Blind System Identification and Channel Equalization of IIR Systems without Statistical Information

Erwei Bai and Minyue Fu, *Member, IEEE*

***Abstract*—A novel approach is proposed to blindly identify an unknown IIR system. The approach is based on faster sampling at the system output and requires neither *a priori* statistical information on the unknown input nor training signals. The methods presented are linear in the parameters of the unknown system so that many standard recursive algorithms can be readily applied. It is also shown in the paper that under a generic condition, any finite-order IIR system is identifiable, provided the oversampling ratio is appropriately chosen.**

***Index Terms*—Blind system identification, equalization, multirate systems, oversampling, system identification, wireless communications.**

I. INTRODUCTION

THE BLIND system identification (BSI) and blind channel equalization (BCE) problems addressed in this paper can be formulated as follows. A sequence of input signal $u[kh_i]$ is transmitted at sampling rate $f_i = 1/h_i$ to a continuous time system via an impulse generator or a zero-order hold. The received output signal $y[nh_o]$ is sampled at the rate $f_o = 1/h_o$. The BSI and BCE problems identify from $y[nh_o]$ both the system transfer function and the input $u[kh_i]$. These problems have received a lot of attention in recent years for applications in digital image processing, speech coding, and mobile communications such as cellular and cordless telephony, where communication channels drift constantly.

A traditional way to solve the BSI problem is to use statistical properties of the unknown input signal, e.g., high-order statistics [5], [8]. Although this method provides satisfactory results in many cases, it does require enough *a priori* statistical information on the unknown input, which is an assumption that may not be valid in certain applications. An alternative approach is to use training signals. For example, in the GSM standard, every 28 bits in a 116-bit sequence are used for the receiver to identify the channel; see Goodman [3]. This method, although it is simple, significantly reduces the transmission efficiency. To overcome this difficulty, the so-called oversampling (or multipath) approach has been proposed recently by Tong *et al.* [11], Slock [9], Xu *et al.* [12], Moulines *et al.* [7], Johnson *et al.* [4], and many others. It is shown that equalization can be achieved without training signals by using a multiple sampling rate for the receiver or, equivalently, using multichannels. This new approach, however, has to assume that the channel transfer function has a finite impulse response (FIR). Although it may often be reasonable to approximate a communication channel by an FIR function, such an approximation is often of very high order (up to 70th order FIR's are used [10]). Further, the order of the filter tends to increase as the sampling rate increases. In contrast, infinite impulse response (IIR) representation often requires far fewer parameters and, thus, leads to a simpler equalizer in some applications.

In this paper, we present a novel approach to the BSI and BCE problems that allows us to develop efficient algorithms without the FIR assumption on the channel, without any statistical information on the input signal, and without training signals. We also use the idea of oversampling, i.e., $f_o > f_i$. The main results of the paper can be summarized as follows:

1) First, we show that doubling the sampling rate for the receiver is sufficient for BSI and BCE when the channel is an all-pole filter. An algorithm is presented for blind identification of these systems. The algorithm is least-squares based and is convergent if the input is sufficiently rich.
2) Second, we show that BSI and BCE with an IIR channel can always be achieved when $f_o/f_i \geq n+1$, where n is the order of the channel transfer function. Since IIR models of the channel can often be of low order, this result offers an attractive alternative for the BSI and BCE problems. Again, least-squares-based algorithms that guarantee convergence under some persistent excitation (PE) conditions are given.
3) Finally, we address a general identifiability issue and give necessary and sufficient conditions for unique identification of both the channel transfer function and the input signal. We show that these conditions are generically satisfiable.

Simulations of the presented algorithms are provided to demonstrate the potential ability of these algorithms in applications involving fast channel variations. The structure of the paper is as follows. In Section II, the problems of BSI and BCE are formulated, and some preliminary results are given. Section III deals with allpole systems. Identification algorithms are presented along with their convergence analysis. Section IV discusses identification algorithms for arbitrary IIR systems. The identifiability problem is formulated and studied in Section V. Finally, some remarks are given in Section VI.

Manuscript received May 12, 1997; revised January 26, 1999. This work was supported in part by Grants from the National Science Foundation and the Australian Research Council. The associate editor coordinating the review of this paper and approving it for publication was Prof. José M. F. Moura.

E. Bai is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA.

M. Fu is with the Department of Electrical and Computer Engineering, The University of Newcastle, Callaghan, Australia.

Publisher Item Identifier S 1053-587X(99)04715-7.

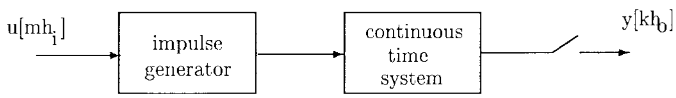

Fig. 1. Sampled data system.

II. Problem Statement and Preliminary

In this paper, the sampled data system consists of an ideal impulse generator, as shown in Fig. 1. The input to the continuous time system is a sequence of δ-functions whose magnitudes vary according to the input sequence $u[mh_i]$ with the sampling interval h_i. This model is very common for digital signal processing and communication systems. Although we focus on the sampled data systems represented in Fig. 1, it should be pointed out that with minor modifications, all the results derived in this paper apply to the sampled data systems with a zero-order hold, which is very common in the control systems literature.

The output $y[kh_o]$ is also a discrete sequence with the sampling interval h_o. The sampling intervals h_i and h_o are usually different. It is assumed in this paper that $h_o = h_i/p$ for some positive integer $p \geq 1$, which is referred to as the oversampling ratio. The continuous time system is assumed to be represented by an unknown linear time-invariant nth-order state space equation

$$\begin{aligned}\dot{x}(t) &= Ax(t) + bu(t-\tau), \quad u, y \in R, A \in R^{n\times n}\\ y &= cx(t)\end{aligned} \tag{2.1}$$

where τ is some possible unknown delay.

In general, if the output sampling frequency is different from the input sampling frequency, the overall system is time varying. However, because the output sampling frequency $f_o = 1/h_o$ is an integer multiple of the input sampling frequency $f_o = 1/h_o = pf_i$, the resultant discrete time system is still linear and time invariant in terms of the output sampling interval h_o.

Remark 2.1: The purpose of the paper is to present a novel approach to solving the BSI and BCE problems for an IIR system. To focus on this point, we do not include any noise at the channel outputs. However, analysis of the noise effect is standard and can be found in many control and signal processing textbooks; see, e.g., [6]. In addition, noises are added in the simulations.

Remark 2.2: By the same token, we will concentrate on the zero delay ($\tau = 0$) case. Extensions to the case with unknown delay τ will be discussed, but only briefly.

Now, if $\tau = 0$ and the input is a sequence of the δ function, the system equation (2.1) can be solved as

$$\begin{aligned}x[(k+1)h_o] &= e^{Ah_o}x[kh_o] + \int_{kh_o}^{(k+1)h_o} \\ &\quad \cdot e^{A((k+1)h_o - v)}\, bu(v)\, dv\\ &= e^{Ah_o}x[kh_o] + e^{Ah_o}bu[kh_o].\end{aligned} \tag{2.2}$$

Thus, by defining

$$F = e^{Ah_o}, \quad \Gamma = e^{Ah_o}\cdot b$$

we have the sampled system

$$\begin{aligned}x[(k+1)h_o] &= Fx[kh_o] + \Gamma u[kh_o]\\ y[kh_o] &= cx[kh_o].\end{aligned} \tag{2.3}$$

Accordingly, the transfer function from the input u to the output y, which is considered at the output sampling period h_o, is given by an nth-order rational transfer function

$$\begin{aligned}G(z) &= c(zI - F)^{-1}\Gamma\\ &= \frac{b(z)}{a(z)} = \frac{b_1 z^{-1} + \cdots + b_n z^{-n}}{1 - a_1 z^{-1} - \cdots - a_n z^{-n}}\end{aligned} \tag{2.4}$$

for some constants a_i and b_i.

Remark 2.3: From the derivation of the discrete time transfer function $G(z)$, we see that the order of the sampled data system is n, which is the same as the order of the continuous time system and is independent of the oversampling ratio p, except in some pathological cases for which pole-zero cancellation happens. This is one of the advantages to model the channel as an IIR system. In contrast, if an FIR model is used to approximate an IIR channel, the number of parameters in the model needs to increase as the oversampling ratio p increases.

Our goal is to identify the parameters in (2.4) as well as the input signal $u[mh_i]$ based on the output measurements $y[kh_o]$ with as little knowledge as possible on the unknown input sequence. Let $Y(z)$ and $U(z)$ be the z-transforms of output and input sequences, respectively. Then

$$Y(z) = G(z)U(z) = (\alpha G(z))\left(\frac{1}{\alpha}\, U(z)\right)$$

for any nonzero constant α. Therefore, the best we can do is to identify $G(z)$ and $U(z)$ up to a scaling constant. The BSI and BCE problems are formally defined as follows.

The Blind System Identification (BSI) and Blind Channel Equalization (BCE) Problems: Consider the sampled data systems in Fig. 1, assuming that the input is an unknown nonzero bounded sequence. Identify $G(z)$ and $U(z)$ up to a constant for some oversampling ratio $p \geq 1$ based only on the output observation $y[kh_o]$.

Let $Y(z)$ and $U(z)$ denote the z transforms of $y[kh_o]$ and $u[kh_o]$, respectively. For $h_i/h_o = p$, note that the input to the system in Fig. 1 is nonzero only once for every p samples, i.e.,

$$Y(z) = G(z)U(z) = G(z)U_0(z^p). \tag{2.5}$$

Before closing this section, we make an assumption on the minimality (reachability and observability) of the discrete time systems:

Assumption 1: It is assumed throughout the paper that the sampled data system represented in Fig. 1 is minimal (reachable and observable) when the oversampling ratio $p = 1$, i.e. when $h_i = h_o$.

The following lemma can be easily verified.

Lemma 2.1: Consider the sampled data system in Fig. 1. Then, we have the following.

1) The discrete time system is minimal (reachable and observable) for a given $h_o = h_i/p, p \geq 1$ if and only if $\text{Im}(\lambda_i - \lambda_j) \neq (2m\pi/h_o), m = \pm 1, \pm 2, \cdots$ whenever

$\mathrm{Re}(\lambda_i - \lambda_j) = 0$, where λ_i's are the eigenvalues of the continuous time system.

2) The discrete time system is minimal for any oversampling ratio $p = h_i/h_o \geq 1$ if it is minimal for $p = 1$, i.e., $h_i = h_o$.

Proof: The proof is a straightforward extension of the theorem in [2, p. 561]. ■

III. Identification of All-Pole Systems

We first consider a simple case of IIR systems, namely, an all-pole filter, to convey the idea. Although all-pole filters rarely resembles the sampled-data system in Fig. 1 precisely, they are often used to approximate IIR models.

Let the oversampling ratio $p = 2$, i.e., $h_i = 2h_o$. Suppose the resultant sampled data system is an all-pole system in terms of the output sampling interval h_o, i.e.,

$$G(z) = \frac{b_1}{1 - a_1 z^{-1} - \cdots - a_n z^{-n}}. \tag{3.1}$$

We have dropped the delay z^{-1} in the numerator for notational simplicity. The result in this section remains valid with some minor notational changes if z^{-1} is present.

The time domain expression of (3.1) is

$$y[kh_o] = \sum_{i=1}^{n} a_i y[(k-i)h_o] + b_1 u[kh_o].$$

Observe that $p = 2$, and the input $u[kh_o]$ to the impulse generator as shown in Fig. 1 is zero for odd k. It follows that

$$y[kh_o] = \sum_{i=1}^{n} a_i y[(k-i)h_o] \tag{3.2}$$

for $k = 2l+1, l = 0, 1, 2, \cdots$. In other words, the coefficients a_i's are the solution of the linear equation

$$\Phi[a_1 \quad a_2 \quad \cdots \quad a_n]' = q \tag{3.3}$$

where

$$q = (y[(2l+1)h_o], y[(2l+3)h_o], y[(2l+5)h_o], \cdots)' \tag{3.4}$$

and Φ is shown in (3.5) at the bottom of the page.

Thus, a_i's can be uniquely solved by many standard methods if the matrix Φ has full column rank. In fact, a_i's can also be calculated on-line recursively by employing recursive least squares or other recursive algorithms. In this case, to guarantee the asymptotical convergence of the estimates $\hat{a}_i$'s to the true but unknown a_i's, some persistent excitation (PE) condition is required [6]. Define

$$\phi[l] = [y[(2l+n)h_o] \quad y[(2l+n-1)h_o] \quad \cdots \quad y[(2l+1)h_o]]'. \tag{3.6}$$

The PE condition means that the condition below holds uniformly in l_0 for some constants $\alpha, m > 0$

$$\sum_{l=l_0+1}^{l_0+m} \phi[l]\phi'[l] \geq \alpha I > 0. \tag{3.7}$$

Theorem 3.1: Consider the system in Fig. 1 with $p = 2$ and $G(z)$ given by (3.1). Then, we have the following.

1) $\phi[l]$ is PE if the spectral measure of the input $u[kh_i]$ (in h_i not in h_o, which is a weaker condition) is not concentrated on $k < n$ points.

2) The recursive least squares estimates $\hat{a}_i$'s of (3.3) converge asymptotically to the true value a_i's if the PE condition in 1) is satisfied.

Proof: See Section VII. ■

Remark 3.1: It is important to point out that the idea of fast sampling at the output is not new. It is a common practice in signal processing area to have several sampling rates. This idea has also recently applied to blind system identification for the FIR system [11]. Let $h_i = 2h_o$ and $G_1(z)$ and $G_2(z)$ be the transfer functions from the input $u[kh_i]$ to the outputs $y[2kh_o]$ and $y[(2k+1)h_o]$, respectively. Denote the Z transforms of $u[kh_i]$, $y[2kh_o]$, and $y[(2k+1)h_o]$ by $U(z)$, $Y_1(z)$, and $Y_2(z)$, respectively. Then

$$Y_1(z) = G_1(z)U(z), \quad Y_2(z) = G_2(z)U(z)$$

and this implies

$$G_2(z)Y_1(z) - G_1(z)Y_2(z) = 0.$$

If $G_1(z)$ and $G_2(z)$ are FIR systems, the coefficients can be estimated from the above equation modulo a scaling constant. If $G_1(z)$ and $G_2(z)$ are IIR systems, however, they share the same denominator, and the above equation does not provide any information about the denominator at all. Thus, the denominator cannot be identified in this way. This difficulty is inherent in the scheme and cannot be easily removed.

Remark 3.2: The PE condition imposed in Theorem 3.1 is very mild in most communications systems. For instance, $\phi[l]$ is PE if $u[kh_i]$ is white or has at least $n/2$ sinusoids.

To illustrate the result of Theorem 3.1, we give a simulation example here.

Simulation Example 1: Suppose the (unknown) sampled data system is represented by a third-order transfer function

$$G(z) = \frac{1}{(1-0.5z^{-1})(1+0.2z^{-1})(1-0.3z^{-1})} = \frac{1}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2} - \alpha_3 z^{-3}}$$

where

$$\alpha_1 = 0.6, \quad \alpha_2 = 0.01, \quad \alpha_3 = -0.03.$$

$$\Phi = \begin{pmatrix} y[(2l)h_o] & y[(2l-1)h_o] & \cdots & y[(2l+1-n)h_o] \\ y[(2l+2)h_o] & y[(2l+1)h_o] & \cdots & y[(2l+3-n)h_o] \\ y[(2l+4)h_o] & y[(2l+3)h_o] & \cdots & y[(2l+5-n)h_o] \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{3.5}$$

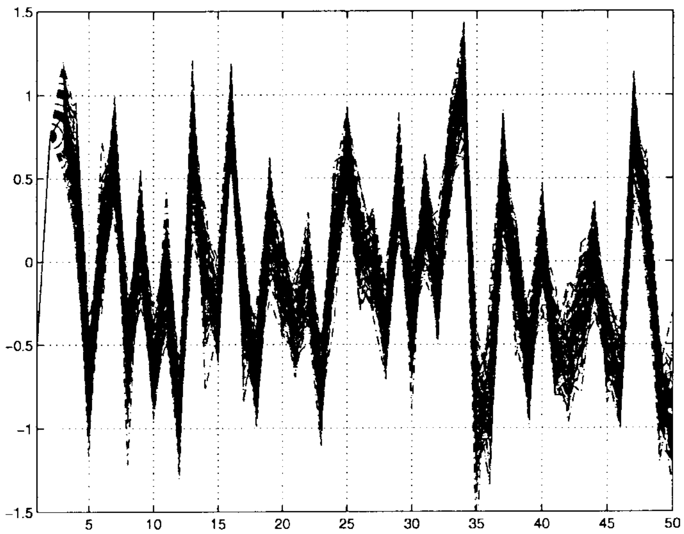

Fig. 2. One hundred estimates of the input signal (SNR $= 25$ dB).

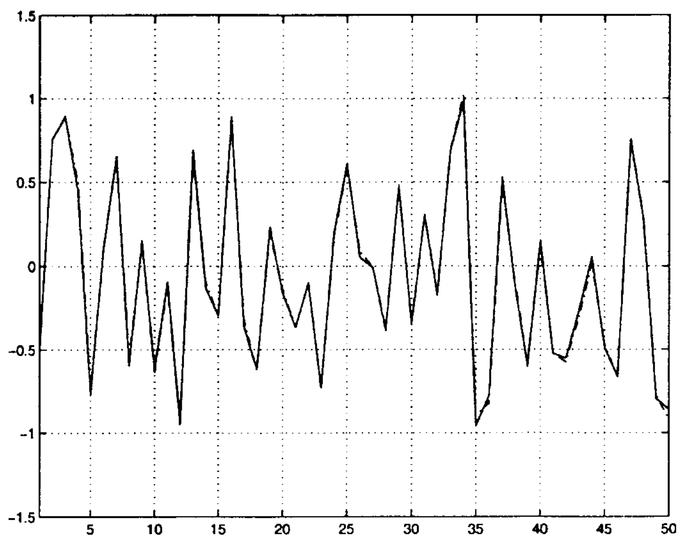

Fig. 3. Actual input (solid) and the sample mean of 100 estimates (dash-dot) (SNR $= 25$ dB).

To obtain a performance measure of the blind channel identification, the normalized root-mean-square error (NRMSE) of the estimate is defined as

$$\text{NRMSE} = \frac{1}{\|(\alpha_1, \alpha_2, \alpha_3)\|} \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left\| \begin{pmatrix} \alpha_1 - \hat{\alpha}_1(i) \\ \alpha_2 - \hat{\alpha}_2(i) \\ \alpha_3 - \hat{\alpha}_3(i) \end{pmatrix} \right\|^2}$$

where $(\hat{\alpha}_1(i), \hat{\alpha}_2(i), \hat{\alpha}_3(i))'$ is the estimate of $(\alpha_1, \alpha_2, \alpha_3)'$ from the ith run, and M is the number of Monte Carlo trials and was set to be 100 in the simulation. Fifty symbols were used in each trial to estimate $(\alpha_1, \alpha_2, \alpha_3)'$. The input signal is estimated by feeding the received output signal into the inverse of the estimated transfer function.

For the simulation, both the input and the noise signals are assumed to be independent random variables in $[-1, 1]$. Fig. 2 shows the 100 estimates of the input signal (SNR $= 25$ dB). Fig. 3 shows the actual input signal (solid line) and the sample mean of its 100 estimates (dash-dot line) with SNR $= 25$ dB. We see that the sample mean is very close to the actual input signal. Fig. 4 shows the NRMSE versus SNR (in decibels) in a series of 100 Monte Carlo runs for different SNR's.

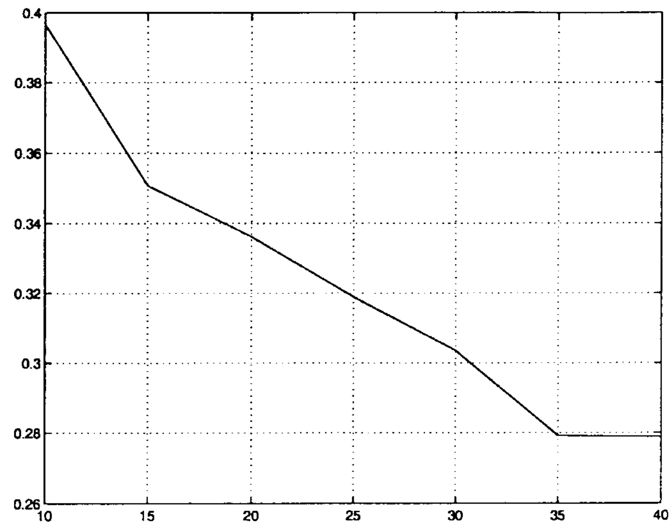

Fig. 4. NRMSE versus SNR.

Remark 3.3: In reality, the order of the channel may be unknown. In this case, the order can be estimated by using some standard methods in the system identification literature. For instance, if the order is overestimated, the matrix Φ in (3.5) is no longer full column rank. Thus, the order can be estimated by testing the rank of $\Phi'\Phi$. For details, e.g., see [6]. Alternatively, the order can be estimated using the standard subspace method.

It should be pointed out that Theorem 3.1 assumes that the input and the output are synchronized (i.e., the channel has zero delay). In reality, the input and the output are not synchronized, and some unknown delay is always there. This corresponds to the case where $\tau \neq 0$ in (2.1) or in the discrete time domain, the transfer function is given by

$$G(z) = \frac{b_1 z^{-\text{del}}}{1 - a_1 z^{-1} - \cdots - a_n z^{-n}}$$

for some unknown delay del $= 2d$ or $2d+$ 1 for some integer d. However, the oversampling ratio $p = 2$ implies that one of the following equations should hold:

$$y[2kh_o] = \sum_{i=1}^{n} a_i y[(2k - i)h_o]$$
$$y[(2k+1)h_o] = \sum_{i=1}^{n} a_i y[(2k + 1 - i)h_o].$$

The problem is that we do not know which one is the right one. To this end, let

$$(\hat{a}_{11}, \hat{a}_{12}, \cdots, \hat{a}_{1n})', \quad \text{and} \quad (\hat{a}_{21}, \hat{a}_{22}, \cdots, \hat{a}_{2n})'$$

be the estimates of a_i's by using the odd samples $(2k+1)$ and the even samples $(2k)$, respectively. Define

$$v_1[kh_o] = (1 - \hat{a}_{11} z^{-1} - \hat{a}_{12} z^{-2} - \cdots - \hat{a}_{1n} z^{-n}) y[kh_o] \tag{3.8}$$

$$v_2[kh_o] = (1 - \hat{a}_{21} z^{-1} - \hat{a}_{22} z^{-1} - \cdots - \hat{a}_{2n} z^{-n}) y[kh_o]. \tag{3.9}$$

Clearly, v_1 and v_2 are the estimates of the input u by using the estimates $(\hat{a}_{11}, \cdots, \hat{a}_{1n})'$ and $(\hat{a}_{21}, \cdots, \hat{a}_{2n})'$, respectively. In the absence of noise, one of $v_i[kh_o]$ recovers, up to a scaling constant, exactly the message signal $u[kh_o]$ modulo a delay due to unknown del in the transfer function $G(z)$. Thus, this signal will be zero for every alternative sample. The other signal, however, does not have this property. Subsequently, if we have

$$\min\left\{\sum_{k=\text{even}} v_1^2[kh_o], \sum_{k=\text{odd}} v_1^2[kh_o]\right\} < \min\left\{\sum_{k=\text{even}} v_2^2[kh_o], \sum_{k=\text{odd}} v_2^2[kh_o]\right\} \quad (3.10)$$

it implies that $v_1[kh_o]$ is $u[kh_o]$ modulo a delay d and, in turn, that $(\hat{a}_{11}, \hat{a}_{12}, \cdots, \hat{a}_{1n})'$ is the estimate of a's. On the other hand,

$$\min\left\{\sum_{k=\text{even}} v_1^2[kh_o], \sum_{k=\text{odd}} v_1^2[kh_o]\right\} < \min\left\{\sum_{k=\text{even}} v_2^2[kh_o], \sum_{k=\text{odd}} v_2^2[kh_o]\right\} \quad (3.11)$$

implies that $v_2[kh_o]$ is $u[kh_o]$ modulo a delay d and, in turn, that the estimate $(\hat{a}_{21}, \hat{a}_{22}, \cdots, \hat{a}_{2n})'$ is correct.

Simulation Example 2: To illustrate the performance of the above algorithm for an unknown delay del, we consider the system in Simulation Example 1 again with an unknown delay del

$$G(z) = \frac{z^{-\text{del}}}{1 - 0.6z^{-1} - 0.01z^{-2} + 0.03z^{-3}}.$$

All the setups are identical as in Simulation Example 1, except the delay del is unknown. We first set the unknown del $= 2$. After 100 Monte Carlo trials with SNR $= 30$ dB, the sample means of

$$\left[\sum_{k=\text{even}} v_1^2[kh_o], \sum_{k=\text{odd}} v_1^2[kh_o], \sum_{k=\text{even}} v_2^2[kh_o] \sum_{k=\text{odd}} v_2^2[kh_o]\right] = [3.8510, 0.1101, 3.7333, 1.9539].$$

Clearly, 0.1101 is the minimum. According to the above discussion, v_1 recovers the input considered at the output sampling interval h_o, and in fact, the even part of v_1 recovers the input signal $u[kh_i]$. Fig. 5 shows the actual input signal $u[kh_i]$ (solid line) and the sample mean of the even part of v_1 (dash-dot line). We see that the even part of v_1 estimates the input well except for some delay due to the unknown del. Fig. 6 shows the NRMSE versus SNR in a series of 100 Monte Carlo runs.

We now change the unknown delay del $= 3$. After 100 runs with SNR $= 30$ dB, the sample means of

$$\left[\sum_{k=\text{even}} v_1^2[kh_o], \sum_{k=\text{odd}} v_1^2[kh_o], \sum_{k=\text{even}} v_2^2[kh_o] \sum_{k=\text{odd}} v_2^2[kh_o]\right] = [2.4230, 4.0469, 0.4250, 4.2141].$$

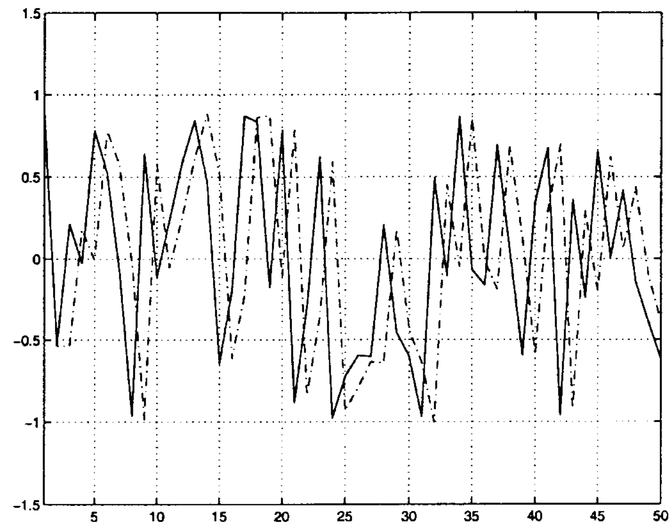

Fig. 5. Actual input (solid) and the sample mean of its estimates (dash-dot) when unknown del = 2.

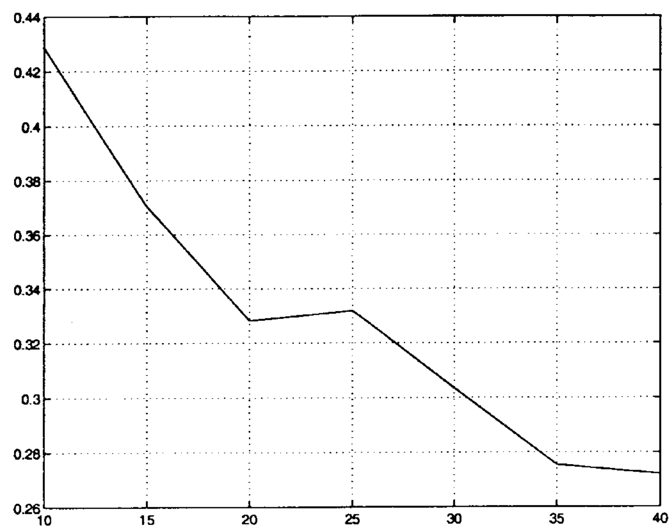

Fig. 6. NRMSE versus SNR when del is unknown.

Obviously, the odd part of v_2 recovers the input. Fig. 7 shows the actual input signal $u[kh_i]$ (solid line) and the sample mean of the odd part of v_2 (dash-dot line). We see that again the proposed scheme estimates the input well except some delay.

IV. Identification of IIR Systems

We now consider arbitrary IIR systems characterized by (2.4). Two algorithms will be provided. The first algorithm requires $p = 2(n+1)$ and identifies a set of parameters α_i and β_i, which are equivalent to a_i and b_i, respectively. The identification of α_i and β_i will be done separately with the advantage that estimation errors of α_i and β_i are independent. The second algorithm uses $p = n + 1$. This algorithm identifies a_i first and then uses the estimate of a_i to identify b_i. With the lower oversampling ratio, the tradeoff of this algorithm is that the estimation error in a_i affects the estimate for b_i.

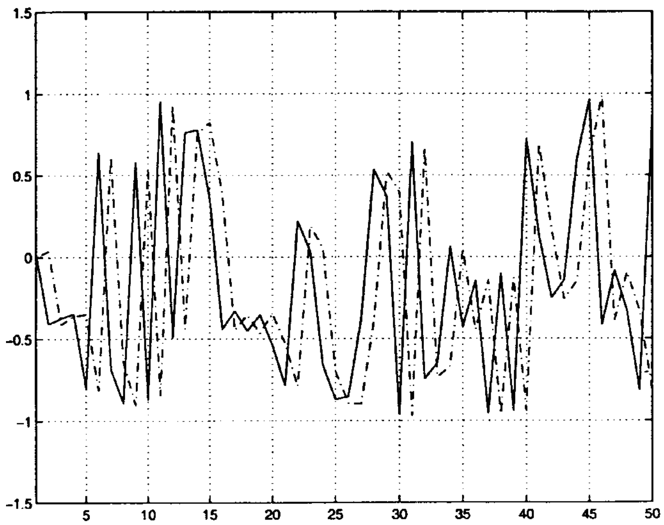

Fig. 7. Actual input (solid) and the sample mean of its estimates (dash-dot) when unknown del = 3.

Algorithm 4.1: Let $p = 2(n+1)$. The input $U(z)$ is given by

$$U(z) = U_0(z^{2(n+1)}). \tag{4.1}$$

Decompose the output

$$Y(z) = Y_0(z^2) + z^{-1}Y_1(z^2)$$

and let $G_0(z)$ and $G_1(z)$ be, respectively, the transfer functions from $U_0(z^{n+1})$ to $Y_0(z)$ and $Y_1(z)$, i.e.,

$$\begin{aligned} Y_0(z) &= G_0(z)U_0(z^{n+1}) \\ Y_1(z) &= G_1(z)U_0(z^{n+1}). \end{aligned} \tag{4.2}$$

Then, we have

$$Y(z) = G_0(z^2)U_0(z^{2(n+1)}) + z^{-1}G_1(z^2)U_0(z^{2(n+1)}).$$

On the other hand

$$Y(z) = G(z)U_0(z^{2(n+1)}) = \frac{b(z)a(-z)}{a(z)a(-z)}\, U_0(z^{2(n+1)}) \tag{4.3}$$

where $a(z)a(-z)$ is an even polynomial. Denote

$$\alpha(z^2) = a(z)a(-z) \tag{4.4}$$

where $\alpha(z) = 1 - \alpha_1 z^{-1} - \cdots - \alpha_n z^{-n}$, and decompose

$$b(z)a(-z) = \beta_0(z^2) + z^{-1}\beta_1(z^2) \tag{4.5}$$

with

$$\begin{aligned} \beta_0(z) &= \beta_{01}z^{-1} + \cdots + \beta_{0n}z^{-n} \\ \beta_1(z) &= \beta_{11} + \cdots + \beta_{1n}z^{-n-1}. \end{aligned}$$

We obtain

$$Y(z) = \frac{\beta_0(z^2)}{\alpha(z^2)}\, U_0(z^{2(n+1)}) + z^{-1}\frac{\beta_1(z^2)}{\alpha(z^2)}\, U_0(z^{2(n+1)}).$$

Hence

$$G_0(z) = \frac{\beta_0(z)}{\alpha(z)}; \quad G_1(z) = \frac{\beta_1(z)}{\alpha(z)}. \tag{4.6}$$

Now, because $G_0(z)$ and $G_1(z)$ are the transfer functions from $U_0(z^{(n+1)})$ to the outputs $Y_0(z)$ and $Y_1(z)$, respectively, in terms of the sampling period $h = 2h_o$, $G_0(z)$ and $G_1(z)$ share the common nth-order denominator. In addition, as far as equalization is concerned, either $G_0(z)$ or $G_1(z)$ is sufficient to recover the input because

$$U(z^{(n+1)}) = G_0^{-1}(z)Y_0(z) = G_1^{-1}(z)Y_1(z).$$

Note that the equations in (4.2) are expressed in terms of the sampling interval $h = 2h_o$. Its time domain representation is given by

$$y_j[kh] = \sum_{i=1}^{n} \alpha_i y[(k-i)h] + \sum_{i=1}^{n} \beta_{ji}u[(k-i)h]. \tag{4.7}$$

The input sequence $u[kh]$ is nonzero only if $k = l(n+1)$, i.e.,

$$\begin{aligned} u[(l(n+1)-1)h] &= u[(l(n+1)-2)h] \\ &= \cdots = u[(l(n+1)-n)h] = 0 \end{aligned}$$

for all l. Now, consider (4.7) at the sampling interval h for $k = l(n+1)$

$$y_j[l(n+1)h] = \sum_{i=1}^{n} \alpha_i y_j[(l(n+1)-i)h].$$

As this holds for $j = 0, 1$, we can write more generally that

$$y[l(n+1)h] = \sum_{i=1}^{n} \alpha_i y[(l(n+1)-i)h]. \tag{4.8}$$

This equation is linear in α_i and can be written as [similar to (3.3)]

$$\Phi[\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_n]' = q \tag{4.9}$$

where we have (4.10) and (4.11), shown at the bottom of the next page. Again, any standard identification method (e.g., least squares) can be used to estimate α_i's.

To identify $\beta_0(z)$ and $\beta_1(z)$, we have by substituting (4.6) into (4.2), that

$$\beta_1(z)Y_0(z) - \beta_0(z)Y_1(z) = 0. \tag{4.12}$$

From the equations $\beta_0(z) = \beta_{01}z^{-1} + \cdots + \beta_{0n}z^{-n}$ and $\beta_1(z) = \beta_{11} + \cdots + \beta_{1n}z^{-n-1}$, we have in the time domain

$$\bar{\Phi}(\beta_{11} \quad \cdots \quad \beta_{1n} \quad \beta_{01} \quad \cdots \quad \beta_{0n})' = 0 \tag{4.13}$$

where we have (4.14), shown at the bottom of the next page. Again, β_{0i} and β_{1i} appear linearly in (4.13). The difference between this equation and (4.9) is that the solution to this equation is unique only up to a scaling constant, provided that $\bar{\Phi}$ has full column rank. This is consistent with our goal of blind system identification, i.e., to find $G(z)$ up to a scaling constant. In fact, any vector in the null space of $\bar{\Phi}$ is a solution. There are many ways to solve this equation. The simpliest one is probably to normalize one of the components of $\beta_0(z)$ or $\beta_1(z)$. For instance, let $\beta_{01} = 1$. Then, we have

$$\begin{aligned} &\phi'[kh](\beta_{11}\cdots\beta_{1n} \quad \beta_{02}\cdots\beta_{0n})' \\ &\quad = y[(k-1)h + h_o] \end{aligned} \tag{4.15}$$

with

$$\begin{aligned}\phi'[kh] = (&y[kh], \cdots, y[(k-n+1)h]\\ &- y[(k-2)h+h_o], \cdots\\ &- y[(k-n)h+h_o])\end{aligned}$$

or, in a compact form

$$\tilde{\Phi}(\beta_{11} \quad \cdots \quad \beta_{1n} \quad \beta_{02} \quad \cdots \quad \beta_{0n})' = q \tag{4.16}$$

where we have

$$q = (y[(k-1)h+h_o], y[kh+h_o], y[(k+1)h+h_o], \cdots)' \tag{4.17}$$

and (4.18), shown at the bottom of the page. Thus, β_{0i} and β_{1i} can be solved in a similar way as for (4.9). Moreover, the full rankness of $\tilde{\Phi}$ can be established in a similar way as for allpole systems. We summarize the results for blind system identification of the IIR system in the following theorem (see Section VII for proof).

Theorem 4.1: Consider the sampled data system in Fig. 1 with $p = 2(n+1)$. Let $G_0(z)$ and $G_1(z)$ be the transfer functions described above. Then, we have the following.

1) The recursive least squares estimates $\hat{\alpha}_i$'s from (4.9) converge asymptotically to α_i's if the spectral measure of the input $u[kh_i]$ is not concentrated on $k < n$ points.
2) The recursive least squares estimate $(\hat{\beta}_{01}, \cdots, \hat{\beta}_{0n})$ and $(\hat{\beta}_{11}, \cdots, \hat{\beta}_{1n})$ from (4.16) converges asymptotically to $(\beta_{01}, \cdots, \beta_{0n})$ and $(\beta_{11}, \cdots, \beta_{1n})$ up to a scaling constant, provided that the spectral measure of the input $u[kh_i]$ is not concentrated on $k < 2n-1$ points and that the numerator of the transfer functions $G(z)$ has coprime even and odd components.

Algorithm 4.2

Let $p = n+1$. Our first step is to identify $a(z)$. To this end, we note that

$$a(z)Y(z) = b(z)U_0(z^{n+1}). \tag{4.19}$$

The time domain expression for the above is

$$y[kh_o] = \sum_{i=1}^{n} a_i y[(k-i)h_o] + \sum_{i=0}^{n-1} b_i u[(k-i)h_o]. \tag{4.20}$$

Taking $k = lp+n$, we have $u[(k-i)h_o] = 0, \forall i = 0, \cdots n-1$. Hence,

$$y[(lp+n)h_o] = \sum_{i=1}^{n} a_i y[(lp+n-i)h_o]. \tag{4.21}$$

Once again, we have

$$\Phi[a_1 \quad a_2 \quad \cdots \quad a_n]' = q \tag{4.22}$$

where

$$q = (y[(lp+n)h_o], y[((l-1)p+n)h_o] y[((l-2)p+n)h_o], \cdots)' \tag{4.23}$$

and Φ is in (4.24), shown at the bottom of the next page. Therefore, recursive least-square algorithms can be used to estimate a_i with asymptotic convergence, provided Φ is PE.

Once the estimate $\hat{a}(z)$ of $a(z)$ are obtained, our next step is to use them to further estimate b_i. For this purpose, we compute

$$V(z) = \hat{a}(z)Y(z) \tag{4.25}$$

and decompose it into

$$V(z) = V_0(z^p) + z^{-1}V_1(z^p) + \cdots z^{-n}V_n(z^p). \tag{4.26}$$

Assume now that $\hat{a}_i = a_i$ for all i. Then, $V(z) = b(z)U_0(z^p)$, meaning that

$$V_0(z^p) = 0; \quad V_i(z) = b_i U_0(z), \quad i = 1, \cdots, n. \tag{4.27}$$

Normalizing $b_1 = 1$, the above becomes $V_1(z) = U_0(z)$, and

$$V_i(z) = b_i V_1(z), \quad i = 2, \cdots, n. \tag{4.28}$$

$$q = (y[l(n+1)h], y[(l+1)(n+1)h], y[(l+2)(n+1)h], \cdots)' \tag{4.10}$$

$$\Phi = \begin{pmatrix} y[(l(n+1)-1)h] & y[(l(n+1)-2)h] & \cdots & y[(l(n+1)-n)h] \\ y[((l+1)(n+1)-1)h] & y[((l+1)(n+1)-2)h] & \cdots & y[((l+1)(n+1)-n)h] \\ y[((l+2)(n+1)-1)h] & y[((l+2)(n+1)-2)h] & \cdots & y[((l+2)(n+1)-n)h] \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{4.11}$$

$$\overline{\Phi} = \begin{pmatrix} y[kh] & \cdots & y[(k-n+1)h] & -y[(k-1)h+h_o] & \cdots & -y[(k-n)h+h_o] \\ y[(k+1)h] & \cdots & y[(k-n+2)h] & -y[kh+h_o] & \cdots & -y[(k-n+1)h+h_o] \\ y[(k+2)h] & \cdots & y[(k-n+3)h] & -y[(k+1)h+h_o] & \cdots & -y[(k-n+2)h+h_o] \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{4.14}$$

$$\tilde{\Phi} = \begin{pmatrix} y[kh] & \cdots & y[(k-n+1)h] & -y[(k-2)h+h_o] & \cdots & -y[(k-n)h+h_o] \\ y[(k+1)h] & \cdots & y[(k-n+2)h] & -y[(k-1)h+h_o] & \cdots & -y[(k-n+1)h+h_o] \\ y[(k+2)h] & \cdots & y[(k-n+3)h] & -y[kh+h_o] & \cdots & -y[(k-n+2)h+h_o] \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{4.18}$$

The time domain expression of the above gives a set of linear equations for b_i

$$q_1 b_i = q_i, \quad i = 2, \cdots, n \tag{4.29}$$

where

$$q_i = (v(lp-i), v((l-1)p-i), \cdots)' \quad i = 1, \cdots, n. \tag{4.30}$$

Subsequently

$$b_i = (q_1' q_1)^{-1} q_1' q_i. \tag{4.31}$$

Note that b_i contains a scalar factor as a result of normalizing $b_1 = 1$. The inverse above exists because $U_0(z^p) \neq 0$ (following from the assumption $Y(z) \neq 0$).

Theorem 4.2: Consider the sampled data system in Fig. 1 with $p = n+1$. Then, we have the following.

1) The recursive least squares estimates $\hat{a}_i$'s from (4.22) converge asymptotically to a_i's if the spectral measure of the input $u[kh_i]$ is not concentrated on $k < 2n-1$ points.
2) Suppose the PE condition above holds. Then, the recursive least squares estimate $\hat{b}_i$ from (4.29) converges asymptotically to b_i up to a scaling constant.

Proof: See Section VII. ■

Algorithm 4.1 needs a faster sampling rate than Algorithm 4.2 does. However, in Algorithm 4.2, the estimates of b_i depend on the estimates of a_i. Therefore, the estimation errors of a_i can propagate to that of b_i. This implies that the overall estimation accuracy of Algorithm 4.1 may be, in general, better than Algorithm 4.1. The following numerical simulations show this fact.

Simulation Example 3: Let the unknown continuous time channel be represented by a second-order transfer function

$$G(s) = \frac{s+10}{s^2+3s+3}.$$

In the simulation, the discretized channel model is taken to be a second-order IIR system. Let the input sampling interval $h_i = 0.3$ be fixed for simulations. The input signal and noise are taken to be independent random variables in $[-1, 1]$. In applying Algorithm 4.1, $h_o = h_i/(2(2+1)) = 0.05$, and in applying Algorithm 4.2, $h_o = h_i/(2+1) = 0.1$. Fig. 7 shows the actual input signal (solid line) and the sample mean of its estimates by 100 Monte Carlo runs (dash-dot line) that were obtained by applying Algorithm 4.1. Similarly, Fig. 8 shows the actual input signal (solid line) and the sample mean of its estimates by 100 Monte Carlo runs (dash-dot line) that were obtained by applying Algorithm 4.2 (see Fig. 9 as well). In both figures, SNR = 30 dB and 50 transmitted output symbols are used to identify the channel and then to equalize the input.

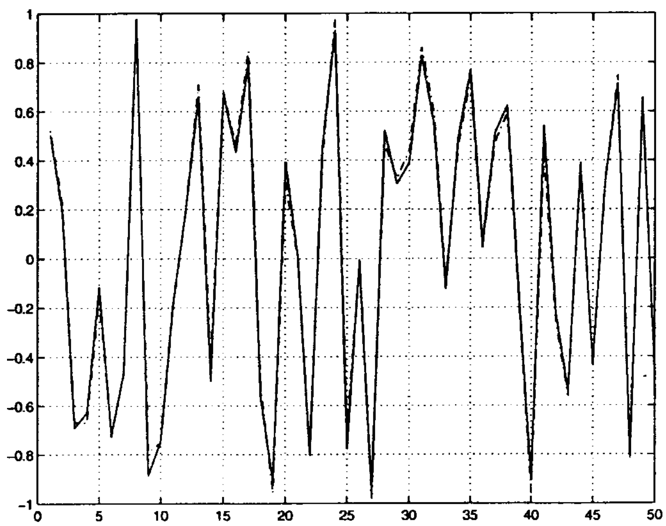

Fig. 8. Actual input (solid) and the sample mean (dash-dot) of 100 estimates obtained by applying Algorithm 4.1.

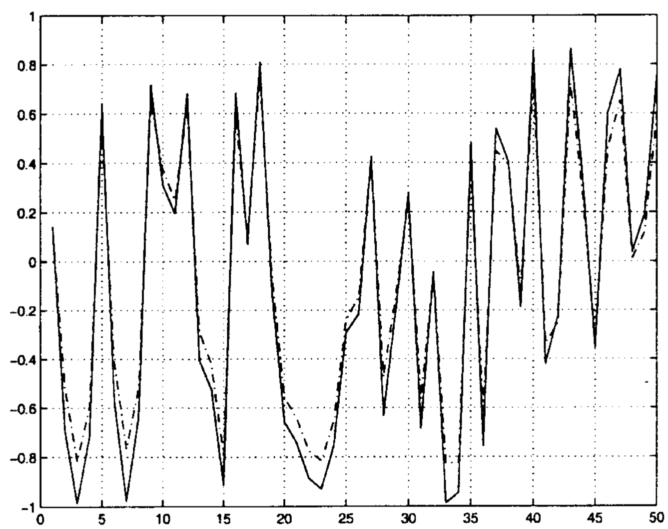

Fig. 9. Actual input (solid) and the sample mean (dash-dot) of 100 estimates obtained by applying Algorithm 4.2.

Remark 4.1: As in the case of all-pole systems, Theorems 4.1 and 4.2 assume $\tau = 0$ in (2.1). If $\tau \neq 0$ and is unknown, a modification similar to the all-pole case is needed. Namely, p estimators should be used simultaneously for all possible delays, and a decision maker is added to select the best estimator.

V. Identifiability

We have seen in the previous sections that $p = 2$ is sufficient for all-pole or FIR systems and that $p = n+1$ for IIR systems with order equal to n, provided that some very

$$\Phi = \begin{pmatrix} y[(lp+n-1)h_o] & y[(lp+n-2)h_o] & \cdots & -y[lph_o] \\ y[((l-1)p+n-1)h_o] & y[((l-1)p+n-2)h_o] & \cdots & -y[(l-1)ph_o] \\ \vdots & \vdots & \cdots & \vdots \end{pmatrix}. \tag{4.24}$$

mild PE conditions are satisfied. In this section, we study a more general identifiability problem: *Under what conditions can $G(z)$ be uniquely identified based on the observation of $\{y(kh_o)\}_0^\infty$ or, equivalently, its z transform $Y(z)$, regardless of the input sequence $\{u(kh_i)\}_0^\infty$?* Here, unique identifiability, or simply identifiability, means the following. Suppose there exists $G(z)$, $\overline{G}(z)$, $U_0(z^p)$, and $\overline{U}_0(z^p)$ such that the degree and relative degree of $\overline{G}(z)$ are the same as those of $G(z)$ and that

$$Y(z) = G(z)U_0(z^p) = \overline{G}(z)\overline{U}_0(z^p).$$

Then

$$G(z) = \alpha\overline{G}(z), \quad U_0(z^p) = \overline{U}_0(z^p)/\alpha$$

for some $\alpha \neq 0$. Implicitly assumed above is that $Y(z) \neq 0$ because otherwise, the problem of identifiability cannot be defined.

The main results of this section provide necessary and sufficient conditions for the identifiability problem. Note that the setting of the identifiability problem requires the availability of the whole sequence of $\{y[kh_o]\}_0^\infty$. Therefore, the necessary and sufficient condition for the identifiability problem will become necessary only for the identifiability of $G(z)$ when a finite sequence of $y[kh_o]$ is available. Nevertheless, the identifiability conditions will help us understand and possibly generalize the results in previous sections.

Since $Y(z) = G(z)U_0(z^p)$, roughly speaking, $G(z)$ can be identified if and only if $G(z)$ is not confused with the input $U_0(z^p)$ and that the numerator $b(z)$ and the denominator $a(z)$ are coprime. In the sequel, we will address these two issues separately.

Definition 5.1—p Factor and p Cofactor): A polynomial (respectively, rational function) $f(z)$ is called a p factor if it can be written as $g(z^p)$ for some nontrivial polynomial (respectively, rational function) $g(\cdot)$. Given a polynomial $f(z)$ that does not contain any p factor, its p cofactor, which is denoted by $f^c(z)$, is a polynomial such that $f^c(z)$ contains no p factor, $f(z)$ and $f^c(z)$ are coprime, and $f(z)f^c(z)$ is a p factor.

Note that the p cofactor is unique up to a constant. In addition, $(f^c(z))^c = f(z)$.

Theorem 5.1: Consider the system in Fig. 1 with $p = 2$.

1) Suppose $G(z)$ is an all-pole system, i.e., $G(z) = b_1/a(z)$. Then, $G(z)$ is identifiable if and only if the odd component of $a(z)$ is nonzero and shares no common factors with the even component of $a(z)$.
2) Similarly, suppose $G(z)$ is an FIR system, i.e., $G(z) = b(z)$. Then, $G(z)$ is identifiable if and only if the odd component of $b(z)$ is nonzero and shares no common factors with the even component of $b(z)$.

Proof: See Section VII. ■

Unfortunately, it is not possible to identify a general nth-order IIR system with the oversampling ratio equal to 2. To illustrate this point, we consider the sampled data system in Fig. 1 with $p = 2$ and $G(z) = (1 - \alpha z^{-1})/(1 - \beta z^{-1})(1 + \gamma z^{-1})$. Then

$$\begin{aligned} Y(z) &= G(z)U_0(z^2) = \frac{(1 - \alpha z^{-1})}{(1 - \beta z^{-1})(1 + \gamma z^{-1})}\, U_0(z^2) \\ &= \frac{(1 + \beta z^{-1})}{(1 + \alpha z^{-1})(1 + \gamma z^{-1})} \left[\frac{(1 - \alpha^2 z^{-2})}{(1 - \beta^2 z^{-2})}\, U_0(z^2)\right] \\ &= \overline{G}(z)\overline{U}_0(z^2). \end{aligned}$$

Since the input is unknown, the transfer function $G(z)$ is clearly not identifiable.

To find out the necessary and sufficient conditions for the identifiability of a general IIR system, we define a notion of Q_p set.

Definition 5.2—Q_p Set: For any oversampling ratio $p \geq 2$, let the set Q_p be all nth-order strictly proper and stable transfer functions $G(z) = b(z)/a(z)$ in terms of the output sampling interval h_o satisfying the following.

1) $a(z)$ and $b(z)$ are coprime.
2) $a(z)$ and $b(z)$ do not contain any p factor.
3) Given any $G(z) = b(z)/a(z)$ and $\overline{G}(z) = \overline{b}(z)/\overline{a}(z)$ in Q_p, express them as

$$b(z) = b_1(z)b_2(z); \quad \overline{b}(z) = b_1(z)\overline{b}_2(z) \tag{5.1}$$

$$a(z) = a_1(z)a_2(z); \quad \overline{a}(z) = a_1(z)\overline{a}_2(z) \tag{5.2}$$

where $b_2(z)$ and $\overline{b}_2(z)$ are coprime, and so are $a_2(z)$ and $\overline{a}_2(z)$. Then $deg\ a_2(z) = deg\ \overline{a}_2(z)$is not satisfied, and

$$\overline{a}_2(z) = b_2^c(z); \quad \overline{b}_2(z) = a_2^c(z) \quad \text{modulo a constant.} \tag{5.3}$$

With the above definition, we have the following result (see Section VII for proof).

Theorem 5.2: Consider the systems in Fig. 1. We have the following:

1) For any oversampling ratio $p \geq 2$, $G(z)$ is identifiable if and only if $G(z) \in Q_p$.
2) An nth-order IIR $G(z)$ is always identifiable for $p \geq 2n$, provided that $G(z)$ is coprime.
3) An nth-order IIR $G(z)$ is always identifiable for $p \geq n+1$ under Assumption 1.

Remark 5.1: Note that the identifiability conditions above impose no assumptions on the input signal. If some *a priori* information on the input signal is available, i.e, the input signal is restricted to be in a certain class, then the result above can be made stronger. For example, $p = 2$ is sufficient, provided that the input signal is such that it does not cancel the zeros and poles of the the channel transfer function.

Remark 5.2: For $p \geq 3$, the possibility that $G(z) \notin Q_p$ (therefore, the system being identifiable) is pathological, i.e., basically $p = 3$ is sufficient for the identifiability condition, although we do not have a practical algorithm for it. To argue that $G(z) \notin Q_p$ is pathological for $p = 3$, we note first that the possibility that $G(z)$ is coprime is generic. Similarly, the possibility that $a(z)$ and $b(z)$ contain a p factor is also generic. Further, if condition (3) in the definition of Q_p is violated, then $b_2^c(z) = \overline{a}_2(z)$ in particular. Then, either $b_2(z)$ or

$\overline{a}_2(z)$ will have two roots with the same magnitude but phase separated by $2\pi k/p$ for some integer $k > 0$. Such a chance is pathological. This means that p must be pathological if both $G(z)$ and $\overline{G}(z)$ can be admitted as possible transfer functions for the channel. Hence, the satisfaction of condition (3) is also nonpathological for $p \geq 3$.

VI. Concluding Remarks

In this paper, a novel approach has been proposed to blindly identify an IIR system. The key is to sample the output faster. We feel that the work reported in this paper is just a preliminary result on the problem of blind identification of an arbitrary IIR system, which we believe is largely an untouched field and deserves more work. For instance, an integer oversampling ratio p is assumed in this paper. It is interesting to see how this can be relaxed to allow any p and what the corresponding identification algorithms are. One of the key motivations for this paper is to find a minimal set of assumptions on the input signal for BSI and BCE. In reality, the input signal is much "richer" than simply being PE. Further, the input signal is restricted to be in a special class, depending on the coding technique. Naturally, our next task is to see how BSI and BCE can be done more effectively in these circumstances.

Proofs

Proof of Theorem 3.1: The key is to establish that $\phi[l]$ is PE when $u[kh_i]$ has at least n spectral lines. Then, the rest of the theorem follows from standard results on Least squares; see [1] and [6]. To this end, note that $u[(2l-1)h_o] = 0$, and

$$\begin{pmatrix} y[(2l+n)h_o] \\ y[(2l+n-1)h_o] \\ \vdots \\ y[(2l+1)h_o] \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} y[(2l+n-1)h_o] \\ y[(2l+n-2)h_o] \\ \vdots \\ y[(2l)h_o] \end{pmatrix} + \begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} u[2lh_o]$$

$$= \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} y[(2l+n-2)h_o] \\ y[(2l+n-3)h_o] \\ \vdots \\ y[(2l-1)h_o] \end{pmatrix} + \begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} u[2lh_o].$$

Let

$$\phi[l] = \begin{pmatrix} y[(2l+n)h_o] \\ y[(2l+n-1)h_o] \\ \vdots \\ y[(2l+1)h_o] \end{pmatrix}, \quad v[l] = u[2lh_o]$$

$$\overline{A} = \begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \overline{b} = \begin{pmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

It follows that

$$\phi[l] = \overline{A}^2 \phi[l-1] + \overline{b} v[l].$$

We now show that the system $\phi[l] = \overline{A}^p \phi[l-1] + \overline{b} v[l]$ is reachable for any $p \geq 1$. Of course, this implies that when $p = 2$, the system is reachable. Let $\overline{\Lambda} = T\overline{A}T^{-1}$ be the Jordan form of $\overline{A}$ and $b = T\overline{b}$. Since $(\overline{A}, \overline{b})$ is reachable, each eigenvalue, including those repeated, can only have one Jordan block. Now, $(\overline{A}^p, \overline{b})$ is reachable $\Leftrightarrow$ (Λ^p, b) is reachable $\Leftrightarrow$ rank$(\lambda I - \Lambda^p, b) = n$, for all $\lambda = e^{ph_o\lambda_1}, \cdots e^{ph_o\lambda_n}$, where λ_i's are the eigenvalues of the continuous time system $\Leftrightarrow$ $e^{ph_o\lambda_i} \neq e^{ph_o\lambda_k}$ $i \neq k$ $\Leftrightarrow$ Im$(\lambda_i - \lambda_k) \neq (2\alpha\pi/ph_o) = (2\alpha\pi/h_i)$, $\alpha = \pm 1, \pm 2, \cdots$ when Re$(\lambda_i - \lambda_k) = 0$ $\Leftrightarrow$ the discrete time system is minimal when $p = 1$. Therefore, the system is reachable under Assumption 1. Now, the PE condition follows from the sufficient richness of $u[2lh_o]$ [1]. This completes the proof.

Proof of Theorem 4.1: The proof of the first part is reminiscent to the proof of Theorem 3.1. Let $v[l] = u[l(n+1)h]$ and

$$\phi[l] = (y[(l(n+1)-1)h], y[(l(n+1)-2)h] y[(l(n+1)-n)h])'.$$

Then, by simple calculation, we have

$$\phi[l] = \overline{A}^p \phi[l-1] + \overline{b} v[l-1]$$

where

$$\overline{A} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} & \alpha_n \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

and

$$\overline{b} = \overline{A}^{n-1} \begin{pmatrix} \beta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \overline{A}^{n-2} \begin{pmatrix} \beta_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \overline{A} \begin{pmatrix} \beta_{n-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \beta_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

From Assumption 1 and Lemma 1, the discrete time system is minimal and therefore, each eigenvalue of the matrix $\overline{A}$, including those of repeated, can only have one Jordan block.

Let

$$T\overline{A}T^{-1} = \Lambda = \begin{pmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_2 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & \Lambda_q \end{pmatrix}$$

where each Λ_i is a Jordan block with the dimension l_i. What we have to show is that the system is reachable; therefore, sufficient richness of the input implies PE of $\phi[l]$ and, consequently, the convergence. From the proof of the previous theorem and Assumption 1, we know that $(\overline{A}^p, \bar{b})$ is reachable if $(\overline{A}, \bar{b})$ is reachable. Note that $(\overline{A}, \bar{b})$ is reachable $\Leftrightarrow$ $(\Lambda, T\bar{b})$ is reachable $\Leftrightarrow$ $\text{rank}(\lambda I - \Lambda, T\bar{b}) = n$ for all $\lambda = e^{h\lambda_1}, \cdots e^{h\lambda_n}$, where λ_i's are the eigenvalues of the continuous time system. Now, let the first column of the matrix T be

$$(t_1, \cdots, t_{l_1}, t_2, \cdots, t_{l_2}, \cdots, t_q, \cdots, t_{l_q})'.$$

It follows that we have $T\bar{b}$, shown at the bottom of the page. Since $e^{\lambda_i h} \neq e^{\lambda_m h}$ if $i \neq m$ by Assumption 1, $(\Lambda, T\bar{b})$ is reachable $\Leftrightarrow$ the last row of each

$$\begin{aligned} &\Lambda_i^{n-1}\begin{pmatrix} t_i \\ \vdots \\ t_{l_i} \end{pmatrix}\beta_1 + \Lambda_i^{n-2}\begin{pmatrix} t_i \\ \vdots \\ t_{l_i} \end{pmatrix}\beta_2 \\ &\qquad + \cdots \Lambda_i \begin{pmatrix} t_i \\ \vdots \\ t_{l_i} \end{pmatrix}\beta_{n-1} + \begin{pmatrix} t_i \\ \vdots \\ t_{l_i} \end{pmatrix}\beta_n \\ &\qquad = t_{l_i}(\beta_1 (e^{\lambda_i h})^{n-1} + \beta_2 (e^{\lambda_i h})^{n-2} \\ &\qquad\quad + \cdots + \beta_{n-1} e^{\lambda_i h} + \beta_n) \end{aligned}$$

is not zero. The term inside the bracket is not zero; otherwise, $e^{\lambda_i h}$ would be a zero of the discrete time system as well as a pole, and this contradicts Assumption 1 and Lemma 1. Thus, $(\overline{A}, \bar{b})$ is reachable if t_{l_i} is not zero, and $i = 1, 2, \cdots, q$. To show that those t_{l_i}'s are not zero, observe that any eigenvalue $e^{\lambda_m h}$ of $\overline{A}$ is not zero, and moreover, if $e^{\lambda_m h}$ is an eigenvalue associated with the mth Jordan block Λ_m with multiplicity l_m, the corresponding eigenvector and generalized eigenvectors are the columns of the matrix T_m, shown at the bottom of the page (see, e.g., [2, p. 65]). Note that $T^{-1} = (T_1, T_2, \cdots, T_q)$ and

$$T = (T^{-1})^{-1} = \frac{\text{adj}(T^{-1})}{\det(T^{-1})} = \frac{C}{\det(T^{-1})}$$

with C^T the adjoint matrix formed by the cofactors c_{ij}. Thus, $t_{l_i} = c_{1l_i}/\det(T^{-1})$. c_{1l_i} is the determinant of some nonsingular matrix formed by deleting the first row and l_i column of T^{-1}, and this implies $t_{l_i} \neq 0$ for all $i = 1, 2, \cdots, q$. This completes the proof. For the second part, we notice that the hypothesis implies that the numerators $\beta_0(z)$ and $\beta_1(z)$ of the transfer functions $G_0(z)$ and $G_1(z)$ do not share any common zeros. Therefore, the conclusion follows from the proof of [12, Th. 2].

Proof of Theorem 4.2: The proof of the first part is identical to that of Theorem 4.1. The second part is a direct consequence of the first part.

Proof of Theorem 5.1: It is clear that at $p = 2$, $G(z)$ contains a p factor if and only if the even and odd components of $a(z)$ (for the all-pole case) or $b(z)$ (for the FIR case) are not coprime. Therefore, the coprimeness is a necessary condition for identifiability (see the remark before Theorem 5.1). In the FIR case, this condition is also sufficient because the ratio of the even and odd components of $Y(z)$ is exactly the ratio of the even and odd components of $b(z)$. Subsequently, $b(z)$ can be uniquely identified from $Y(z)$ modulo a constant. Similarly, in the all-pole case, the ratio of the even and odd components of $Y(z)$ is the ratio of the even and odd components

$$T\bar{b} = \Lambda^{n-1}T\begin{pmatrix} \beta_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \Lambda^{n-2}T\begin{pmatrix} \beta_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \Lambda T\begin{pmatrix} \beta_{n-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + T\begin{pmatrix} \beta_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} \Lambda_1^{n-1}\begin{pmatrix} t_1 \\ \vdots \\ t_{l_1} \end{pmatrix}\beta_1 + \Lambda_1^{n-2}\begin{pmatrix} t_1 \\ \vdots \\ t_{l_1} \end{pmatrix}\beta_2 + \cdots + \Lambda_1\begin{pmatrix} t_1 \\ \vdots \\ t_{l_1} \end{pmatrix}\beta_{n-1} + \begin{pmatrix} t_1 \\ \vdots \\ t_{l_1} \end{pmatrix}\beta_n \\ \Lambda_2^{n-1}\begin{pmatrix} t_2 \\ \vdots \\ t_{l_2} \end{pmatrix}\beta_1 + \Lambda_2^{n-2}\begin{pmatrix} t_2 \\ \vdots \\ t_{l_2} \end{pmatrix}\beta_2 + \cdots + \Lambda_2\begin{pmatrix} t_2 \\ \vdots \\ t_{l_2} \end{pmatrix}\beta_{n-1} + \begin{pmatrix} t_2 \\ \vdots \\ t_{l_2} \end{pmatrix}\beta_n \\ \vdots \\ \Lambda_q^{n-1}\begin{pmatrix} t_q \\ \vdots \\ t_{l_q} \end{pmatrix}\beta_1 + \Lambda_q^{n-2}\begin{pmatrix} t_q \\ \vdots \\ t_{l_q} \end{pmatrix}\beta_2 + \cdots + \Lambda_q\begin{pmatrix} t_q \\ \vdots \\ t_{l_q} \end{pmatrix}\beta_{n-1} + \begin{pmatrix} t_q \\ \vdots \\ t_{l_q} \end{pmatrix}\beta_n \end{pmatrix}.$$

$$T_m = \begin{pmatrix} -e^{(n-1)\lambda_m h} & -(n-1)e^{(n-2)\lambda_m h} & -\binom{n-1}{2}e^{(n-3)\lambda_m h} & \cdots & -\binom{n-1}{l_m-1}e^{(n-l_m)\lambda_m h} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -e^{\lambda_m h} & -1 & 0 & \cdots & 0 \\ -1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

of $a(-z)/a(z)a(-z)$. Once this function is identified from $Y(z)$, $a(z)$ can be uniquely decomposed, provided that the coprimeness condition for $a(z)$ holds.

Proof of Part 1 of Theorem 5.2: Obviously, from the example given before the definition of the Q_p set, conditions (1)–(3) in the definition of Q_p set are necessary for the unique identification of $G(z)$. Now, we show that they are sufficient. To this end, we suppose $G(z) = (b(z)/a(z)) \in Q_p$, and there exist some $\overline{G}(z) = (\overline{b}(z)/\overline{a}(z))$ with the same degree as $G(z)$ such that $Y(z) = G(z)U(z^p) = \overline{G}(z)\overline{U}(z^p)$ for some $U(z^p)$ and $\overline{U}(z^p)$. Then, let $b_1(z)$ be the greatest common divider of $b(z)$ and $\overline{b}(z)$. $a_1(z)$ and $U_1(z^p)$ are similarly defined. That is, (5.1) and (5.2) hold, and

$$U(z^p) = U_1(z^p)U_2(z^p), \quad \overline{U}(z^p) = U_1(z^p)\overline{U}_2(z^p).$$

It is implied that $G(z)U(z^p) = \overline{G}(z)\overline{U}(z^p)$, and

$$\frac{b_2(z)}{a_2(z)}\, U_2(z^p) = \frac{\overline{b}_2(z)}{\overline{a}_2(z)}\, \overline{U}_2(z^p). \tag{7.1}$$

Obviously, $\deg a_2(z) = \deg \overline{a}_2(z)$ is necessary for unique identification of $G(z)$. Now, we show that $b_2(z)$ and $a_2(z)$ have to be constants if $G(z) \in Q_p$. Suppose they are not. Due to the coprimeness of $b_2(z)$ and $\overline{b}_2(z)$, $\overline{U}_2(z^p)$ must contain $b_2(z)$ and, thus, $b_2^c(z)$. However, the term $b_2^c(z)$ in $\overline{U}_2(z^p)$ must be canceled by $\overline{a}_2(z)$ due to the coprimeness of $U_2(z)$ and $\overline{U}_2(z)$. Therefore, $\overline{a}_2(z)$ contains $b_2^c(z)$. By the same token, $U_2(z^p)$ contains $\frac{1}{\overline{a}_2(z)}$ and $1/\overline{a}_2^c(z)$, and $1/\overline{a}_2^c(z)$ has to be canceled by $b_2(z)$. Thus, $b_2(z)$ also contains $\overline{a}_2^c(z)$. Hence, $b_2^c(z) = \overline{a}_2(z)$. Similarly, we have $a_2^c(z) = \overline{b}_2(z)$. These two co-factor conditions are exactly in violation of condition (3) in the definition of Q_p. Therefore, we conclude that $b_2(z)$ and $a_2(z)$ are constant if $G(z) \in Q_p$. Suppose that now, $b_2(z)$ is a constant and must therefore be $\overline{b}_2(z)$. Subsequently, both $a_2(z)$ and $\overline{a}_2(z)$ must be constant as well due to their coprimeness and condition (2). In this case, $G(z)$ and $\overline{G}(z)$ differ by a constant. Similarly, if $a_2(z)$ is a constant, $G(z)$ and $\overline{G}(z)$ differ by a constant as well. Subsequently, $G(z)$ and $\overline{G}(z)$ differ by a constant if $G(z) \in Q_p$. The proof is thus completed.

Proof of Part 2 of Theorem 5.2: Let $p \geq 2n$. Using the result in Part 1, we argue that if $G(z)$ is coprime, $G(z)$ must be a member of Q_p. Obviously, $a(z)$ and $b(z)$ do not contain any p factor. Therefore, we need to show that condition (3) in the definition of Q_p is always satisfied as well, argued by contradiction. That is, we assume that there exists some coprime $G(z)$ that satisfies (5.3) and $\deg a_2(z) = \deg b_2^c(z)$, i.e., $G(z) \notin Q_p$. We use the notation in condition (3). Denote the degree of $b_2(z)$ by k, and then, $\deg a_2(z) \geq p - k$. On the other hand, $\deg a_2(z) \leq n$ is constrained. Therefore, we have $p - k \leq n$. However, k is restricted to be $n-$ 1. The above leads to $p - (n-1) \leq n$ or, equivalently, $p \leq 2n-$ 1, which violates the assumption that $p \geq 2n$. Hence, $G(z) \in Q_p$ must hold. Subsequently, the Part 1 of the theorem implies that $G(z)$ is identifiable.

Proof of Part 3 of Theorem 5.2: $G(z)$ is not identifiable only if (7.1) holds for some nonconstant $b_2(z)$, $a_2(z)$, $\overline{b}_2(z)$, and $\overline{a}_2(z)$. In other words, $b_2(z)\overline{a}_2(z)$ is a p factor, and so is $a_2(z)\overline{b}_2(z)$. Note that max (degree of $b_2(z)$ degree of $\overline{b}_2(z)) = n-$ 1 and $p \geq n+$ 1. Thus, $a_2(z)$ and $\overline{a}_2(z)$ must contribute to p factors. Let $e^{(\lambda_i + j\beta_i)h_o}$ be the roots of $a(z) = a_1(z)a_2(z)$, where $\lambda_i + j\beta_i$'s are the continuous time system poles. Since $a_2(z)\overline{b}_2(z)$ is a p factor, at least two roots $e^{(\lambda_i + j\beta_i)h_o}$ and $e^{(\lambda_m + j\beta_m)h_o}$ of $a_2(z)$ satisfy $\lambda_i = \lambda_m$ and $h_o(\beta_i - \beta_m) = 2\pi l/p$ for some integer $l \geq 1$. However, this implies $\beta_i - \beta_m = 2\pi l/h_i$, which is a contradiction to the minimality of the system at $p = 1$; see Assumption 1 and Lemma 1. This completes the proof.

REFERENCES

[1] E. W. Bai and S. S. Sastry, "Persistency of excitation, sufficient richness and parameter convergence in discrete time adaptive control," *Syst. Contr. Lett.*, vol. 6, pp. 153–163, 1985.

[2] C. T. Chen, *Linear System Theory and Design*, 2nd ed. New York: Holt, Rinehart, and Winston, 1984.

[3] D. J. Goodman, "Trends in cellular and cordless communications," *IEEE Commun. Mag.*, pp. 31–40, June 1991.

[4] C. R. Johnson *et al.*, "On fractionally-spaced equalizer design for digital microwave radio channels," submitted for publication.

[5] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary satistics," in *Proc. ICASSP*, Minneapolis, MN, 1993, pp. 81–84.

[6] L. Ljung, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[7] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[8] A. P. Petropulu and C. L. Nikias, "Blind deconvolution using signal reconstruction form partial higher order cepstral information," *IEEE Trans. Signal Processing*, vol. 41, pp. 2088–2095, 1993.

[9] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction," *Proc. IEEE ICASSP*, Apr. 1994.

[10] L. Tong, "Blind sequence estimation," *IEEE Trans. Commun.*, vol. 43, pp. 2986–2994, Dec. 1995.

[11] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channel," in *Proc. 25th Asilomar Conf. Signal, Syst., Comput.*, Pacific Grove, CA, Nov. 1991.

[12] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

Erwei Bai was educated at Fudan University, Shanghai Jiaotong University, and the University of California, Berkeley.

Since 1987, he has been with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, where he teaches and conducts research in the area of system identications and signal processing.

Minyue Fu (M'88) received the B.S. degree in electrical engineering from the China University of Science and Technology, Hefei, in 1982 and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin, Madison, in 1983 and 1987, respectively.

From 1983 to 1987, he held a teaching assistantship and a research assistantship at the University of Wisconsin. He worked as a Computer Engineering Consultant at Nicolet Instruments, Inc., Madison, during 1987. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI. For the summer of 1989, he was employed by the Universite Catholoque de Louvain, Louvain, Belgium, as a Maitre de Conferences Invited. He joined the Department of Electrical and Computer Engineering, the University of Newcastle, Callaghan, Australia, in 1989. Currently, he is an Associate Professor and Head of the Department. His main research interests include control systems and signal processing. He has been an Associate Editor of the IEEE TRANSACTIONS ON AAUTOMATIC CONTROL, and an Associate Editor for the Conference Editorial Board of the IEEE Control Systems Society. He is currently an Associate Editor for the journal *Optimization and Engineering*.