The Annals of Statistics
1995, Vol. 23, No. 2, 690–704

BLIND DECONVOLUTION OF LINEAR SYSTEMS WITH MULTILEVEL NONSTATIONARY INPUTS

BY TA-HSIN LI[1]

Texas A & M University

A method is proposed to deal with the problem of blind deconvolution of a special non-Gaussian linear process, in which the input to the linear system is a real- or complex-valued multilevel random sequence that satisfies certain regularity conditions. The gist of the method is to apply a linear filter to the observed process and adjust the filter until a multilevel output is obtained. It is shown that the deconvolution problem can be solved (with only scale/rotation and shift ambiguities) if the output sequence of the filter contains a subsequence that converges weakly to a multilevel random variable. A cost function is proposed so that any minimizer of the cost function provides a solution to the deconvolution problem. Moreover, when the process is parametric, it is shown that a consistent estimator of the parameter can be obtained by minimizing an empirical criterion. The estimation accuracy is shown to depend on the tail behavior of the inverse system, which in many cases decays exponentially as the sample size grows. Special consideration is given to applications in the equalization of digital communications systems.

1. Introduction. Let us consider the linear process $\{Y_k\}$ as defined by

$$Y_k = \sum_{j=-\infty}^{\infty} s_j X_{k-j}. \tag{1}$$

In this expression, $\mathscr{S} := \{s_k\}$ is an unknown deterministic sequence or a linear time-invariant system and $\{X_k\}$ is the system's input consisting of unobservable independent random variables. The blind deconvolution problem in general is to simultaneously estimate the system $\mathscr{S}$ and its input $\{X_k\}$ from the observed process $\{Y_k\}$ alone. Throughout the paper, we assume that $\mathscr{S}$ satisfies the following conditions:

C1. The inverse of $\mathscr{S}$, denoted by $\mathscr{S}^{-1} := \{s_k^{-1}\}$, exists so that $\sum s_j s_{k-j}^{-1} = \delta_k$, where $\delta_k = 0$ for $k \neq 0$ and $\delta_0 = 1$.

C2. Both $\mathscr{S}$ and $\mathscr{S}^{-1}$ are stable, that is, $\{s_k\}, \{s_k^{-1}\} \in l_1$, where l_1 is the set of all absolutely summable sequences.

Several methods exist in the literature for blind deconvolution. In particular, the classical technique of linear prediction can be employed if $\mathscr{S}$ is minimum phase and $\{X_k\}$ is stationary [e.g., Brockwell and Davis (1991)]. For

Received January 1992; revised August 1994.

[1]The author was with the Department of Mathematics and the Institute for Systems Research, University of Maryland, College Park, MD.

AMS 1991 *subject classifications.* Primary 62M10; secondary 60G35, 93E12.

Key words and phrases. Deconvolution, equalization, linear filter, linear process, multilevel random sequence, non-Gaussian, system identification.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

nonminimum phase systems, some procedures are also available that utilize higher-order moments (cumulants, polyspectra, etc.) of the observed data to estimate the phase information that cannot be recovered from the second-order moments [e.g., Benveniste, Goursat and Ruget (1980), Donoho (1981), Lii and Rosenblatt (1982), Giannakis and Mendel (1989), Cheng (1990), Shalvi and Weinstein (1990)]. These procedures usually assume that the X_k are i.i.d. or higher-order stationary, and many of them require the moments of $\{X_k\}$ to be available.

Although the blind deconvolution problem can be encountered in a variety of engineering and scientific applications, the present paper focuses on a special problem of blind deconvolution in which $\{X_k\}$ is known a priori to be a *multilevel* random sequence that takes on values in a common finite alphabetical set. This problem stems from an application in digital communications [e.g., Feher (1987)], where $\{X_k\}$ stands for the transmitted digital signal, $\mathscr{S}$ for the linear system that characterizes the distortion introduced by the communication channel (e.g., telephone lines) and $\{Y_k\}$ for the signal observed at the receiver. The problem is also known as blind equalization of communication channels, especially for on-line implementation [e.g., Shalvi and Weinstein (1990)].

For the simplest case of binary inputs and real systems, a blind deconvolution method was recently proposed by the author [Li (1992)]. This method is able to handle nonminimum phase systems and nonstationary inputs, without utilizing any distributional information of $\{X_k\}$ other than binariness and independence. Furthermore, if $\mathscr{S}$ is a parametric system that can be characterized by a finite dimensional parameter, the proposed method leads to a consistent estimator for the parameter on the basis of a finite number of observations, and when $\mathscr{S}$ is an AR system in particular, it was shown [Li (1993a)] that the true parameter values can be obtained *without error*, at least with probability tending to unity, as the data length increases. All these results are generalized and extended in this paper to real- and complex-valued multilevel random sequences as well as to real- and complex-valued linear time-invariant systems.

In addition, the present paper also investigates the issue of estimation accuracy for parametric systems under some more general conditions. It is shown in particular that the estimation error for a parametric system depends on the *tail behavior* of the true inverse system which in many cases decays *exponentially* as the data length increases. This confirms, from a different perspective, the recent results of Davis and Rosenblatt (1991), claiming that a "super-efficiency"—a convergence rate higher than the usual reciprocal of square root of the sample size—can be achieved for system estimation if the distribution of the input sequence possesses singularities.

Following the same ideas as in the binary case [Li (1992)], we apply a linear filter $\mathscr{H} := \{h_k\}$ to the observed process $\{Y_k\}$ and obtain the output

$$Z_k := \sum_{j=-\infty}^{\infty} h_j Y_{k-j}. \tag{2}$$

Let $\{t_k\}$ be the impulse response of the cascaded system $\mathscr{T} := \mathscr{H} * \mathscr{S}$, namely,

$$t_k := \sum_{j=-\infty}^{\infty} h_j s_{k-j}.$$

Then the output sequence $\{Z_k\}$ can be written as a convolution of $\{X_k\}$ with $\{t_k\}$, that is,

$$Z_k = \sum_{j=-\infty}^{\infty} t_j X_{k-j}. \tag{3}$$

Thus the deconvolution problem can be solved by seeking a filter $\mathscr{H} \in l_1$ so that $t_k = r\delta_{k-K}$ for some $r \neq 0$ and K. If this can be done, one would obtain

$$h_k = r s_{k-K}^{-1} \quad \text{and} \quad Z_k = r X_{k-K} \quad \text{for all } k, \tag{4}$$

which simply means that $\mathscr{H}$ is identical to $\mathscr{S}^{-1}$ and $\{Z_k\}$ to $\{X_k\}$, except for a possible multiplier r and a possible shift K. Therefore, the objective of blind deconvolution is to find such a filter on the basis of $\{Y_k\}$ and partial statistical information of $\{X_k\}$.

2. General results. In this section, a general situation is considered where the knowledge about $\{X_k\}$ is limited to the cardinality of its common alphabet. The main results are stated separately for real and complex systems. In both cases, stationarity is *not* a requirement, so that the marginal distribution of X_k may *depend* on k and could even be *unknown*. This relaxed requirement on $\{X_k\}$ is one of the features that distinguish the present method from many others, and its impact on deconvolution results can be found in Li (1993b).

2.1. *Real systems.* Consider the case where both $\{X_k\}$ and $\{s_k\}$ are real-valued. Suppose that $\{X_k\}$ is a multilevel sequence, taking on discrete values in a finite real alphabet $\mathbf{A} := \{a(u),\ u \in \mathbf{S}_A\}$, where the $a(u)$ are distinct real numbers and $\mathbf{S}_A$ is a finite set of integers. The *cardinality* of $\mathbf{A}$, denoted by $|\mathbf{A}|$, is defined as the number of elements in $\mathbf{A}$. It is always assumed in the sequel that $0 \in \mathbf{S}_A$ and $|\mathbf{A}| \geq 2$. It is further assumed that $\mathbf{A}$ satisfies the following *regularity condition* that can be easily satisfied by many signal constellations in digital communications systems [e.g., Korn (1985)].

DEFINITION 1. A real alphabet $\mathbf{A} = \{a(u),\ u \in \mathbf{S}_A\}$ is said to be *regular* if there exists a real number $\gamma_A > 0$ such that

$$a(u) = a(0) + \gamma_A u \qquad \forall\, u \in \mathbf{S}_A, \tag{5}$$

namely, if $\mathbf{A}$ is a finite subset of the lattice $a(0) + \gamma_A \mathbb{Z}$, where $\mathbb{Z}$ is the set of all integers.

For any real alphabet $\mathbf{C} := \{c(u),\ u \in \mathbf{S}_C\}$, let $M(\mathbf{C}; \mathbf{p})$ denote the probability distribution of a real-valued multilevel random variable X for which $\mathbf{p} := \{p(u),\ u \in \mathbf{S}_C\}$ and $p(u) := \mathrm{pr}\{X = c(u)\}$. As a convention, once the notation $M(\mathbf{C}; \mathbf{p})$ is employed, it always implies that $p(u) > 0$ for *all* $u \in \mathbf{S}_C$

(symbolically denoted by $\mathbf{p} > 0$) and $\Sigma p(u) = 1$. The shorthand notation $\mathbf{p} \ge \varepsilon$ is also used to represent $p(u) \ge \varepsilon$, $\forall\, u \in \mathbf{S}_C$. Note that if $\mathbf{C}$ is regular, $M(\mathbf{C};\mathbf{p})$ is known as the *lattice distribution* [Feller (1971)].

With this notation, one can write $X_k \sim M(\mathbf{A};\mathbf{p}_k)$ for some $\mathbf{p}_k := \{p_k(u), u \in \mathbf{S}_A\}$, where the $\mathbf{p}_k$ are assumed to be *unknown functions* of k. To further generalize the situation, the alphabet $\mathbf{A}$ is assumed unknown except for its regularity and its cardinality $|\mathbf{A}|$. Under these mild conditions, the following results can be obtained.

THEOREM 1. *Let the real-valued random sequence $\{X_k\}$ be independently distributed according to $X_k \sim M(\mathbf{A};\mathbf{p}_k)$ for all k, where $\mathbf{A}$ is regular and $\mathbf{p}_k \ge \varepsilon$ for some $\varepsilon > 0$ and for all k. If a real-valued filter $\mathscr{H} \in l_1$ with $\mathscr{H} \ne 0$ can be found such that a subsequence of $\{Z_k\}$ converges in distribution to $M(\mathbf{B};\mathbf{q})$ for some regular $\mathbf{B}$ with $|\mathbf{B}| \le |\mathbf{A}|$, then there exist an integer K and a real number $r \ne 0$ such that $t_k = r\delta_{k-K}$.*

PROOF. Since $\mathbf{B}$ is regular, one can write $\mathbf{B} = \{b(v), v \in \mathbf{S}_B\}$ with $b(v) = b(0) + \gamma_B v$ for all $v \in \mathbf{S}_B$. Let $\psi_k(\lambda)$ and $\phi_k(\lambda)$ be the characteristic functions of X_k and Z_k, respectively, that is, $\psi_k(\lambda) := E\{\exp(iX_k\lambda)\}$ and $\phi_k(\lambda) := E\{\exp(iZ_k\lambda)\}$. Then the assumption that $X_k \sim M(\mathbf{A};\mathbf{p}_k)$ leads to $\psi_k(\lambda) = \Sigma p_k(u)\exp\{ia(u)\lambda\}$. For notational simplicity, let the convergent subsequence in the theorem be denoted by $\{Z_k\}$ itself. Then the continuity theorem of characteristic functions guarantees that

$$\lim_{k\to\infty} \phi_k(\lambda) = \phi(\lambda) := \sum q(v)\exp\{ib(v)\lambda\} \tag{6}$$

for any λ, where $\mathbf{q} = \{q(v), v \in \mathbf{S}_B\}$. On the other hand, since $\mathscr{H} \in l_1$ and thus $\mathscr{T} = \{t_k\} \in l_1$, the infinite sum in (3) converges in mean square. Using the continuity theorem along with the independence of $\{X_k\}$, one further obtains $\phi_k(\lambda) = \prod \psi_{k-j}(t_j\lambda)$ for all k and λ, so that (6) becomes

$$\lim_{k\to\infty} \prod \psi_{k-j}(t_j\lambda) = \phi(\lambda). \tag{7}$$

Now, it is crucial to note that $|\phi(\lambda_0)| = 1$ with $\lambda_0 := 2\pi/\gamma_B$, as ensured by the regularity of $\mathbf{B}$. This, combined with (7), implies that

$$\lim_{k\to\infty} \prod |\psi_{k-j}(t_j\lambda_0)| = 1. \tag{8}$$

Since $|\psi_{k-j}(\lambda)| \le 1$ for all k, j and λ, then, for any fixed j, the product on the left-hand side of (8) can be bounded above by $w_k := |\psi_{k-j}(t_j\lambda_0)|$. Moreover, for any fixed j, the bounded sequence $\{\mathbf{p}_{k-j}\}$ contains convergent subsequences. Therefore, with j being fixed, one can assume for simplicity that $\{\mathbf{p}_{k-j}\}$ itself is convergent and $\mathbf{p}_{k-j} \to \mathbf{p} := \{p(u), u \in \mathbf{S}_A\}$. Since $\mathbf{p}_k \ge \varepsilon > 0$ for all k, it follows that $p(u) > 0$ for all $u \in \mathbf{S}_A$ and $\Sigma p(u) = 1$. This, coupled with (8), implies that

$$\lim_{k\to\infty} w_k = |\psi(\lambda_0)| \ge 1, \tag{9}$$

where $\psi(\lambda)$ is the characteristic function of $M(\mathbf{C};\mathbf{p})$ with $c(u) := a(u)t_j$. Since

$|\psi(\lambda)| \le 1$ for any λ, it follows from (9) that $|\psi(\lambda_0)| = 1$. Invoking Lemma 2 with $u_0 = 0$ proves that there exist integers $n_j(u)$ such that $n_j(u) = (\gamma_A/\gamma_B)ut_j$ for all $u \in \mathbf{S}_A$ and all j, implying that $t_j \equiv (\gamma_B/\gamma_A)n_j(u)/u$ for *all* $u \in \mathbf{S}_A$ with $u \ne 0$. Therefore, by fixing $u_1 \in \mathbf{S}_A$ with $u_1 \ne 0$, one obtains $t_j = \beta n_j(u_1)$ for all j, where $\beta := (\gamma_B/\gamma_A)/u_1$. Since $\mathscr{T} \in l_1$, the *integer* sequence $\{n_j(u_1)\}$ contains at most a finite number of nonzero elements, and so does the sequence $\{t_j\}$.

To proceed with the proof, let $\mathbf{J}$ be the collection of j for which $t_j \ne 0$. Then $\mathbf{J}$ is a finite set of integers, and thus on the left-hand side of (7) is actually a finite product for $j \in \mathbf{J}$. Since for each $j \in \mathbf{J}$ the bounded sequence $\{\mathbf{p}_{k-j}\}$ has convergent subsequences and $\mathbf{J}$ is a finite set, there must be a single subsequence of k that converges for *all* $j \in \mathbf{J}$. Assume, without loss of generality, that $\mathbf{p}_{k-j} \to \mathbf{q}_j := \{q_j(u),\ u \in \mathbf{S}_A\}$ for all $j \in \mathbf{J}$. Then it follows from (7) that

$$\sum q(\mathbf{w})\exp\{ia(\mathbf{w})\lambda\} = \phi(\lambda), \tag{10}$$

where $\mathbf{w} := \{u_j,\ j \in \mathbf{J}\}$ with $u_j \in \mathbf{S}_A$, $a(\mathbf{w}) := \Sigma a(u_j)t_j$ and $q(\mathbf{w}) := \Pi q_j(u_j) > 0$. Since $\Sigma q(\mathbf{w}) = 1$, the function on the left-hand side of (10) is the characteristic function of a discrete random variable taking on values in the alphabet $\{a(\mathbf{w})\}$, whereas the function $\phi(\lambda)$, on the right-hand side, is the characteristic function of a $\mathbf{B}$-valued random variable. According to the uniqueness theorem of characteristic functions, these two random variables must have the same distribution, and hence $\{a(\mathbf{w})\} = \mathbf{B}$. Furthermore, for any given $k, k' \in \mathbf{J}$ with $k \ne k'$, consider a subset of $\{a(\mathbf{w})\}$ corresponding to a special choice of $\mathbf{w}$ for which $u_k = u$, $u_{k'} = u'$ and $u_j \equiv 0$ for $j \ne k, k'$. In this case, $a(\mathbf{w})$ can be written as $a(u)t_k + a(u')t_{k'} + c$, where $c := a(0)\Sigma_{j \ne k, k'} t_j$, and hence $\{a(u)t_k + a(u')t_{k'} + c\colon u, u' \in \mathbf{S}_A\} \subseteq \mathbf{B}$. By Lemma 1, t_k and $t_{k'}$ cannot both be nonzero. This, together with the fact that $\mathscr{T} \ne 0$, ensures the existence of an integer K such that $t_k = 0$ for all $k \ne K$ and $r := t_K \ne 0$. The proof is thus complete. □

REMARK 1. Since the $\mathbf{p}_k$ may depend on k and hence the X_k may have *different* marginal distributions, Theorem 1 only requires that the X_k be independent and have a common regular alphabet. In this sense, the theorem holds for *nonstationary* multilevel random sequences. Examples can be found in Li (1993b) that demonstrate the advantages of this method over some existing procedures in dealing with nonstationary inputs.

REMARK 2. In the special case where some k can be found such that $Z_k \sim M(\mathbf{B}; \mathbf{q})$, Theorem 1 remains valid without assuming $\mathbf{p}_k \ge \varepsilon > 0$ for all k, since, in this case, it is no longer necessary to consider convergent subsequences of $\{\mathbf{p}_{k-j}\}$ in the proof. The theorem thus reveals a characterization property of multilevel random variables: *If a nontrivial linear combination of infinitely many independent multilevel random variables with a common regular alphabet* $\mathbf{A}$ *is also regular with an alphabet* $\mathbf{B}$ *that satisfies*

$|\mathbf{B}| \le |\mathbf{A}|$, *then there must be one and only one nonzero coefficient in the combination.*

2.2. *Complex systems.* Consider the case of complex-valued input $X_k := X_k^R + iX_k^I$ and assume that $\mathscr{S} = \{s_k\}$ and $\mathscr{H} = \{h_k\}$ are also complex-valued so that t_k can be written as $t_k = t_k^R + it_k^I$. In this case, one obtains $Z_k = Z_k^R + iZ_k^I$ with

$$(11)\quad Z_k^R := \sum_{j=-\infty}^{\infty} \left(t_j^R X_{k-j}^R - t_j^I X_{k-j}^I\right) \quad \text{and} \quad Z_k^I := \sum_{j=-\infty}^{\infty} \left(t_j^I X_{k-j}^R + t_j^R X_{k-j}^I\right).$$

Let $\mathbf{A} := \{a(u,v),\ (u,v) \in \mathbf{S}_A\}$ be the complex alphabet of X_k, where the $a(u,v)$ are distinct complex numbers and $\mathbf{S}_A$ is a finite set of integer pairs (u,v) that includes the pair $(u,v) = (0,0)$. As in the case of real systems, it is assumed that $|\mathbf{A}| \ge 2$ and that $\mathbf{A}$ satisfies the following regularity condition. [Examples of regular complex alphabet in digital communications can be found in, e.g., Korn (1985).]

DEFINITION 2. A complex alphabet $\mathbf{A}$ is said to be *regular* if there exist real numbers $\gamma_A^R > 0$ and $\gamma_A^I > 0$ such that

$$(12)\qquad a(u,v) - a(0,0) = \gamma_A^R u + i\gamma_A^I v \qquad \forall\,(u,v) \in \mathbf{S}_A,$$

namely, if $\mathbf{A}$ is a finite subset of the two-dimensional lattice $(\alpha_R + \beta_R\mathbb{Z}) + i(\alpha_I + \beta_I\mathbb{Z})$ with $\alpha_R + i\alpha_I = a(0,0)$, $\beta_R = \gamma_A^R$ and $\beta_I = \gamma_A^I$.

The distribution of a multilevel random variable $X = X_R + iX_I$ with alphabet $\mathbf{C} := \{c(u,v),\ (u,v) \in \mathbf{S}_C\}$ is denoted by $M(\mathbf{C};\mathbf{p})$, where $\mathbf{p} := \{p(u,v),\ (u,v) \in \mathbf{S}_C\}$, $p(u,v) := \text{pr}\{X_R = c_R(u,v),\ X_I = c_I(u,v)\}$ and $c(u,v) = c_R(u,v) + ic_I(u,v)$. Assume that $p(u,v) > 0$, $\forall\ (u,v) \in \mathbf{S}_C$ and $\sum p(u,v) = 1$. Equipped with these assumptions, the following theorem extends the results in Theorem 1 to the complex case.

THEOREM 2. *Let the complex-valued random sequence $\{X_k\}$ be independently distributed with $X_k \sim M(\mathbf{A};\mathbf{p}_k)$ for all k, where $\mathbf{A}$ is regular and $\mathbf{p}_k \ge \varepsilon > 0$ for all k. If there exists a complex-valued filter $\mathscr{H} \in l_1$ with $\mathscr{H} \ne 0$ such that a subsequence of $\{Z_k\}$ converges in distribution to $M(\mathbf{B};\mathbf{q})$, where $\mathbf{B}$ is regular with $|\mathbf{B}| \le |\mathbf{A}|$, then* (4) *holds for an integer K and a complex number $r \ne 0$.*

PROOF. As in the proof of Theorem 1, it is convenient to assume that $\{Z_k\}$ itself converges weakly to $M(\mathbf{B};\mathbf{q})$, where $\mathbf{B} = \{b(\mu,\nu),\ (\mu,\nu) \in \mathbf{S}_B\}$ and $\mathbf{q} = \{q(\mu,\nu),\ (\mu,\nu) \in \mathbf{S}_B\}$. Due to (11) and the independence of $\{(X_k^R, X_k^I)\}$, the characteristic function $\phi_k(\lambda_R, \lambda_I)$ of Z_k can be written as

$$(13)\qquad \phi_k(\lambda_R, \lambda_I) = \prod \psi_{k-j}\left(t_j^R\lambda_R + t_j^I\lambda_I,\ t_j^R\lambda_I - t_j^I\lambda_R\right),$$

where $\psi_k(\lambda_R, \lambda_I)$ is the characteristic function of X_k. The convergence of $\{Z_k\}$ implies that $\phi_k(\lambda_R, \lambda_I) \to \phi(\lambda_R, \lambda_I) := \sum q(\mu,\nu)\exp\{ib_R(\mu,\nu)\lambda_R +$

$ib_I(\mu, \nu)\lambda_I\}$ for all λ_R and λ_I and the regularity of **B** further implies that $|\phi(\lambda_R^0, 0)| = |\phi(0, \lambda_I^0)| = 1$ with $\lambda_R^0 := 2\pi/\gamma_B^R$ and $\lambda_I^0 := 2\pi/\gamma_B^I$. Combining these results leads to

$$(14) \qquad \lim_{k\to\infty} \prod \left| \psi_{k-j}\left(t_j^R\lambda_R^0, -t_j^I\lambda_R^0\right) \right| = \lim_{k\to\infty} \prod \left| \psi_{k-j}\left(t_j^I\lambda_I^0, t_j^R\lambda_I^0\right) \right| = 1.$$

For any given j, assume $\mathbf{p}_{k-j} \to \mathbf{p}$. Then, by the same argument as in the proof of Theorem 1, it can be shown from (14) that $|\psi_R(\lambda_R^0)| = |\psi_I(\lambda_I^0)| = 1$, where $\psi_R(\lambda)$ and $\psi_I(\lambda)$ are the characteristic functions of $M(\mathbf{C}_R; \mathbf{p})$ and $M(\mathbf{C}_I; \mathbf{p})$, respectively, with $\mathbf{C}_R := \{a_R(u,v)t_j^R - a_I(u,v)t_j^I\}$ and $\mathbf{C}_I := \{a_R(u,v)t_j^I + a_I(u,v)t_j^R\}$. By Lemma 2, there must be integers $m_j(u,v)$ and $n_j(u,v)$ such that

$$(15) \qquad \begin{aligned} m_j(u,v) &= \left(\gamma_A^R t_j^R u - \gamma_A^I t_j^I v\right)/\gamma_B^R \quad \text{and} \\ n_j(u,v) &= \left(\gamma_A^R t_j^I u + \gamma_A^I t_j^R v\right)/\gamma_B^I \end{aligned}$$

for all $(u,v) \in \mathbf{S}_A$ and for all j. Moreover, for any fixed $(u,v) \in \mathbf{S}_A$, the integer sequences $\{m_j(u,v)\}$ and $\{n_j(u,v)\}$ are absolutely summable, so that the set $\mathbf{J} := \{j\colon m_j^2(u_0,v_0) + n_j^2(u_0,v_0) > 0\}$ is finite for any fixed $(u_0,v_0) \in \mathbf{S}_A$ with $u_0^2 + v_0^2 > 0$. On the other hand, solving (15) for t_j^R and t_j^I yields $t_j^R = m_j(u_0,v_0)\alpha_R + n_j(u_0,v_0)\beta_R$ and $t_j^I = m_j(u_0,v_0)\alpha_I + n_j(u_0,v_0)\beta_I$, where $\alpha_R := \gamma_A^R\gamma_B^R u_0/\sigma$, $\beta_R := \gamma_A^I\gamma_B^I v_0/\sigma$, $\alpha_I := -\gamma_A^I\gamma_B^R v_0/\sigma$, $\beta_I := \gamma_A^R\gamma_B^I u_0/\sigma$ and $\sigma := (\gamma_A^R u_0)^2 + (\gamma_A^I v_0)^2$. This, combined with (15), implies that **J** is identical to the collection of j for which $t_j \neq 0$, so the product in (13) can be considered only for $j \in \mathbf{J}$. Assuming $\mathbf{p}_{k-j} \to \mathbf{q}_j = \{q_j(u,v), (u,v) \in \mathbf{S}_A\}$ for all $j \in \mathbf{J}$ leads to

$$(16) \qquad \sum q(\mathbf{w})\exp\{ia_R(\mathbf{w})\lambda_R + ia_I(\mathbf{w})\lambda_I\} = \phi(\lambda_R, \lambda_I),$$

where $\mathbf{w} := \{(u_j, v_j),\ j \in \mathbf{J}\}$, $(u_j, v_j) \in \mathbf{S}_A$, $a_R(\mathbf{w}) := \sum\{a_R(u_j, v_j)t_j^R - a_I(u_j,v_j)t_j^I\}$, $a_I(\mathbf{w}) := \sum\{a_R(u_j,v_j)t_j^I + a_I(u_j,v_j)t_j^R\}$ and $q(\mathbf{w}) := \prod q_j(u_j, v_j) > 0$. The rest of the proof is analogous to that of Theorem 1. □

REMARK 3. In addition to Remarks 1 and 2, one should also note that for complex systems the undetermined multiplier r in Theorem 2 could be a *complex* number. This means that in addition to the scale and shift ambiguities, as in Theorem 1, it is also possible that the deconvolution results be subject to an unknown *rotation*. Because of the regularity requirement on **B**, however, the number of possible angles of rotation is finite and limited to those that yield regular alphabets.

3. Applications. In this section, the general results of deconvolution are applied to two special cases of practical interest. The first case, typical in digital communications, assumes that the alphabet **A** is *completely* known. With this assumption, the problem is simplified to that of minimizing a cost function that measures the deviation of a random sequence from being **A**-valued. The second case further assumes that the system $\mathscr{S}$ is *parametric* (e.g., ARMA). In this case, an empirical criterion is designed to produce a consistent estimator for the parameter. It turns out that in many situations the convergence rate of the estimator can be much higher than the usual reciprocal of square root of the sample size, thanks to the utilization of the discreteness in the input distribution.

To remove the ambiguity of r in the previous theorems, assume in the sequel that the *value* of Σs_k is available and nonzero. Without loss of generality, let $\Sigma s_k = 1$, and denote by l_1^0 the totality of absolutely summable sequences that possess this property.

3.1. *Case of known alphabet.* When $\mathbf{A}$ is known and $\Sigma s_k = 1$, one can introduce a "cost function" $J(\mathscr{H})$ so that the deconvolution problem becomes a problem of minimizing $J(\mathscr{H})$ with respect to $\mathscr{H} \in l_1^0$. For brevity, let $\mathbf{A} = \{a(u), u \in \mathbf{S}_A\}$ be used to represent *both* real and complex alphabet. With this notation, one can define

$$(17) \qquad g(Z) := \prod |Z - a(u)|^2 \quad \text{with } L(Z) := E\{g(Z)\},$$

where the product is over $u \in \mathbf{S}_A$ and Z is a random variable of finite moments up to the order $2|\mathbf{A}|$. It is evident that $L(Z) \geq 0$ for any Z and $L(Z) = 0$ if and only if Z is a $\mathbf{B}$-valued random variable for some $\mathbf{B} \subseteq \mathbf{A}$. A cost function can be defined by

$$(18) \qquad J(\mathscr{H}) := \inf_{-\infty < k < \infty} L(Z_k(\mathscr{H})),$$

where $Z_k(\mathscr{H})$ is given by (2) and written explicitly as being a function of $\mathscr{H}$. Clearly, $J(\mathscr{H})$ attains its minimum value zero with $\mathscr{H} = \mathscr{S}^{-1}$ (or any shifted version of $\mathscr{S}^{-1}$) since $L(Z_k(\mathscr{S}^{-1})) = L(X_k) = 0$ for all k. The converse statement is guaranteed by the following theorem that extends some earlier results for binary sequences [Li (1992)].

THEOREM 3. *Let $\{X_k\}$ be a real-valued (or complex-valued) independent random sequence satisfying the conditions in Theorem 1 (Theorem 2). Assume in addition that $\mathbf{A}$ is known and $\Sigma s_k = 1$. If there exists a real (complex) filter $\mathscr{H} \in l_1^0$ such that $J(\mathscr{H}) \leq J(\mathscr{H}')$ for any real (complex) $\mathscr{H}' \in l_1^0$, then (4) holds with $r = 1$ and some integer K. An equivalent requirement on $\mathscr{H} \in l_1^0$ is that $J(\mathscr{H}) = 0$.*

PROOF. The equivalence is trivial. To prove the rest of the assertion, it suffices to note that if $J(\mathscr{H}) = 0$, one can always find, by (18), a subsequence of $\{L(Z_k)\}$, denoted by $\{L(Z_k)\}$ itself for simplicity, that converges to zero, where $Z_k := Z_k(\mathscr{H})$. Using Chebyshev's inequality, it is easy to show that $g(Z_k) \to_P 0$. Further, since $\{Z_k\}$ is almost surely (and hence stochastically) bounded, one can always find a subsequence from $\{Z_k\}$ that converges weakly to a random variable [Feller (1971), Theorem 1, page 267]. Denoting the convergent subsequence by $\{Z_k\}$ itself for brevity and its limit by Z, one obtains $Z_k \to Z$ in distribution. This in turn implies that $g(Z_k) \to g(Z)$ and hence $g(Z) = 0$, in distribution. Since $L(Z) = E\{g(Z)\} = 0$, then Z must be a $\mathbf{B}$-valued random variable for some $\mathbf{B} \subseteq \mathbf{A}$. The proof is complete upon invoking Lemma 3. □

REMARK 4. If $\mathbf{A}$ is *not* rotation-invariant, that is, if $r\mathbf{A} = \mathbf{A}$ with $|r| = 1$ implies $r = 1$, then the same result in Theorem 3 can be obtained without assuming $\sum s_k = 1$. In this case, it suffices to search for $\mathscr{H} \neq 0$ in l_1 (not in l_1^0) that minimizes $J(\mathscr{H})$.

3.2. *Case of parametric systems.* Suppose $\mathscr{S} := \mathscr{S}(\theta)$ is a parametric system characterized by a real-valued parameter vector $\theta := [\theta_1, \dots, \theta_m]^T$. Given a finite time series $\{Y_k,\ k = 0, \pm 1, \dots, \pm 2n\}$, the deconvolution problem becomes the estimation of the true value of θ, denoted by θ^*, that generates the Y_k and the reconstruction of $\{X_k\}$ using the estimated system.

For simplicity, the parameter θ is constrained in a neighborhood $\Theta := \{\theta: \|\theta - \theta^*\| \le \rho\}$ of the true value θ^*. Assume further that $\mathscr{S}(\theta)$ satisfies the following conditions:

C3. The system is normalized so that $\mathscr{S}(\theta) \in l_1^0$ for all $\theta \in \Theta$, and the parameterization is *unique* so that $s_k(\theta) = s_{k-K}(\theta')$ for all k implies $K = 0$ and $\theta = \theta'$.

C4. The series $\sum |s_k^{-1}(\theta)|$ converges *uniformly* in $\theta \in \Theta$.

C5. The sequence $\mathscr{S}^{-1}(\theta) = \{s_k^{-1}(\theta)\}$ has continuous and absolutely summable derivatives up to the third order.

C6. The sequences $\partial_p \mathscr{S}^{-1}(\theta^*)$ $(p = 1, \dots, m)$ are linearly independent, where ∂_p stands for the partial differentiation with respect to θ_p.

Under these conditions, θ^* can be estimated by minimizing the empirical criterion

$$\hat{J}_n(\theta) := \frac{1}{2n+1} \sum_{k=-n}^{n} g\big(\hat{Z}_k(\theta)\big)$$

with respect to θ, where $g(\cdot)$ is given by (17) and $\hat{Z}_k(\theta) := \sum_{j=-2n}^{2n} s_{k-j}^{-1}(\theta) Y_j$. Note that $\{\hat{Z}_k(\theta)\}$ depends solely on the observed time series $\{Y_k,\ k = 0, \pm 1, \dots, \pm 2n\}$.

Denote by $\hat{\theta}_n$ the minimizer of $\hat{J}_n(\theta)$ in a closed neighborhood $\Theta_0 \subseteq \Theta$ of θ^*, that is,

$$\hat{\theta}_n := \arg\min\big\{\hat{J}_n(\theta) : \theta \in \Theta_0\big\}. \tag{19}$$

The existence of $\hat{\theta}_n$ is ensured by the fact that $\hat{J}_n(\theta)$ is continuous in θ. The computation of $\hat{\theta}_n$ usually relies upon solving the system of (nonlinear) equations

$$\partial_p \hat{J}_n(\theta) = 0, \qquad p = 1, \dots, m, \tag{20}$$

by, for example, Newton–Raphson-type algorithms. Asymptotic behavior of $\hat{\theta}_n$ as an estimator of θ^* can be summarized in the following theorem.

THEOREM 4. *Suppose that the conditions in Theorem* 3 *and the additional assumptions C*3–*C*6 *are satisfied. Then the estimator* $\hat{\theta}_n$ *in* (19) *has the following properties.*

(i) *There exists a neighborhood* $\Theta_0 := \{\theta\colon \|\theta - \theta^*\| \le \rho_0\}$ *in which* $\hat{\theta}_n$ *is, with probability tending to unity as* $n \to \infty$, *the unique minimizer of* $\hat{J}_n(\theta)$ *and the unique solution to* (20). *It also converges to* θ^* *in probability.*
(ii) *If* $\{X_k\}$ *is stationary, then* $\hat{\theta}_n$ *is almost surely the unique solution to* (20) *in* Θ_0 *for sufficiently large n and converges to* θ^* *almost surely as* $n \to \infty$.
(iii) *There exists a constant* $c > 0$ *such that*

$$\|\hat{\theta}_n - \theta^*\| \le c \sum_{|k|>n} |s_k^{-1}(\theta^*)| \tag{21}$$

with probability tending to unity as $n \to \infty$ (*or, in the stationary case, almost surely for sufficiently large n*).

PROOF. The proof of (i) and (ii) is omitted since it closely follows the lines of Lehmann [(1983), Chapter 6] for maximum likelihood estimators and of the proof of Theorem 3 in Li (1992). The following facts are found to be helpful to note:

(a) The difference between $\hat{J}_n(\theta)$ and $\tilde{J}_n(\theta)$ tends to zero almost surely up to the second derivatives.
(b) $\partial_p \tilde{J}_n(\theta^*) = 0$.
(c) The third derivatives of $\hat{J}_n(\theta)$ are bounded.
(d) As ensured by Lemma 4(iii), the Hessian matrix of $\hat{J}_n(\theta)$ is uniformly positive definite in a suitable neighborhood $\Theta_0 \in \Theta$.
(e) The consistency is ensured by Lemma 4(ii) and the fact that $\tilde{J}_n(\hat{\theta}_n) \to 0$.

To prove part (iii) of the theorem, it is sufficient to consider at θ^* the second-order Taylor expansion of the gradient vector, denoted by $\nabla\hat{J}_n(\theta)$, of $\hat{J}_n(\theta)$. Since $\nabla\hat{J}_n(\hat{\theta}_n) = 0$ and $\hat{\theta}_n \to \theta^*$, as guaranteed by parts (i) and (ii) of the theorem, it is easy to show that

$$\hat{\Sigma}_n\big(\hat{\theta}_n - \theta^*\big) = -\nabla\hat{J}_n(\theta^*), \tag{22}$$

where $\hat{\Sigma}_n$ is an m-by-m matrix such that $\hat{\Sigma}_n - \Sigma_n \to_P 0$ with $\Sigma_n := \Phi_n + \overline{\Phi}_n$ and $\Phi_n := [\phi_{pq}]$ is an m-by-m matrix with ϕ_{pq} defined by (26) in Section 4. To evaluate $\nabla\hat{J}_n(\theta^*)$, let $f(Z) := \prod\{Z - a(u)\}$ and $\overline{W}_{pk} := f'(\hat{Z}_k(\theta^*))\,\partial_p\hat{Z}_k(\theta^*)$. Then it follows that

$$\partial_p \hat{J}_n(\theta^*) = \frac{1}{2n+1}\sum_{k=-n}^{n} 2\Re\Big[W_{pk}\Big\{f\big(\hat{Z}_k(\theta^*)\big) - f\big(Z_k(\theta^*)\big)\Big\}\Big],$$

where $Z_k(\theta)$ is defined by (24) in Section 4. Due to the boundedness of $\hat{Z}_k(\theta)$, $\partial_p\hat{Z}_k(\theta)$ and hence $f'(\hat{Z}_k(\theta))$, one can find constants $c_1 > 0$ and $c_2 > 0$ such that $|W_{pk}| \le c_1$ and $|f(\hat{Z}_k(\theta^*)) - f(Z_k(\theta^*))| \le c_2|\hat{Z}_k(\theta^*) - Z_k(\theta^*)|$ almost surely for all n, k and p. The boundedness of Y_k further implies the existence of a constant $c_3 > 0$ such that

$$|\hat{Z}_k(\theta^*) - Z_k(\theta^*)| \le c_3 \sum_{|j-k|>2n} |s_j^{-1}(\theta^*)| \le c_3 \sum_{|j|>n} |s_j^{-1}(\theta^*)|$$

almost surely for any $|k| \le n$ and for all n. Combining these results yields

$$\|\nabla \hat{J}_n(\theta^*)\| \le c_4 \sum_{|k|>n} |s_k^{-1}(\theta^*)| \quad \text{almost surely for all } n, \tag{23}$$

where $c_4 := 2c_1c_2c_3$. On the other hand, since Lemma 4(iii) guarantees that $\|\Sigma_n(\hat{\theta}_n - \theta^*)\| \ge \lambda_0\|\hat{\theta}_n - \theta^*\|$ with probability tending to unity and since $\|\hat{\Sigma}_n - \Sigma_n\| \to_P 0$, it follows from (22) that $\|\nabla\hat{J}_n(\theta^*)\| \ge \lambda_1\|\hat{\theta}_n - \theta^*\|$ with probability tending to unity for any $0 < \lambda_1 < \lambda_0$. This, combined with (23), yields (21) with $c = c_4/\lambda_1$. A similar argument applies to the stationary case. □

REMARK 5. Part (iii) of the theorem claims that the estimation error $\|\hat{\theta}_n - \theta^*\|$ depends on the tail of the true inverse system $\mathscr{S}^{-1}(\theta^*)$. When $\mathscr{S}(\theta^*)$ is an ARMA filter, for example, the right-hand side of (21) approaches zero *exponentially* as $n \to \infty$, so that the estimation error becomes $O(\varrho^n)$ for some $0 < \varrho < 1$. Recall that a typical error of maximum likelihood estimation is only $O(n^{-1/2})$. The higher-order accuracy, or "super-efficiency" as denoted by Davis and Rosenblatt (1991), of the estimator $\hat{\theta}_n$ is entirely due to the utilization of the discreteness of $\{X_k\}$ in the method. When $\mathscr{S}(\theta^*)$ is an AR system, in particular, the right-hand side of (21) vanishes for large n, leading to the conclusion that $\text{pr}\{\hat{\theta}_n = \theta^*\} \to 1$ as $n \to \infty$ [see also Li (1993a) for the binary case].

4. Lemmas.

LEMMA 1. *Let* $\mathbf{B}$ *and* $\mathbf{C} := \{c(u),\ u \in \mathbf{S}_C\}$ *be finite real (complex) alphabets with* $|\mathbf{C}| \ge |\mathbf{B}| \ge 2$. *If there exist real (complex) numbers* t *and* t' *such that* $\{c(u)t + c(u')t' + c\colon u, u' \in \mathbf{S}_C\} \subseteq \mathbf{B}$ *for some constant* c, *it is impossible for both* t *and* t' *to be nonzero.*

PROOF. By contradiction. Let both t and t' be nonzero and let $\mathbf{C}(u) := \{c(u)\,t + c(u')t' + c\colon u' \in \mathbf{S}_C\}$ for any fixed u. Then $\mathbf{C}(u) \subseteq \mathbf{B}$ for any $u \in \mathbf{S}_C$. Since $|\mathbf{C}(u)| = |\mathbf{C}|$ and $|\mathbf{B}| \le |\mathbf{C}|$, then $\mathbf{C}(u) = \mathbf{B}$ for any $u \in \mathbf{S}_C$. In particular, for fixed $u_1, u_2 \in \mathbf{S}_C$ with $u_1 \ne u_2$, one obtains $\mathbf{C}(u_1) = \mathbf{C}(u_2)$ and hence $c(u_1)t + c(u)t' = c(u_2)t + c(v(u))t'$ for all $u \in \mathbf{S}_C$, where $\{v(u),\ u \in \mathbf{S}_C\}$ is a permutation of $\mathbf{S}_C$. This implies that $(c(v(u)) - c(u))t' \equiv (c(u_1) - c(u_2))t$ for all u so that $c(v(u)) = c(u) + c_0$ for all $u \in \mathbf{S}_C$, where $c_0 := (c(u_1) - c(u_2))t/t'$. Adding up the quantities on each side of the expression gives $\sum c(v(u)) = \sum c(u) + |\mathbf{C}|c_0$, which in turn yields $c_0 = 0$ and hence $c(u_1) = c(u_2)$. This contradicts the assumption that $u_1 \ne u_2$. □

LEMMA 2. *Let* $\psi(\lambda)$ *be the characteristic function of* $M(\mathbf{C};\mathbf{p})$, *where* $\mathbf{C} = \{c(u),\ u \in \mathbf{S}_C\}$ *is a real alphabet. If* $|\psi(\lambda_0)| = 1$ *for some real number* λ_0, *then* *for any* $u_0 \in \mathbf{S}_C$ *there exist integers* $n(u)$ *such that* $(c(u) - c(u_0))\lambda_0 = 2\pi n(u)$ *for all* $u \in \mathbf{S}_C$.

PROOF. The result is trivial if $\lambda_0 = 0$. When $\lambda_0 \neq 0$, since $|\psi(\lambda_0)| = 1$, there exists a real number ϑ such that $\psi(\lambda_0) = \exp(i\vartheta)$, where $i := \sqrt{-1}$. Applying Lemma 3 of Feller [(1971), pages 500–501] to the random variable $X - \vartheta$ with $X \sim M(\mathbf{C}; \mathbf{p})$ yields $c(u) = \vartheta + m(u)\eta$ for all $u \in \mathbf{S}_C$, where the $m(u)$ are integers and $\eta := 2\pi/\lambda_0$. In particular, $c(u_0) = \vartheta + m(u_0)\eta$ and hence $c(u) - c(u_0) = (m(u) - m(u_0))\eta$ for all $u \in \mathbf{S}_C$. Taking $n(u) := m(u) - m(u_0)$ completes the proof. □

LEMMA 3. *Let* $\{X_k\}$ *be a real-valued (complex-valued) independent random sequence satisfying the conditions in Theorem* 1 *(Theorem* 2*). Assume in addition that* $\mathbf{A}$ *is known and* $\sum s_k = 1$. *If a real-valued (complex-valued) filter* $\mathscr{H} \in l_1^0$ *can be found such that a subsequence of* $\{Z_k\}$ *converges in distribution to* $M(\mathbf{B}; \mathbf{q})$ *for some* $\mathbf{B} \subseteq \mathbf{A}$, *then* (4) *holds with* $r = 1$ *and some integer* K.

PROOF. Since a subset of a regular alphabet is also regular, $\mathbf{B} \subseteq \mathbf{A}$ implies that $\mathbf{B}$ is regular with $|\mathbf{B}| \leq |\mathbf{A}|$. The assertion follows from Theorem 1 and Theorem 2 along with the fact that $r = \sum t_k = (\sum h_k)(\sum s_k) = 1$. □

To investigate the asymptotic behavior of $\hat{J}_n(\theta)$, it is helpful to introduce

$$Z_k(\theta) := \sum_{j=-\infty}^{\infty} s_{k-j}^{-1}(\theta) Y_j. \tag{24}$$

By assumption C4, it is easy to show that $\hat{Z}_k(\theta)$ converges to $Z_k(\theta)$ almost surely and *uniformly* in both $|k| \leq n$ and $\theta \in \Theta$. This in turn implies that $\hat{J}_n(\theta) - \tilde{J}_n(\theta) \to 0$ almost surely and *uniformly* in θ, where $\tilde{J}_n(\theta) := (2n + 1)^{-1}\sum_{k=-n}^{n} g(Z_k(\theta))$ is the counterpart of $\hat{J}_n(\theta)$ with $Z_k(\theta)$ in place of $\hat{Z}_k(\theta)$. The following lemma presents some properties of $\tilde{J}_n(\theta)$ that are found to be useful in the proof of Theorem 4:

LEMMA 4. *Suppose that the conditions in Lemma* 3 *and the assumptions* C3–C6 *are satisfied. Then* $\tilde{J}_n(\theta)$ *has the following properties.*

(i) *As* $n \to \infty$, $\sup\{|\tilde{J}_n(\theta) - E\{\tilde{J}_n(\theta)\}|: \theta \in \Theta\} \to 0$ *in probability. The limit exists almost surely if* $\{X_k\}$ *is stationary, that is,* $\mathbf{p}_k \equiv \mathbf{p}$ *for all* k.

(ii) *For any* $\eta > 0$, *there exists a constant* $\sigma > 0$ *such that* $\tilde{J}_n(\theta) \geq \sigma$ *for all* $\theta \in \Omega_\eta$ *with probability tending to unity as* $n \to \infty$, *where* $\Omega_\eta := \{\theta \in \Theta: \|\theta - \theta^*\| \geq \eta\}$.

(iii) *Let* $\hat{\lambda}_n(\theta)$ *be the smallest eigenvalue of the Hessian matrix of* $\tilde{J}_n(\theta)$. *Then there exist constants* $\lambda_0 > 0$ *and* $\rho_1 > 0$ *such that* $\hat{\lambda}_n(\theta) \geq \lambda_0$ *for all* $\theta \in \Theta_1 := \{\theta: \|\theta - \theta^*\| \leq \rho_1\} \subseteq \Theta$ *with probability tending to unity as* $n \to \infty$.

If $\{X_k\}$ *is stationary,* (ii) *and* (iii) *hold almost surely for sufficiently large* n.

PROOF OF (i). It suffices to show that $\sup|I_n(\theta)| \to 0$ in probability (or almost surely), where

$$I_n(\theta) := \frac{1}{2n+1} \sum_{k=-n}^{n} \left[\bar{Z}_k^p(\theta) Z_k^q(\theta) - E\{\bar{Z}_k^p(\theta) Z_k^q(\theta)\} \right]$$

and $p, q = 0, 1, \dots, |\mathbf{A}|$. To this end, write $I_n(\theta) = \Sigma \tau_{\mathbf{w}}(\theta) \zeta_n(\mathbf{w}) = \Sigma_{\mathbf{w} \in D} + \Sigma_{\mathbf{w} \notin D} := T_1(\theta) + T_2(\theta)$, where $\mathbf{w} := \{i(1), \dots, i(p); j(1), \dots, j(q)\}$,

$$\zeta_n(\mathbf{w}) := \frac{1}{2n+1} \sum_{k=-n}^{n} [\xi_k(\mathbf{w}) - E\{\xi_k(\mathbf{w})\}],$$

$$\xi_k(\mathbf{w}) := \prod_{u=1}^{p} \bar{X}_{k-i(u)} \prod_{v=1}^{q} X_{k-j(v)},$$

$$\tau_{\mathbf{w}}(\theta) := \prod_{u=1}^{p} \bar{t}_{i(u)}(\theta) \prod_{v=1}^{q} t_{j(v)}(\theta), \qquad t_k(\theta) := \sum_{j=-\infty}^{\infty} s_j^{-1}(\theta) s_{k-j}(\theta^*)$$

and $D = D_N := \{\mathbf{w}: |i(u)| \le N,\ |j(v)| \le N,\ \forall\ 1 \le u \le p,\ 1 \le v \le q\}$. For fixed $\mathbf{w}$, it is easy to show that $E|\zeta_n(\mathbf{w})|^2 = O(1/n)$ and hence $\zeta_n(\mathbf{w}) \to_P 0$. Since $\tau_{\mathbf{w}}(\theta)$ is uniformly bounded, it follows that $\sup|T_1(\theta)| \to_P 0$ as $n \to \infty$ for any fixed N. This implies that

$$\lim_{N \to \infty} \lim_{n \to \infty} \sup_{\theta \in \Theta} |T_1(\theta)| = 0 \quad \text{in probability.} \tag{25}$$

Further, since $|\zeta_n(\mathbf{w})| \le c$ for some constant $c > 0$ and all n so that $|T_2(\theta)| \le c_N(\theta) := c \Sigma_{w \notin D} |\tau_{\mathbf{w}}(\theta)|$ for all n, and since $c_N(\theta) \to 0$ uniformly in θ as $N \to \infty$, then

$$\lim_{N \to \infty} \limsup_{n \to \infty} \sup_{\theta \in \Theta} |T_2(\theta)| = 0 \quad \text{almost surely.}$$

This, coupled with (25), proves $\sup|I_n(\theta)| \to_P 0$. If $\{X_k\}$ is (strictly) stationary, so is $\{\xi_k(\mathbf{w})\}$. By the strong ergodic theorem [Karlin and Taylor (1975)], $\zeta_n(\mathbf{w}) \to 0$ a.s. so that (25) holds almost surely and $\sup|I_n(\theta)| \to 0$ a.s. □

PROOF OF (ii). Let $J_n(\theta) := E\{\tilde{J}_n(\theta)\} = (2n+1)^{-1} \Sigma_{k=-n}^{n} L(Z_k(\theta))$. Then, in the stationary case, $J_n(\theta) \equiv L(Z_0(\theta))$ and the uniform convergence in part (i) implies that $\inf\{\tilde{J}_n(\theta): \theta \in \Omega_\eta\} \to \inf\{L(Z_0(\theta)): \theta \in \Omega_\eta\}$ almost surely. Since $L(Z_0(\theta))$ is continuous in θ, the infimum of $L(Z_0(\theta))$ can be attained by some $\theta_0 \in \Omega_\eta$, and by Theorem 3, the minimum must be strictly positive. In the nonstationary case, it suffices to show, according to part (i) of the lemma, that there exists a constant $\sigma_0 > 0$ such that $J_n(\theta) \ge \sigma_0$ for all $\theta \in \Omega_\eta$ and for sufficiently large n. The assertion is valid, since otherwise one could find $\theta_0 \in \Omega_\eta$ and $\{k(N)\}$ with $k(N) \to \infty$ as $N \to \infty$ such that $L(Z_{k(N)}(\theta_0)) \to 0$, and this would lead to $J(\theta_0) := \inf\{L(Z_k(\theta_0))\} = 0$, which contradicts Theorem 3. □

PROOF OF (iii). Let $\Sigma_n := [\partial_{pq}^2 \tilde{J}_n(\theta^*)]$ be the Hessian matrix of $\tilde{J}_n(\theta^*)$. Then it is easy to verify that $\Sigma_n = \Phi_n + \bar{\Phi}_n$, where $\Phi_n := [\phi_{pq}]$ is an Hermitian matrix with

$$\phi_{pq} := \frac{1}{2n+1} \sum_{k=-n}^{n} \bar{\xi}_{pk} \xi_{qk}, \qquad p, q = 1, \dots, m, \tag{26}$$

$\xi_{pk} := f'(X_k)\, \partial_p Z_k(\theta^*)$ and $f(Z) := \prod\{Z - a(u)\}$. In addition, since $[\partial^2_{pq} \tilde{J}_n(\theta)] = \Sigma_n + O(\|\theta - \theta^*\|)$ almost surely and uniformly in θ, it suffices to show that the smallest eigenvalue of Φ_n, denoted by $\tilde{\lambda}_n$, is bounded below by a positive constant. To this end, let $t_{pk} := \sum(\partial_p s_j^{-1}(\theta^*)) s_{k-j}(\theta^*)$ so that $\partial_p Z_k(\theta^*) = \sum t_{pj} X_{k-j}$. For any complex-valued vector $\gamma := [\gamma_1, \dots, \gamma_m]^T$, consider $\gamma^H E(\Phi_n)\gamma = (2n+1)^{-1} \sum_{k=-n}^{n} E|U_k|^2$, where $U_k := \sum_{p=1}^{m} \gamma_p \xi_{pk} = f'(X_k) V_k$, $V_k := \sum \tau_j X_{k-j}$ and $\tau_j := \sum_{p=1}^{m} \gamma_p t_{pj}$. Since $|f'(X_k)|^2 \geq \alpha$ for some constant $\alpha > 0$ and for all k, then $E|U_k|^2 \geq \alpha\ \mathrm{var}(V_k) = \alpha \sum |\tau_j|^2\ \mathrm{var}(X_{k-j})$. Moreover, since $\mathbf{p}_k \geq \varepsilon > 0$, it follows that $\mathrm{var}(X_k) \geq \beta > 0$ for some β and all k so that $E|U_k|^2 \geq \alpha\beta \sum |\tau_j|^2$. It can also be shown under assumption C6 that $\sum |\tau_j|^2$ is a positive definite quadratic function of γ [Li (1992)] so that $\sum |\tau_j|^2 \geq \lambda_1 \|\gamma\|^2$ for some constant $\lambda_1 > 0$ and for all γ. Combining these results yields $\gamma^H E(\Phi_n)\gamma \geq \lambda_2$ for all γ with $\|\gamma\| = 1$, where $\lambda_2 := \alpha\beta\lambda_1 > 0$. Since $\Phi_n - E(\Phi_n) \to_P 0$ (or almost surely in the stationary case), the assertion follows from the identity $\tilde{\lambda}_n = \gamma_n^H \Phi_n \gamma_n = \gamma_n^H \{\Phi_n - E(\Phi_n)\}\gamma_n + \gamma_n^H E(\Phi_n)\gamma_n$, where γ_n is the unit-norm eigenvector associated with $\tilde{\lambda}_n$. □

REFERENCES

BENVENISTE, A., GOURSAT, M. and RUGET, G. (1980). Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications. *IEEE Trans. Automat. Control* **25** 385–399.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

CHENG, Q. (1990). Maximum standardized cumulant deconvolution of non-Gaussian linear processes. *Ann. Statist.* **18** 1745–1783.

DAVIS, R. A. and ROSENBLATT, M. (1991). Parameter estimation for some time series models without contiguity. *Statist. Probab. Lett.* **11** 515–521.

DONOHO, D. (1981). On minimum entropy deconvolution. In *Applied Time Series Analysis II* (D. F. Findley, ed.) 564–608. Academic Press, New York.

FEHER, K., ED. (1987). *Advanced Digital Communications—Systems and Signal Processing Techniques*. Prentice-Hall, Englewood Cliffs, NJ.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

GIANNAKIS, G. B. and MENDEL, J. M. (1989). Identification of non-minimum phase systems using higher order statistics. *IEEE Trans. Acoust. Speech Signal Process.* **37** 665–674.

KARLIN, S. and TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*. Academic Press, New York.

KORN, I. (1985). *Digital Communications*. Van Nostrand Reinhold, New York.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

LI, T. H. (1992). Blind identification and deconvolution of linear systems driven by binary random sequences. *IEEE Trans. Inform. Theory* **38** 26–38.

LI, T. H. (1993a). Estimation and blind deconvolution of AR systems with nonstationary binary inputs. *J. Time Ser. Anal.* **14** 575–588.

LI, T. H. (1993b). Blind deconvolution of discrete-valued signals. In *Proceedings of the IEEE 27th Asilomar Conference on Signals, Systems and Computers* **2** 1240–1244. IEEE, New York.

LII, K. S. and ROSENBLATT, M. (1982). Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. *Ann. Statist.* **10** 1195–1208.

SHALVI, O. and WEINSTEIN, E. (1990). New criteria for blind deconvolution of non-minimum phase systems (channels). *IEEE Trans. Inform. Theory* **36** 312–321.

DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143