Cyclic Prefix based Enhanced Data Recovery in OFDM*

T. Y. Al-Naffouri[1] and A. Quadeer[2]

Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Saudi Arabia, Email: {naffouri[1], aquadeer[2]}@kfupm.edu.sa

Abstract

In this correspondence, we show how the cyclic prefix can be used to enhance the performance of an OFDM receiver. Specifically, we show how an OFDM symbol can be blindly detected using output symbol and associated cyclic prefix. The algorithm boils down to a nonlinear relationship involving the input and output data only that can be searched for the maximum likelihood (ML) estimate of the input.This relationship becomes much simpler in case of constant modulus (CM) data. We also propose iterative methods to reduce the computational complexity involved in the ML search of input for CM data.

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) has emerged as a modulation scheme that can achieve high data rate by efficiently handling multipath effects. The additional advantages of simple implementation and high spectral efficiency due to orthogonality contribute towards the increasing interest in OFDM. In order to achieve high data rates in OFDM, the receiver must estimate the channel efficiently and subsequently the data. Many techniques have been proposed in the literature for channel estimation in OFDM systems. These techniques fall into three distinct classes: 1) training based, 2) semiblind and 3) blind methods. In training based methods, pilots i.e. symbols which are known to the receiver are sent with the data symbols [4] - [6]. Use of training sequences decreases the system bandwidth efficiency [7]. The limitations in training based estimation techniques motivated interest in the spectrally efficient blind approaches. Only natural constraints

*Part of this work has been presented at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, USA, Apr. 2008 [1]

are used for estimation in blind algorithms [8]-[11] while semi-blind techniques make use of both pilots and the natural constraints to efficiently estimate the channel [12]-[14].

A. Paper Contribution and Organization

In this paper, we perform blind channel identification and equalization from output data only (i.e. OFDM output symbol and associated CP). The advantage of our approach is three fold:

1. The method provides a blind estimate of the data from one output symbol without the need for training or averaging (contrary to the common practice in blind methods where averaging over several symbols is required). Thus, the method lends itself to block fading channels.

2. Data detection is done without any restriction on the channel (as long as the delay spread is shorter than the CP). In fact, data detection can be performed even in the presence of zeros on the FFT grid.[1]

3. The fact that we use two observations (the OFDM symbol and CP) to recover the input symbol enhances the diversity of the system as can be seen from simulations.

Our approach is based on the transformation of the OFDM channel into two parallel subchannels due to the presence of CP at the input (see Section II). One is a circular subchannel that relates the input and output OFDM symbols and thus is free of any intersymbol interference (ISI) effects and is best described in the frequency domain (Section II-A). The other one is a linear subchannel that carries the burden of ISI and that relates the input and output prefixes through linear convolution (Section II-B). This subchannel is best studied in the time domain. It can be shown that the two subchannels are characterized by the same set of parameters (or impulse response(IR)) and are driven by the same stream of data. They only differ in the way in which they operate on the data (i.e. linear vs circular convolution). This fact enables us in Section III to estimate the IR from one subchannel and eliminate its effect from the other, thus obtaining a nonlinear least squares relationship that involves the input and output data only. This relationship can in turn be optimized for the ML data estimate, something that can be achieved through exhaustive search (in the worst case scenario). The

[1]This comes contrary to the common belief that OFDM using CP cannot be equalized for channels with zeros on the FFT grid [2] and [3]

relationship takes a particularly simple form in CM case. Exhaustive search is computationally very expensive. We thus suggest in Section IV, two approaches to reduce the computational complexity in the CM case. In the first approach, we use the Genetic Algorithm (GA) [15], [16], [17], to directly solve the nonlinear problem. The second approach describes a semiblind algorithm in which we use Newton's method to estimate the data when it is initialized with an estimate obtained using a few number of pilots. To setup the stage, we introduce our notation in the following subsection and the system overview in Section II.

B. Notation

We denote scalars with small-case letters, vectors with small-case boldface letters, and matrices with uppercase boldface letters. Calligraphic notation (e.g. $\boldsymbol{\mathcal{X}}$) is reserved for vectors in the frequency domain. The individual entries of a vector like $\boldsymbol{h}$ are denoted by $h(l)$. A hat over a variable indicates an estimate of the variable (e.g., $\hat{\boldsymbol{h}}$ is an estimate of $\boldsymbol{h}$). When any of these variables becomes a function of time, the time index i appears as a subscript. Now consider a length-N vector $\boldsymbol{x}_i$. We deal with three derivatives associated with this vector. The first two are obtained by partitioning $\boldsymbol{x}_i$ into a lower (trailing) part $\underline{\boldsymbol{x}}_i$ (known as the CP) and an upper vector $\tilde{\boldsymbol{x}}_i$ so that $\boldsymbol{x}_i = [\tilde{\boldsymbol{x}}_i^T \quad \underline{\boldsymbol{x}}_i^T]^T$. The third derivative, $\overline{\boldsymbol{x}}_i$, is created by concatenating $\boldsymbol{x}_i$ with a copy of CP i.e. $\underline{\boldsymbol{x}}_i$. Thus we have, $\overline{\boldsymbol{x}}_i = [\underline{\boldsymbol{x}}_i^T \quad \boldsymbol{x}_i^T]^T = [\underline{\boldsymbol{x}}_i^T \quad \tilde{\boldsymbol{x}}_i^T \quad \underline{\boldsymbol{x}}_i^T]^T$. In line with the above notation, a matrix $\boldsymbol{Q}$ having N rows will have the natural partitioning, $\boldsymbol{Q} = [\tilde{\boldsymbol{Q}}^T \quad \underline{\boldsymbol{Q}}^T]^T$, where the number of rows in $\tilde{\boldsymbol{Q}}$ and $\underline{\boldsymbol{Q}}$ are understood from the context and when it is not clear, the number of rows will appear as a subscript.

II. System Overview

In an OFDM system, data is transmitted in symbols $\boldsymbol{\mathcal{X}}_i$ of length N each. The symbol undergoes an IDFT operation to produce the time domain symbol $\boldsymbol{x}_i$, i.e. $\boldsymbol{x}_i = \boldsymbol{Q}\boldsymbol{\mathcal{X}}_i$, where $\boldsymbol{Q}$ is the $N \times N$ IDFT matrix. When juxtaposed, these symbols result in the sequence $\{\boldsymbol{x}_k\}$.[2] We assume a channel $\underline{\boldsymbol{h}}$ of maximum length $L+1$. To avoid ISI caused by passing the symbol through a channel, a CP $\underline{\boldsymbol{x}}_i$ (of length L) is appended to $\boldsymbol{x}_i$, resulting in super-symbol $\overline{\boldsymbol{x}}_i$ as defined in Section I-B. The concatenation

[2]The time indices in the sequence $\boldsymbol{x}_i$ and the underlying sequence $\{\boldsymbol{x}_k\}$ are dummy variables. Nevertheless, we chose to index the two sequences differently to avoid any confusion that might arise from choosing identical indices.

of these symbols produces the underlying sequence $\{\overline{\boldsymbol{x}}_k\}$. When passed through the channel $\underline{\boldsymbol{h}}$, the sequence $\{\overline{\boldsymbol{x}}_k\}$ produces the output sequence $\{\overline{\boldsymbol{y}}_k\}$ i.e. $\overline{\boldsymbol{y}}_k = \underline{\boldsymbol{h}}_k * \overline{\boldsymbol{x}}_k + \overline{\boldsymbol{n}}_k$, where $\overline{\boldsymbol{n}}_k$ is the additive white Gaussian noise and $*$ stands for linear convolution. Motivated by the symbol structure of the input, it is convenient to partition the output into length $N + L$ symbols, $\overline{\boldsymbol{y}}_i = [\underline{\boldsymbol{y}}_i^T \;\; \boldsymbol{y}_i^T]^T$. This is a natural way to partition the output because the prefix $\underline{\boldsymbol{y}}_i$ actually absorbs all ISI that takes place between the adjacent symbols $\overline{\boldsymbol{x}}_{i-1}$ and $\overline{\boldsymbol{x}}_i$. Moreover, the remaining part $\boldsymbol{y}_i$ of the symbol depends on the ith input OFDM symbol $\boldsymbol{x}_i$ only. Because of the redundancy in the input, the convolution above can be decomposed into two distinct constituent convolution operations or subchannels, as we described in Section I above.

A. Circular Convolution (Subchannel)

Due to the presence of CP, the input and output OFDM symbols are related by circular convolution which reduces to an element by element operation in frequency domain

$$\boxed{\boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}}_i \odot \boldsymbol{\mathcal{X}}_i + \boldsymbol{\mathcal{N}}_i} \tag{1}$$

where $\boldsymbol{\mathcal{H}}_i$, $\boldsymbol{\mathcal{X}}_i$, $\boldsymbol{\mathcal{N}}_i$, and $\boldsymbol{\mathcal{Y}}_i$, are the DFT's of $\boldsymbol{h}_i$ (length-N zero-padded version of $\underline{\boldsymbol{h}}_i$), $\boldsymbol{x}_i$, $\boldsymbol{n}_i$, and $\boldsymbol{y}_i$ respectively. Since $\underline{\boldsymbol{h}}_i$ corresponds to the first $L + 1$ elements of $\boldsymbol{h}_i$, we can show that

$$\boldsymbol{\mathcal{H}}_i = \boldsymbol{Q}^*_{L+1}\underline{\boldsymbol{h}}_i \quad \text{and} \quad \underline{\boldsymbol{h}}_i = \boldsymbol{Q}_{L+1}\boldsymbol{\mathcal{H}}_i \tag{2}$$

where $\boldsymbol{Q}^*_{L+1}$ consists of the first $L + 1$ columns of $\boldsymbol{Q}^*$ and $\boldsymbol{Q}_{L+1}$ the first $L + 1$ rows of $\boldsymbol{Q}$. This allows us to rewrite (1) as, $\boldsymbol{\mathcal{Y}}_i = \operatorname{diag}(\boldsymbol{\mathcal{X}}_i)\boldsymbol{Q}^*_{L+1}\underline{\boldsymbol{h}}_i + \boldsymbol{\mathcal{N}}_i$.

B. Linear Convolution (Subchannel)

We can also note that the cyclic prefixes at the input and output are related by linear convolution in OFDM. Specifically, if we concatenate all cyclic prefixes at the input into a sequence $\{\underline{\boldsymbol{x}}_k\}$ and the cyclic prefixes at the output into the corresponding sequence $\{\underline{\boldsymbol{y}}_k\}$, then we can show that the two sequences are related by linear convolution [18], $\underline{\boldsymbol{y}}_k = \underline{\boldsymbol{h}}_k * \underline{\boldsymbol{x}}_k + \underline{\boldsymbol{n}}_i$. From this we deduce that the cyclic prefix of OFDM symbol, $\boldsymbol{y}_i$, is related to the input cyclic prefixes $\underline{\boldsymbol{x}}_{i-1}$ and $\underline{\boldsymbol{x}}_i$ by

$$\boxed{\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \underline{\boldsymbol{h}}_i + \underline{\boldsymbol{n}}_i} \tag{3}$$

where $\underline{\boldsymbol{X}}_i$ is constructed from $\underline{\boldsymbol{x}}_{i-1}$ and $\underline{\boldsymbol{x}}_i$ according to $\underline{\boldsymbol{X}}_i = \underline{\boldsymbol{X}}_{Ui-1} + \underline{\boldsymbol{X}}_{Li}$, and where

$$\underline{\boldsymbol{X}}_{\mathsf{U}i-1} = \begin{bmatrix} 0 & \underline{x}_{i-1}(L-1) & \cdots & \underline{x}_{i-1}(0) \\ 0 & 0 & \cdots & \underline{x}_{i-1}(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \underline{x}_{i-1}(L-1) \end{bmatrix} \text{ and } \underline{\boldsymbol{X}}_{\mathsf{L}i} = \begin{bmatrix} \underline{x}_i(0) & 0 & \cdots & 0 \\ \underline{x}_i(1) & \underline{x}_i(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \underline{x}_i(L-1) & \cdots & \underline{x}_i(0) & 0 \end{bmatrix} \tag{4}$$

This fact together with the FFT relationship (2) yields the time-frequency input/output equation

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \boldsymbol{Q}_{L+1} \boldsymbol{\mathcal{H}}_i + \underline{\boldsymbol{n}}_i \tag{5}$$

III. Maximum-Likelihood Estimation

Consider the frequency domain description of the circular subchannel (1). To obtain the maximum-likelihood (ML) estimate of $\boldsymbol{\mathcal{H}}_i$, we assume that the sequence $\boldsymbol{\mathcal{X}}_i$ is deterministic and perform an element-by-element division of (1) by $\boldsymbol{\mathcal{X}}_i$ to get

$$\boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}}_i + \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{N}}_i \tag{6}$$

where $\boldsymbol{D}_{\mathcal{X}} = \operatorname{diag}(\boldsymbol{\mathcal{X}}_i)$. Equivalently, we can write it as $\boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i = \boldsymbol{\mathcal{H}}_i + \boldsymbol{\mathcal{N}}'_i$, where $\boldsymbol{\mathcal{N}}'_i$ is Gaussian distributed with zero mean and autocorrelation matrix $\boldsymbol{R}_{n'} = \sigma_n^2 \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{D}_{\mathcal{X}}^{-*} = \sigma_n^2 |\boldsymbol{D}_{\mathcal{X}}|^{-2}$. The maximum-likelihood estimate of $\boldsymbol{\mathcal{H}}$ can now be obtained by solving the system of equations (6) in the least-squares (LS) sense subject to the constraint $\tilde{\boldsymbol{Q}}_{N-L-1} \boldsymbol{\mathcal{H}}_i \triangleq \tilde{\boldsymbol{Q}} \boldsymbol{\mathcal{H}}_i = 0$. We can show that the ML estimate is given by [20]

$$\hat{\boldsymbol{\mathcal{H}}}_i^{ML} = \left[\boldsymbol{I} - \boldsymbol{R}_{n'} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} \boldsymbol{R}_{n'} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i = \left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i \tag{7}$$

The ML estimate (7) was obtained solely from the circular convolution subchannel. Upon replacing $\boldsymbol{\mathcal{H}}_i$ that appears in the time-frequency input/output equation (5) (corresponding to the linear subchannel) with its ML estimate (7), we obtain

$$\underline{\boldsymbol{y}}_i = \underline{\boldsymbol{X}}_i \boldsymbol{Q}_{L+1} \left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i + \underline{\boldsymbol{n}}_i \tag{8}$$

This is an input/output relationship that does not depend on any channel information whatsoever. Since the data is assumed deterministic, maximum-likelihood estimation is the optimum way to detect it

$$\hat{\boldsymbol{\mathcal{X}}}_i^{ML} = \arg \min_{\boldsymbol{\mathcal{X}}_i} \left\| \underline{\boldsymbol{y}}_i - \underline{\boldsymbol{X}}_i \boldsymbol{Q}_{L+1} \left[\boldsymbol{I} - |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \left(\tilde{\boldsymbol{Q}} |\boldsymbol{D}_{\mathcal{X}}|^{-2} \tilde{\boldsymbol{Q}}^* \right)^{-1} \tilde{\boldsymbol{Q}} \right] \boldsymbol{D}_{\mathcal{X}}^{-1} \boldsymbol{\mathcal{Y}}_i \right\|^2 \tag{9}$$

This is a nonlinear least-squares problem in the data. To gain more insight into this problem, we now treat the case of CM data which leads to more explicit results.

In the CM case, we have $|\boldsymbol{D}_{\mathcal{X}}|^{-2} = \frac{1}{\mathcal{E}_{\mathcal{X}}} \boldsymbol{I}$. As a consequence, we can also write $\boldsymbol{D}_{\mathcal{X}}^{-1} = \frac{1}{\mathcal{E}_{\mathcal{X}}} \boldsymbol{D}_{\mathcal{X}}^*$, where $\mathcal{E}_{\mathcal{X}}$ stands for energy of data. Using this and the unitary nature of $\boldsymbol{Q}$, $\boldsymbol{I} = \boldsymbol{Q}\boldsymbol{Q}^*$ $= [\tilde{\boldsymbol{Q}}_{N-L}^T \; \underline{\boldsymbol{Q}}_L^T]^T \; [\underline{\boldsymbol{Q}}_{L+1}^* \; \tilde{\boldsymbol{Q}}_{N-L-1}^*]$, simplifies equation (9) to (after some algebraic manipulations)

$$\hat{\boldsymbol{\mathcal{X}}}_i^{\mathsf{ML}} = \arg \min_{\mathcal{X}_i} \left\| \underline{\boldsymbol{y}}_i - \frac{1}{\mathcal{E}_{\mathcal{X}}} \underline{\boldsymbol{X}}_i \underline{\boldsymbol{Q}}_{L+1} \boldsymbol{\mathcal{Y}}_i \odot \boldsymbol{\mathcal{X}}_i^* \right\|^2 \tag{10}$$

Notice that the only unknowns in this minimization are $\underline{\boldsymbol{X}}_i$ and $\boldsymbol{\mathcal{X}}_i$, i.e. the input data sequence. This minimization is nothing but a *nonlinear least-squares* problem in the data. In the worst case scenario, we can obtain the ML estimate through an exhaustive search.

IV. Approximate Methods to Reduce Computational Complexity

The search for the optimal $\boldsymbol{\mathcal{X}}_i$ in (10) is computationally very complex. In the following, we describe two approaches to reduce this complexity:

A. The Genetic Algorithm

We can use the search algorithms like the Genetic Algorithm (GA), to directly solve the nonlinear problem (10). GA was first introduced by [15]. GA is a robust, population based iterative stochastic search algorithm that is based on natural selection (survival of the fittest) and evolution [16], [17]. The reason behind GA being widely used in optimization problems is its ability to avoid local minima.

A *population* of *chromosomes* (candidate solutions to the problem of size N in our case) is generated. Each chromosome has a *fitness* (a positive number) associated to it which represents the goodness of the solution. The fitness in our case is calculated by evaluating the cost function (10) for a particular chromosome. This fitness is used to determine the parent chromosomes which will produce the offsprings in the next generation. This process is called *selection*. The selected parents are allowed to reproduce using the genetic operators called *cross-over* and *mutation*. The parent chromosomes with the highest fitness values, known as *elite* chromosomes, are transferred to the next generation without any change to be utilized again in reproduction. It is important to note that during selection process, it is necessary to prevent *incest* (i.e. the two parents being selected for reproduction should not be same) to avoid local minima [17]. The process is terminated when a fixed number of iterations called *generations* are completed.

B. Newton's Method

Pilots can also be used to reduce the computational complexity. Specifically, a few pilots are used to obtain an initial channel estimate which is then enhanced using Newton's method. To this end, the objective function of concern here is, $\boldsymbol{Z} = \left\| \underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{C}\boldsymbol{\mathcal{X}}_i^* \right\|^2$ subject to the CM constraint, $\Psi_j = |\boldsymbol{\mathcal{X}}_i(j)|^2 = \mathcal{E}_{\mathcal{X}}$ $(j = 1, 2, ..., N)$, where $\boldsymbol{B} = \underline{\boldsymbol{X}}_{\mathsf{U}i-1}\frac{1}{\mathcal{E}_{\mathcal{X}}}\boldsymbol{Q}_{L+1}\boldsymbol{D}_{\mathcal{Y}}$, and $\boldsymbol{C} = \underline{\boldsymbol{X}}_{\mathsf{L}i}\frac{1}{\mathcal{E}_{\mathcal{X}}}\boldsymbol{Q}_{L+1}\boldsymbol{D}_{\mathcal{Y}}$. We shall apply Newton's method to the cost function, $\boldsymbol{Z} + \frac{1}{\sigma_n^2}\sum_{j=1}^{N}\Psi_j$, where $\Psi_j = \left\| \mathcal{E}_{\mathcal{X}} - \boldsymbol{\mathcal{X}}_i^H \boldsymbol{E}_j \boldsymbol{\mathcal{X}}_i \right\|^2$. If initial estimate of data $\boldsymbol{\mathcal{X}}_{-1}$ is available, then it can be refined by applying Newton's Method [20]

$$\boldsymbol{\mathcal{X}}_k = \boldsymbol{\mathcal{X}}_{k-1} - \mu \left[\nabla^2 \boldsymbol{Z}(\boldsymbol{\mathcal{X}}_{k-1})\right]^{-1} \left[\nabla \boldsymbol{Z}(\boldsymbol{\mathcal{X}}_{k-1})\right]^*, \quad k \geq 0 \tag{11}$$

where μ is the step size, ∇ is the gradient, and ∇^2 is the Hessian of cost function $\boldsymbol{Z}$ subjected to the CM constraint. The algorithm runs iteratively until a maximum number of iterations or a stopping criteria is reached. To implement Newton's method, we need to calculate gradient and Hessian of the cost function.

B.1. Evaluating the Gradient. We evaluate the gradient of cost function and the CM constraint on data separately. The cost function $\boldsymbol{Z}$ can be written as, $\boldsymbol{Z} = \left\|\underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{C}\boldsymbol{\mathcal{X}}_i^*\right\|^2 = \left\|\underline{\boldsymbol{y}}_i - \boldsymbol{B}\boldsymbol{\mathcal{X}}_i^* - \boldsymbol{c}\right\|^2 = \|\boldsymbol{a}\|^2 = \boldsymbol{a}^H\boldsymbol{a}$, where each element of $\boldsymbol{c}$ is, $c(j) = \|\boldsymbol{\mathcal{X}}_i\|^2_{\boldsymbol{W}_j} \triangleq \boldsymbol{\mathcal{X}}_i^*\boldsymbol{W}_j\boldsymbol{\mathcal{X}}_i$, for some weighted matrix $\boldsymbol{W}_j$ that is independent from input $\boldsymbol{\mathcal{X}}_i$. First we define a matrix, $\boldsymbol{F}(\boldsymbol{y}_i^T, \boldsymbol{Y}) = [(\boldsymbol{y}_i{}^T\boldsymbol{Y}_1)^T\ (\boldsymbol{y}_i{}^T\boldsymbol{Y}_2)^T\ \cdots\ (\boldsymbol{y}_i{}^T\boldsymbol{Y}_j)^T]^T$, where $\boldsymbol{y}$ represents a vector and $\boldsymbol{Y}$ a matrix. The gradient of $\boldsymbol{Z}$ is obtained by using chain rule for complex matrices [19], given by, $\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} = -\boldsymbol{a}^H\boldsymbol{B} - \boldsymbol{a}^H\boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^T, \boldsymbol{W}) - \boldsymbol{a}^T\boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^T, \boldsymbol{W})$. The CM constraint on data can be written as $\Psi_j = \left\|\mathcal{E}_{\mathcal{X}} - \boldsymbol{\mathcal{X}}_i^H\boldsymbol{E}_j\boldsymbol{\mathcal{X}}_i\right\|^2 = \|\boldsymbol{b}\|^2 = \boldsymbol{b}^H\boldsymbol{b}$, where $\boldsymbol{E}_j$ is a $N \times N$ matrix which is all zero except for one nonzero diagonal element $e_{jj} = 1$. The gradient of CM constraint is calculated in a similar manner to that of the cost function and is given by, $\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*} = -\boldsymbol{b}^H\boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^T, \boldsymbol{E}^T) - \boldsymbol{b}^T\boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^T, \boldsymbol{E}^*)$. Finally, the gradient of cost function subject to the constraint is a vector of size $1 \times N$, given by, $\nabla(\boldsymbol{Z}) = \frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*} + \frac{1}{\sigma_n^2}\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*}$.

B.2. Evaluating the Hessian. Similar to gradient, we evaluate the Hessian of cost function and constraint separately. The Hessian of the cost function is given by, $\frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T = -\boldsymbol{B}^T\left(-\boldsymbol{B}^* - \boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^H, \boldsymbol{W}^*)\right) - \boldsymbol{F}(\boldsymbol{W}, \boldsymbol{\mathcal{X}}_i)\left(-\boldsymbol{B}^* - \boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^H, \boldsymbol{W}^*)\right) - \boldsymbol{F}(\boldsymbol{W}^*, \boldsymbol{\mathcal{X}}_i)\boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^H, \boldsymbol{W}^T)$. The Hessian of CM constraint on data is calculated in a similar fashion to cost function and the final result is given as, $\frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T = \boldsymbol{F}(\boldsymbol{E}, \boldsymbol{\mathcal{X}}_i)\boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^H, \boldsymbol{E}^H) + \boldsymbol{F}(\boldsymbol{E}^H, \boldsymbol{\mathcal{X}}_i)\boldsymbol{F}(\boldsymbol{\mathcal{X}}_i^H, \boldsymbol{E})$. Thus, the Hessian of the cost function subject to the CM constraint is a matrix of size $N \times N$, given by, $\nabla^2(\boldsymbol{Z}) = \frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T + \frac{1}{\sigma_n^2}\frac{\partial}{\partial \boldsymbol{\mathcal{X}}_i}\left(\frac{\partial \Psi}{\partial \boldsymbol{\mathcal{X}}_i^*}\right)^T$.

V. Simulation Results

We consider an OFDM system with $N = 16$ and CP of length $L = 4$. The OFDM symbol consists of BPSK or 4-QAM symbols. The channel IR consists of $L + 1 = 5$ iid Rayleigh fading taps.

A. Bench Marking. To bench mark our blind algorithm, we compared it with the blind algorithm proposed by Muquet et al. in [12], a method in which channel is estimated using $L + 1$ pilots, and perfectly known channel case. As opposed to our algorithm in which estimation is done in a single block, the subspace method in [12] requires large memory to collect sufficiently enough data blocks to render the covariance matrix full rank. It also suffers from sign ambiguity and cannot be implemented without transmitting a pilot in the first symbol of the block. Moreover, it fails in the presence of nulls on subcarriers while our algorithm is robust to channel nulls.

In Figure 1, the performance of both algorithms is compared for BPSK modulated data over a Rayleigh channel. To make the covariance matrix full rank, 50 blocks of data were considered to implement the subspace algorithm while only a single block of data was used for channel estimation in our algorithm. As expected, the best performance is achieved by the perfectly known channel, followed by that obtained by training based estimated channel. It can be observed that both the blind algorithms perform close to each other at low SNR but our algorithm outperforms the subspace algorithm at high SNR. It should also be noted that in the high SNR region, the BER curve of our blind algorithm exhibits steeper slope (higher diversity) which can be explained from the fact that the two subchannels (linear and circular) are used to detect the data in our case when only the circular subchannel is used in the pilot based and known channel cases. An alternative way to see this is to note that a Rayleigh fading channel will occasionally hit a (near) zero on the FFT grid resulting in a loss of the corresponding BPSK symbol. Our blind algorithm does not suffer from this and thus demonstrates improved performance in higher SNR.

In Figure 2, it can be seen that the performance of subspace algorithm becomes worse when only 20 data blocks are used as the covariance matrix is not full rank. Figure 3 compares the performance of the algorithms when the channel has persistent nulls. The number of blocks used to implement the subspace algorithm is fixed at 50. Our algorithm is robust to channel nulls and thus in this case, it easily outperforms the subspace one. It should also be noted that at high SNR, the BER for perfectly known channel and that of the estimated channel reach an error floor but our algorithm does not suffer from it and outperforms them too.

B. Performance using Genetic Algorithm. Figure 4 shows the performance of GA for BPSK modulated data over a Rayleigh fading channel. The parameters used in implementing the GA are listed in Table 1. It can be observed from Figure 4 that GA performs quite close to the blind exhaustive search.

TABLE I
Simulation Parameters used to implement GA

Symbol Initialization	Random
Population Size	200
Number of Generations	150
Selection Method	Fitness−Proportionate
Incest Prevention	Yes
Cross-over Scheme	Uniform Cross-over
Mutation Scheme	Uniform Mutation
Mutation Probability	0.15
Number of Elite Chromosomes	20

C. Semiblind Recovery using Newton's Method. The Newton's method described in equation (11) was implemented using a step size of 0.5. The iterative algorithm was run until the difference between the value of current and previous cost function becomes less than 10^{-6}. Figure 5 shows the performance of Newton's method for 4-QAM with $N = 16$ and $L = 4$ when it is initialized by the estimate obtained by using 3 pilots and channel correlation. It can be seen that the 3 pilots based method reaches an error floor at high SNR while the Newton's method performs quite close to the blind exhaustive search. In Figure 6, the performance of Newton's method is compared with the $L+1$ pilots case and perfectly known channel for more realistic OFDM system using 4-QAM with $N = 64$ and $L = 16$. The Newton's method is initialized with an estimate obtained by using 12 pilots and channel correlation. Figure 6 clearly indicates that Newton's method performs quite well even for higher number of carriers.

VI. Conclusion

In this paper, we demonstrated how to perform blind ML data recovery in OFDM transmission. This is done using a single output OFDM symbol and associated CP. In particular, it was shown that the ML data estimate is the solution of an integer nonlinear-least squares problem which becomes simpler in the case of CM data. We further demonstrated that data recovery is possible from output data only, irrespective of the channel zero locations and irrespective of the quality of the channel estimates or of its exact order. We have also proposed approximate methods to reduce the exponential complexity entailed in the algorithm developed in the paper. As is evident from the simulation results, GA performs quite close to the exhaustive search method. As all standard-based OFDM systems involve some form of training, we have also studied the behavior of the blind receiver in the presence of pilots and channel frequency correlation. It was found that Newton's method performs quite well at all values of SNR even for higher number of carriers.

References

[1] T. Y. Al-Naffouri and Ahmed Abdul Quadeer, "Blind maximum-likelihood data recovery in OFDM," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 1, pp. 2829-2832, Apr. 2008.

[2] Z. Wang and G. B. Giannakis, "Wireless Multicarrier Communications: Where Fourier Meets Shannon," *IEEE Signal Proc. Mag.*, vol. 17, no. 3, pp. 29-48, May 2000.

[3] E. Chen, R. Tao, and X. Zhao, "Channel Equalization for OFDM System Based on the BP Neural Network," *Int. Conf. Signal Proc.*, vol. 3, 2006.

[4] 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.

[5] 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.

[6] M. Ghogho, D. McLernon, E. Alameda-Hemandez, A. Swami, "SISO and MIMO channel estimation and symbol detection using data-dependent superimposed training," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 3, pp. iii/461-iii/464, Mar. 2005.

[7] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951 1968, Oct. 1998.

[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] T. Kim and I. Eo, "Reliable blind channel estimation scheme based on cross-correlated cyclic prefix for OFDM system," *Int. Conf. Adv. Commun. Technol.*, vol. 1, Feb. 2006.

[10] 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 Comm. Lett.*, vol. 7, no. 8, pp. 388-390, Aug. 2003.

[11] 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.

[12] B. Muquet, M. de Courville and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Proc.*, vol. 50, no. 7, pp. 1699-1712, Jul. 2002.

[13] 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.

[14] R. C. Cannizzaro, P. Banelli and G. Leus, "Adaptive Channel Estimation for OFDM Systems with Doppler spread," *IEEE Signal Proc. Workshop Signal Process. Adv. Wireless Comm.*, pp. 15, Jul. 2006.

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

[16] K. C. Sharman and G. D. McClurkin, "Genetic algorithms for maximum likelihood parameter estimation," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 4, pp. 2716-2719, May 1989.

[17] L. Hanzo, L-L. Yang, E-L. Kaun, and K. Yen, *Single- and Multi-Carrier DS-CDMA Multi-User Detection, Space-Time Spreading, Synchronization and Standards.* John Wiley & Sons, Ltd., 2003.

[18] X. Wang and R. Liu, "Adaptive channel estimation using cyclic prefix in multicarrier modulation system," *IEEE Comm. Lett.*, vol. 3, no. 10, pp. 291-293, Oct. 1999.

[19] A. Hjorungnes, D. Gesbert, D. P. Palomar, "Unified Theory of Complex-Valued Matrix Differentiation," *IEEE Int. Conf. on Acoust. Speech and Signal Proc.*, vol. 3, pp. III-345−III-348, Apr. 2007.

[20] Ali H. Sayed, *Fundamentals of Adaptive Filtering.* John Wiley & Sons, Ltd., 2003.

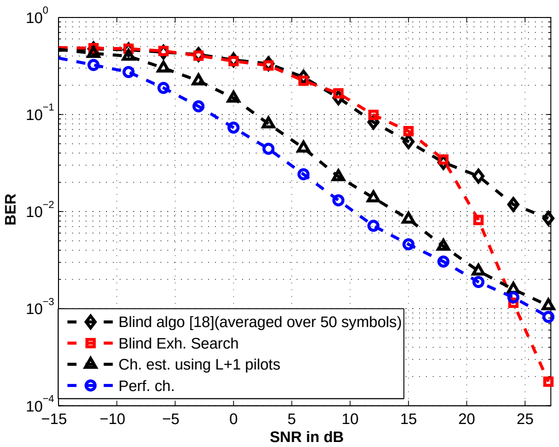

Fig. 1. BER performance of Kalman and FB-Kalman using Hard and Soft estimate of data

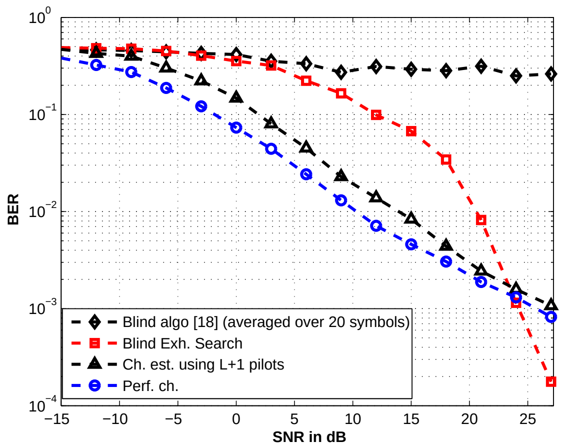

Fig. 2. BER performance of Kalman and FB-Kalman using Hard and Soft estimate of data

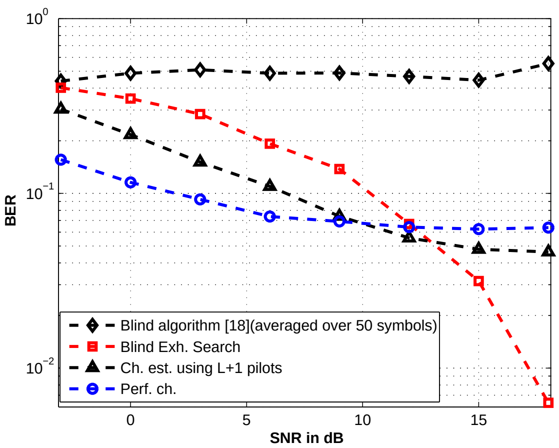

Fig. 3. BER performance of Kalman and FB-Kalman using Hard and Soft estimate of data

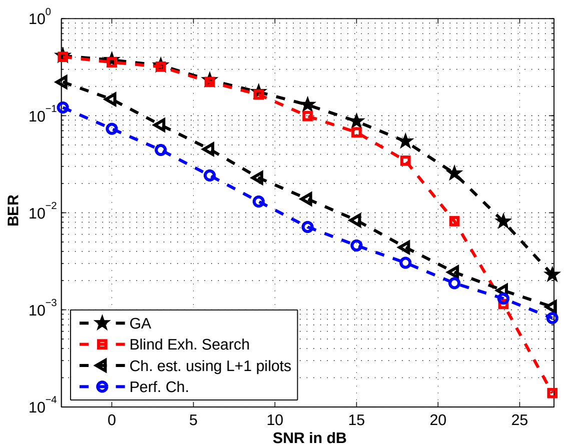

Fig. 4. Comparison of low complexity algorithms for BPSK-OFDM over a Rayleigh channel

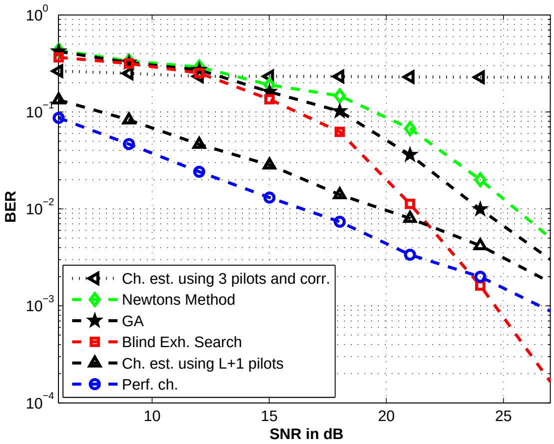

Fig. 5. Comparison of Newton's Method for 4QAM-OFDM with $N = 16$ and $L = 4$ over a Rayleigh channel

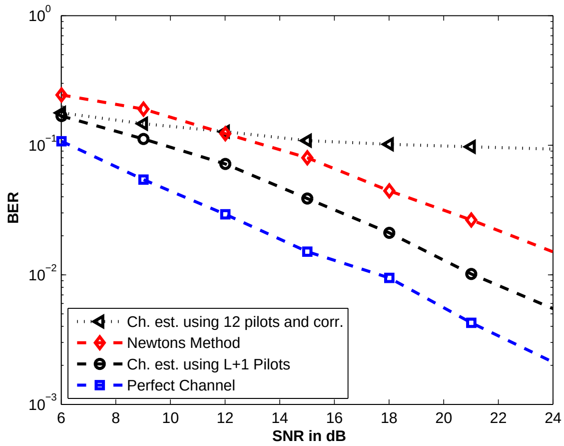

Fig. 6. Comparison of Newton's Method for 4QAM-OFDM with $N = 64$ and $L = 16$ over a Rayleigh channel