Blind MMSE Equalization of FIR/IIR Channels Using Oversampling and Multichannel Linear Prediction

Fangjiong Chen, Sam Kwong, and Chi-Wah Kok

A linear-prediction-based blind equalization algorithm for single-input single-output (SISO) finite impulse response/infinite impulse response (FIR/IIR) channels is proposed. The new algorithm is based on second-order statistics, and it does not require channel order estimation. By oversampling the channel output, the SISO channel model is converted to a special single-input multiple-output (SIMO) model. Two forward linear predictors with consecutive prediction delays are applied to the subchannel outputs of the SIMO model. It is demonstrated that the partial parameters of the SIMO model can be estimated from the difference between the prediction errors when the length of the predictors is sufficiently large. The sufficient filter length for achieving the optimal prediction is also derived. Based on the estimated parameters, both batch and adaptive minimum-mean-square-error equalizers are developed. The performance of the proposed equalizers is evaluated by computer simulations and compared with existing algorithms.

Keywords: Equalizers, least-mean-square methods, adaptive equalizers, estimation.

Manuscript received Mar. 6, 2008; revised Jan. 13, 2009; accepted Feb. 10, 2009.

Part of this work has been presented in the IEEE VTC2006 Spring [1].

This work was supported by the NSF of China (Grant No. 60672065), the Natural Science Foundation of Guangdong Province, China (Grant No. 8151064101000066), and the City University Strategic Grant (Grant No. 7002294).

Fangjiong Chen (phone: +86 20 87112114 12, email: eefjchen@scut.edu.cn) is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China.

Sam Kwong (email: cssamk@cityu.edu.hk) is with the Department of Computer Science, City University of Hong Kong, Kowloon Tong, Kowloon, Hong Kong.

Chi-Wah Kok (email: eekok@ieee.org) is with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong.

I. Introduction

Blind equalization has become an important research problem in digital signal processing. If a training sequence is available, an adaptive equalizer can be easily developed using the standard least-mean-square (LMS) algorithm. However, there are many conditions, such as high data rates or bandlimited digital communication systems, in which the transmission of a training sequence is impractical or very costly. Therefore, blind adaptive equalization algorithms that do not rely on training signals need to be developed [2].

Previous research on blind equalization has been based on higher (than second) order statistics (HOS) of the channel outputs. These kinds of algorithms usually are derived via optimization criteria involving HOS. Various gradient-based algorithms are then employed to achieve optimization [3], [4]. The HOS-based methods can be divided into two classes: implicit and explicit [4]. Implicit methods are adaptive and have simple implementation. Explicit methods need to explicitly estimate the HOS from a received block of signal. The major problem with HOS-based algorithms is their slow convergence and multiple local minimizers. Moreover, the global convergence may be jeopardized when the channel has finite impulse response [3].

In the 1990s, Tong and others showed that second-order statistics (SOS) of the channel outputs contain sufficient information to estimate the parameters of most communication channels when the outputs are sampled faster than the symbol rate (oversampling) [5]. The SOS method provides satisfactory tradeoff between complexity and estimation performance. Since the work of Tong, many efficient algorithms have been proposed for SOS-based blind channel estimation and equalization [6]-[10]. However, most SOS-based channel

estimation algorithms and direct equalization algorithms require the assumption that the channel has finite impulse response (FIR) with known channel order. Although it is reasonable to approximate a practical communication channel using an FIR filter, such an approximation may result in a very high channel order. For instance, the delay spread of an outdoor wireless channel is on the order of μs [11]. The typical symbol rate of wireless systems is on the order of 10 million symbols per second (20M symbols/s in 802.11a). As a result, the FIR channel order can be on the order of 10, or more than 100 when oversampling is applied. When the transmission rate becomes higher (1 Gbps has been targeted in WLAN [12]), we conjecture that current equalization algorithms for FIR channel models will not be efficient due to the potentially high orders. Another example is the channels in digital subscriber line (DSL), where the channel response typically has a long tail [13], [14]. In some studies, the channel response was truncated and modeled as an FIR filter [14], which is not accurate and only practical at low to medium data transmission rates.

The estimation and equalization of high-order FIR channel models require heavy computational cost. In contrast, an infinite impulse response (IIR) approximation is of a much smaller order and, consequently, incurs less computational cost [15]. Therefore, it is interesting to model communication channels by using IIR filters. An IIR channel model was applied to DSL channel equalization in [16], however the equalization algorithm applied is based on training symbols. Deterministic approaches, (that is, approaches not based on statistics) were proposed for blind parameter estimation of single-input single-output (SISO) IIR channels and single-input multiple-output (SIMO) IIR channels in [15] and [17], respectively. Since estimation of statistics is not required, only a short block of received data (tens of symbols) is required for blind parameter estimation. However, the algorithms are sensitive to channel noise. In this paper, we consider direct channel equalization instead. That is, we estimate the equalizer rather than the channel parameters. The SOSs of the channel output is exploited to alleviate the effect of channel noise. The SOSs were also exploited in [18] for direct IIR channel equalization. In [18], multi-step predictors were applied to estimate partial channel impulse response based on which the channel equalizer was calculated. In this paper we generalize the prediction model in [18] to ensure that the same partial channel response can be estimated by a one-step predictor.

We consider direct equalization of SISO IIR channels. By oversampling the channel output, we show that the SISO IIR channel model can be uniquely transformed into a special SIMO model [19]; therefore, we can alternatively consider blind channel equalization of this SIMO IIR model. We consider Li's linear prediction (LP) problems [8], [9] in the SIMO model and demonstrate that the prediction error derivations in [8] are still valid. Hence, as in [9], partial model parameters can be estimated from the prediction errors. Based on the estimated model parameters, a batch MMSE equalizer with finite delay is developed. In addition, adaptive approaches are developed for tracking time varying channels. The batch algorithm requires the computation of two eigenvalue decompositions (EVD), while the adaptive algorithm does not require explicit computation of EVD and matrix inversion, and can be efficiently implemented by a recursive least-square (RLS) algorithm or an LMS algorithm.

The remainder of this paper is organized as follows. In section II, we review the problem of blind equalization by oversampling the received signal. A new MMSE equalizer is presented in section III, and adaptive algorithms are introduced in section V. Computer simulation results are given in section VI, and finally, conclusions are drawn in section VII.

II. Problem Model

Let $x(n)$ denote the discrete signal to be transmitted, $p(t)$ denote the pulse shaping filter at the transmitter, and $h_c(t)$ denote the continuous-time channel response. The channel output is given by

$$\begin{aligned} y(t) &= u(t) * h_c(t) + v(t), \\ u(t) &= \sum_{n=-\infty}^{+\infty} x(n)\, p(t - nT_i), \end{aligned} \tag{1}$$

where T_i is the transmitting symbol interval, and $v(t)$ is the channel noise, which is assumed to be white and Gaussian in this paper. Similar to [10], [15], [20], and [21], we define $h(t)=p(t)*h_c(t)$ as the equivalent channel to be equalized. The channel output is represented as

$$y(t) = \sum_{-\infty}^{\infty} x(n) h(t - nT_i) + v(t). \tag{2}$$

We assume that the output sampling rate is an integral multiple of the input baud rate. Let T_o denote the output sampling interval. We obtain $p=T_i/T_o$ as the oversampling ratio. Without loss of generality, we assume that T_o is normalized to $T_o = 1$. The discrete output then can be denoted as $y(k)$, and its z-transform is given by

$$Y(z) = H(z) X(z^p) + V(z). \tag{3}$$

In our derivations, the uppercase quantities are the z-transforms of the respective lowercase quantities. We assume that the input signal is an uncorrelated time sequence with zero mean and unit power. Consider a causal IIR channel with transfer function $H(z) = B(z)/A(z)$. It was proved that, in the

noiseless case, the SISO model can be uniquely transformed to a SIMO model as in [19] as

$$Y^l(z) = F^l(z)X(z) = \frac{D^l(z)}{G(z)}X(z), \quad l = 0, \cdots, p-1, \quad (4)$$

where $D^l(z)/G(z)$ and $Y^l(z)$ are the transfer function and output of the l-th subchannel, respectively. Here, $Y^l(z)$ is obtained from the polyphase decomposition of $Y(z)$: $Y(z) = Y^0(z^P) + z^{-1}Y^{-1}(z^P) + \cdots + z^{-P+1}Y^{P-1}(z^P)$. Since the transformation from (2) to (3) is unique and $y^l(k)$ can be obtained by decimating $y(k)$, we can equivalently equalize the SIMO model in (4). Channel noise in (4) is obtained from $v(t)$ by sampling and decimating. Since $v(t)$ is white and Gaussian, the noise in the subchannels of (4) are white, Gaussian, and mutually uncorrelated. When a match filter is applied at the receiver, the discrete noise of the subchannels becomes time-correlative due to oversampling. Since the power spectrum density of the channel noise is known, we can whiten the noise before channel estimation. For derivation simplicity, we consider only white noise in this paper.

Collecting p output samples in a vector as $\mathbf{y}(k) = [y^0(k) \ \cdots \ y^{p-1}(k)]^T$ and temporarily ignoring the effect of noise, we have $\mathbf{y}(k) = [\mathbf{f}_0 \ \mathbf{f}_1 \ \cdots]\mathbf{x}(k)$, where $\mathbf{f}_i = [f^0(i) \ \cdots \ f^{p-1}(i)]^T$ and $\mathbf{x}(k) = [x(k) \ x(k\text{-}1) \ \cdots]^T$. Denote $\mathbf{y}_K(k) = [\mathbf{y}^T(k) \ \cdots \ \mathbf{y}^T(k-K+1)]^T$ as the output vector collected at time slot k. We obtain

$$\mathbf{y}_K(k) = \mathbf{F}\mathbf{x}(k), \quad (5)$$

where

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}_0 & \mathbf{f}_1 & & \cdots & \cdots & \cdots \\ 0 & \mathbf{f}_0 & \mathbf{f}_1 & \cdots & \cdots & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \mathbf{f}_0 & \mathbf{f}_1 & \cdots \end{pmatrix}. \quad (6)$$

In the following section we derive an MMSE equalizer based on the SIMO IIR channel model of (5).

III. Blind MMSE Equalization

Denote $\mathbf{w}_d$ as the MMSE equalizer with delay d which minimizes $E \| \mathbf{w}_d^H \mathbf{y}_K - x(n-d) \|^2$. Since the input signal is an uncorrelated white process, one can easily derive that

$$\mathbf{w}_d = \mathbf{R}_K^{-1}\mathbf{f}^d, \quad (7)$$

where $\mathbf{f}^d$ is the $(d+1)$th column of $\mathbf{F}$, and $\mathbf{R}_K$ is the autocorrelation matrix of $\mathbf{y}_K(k)$, which can be estimated from the observed data. Our task is to estimate $\mathbf{f}^d$. Note that $\mathbf{f}^d$ contains partial parameters of the channel impulse response. The estimation of $\mathbf{f}^d$ is called a partial model estimation. A similar approach was applied in [18], where the applied problem model is a special case of (5), in which $\mathbf{y}_1(k) = [\mathbf{f}_0 \ \mathbf{f}_1 \ \cdots]\mathbf{x}(k)$. Multi-step predictors were applied to estimate $[\mathbf{f}_0 \ \mathbf{f}_1 \ \cdots \ \mathbf{f}_d]$ separately and then to form the estimate of $\mathbf{f}^d$. In this paper, we generalize the problem model and show that $\mathbf{f}^d$ can be estimated by one-step linear prediction. Consider the following linear prediction problem:

$$\boldsymbol{\varepsilon}^{(d)}(k) = \begin{pmatrix} \mathbf{I}_{pK} & -\mathbf{P}_d \end{pmatrix} \begin{pmatrix} \mathbf{y}_K(k) \\ \mathbf{y}_M(k-d-1) \end{pmatrix}, \quad (8)$$

where $\mathbf{y}_M(k-d-1) = [\mathbf{y}^T(k-d-1), \cdots, \mathbf{y}^T(k-d-M)]^T$, $\mathbf{I}_{pK}$ is the $pK \times pK$ identity matrix, and $\mathbf{P}_d$ is the $pK \times pM$ prediction matrix. We can estimate $\mathbf{P}_d$ from

$$\mathbf{P}_d = \arg\min_{\mathbf{P}_d} tr\{E[\boldsymbol{\varepsilon}^{(d)}(k)][\boldsymbol{\varepsilon}^{(d)}(k)]^H\}. \quad (9)$$

We first consider the noiseless case. Decompose $\mathbf{y}_K(k)$ in (5) as

$$\mathbf{y}_K(k) = \mathbf{F}\mathbf{x}(k) = \begin{pmatrix} \mathbf{F}_d & \hat{\mathbf{F}}_d \end{pmatrix} \begin{pmatrix} \mathbf{x}_d(k) \\ \hat{\mathbf{x}}_d(k) \end{pmatrix}, \quad (10)$$

where $\mathbf{F}_d$ contains the first $d+1$ columns of $\mathbf{F}$, and $\hat{\mathbf{F}}_d$ contains the rest of the columns of $\mathbf{F}$. The subvector $\mathbf{x}_d$ contains the first $d+1$ elements of $\mathbf{x}$, while $\hat{\mathbf{x}}_d(k)$ contains the rest of the elements in $\mathbf{x}$. We have the following theorem.

Theorem 1. If $\mathbf{P}_d$ is the optimal prediction matrix that minimizes $tr\{E[\boldsymbol{\varepsilon}^{(d)}(k)][\boldsymbol{\varepsilon}^{(d)}(k)]^H\}$, then

$$\boldsymbol{\varepsilon}^{(d)}(k) = \mathbf{F}_d \mathbf{x}_d(k), \quad (11)$$

$$\min E\{[\boldsymbol{\varepsilon}^{(d)}(k)][\boldsymbol{\varepsilon}^{(d)}(k)]^H\} = \mathbf{F}_d \mathbf{F}_d^H, \quad (12)$$

where the length of the predictor is pM with $M = \max(n_d - 1; n_g - n_d)$ being sufficient to achieve the optimal prediction, while n_g and n_d denote the order of $G(z)$ and the maximal order of $D^l(z)$, respectively.

Proof) See the appendix. □

Remark 1. An FIR SIMO model is equivalent to an IIR SIMO model with $n_g = 0$. Blind equalization of such FIR SIMO models was considered in [8] and [9], where a similar theorem was derived (Theorem 2 in [8]) under the assumption that $\mathbf{F}$ has full column rank. On the other hand, theorem 1 in this paper is derived differently, where a general IIR channel is considered, and $\mathbf{F}$ may not have full column rank.

Remark 2. The derivation of the optimal predictor is motivated by [18], and we derive an optimal predictor with tighter sufficient predictor length. This is because a more efficient predictor is used. The sufficient length of the optimal predictor obtained in this paper is smaller than that in [18] which equals $p \cdot (n_g + n_d - 1)$.

Remark 3. Theorem 1 suggests a method for estimating $\mathbf{f}_d$. Consider two predictors with delays d and d-1. From (11), we have $\boldsymbol{\varepsilon}^{(d)}(k)-\boldsymbol{\varepsilon}^{(d-1)}(k)=\mathbf{f}^d x(k-d)$, which is an estimate of $\mathbf{f}^d$ with unknown scalar $x(k–d)$. Statistics can be exploited to reduce the effect of noise. From (12), we have

$$E\{\boldsymbol{\varepsilon}^{(d)}(k)[\boldsymbol{\varepsilon}^{(d)}(k)]^H\}-E\{\boldsymbol{\varepsilon}^{(d-1)}(k)[\boldsymbol{\varepsilon}^{(d-1)}(k)]^H\}$$
$$=\mathbf{f}^d[\mathbf{f}^d]^H \overset{def}{=} \mathbf{E}. \tag{13}$$

Then, $\mathbf{f}^d$ can be estimated as the eigenvector of $\mathbf{E}$ which corresponds to the unique non-zero eigenvalue.

Remark 4. The selection of predictor length requires the values of n_g and n_d, which are unknown in practice. A longer predictor will not significantly increase the computation complexity. We argue that a very rough estimate of model order is enough. According to the model transfer in [19], we have $\{n_d, n_g\} \le n_h$ with n_h as the order of the original SISO channel. Usually, a third-order IIR model can be applied to model most practical channels. When applying the proposed algorithm, it is safe to set the predictor length to M=3 or M=4.

We estimate $\mathbf{f}^d$ based on (13). Consider another LP problem with delay $d-1$:

$$\boldsymbol{\varepsilon}^{(d-1)}(k)=\begin{pmatrix}\mathbf{I}_{pK} & -\mathbf{P}_{d-1}\end{pmatrix}\begin{pmatrix}\mathbf{y}_K(k)\\ \mathbf{y}_M(k-d)\end{pmatrix}. \tag{14}$$

Equations (8) and (14) imply that

$$E\{[\boldsymbol{\varepsilon}^{(d)}(k)][\boldsymbol{\varepsilon}^{(d)}(k)]^H\}$$
$$=\begin{pmatrix}\mathbf{I}_{pK} & -\mathbf{P}_{d-1}\end{pmatrix}\mathbf{R}^{(d)}\begin{pmatrix}\mathbf{I}_{pK} & -\mathbf{P}_{d-1}\end{pmatrix}^H, \tag{15}$$

$$E\{[\boldsymbol{\varepsilon}^{(d-1)}(k)][\boldsymbol{\varepsilon}^{(d-1)}(k)]^H\}$$
$$=\begin{pmatrix}\mathbf{I}_{pK} & -\mathbf{P}_{d-1}\end{pmatrix}\mathbf{R}^{(d\text{-}1)}\begin{pmatrix}\mathbf{I}_{pK} & -\mathbf{P}_{d-1}\end{pmatrix}^H, \tag{16}$$

where

$$\mathbf{R}^{(\mathrm{d})}=E\{\begin{pmatrix}\mathbf{y}_K(k)\\ \mathbf{y}_M(k-d-1)\end{pmatrix}\begin{pmatrix}\mathbf{y}_K^H & \mathbf{y}_M^H(k-d-1)\end{pmatrix}\}, \tag{17}$$

$$\mathbf{R}^{(d\text{-}1)}=E\{\begin{pmatrix}\mathbf{y}_K(k)\\ \mathbf{y}_M(k-d)\end{pmatrix}\begin{pmatrix}\mathbf{y}_K^H & \mathbf{y}_M^H(k-d)\end{pmatrix}\}. \tag{18}$$

Partition the matrices $\mathbf{R}^{(d)}$ and $\mathbf{R}^{(d\text{-}1)}$ as

$$\mathbf{R}^{(d)}=\begin{pmatrix}\mathbf{U}_{11} & \mathbf{U}_{12}\\ \mathbf{U}_{21} & \mathbf{U}_{22}\end{pmatrix},\quad \mathbf{R}^{(d\text{-}1)}=\begin{pmatrix}\mathbf{W}_{11} & \mathbf{W}_{12}\\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{pmatrix}, \tag{19}$$

where $\mathbf{U}_{11}$ and $\mathbf{W}_1$ are of size $pK\times pK$, while $\mathbf{U}_{12}$ and $\mathbf{W}_{12}$ are of size $pM\times pM$. The optimal prediction matrices that minimize the prediction error variance are given as in [9] by

$$\mathbf{P}_d=\mathbf{U}_{12}\mathbf{U}_{22}^{+},\quad \mathbf{P}_{d-1}=\mathbf{W}_{12}\mathbf{W}_{22}^{+}. \tag{20}$$

Note that $\mathbf{U}_{11}=\mathbf{W}_{11}$ and $\mathbf{U}_{22}=\mathbf{W}_{22}$. Substituting (20) into (15) and (16) yields

$$\min E\{[\boldsymbol{\varepsilon}^{(d)}(k)][\boldsymbol{\varepsilon}^{(d)}(k)]^H\}$$
$$=\mathbf{U}_{11}-\mathbf{U}_{12}\mathbf{U}_{22}\mathbf{U}_{21}=\mathbf{F}_d\mathbf{F}_d^H, \tag{21}$$

$$\min E\{[\boldsymbol{\varepsilon}^{(d-1)}(k)][\boldsymbol{\varepsilon}^{(d-1)}(k)]^H\}$$
$$=\mathbf{U}_{11}-\mathbf{W}_{12}\mathbf{W}_{22}\mathbf{W}_{21}=\mathbf{F}_{d-1}\mathbf{F}_{d-1}^H. \tag{22}$$

It follows that

$$\mathbf{E}=\mathbf{f}^d[\mathbf{f}^d]^H=\mathbf{W}_{12}\mathbf{U}_{22}^{+}\mathbf{W}_{21}-\mathbf{U}_{12}\mathbf{U}_{22}^{+}\mathbf{U}_{21}. \tag{23}$$

Since $\mathbf{W}_{12}$, $\mathbf{U}_{22}$, $\mathbf{W}_{21}$, $\mathbf{U}_{12}$, and $\mathbf{U}_{21}$ are all estimated from the output signal $y(k)$, $\mathbf{f}^d$ can be estimated from (23) by Chelosky decomposition or eigendecomposition. Let λ denote the unique non-zero eigenvalue of $\mathbf{E}$, and let $\boldsymbol{\beta}$ denote the corresponding eigenvector. Then, $\mathbf{f}^d$ can be estimated as

$$\mathbf{f}^d=\lambda\boldsymbol{\beta}. \tag{24}$$

The MMSE equalizer with delay d can be computed from (7) with the above partial model estimation. Note that an FIR SIMO model is a special case of the IIR SIMO model. The MMSE equalizer is also applicable to SISO FIR channels.

To obtain a more reliable estimate of $\mathbf{f}^d$ in noisy cases, we need to remove the noise effect from $\mathbf{W}$, $\mathbf{R}_M$, and $\mathbf{U}$. The method in [18] is applied (see (3.58) in [18]), which requires estimation of the noise variance. If the noise free correlation matrix is rank deficient, then the noise variance can be estimated as the average of the smallest eigenvalues of the noise corrupted correlation matrix. We have the following theorem regarding the rank of the noise free correlation matrix.

Theorem 2. The rank of the $pN\times pN$ noiseless correlation matrix $\mathbf{R}_N$ equals $\rho(\mathbf{R}_N)\le N+n_d$.

Proof. See [1]. □

In the following simulations, the noise variance is estimated by performing EVD on $\mathbf{U}_{22}$, which is of size $pM\times pM$. It was shown in [18] that the rank of noiseless $\mathbf{R}_{pM\times pM}$ is upper bounded by $p(M–1)+1$. As a result, only $M–1$ eigenvalues were used to estimate the noise variance. In the following simulations, M is selected so as to satisfy $M \ge n_d-1$. Theorem 2 implies that we can use at least $(p–2)M+1$ eigenvalues to estimate the noise variance, where more precise estimation than that presented in [18] is expected.

IV. Adaptive Algorithms

In the previous section, we presented the derivation of a blind MMSE equalizer for an IIR SISO channel using oversampling and LP by batch processing. The implementation of (7) requires the computation of EVD for $\mathbf{E}$ and $\mathbf{U}_{22}$ and the matrix inversion of $\mathbf{R}_K$ and $\mathbf{U}_{22}$. Note that the inversion of $\mathbf{U}_{22}$

can be computed via EVD. Further computational complexity reduction can be achieved if we set $K = M$, where only two EVDs (for $\mathbf{E}$ and $\mathbf{R}_K$) are required. However, the computational complexity is still too high, and this limits its application in time-varying channels. In the following subsection, we derive the blind MMSE equalizer for IIR SISO channel by adaptive processing to avoid the computation of EVD.

1. RLS Adaptive Equalizer

The blind MMSE equalizer can be implemented by adaptive processing such that the equalizer is updated in a similar fashion as the RLS adaptive filter. This can be done by updating the SOS of $\mathbf{R}_K$, $\mathbf{U}_{12}$, $\mathbf{U}_{21}$, $\mathbf{W}_{12}$, $\mathbf{W}_{21}$, and $\mathbf{U}_{22}$ using an RLS algorithm. For each time slot k, the sample vectors $\mathbf{y}_K(k)$, $\mathbf{y}_K(k–d)$, and $\mathbf{y}_K(k–d–1)$ are used to update the statistics of $\mathbf{U}_{12}$, $\mathbf{U}_{21}$, $\mathbf{W}_{12}$, $\mathbf{W}_{21}$, and $\mathbf{U}_{22}$ as well as the partial model parameters according to (24). Then, $\mathbf{R}_K$ is updated, and the equalizer is re-estimated according to (7). The algorithm comprises the following steps.

Step 1. Initialization.

$$\mathbf{U}_{12}(0)=0,\quad \mathbf{U}_{21}(0)=0,\quad \mathbf{W}_{12}(0)=0,\quad \mathbf{W}_{21}(0)=0$$
$$\mathbf{R}_K^{-1}(0)=\delta\mathbf{I}_{pK\times pK},\quad \mathbf{U}_{22}^{-1}(0)=\delta\mathbf{I}_{pM\times pM}$$

where δ is a small positive constant.

Step 2. For each time instant $k = 1, 2, \cdots$, update the matrices $\mathbf{U}_{12}$, $\mathbf{U}_{21}$, $\mathbf{W}_{12}$, $\mathbf{W}_{21}$, and $\mathbf{U}_{22}$ as

$$\mathbf{U}_{12}(k)=\lambda\mathbf{U}_{12}(k-1)+\mathbf{y}_K^H(k)\mathbf{y}_M(k-d-1), \tag{25}$$

$$\mathbf{U}_{21}(k)=\lambda\mathbf{U}_{21}(k-1)+\mathbf{y}_M^H(k-d-1)\mathbf{y}_K(k), \tag{26}$$

$$\mathbf{W}_{12}(k)=\lambda\mathbf{W}_{12}(k-1)+\mathbf{y}_K^H(k)\mathbf{y}_K(k-d), \tag{27}$$

$$\mathbf{W}_{21}(k)=\lambda\mathbf{W}_{21}(k-1)+\mathbf{y}_M^H(k-d)\mathbf{y}_K(k), \tag{28}$$

$$\mathbf{v}_2(k)=\frac{\lambda^{-1}\mathbf{U}_{22}^{-1}(k-1)\mathbf{y}_M(k-d)}{1+\lambda^{-1}\mathbf{y}_M^H(k-d)\mathbf{U}_{22}^{-1}(k-1)\mathbf{y}_M(k-d)}, \tag{29}$$

$$\begin{aligned}\mathbf{U}_{22}^{-1}(K)=&\lambda^{-1}\mathbf{U}_{22}^{-1}(k-1)\\&-\lambda^{-1}\mathbf{v}_2^H(k)\mathbf{y}_M(k-d)\mathbf{U}_{22}^{-1}(k-1).\end{aligned} \tag{30}$$

Step 3. Estimate the partial channel parameter.

$$\mathbf{E}(k)=\mathbf{U}_{12}(k)\mathbf{U}_{22}^{-1}(k)\mathbf{U}_{21}(k)-\mathbf{W}(k)\mathbf{U}_{22}^{-1}(k)\mathbf{W}_{21}(k). \tag{31}$$

Estimate $\mathbf{f}^d(k)$ as the column in $\mathbf{E}$ with the largest norm [8].

Step 4. Update $\mathbf{R}_K$ and estimate the equalizer.

$$\mathbf{v}_1(k)=\frac{\lambda^{-1}\mathbf{R}_K^{-1}(k-1)\mathbf{y}_K(k)}{1+\lambda^{-1}\mathbf{y}_K^H(k)\mathbf{R}_K^{-1}(k-1)\mathbf{y}_K(k)}, \tag{32}$$

$$\mathbf{R}_K^{-1}(k)=\lambda^{-1}\mathbf{R}_K^{-1}(k-1)-\lambda^{-1}\mathbf{v}_1^H(k)\mathbf{y}_K(k)\mathbf{R}_K^{-1}(k-1), \tag{33}$$

$$\mathbf{w}_{RLS}(k)=\mathbf{R}_K^{-1}(k)\mathbf{f}^d(k). \tag{34}$$

Here, λ is the forgetting factor. The proper selection of λ was discussed in [8]. Although different λ values can be used in each update equation from (25) to (33), for simplicity, the same λ value is used. We found that the performance gain achieved by using different λ values is close to that of using the same λ in the above algorithm.

An RLS adaptive equalizer based on the zero-forcing criterion was also proposed in [8], while the proposed algorithm is MMSE optimal. Furthermore, the algorithm in [8] requires the computation of two prediction matrices, whereas the proposed RLS algorithm directly updates the correlation matrices. The computational complexity of the above algorithm is approximately on the order of $O(2(K+M)KMp^3+4(M^2+KM)p^2)$, which is similar to that of the algorithm in [8].

2. LMS Adaptive Equalizer

The blind MMSE equalizer can also be updated in a similar fashion as that in the LMS adaptive filter, where the prediction matrix, the partial channel parameters, and the equalizer are updated in a sequential manner. The algorithm comprises the following steps.

Step 1. Update the prediction matrices.

$$\boldsymbol{\varepsilon}_d(k)=\mathbf{y}_K(k)-\mathbf{P}_d(k)\mathbf{y}_K(k-d-1), \tag{35}$$

$$\boldsymbol{\varepsilon}_{d-1}(k)=\mathbf{y}_K(k)-\mathbf{P}_{d-1}(k)\mathbf{y}_K(k-d), \tag{36}$$

$$\mathbf{P}_d(k)=\mathbf{P}_d(k-1)+\mu_1\boldsymbol{\varepsilon}_d(k)\mathbf{y}_K(k-d-1), \tag{37}$$

$$\mathbf{P}_{d-1}(k)=\mathbf{P}_{d-1}(k-1)+\mu_1\boldsymbol{\varepsilon}_{d-1}(k)\mathbf{y}_K(k-d). \tag{38}$$

Step 2. Update the partial channel estimate.

$$\mathbf{E}(k)=(\boldsymbol{\varepsilon}_d(k)\boldsymbol{\varepsilon}_d^H(k)-\boldsymbol{\varepsilon}_{d-1}(k)\boldsymbol{\varepsilon}_{d-1}^H(k)). \tag{39}$$

$$\begin{aligned}\mathbf{f}^d(k)=&\mathbf{f}^d(k-1)\\&+2\mu_2\left[\mathbf{E}(k)\mathbf{f}^d(k-1)-\frac{[\mathbf{f}^d(k-1)]^H\mathbf{E}(k)\mathbf{f}^d(k-1)\mathbf{f}^d(k-1)}{[\mathbf{f}^d(k-1)]^H\mathbf{f}^d(k-1)}\right].\end{aligned} \tag{40}$$

Note that $\mathbf{f}^d$ is estimated as the eigenvector corresponding to the largest eigenvalue of $\mathbf{E}(k)$.

$$\mathbf{f}^d(k)=\max_f\frac{\mathbf{f}^H\mathbf{E}(k)\mathbf{f}}{\mathbf{f}^H\mathbf{f}}, \tag{41}$$

which immediately leads to the update equation (40).

Step 3. Update the equalizer.

$$\mathbf{w}_{LMS}(k) = \mathbf{w}_{LMS}(k-1) - 2\mu_3 \mathbf{y}_K(k)\mathbf{y}_K^H(k-d)\mathbf{w}_{LMS}(k-1) - \mathbf{f}^d(k), \quad (42)$$

where μ_1, μ_2, and μ_3 are the learning rates, which should be selected according to the landscape of the cost functions.

In our simulations, we use the same learning rate for simplicity. The complexity of the LMS equalizer is approximately in the order of $O(5K^2p^2 + 4KMp^2)$, which is much lower than that of the RLS equalizer, and the LME equalizer in [9].

V. Simulation Results

The performance of the proposed MMSE equalizer is compared to the zero-forcing equalizer presented in [9], and the MMSE equalizer presented in [18]. Simulations are performed with 16 QAM i.i.d. input signal $x(n)$ on random channels corrupted by AWGN at various signal to noise ratios (SNRs), where the SNR is defined as

$$SNR = 10\log(E\|x(k)\|^2 / E\|v(k)\|^2) \text{ (dB)}. \quad (43)$$

The performance of the equalizers is measured by the mean-square error (MSE) between the input signal and the equalized signals, which is defined as

$$MSE = E[\|x(k) - \hat{x}(k)\|^2], \quad (44)$$

where $\hat{x}(k)$ is the equalized signal.

All the equalizers considered in the simulations suffer from a scalar ambiguity problem, which makes it impossible to compute the MSE. In practice, the scalar ambiguity problem can be resolved by differential coding. For simplicity, we assume that the first transmitted symbol is known to the receiver, that is, that one pilot symbol is being sent to resolve the scalar ambiguity problem.

All the batch equalizers considered in the simulations require the estimation of the noise free correlation matrices, such as $\mathbf{R}_d$ and $\mathbf{R}_{d-1}$ in (19), where the method presented in [18] is used. To implement the proposed equalizer, the noise variance is estimated as the average of the $(p-2)M+1$ smallest eigenvalues of the $pM \times pM$ correlation matrix $\mathbf{U}_{22}$. When implementing the method in [18], only the M-1 smallest eigenvalues are used (as suggested in [18]). In a noisy case, the estimated $\mathbf{U}_{12}$, $\mathbf{U}_{21}$, $\mathbf{W}_{12}$, and $\mathbf{W}_{21}$ may also be corrupted by AWGN. In this case, the estimated noise variance is used to suppress the noise component.

1. FIR Channels

The proposed algorithms are applicable to both FIR and IIR channels. We first consider FIR channels. An FIR channel obtained from microwave channel measurement with real and imaginary parts, as shown in Fig. 1, is applied in this set of simulations. The SIMO channel model was obtained by oversampling the FIR channel output as discussed in section I. In the simulations, the oversampling ratio $p = 3$, such that the FIR channel is transformed to a SIMO system with 3 subchannels. The delay was selected to be 1. That is, $d = 1$. The length of the predictor was set to $pM = 3\times4=12$, and the length of the equalizer was set to $pK = 3\times3 = 9$.

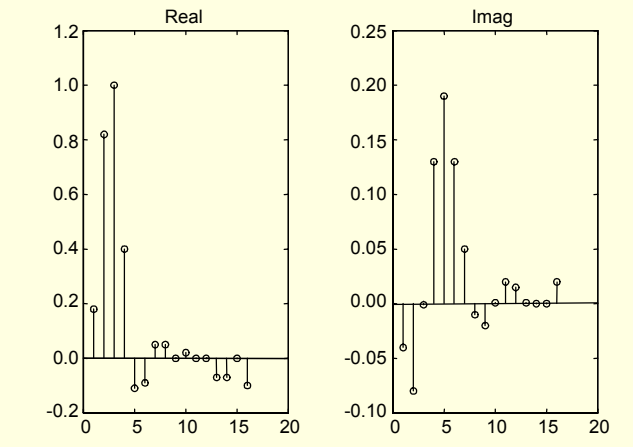

Fig. 1. Real and imaginary parts of the measured FIR channel.

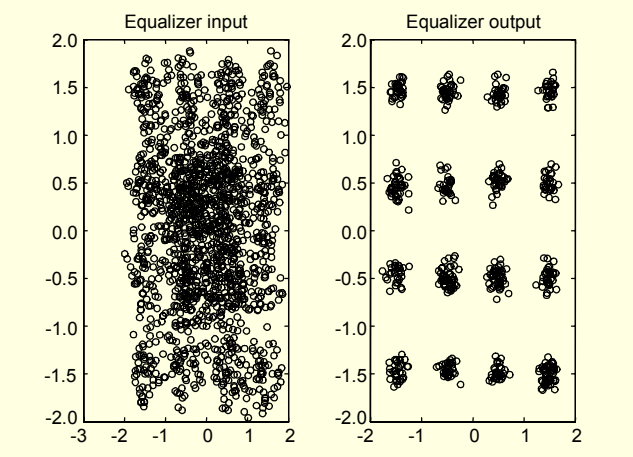

Fig. 2. Eye diagram of the proposed MMSE equalizer with 500 symbols (SNR=25 dB).

The received signal constellation and the equalized signal constellation for 500 symbols at an SNR of 25 dB are shown in Fig. 2. The proposed equalizer successfully suppresses the ISI caused by the channel.

To compare the performance of the batch equalizers, the average of the MSE obtained from 150 independent trails are plotted in Fig. 3. The zero-forcing equalizer is the most sensitive to additive noise. The proposed MMSE equalizer outperforms the zero-forcing equalizer at low SNRs. At high

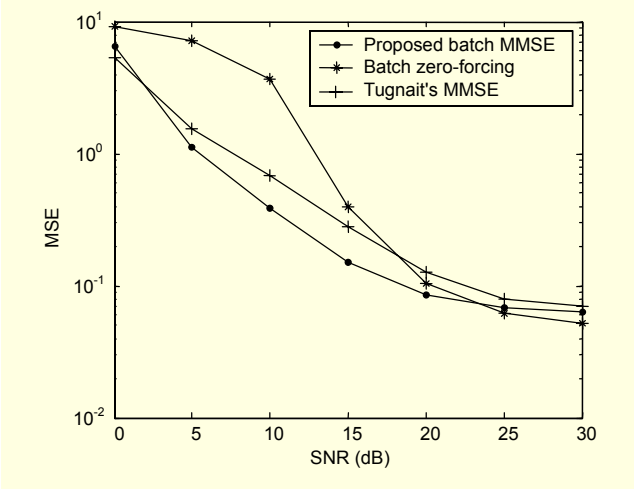

Fig. 3. MSE comparison of the batch equalizers in the measured FIR channel for 150 Monte Carlo runs with 500 symbols.

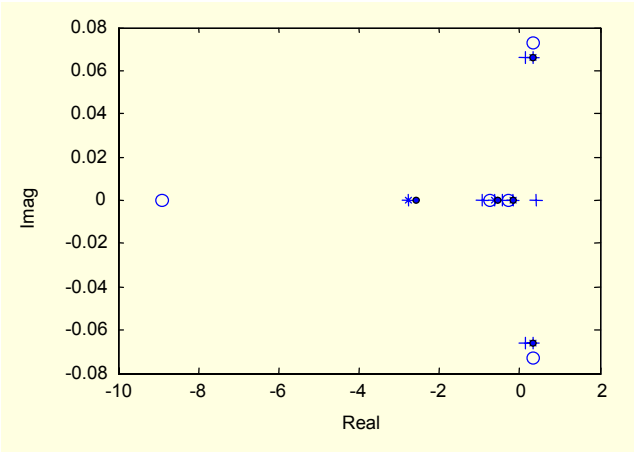

Fig. 5. Zeros of the subchannel of the ill-conditioned FIR channel, ∗, o, +, and · denote the zero locations of each subchannel.

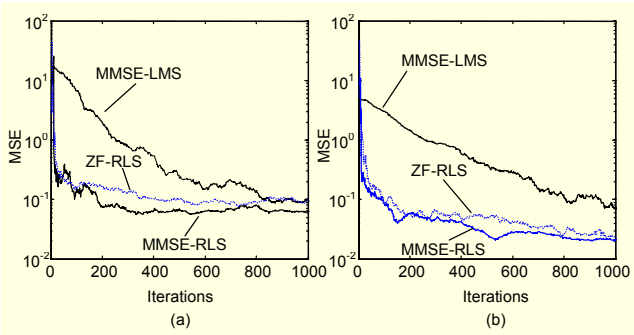

Fig. 4. Learning curves of the adaptive equalizers for the measured FIR channel: (a) SNR=15 dB and (b) SNR=25 dB.

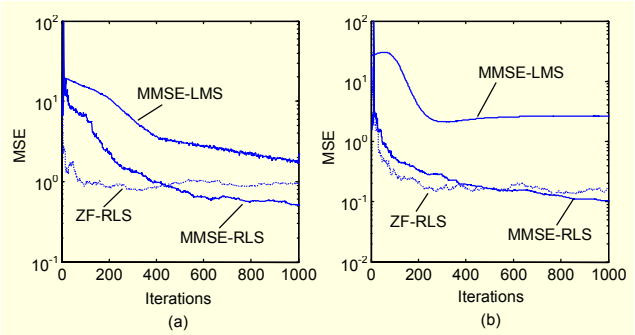

Fig. 6. Learning curves of the adaptive equalizers in the ill-conditioned FIR channel: (a) SNR=15 dB and (b) SNR=25 dB.

SNRs, all equalizers considered in the simulations show similar performance.

It can also be observed that the proposed MMSE equalizer exhibit better performance than that of the MMSE equalizer in [18]. Even though both algorithms are MMSE optimal, different approaches are used to estimate the partial channels. The better performance of the proposed algorithm is because a general prediction model is used in (8); however, the prediction model in [18] can be considered a special case of that in (7) with K=1. As a result, the columns of the channel transfer matrix $\mathbf{F}$ considered in [18] contain only one slot of the impulse response. Multi-step predictions are required to estimate the impulse response of multiple time slots; therefore, the method in [18] is more complicated. Secondly, more eigenvalues are used to estimate the noise variance, which results in a more accurate estimation [18].

The performance of the proposed MMSE-based RLS and LMS adaptive algorithms at SNRs of 15 dB and 25 dB is shown in Fig. 4. The forgetting factor λ = 0.998 and the learning rate μ_1=μ_2=μ_3=0.0009. Compared with that obtained from the RLS equalizer in [8] (denoted as ZF-RLS), both RLS equalizers are able to achieve faster convergence than the LMS equalizer. In particular, the proposed RLS equalizer outperforms the ZF RLS equalizer in [8] at low SNRs. However, they have similar MSE performance at high SNRs.

We further evaluate the performance of each equalizer with an ill-conditioned channel, that is, a channel with near common zeros among subchannels. The ill-conditioned channel in [9] is considered in the simulations, where the impulse response of the channel is given by

$$h(t) = c(t, 0.45)W(t) + 0.8c(t - 0.25T, 0.45)W(t - 0.25T) - 0.4c(t - 2T, 0.45)W(t - 2T), \quad (45)$$

where $c(t, 0.45)$ is the raised roll-off cosine window with a roll-off factor of 0.45, and $W(t)$ is a rectangular truncation window spanning [-0.85T, 5.14T]. A discrete channel of order 23 is obtained by sampling $h(t)$ with a interval of 1. The oversampling ratio was set to p=4. The zeros of each

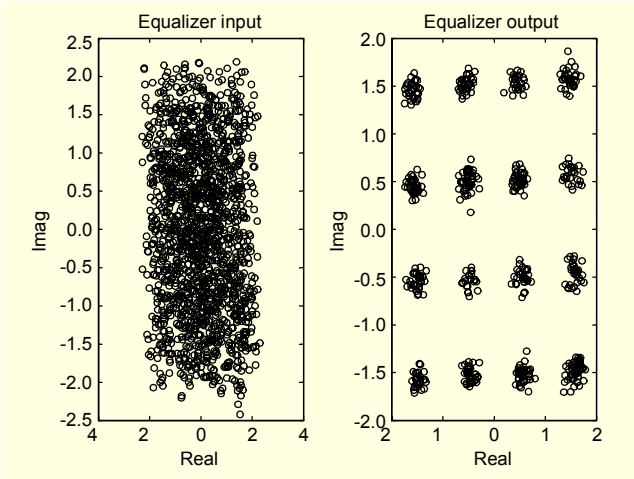

Fig. 7. Eye diagram of the proposed MMSE equalizer in the IIR channel with 500 symbols (SNR=25 dB).

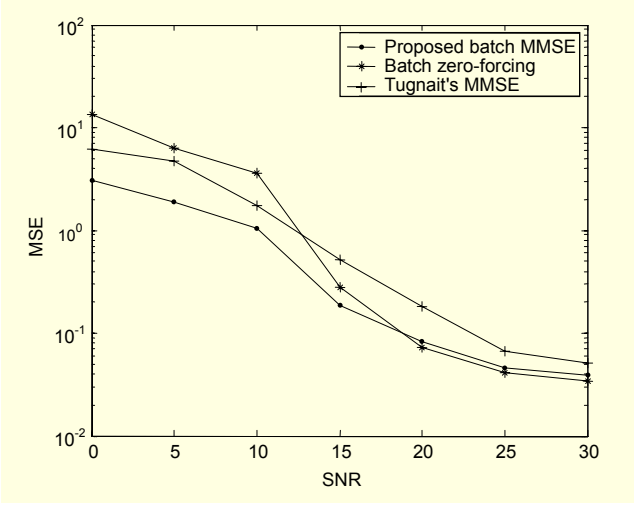

Fig. 8. MSE comparison of the batch equalizers in the IIR channel for 150 Monte Carlo runs with 500 symbols.

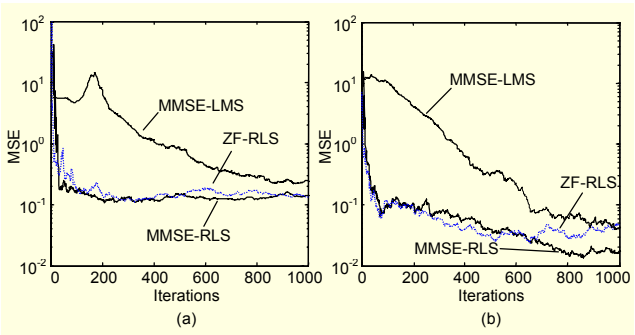

Fig. 9. Learning curves of the adaptive equalizers in the IIR channel: (a) SNR=15 dB and (b) SNR=25 dB.

subchannel are plotted in Fig. 5.

The learning curves of the three adaptive algorithms considered in this paper are shown in Fig. 6 for SNRs of 15 dB and 25 dB. It can be observed that the LMS equalizer has slow convergence in the ill-conditioned channel, and it does not converge to the correct equalizer at an SNR of 25 dB. The ZF-RLS equalizer has faster convergence than the proposed MMSE-RLS equalizer, but it has a higher steady-state MSE.

2. IIR Channels

Simulations of the proposed algorithms using randomly selected IIR channel are performed. An example of the random IIR channel is the following:

$$H(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} = \frac{1-0.5z^{-1}+0.14z^{-2}+0.1z^{-3}}{1+0.3z^{-1}-0.22z^{-2}+0.1z^{-3}}. \quad (46)$$

The constellation diagram before and after equalization at an SNR of 25 dB is shown in Fig. 7. The input to the equalizer is severely corrupted by ISI when compared to that of the FIR channel case. The proposed equalizer can efficiently suppress the ISI effect as observed in Fig. 7.

The MSE performance of the proposed batch equalizers are shown in Fig. 8 with oversampling ratio $p = 3$ at various SNRs. The length of the equalizers is $pK = 3\times3 = 9$, and the length of the predictors is $pM = 3\times4 = 12$. The packet size of the channel input is 500. The corresponding channel output is used to estimate the correlation matrices. Theorem 1 implies that the ZF equalizer in [9] is also applicable to IIR channels, and its result is also shown in Fig. 8. The MSE performance shown in Fig. 8 is similar to that shown in Fig. 3. We can conclude that the zero-forcing equalizer in [9] is the most sensitive to noise for both FIR and IIR channels, while the proposed MMSE equalizer can achieve the best performance at low to medium SNRs for both FIR and IIR channels.

The learning curves of different adaptive equalizers are shown in Fig. 9 at SNRs of 15 dB and 25 dB with the same forgetting factor and learning rate as in the FIR cases. The LMS equalizer has slow convergence and high steady state MSE. At low SNRs, the ZF-RLS and MMSE-RLS equalizers achieve fast convergence and have similar levels of steady-state MSE. At high SNRs, the MMSE-RLS has slower convergence than ZF-RLS but can achieve a lower steady-state MSE.

VI. Conclusion

A new blind equalization algorithm was proposed for oversampled SISO FIR/IIR channels. Both batch and adaptive implementations were presented. This paper also proposed a general partial model estimation algorithm for IIR channels as well as for FIR channels, which in fact can be applied to other applications, such as the derivation of the ZF equalizer in [8]. More importantly, this paper extended the work of [8], [9], and

[18] to the SISO IIR channel model by applying a more efficient presentation of linear prediction. We demonstrated that the algorithms in [8], [9], and [18] are applicable to blind equalization of SISO IIR channels.

Appendix. Proof of Theorem 1

We first consider K=1 and d=0. Denote $U(z) = [1/G(z)]X(z)$. We assume that $G(z)$ is monic here. Based on (4), $\mathbf{y}(k)$ can be represented as

$$\mathbf{y}(k) = \mathbf{d}_0(-\sum_{i=1}^{n_g} g_i u(k-i) + x(k)) + \sum_{i=1}^{n_d} \mathbf{d}_i u(k-i) \quad \text{(A1)}$$

$$= \sum_{i=1}^{n_g} \mathbf{r}_i u(k-i) + \sum_{i=1}^{n_d} \mathbf{d}_i u(k-i) + \mathbf{d}_0 x(k) \quad \text{(A2)}$$

$$= \sum_{i=1}^{\max(n_g, n_d)} \boldsymbol{\theta}_i u(k-i) + \mathbf{d}_0 x(k) \quad \text{(A3)}$$

$$\overset{def}{=} \mathbf{s}(k) + \mathbf{d}_0 x(k), \quad \text{(A4)}$$

where $\mathbf{d}_i = [d^0(i), \cdots, d^{p-1}(i)]^T$, $i = 0, \cdots, n_d$, $\mathbf{r}_i = -\mathbf{d}_0 g_i$, and $\boldsymbol{\theta}_i = \mathbf{r}_i + \mathbf{d}_i$. We show that $\mathbf{s}(k)$ is the optimal prediction of $\mathbf{y}(k)$ and $\mathbf{d}_0 x(k)$ is the prediction error.

Denote $\mathbf{y}_N(k) = [\mathbf{y}^T(k), \cdots, \mathbf{y}^T(k-N+1)]^T$ and $\mathbf{u}_{N+n_d}(k) = [u(k), \cdots, u(k-N-n_d+1)]$. It has been proved that if $D^i(z^{-1})$'s do not share a common zero and $N \geq n_d - 1$ then $\mathbf{y}_N(k)$ and $u_{N+n_d}(k)$ span the same Hilbert space (theorem 1 in [6]).

$$H_{k,N}(\mathbf{y}) = H_{k,N+n_d}(u), \quad \text{(A5)}$$

where $H_{k,N}(\mathbf{y}) = \text{sp}\{\mathbf{y}(k-i) \mid i = 0, \cdots, N-1\}$, $H_{k,N+n_d}(u) = \text{sp}\{u(k-i) \mid i = 0, \cdots, N+n_d-1\}$. Let $N = n_d - 1$, and apply (A3) to $\mathbf{y}(k-1)$. It is straightforward that if $\max(n_g, n_d) = n_d$ or $n_d \leq n_g \leq 2n_d - 1$, $\mathbf{s}(k) \in H_{k-1,2n_d-1}(u) = H_{k-1,N+n_d}(u)$, and consequently, $\mathbf{s}(k) \in H_{k-1,n_d-1}(\mathbf{y})$. It follows that

$$\mathbf{s}(k) = \sum_{i=1}^{n_d-1} \mathbf{A}_i \mathbf{y}(k-i), \quad \text{(A6)}$$

for some parameter matrices $\mathbf{A}_i$. When $n_g \geq 2n_d$, let $N = n_g - n_d$, and we have $\mathbf{s}(k) \in H_{k-1,n_g}(u) = H_{k-1,N+n_d}(u)$. It follows that $\mathbf{s}(k) \in H_{k-1,n_g-n_d}(y)$, that is,

$$\mathbf{s}(k) = \sum_{i=1}^{n_g-n_d} \mathbf{A}_i \mathbf{y}(k-i) \quad \text{(A7)}$$

for some parameter matrices $\mathbf{A}_i$. The relation of $\mathbf{s}(k)$ and $\mathbf{y}(k)$ finally can be presented as

$$\mathbf{s}(k) = \sum_{i=1}^{\max(n_d-1, n_g-n_d)} \mathbf{A}_i \mathbf{y}(k-i) \quad \text{(A8)}$$

for properly selected parameter matrices $\mathbf{A}_i$. Equation (A8) presents the prediction of $\mathbf{y}(k)$ using the past data. It is straightforward that the prediction error $\mathbf{d}_0 x(k)$ is orthogonal to $\mathbf{y}(k-i)$ for i = 1, 2,⋯. Therefore, $\mathbf{P}_0 = [\mathbf{A}_1, \ \mathbf{A}_2, \ \cdots]$ is the optimal prediction matrix and $M = \max(n_d-1, n_g - n_d)$ is a sufficient predictor length to achieve the optimal prediction.

Now, we investigate the relation between $\mathbf{d}_0$ and $\mathbf{f}_0$. From (A2) and (A8) we have $E[x(k)\mathbf{y}(k)] = E[x(k)\mathbf{d}_0 x(k)]$. From (4) we have $E[x(k)\mathbf{y}(k)] = \mathbf{f}_0$, which follows $\mathbf{d}_0 = \mathbf{f}_0$. Next, we consider $K = 1$ and $d > 0$. Replacing $u(k-1)$ in (A2) with $u(k-1) = -\sum_{i=2}^{n_g+1} g_i u(k-i) + x(k-1)$, we have

$$\mathbf{y}(k) = \sum_{i=2}^{n_g+1} \mathbf{r}_i^{(1)} u(k-i) + \sum_{i=2}^{n_d} \mathbf{d}_i u(k-i) + \mathbf{d}_0 x(k) + \mathbf{r}_1^{(1)} x(k-1) \quad \text{(A9)}$$

for some properly selected parameter vectors $\mathbf{r}_i^{(1)}$. Similarly, replacing $u(k-i)$ for $i = 2, \cdots, d$, we have, in general, for appropriate choices of $\mathbf{r}_i^{(d)}$

$$\mathbf{y}(k) = \sum_{i=d+1}^{n_g+d} \mathbf{r}_i^{(d)} u(k-i) + \sum_{i=d+1}^{n_d} \mathbf{d}_i u(k-i) + \sum_{i=0}^{d} \mathbf{r}_i^{(d)} x(k-i)$$
$$\overset{def}{=} \mathbf{s}^{(d)}(k) + \boldsymbol{\varepsilon}^{(d)}(k), \quad \text{(A10)}$$

where $\mathbf{s}^{(d)}(k) = \sum_{i=d+1}^{\max(n_g+d, n_d)} \boldsymbol{\theta}_i^{(d)} u(k-i)$ with $\boldsymbol{\theta}_i^{(d)} = \mathbf{r}_i^{(d)} + \mathbf{d}_i$.

Similar to (A8), it can be derived that

$$\mathbf{s}^{(d)}(k) = \sum_{i=d+1}^{\max(n_d-1, n_g-n_d)+d} \mathbf{A}_i^{(d)} \mathbf{y}(k-i) \quad \text{(A11)}$$

for some properly chosen parameter matrix $\mathbf{A}_i^{(d)}$. It is easy to verify that $\boldsymbol{\varepsilon}^{(d)}(k)$ is orthogonal to $\mathbf{y}(k-i)$ for $i = d+1, d+2, \cdots$. Therefore, $\mathbf{s}^{(d)}(k)$ is the optimal prediction and $\boldsymbol{\varepsilon}^{(d)}(k)$ is the prediction error. By calculating $E[x(k-i)\mathbf{y}(k)]$ for $i = 0, \cdots, d$, we have $\mathbf{f}_i = \mathbf{r}_i^{(d)}$ (i = 0,⋯, d). Next, we consider $K > 1$ and $d > 0$. Applying (A8) for $\mathbf{y}(k-j)$ (j = 0,⋯, K–1) and using different delays, we have

$$\mathbf{y}(k-j) = \mathbf{s}^{(d-j)}(k-j) + \boldsymbol{\varepsilon}^{(d-j)}(k-j), \ j = 0, \cdots, K-1. \quad \text{(A12)}$$

The prediction can be presented as

$$\mathbf{s}^{(d-j)}(k-j) = \sum_{i=d-j+1}^{\max(n_d-1, n_g-n_d)+d-j} \mathbf{A}_i^{(d-j)} \mathbf{y}(k-j-i)$$
$$= \sum_{i=d+1}^{\hat{M}+d} \mathbf{A}_i^{(d-j)} \mathbf{y}(k-i), \ j = 0, \cdots, K-1, \quad \text{(A13)}$$

where $\hat{M} = \max(n_d - 1, n_g - n_d)$. Equation (A13) represents the prediction of $\mathbf{y}(k)$ to $\mathbf{y}(k-K+1)$, that is, $\mathbf{y}_K(k)$, using the same past data. We combine the equations in (A13) by defining the following matrices:

$$\bar{\mathbf{A}}_i = \begin{pmatrix} \mathbf{A}_i^{(d)} \\ \mathbf{A}_i^{(d-1)} \\ \vdots \\ \mathbf{A}_i^{(d-K+1)} \end{pmatrix}, \quad \bar{\mathbf{r}}_i = \begin{pmatrix} \mathbf{r}_i^{(d)} \\ \mathbf{r}_{i-1}^{(d-1)} \\ \vdots \\ \mathbf{r}_{i-K+1}^{(d-K+1)} \end{pmatrix}. \tag{A14}$$

Combining the equations in (A14) yields

$$\mathbf{y}_K(k) = \sum_{i=d+1}^{\hat{M}+d} \bar{\mathbf{A}}_i \mathbf{y}(k-i) + \sum_{i=0}^{d} \bar{\mathbf{r}}_i x(k-i) \tag{A15}$$

$$\stackrel{def}{=} \hat{\mathbf{s}}(k) + \mathbf{\Phi}\mathbf{x}_d(k), \tag{A16}$$

where $\mathbf{\Phi} = [\bar{\mathbf{r}}_0 \cdots \bar{\mathbf{r}}_d]$. Note that the channels considered in this paper are causal. We have $\mathbf{r}_{i-j}^{d} = 0$ for arbitrary d if $i-j < 0$. Hence, based on (5), It can be derived that $\mathbf{F}_d = \mathbf{\Phi}$, and consequently, we have $[\varepsilon^{(d)}(k)][\varepsilon^{(d)}(k)]^H = \mathbf{F}_d \mathbf{F}_d^H$. Note that (A13) also shows that the sufficient length of the optimal predictor is decided by $M = \hat{M} = \max(n_d - 1, n_g - n_d)$, which completes the proof.

References

[1] F. Chen et al., "Blind IIR Channel Equalization Based on Second-Order Statistics," *IEEE Vehicular Technology Conference*, May 2006, pp. 2334-2338.

[2] Z. Ding and Y. Li, *Blind Equalization and Identification*, Marcel Dekker, 2001.

[3] R. Johnson, "Admissibility in Blind Adaptive Channel Equalization," *IEEE Contr. Syst. Mag*., Jan. 1991, pp. 3-15.

[4] C.Y. Chi et al., "Batch Processing Algorithms for Blind Equalization Using Higher-Order Statistics," *IEEE Signal Processing Magazine*, vol. 20, no. 1, Jan. 2003, pp. 25-49.

[5] L. Tong et al., "Blind Identification and Equalization Based on Second-Order Statistics: A Frequency Domain Approach," *IEEE Tran. Information Theory*, vol. 41, no. 1, Jan. 1995, pp. 329-334.

[6] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction Error Method for Second-Order Blind Identification," *IEEE Trans. on Signal Processing*, vol. 45, no. 3, Mar. 1997, pp. 694-705.

[7] J. Shen and Z. Ding, "Direct Blind MMSE Channel Equalization Based on Second-Order Statistics," *IEEE Trans. on Signal Processing*, vol. 48, no. 4, Apr. 2000, pp. 1015-1023.

[8] X. Li and H. Fan, "Direct Estimation of Blind Zero-Forcing Equalizers Based on Second-Order Statistics," *IEEE Trans. on Signal Processing,* vol. 48, no. 8, Aug. 2000, pp. 2211-2218.

[9] X. Li and H. Fan, "Linear Prediction Methods for Blind Fractionally Spaced Equalization," *IEEE Trans. Signal Processing,* vol. 48, June 2000, pp. 1667-1675.

[10] G.B. Giannakis and S.D.Halford, "Blind Fractionally Spaced Equalization of Noisy FIR Channels: Direct and Adaptive *Solutions*," *IEEE Trans. Signal Processing,* vol. 45, Sept. 1997, pp. 2277-2292.

[11] W.B. Braun and U. Dersch, "A Physical Mobile Radio Channel Model," *IEEE Trans. Vehicular Technology*, vol. 40, no. 2, May 1991, pp. 472-482.

[12] A.J. Paulraj et al., "An Overview of MIMO Communication: A Key to Gigabit Wireless." *Proc. of the IEEE,* vol. 92, no. 2, Feb. 2004, pp. 198-218.

[13] A. Tkacenko and P.P. Vaidyanathan, "A Low-Complexity Eigenfilter Design Method for Channel Shortening Equalizers for DMT Systems," *IEEE Trans. Communication*, vol. 51, no. 7, July 2003, pp. 1069-1072, July 2003.

[14] S.E. El-Khamy, H. El-Ragal, and A.A, El-Sherif, "A Modified Elgen Approach Method for the Design of Channel Shortening Equalizers in Discrete Multi-tone Systems," *The 22th National Radio Science Conference, NRSC*, Mar. 2005, pp. 367-376.

[15] E. Bai and M. Fu, "Blind System Identification and Channel Equalization of IIR Systems Without Statistical Information," *IEEE Trans. Signal Process*, vol. 47, no. 7, July 1999, pp. 1910-1921.

[16] K. Vanbleu, M. Moonen, and G. Leus, "Linear and Decision Feedback Per Tone Equalization for DMT-Based Transmission Over IIR Channels," *IEEE Trans. Signal Processing*, vol. 54, no. 1, Jan. 2006, pp. 258-273.

[17] G.M. Raz and B.D. Van Veen, "Blind Equalization and Identification of Nonlinear and IIR Systems: A Least Squares Approach," *IEEE Trans. Signal Process*, vol. 48, no. 1, Jan. 2000, pp. 192-200.

[18] J.K. Tugnait, "Multistep Linear Predictors-Based Blind Equalization of FIR/IIR Single-Input Multiple-Output Channels with Common Zeros," *IEEE Trans. Signal Processing*, vol. 47, June 1999, pp. 1689-1700.

[19] F. Chen et al., "Blind Linear Channel Estimation Using Genetic Algorithm and SIMO Model," *Signal Processing*, vol. 83, no. 9, Sept. 2003, pp. 2021-2035.

[20] V. Sharma and V.N. Raj, "Convergence and Performance Analysis of Godard Family and Multimodulus Algorithms for Blind Equalization," *IEEE Trans. Signal Processing*, vol. 53, no. 4, Apr. 2005, pp. 1520-1533

[21] J. Yang, J.J. Werner, and G.A. Dumont, "The Multimodulus Blind Equalization and Its Generalized Algorithms," *IEEE J. Selected Areas in Commu.*, vol. 20, no. 5, June 2002, pp. 997-1015.

Fangjiong Chen received the BEng degree in information and electronics engineering from Zhejiang University, Hangzhou, Zhejiang, China, in 1997, and the PhD degree in electronics and information engineering from South China University of Technology, Guangzhou, Guangdong, China, in 2002. After graduation, Dr. Chen joined the school of electronics and information engineering, South China University of Technology. He was a lecturer from 2002 to 2005 and has been an associate professor since 2005. He worked as a senior research associate with the Department of Computer Science of the City University of Hong Kong from July 2006 to July 2007. His research focuses on physical layer technologies of wireless systems, including multicarrier modulation, intelligent antennas, MIMO, as well as channel estimation and equalization.

Sam Kwong received his BSc degree in electrical engineering from the State University of New York at Buffalo, USA, in 1983, and MASc degree in electrical engineering from the University of Waterloo, Canada, in 1985. In 1996, he obtained his PhD from the University of Hagen, Germany. From 1985 to 1987, he was a diagnostic engineer with Control Data Canada, where he designed the diagnostic software to detect manufacturing faults of the VLSI chips in the Cyber 430 machine. He later joined Bell Northern Research Canada as a member of scientific staff. In 1990, he joined the City University of Hong Kong as a lecturer with the Department of Electronic Engineering. He is currently an associate professor with the Department of Computer Science.

Chi-Wah Kok received the PhD degree in electrical engineering in 1996 from the University of Wisconsin, Madison, USA. He has been affiliated with Sony US Research Laboratory, Stanford University, Hong Kong Polytechnic University, Hong Kong University of Science and Technology, and the City University of Hong Kong. Currently, he is the executive director of Canaan Microelectronics Corp., Ltd. He has published more than 100 archival papers in the area of signal processing with applications in multimedia and digital communications and in the area of microelectronics in devices and circuits theories.