\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1})
\end{bmatrix}
\end{aligned}$$

which is our required gradient of size $(L+1)\times 1$. This gradient can be used in equation (10) to estimate the channel using the steepest descent algorithm.

VI. Simulation Results

We consider an OFDM system with $N = 64$ and cyclic prefix of length $L = 8$. The OFDM symbol consists of BPSK or 16-QAM symbols. The channel IR consists of 9 iid Rayleigh fading taps. The parameters used in implementing the blind approach using GA are listed in Table 2. The proposed semi-blind algorithm was run for 20 iterations in all cases.

We compare the BER performance of the proposed algorithms with the following two cases: (i) Perfectly known channel, and (ii) Channel estimated using $L+1$ pilots. The simulation results for BPSK and 16-QAM modulated data are discussed below:

Table 2. Simulation Parameters used to implement GA

Population Size	100
Number of Generations	50
Cross-over Scheme	BLX-α Cross-over ($\alpha = 0.5$)
Cross-over Probability	0.8
Mutation Scheme	Non-Uniform Mutation
Mutation Probability	0.08
Number of Elite Chromosomes	5

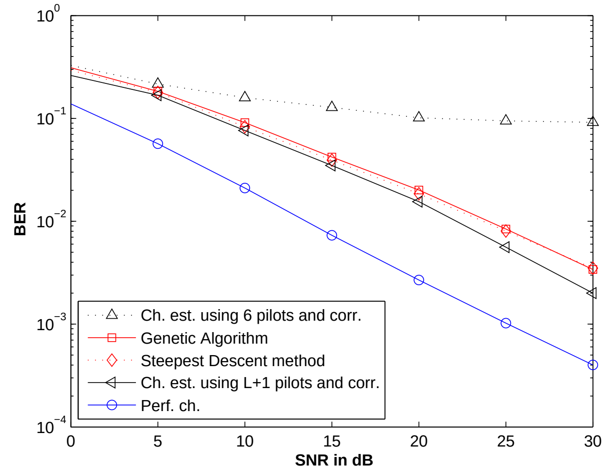

Fig. 3. BER vs SNR comparison for BPSK modulated data

A. BER vs SNR Comparison for BPSK Modulated Data

In Figure 3, the proposed algorithms are compared with the above mentioned methods when the input data is BPSK modulated. In this case, the semi-blind algorithm is initialized with an estimate obtained by using 6 pilots and channel correlation. The step size μ used in this case was 7.5×10^{-3}. This figure clearly indicates that both the proposed algorithms perform quite close to the case when channel is estimated using $L+1$ pilots.

B. BER vs SNR Comparison for 16-QAM Modulated Data

Figure 4 shows the performance of the proposed algorithms for the case of 16−QAM modulated data. Similar to the BPSK case, the semi-blind algorithm is again initialized with an estimate obtained by using 6 pilots and channel correlation. The step size μ used in this case was 2.5×10^{-4}. It can be seen from the figure that both the proposed algorithms perform quite well for the case of non-constant modulus data specially at high SNR.

VII. Conclusions

In this paper, we presented two methods for channel estimation and data recovery in OFDM transmission. It was argued in this paper that the transmitted data in OFDM is Gaussian. Thus the output is also Gaussian and its pdf can be evaluated easily. The channel can then be estimated by

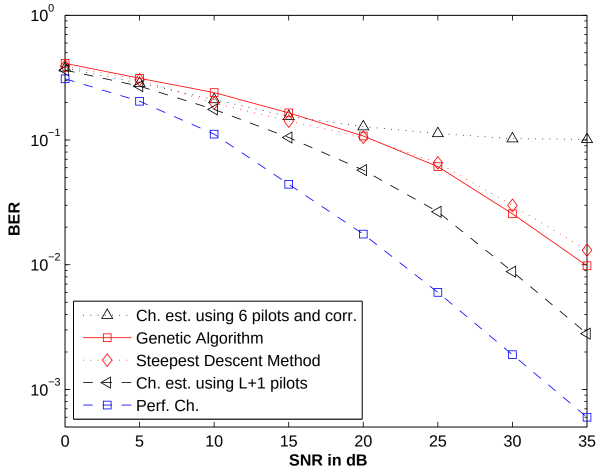

Fig. 4. BER vs SNR comparison for $16-$QAM modulated data

maximizing the likelihood function given the output pdf. The experimental results unfortunately demonstrated that the log likelihood function does not have a unique maxima when it is plotted against the channel taps. Therefore, it is difficult to come up with a convex formulation to pursue a blind approach. However, a blind channel estimation algorithm using GA was proposed due to its ability to avoid local minima. A semi-blind algorithm was also presented using the steepest descent algorithm initialized by a rough estimate of channel obtained by using a few of pilots and channel correlation. Simulation results show the favorable performance of the two proposed algorithms for constant as well as non-constant modulus data.

REFERENCES

[1] Z. Wang and G. B. Giannakis, "Wireless Multicarrier Communications: Where Fourier Meets Shannon," IEEE Signal Processing Mag., vol. 17, May 2000.

[2] G. Stuber, J. Barry, S. Mclaughlin, Y. Li, M. Ingram, and T. Pratt, "Broadband MIMO-OFDM wireless communications,"*Proc. IEEE*, vol. 92, no. 2, pp. 271−294, Feb. 2004.

[3] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Vehicular Tech. Conf.*, vol. 2, pp. 1131−1135, May 1999.

[4] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation,"*IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158−1168, May 2006.

[5] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-Assisted Time-Varying Channel Estimation for OFDM Systems," *IEEE Trans. Signal Proc.*, vol. 55, no. 5, pp. 2226-2238, Mar. 2007.

[6] S. Zhou and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1402-1414, Aug. 2001.

[7] T. Y. Al-Naffouri, "An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM," *IEEE Trans. Signal Proc.*, vol. 55, no. 7, pp. 3924−3930, Jul. 2007.

[8] R. W. Heath and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems,"*IEEE Trans. Signal Proc.*, vol. 47, no. 3, pp. 848−856, Mar. 1999.

[9] G. A. Al-Rawi, T. Y. Al-Naffouri, A. Bahai, and J. Cioffi, "Exploiting error-control coding and cyclic-prefix in channel estimation for coded OFDM systems,"*IEEE Commun. Lett.*, vol. 7, no. 8, pp. 388−390, Aug. 2003.

[10] T. Y. Al-Naffouri and A. A. Quadeer, "Cyclic Prefix based enhanced data recovery in OFDM,"*accepted in IEEE Trans. Signal Proc.*, Jul. 2009.

[11] T. Y. Al-Naffouri and A. A. Quadeer, "A Forward-Backward Kalman Filter-based STBC MIMO OFDM Receiver," *accepted in EURASIP Journal on Advances in Signal Processing*, Dec. 2008.

[12] T. Cui and C. Tellambura, "Semi-blind channel estimation and data detection for OFDM systems over frequency-selective fading channels," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 3, pp. iii/597−iii/600, Mar. 2005.

[13] W. Kunji, Z. Jianhua, L. Chaojun, and H. Chen, "Semi-blind OFDM channel estimation using receiver diversity in the presence of virtual carriers," *Int. Conf. Commun. and Networking*, pp. 1−4, Oct. 2006.

[14] H. Ochiai and H. Imai, "On the distribution of the peak-to-average power ratio in OFDM signals" IEEE Transactions on Communications, vol. 50, no. 1, pp. 89 101, Jan. 2001.

[15] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications,"*IEEE Trans. on Automatic Control*, vol. 25, no. 3, pp. 385−399, Jun. 1980.

[16] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics,"*IEEE Trans. on Acoust., Speech, and Signal Proc.*, vol. 37, no. 3, pp. 360−377, Mar. 1989.

[17] J. H. Holland, *Adaptation in natural and artificial systems.* The University of Michigan Press, 1975.

[18] F. Herrera, M. Lozano, JL. Verdegay, *Tackling real-coded genetic algorithms: operators and tools for behavioural analysis,* Kluwer Academic Publishers, 1998.

[19] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley and Sons, Inc., 1999.