Asymptotically Optimal Blind Fractionally Spaced Channel Estimation and Performance Analysis

Georgios B. Giannakis, *Fellow, IEEE,* and Steven D. Halford, *Student Member, IEEE*

***Abstract*— When the received data are fractionally sampled, the magnitude and phase of most linear time-invariant FIR communications channels can be estimated from second-order output-only statistics. We present a general cyclic correlation matching algorithm for known order FIR blind channel identification that has closed-form expressions for calculating the asymptotic variance of the channel estimates. We show that for a particular choice of weights, the weighted matching estimator yields (at least for large samples) the minimum variance channel estimator among all unbiased estimators based on second-order statistics. Furthermore, the matching approach, unlike existing methods, provides a useful estimate even when the channel is not uniquely identifiable from second-order statistics.**

***Index Terms*— Cyclostationarity, fractional sampling, system identification.**

I. Introduction

In HIGH-SPEED digital communications, the channel often introduces a spreading of the transmitted symbols across time. This spreading or, as it is better known, intersymbol interference (ISI), can severely limit performance even when the overall noise level is low. At the receiver, equalizers can use (either explicitly or implicitly) knowledge of the channel impulse response to remove or reduce the ISI [18, p. 583]. Unfortunately, the channel impulse response is rarely known *a priori* by the receiver. This impulse response therefore, must be estimated before equalization can be used to remove the ISI. When "training" sequences are transmitted, the receiver can use the exact knowledge of the transmitted sequence to estimate the channel. As an alternative to using training sequences, the receiver can estimate the channel using received data only (i.e., without exact knowledge of the input). Channel estimation based only on the received data is known as *blind* channel identification and is advantageous because it does not waste the bandwidth required to transmit a training sequence.

When the receiver's matched filter output is sampled at the symbol rate, the resulting sequence is stationary and blind channel identification techniques must use higher-(than second-) order statistics (HOS) to estimate the phase of the channel [7], [17]. If the output of the matched filter is sampled at a rate greater than the symbol rate, the resulting time-series is *cyclostationary*. Unlike symbol rate outputs, Tong *et al*., [20] showed that second-order cyclostationary data provides sufficient information for blind identification of both magnitude and phase for most channels. A blind estimator based on second-order statistics has the advantage over a HOS-based estimator in that it needs fewer observations to obtain reliable (low variance) estimates.

Because oversampling implies samples are taken at some fraction of the symbol rate, these systems are generally called *fractionally spaced* or *fractionally sampled* (FS) systems, and their ability to use second-order statistics for blind identification has generated considerable research interest. Motivated by [20], many effective subspace and least-squares blind methods have been proposed for estimating the channel from the output second-order statistics [13], [21], [25], [11], [19]. These methods rely on the structure and rank properties of the correlation (or data) matrix induced by fractional sampling. Alternatively, [12], [6], and [22] present methods that are based on correlation matching. Reference [12] proposes a cross-correlation matching approach, whereas [22] and [6] present unweighted and weighted *cyclic* correlation matching approaches, respectively.

Matching approaches offer four main advantages over the subspace and least-squares methods:

i) asymptotic (or large sample) minimum variance estimation of the channel impulse response [6], [9];
ii) closed-form expressions for the asymptotic variance of the channel estimates [6], [9];[1]
iii) closed-form expressions for the asymptotic minimum variance achievable by any second-order based method [6];
iv) robustness to channels that cannot be estimated strictly from the second-order output statistics (see also [26]).

The price paid for these advantages is the additional computational complexity due to the nonlinear minimization required by the matching approaches.

In this paper, we analyze a *weighted nonlinear matching* method for FS blind channel identification. We give exact closed-form expressions for calculating the asymptotic variance with arbitrary choice of weights and provide the lower bound on the achievable variance for the class of blind identification methods that use second-order statistics. We

Manuscript received October 16, 1995, revised December 9, 1996. This work was supported by ONR Grant N00014-93-0485. Part of the work in this paper was presented at the Conference of Information Science and Systems, Princeton, NJ, March 1994 and at Milcom '94, Ft. Monmouth, NJ, October 1994. The associate editor coordinating the review of this paper and approving it for publication was Dr. Zhi Ding.

G. B. Giannakis is with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903-2442 USA.

S. D. Halford is with the Government Communications Systems Division, Harris Corporation, Melbourne, FL USA.

Publisher Item Identifier S 1053-587X(97)04959-3.

[1] While this paper was in review, preliminary performance analysis for some subspace methods was presented also in [14].

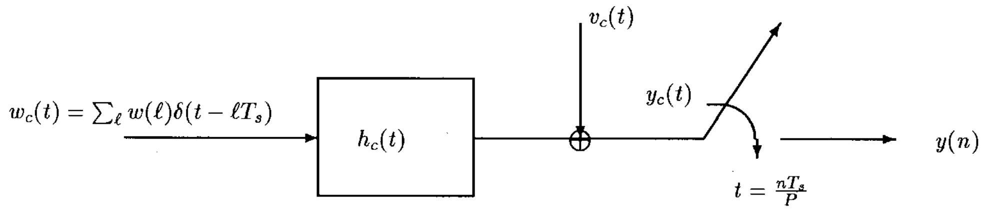


Fig. 1. Continuous-time communication channel.

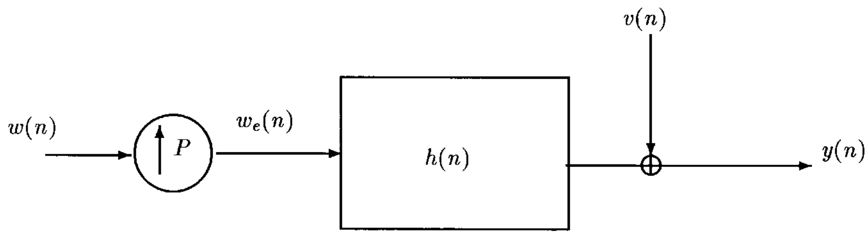


Fig. 2. Discrete time equivalent channel model.

show that this lower bound can be achieved by the weighted second-order cyclic correlation matching approach when the additive noise is Gaussian. Finally, we show that in contrast with other methods, matching methods are consistent even when the channels are not strictly identifiable from second-order cyclostationary statistics.

While there are many complementary results in this paper and the independently derived results of [27], the focus of this work is on the *method* of channel estimation and general expressions for the *variance* of the matching approaches. Zeng and Tong [27] present, for the real channel case, a lower bound on the root-mean square error. In addition, by assuming the input source to be Gaussian in distribution, they present a similar bound for the subspace and least-squares approaches. These lower bounds are then used to study the behavior of the different classes of approaches when channels are nonidentifiable or close to nonidentifiable. The superiority of the matching methods demonstrated in [27] provides additional motivation for the method presented in this paper.

II. Mathematical Background

We begin by presenting the mathematical model for FS systems. Based on this model, we show the cyclic correlations and briefly discuss identifiability of FS systems. The results presented in this section provide the framework for the channel estimation method described in Section III.

A. Mathematical Model for FS

A continuous-time FS system is shown in Fig. 1, where

- $w(\ell)$ information symbols;
- $h_c(t)$ continuous-time "composite" channel;
- $v_c(t)$ additive noise that is assumed to be uncorrelated with $w(\ell)$;
- T_s symbol duration;
- P integer that denotes the amount of oversampling.

The constituents of the composite channel $h_c(t)$ include the known transmit and receive filters as well as the unknown propagation channel. For this system, the signal at the sampler is

$$\begin{aligned} y_c(t) &= \int_{-\infty}^{\infty} w_c(\tau)h_c(t-\tau-\epsilon)\,d\tau + v_c(t) \\ &= \sum_\ell w(\ell)h_c(t-\ell T_s-\epsilon)+v_c(t) \end{aligned}$$

where we have used ϵ to denote the propagation delay, and

$$w_c(t) = \sum_\ell w(\ell)\delta(t-\ell T_s).$$

If $y_c(t)$ is sampled at $t = nT_s/P$, the received data are

$$\begin{aligned} y(n) &\triangleq y_c(nT_s/P) \\ &= \sum_\ell w(\ell)h_c(nT_s/P-\ell T_s-\epsilon)+v_c(nT_s/P). \end{aligned}$$

In many cases, the continuous-time channel $h_c(t)$ can be well-modeled as a causal FIR system [18, p. 586]. In the sequel, we will assume that $h_c(t)$ is FIR of order L_hT_s (i.e., $h_c(t) = 0\ \forall t > L_hT_s$).

If we consider the discrete time sequences $w(n)$ and $y(n)$ as the input and output, respectively, it is convenient to rewrite the input–output relationship as an equivalent discrete-time system

$$y(n) = \sum_{\ell=0}^{L} w(\ell)h(n-\ell P)+v(n) = x(n)+v(n) \qquad (1)$$

where $h(n)$ and $v(n)$ are the discrete-time equivalents of $h_c(t-\epsilon)$ and $v_c(t)$, respectively, $x(n)$ is the discrete-time equivalent of the noise-free received signal, and $L = L_hP$ is the discrete time channel order. From (1), it is easy to verify that Fig. 2 is the discrete-time equivalent of Fig. 1, where

$$w_e(n) = \sum_\ell w(\ell)\delta(n-\ell P).$$

Based on the finite set observations $\{y(n)\}_{n=0}^{T-1}$ from an oversampled (by P) system, we wish to estimate the Lth-order FIR channel $\{h(n)\}_{n=0}^{L}$ and evaluate the statistical performance of the channel estimator based on (even large sample) variance expressions. As will be seen, the main tool used in this paper for blind identification will be the second-order cyclostationary statistics of $\{y(n)\}_{n=0}^{T-1}$, which are described in the next subsection.

B. Time-Varying and Cyclic Cumulants

The $(k+\ell)$th-order cumulant of the (possibly) complex time-varying observations $y(n)$ will be denoted by (see, e.g., [1, p. 19])

$$\begin{aligned} &c_{k\ell y}(n; m_1, m_2, \cdots, m_{k-1}; r_1, r_2, \cdots, r_\ell) \\ &\quad \triangleq \text{cum}\,\{y(n), y(n+m_1), y(n+m_2), \cdots \\ &\qquad y(n+m_{k-1}), y^*(n+r_1), \cdots, y^*(n+r_\ell)\} \end{aligned} \quad (2)$$

where * indicates complex conjugate, and subscript $k\ell$ indicates that k copies are unconjugated and ℓ are conjugated. In the sequel, we will primarily be concerned with the second-order cumulants. In particular, we will need

$$\begin{aligned} c_{11y}(n; m) &= \text{cum}\,\{y(n), y^*(n+m)\} \\ &= E\{y(n)y^*(n+m)\} \end{aligned} \quad (3)$$

$$\begin{aligned} c_{20y}(n; m) &= \text{cum}\,\{y(n), y(n+m)\} \\ &= E\{y(n)y(n+m)\} \end{aligned} \quad (4)$$

$$\begin{aligned} c_{02y}(n; m) &\triangleq c_{20y^*}(n; m) \\ &= \text{cum}\,\{y^*(n), y^*(n+m)\} \\ &= E\{y^*(n)y^*(n+m)\} \end{aligned} \quad (5)$$

where we have used $E\{y(n)\} = 0$ to simplify the second-order cumulants. Note that $c_{11y}(n; m)$ is the time-varying covariance at time n and lag m. If $w(n)$ is stationary independent identically distributed (i.i.d.) and the noise $v(n)$ is stationary, (3) and (1) give

$$c_{11y}(n; m) = \sigma_w^2 \sum_\ell h(n-\ell P)h^*(n+m-\ell P) + c_{11v}(m) \quad (6)$$

where

$$\sigma_w^2 \triangleq E\{w(n)w^*(n)\} \quad \text{and}$$
$$c_{11v}(m) \triangleq E\{v(n)v^*(n+m)\}.$$

From (6), we can verify that $c_{11y}(n; m)$ is periodic in n with period P, i.e.,

$$c_{11y}(n; m) = c_{11y}(n + \ell P; m)$$

for any integer ℓ. Therefore, the data $y(n)$ is indeed cyclostationary. Because the correlation $c_{11y}(n; m)$ is periodic, it accepts a Fourier series expansion (for $k = 0, 1, \cdots, P-1$)

$$C_{11y}(k; m) \triangleq \frac{1}{P} \sum_{n=0}^{P-1} c_{11y}(n; m) e^{-j(2\pi/P)kn} \quad (7)$$

where the kth Fourier component $C_{11y}(k; m)$ is known as the cyclic correlation at cycle k and lag m. From [3] or [21], it can be shown that substituting (6) into (3) and simplifying (see also Appendix A) yields

$$\begin{aligned} C_{11y}(k; m) = &(\sigma_w^2/P) \sum_{n=0}^{L} h(n)h^*(n+m) \\ &\cdot \exp\{-j(2\pi/P)kn\} + c_{11v}(m)\delta(k). \end{aligned}$$

As in [27], we will assume in the sequel that the additive noise is white with finite variance σ_v^2 or $c_{11v}(m) = \sigma_v^2 \delta(m)$. A receive filter of bandwidth P/T_s guarantees that $v(n)$ is white. Using the whiteness of the noise, the cyclic correlation becomes

$$\begin{aligned} C_{11y}(k; m) = &\frac{\sigma_w^2}{P} \sum_{n=0}^{L} h(n)h^*(n+m)e^{-j(2\pi/P)kn} \\ &+ \sigma_v^2 \delta(m)\delta(k). \end{aligned} \quad (8)$$

From (8), it is straightforward to show the following relationships:

$$C_{11y}(k; m) = e^{j(2\pi/P)km} C_{11y}^*(P-k; -m) \quad (9)$$
$$C_{11y}(k; m) = 0 \quad \forall |m| > L. \quad (10)$$

The conjugate symmetry in the lags (9) and memory constraint (10) imply that all the nonredundant statistical information about the complex-valued channel is contained, at most, in the set

$$\mathcal{K} \triangleq \{C_{11y}(k; m); k = 0, 1, 2, \cdots, P-1, m = 0, 1, \cdots, L\}.$$

When $h(n)$ is real, there is also a conjugate symmetry in the cycles, i.e.,

$$C_{11y}(k; m) = C_{11y}^*(P-k; m).$$

For real channels, the nonredundant statistical information is contained in the set

$$\overline{\mathcal{K}} \triangleq \{C_{11y}(k; m); k = 0, 1, 2, \cdots, \lfloor P/2 \rfloor, m = 0, 1, \cdots, L\}$$

where $\lfloor \cdot \rfloor$ indicates rounding down to the nearest integer. It is important to note that for $P > 2$, $C_{11y}(k; m)$ is in general a complex quantity even when $y(n)$ is real.

C. Identifiability for FS Systems

A set of parameters is said to be identifiable from its output correlations if those correlations could have only been generated by that particular set of parameters, i.e., the correlations uniquely specify the parameters. For FS systems, the unknown parameters, that is, the true channel impulse response $\{h_0(n)\}_{n=0}^{L}$ and the variance of the additive noise σ_v^2, can be collected into a vector:

$$\boldsymbol{\theta}_0 \triangleq [h_0(0), h_0(1), \cdots, h_0(L), \sigma_v^2]'$$

where $'$ indicates transpose. From the set of cyclic correlations $\mathcal{K}$, we can collect in a vector the nonredundant cyclic correlations to form[2]

$$\begin{aligned}\tilde{\boldsymbol{c}}(\boldsymbol{\theta}_0) &\triangleq [C_{11y}(0;0), \cdots C_{11y}(0;L), C_{11y}(1;0), \cdots \\ &\qquad \cdots, C(P-1;L)]' \\ \boldsymbol{c}(\boldsymbol{\theta}_0) &\triangleq [\mathrm{Re}\,\{\tilde{\boldsymbol{c}}(\boldsymbol{\theta}_0)\}, \mathrm{Im}\,\{\tilde{\boldsymbol{c}}(\boldsymbol{\theta}_0)\}]' \end{aligned} \tag{11}$$

where we have split the generally complex cyclic correlations into the real and imaginary parts for notational convenience in the sequel. A system $\boldsymbol{\theta}_0$ is identifiable if, for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \boldsymbol{\theta} \in \mathcal{C}^L$, then

$$\boldsymbol{c}(\boldsymbol{\theta}_0) \neq \boldsymbol{c}(\boldsymbol{\theta}).$$

In other words, the cyclic correlations are unique to the parameters $\boldsymbol{\theta}_0$. To quantify which channels are identifiable, we first define the transfer function of the channel $h(n)$ as

$$H(z) \triangleq \sum_n h(n) z^{-n}.$$

Now, we state the indentifiability (ID) condition.

ID Condition [20], [3], [24]: No subset of the zeros of $H(z)$ lies equispaced on a circle with angle $2\pi/P$ separating each zero from the next.

In a blind identification problem, it is not possible to determine *a priori* whether the ID condition is satisfied. However, [21] gives a procedure for determining from the output spectra whether the ID condition is satisfied. In addition, Tugnait [24] describes a class of channels that never satisfy the ID condition. For channels that do not meet the ID condition, HOS methods must be employed (see, e.g., [8]) to uniquely identify the channel.

D. Sample Cyclic Correlations

Given the observations $\{y(n)\}_{n=0}^{T-1}$, the cyclic correlation at cycle k and lag m [see (3)] can be estimated using (see, e.g., [2]):

$$\hat{C}_{11y}^{(T)}(k;m) \triangleq \frac{1}{T} \sum_{\ell=0}^{T-1} y(\ell) y^*(\ell+m) e^{-j(2\pi/P)k\ell}. \tag{12}$$

Similar to $\boldsymbol{c}(\boldsymbol{\theta}_0)$ in (11), we can collect the nonredundant estimated cyclic correlations in a vector:

$$\begin{aligned}\hat{\tilde{\boldsymbol{c}}}_y^{(T)} &\triangleq [\hat{C}_{11y}^{(T)}(0;0)\hat{C}_{11y}^{(T)}(0;1) \cdots \hat{C}_{11y}^{(T)}(0;L) \\ &\qquad \hat{C}_{11y}^{(T)}(1;0) \cdots \hat{C}_{11y}^{(T)}(P-1;L)]' \\ \hat{\boldsymbol{c}}_y^{(T)} &\triangleq [\mathrm{Re}\,\{\hat{\tilde{\boldsymbol{c}}}_y^{(T)}\}, \mathrm{Im}\,\{\hat{\tilde{\boldsymbol{c}}}_y^{(T)}\}]'.\end{aligned}$$

In [2], the estimate $\hat{C}_{11y}^{(T)}(k;m)$ is shown to be mean-square consistent and asymptotically normal for the linear relationship of (1), provided $w(n)$ has finite moments. If the true system that generates $\{y(n)\}_{n=0}^{T-1}$ is $\boldsymbol{\theta}_0$, then the *normalized* asymptotic covariance of $\hat{\boldsymbol{c}}_y^{(T)}$ is defined as

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) \triangleq \lim_{T\to\infty} TE\{[\hat{\boldsymbol{c}}_y^{(T)} - \boldsymbol{c}(\boldsymbol{\theta}_0)][\hat{\boldsymbol{c}}_y^{(T)} - \boldsymbol{c}(\boldsymbol{\theta}_0)]'\}. \tag{13}$$

[2] We show only the complex parameter case. The same results will hold for real channels where $\boldsymbol{c}(\cdot)$ is formed by elements of $\overline{\mathcal{K}}$.

In Section IV-A, we simplify general expressions for (13) from [2] to find $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ for the FS system of (1).

III. Parameter Estimation and Performance Analysis

In this section, we first describe a weighted nonlinear (NL) matching approach for estimating the channel impulse response and noise variance $\boldsymbol{\theta}_0$ from the output data $\{y(n)\}_{n=0}^{T-1}$. Next, we analyze the asymptotic variance of the channel estimates for various choices of weights. We conclude the performance analysis with a brief discussion of the robustness of this estimator in the presence of model mismatch. Finally, we give a brief summary of the NL matching algorithm. For clarity of presentation, we will restrict our attention to channels with real impulse response coefficients, although extension to complex channels is straightforward if one adopts the real vector $[\mathrm{Re}\,\{\boldsymbol{\theta}_0\}, \mathrm{Im}\,\{\boldsymbol{\theta}_0\}]$. All blind identification techniques suffer from a possible scale and shift ambiguity. In order to remove the shift ambiguity, we have assumed that the channel is strictly causal. To remove the scale ambiguity, there are two commonly used assumptions (see e.g., [13]): i) $\Sigma_n |h(n)|^2 = 1$, or ii) $h(0) = 1$.

A. NL Weighted Matching Estimator

As in [19], [21], and [22], we assume that the channel order L is known. Methods for estimating L can be found in [10] and [25]. Alternatively, the parameter estimation method described herein could be modified to include the channel order L as an unknown parameter. This approach is described in [10]. The nonlinear weighted matching estimator finds the $\hat{\boldsymbol{\theta}}$ whose cyclic correlations $\boldsymbol{c}(\hat{\boldsymbol{\theta}})$ are closest to the observed cyclic correlations $\hat{\boldsymbol{c}}_y^{(T)}$ in a weighted least-squares sense. More precisely

$$\begin{aligned}\hat{\boldsymbol{\theta}} &\triangleq \arg\min_{\boldsymbol{\theta}} J_{\mathcal{Q}}[\hat{\boldsymbol{c}}^{(T)}; \boldsymbol{\theta}], \\ J_{\mathcal{Q}}[\hat{\boldsymbol{c}}_y^{(T)}; \boldsymbol{\theta}] &\triangleq [\hat{\boldsymbol{c}}_y^{(T)} - \boldsymbol{c}(\boldsymbol{\theta})]' \mathcal{Q} [\hat{\boldsymbol{c}}_y^{(T)} - \boldsymbol{c}(\boldsymbol{\theta})]\end{aligned} \tag{14}$$

where $\mathcal{Q}$ is a positive-definite weighting matrix. Note that $\mathcal{Q}$ can be a function of $\boldsymbol{\theta}_0$ or $\boldsymbol{\theta}$. Before stating a result on the consistency of this estimator, we define the set $\boldsymbol{\eta}$ to be the set of parameters $\boldsymbol{\theta}$ such that $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{c}(\boldsymbol{\theta}_0)$, i.e., $\boldsymbol{\eta} \triangleq [\boldsymbol{\theta} | \boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{c}(\boldsymbol{\theta}_0)]$.

*Theorem 1—*Consistency*:* The estimate $\hat{\boldsymbol{\theta}}$ obtained by minimizing $J_{\mathcal{Q}}[\hat{\boldsymbol{c}}_y^{(T)}; \boldsymbol{\theta}]$ over a compact set converges in the mean-square sense to the set $\boldsymbol{\eta}$ provided that i) $w(n)$ and $v(n)$ in (1) have finite moments, and ii) $\mathcal{Q}$ is positive definite.

Proof: (See also [15, p. 82–88] and [23].) If $w(n)$ and $v(n)$ have finite moments, then from [2], we have $\lim_{T\to\infty} \hat{\boldsymbol{c}}^{(T)} \overset{\mathrm{m.s.s}}{=} \boldsymbol{c}(\boldsymbol{\theta}_0)$, where $\overset{\mathrm{m.s.s}}{=}$ indicates convergence in the mean-square sense. Since $J_{\mathcal{Q}}[\cdot,\cdot]$ is a continuous function on a compact set of a consistent estimator, it follows that it is itself consistent, i.e.,

$$\begin{aligned}\lim_{T\to\infty} J_{\mathcal{Q}}[\hat{\boldsymbol{c}}^{(T)}; \boldsymbol{\theta}] &\overset{\mathrm{m.s.s}}{=} J_{\mathcal{Q}}[\boldsymbol{c}(\boldsymbol{\theta}_0); \boldsymbol{\theta}] \\ J_{\mathcal{Q}}[\boldsymbol{c}(\boldsymbol{\theta}_0); \boldsymbol{\theta}] &\triangleq [\boldsymbol{c}(\boldsymbol{\theta}_0) - \boldsymbol{c}(\boldsymbol{\theta})]' \quad \mathcal{Q}[\boldsymbol{c}(\boldsymbol{\theta}_0) - \boldsymbol{c}(\boldsymbol{\theta})].\end{aligned} \tag{15}$$

As a result of $\mathcal{Q}$ being positive definite, the minimum of $J_{\mathcal{Q}}[\boldsymbol{c}(\boldsymbol{\theta}_0); \boldsymbol{\theta}]$ is obtained if and only if $[\boldsymbol{c}(\boldsymbol{\theta}_0) - \boldsymbol{c}(\boldsymbol{\theta})] = 0$.

Equivalently, the minimum occurs for any $\boldsymbol{\theta}$ such that $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{c}(\boldsymbol{\theta}_0)$. The requirement $\boldsymbol{c}(\boldsymbol{\theta}) = \boldsymbol{c}(\boldsymbol{\theta}_0)$ in turn implies that any minimizer of $J_{\mathcal{Q}}[\boldsymbol{c}(\boldsymbol{\theta}_0); \boldsymbol{\theta}]$ is a member of $\boldsymbol{\eta}$, i.e., $\boldsymbol{\theta} \in \boldsymbol{\eta}$. Therefore, m.s.s. convergence of $\hat{\boldsymbol{\theta}}$ to the set $\boldsymbol{\eta}$ is established.□

Since $C_{11y}(k;m)$ and, thus, $J_{\mathcal{Q}}[\hat{\boldsymbol{c}}_y^{(T)}; \boldsymbol{\theta}]$ is nonconvex with respect to the parameters [c.f. (6)], a good initial estimate is *essential* in order to both speed the convergence and avoid spurious solutions (local minima) when seeking the global minimum. Although any of the second-order identification methods mentioned can be used to obtain an initial estimate, the linear cyclic approach of [5] or [21] makes a good choice because of its low additional computational complexity. Once the initial estimate is obtained, a nonlinear optimization method must be used in order to find the minimum of (14). Candidate optimization schemes include Gauss–Newton, the Marquardt–Levenberg, or the Nelder–Mead algorithms. The minimization will involve iterative search for the minimizer of $J_{\mathcal{Q}}[\hat{\boldsymbol{c}}_y^{(T)}; \boldsymbol{\theta}]$, where the ith update will be of the form

$$\hat{\boldsymbol{\theta}}^{(i+1)} = \hat{\boldsymbol{\theta}}^{(i)} + \mu_i \, \boldsymbol{g}_T^{(i)}\left(\hat{\boldsymbol{\theta}}^{(i)}\right) \tag{16}$$

where $\mu_i > 0$ is the stepsize, and $\boldsymbol{g}_T^{(i)}(\cdot)$ is a search direction that depends on the minimization algorithm being used. The computation of $\boldsymbol{g}_T^{(i)}(\cdot)$ may require the gradient $\nabla_{\boldsymbol{\theta}} J_{\mathcal{Q}}[\hat{\boldsymbol{c}}_y^{(T)}; \boldsymbol{\theta}]$ and, perhaps, the Hessian $\nabla^2_{\boldsymbol{\theta}} J_{\mathcal{Q}}[\hat{\boldsymbol{c}}_y^{(T)}; \boldsymbol{\theta}]$. Clearly, the speed of convergence (and hence the complexity of the NL matching approach) is dependent on the minimization method used and the quality of the initial estimate.

B. Asymptotic Covariance of Parameter Estimate

The normalized asymptotic covariance of any general parameter estimate $\hat{\boldsymbol{\xi}}$ of $\boldsymbol{\theta}_0$ is defined as

$$\boldsymbol{P}(\hat{\boldsymbol{\xi}}, \boldsymbol{\theta}_0) \triangleq \lim_{T\to\infty} TE\{(\hat{\boldsymbol{\xi}} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\xi}} - \boldsymbol{\theta}_0)'\}. \tag{17}$$

When the parameter estimates are functions of the cyclic correlation estimates $\hat{C}_{11y}(k;m)$, a bound on $\boldsymbol{P}(\hat{\boldsymbol{\xi}}, \boldsymbol{\theta}_0)$ follows readily from the general theorems in [15, p. 82] and (see also [16]) Theorem 2.

Theorem 2—Asymptotic Performance Bound*:* Let $\hat{\boldsymbol{c}}_y^{(T)}$ be the estimated cyclic correlations based on $\{y(n)\}_{n=0}^{T-1}$ with asymptotic covariance $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ defined in (13), and let $\hat{\boldsymbol{\xi}} = f(\hat{\boldsymbol{c}}_y^{(T)})$ be any general estimate of $\boldsymbol{\theta}_0$ based on the cyclic correlations. Provided $f(\cdot)$ is continuous and has continuous bounded partial derivatives of the first and second order, $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is full rank, and $\boldsymbol{F}(\boldsymbol{\theta}_0)$ has full-column rank, the asymptotic covariance of $\hat{\boldsymbol{\xi}}$ is bounded by

$$\boldsymbol{P}(\hat{\boldsymbol{\xi}}, \boldsymbol{\theta}_0) \geq \boldsymbol{B}(\boldsymbol{\theta}_0) \triangleq [\boldsymbol{F}'(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{F}(\boldsymbol{\theta}_0)]^{-1} \tag{18}$$

where

$$\boldsymbol{F}(\boldsymbol{\theta}_0) \triangleq [\nabla_{\theta(1)}\boldsymbol{c}(\boldsymbol{\theta}_0) \cdots \nabla_{\theta(L+2)}\boldsymbol{c}(\boldsymbol{\theta}_0)]. \tag{19}$$

When the estimator $f(\hat{\boldsymbol{c}}_y^{(T)})$ is of the form of (14), i.e., $\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\theta}}$, the asymptotic covariance for any weight matrix is given by the following theorem (see, e.g., [15, pp. 82–83] and [16]):

Theorem 3—Asymptotic Covariance for Weighted Matching Estimator*:* The asymptotic covariance $\boldsymbol{P}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)$ for the weighted matching approach in (14) is

$$\boldsymbol{P}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = \boldsymbol{G}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{G}'(\boldsymbol{\theta}_0) \tag{20}$$

where

$$\boldsymbol{G}(\boldsymbol{\theta}_0) \triangleq [\boldsymbol{F}'(\boldsymbol{\theta}_0)\mathcal{Q}\boldsymbol{F}(\boldsymbol{\theta}_0)]^{-1}\boldsymbol{F}'(\boldsymbol{\theta}_0)\mathcal{Q} \tag{21}$$

provided $\boldsymbol{F}(\boldsymbol{\theta}_0)$ is full column rank, and $\mathcal{Q}$ is positive definite.

Theorem 3 applies for any choice of weight matrix. However, if $\mathcal{Q} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$, Theorem 3 can be used to show that $\boldsymbol{P}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)$ achieves the lower bound $\boldsymbol{B}(\boldsymbol{\theta}_0)$. We summarize this result in the following theorem.

Theorem 4—Asymptotic Minimum Variance Unbiased Channel Estimator*:* The minimum asymptotic covariance $\boldsymbol{B}(\boldsymbol{\theta}_0)$ defined in Theorem 2 is achieved with equality for $f(\hat{\boldsymbol{c}}_y^{(T)})$ of the form of (14) when $\mathcal{Q} = \mathcal{Q}_{\text{opt}} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is full rank.

Proof: The proof follows from Theorem 2 with $\mathcal{Q} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ (see Appendix B). □

Theorem 4 implies that by choosing $\mathcal{Q} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$, the NL matching estimator in (14) is the *asymptotically minimum variance* unbiased estimator among the class of estimators based on second-order statistics (see Remark 1). Although it may appear that $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ depends on higher order moments, we show in Proposition 1 of Section IV-B that is not necessarily true. In fact, we show that in Gaussian noise, $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ only depends on second-order statistics. In other words, it is possible to achieve the minimum variance bound using an estimator that only relies on second-order statistics.

For the optimal weights, the evaluation of $\boldsymbol{g}_T^{(i)}(\cdot)$ in (16) will require the estimation of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ and its inverse at each step of the iteration. This is very computationally intensive. However, arguing as in [15, p. 84], it follows that the same asymptotic performance can be achieved when

$$\mathcal{Q} = [\hat{\boldsymbol{\Sigma}}^{(T)}]^{-1} \tag{22}$$

where $\hat{\boldsymbol{\Sigma}}^{(T)}$ is any consistent estimate of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ based on T data. The choice in (22) obviates the need to calculate a new matrix inverse at each step of the search. In Section IV-B, we discuss methods for estimating $\hat{\boldsymbol{\Sigma}}^{(T)}$ directly from the data. As with other cumulant matching methods [16], [17], if ill conditioning appears with $\mathcal{Q} = [\hat{\boldsymbol{\Sigma}}^{(T)}]^{-1}$ due to finite sample effects only, we recommend replacing any negative eigenvalues of $\mathcal{Q}$ with zero and using the pseudo-inverse in Theorems 2–4.

Since Theorem 3 applies for any choice of weight matrix $\mathcal{Q}$, another natural choice is to use the unweighted least-squares criterion in (14), i.e., $\mathcal{Q} = \boldsymbol{I}$. Note that $\mathcal{Q} = \boldsymbol{I}$ is equivalent (see Remark 1) to the methods of [12] and [22]. While the unweighted choice does not yield the minimum variance channel estimate, it is the simplest matching approach. The unweighted choice therefore retains the advantages inherent in matching approaches (see, e.g., Theorem 1) with the minimum additional computational complexity. Unlike the results of [12] and [27], Theorem 3 allows prediction of the asymptotic covariance $\boldsymbol{P}_{uw}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)$ of the unweighted estimator. By straightforward manipulation, $\boldsymbol{P}_{uw}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)$ is given by

$$\boldsymbol{P}_{uw}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = [\boldsymbol{F}'(\boldsymbol{\theta}_0)\boldsymbol{F}(\boldsymbol{\theta}_0)]^{-1}\boldsymbol{F}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\boldsymbol{F}'(\boldsymbol{\theta}_0) \times ([\boldsymbol{F}'(\boldsymbol{\theta}_0)\boldsymbol{F}(\boldsymbol{\theta}_0)]^{-1})'. \tag{23}$$

Remark 1: Although Theorems 2–4 only directly address estimators based on cyclic correlations, the one-to-one correspondence between $c_{11y}(n;m)$ and $C_{11y}(k;m)$ imply that these theorems also include estimators such as [13], [19], [20], and [21], which are based on $c_{11y}(n;m)$. The asymptotic covariance of estimates based on $c_{11y}(n;m)$ is also covered in [27] for additive white Gaussian noise and real channels.

C. Model Mismatch

In contrast with subspace and least-squares methods, if there is any deviation from the model assumptions, the matching approach retains a sense of optimality in that it finds the parameter vector $\hat{\boldsymbol{\theta}}$ whose cyclic correlations are closest to the observed cyclic correlations. The importance of this is most evident when examining the ID condition (see Section II-C). For channels that do not satisfy the ID condition, the second-order statistics fail to uniquely specify the channel. For subspace and least-squares methods, the estimates obtained from such statistics have no meaningful relationship to the true system $\boldsymbol{\theta}_0$. Theorem 1, however, shows that the matching method can still provide a useful estimate. In fact, Theorem 1 states that the matching methods are consistent in the sense that they always provide a member of $\boldsymbol{\eta}$. Furthermore, it was demonstrated in [4] and justified theoretically in [26] that even when channels satisfy the ID condition but are close to nonidentifiable, the variance of subspace and least-squares estimates increases dramatically. As was shown in [27], the matching methods do not suffer the same fate.

As stated previously, when a channel does not satisfy the ID condition, the observed cyclic correlations are not unique to that channel; rather, there is a finite set of channels that are capable of generating the same output cyclic correlations (see also [8]). More specifically, if a channel has N sets of $2\pi/P$ angularly spaced zeros, then there are $M = 2^N$ channels $\boldsymbol{\eta} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{M-1}\}$, which yield the same cyclic correlations, i.e.,

$$\boldsymbol{c}(\boldsymbol{\theta}_0) = \boldsymbol{c}(\boldsymbol{\theta}_i) \quad \forall \boldsymbol{\theta}_i \in \boldsymbol{\eta}.$$

To make the point clearer, consider the following example.

Example 1: Consider a FS system with $P = 2, \sigma_w^2 = 1$ and with a channel whose transfer function $H_0(z)$ is

$$H_0(z) = z^{-3}(z + 1.2)(z - 0.5j)(z + 0.5j).$$

In the absence of noise (i.e., $\sigma_v^2 = 0$), the parameter vector is given by $\boldsymbol{\theta}_0 = [1, 1.2, 0.25, 0.3]'$, which generates the cyclic correlation vector

$$\boldsymbol{c}(\boldsymbol{\theta}_0) = [1.2962, 0.7875, 0.3050, 0.1500, -0.2337, 0.4875, -0.0550, 0.1500].$$

However, the same cyclic correlations are generated for

$$\boldsymbol{\theta}_1 = [0.25, 0.3, 1.0, 1.2]'$$

which has a transfer function

$$H_1(z) = 0.25z^{-3}(z + 1.2)(z - 2j)(z + 2j).$$

In this example, $N = 1, M = 2$, and $\boldsymbol{\eta} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$. Provided the weight matrix $\mathcal{Q}$ is positive definite, there are only two possible solutions ($\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ for this example), which are global minima of $J_{\mathcal{Q}}[\hat{\boldsymbol{c}}^{(T)}; \boldsymbol{\theta}]$. The member of $\boldsymbol{\eta}$ found by (16) will depend, of course, on the initialization as well as the search algorithm used in the minimization. In contrast, linear and subspace methods [5], [13], [19], [20], [21], [25] are not guaranteed to converge to any of the members of $\boldsymbol{\eta}$ when the ID condition is not satisfied. Since an estimate obtained from these matching methods is at least member of $\boldsymbol{\eta}$, this estimate can be used to generate all the possible members of $\boldsymbol{\eta}$. Once the set $\boldsymbol{\eta}$ is known, a HOS-based *classification* method can be used to determine which member of $\boldsymbol{\eta}$ is the true channel [8]. HOS-based linear methods can also provide reliable initialization in order to facilitate convergence to the global minimum.

Remark 2: Zeng and Tong [26], [27] studied conditions under which the gradient matrix $\boldsymbol{F}(\boldsymbol{\theta}_0)$ is not full rank. If $\boldsymbol{F}(\boldsymbol{\theta}_0)$ and/or $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ are rank deficient, Theorems 2–4 and, hence, our asymptotic optimality claims do not hold. For any positive definite $\mathcal{Q}$, however, the matching estimator remains consistent in that it finds a member of the set $\boldsymbol{\eta}$. Hence, the choice $\mathcal{Q} = \boldsymbol{I}$ is recommended.

D. Algorithm

In summary, the steps of the proposed matching algorithm for the parameter estimation are as follows:

Step 1: Compute $\hat{C}_{11y}(k;m)$ for the nonredundant values (e.g., for the sets $\mathcal{K}$ or $\overline{\mathcal{K}}$) according to (12).

Step 2: Calculate the weight matrix from the data. For the optimal choice, estimate $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ (see Section V) and calculate its inverse. For the unweighted choice, simply use $\mathcal{Q} = \boldsymbol{I}$.

Step 3: Find $\hat{\boldsymbol{\theta}}_0$ by solving (14) according to (16).

IV. Asymptotic Covariance of Cyclic Correlations

From the previous sections, we see that the asymptotic covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ of the sample cyclic correlation vector $\hat{\boldsymbol{c}}_y^{(T)}$ plays an important role in

i) selection of optimal weights (c.f. Theorem 4),
ii) performance analysis of the channel estimates (c.f. Theorem 3),
iii) calculation of the bound on performance for all second-order-based blind estimation methods (c.f. Theorem 2).

In addition to proving mean-square consistency and asymptotic normality, Dandawaté and Giannakis [2] provided equations to calculate the covariance for cyclic cumulant estimators of any order. In this section, we specialize the general expressions of [2] for the cyclic correlation estimator used in (12). In addition, we present consistent estimators $\hat{\boldsymbol{\Sigma}}^{(T)}$ of the asymptotic covariance matrix.

A. Asymptotic Covariance Matrix: Theoretical Calculations

Since the cyclic correlation is in general a complex quantity, it converges asymptotically to a complex Gaussian distribution [1, p. 89]. It is important to carefully define the covariance when dealing with complex distributions. This was the motivation for splitting the cyclic correlation vector into its real and imaginary parts, as in (11). For such a choice, the asymptotic covariance in (13) can be written as a block matrix:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \begin{bmatrix} \mathrm{Re}\left\{\dfrac{\boldsymbol{\Gamma}+\overline{\boldsymbol{\Gamma}}}{2}\right\} & \mathrm{Im}\left\{\dfrac{\boldsymbol{\Gamma}-\overline{\boldsymbol{\Gamma}}}{2}\right\} \\ \mathrm{Im}\left\{\dfrac{\boldsymbol{\Gamma}+\overline{\boldsymbol{\Gamma}}}{2}\right\} & \mathrm{Re}\left\{\dfrac{\overline{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}}{2}\right\} \end{bmatrix} \tag{24}$$

with the matrices $\boldsymbol{\Gamma}$ and $\overline{\boldsymbol{\Gamma}}$ representing, respectively, the "unconjugated"

$$\boldsymbol{\Gamma} \triangleq \lim_{T\to\infty} TE\{[\hat{\tilde{\boldsymbol{c}}}_y^{(T)} - \tilde{\boldsymbol{c}}(\boldsymbol{\theta}_0)][\hat{\tilde{\boldsymbol{c}}}_y^{(T)} - \tilde{\boldsymbol{c}}(\boldsymbol{\theta}_0)]'\} \tag{25}$$

and the "conjugated" asymptotic covariance

$$\overline{\boldsymbol{\Gamma}} \triangleq \lim_{T\to\infty} TE\{[\hat{\tilde{\boldsymbol{c}}}_y^{(T)} - \tilde{\boldsymbol{c}}(\boldsymbol{\theta}_0)][\hat{\tilde{\boldsymbol{c}}}_y^{(T)} - \tilde{\boldsymbol{c}}(\boldsymbol{\theta}_0)]^{*'}\}. \tag{26}$$

Note that in general, $\boldsymbol{\Gamma} \neq \overline{\boldsymbol{\Gamma}}^*$, but $\boldsymbol{\Gamma} = \overline{\boldsymbol{\Gamma}}$ when the cyclic correlations are real.

Defining the ith entry of the vector $\hat{\tilde{\boldsymbol{c}}}_y^{(T)}$ as $[\hat{\tilde{\boldsymbol{c}}}_y^{(T)}]_i$, we note that

$$[\hat{\tilde{\boldsymbol{c}}}_y^{(T)}]_i = \hat{C}_{11y}(k_i; m_i)$$

where

$$k_i = \lfloor i/(L+1) \rfloor, \quad m_i = i - k_i(L+1).$$

Therefore, the (i,j)th entry of $\boldsymbol{\Gamma}$ is

$$\begin{aligned}[\boldsymbol{\Gamma}]_{ij} &= \lim_{T\to\infty} TE\{[\hat{C}_{11y}(k_i;m_i) - C_{11y}(k_i;m_i)] \\ &\quad \times [\hat{C}_{11y}(k_j;m_j) - C_{11y}(k_j;m_j)]\} \\ &= \lim_{T\to\infty} T\,\mathrm{cum}\{\hat{C}_{11y}(k_i;m_i), \hat{C}_{11y}(k_j;m_j)\} \end{aligned} \tag{27}$$

$$\begin{aligned}[\overline{\boldsymbol{\Gamma}}]_{ij} &= \lim_{T\to\infty} TE\{[\hat{C}_{11y}(k_i;m_i) - C_{11y}(k_i;m_i)] \\ &\quad \times [\hat{C}^*_{11y}(k_j;m_j) - C^*_{11y}(k_j;m_j)]\} \\ &= \lim_{T\to\infty} T\,\mathrm{cum}\{\hat{C}_{11y}(k_i;m_i), \hat{C}^*_{11y}(k_j;m_j)\}. \end{aligned} \tag{28}$$

In order to find $[\boldsymbol{\Gamma}]_{ij}$ and $[\overline{\boldsymbol{\Gamma}}]_{ij}$, we first define

$$\begin{aligned} S_{2y}(k,\ell;m_1,m_2) &\triangleq \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T-1} \sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn} e^{-j(2\pi/P)\ell\xi} \\ &\quad \times \mathrm{cum}\{y(n)y^*(n+m_1), \\ &\quad \times y(n+\xi)y^*(n+\xi+m_2)\} \end{aligned} \tag{29}$$

$$\begin{aligned} \overline{S}_{2y}(k,\ell;m_1,m_2) &\triangleq \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T-1} \sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn} e^{-j(2\pi/P)\ell\xi} \\ &\quad \times \mathrm{cum}\{y(n)y^*(n+m_1) \\ &\quad \times y^*(n+\xi)y(n+\xi+m_2)\}. \end{aligned} \tag{30}$$

Following directly from [2, Theorem 2.1] for the estimator of (12), we have the following result:

$$\begin{aligned} S_{2y}(k+\ell,\ell;m_1,m_2) &= \lim_{T\to\infty} \{\hat{C}_{11y}(k;m_1), \hat{C}_{11y}(\ell;m_2)\} \end{aligned} \tag{31}$$

$$\begin{aligned} \overline{S}_{2y}(k-\ell,-\ell;m_1,m_2) &= \lim_{T\to\infty} T\,\mathrm{cum}\{\hat{C}_{11y}(k;m_1), \hat{C}^*_{11y}(\ell;m_2)\}. \end{aligned} \tag{32}$$

Using the structure of the FS system, we establish the following result for the estimator in (12):

Proposition 1: For data $y(n)$ that satisfy the mixing conditions of [2] and are generated according to (1), the asymptotic covariance terms (29) and (30) can be found from

$$\begin{aligned} S_{2y}(k,\ell;m_1,m_2) &= \frac{Pc_{22w}(0;0,0)}{[\sigma_w^2]^2} \\ &\quad \times C_{11x}(\ell;m_2)C_{11x}(k-\ell;m_1) \\ &\quad + c_{22v}(0;0,0)\delta(m_1)\delta(m_2)\delta(k) \\ &\quad + \sum_{\xi} e^{-j(2\pi/P)\ell\xi} \sum_{q=0}^{P-1} [C_{02y}(q;m_1-m_2-\xi) \\ &\quad \times C_{20y}(k-q;\xi)e^{j(2\pi/P)q(\xi+m_2)} + C_{11y}(q;m_1-\xi) \\ &\quad \times C_{11y}(k-q;\xi+m_2)e^{j(2\pi/P)q\xi}] \end{aligned} \tag{33}$$

and

$$\begin{aligned} \overline{S}_{2y}(k,\ell;m_1,m_2) &= \frac{Pc_{22w}(0;0,0)}{[\sigma_w^2]^2} \\ &\quad \times C_{11x}(\ell;-m_2)C_{11x}(k-\ell;m_1)e^{j(2\pi/P)\ell m_2} \\ &\quad + c_{22v}(0;0,0)\delta(m_1)\delta(m_2)\delta(k) \\ &\quad + \sum_{\xi} e^{-j(2\pi/P)\ell\xi} \sum_{q=0}^{P-1} [C_{11y}(q;m_1-m_2-\xi) \\ &\quad \times C_{11y}(k-q;\xi)e^{j(2\pi/P)q(\xi+m_2)} \\ &\quad + C_{20y}(q;m_1-\xi)C_{02y}(k-q;\xi+m_2) \\ &\quad \times e^{j(2\pi/P)q\xi}] \end{aligned} \tag{34}$$

where the $(k+\ell)$th-order cyclic cumulant is defined as

$$\begin{aligned} C_{k\ell y}(k;m_1,\cdots,m_{k-1};m'_1,\cdots,m'_\ell) &\triangleq \frac{1}{P} \sum_{n=0}^{P-1} c_{k\ell y}(n;m_1,\cdots,m_{k-1};m'_1,\cdots \\ &\quad m'_\ell)e^{-j(2\pi/P)kn} \end{aligned}$$

and the time-invariant fourth-order cumulants of input and noise are defined as

$$\begin{aligned} c_{22w}(m_1;m_2,m_3) &\triangleq \mathrm{cum}\,(w(n),w(n+m_1),w^*(n+m_2)w^*(n+m_3)) \\ c_{22v}(m_1;m_2,m_3) &\triangleq \mathrm{cum}\,(v(n),v(n+m_1),v^*(n+m_2)v^*(n+m_3)). \end{aligned}$$

Proof: See Appendix C. □

In many practical communications situations, it is possible to simplify (33) and (34).

1) Gaussian Noise: In communications, the additive noise is frequently modeled as Gaussian. When the additive noise is Gaussian, it is possible to simplify (33) and (34) since the higher (than second)-order cumulants of Gaussian processes are zero [1, ch. 2] $c_{22v}(m_1;m_2,m_3)=0\ \forall m_1,m_2,m_3$. We note that in this case, no higher order terms are required.

2) Real Input/Real Systems: If $y(n)$ is real, then

$$C_{20y}(k;m)=C_{02y}(k;m)=C_{11y}(k;m).$$

If $P=2$, we note that

$$\exp\{-j(2\pi/P)kn\}=\exp\{-j\pi kn\}=(-1)^{kn}$$

and, hence, the cyclic correlations are real [see (3)]. This implies that

$$S_{2y}(k,\ell;m_1,m_2)=\overline{S}_{2y}(k,\ell;m_1,m_2).$$

3) Complex Symmetric Input: If the input $w(n)$ is an i.i.d. complex symmetric random process (e.g., QAM or QPSK signal constellation), then it is straightforward to show that

$$C_{02x}(k;m)=C_{20x}(k;m)=0 \quad \forall k,\ell.$$

The case of white Gaussian noise and real input/systems is also considered in [27].

B. Sample Estimation of the Asymptotic Covariance

In (22), we saw that it made sense computationally to choose $\mathcal{Q}=[\hat{\boldsymbol{\Sigma}}^{(T)}]^{-1}$, where $[\hat{\boldsymbol{\Sigma}}^{(T)}]$ was a consistent estimate of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. In this section, we describe three possible methods for estimating $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ from the output data.

Method 1—Data Based: When estimates of the $c_{22v}(0;0,0)$ and σ_v^2 are available or if the noise is Gaussian and an estimate of σ_v^2 is available (e.g., when noise only data is available), then (33) and (34) provide a method for data-based estimation of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. To estimate $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ from $\{y(n)\}_{n=0}^{T-1}$ using (33) and (34), estimates of the quantities $C_{11x}(k;m), C_{11y}(k;m), C_{20y}(k;m)$, and $C_{02y}(k;m)$ are needed. An estimate of $C_{11y}(k;m)$ is obtained from (12), whereas consistent estimates of $C_{20y}(k;m)$ and $C_{02y}(k;m)$ are found by (see, e.g., [2])

$$\hat{C}_{02y}(k;m) \triangleq \frac{1}{T}\sum_{n=0}^{T-1} y^*(n)y^*(n+m)e^{-j(2\pi/P)kn} \tag{35}$$

$$\hat{C}_{20y}(k;m) \triangleq \frac{1}{T}\sum_{n=0}^{T-1} y(n)y(n+m)e^{-j(2\pi/P)kn}. \tag{36}$$

Finally, an estimate of $C_{11x}(k;m)$ can be found by using [see (3)]

$$\hat{C}_{11x}(k;m) \triangleq \hat{C}_{11y}(k;m)-\sigma_v^2\delta(m)\delta(k). \tag{37}$$

Then, (33) and (34) can be used to generate $\hat{S}_{2y}(k;\ell,m_1,m_2)$ and $\hat{\overline{S}}_{2y}(k,\ell;m_1,m_2)$, respectively. From these values, $\boldsymbol{\Gamma}$ and $\overline{\boldsymbol{\Gamma}}$ can be found by applying (27) and (28). Finally, $\hat{\boldsymbol{\Sigma}}^{(T)}$ can be found from (24). Note that for high SNR, $c_{22v}(0;0,0)\approx 0$, and $\sigma_v^2\approx 0$. As a consequence, $\hat{C}_{11x}(k;m)\approx\hat{C}_{11y}(k;m)$. In this case, it is sufficient to use $\hat{C}_{11y}(k;m)$ in place of $\hat{C}_{11x}(k;m)$, which obviates the need for (37) or knowledge of the noise statistics. This approximation is justified experimentally in Experiment 1 of Section V.

To summarize, we have the following procedure for estimating $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ from the observations $\{y(n)\}_{n=0}^{T-1}$.

Step 1: From the observations $\{y(n)\}_{n=0}^{T-1}$, calculate $\hat{C}_{11y}(k;m), \hat{C}_{02y}(k;m)$ and $\hat{C}_{20y}(k;m)$ according to (12), (35), and (36), respectively. Using $\hat{C}_{11y}(k;m)$ and σ_v^2, estimate $C_{11x}(k;m)$ from (37).

Step 2: Form $\hat{S}_{2y}(k;\ell;m_1,m_2)$ and $\hat{\overline{S}}_{2y}(k,\ell;m_1,m_2)$ using the estimated values from Step 1 in (33) and (34).

Step 3: Form the matrices $\boldsymbol{\Gamma}$ and $\overline{\boldsymbol{\Gamma}}$ according to (27) and (28), and form $\hat{\boldsymbol{\Sigma}}^{(T)}$ by using (24).

Method 2—Parameter Based: The initial estimate $\hat{\boldsymbol{\theta}}_0^{(0)}$ of $\boldsymbol{\theta}_0$, which is generally used to speed convergence of the nonlinear minimization, provides a convenient parameter-based method for estimation of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. In other words, using $\{\hat{h}(n)\}_{n=0}^{L}$ and $\hat{\sigma}_v^2$ from $\hat{\boldsymbol{\theta}}_0^{(0)}$, (6) can be used to find $\hat{C}_{11y}(k;m)$. Similar to (6), $C_{20y}(k;m)$ and $C_{02y}(k;m)$ can be expressed in terms of the channel as

$$C_{20y}(k;m)=\frac{c_{20w}(0)}{P}\sum_{n=0}^{L}h(n)h(n+m)e^{-j(2\pi/P)kn} + \sigma_v^2\delta(m)\delta(k) \tag{38}$$

$$C_{02y}(k;m)=\frac{c_{20w^*}(0)}{P}\sum_{n=0}^{L}h^*(n)h^*(n+m)e^{-j(2\pi/P)kn} + \sigma_v^2\delta(m)\delta(k). \tag{39}$$

Estimates $\hat{C}_{20y}(k;m)$ and $\hat{C}_{02y}(k;m)$ are obtained by using $\{\hat{h}(n)\}_{n=0}^{L}$ and $\hat{\sigma}_v^2$ in place of the ensemble quantities in (38) and (39). In addition, $\hat{C}_{11x}(k;m)$ is obtained from (37). Following the same procedures as described in Steps 2 and 3 of Method 1, an estimate of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ can be found. Note that if $x(n)$ and $v(n)$ are complex (circularly) symmetric, $C_{20x}=C_{02x}=C_{20v}=C_{02v}=0$.

Method 3—Higher Order Statistics Based: When neither estimates of the noise statistics nor initial estimates of $\boldsymbol{\theta}_0$ are available, it is still possible to estimate $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ from $\{y(n)\}_{n=0}^{T-1}$; however, such a procedure relies on higher order statistics. Since the focus of this paper is on blind identification from second-order statistics, we will only briefly describe the HOS-based method. Starting from the general expressions for $S_{2y}(k,\ell;m_1,m_2)$ and $\overline{S}_{2y}(k,\ell;m_1,m_2)$ (see, e.g., (46) and (47) in the Appendix), estimates of $C_{11y}(k;m), C_{20y}(k;m), C_{02y}(k;m)$, and $C_{22y}(k;m_1;m_2,m_3)$ are needed. The first three $C_{11y}(k;m), C_{20y}(k;m)$, and $C_{02y}(k;m)$ can be estimated from (12), (36), and (35), respectively, whereas the estimation of $C_{22y}(k;m_1;m_2,m_3)$ is described in [2]. Once estimates of $S_{2y}(k,\ell;m_1,m_2)$ and $\overline{S}_{2y}(k,\ell;m_1,m_2)$ are obtained, the

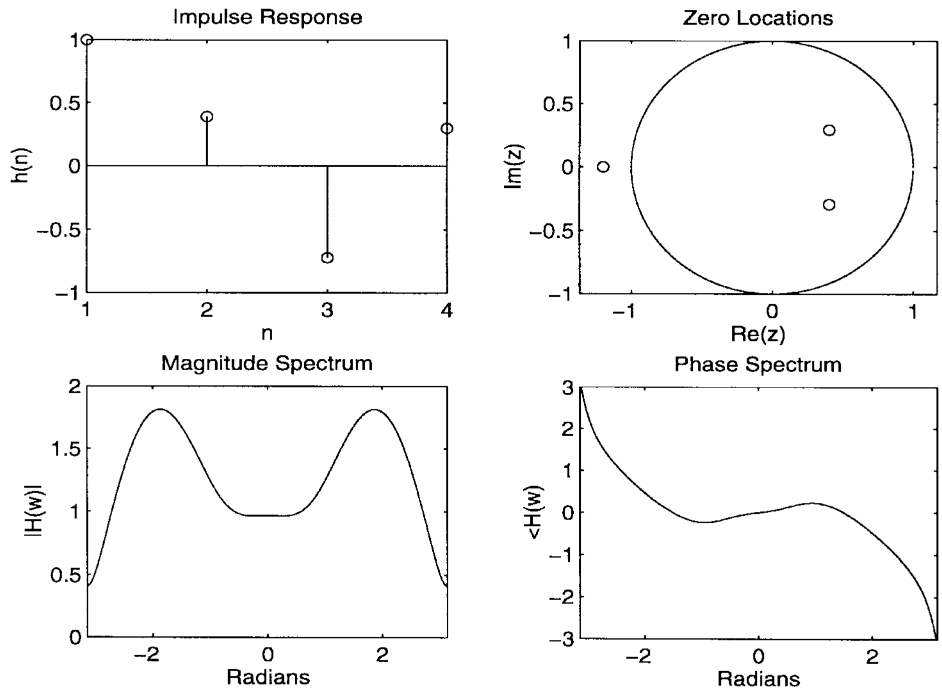


Fig. 3. Channel characteristics for experiments 1–3.

same procedure as described in Step 3 of Method 1 can be used to find $\hat{\boldsymbol{\Sigma}}^{(T)}$.

V. Simulations

In this section, we examine the asymptotic covariance both for the cyclic correlation estimates, $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, and for the NL weighted channel estimates, $\boldsymbol{P}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)$, through simulations. In addition, we compare the NL weighted channel estimates with several existing FS channel estimators at different SNR's. The noise statistics could be part of the unknown parameters as we indicated in Section II, but in order to focus on parameter identifiability issues, we assumed them known in our experiments. Finally, we look at the performance for channels which are close to nonidentifiable and those that are nonidentifiable from second-order statistics.

For all experiments, the input sequence $w(n)$ is an i.i.d. BPSK $\pm$ 1 sequence while the additive noise, $v(n)$, is a white Gaussian sequence. The oversample rate used was 2 (i.e., $P = 2$). For Experiments 1–3, the channel is order 3 (i.e., $L = 3$), and the impulse response vector $\boldsymbol{\theta}_0 = [1, 0.3910, -0.7208, 0.3000]'$. A plot of this channel and its frequency domain characteristics is shown in Fig. 3. In all cases, the NL minimization used is a Gauss-Newton gradient method which was chosen for simplicity in implementation. To resolve the scale ambiguity inherent with any blind channel estimator, all estimates $\hat{\boldsymbol{\theta}}$ were normalized so that $\hat{h}(0) = 1$.

Experiment 1: In this experiment, we examine the asymptotic covariance of sample estimates of the cyclic correlation. The cyclic correlations $C_{11y}(k;m)$ were estimated according to (12) for $k = 0, 1$ and $m = -3, -2, \cdots, 2, 3$ for T data observations. The variance of the cyclic correlation estimates was computed from 1000 Monte Carlo runs. Table I shows the observed variances of the cyclic correlation estimates for

TABLE I
Large Sample Variance of $\hat{C}_{11y}(k;m)$: $T = 600$

Parameter	Variance at 5 dB		Variance at 10 dB	
	Experimental	Theoretical	Experimental	Theoretical
$C_{11y}(0;0)$	1.8872	1.8661	1.0968	1.0546
$C_{11y}(0;1)$	0.4748	0.4642	0.1690	0.1735
$C_{11y}(0;2)$	2.2752	2.3040	1.8813	1.9883
$C_{11y}(0;3)$	0.6721	0.6523	0.2221	0.2195
$C_{11y}(1;0)$	2.5041	2.5425	1.7555	1.7310
$C_{11y}(1;1)$	0.7838	0.7942	0.2819	0.2735
$C_{11y}(1;2)$	1.7651	1.7350	1.2992	1.3293
$C_{11y}(1;3)$	0.8429	0.8220	0.4182	0.4165

$T = 600$ and an SNR of 5 and 10 dB. The predicted variances were found by solving (33) for the appropriate lags m and cycles k. Fig. 4 shows the predicted (dashed line) and the observed variances for $T = 1000$ across a range of SNR's. The values shown are $C_{11y}(0;0), C_{11y}(0;3), C_{11y}(1;0)$, and $C_{11y}(1;2)$. From Table I, we see that for as few as 600 samples, our estimates are very close to their asymptotic predicted values. Fig. 4 shows a similar result for a wide range of SNR's and shows that the asymptotic variance above 15 dB is not significantly affected by the noise level.

Experiment 2: In this experiment, we examine the predicted asymptotic covariance of the channel estimates obtained from the cyclic matching algorithm. We will look at both the unweighted case and the optimal weighted case. The sample estimates $\hat{C}_{11y}(k;m)$ were obtained from (12) to form

$$\hat{\boldsymbol{c}}_y^{(T)} = [\hat{C}_{11y}(0;0), \hat{C}_{11y}(0;1), \hat{C}_{11y}(0;2), \hat{C}_{11y}(0;3) \\ \hat{C}_{11y}(1;0), \cdots, \hat{C}_{11y}(1;2)]'.$$

Note that we have omitted the $\hat{C}_{11y}(1;3)$ term from the parameter vector since for this particular system $\hat{C}_{11y}(0;3) = \hat{C}_{11y}(1;3)$.

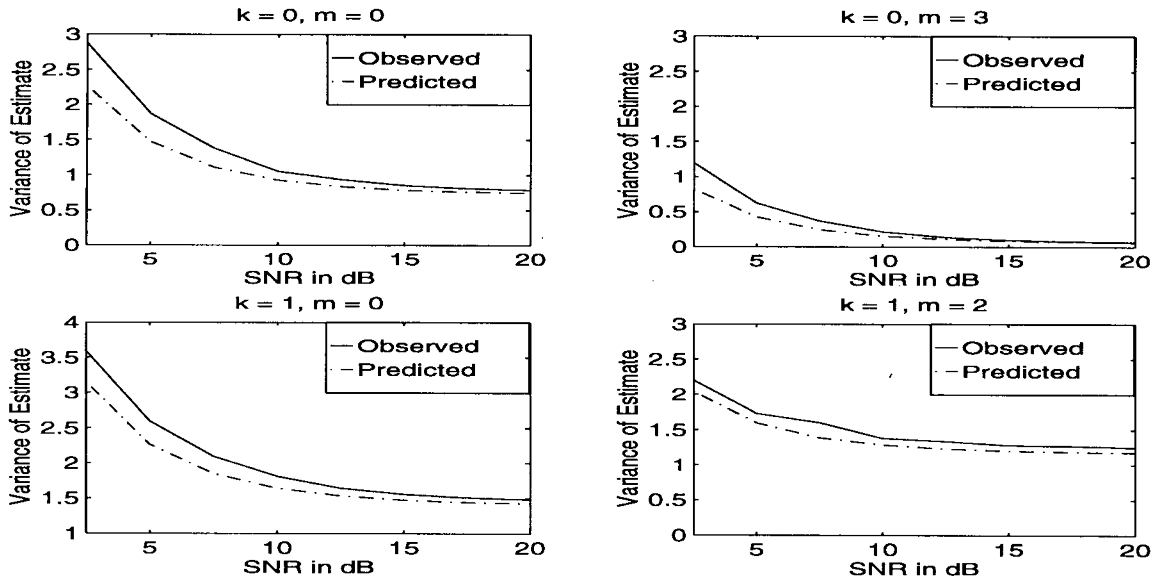


Fig. 4. Large sample variance of cyclic correlation: $\hat{C}_{11y}(k;m), T = 1000$.

TABLE II
EXPERIMENT 2: VARIANCE OF CHANNEL ESTIMATES AT SNR = 5 dB, $T = 1000$

Parameter	Optimal Weighted		Unweighted	
	Experimental	Theoretical	Experimental	Theoretical
$h(1)$	0.69	0.69	0.98	0.87
$h(2)$	2.07	2.03	4.58	3.34
$h(3)$	0.96	0.93	1.14	0.81

TABLE III
EXPERIMENT 2: VARIANCE OF CHANNEL ESTIMATES FOR SNR = 10 dB, $T = 1000$

Parameter	Optimal Weighted		Unweighted	
	Experimental	Theoretical	Experimental	Theoretical
$h(1)$	0.21	0.20	0.25	0.26
$h(2)$	0.98	0.82	2.89	2.82
$h(3)$	0.31	0.31	0.39	0.35

For the optimal approach, $\hat{\boldsymbol{\Sigma}}^{(T)}$ was estimated as described in Steps 1-3 of Section IV-B to form the weight matrix $\mathcal{Q} = [\hat{\boldsymbol{\Sigma}}^{(T)}]^{-1}$. For the unweighted approach, we used $\mathcal{Q} = \boldsymbol{I}$. The predicted asymptotic variance values were calculated for the unknown channel impulse response coefficients $h(1), h(2)$, and $h(3)$ using (18) and (23). The observed variances were calculated from 100 Monte Carlo runs for $T = 1000$ at SNR $= 5$ dB (see Table II) and SNR $= 10$ dB (see Table III) according to

$$\hat{\text{var}}\,\{\hat{h}(i)\} = T\frac{1}{100}\sum_{r=1}^{100}[\hat{h}_r(i) - h(i)]^2$$

where $\hat{h}_r(i)$ is the estimate of $h(i)$ obtained on the rth Monte Carlo run.

Tables II and III show that the predicted asymptotic covariances for the channel estimates can be obtained for practical data lengths. In addition, we see improvement obtained by the weighted approach over the unweighted approach.

Experiment 3: In this experiment, we compare the performance of the cyclic matching algorithms with existing FS channel estimators. Using the same procedures as in Experiment 2, we compare the unweighted and optimally weighted estimators with the subspace method of Moulines *et al.* [13] for a window size of 4, the deterministic method of Xu *et al.* [25], and the linear cyclic method [5] using cyclic correlations with lags $m = 0, 1, 2,$ and 3. Tables IV and V show the observed and predicted variances for the matching algorithms as well as the observed variances for the other methods calculated from 200 Monte Carlo runs. For Table IV, the data length used was $T = 250$, and the SNR was 12.5 dB. For Table V, the data length used was $T = 500$, and the SNR was again 12.5 dB. Fig. 5 shows the average mean bias across a range of data lengths. The mean bias was calculated from [13]:

$$\bar{b}_h \triangleq \frac{1}{100}\sum_{i=1}^{3}\left|\sum_{r=1}^{100}[h_r(i) - h(i)]\right|.$$

Note that Fig. 5 does not include the mean bias for the Xu *et al.* [25] estimator. Its observed bias was much larger than the values shown in Fig. 5. Tables IV and V show that the weighted matching approach for this data length does at least as well as the subspace method and did better than the other methods. The unweighted one performs well, although the var $\{\hat{h}(2)\}$ is large. From Fig. 5, we conclude that the subspace method and the weighted approach have approximately the same bias for large sample sizes and that both methods outperform the other methods for small data sizes.

Experiment 4: In this experiment, we study the behavior of the proposed algorithms and the existing algorithms for channels that do not satisfy the ID condition. For the nonidentifiable channel described in Example 1, Table VI shows the variance of the various algorithms for $T = 2000$ and at SNR $= 10$ dB. Table VII shows the mean bias for the different algorithms. In each Monte Carlo run and for every method

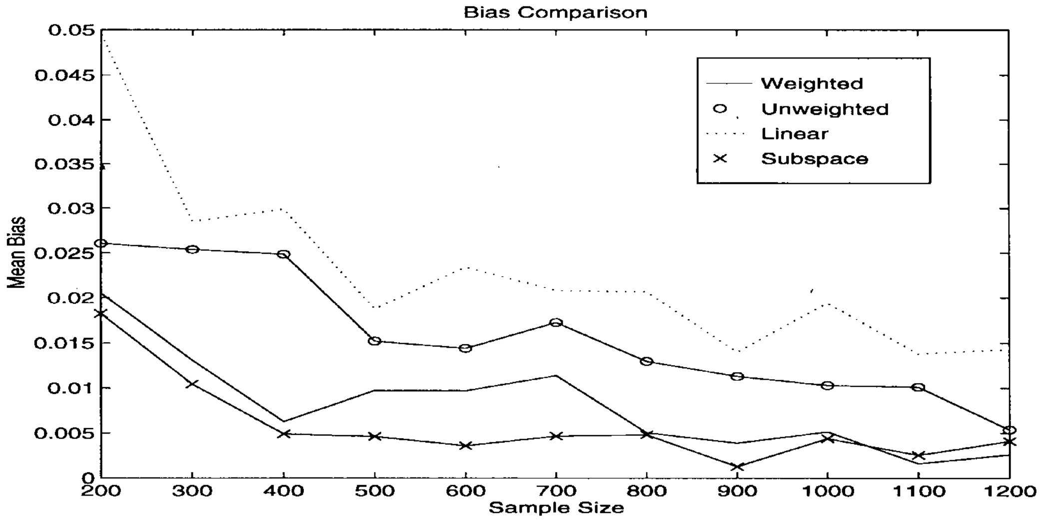


Fig. 5. Experiment 3: Bias of channel estimates.

TABLE IV
EXPERIMENT 3: VARIANCE OF CHANNEL ESTIMATES FOR $T = 250$

Parameter	Theoretical Variances		Estimated Variances				
	Optimal	Unweighted	Optimal	Unweighted	Linear [5]	XLTK [25]	Moulines [13]
$h(1)$	0.11	0.16	0.11	0.17	0.23	0.16	0.14
$h(2)$	0.50	2.72	0.55	2.61	1.12	0.56	0.51
$h(3)$	0.18	0.27	0.18	0.24	0.22	0.18	0.19

TABLE V
EXPERIMENT 3: VARIANCE OF CHANNEL ESTIMATES FOR $T = 500$

Parameter	Theoretical Variances		Estimated Variances				
	Optimal	Unweighted	Optimal	Unweighted	Linear [5]	XLTK [25]	Moulines [13]
$h(0)$	0.11	0.16	0.12	0.14	0.32	0.17	0.13
$h(1)$	0.50	2.72	0.60	2.78	0.85	0.55	0.50
$h(2)$	0.18	0.27	0.18	0.24	0.22	0.18	0.18

tested, $h(i)$ was considered to be the element of $\boldsymbol{\eta}$ closest to the obtained $\boldsymbol{\theta}$. To study the performance of these algorithms as the channel approaches nonidentifiability, consider a channel $H_\beta(z)$ whose transfer function has zeros at

$$\zeta_0 = -1.2, \quad \zeta_1 = 0.5\exp\{j\beta\}, \quad \zeta_2 = 0.5\exp\{-j\beta\}.$$

This channel is parameterized by β. When $\beta = \pi/2$, the channel does not satisfy the identifiability condition and coincides with the nonidentifiable channel in Example 1. Figs. 6–8 show the predicted asymptotic variance of the weighted and unweighted approach for the impulse response coefficient estimates $\hat{h}(1), \hat{h}(2)$, and $\hat{h}(3)$, respectively, as β varies.

Tables VI and VII show that the estimates from existing approaches are strongly biased (i.e., have large mean-square error) and/or strongly inconsistent for nonidentifiable channels. Estimates from the matching approaches, however, have small variance and are consistent as described in Section III-C. Figs. 6–8 show that the variance of the estimates increases but remains finite as the channels approach nonidentifiability. The true model here had $h_0(0) = 1$, and the normalization $\hat{h}(0) = 1$ was used in each iteration of the search. To cover the general case though, where one may have $h_0(0) \neq 1$, the normalization Im $[\hat{h}(0) = 1]$ is recommended.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

This paper presents a cyclic correlation weighted matching algorithm for estimating an unknown channel from FS output data only. Due to the nonlinear minimization involved, the weighted matching estimator may be more computationally intensive than some existing approaches. However, with a proper choice of weights, the weighted matching estimator is optimal in the sense that it yields the smallest asymptotic variance estimate of the channel impulse response. For this optimal case and for the more general weighted case, exact expressions are given for finding the asymptotic variance of the channel estimates. In the optimal case, these variance calculations serve as a lower bound on the variance of any channel estimate based on second-order statistics. In addition, for any choice of weights, the matching approach provides useful estimates when the channel is not identifiable from only the output correlations. Existing algorithms do not have this sense of optimality when the identifiability condition fails.

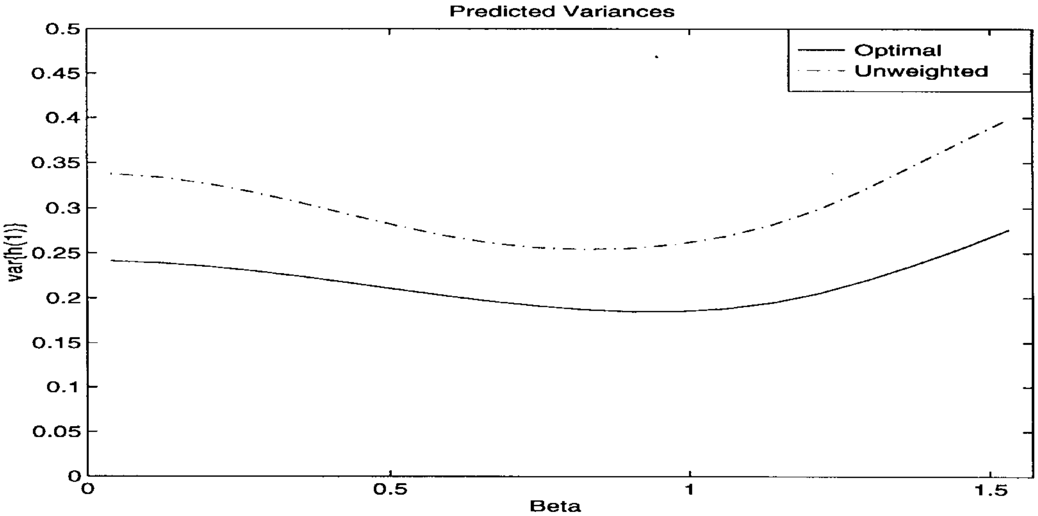


Fig. 6. Theoretical asymptotic variance: SNR = 10 dB.

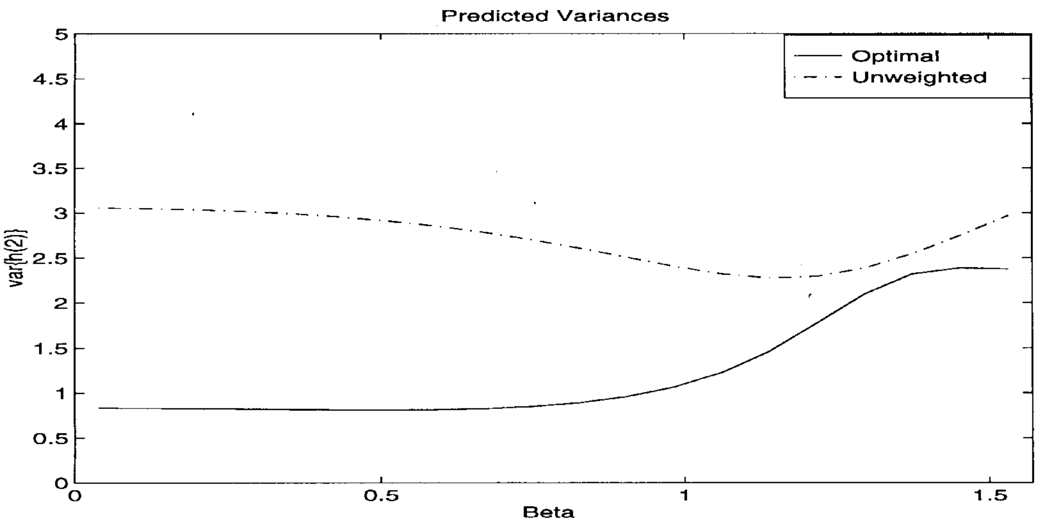


Fig. 7. Theoretical asymptotic variance: SNR = 10 dB.

While the minimum variance property is only assured for large sample sizes, simulations show that the optimal algorithm performs at least as well as a successful subspace method for short data lengths.

We present equations for the individual entries of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$, which can be expressed in terms of the channel, the input statistics, and the noise statistics. An interesting topic for future research is to express the matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ in terms of the zeros of the channel. This would allow one to analyze the asymptotic covariance [e.g., either $\boldsymbol{B}(\boldsymbol{\theta}_0)$ or $\boldsymbol{P}_{uw}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)$] of the channel estimates as the channel approaches nonidentifiable. In addition, it is possible to extend this weighted matching approach to include colored input. Finally, comparisons between the performance of the matching approaches and the conditional maximum likelihood approach of [11] make interesting topics for further research.

Appendix A
Cyclic Correlations

For the convenience of the reader, we show the derivation of the cyclic correlation. For an alternative derivation, see [3] and [21]. Substitution of (3) into (7) gives

$$C_{11y}(k;m) = \frac{\sigma_w^2}{P}\sum_{n=0}^{P-1}\sum_{\ell} h(n-\ell P)h^*(n-\ell P+m) \times e^{-j(2\pi/P)kn} + \frac{1}{P}\sum_{n=0}^{P-1} c_{11v}(m)e^{-j(2\pi/P)kn}. \tag{40}$$

Since $c_{11v}(m)$ is independent of n, the second term will become $c_{11v}(m)\delta(k)$, where $\delta(k)$ is the Kronecker delta and where we have used

$$(1/P)\sum_{n=0}^{P-1}\exp\left\{-j\frac{2\pi}{P}nk\right\} = \delta(k).$$

In order to simplify the first term in (40), consider the property of integers that allows any integer s to be written as $s = (-\ell)P + n$, where $n = 0, 1, 2, \cdots, P-1$, and ℓ and P are integers. Furthermore, we can rewrite the sum of any function

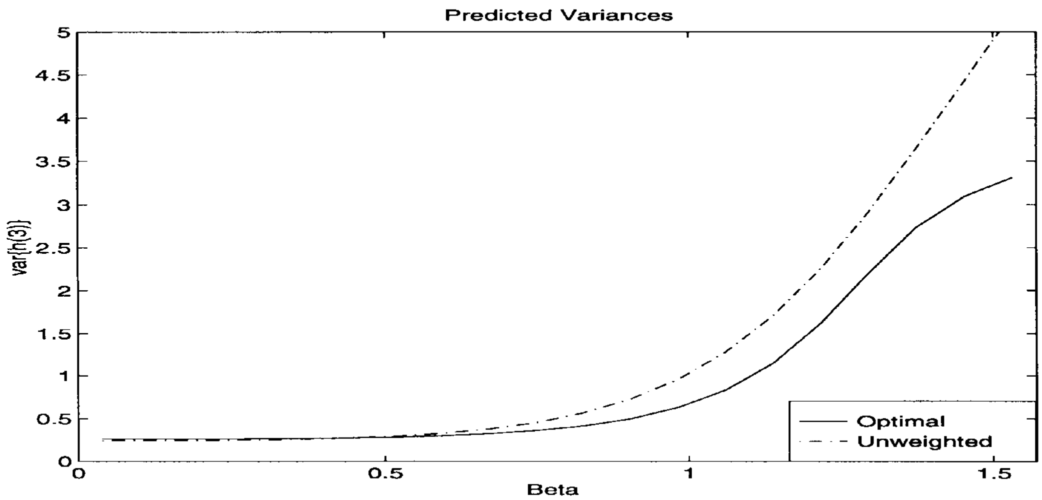


Fig. 8. Theoretical asymptotic variance: SNR = 10 dB.

TABLE VI
EXPERIMENT 4: VARIANCE OF CHANNEL ESTIMATES FOR $\beta = \pi/2$ (NONIDENTIFIABLE CHANNEL)

Parameter	Theoretical Variances		Estimated Variances				
	Weighted	Unweighted	Weighted	Unweighted	Linear [5]	XLTK [25]	Moulines [13]
$h(1)$	0.28	0.41	0.34	0.41	0.59	0.42	1,989.2
$h(2)$	2.35	3.08	2.82	3.32	8.01	1.99	1,540.40
$h(3)$	3.32	5.61	4.11	5.93	11.62	2.71	2,190.6

of an integer $f(s)$ as follows:

$$\sum_s f(s) = \sum_{n=0}^{P-1} \sum_\ell f(n - \ell P). \qquad (41)$$

If

$$f(s) \triangleq h(s)h(s+m) \exp\left\{-j\frac{2\pi}{P}ks\right\}$$

then

$$f(n-\ell P) = h(n-\ell P)h^*(n-\ell P+m)\exp\left\{-j\frac{2\pi}{P}kn\right\}$$

since

$$\exp\left\{-j\frac{2\pi}{P}k\ell P\right\} = 1 \quad \forall k, \ell.$$

Recognizing the first term in (40) as the right-hand side of (41), we have

$$C_{11y}(k;m) = \frac{\sigma_w^2}{P}\sum_s h(s)h^*(s+m)e^{-j(2\pi/P)ks} + c_{11v}(m)\delta(k). \qquad (42)$$

With $s = n$, (42) becomes (6).

APPENDIX B
ASYMPTOTIC COVARIANCE FOR THE OPTIMAL WEIGHT ESTIMATOR

If $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is full rank, we choose $\mathcal{Q} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$, and the matrix $\boldsymbol{G}(\boldsymbol{\theta}_0)$ in (21) becomes

$$\boldsymbol{G}(\boldsymbol{\theta}_0) = (\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1} \qquad (43)$$

TABLE VII
EXPERIMENT 4: MEAN BIAS OF CHANNEL ESTIMATES FOR $\beta = \pi/2$ (NON-IDENTIFIABLE CHANNEL)

Mean Bias				
Weighted	Unweighted	Linear [5]	XLTK [25]	Moulines [13]
0.05	0.05	0.54	1.16	0.68

where we have dropped the dependency on $\boldsymbol{\theta}_0$ for notational convenience. Substitution of (43) into (20) will give

$$\begin{aligned}\boldsymbol{P}_{\rm opt}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}_0) &= (\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}[(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}]' \\ &= (\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})(\boldsymbol{F}'\boldsymbol{\Sigma}^{-1}\boldsymbol{F})^{-1}.\end{aligned}$$

Canceling terms and using the fact that $\boldsymbol{P}_{\rm opt}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}_0)$ is Hermitian (i.e., $\boldsymbol{P}_{\rm opt}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}_0) = \boldsymbol{P}^{*'}_{\rm opt}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}_0)$) gives $\boldsymbol{B}(\boldsymbol{\theta}_0)$ in (18).

APPENDIX C
ASYMPTOTIC COVARIANCE OF $\hat{C}_{11y}(k;m)$

First, we state the following general result for estimation of cyclic correlations.

Result 1: For any cyclostationary time-series satisfying the mixing conditions of [2], the estimate of the cyclic correlation has asymptotic covariance defined in (29) and (30)

$$\begin{aligned}&S_{2y}(k,\ell;m_1,m_2) \\ &= \lim_{T\to\infty}\frac{1}{T}\sum_{n=0}^{T-1}\sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn}e^{-j(2\pi/P)\ell\xi} \\ &\quad\times [c_{22y}(n;\xi;m_1,\xi+m_2) + c_{20y}(n;\xi) \\ &\quad\times c_{02y}(n+\xi+m_2;m_1-m_2-\xi) \\ &\quad+ c_{11y}(n;\xi+m_2)c_{11y}(n+\xi;m_1-\xi)] \qquad (44)\end{aligned}$$

and

$$\begin{aligned}\overline{S}_{2y}(k,\ell;m_1,m_2) &= \lim_{T\to\infty}\frac{1}{T}\sum_{n=0}^{T-1}\sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn}e^{-j(2\pi/P)\ell\xi} \\ &\quad\times [c_{22y}(n;\xi+m_2;m_1,\xi)+c_{11y}(n;\xi) \\ &\quad\times c_{11y}(n+\xi+m_2;m_1-m_2-\xi) \\ &\quad+ c_{02y}(n;\xi+m_2)\times c_{20y}(n+\xi;m_1-\xi)]. \qquad (45)\end{aligned}$$

where $c_{k\ell y}(\cdot)$ denotes the $k+\ell$th-order cumulant as defined in (2).

Proof: Equations (44) and (45) follow from (29) and (30) by using the Leonov–Shiryaev theorem [1, p. 21] to break the cumulant of a product into sum of the product of cumulants. □

It is important to note that Result 1 is not limited to output sequences of FS systems. Rather, it applies to any cyclic correlation estimate, provided $y(n)$ satisfies the regularity conditions of [2]. Taking advantage of the FS structure will allow us to simplify (44) and (45) further. As in the second-order case, it is straightforward to show that $c_{k\ell y}(n;\cdot)$ is strictly periodic for FS systems, i.e., $c_{k\ell y}(n;\cdot)=c_{k\ell y}(n+sP;\cdot)$ for any integer s. This strict periodicity allows us to simplify (44) and (45) by recognizing "$\lim_{T\to\infty}(1/T)\sum_{n=0}^{T-1}\exp\{-j(2\pi/P)kn\}$" operation as a Fourier series expansion. Combined with the shifting and circular convolution properties of a Fourier series, we have the following result.

Result 2: The asymptotic covariance defined in (29) and (30) can be written as

$$\begin{aligned}S_{2y}(k,\ell;m_1,m_2) &= \sum_\xi e^{-j(2\pi/P)\ell\xi}C_{22y}(k;\xi;m_1,m_2+\xi) \\ &\quad+\sum_\xi e^{-j(2\pi/P)\ell\xi}\times\sum_{q=0}^{P-1}[C_{02y}(q;m_1-m_2-\xi) \\ &\quad\times C_{20y}(k-q;\xi)e^{j(2\pi/P)q(\xi+m_2)} \\ &\quad+C_{11y}(q;m_1-\xi)C_{11y}(k-q;\xi+m_2) \\ &\quad\times e^{j(2\pi/P)q\xi}] \qquad (46)\end{aligned}$$

and

$$\begin{aligned}\overline{S}_{2y}(k,\ell;m_1,m_2) &= \sum_\xi e^{-j(2\pi/P)\ell\xi}C_{22y}(k;\xi+m_2;m_1,\xi) \\ &\quad+\sum_\xi e^{-j(2\pi/P)\ell\xi}\times\sum_{q=0}^{P-1}[C_{11y}(q;m_1-m_2-\xi) \\ &\quad\times C_{11y}(k-q;\xi)e^{j(2\pi/P)q(\xi+m_2)} \\ &\quad+C_{20y}(q;m_1-\xi)C_{02y}(k-q;\xi+m_2) \\ &\quad\times e^{j(2\pi/P)q\xi}] \qquad (47)\end{aligned}$$

where the $(k+\ell)$th-order cyclic cumulant $C_{k\ell y}(k;m_1,\cdots,m_{k-1};m'_1,\cdots,m'_\ell)$ is defined in Proposition 1.

Proof: Initially, consider only the second $[T2]$ and third terms $[T3]$ of (44)

$$\begin{aligned}[T2]+[T3] &= \lim_{T\to\infty}\frac{1}{T}\sum_{n=0}^{T-1}\sum_{\xi=-\infty}^{\infty}e^{-j(2\pi/P)kn}e^{-j(2\pi/P)\ell\xi} \\ &\quad\times[c_{20y}(n;\xi)c_{02y}(n+\xi+m_2;m_1-m_2-\xi) \\ &\quad+c_{11y}(n;\xi+m_2)c_{11y}(n+\xi;m_1-\xi)] \qquad (48)\\ &= \frac{1}{P}\sum_{\xi=-\infty}^{\infty}e^{-j(2\pi/P)\ell\xi}\sum_{n=0}^{P-1}e^{-j(2\pi/P)kn} \\ &\quad\times[c_{20y}(n;\xi)c_{02y}(n+\xi+m_2;m_1-m_2-\xi) \\ &\quad+c_{11y}(n;\xi+m_2)c_{11y}(n+\xi;m_1-\xi)] \qquad (49)\end{aligned}$$

where we have used the periodicity to simplify the limit, i.e.,

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\sum_{n=0}^{T-1}g(n) &= \lim_{Q\to\infty}\frac{1}{QP}\sum_{\ell=0}^{Q-1}\sum_{n=P\ell}^{P(\ell+1)}g(n) \\ &= \frac{1}{P}\sum_{n=0}^{P-1}g(n)\end{aligned}$$

if $g(n)=g(n+sP)$ for any integer s and where $Q\triangleq[T/P]$. Now, if we define

$$\begin{aligned}f_1(n;\xi,m_1,m_2) &\triangleq c_{20y}(n;\xi) \\ &\qquad\times c_{02y}(n+\xi+m_2;m_1-m_2-\xi) \\ f_2(n;\xi,m_1,m_2) &\triangleq c_{11y}(n;\xi+m_2) \\ &\qquad\times c_{11y}(n+\xi;m_1-\xi)\end{aligned}$$

(49) becomes

$$[T2]+[T3]=\frac{1}{P}\sum_{\xi=-\infty}^{\infty}e^{-j(2\pi/P)\ell\xi}\left\{\sum_{n=0}^{P-1}[f_1(n;\xi,m_1,m_2)+f_2(n;\xi,m_1,m_2)]e^{-j(2\pi/P)kn}\right\}.$$

Recognizing the sum over n as a discrete Fourier series, we have

$$\begin{aligned}[T2]+[T3] &= \sum_{\xi=-\infty}^{\infty}e^{-j(2\pi/P)\ell\xi}[F_1(k;\xi,m_1,m_2) \\ &\quad+F_2(k;\xi,m_1,m_2)] \qquad (50)\end{aligned}$$

where

$$\begin{aligned}F_1(k;\xi,m_1,m_2) &\triangleq \frac{1}{P}\sum_{n=0}^{P-1}f_1(n;\xi,m_1,m_2)e^{-j(2\pi/P)kn} \\ F_2(k;\xi,m_1,m_2) &\triangleq \frac{1}{P}\sum_{n=0}^{P-1}f_2(n;\xi,m_1,m_2)e^{-j(2\pi/P)kn}.\end{aligned}$$

Now, we need to express $F_1(k;\xi,m_1,m_2)$ and $F_2(k;\xi,m_1,m_2)$ in terms of the cyclic correlations. The following discrete Fourier Series relations are used:

$$g(n) \overset{\text{DFS}}{\longleftrightarrow} G(k) g_1(n)g_2(n) \overset{\text{DFS}}{\longleftrightarrow} \sum_{\ell=0}^{P-1} G_1(\ell)G_2(k-\ell) \quad (51)$$

$$g(n+m) \overset{\text{DFS}}{\longleftrightarrow} G(k)e^{j(2\pi/P)km}. \quad (52)$$

Using these relations, we have

$$\begin{aligned} &F_1(k;\xi,m_1,m_2)\\ &\quad= \sum_{q=0}^{P-1} C_{20y}(q;\xi) \times C_{02y}(k-q;m_1-m_2-\xi)\\ &\quad\quad \times e^{j(2\pi/P)(k-q)(\xi+m_2)}\\ &\quad= \sum_{q=0}^{P-1} C_{20y}(k-q;\xi) \times C_{02y}(q;m_1-m_2-\xi)\\ &\quad\quad \cdot e^{j(2\pi/P)q(\xi+m_2)} \end{aligned}$$

and

$$\begin{aligned} F_2(k;\xi,m_1,m_2) &= \sum_{q=0}^{P-1} C_{11y}(q;\xi+m_2)\\ &\quad \times C_{11y}(k-q;m_1-\xi)e^{j(2\pi/P)(k-q)\xi}\\ &= \sum_{q=0}^{P-1} C_{11y}(q;m_1-\xi)\\ &\quad \times C_{11y}(k-q;m_2+\xi)e^{j(2\pi/P)q\xi}. \end{aligned}$$

Therefore

$$\begin{aligned} &[T2]+[T3]\\ &\quad= \sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)\ell\xi} \sum_{q=0}^{P-1} [C_{02y}(q;m_1-m_2-\xi)\\ &\quad\quad \times C_{20y}(k-q;\xi)e^{j(2\pi/P)q(\xi+m_2)} + C_{11y}(q;m_1-\xi)\\ &\quad\quad \cdot C_{11y}(k-q;\xi+m_2) \times e^{j(2\pi/P)q\xi}]. \end{aligned} \quad (53)$$

In addition, by definition

$$\begin{aligned} &C_{22y}(k;\xi,m_1,m_2+\xi)\\ &\quad \triangleq \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{\infty} e^{-j(2\pi/P)kn} \times c_{22y}(n;\xi,m_1,m_2+\xi). \end{aligned} \quad (54)$$

Hence, combining (53) and (54), we have (46). Equation (47) can be proven similarly. □

The structure of the FS cyclic correlation allows the fourth-order cyclic cumulant $C_{22y}(k;m_1;m_2,m_3)$ in both (46) and (47) to be written as the product of two second-order cyclic cumulants of the signal and a fourth-order cumulant of the noise. To show (33), we focus on the first term in (44), which involves the fourth-order cumulant $c_{22y}(n;\xi;m_1,\xi+m_2)$

$$\begin{aligned} [T1] &= \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T-1} \sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn} e^{-j(2\pi/P)\xi\ell}\\ &\quad \times c_{22y}(n;\xi,m_1,\xi+m_2)\\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T-1} \sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn} e^{-j(2\pi/P)\xi\ell}\\ &\quad \times \text{cum}\,\{y(n), y(n+\xi), y^*(n+m_1)\\ &\quad \cdot y^*(n+\xi+m_2)\}. \end{aligned}$$

Since we assume the noise is independent of the input, (1) and the independence property of cumulants [1, ch. 2] gives

$$\begin{aligned} [T1] &= \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T-1} \sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn} e^{-j(2\pi/P)\xi\ell}\\ &\quad \times [c_{22x}(n;\xi,m_1,\xi+m_2) + c_{22v}(n;\xi;m_1,\xi+m_2)]. \end{aligned}$$

Because $c_{22x}(\cdot)$ is strictly periodic, the limiting sum is replaced by a sum over a period, and since $v(n)$ is white and stationary, a delta appears in the noise term

$$\begin{aligned} [T1] &= \frac{1}{P} \sum_{n=0}^{P-1} \sum_{\xi=-\infty}^{\infty} e^{-j(2\pi/P)kn} e^{-j(2\pi/P)\ell\xi}\\ &\quad \cdot c_{22x}(n;\xi,m_1,\xi+m_2)\\ &\quad + \delta(k)\delta(m_1)\delta(m_2)c_{22v}(0;0,0). \end{aligned} \quad (55)$$

We can simplify the $c_{22x}(n;m_1,\xi,\xi+m_2)$ term by using the multilinearity of cumulants

$$\begin{aligned} &\frac{1}{P} \sum_{n=0}^{P-1} c_{22x}(n;\xi,m_1,\xi+m_2)e^{-j(2\pi/P)kn}\\ &\quad= \frac{c_{22w}(0,0,0)}{P} \sum_{n} \sum_{\ell} h(n-\ell P)h(n+\xi-\ell P)\\ &\quad\quad \times h^*(n+m_1-\ell P)h^*(n+\xi+m_2-\ell P)\\ &\quad\quad \cdot e^{-j(2\pi/P)kn}. \end{aligned}$$

Similar to (41), the double sum on n and ℓ can be reduced to a single sum. Substitution in (55) yields

$$\begin{aligned} [T1] &= \frac{c_{22w}(0,0,0)}{P} \sum_{n} \sum_{\xi} h(n)h(n+\xi)h^*(n+m_1)\\ &\quad \times h^*(n+\xi+m_2)e^{-j(2\pi/P)kn} e^{-j(2\pi/P)\ell\xi}\\ &\quad + \delta(k)\delta(m_1)\delta(m_2)c_{22v}(0;0,0). \end{aligned}$$

We are concerned mainly with the noise-free term. With this in mind, we can rewrite the above as

$$\begin{aligned} [T1] &= \frac{c_{22w}(0,0,0)}{P} \sum_{n} h(n)h^*(n+m_1)e^{-j(2\pi/P)kn}\\ &\quad \times \sum_{\xi=-\infty}^{\infty} h(n+\xi)h^*(n+\xi+m_2)e^{-j(2\pi/P)\ell\xi}\\ &\quad + [\text{noise term}]. \end{aligned}$$

Using (8), this can be written as

$$\begin{aligned}[T1] &= \frac{c_{22w}(0,0,0)}{P} \sum_n h(n)h^*(n+m_1)e^{-j(2\pi/P)kn} \\ &\quad \times \frac{P}{c_{11w}(0)} C_{11x}(\ell; m_2)e^{j(2\pi/P)\ell n} + [\text{noise term}] \\ &= \frac{c_{22w}(0,0,0)}{c_{11w}(0)} C_{11x}(\ell; m_2) \sum_n h(n)h^*(n+m_1) \\ &\quad \times e^{-j(2\pi/P)n(k-\ell)} + [\text{noise term}] \\ &= \frac{c_{22w}(0,0,0)}{[c_{11w}(0)]^2} P C_{11x}(\ell; m_2) C_{11x}(k-\ell; m_1) \\ &\quad + [\text{noise term}].\end{aligned}$$

Combining the above with (55) gives (33). Equation (34) can be proven similarly.

REFERENCES

[1] D. R. Brillinger, *Time Series Data Analysis and Theory*. San Francisco, CA: Holden-Day, 1981.

[2] A. V. Dandawaté and G. B. Giannakis, "Asymptotic theory of mixed time averages and kth-order cyclic-moment and cumulant statistics," *IEEE Trans. Inform. Theory*, vol. 41, pp. 216–239, Jan. 1995.

[3] Z. Ding and Y. Li, "On channel identification based on second-order cyclic spectra," *IEEE Trans. Signal Processing*, pp. 1260–1264, May 1994.

[4] T. J. Endres, S. D. Halford, C. R. Johnson, Jr., and G. B. Giannakis, "Blind adaptive channel equalization using fractionally spaced receivers: A comparison study," in *Proc. 30th Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1996, pp. 1172–1177.

[5] G. B. Giannakis, "Linear cyclic correlation approaches for blind identification of FIR channels," in *Proc. 28th Ann. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Oct. 31–Nov. 2, 1994, pp. 420–424.

[6] G. B. Giannakis and S. D. Halford, "Performance analysis of blind equalizers based on cyclostationary statistics," in *Proc. 28th Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1994, pp. 873–876.

[7] G. B. Giannakis and J. M. Mendel, "Identification of non-minimum phase systems via higher order-statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 360–377, Mar. 1989.

[8] G. B. Giannakis and M. K. Tsatsanis, "Restoring identifiability of fractionally sampled blind channel estimators using higher-order statistics," in *Proc. Int. Conf. Higher Order Statist.*, Barcelona, Spain, June 12–14, 1995, pp. 409–413.

[9] S. Halford and G. B. Giannakis, "Asymptotically optimal blind equalizers based on cyclostationary statistics," in *Proc. IEEE Military Commun. Conf.*, Fort Monmouth, NJ, Oct. 2–5, 1994, pp. 306–310.

[10] ——, "Channel order determination based on cyclic correlations," in *Proc. 29th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 31–Nov. 2, 1994, pp. 425–429.

[11] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.

[12] Y. Li and Z. Ding, "New results on the blind identification of FIR channels based on second-order statistics," in *Proc. IEEE Military Commun. Conf.*, Boston, MA, Oct. 1993, pp. 644–647.

[13] 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.

[14] M. Kristensson and B. Ottersten, "Statistical analysis of a subspace method for blind channel identification," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Atlanta, GA, vol. V, 1996, pp. 243-5-2438.

[15] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[16] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *Int. J. Adap. Contr. Signal Processing*, vol. 3, pp. 191–229, 1989.

[17] ——, "Blind equalization of digital communication channels using higher-order moments," *IEEE Trans. Signal Processing*, vol. 39, pp. 522–526, Feb. 1991.

[18] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1989.

[19] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Adelaide, Australia, vol. IV, 1994, pp. 585–588.

[20] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[21] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: A frequency-domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.

[22] L. Tong and H. Zeng, "Blind channel identification using cyclic spectra," in *Proc. 28th Ann. Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1994, pp. 711–716.

[23] J. Tugnait, "Identification of linear stochastic systems via second- and fourth-order cumulant matching," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 393–407, May 1987.

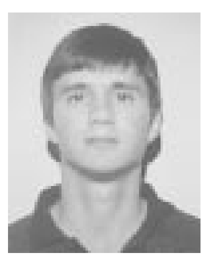
[24] ——, "On blind identifiability of multipath channels using fractional sampling and second-order cyclostationary statistics," *IEEE Trans. Inform. Theory*, vol. 41, pp. 308–311, Jan. 1995.

[25] 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.

[26] H. Zeng and L. Tong, "Some new results on blind channel estimation: Performance and algorithms," in *Proc. 29th Ann. Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., Baltimore, MD, Mar. 1995, pp. 695–700.

[27] H. Zeng and L. Tong, "Blind channel estimation using the second-order statistics: Asymptotic performance and limitations," to be published.

Georgios B. Giannakis (F'97), for photograph and biography, see this issue, p. 1794.



Steven D. Halford (S'86) received the B.S. and M.S. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1988 and 1990, respectively. In January 1997, he received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, where he was a Graduate Research Assistant in the Communications Systems Laboratory.

From 1988 until 1990, he was employed by Southern Bell Telephone Company, Atlanta, where he worked on narrowband ISDN. From 1993 to 1995, he was with the Naval Research Laboratories, Washington, DC, working on ATM for wireless communications. He is now with the Government Communications Systems Division, Harris Corporation, Melbourne, FL. His current research interests are in the general areas of channel equalization, signal extraction, and signal processing to improve wireless communications.