Blind Channel Estimation Using the Second-Order Statistics: Asymptotic Performance and Limitations

Hanks H. Zeng, *Student Member, IEEE*, and Lang Tong, *Member, IEEE*

***Abstract*— In this paper, we consider the asymptotic performance and fundamental limitations of the class of blind estimators that use the second-order statistics. An achievable lower bound of the asymptotic normalized mean-square error (ANMSE) is derived. It is shown that the achievable ANMSE is lower bounded by the condition number of the Jacobian matrix of the correlation function with respect to the channel parameters. It is shown next that the Jacobian matrix is singular if and only if the subchannels share common conjugate reciprocal zeros. This condition is different from the existing channel identification conditions. Asymptotic performance of some existing eigenstructure-based algorithms is analyzed. Closed-form expressions of ANMSE and their lower bounds are derived for the least-squares (LS) and the subspace (SS) blind channel estimators when there are two subchannels. Asymptotic efficiency of LS/SS algorithms is also evaluated, showing that significant performance improvement is possible when the information of the source correlation is exploited.**

I. Introduction

ESTIMATING channels using only output signals is often referred to as blind channel estimation. Earlier techniques in blind channel estimation/equalization rely on some form of higher order statistics. This is natural since the single-input single-output linear time-invariant channel driven by stationary input signal cannot be identified from the second-order statistics of the output. Since the publication of [14] and [15], it becomes clear that such limitation does not apply to the case when the input signal is cyclostationary or, equivalently, to the case when the channel has single input and multiple outputs. The channel identification algorithm presented in [14] and [15] also offers a simple closed-form solution to the identification problem. A set of identifiability conditions is presented in [13].

The blind channel estimator given in [14] and [15] belongs to the class of *moment-based estimators*, i.e., the channel estimator is derived from the estimated correlations. Most recent moment-based blind channel estimators fall into the following two categories.

A. Eigenstructure-Based Approach [2], [7]–[9], [11]–[13], [17]

These eigenstructure-based algorithms share some common characteristics when compared with the original algorithm in [15]. They all exploit the special structure of the channel more efficiently than the original approach and often show improved performance. Second, they do not require the correlation structure of the source, which is of great value. On the other hand, when the source correlation structure is known, the omission of this information may degrade the performance significantly, as we shall demonstrate in this paper.

These algorithms are also developed around two original ideas ignored in [14] and [15]. The first one, presented succinctly by Moulines *et al.* [9], shows that the null space of the channel matrix determines the channel parameter uniquely. This idea forms the basis for the so-called subspace (SS) algorithm. The second idea, which has been presented in [6], [7], and [17], shows that the channel parameter is orthogonal to the null space of a special data matrix. It is shown that under the identifiable condition [13] and the persistent excitation condition, the null space of this special data matrix is colinear with the channel vector, which leads to the least square (LS) algorithm. Slock [12] provides a different interpretation using the linear prediction theory.

B. Optimal Spectral/Correlation Fitting [3]–[5], [16]

These algorithms are based on the idea of the optimal fitting of the output correlation function or spectrum. The so-called asymptotically best consistent (ABC) estimator presented by Giannakis and Halford [5], [4] is of particular theoretical importance. Although its implementation is complicated, the approach in [5], [4] is perhaps the first which addresses limitations of moment-based algorithms. The algorithm presented in [16] offers a simpler suboptimal implementation. These algorithms exploit both the structures of the channel and the source correlation. They are expected to offer better performance when the source correlation is known.

In this paper, algorithms using second-order statistics are compared based on the the asymptotic normalized mean-square error (ANMSE) of channel estimates. We attempt to answer the following basic questions concerning the blind channel estimators that use the second-order statistics:

- Q1) What is the *achievable* ANMSE among all consistent estimators using consistent estimates of second-order moments?
- Q2) What are fundamental limitations to the ANMSE of blind channel estimators using the second-order statistics?
- Q3) What is the ANMSE of the LS/SS estimators?
- Q4) What are the fundamental performance limitations on LS/SS estimators?
- Q5) How much potential improvement can be made over the LS/SS estimators?

Manuscript received May 10, 1995; revised February 28, 1997. This work was supported in part by the National Science Foundation under Contract NCR-9321813 and by the Advanced Research Projects Agency monitored by the Federal Bureau of Investigation under Contract J-FBI-94-221. The associate editor coordinating the review of this paper and approving it for publication was Prof. Hagit Messer-Yaron.

The authors are with the Department of Electrical and Systems Engineering, University of Connecticut, Storrs, CT 06269 USA.

Publisher Item Identifier S 1053-587X(97)05789-9.

Our approach is to apply asymptotic statistical analysis to the optimal and LS/SS algorithms and to derive expressions of ANMSE and their lower bounds. In obtaining the lower bounds, we aim to relate the ANMSE bounds to the channel characteristics such as the location of zeros and the condition number of channel matrix. Consequently, although the lower bounds may not always be tight, they illuminate key factors affecting the ANMSE. For example, our answer to Q4) and Q5) reveals insights into the value of exploiting the source correlation structure. Specifically, we show that despite the fact that LS/SS estimators and subspace techniques are very attractive in performance and especially in implementation, the optimal spectral/correlation fitting may offer better performance when the channel matrix is ill conditioned. In this paper, we do not address how to combine the simplicity of the LS/SS algorithms with the optimal approaches. A new approach motivated by this question is presented in [18] and [19]. Although some of the results in this paper are derived for the real case, similar results hold for the general complex case.

This paper is organized as follows. In Section II, the problem formulation is given. To answer Q1) and Q2) in Section III, the achievable ANMSE for all moment-based estimators is presented. We show that the achievable ANMSE is lower bounded by the condition number of the Jacobian matrix of the correlation function with respect to the channel parameters. A necessary and sufficient condition for nonsingular Jacobian matrix, which ensures the existence of the asymptotically best consistent (ABC) estimators, is given. In Section IV, the asymptotic performance of the LS/SS estimators is investigated. Here, the questions posed in Q3)–Q5) are answered. As an application of the theoretical analysis, we considered a set of two-ray multipath channels in Section V.

II. Problem Formulation

A. Notations

Notations used in this paper are standard. Uppercase and lowercase bold letters denote matrices and vectors, respectively. Other key notations are listed as follows:

$(\cdot)^T$	Transpose.
$(\cdot)^*$	Complex conjugate.
$(\cdot)^H$	Hermitian.
$E\{\cdot\}$	Expectation operator.
$\mathcal{Z}\{h(z)\}$	Zeros of $h(z)$.
$\|\boldsymbol{h}\|$	Two-norm.
$\boldsymbol{I}_n$	$n \times n$ identity matrix.
$\lambda_i(\boldsymbol{A})$	ith singular value of $\boldsymbol{A}$.
$\lambda_{\min}(\boldsymbol{A})(\lambda_{\max}(\boldsymbol{A}))$	Minimum (maximum) singular value of $\boldsymbol{A}$.
$\kappa(\boldsymbol{A})$	Condition number of $\boldsymbol{A}$.
$\mathrm{tr}\{\boldsymbol{A}\}$	Trace of $\boldsymbol{A}$.

B. Models

We consider a discrete-time single-input multiple-output (SIMO) model given by

$$x_i^{(j)} = \sum_k s_k h_{i-k}^{(j)} + n_i^{(j)}, \qquad j = 1, \cdots, M \tag{1}$$

where

s_k input sequence;
$h_i^{(j)}$ impulse response of the jth subchannel;
$n_i^{(j)}$ additive white noise.

In communication applications, the above model may come from a fractionally sampled baseband intersymbol interference (ISI) channel or systems using multiple receivers.

Assuming that the channel impulse response has a finite duration, let $L = \max_j \deg\{h^{(j)}(z)\}$, where $h^{(j)}(z)$ is the Z-transform of $h_i^{(j)}$. Denote $\boldsymbol{h}_j = [h_0^{(j)}, \cdots, h_L^{(j)}]^T$, and let the vector of the entire channel parameters be $\boldsymbol{h} = [\boldsymbol{h}_1^H, \cdots, \boldsymbol{h}_M^H]^H$. The single-input multiple-output (SIMO) model (1) has a matrix form

$$\boldsymbol{x}(i) = \mathcal{H}(\boldsymbol{h})\boldsymbol{s}(i) + \boldsymbol{n}(i) \tag{2}$$

where

$$\boldsymbol{s}(i) = [s_i, \cdots, s_{i-2L}]^T \tag{3}$$
$$\boldsymbol{x}(i) = [x_i^{(1)}, \cdots, x_{i-L}^{(1)}, \cdots, x_i^{(M)}, \cdots, x_{i-L}^{(M)}]^T \tag{4}$$
$$\boldsymbol{n}(i) = [n_i^{(1)}, \cdots, n_{i-L}^{(1)}, \cdots, n_i^{(M)}, \cdots, n_{i-L}^{(M)}]^T \tag{5}$$

are $2L+1, M(L+1)$ and $M(L+1)$ dimensional vectors, respectively, and the multichannel filtering transform (MFT) $\mathcal{H}(\boldsymbol{h})$ is defined by

$$\mathcal{H}(\boldsymbol{h}) = [\mathcal{F}^H(\boldsymbol{h}_1) \cdots \mathcal{F}^H(\boldsymbol{h}_M)]^H \tag{6}$$

$$\mathcal{F}(\boldsymbol{h}_i) = \begin{pmatrix} h_0^{(i)} & \cdots & h_L^{(i)} & & \\ & \ddots & & \ddots & \\ & & h_0^{(i)} & \cdots & h_L^{(i)} \end{pmatrix}_{(L+1)\times(2L+1)}. \tag{7}$$

The LS estimator requires the data window size of each channel to be $L+1$ in (4). Other estimators such as spectra/correlation fitting methods can be formulated from the estimated correlation/cyclic spectra of $x_i^{(j)}$. Since the achievable performance bound developed in Section III is related to the latter schemes, the restriction of the window size does not affect the results related to achievable performance bound.

C. Assumptions and Performance Measure

We assume that the input sequence $\{s_k\}$ is independent, identically distributed (i.i.d.), zero mean, and with unit variance. The noise $\{n_i^{(j)}\}$ is zero mean white Gausian noise with variance σ^2 and is mutually independent to the input. We also assume that the channel is normalized, i.e., $\|\boldsymbol{h}\|_2 = 1$, and the first nonzero element of $\boldsymbol{h}$ is positive. The signal-to-noise ratio (SNR) is defined by

$$\mathrm{SNR} \triangleq \frac{1}{M\sigma^2} E\left\{\sum_{j=1}^{M} |x_i^{(j)}|^2\right\} = \frac{1}{M\sigma^2}. \tag{8}$$

The performance measure considered in this paper is the *asymptotic normalized mean square error* (ANMSE). For an estimator $g(\cdot)$, given N_s data samples $\{\boldsymbol{x}(i), i = 0, \cdots, N_s -$

$1\}$, denote the channel estimate as $\hat{\boldsymbol{h}}(N_s)$. The ANMSE of the estimator $g(\cdot)$ is defined by

$$\text{ANMSE}(g) = \lim_{N_s\to\infty} N_s E\{\|\hat{\boldsymbol{h}}(N_s) - \boldsymbol{h}\|^2\} \tag{9}$$

when the limit exists. ANMSE(g) measures the mean square error of the consistent estimator g for a sufficiently large sample size N_s

$$E\{\|\hat{\boldsymbol{h}}(N_s) - \boldsymbol{h}\|^2\} \approx \frac{\text{ANMSE}(g)}{N_s}.$$

Obviously, small ANMSE(g) is desired. When ANMSE(g) $\to \infty$, it implies that the estimator g does not have convergence rate of $1/N_s$.

III. Asymptotic Performance of Moment-Based Estimators

We begin by asking the following question: *Given consistent estimates of the second-order moments of the received signal, what is the best achievable performance among all consistent channel estimators?* The answer to this question is significant in two aspects. First, it shows a fundamental limitation of all blind channel estimators based on the second-order statistics. Second, it provides a benchmark for the efficiency of existing blind channel estimators. For convenience, we consider in this section the case involving real data, and the input symbols are zero-mean binary signals (± 1).

A. The Autocorrelation and Its Jacobian Matrix

Given the correlation function $r_{ij}(u) \triangleq E\{x_t^{(i)} x_{t+u}^{(j)}\} = \sum_n h_n^{(i)} h_{n+u}^{(j)} + \sigma^2 \delta_{ij}\delta_u$, a consistent estimator of $r_{ij}(u)$ is given by

$$\hat{r}_{ij}(u) = \frac{1}{N_s - u} \sum_{t=0}^{N_s-1-u} x_t^{(i)} x_{t+u}^{(j)}. \tag{10}$$

To eliminate redundancies in $r_{ij}(u)$, define the vector of correlation functions by

$$\boldsymbol{r}_{ij} = \begin{cases} [r_{ij}(0), \cdots, r_{ij}(L)]^T, & i \le j \\ [r_{ij}(1), \cdots, r_{ij}(L)]^T, & i > j \end{cases} \tag{11}$$

$$\boldsymbol{r} = [\boldsymbol{r}_{11}^T, \cdots, \boldsymbol{r}_{1M}^T, \cdots, \boldsymbol{r}_{M1}^T, \cdots, \boldsymbol{r}_{MM}^T]^T. \tag{12}$$

Note that the estimated correlation $\hat{\boldsymbol{r}}$ contains all the second-order statistical information of the received signal.

The Jacobian matrix $\boldsymbol{S}(\boldsymbol{h}) = \partial \boldsymbol{r}/\partial \boldsymbol{h}$ plays a key role in the development. It is constructed from the following set of submatrices:

$$\frac{\partial \boldsymbol{r}_{ij}}{\partial \boldsymbol{h}_i} = \begin{pmatrix} h_0^{(j)} & \cdots & h_{L-1}^{(j)} & h_L^{(j)} \\ h_1^{(j)} & \cdots & h_L^{(j)} & \\ \vdots & & & \\ h_L^{(j)} & & & \end{pmatrix} \triangleq \boldsymbol{F}_j, \quad i < j \tag{13}$$

$$\frac{\partial \boldsymbol{r}_{ij}}{\partial \boldsymbol{h}_j} = \begin{pmatrix} h_0^{(i)} & \cdots & h_{L-1}^{(i)} & h_L^{(i)} \\ & h_0^{(i)} & \cdots & h_{L-1}^{(i)} \\ & & \ddots & \vdots \\ & & & h_0^{(i)} \end{pmatrix} \triangleq \boldsymbol{G}_i, \quad i < j \tag{14}$$

$$\frac{\partial \boldsymbol{r}_{ii}}{\partial \boldsymbol{h}_i} = \begin{pmatrix} h_0^{(i)} & \cdots & h_{L-1}^{(i)} & h_L^{(i)} \\ h_1^{(i)} & \cdots & h_L^{(i)} & \\ \vdots & & & \\ h_L^{(i)} & & & \end{pmatrix} + \begin{pmatrix} h_0^{(i)} & \cdots & h_{L-1}^{(i)} & h_L^{(i)} \\ & h_0^{(i)} & \cdots & h_{L-1}^{(i)} \\ & & \ddots & \vdots \\ & & & h_0^{(i)} \end{pmatrix} \triangleq \boldsymbol{F}_{ii} = \boldsymbol{F}_i + \boldsymbol{G}_i \tag{15}$$

$$\frac{\partial \boldsymbol{r}_{ij}}{\partial \boldsymbol{h}_i} = \begin{pmatrix} h_1^{(j)} & \cdots & h_L^{(j)} & 0 \\ \vdots & & & \\ h_L^{(j)} & & & \end{pmatrix} \triangleq \overline{\boldsymbol{F}}_j, \quad i > j \tag{16}$$

$$\frac{\partial \boldsymbol{r}_{ij}}{\partial \boldsymbol{h}_j} = \begin{pmatrix} 0 & h_0^{(i)} & \cdots & h_{L-1}^{(i)} \\ & & \ddots & \vdots \\ & & & h_0^{(i)} \end{pmatrix} \triangleq \overline{\boldsymbol{G}}_i, \quad i > j. \tag{17}$$

When $M = 2$, the Jacobian $\boldsymbol{S}(\boldsymbol{h})$ is given by

$$\boldsymbol{S}(\boldsymbol{h}) = \begin{pmatrix} \boldsymbol{F}_1 + \boldsymbol{G}_1 & \boldsymbol{0} \\ \boldsymbol{F}_2 & \boldsymbol{G}_1 \\ \overline{\boldsymbol{G}}_2 & \overline{\boldsymbol{F}}_1 \\ \boldsymbol{0} & \boldsymbol{F}_2 + \boldsymbol{G}_2 \end{pmatrix}. \tag{18}$$

When $M = 3$

$$\boldsymbol{S}(\boldsymbol{h}) = \begin{pmatrix} \boldsymbol{F}_1 + \boldsymbol{G}_1 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{F}_2 & \boldsymbol{G}_1 & \boldsymbol{0} \\ \boldsymbol{F}_3 & \boldsymbol{0} & \boldsymbol{G}_1 \\ \overline{\boldsymbol{G}}_2 & \overline{\boldsymbol{F}}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_2 + \boldsymbol{G}_2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{F}_3 & \boldsymbol{G}_2 \\ \overline{\boldsymbol{G}}_3 & \boldsymbol{0} & \overline{\boldsymbol{F}}_1 \\ \boldsymbol{0} & \overline{\boldsymbol{G}}_3 & \overline{\boldsymbol{F}}_2 \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{F}_3 + \boldsymbol{G}_3 \end{pmatrix}. \tag{19}$$

Similar matrices are easily constructed when $M > 3$. For notational convenience, we define the pairwise Jacobian matrix for $i \neq j$

$$\boldsymbol{S}_{ij}(\boldsymbol{h}) \triangleq \begin{pmatrix} \boldsymbol{F}_i + \boldsymbol{G}_i & \boldsymbol{0} \\ \boldsymbol{F}_j & \boldsymbol{G}_i \\ \overline{\boldsymbol{G}}_j & \overline{\boldsymbol{F}}_i \\ \boldsymbol{0} & \boldsymbol{F}_j + \boldsymbol{G}_j \end{pmatrix}, \quad i < j$$

$$\boldsymbol{S}_{ij}(\boldsymbol{h}) \triangleq \begin{pmatrix} \boldsymbol{F}_i + \boldsymbol{G}_i & \boldsymbol{0} \\ \overline{\boldsymbol{F}}_j & \overline{\boldsymbol{G}}_i \\ \boldsymbol{G}_j & \boldsymbol{F}_i \\ \boldsymbol{0} & \boldsymbol{F}_j + \boldsymbol{G}_j \end{pmatrix}, \quad i > j. \tag{20}$$

We note that the Jacobian matrix $\boldsymbol{S}(\boldsymbol{h})$ is specified by the set of $\{\boldsymbol{S}_{ij}(\boldsymbol{h})\}$.[1] $\boldsymbol{S}(\boldsymbol{h})$ also has the following property that is used in our development:

$$\boldsymbol{S}(\boldsymbol{h}) \begin{pmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_M \end{pmatrix} = \boldsymbol{0} \quad \text{iff} \quad \boldsymbol{S}_{ij}(\boldsymbol{h}) \begin{pmatrix} \boldsymbol{v}_i \\ \boldsymbol{v}_j \end{pmatrix} = \boldsymbol{0} \quad \forall i \neq j. \tag{21}$$

[1] Note that $\boldsymbol{S}_{ij}(\boldsymbol{h})$ is not necessarily the (i,j)th block in $\boldsymbol{S}(\boldsymbol{h})$.

B. The Achievable Performance Bound

The main result of this section is summarized in Theorem 1, where two lower bounds on ANMSE are given in (22). The first bound is significant because i) it is a lower bound for *all* consistent estimators using $\{\hat{r}_{ij}(u)\}$;, and ii) it is achievable by an optimal estimator $g_{\text{opt}}(\hat{\boldsymbol{r}})$. The second bound shows that when the Jacobian $\boldsymbol{S}(\boldsymbol{h}) = \partial \boldsymbol{r}/\partial \boldsymbol{h}$ of the autocorrelation vector $\boldsymbol{r}$ is nonsingular, there exists an estimator that achieves the lowest possible ANMSE. On the other hand, when $\boldsymbol{S}(\boldsymbol{h})$ is close to being singular, the performance degrades significantly. In other words, the performance of *all* moment-based blind channel estimation algorithms is limited by the condition number of the Jacobian matrix $\boldsymbol{S}(\boldsymbol{h})$, which is in sharp contrast with the behavior of the least-squares (LS) and subspace (SS) algorithms discussed in Section IV.

Theorem 1: Consider channel estimation in a parameter set Θ, which is open. Assume that $\boldsymbol{S}(\boldsymbol{h})$, which is the Jacobian of the autocorrelation vector $\boldsymbol{r}$ with respect to the channel vector $\boldsymbol{h}$, is full column rank. Let $\hat{\boldsymbol{r}}$ be the estimated autocorrelation vector obtained from $\{\boldsymbol{x}(i), i = 0, \cdots, N_s - 1\}$ with normalized asymptotic covariance $\Sigma(\boldsymbol{h}) > 0$. Let $\hat{\boldsymbol{h}} = g(\hat{\boldsymbol{r}})$ be a channel parameter estimator such that $g(\boldsymbol{r}) = \boldsymbol{h}$. Then, the ANMSE of $g(\hat{\boldsymbol{r}})$ is lower bounded by

$$\text{ANMSE} \geq \text{tr}\{(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1}\} \geq K_* \left(\frac{\kappa(\boldsymbol{S}(\boldsymbol{h}))}{\text{SNR}}\right)^2 \tag{22}$$

where $K_* = 1/4(L+1)M^3$ is a constant, and $\kappa(\boldsymbol{S}(\boldsymbol{h}))$ is the condition number of $\boldsymbol{S}(\boldsymbol{h})$. Moreover, there exists an estimator $g_{\text{opt}}(\hat{\boldsymbol{r}})$ that achieves the lower bound $\text{tr}\{(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1}\}$.

Proof: See Appendix A.

Remarks:

1) The parameter set Θ should be made so that full rank condition of $\boldsymbol{S}(\boldsymbol{h})$ and positive definite condition of $\Sigma(\boldsymbol{h})$ are satisfied for all $\boldsymbol{h}$ in Θ. The choice of Θ will be discussed in Section III-C.
2) In general, the lower bound $K_*(\kappa(\boldsymbol{S}(\boldsymbol{h}))/\text{SNR})^2$ is loose. However, this simple lower bound has some advantages. First, if this lower bound is very large, all second-order moment based estimators will have large ANMSE. Second, this lower bound depends only on several important factors such as the SNR and the condition number of the Jacobian. This allows us to reveal further insights into what kind of channels may have large ANMSE. See Section III-C.

C. The Existence of Asymptotic Best Consistent Estimators

Following [4] and [5], the estimator achieving this bound is referred to as the ABC estimator. Theorem 1 gives the achievable lower bound of ANMSE among all estimators using $\hat{\boldsymbol{r}}$. The lower bound on the achievable ANMSE shows that the condition number of the Jacobian matrix $\boldsymbol{S}(\boldsymbol{h})$ is a fundamental limitation. When $\boldsymbol{S}(\boldsymbol{h})$ is nonsingular, there exists an algorithm that achieves the lowest possible ANMSE. On the other hand, when $\boldsymbol{S}(\boldsymbol{h})$ is close to being singular, the performance of all moment-based algorithms deteriorates. The question addressed in this section is as follows: *When does $\boldsymbol{S}(\boldsymbol{h})$ become singular?* It turns out that the necessary and sufficient condition of $\boldsymbol{S}(\boldsymbol{h})$ being singular can be stated in terms of the locations of the channel zeros.

Theorem 2: Jacobian $\boldsymbol{S}(\boldsymbol{h})$ is singular if and only if $\{h^{(i)}(z)\}_{i=1}^M$ share common conjugate reciprocal zeros (CRZ's) or, equivalently, $\{h^{(i)}(z), h^{(i)}(1/z^*)\}_{i=1}^M$ share common zeros.

Proof: See Appendix C. □

Remarks: A z_0 is a CRZ if $H(z_0) = H(1/z_0^*) = 0$. CRZ's are symmetric with respect to the unit circle. For the real case, the CRZ's are equivalent to the *reciprocal zeros*. This is because zeros of a real polynomial come together with their conjugate counterpart. If z_0 and $1/z_0^*$ are zeros, so are z_0^* and $1/z_0$.

Theorem 2 highlights a striking difference from the channel identifiability condition, which states that the channel is not uniquely determined from the second-order statistics if the subchannels share common zeros [13]. Note that the identifiability condition is *global* in the sense that when the channel is not identifiable, there are multiple channels in the Euclidean space $R^{M(L+1)}$ satisfying the constraints imposed by the second-order moments of the received signal. In contrast, Theorem 2 says that some unidentifiable channels have nonsingular Jacobians. By properly specifying the parameter set Θ, these channels can be estimated with the achievable ANMSE shown in Theorem 1. Two choices of parameter set Θ can be made according to Theorem 2.

- *Identifiable Channels*: Consider the set of all identifiable channels. It is easy to verify that such a set is open. All channels in this set do not have common zeros, and hence, they have nonsingular Jacobians. Therefore, one can choose this set as Θ for which the requirements in Theorem 1 are satisfied.
- *Unidentifiable Channels without Common CRZ*: Consider a channel without CRZ. According to Theorem 2, an open neighborhood Θ around this channel can be chosen such that requirements in Theorem 1 are satisfied. Therefore, there is an ABC estimator for any channel $\boldsymbol{h}$ in Θ. In contrast with the previous case when the specification of this parameter set does not dependent on a particular channel, one must know the whereabouts of the true channel. Such knowledge may come from some prior information of the channel. In [18], we investigated the case when the channel is made of scaled and delayed versions of known pulse shape.

IV. Asymptotic Performance of Least Squares and Subspace Estimators

We consider in this section two recent blind channel estimators that have shown promising performance: the least-squares (LS) [2], [7], [11], [12], [17] and the subspace (SS) estimators [8], [12]. The questions asked here are as follows: *What is the asymptotic performance of the LS and SS estimators? What are fundamental limiting factors? How much improvement is possible?* We give answers for the LS estimators and the SS estimators when $M = 2$.

A. The LS and SS Estimators

The idea of the least-squares (LS) approach is motivated by the following convolution property (noise-free case): Given

$$x_i^{(1)} = h_i^{(1)} \odot s_i, \quad x_i^{(2)} = h_i^{(2)} \odot s_i \tag{23}$$

then

$$h_i^{(2)} \odot x_i^{(1)} = h_i^{(2)} \odot h_i^{(1)} \odot s_i = h_i^{(1)} \odot x_i^{(2)}. \tag{24}$$

Define the data-selection transform (DST) by

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{L+1} \\ -\boldsymbol{I}_{L+1} & \boldsymbol{0} \end{pmatrix}. \tag{25}$$

Using the notation defined in (2), we have

$$\boldsymbol{h}^H \boldsymbol{T} \boldsymbol{x}^*(i) = \boldsymbol{0}. \tag{26}$$

Hence, the LS estimator is given by

$$\begin{aligned} \hat{\boldsymbol{h}}_{LS} &= \arg \min_{\|\hat{\boldsymbol{h}}\|=1} \frac{1}{N_s} \sum_{i=0}^{N_s-1} \|\hat{\boldsymbol{h}}^H (\boldsymbol{T}\boldsymbol{x}^*(i))\|^2 \\ &= \arg \min_{\|\hat{\boldsymbol{h}}\|=1} \hat{\boldsymbol{h}}^H (\boldsymbol{T}\hat{\boldsymbol{R}}^* \boldsymbol{T}^H)\hat{\boldsymbol{h}} \triangleq \arg \min_{\|\hat{\boldsymbol{h}}\|=1} \hat{\boldsymbol{h}}^H \hat{\boldsymbol{Q}} \hat{\boldsymbol{h}} \end{aligned} \tag{27}$$

where

$$\hat{\boldsymbol{R}} \triangleq \frac{1}{N_s} \sum_{i=0}^{N_s-1} \{\boldsymbol{x}(i)\boldsymbol{x}(i)^H\}. \tag{28}$$

It can be seen that the LS approach is based on the orthogonality of the channel vector and the matrix $\hat{\boldsymbol{Q}}$ obtained from observation data.

The SS estimator is derived from a different angle [8], [12]. It exploits the Toeplitz structure of the channel matrix and obtains the null space of the channel vector differently. The singular value decomposition (SVD) of the estimated channel covariance matrix is given by

$$\hat{\boldsymbol{R}} = [\hat{\boldsymbol{U}}_s \hat{\boldsymbol{U}}_n] \begin{pmatrix} \hat{\Sigma}_s & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\Sigma}_n \end{pmatrix} [\hat{\boldsymbol{U}}_s \hat{\boldsymbol{U}}_n]^H \tag{29}$$

where columns of $\hat{\boldsymbol{U}}_s$ span the "signal" subspace and columns of $\hat{\boldsymbol{U}}_n = [\hat{\boldsymbol{u}}_{n1}, \cdots, \hat{\boldsymbol{u}}_{nm}]$ span the "noise" subspace. The SS estimator is given by

$$\hat{\boldsymbol{h}}_{SS} = \arg \min_{\|\hat{\boldsymbol{h}}\|=1} \hat{\boldsymbol{h}}^H \sum_i \mathcal{H}(\hat{\boldsymbol{u}}_{ni})\mathcal{H}(\hat{\boldsymbol{u}}_{ni})^H \hat{\boldsymbol{h}}. \tag{30}$$

Despite the apparent differences, these two estimators turn out to be identical when $M = 2$ with probability one [19], [20]. For the general case $M > 2$, the two algorithms are different in their ways of exploiting the noise subspace.

B. Asymptotic Performance of the LS/SS Estimator

We consider here the case $M = 2$ when the LS and SS estimators are identical. The following theorem gives the exact expression and the lower bound of ANMSE for the LS/SS estimator.

Theorem 3: Assume that the channel matrix $\mathcal{H}(\boldsymbol{h})$ is full rank, i.e., identifiable, and the sample covariance matrix of $\boldsymbol{x}(i)$ has the Wishart distribution. Then, the asymptotic normalized mean-square error of the least-squares blind channel estimation is given by

$$\text{ANMSE}_{\text{LS/SS}} = \sum_{k=1}^{2L+1} \frac{\sigma^2(\lambda_k^2 + \sigma^2)}{(\lambda_k^2)^2} \geq K_{\text{LS/SS}} \frac{\kappa^2(\mathcal{H}(\boldsymbol{h}))}{\text{SNR}} \tag{31}$$

where $\lambda_{\max} = \lambda_1 > \lambda_2 > \cdots > \lambda_{2L+1} = \lambda_{\min}$ are the singular values of the $\mathcal{H}(\boldsymbol{h})$, $\kappa(\mathcal{H}(\boldsymbol{h}))$ is the condition number of $\mathcal{H}(\boldsymbol{h})$, and $K_{\text{LS/SS}} = 1/2(L+1)$.

Proof: see Appendix D.

Remarks:

1) The limitation to the performance of LS/SS estimators is the identifiability condition, i.e., $\mathcal{H}(\boldsymbol{h})$ is nonsingular, or the subchannels do not share common zero. This is very different from the condition given in Theorem 1. It implies that when the subchannels share common zeros but not conjugate reciprocal ones, the LS/SS estimators fail, whereas there are locally consistent estimators that may still obtain accurate estimate of the true channel. Algorithms having this property are developed in [18] and [19].
2) One may question the assumption that the sample covariance matrix of $\boldsymbol{x}(i)$ has the Wishart distribution. This assumption is valid when $\boldsymbol{x}(i)$ is Gaussian and i.i.d., which, in general, is not true. We make this assumption because
   a) it enables a closed-form expression for the ANMSE;
   b) it is shown to be accurate in our simulations (see Section IV-D);
   c) it is valid when the input sequence is i.i.d. Gaussian and $\boldsymbol{x}(i)$ is subsampled to generate an i.i.d. Gaussian subsequence.

Note that the LS/SS methods can be derived when the symbol covariance matrix $\boldsymbol{R}_s \neq \boldsymbol{I}$. By replacing $\mathcal{H}(\boldsymbol{h})\mathcal{H}(\boldsymbol{h})^H$ in (96) by $\mathcal{H}(\boldsymbol{h})\boldsymbol{R}_s\mathcal{H}(\boldsymbol{h})^H$, it is evident that Theorem 3 is easily modified by defining the $\{\lambda_i\}$ as the singular values of the $\mathcal{H}(\boldsymbol{h})\boldsymbol{R}_s^{1/2}$.

C. Asymptotic Efficiency of the LS/SS Estimator

Here, we investigate how well the LS/SS estimator perform in the asymptotic case. When comparing different algorithms, a useful measure is the *asymptotic efficiency* $\eta \in [0, 1]$ defined below.

Definition 1: Given a moment-based consistent channel estimator g: $\hat{\boldsymbol{r}} \rightarrow \hat{\boldsymbol{h}}$ and its ANMSE(g), the asymptotic efficiency of $g(\cdot)$ is defined by

$$\eta = \frac{\text{ANMSE}(g_{\text{opt}})}{\text{ANMSE}(g)} \tag{32}$$

where ANMSE(g_{opt}) is the achievable lower bound on all moment-based estimators, i.e.,

$$\text{ANMSE}(g_{\text{opt}}) = \text{tr}\{(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1}\}. \tag{33}$$

With this definition, the asymptotic efficiency of the LS/SS approaches given below can be obtained easily by applying

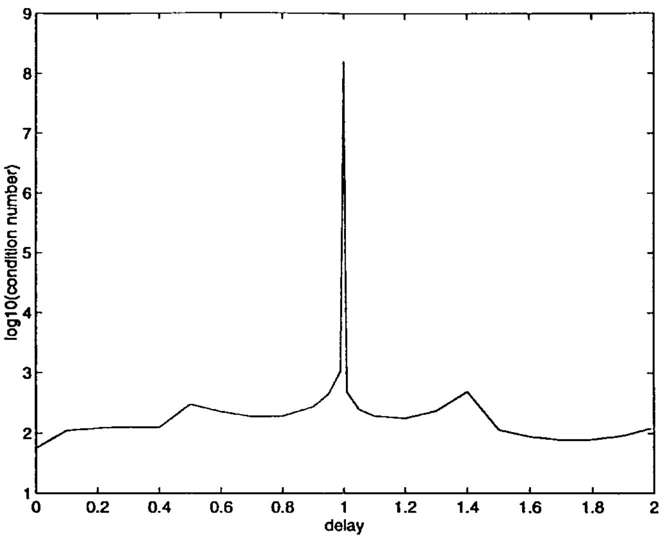

Fig. 1. Two-ray channels: Condition number versus delays.

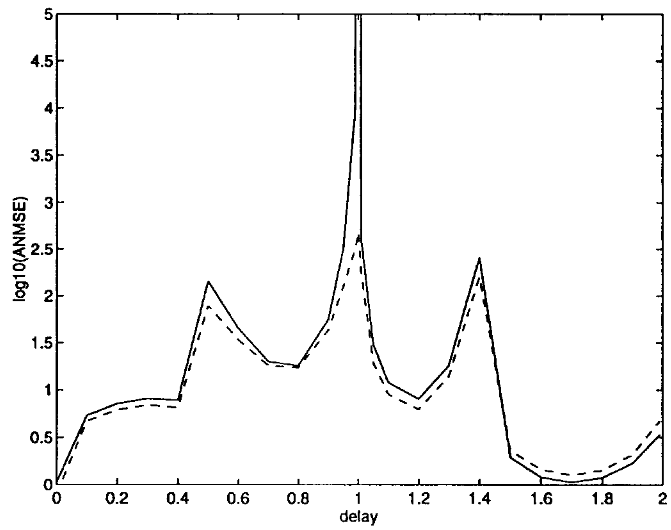

Fig. 2. Performance of the LS estimator: SNR $=35$ dB. Solid: ANMSE_{LS}. Dashed: $\overline{\text{ANMSE}}_{\text{LS/SS}}$ obtained from 50 Monte Carlo trials with $N_s = 1000$.

Theorems 3 and 1. Under assumptions in Theorems 3 and 1, the asymptotic efficiency of the LS/SS estimator is given by

$$\eta = \frac{\text{tr}\{(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1}\}}{\sum_{k=1}^{2L+1} \frac{\sigma^2(\lambda_k^2+\sigma^2)}{(\lambda_k^2)^2}}. \tag{34}$$

The upper bound of η provides much insight into the effect of singularity.

Theorem 4: Under the assumptions in Theorems 1 and 3, the asymptotic efficiency of the LS/SS estimator is upper bounded by

$$\eta \le K_\eta \, \text{SNR} \frac{\kappa^2(\boldsymbol{S}(\boldsymbol{h}))}{\kappa^2(\mathcal{H}(\boldsymbol{h}))} \tag{35}$$

where K_η is a constant depending on L, σ^2.

Proof: see Appendix E.

V. Example: Two-Ray Multipath Channels

In evaluating the ANMSE expression derived above, we considered a set of two-ray multipath channels given by

$$h(t) = p(t) + 0.7p(t-\tau) \tag{36}$$

where $p(t)$ was a raised-cosine function with rolloff factor 0.1 and finite support of $6T$. (T is the symbol interval.) τ varied from $0.1T$ to $1.9T$. Channel vector $\boldsymbol{h}$ was obtained from sampling $h(t)$ at a rate of $T/2$. The condition number of the channel matrix $\kappa(\mathcal{H}(\boldsymbol{h}))$ is shown in Fig. 1. It is expected that the LS/SS methods should fail at $\tau = 1T$. In this simulation, the transmitted signal was BPSK.

A. The ANMSE Expression for LS/SS Estimators

We evaluated first the accuracy of the ANMSE expression given in Theorem 3. Such an evaluation is particularly meaningful because the assumptions made in Theorem 3 may not be satisfied when a non-Gaussian input sequence is used.

In this example, the performance of the LS estimator was evaluated via 50 Monte Carlo runs at SNR $=35$ dB and $N_s = 1000$. The estimated asymptotic normalized mean square error is given by

$$\overline{\text{ANMSE}}_{\text{LS/SS}} = \frac{N_s}{50}\sum_{m=1}^{50} \|\hat{\boldsymbol{h}}^{(m)} - \boldsymbol{h}\|^2 \tag{37}$$

where $\hat{\boldsymbol{h}}^{(m)}$ was the estimated channel vector in the mth trial. The theoretical ANMSE was computed from Theorem 3.

Fig. 2 shows the comparison result. For the finite data size with BPSK input, the theoretical ANMSE predicted the behavior of the LS/SS estimator rather well. At $\tau = 1T$, the channel matrix $\mathcal{H}(\boldsymbol{h})$ is singular, which caused ANMSE to approach infinity. In this case, the LS/SS algorithm should fail, and the $\overline{\text{ANMSE}}_{\text{LS/SS}}$ has a large peak. To illustrate the effect of the condition number, Figs. 1 and 2 show a strong resemblance between the $\kappa(\mathcal{H}(\boldsymbol{h}))$ and the ANMSE curves.

B. ANMSE Bounds

In Fig. 3, ANMSE bounds (22, 31) of the optimal estimator and the LS/SS estimator were compared for the channel at delay $\tau = 0.5T$. It is interesting to observe that as the SNR increases, the performance of LS/SS estimator approaches to that of the optimal estimator. Both lower bounds of $\text{ANMSE}(g_{\text{opt}})$ and $\text{ANMSE}_{LS/SS}$ are not tight, and the lower bound of $\text{ANMSE}(g_{\text{opt}})$ (dash-dotted line) is very loose. However, these simple lower bounds can be used to indicate performance degradations when subchannels are close to sharing common zeros or common conjugate reciprocal zeros.

C. Asymptotic Efficiency

The channel condition is an important factor to the asymptotic efficiency. In this experiment, $\text{ANMSE}(g_{\text{opt}})$ and $\text{ANMSE}_{\text{LS/SS}}$ were compared for the set of two-ray multipath channels with different delays (Fig. 4). At $\tau = 1T$, the LS/SS estimator fails while the optimal estimator still performs well. The asymptotic efficiency of the LS/SS estimators is shown in Fig. 5. It is interesting to note that the LS/SS estimator had

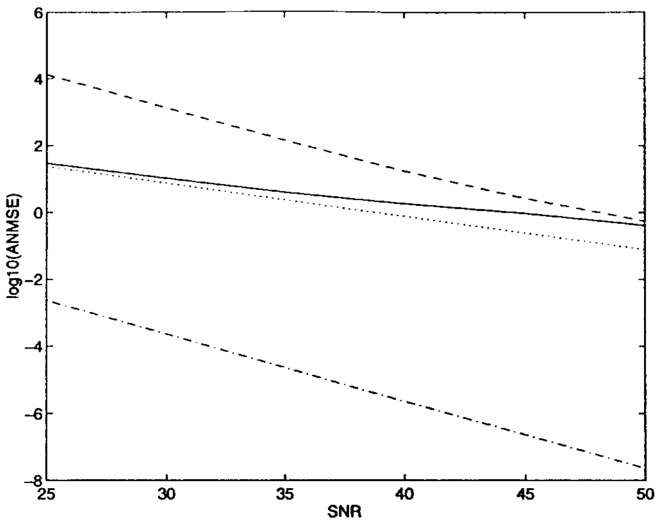

Fig. 3. Comparison of ANMSE's and their lower bounds: Dashed line: ANMSE_{LS}. Solid line: $\text{ANMSE}(g_{\text{opt}})$. Dotted line: lower bound of ANMSE_{LS}. Dash-dotted line: lower bound of $\text{ANMSE}(g_{\text{opt}})$.

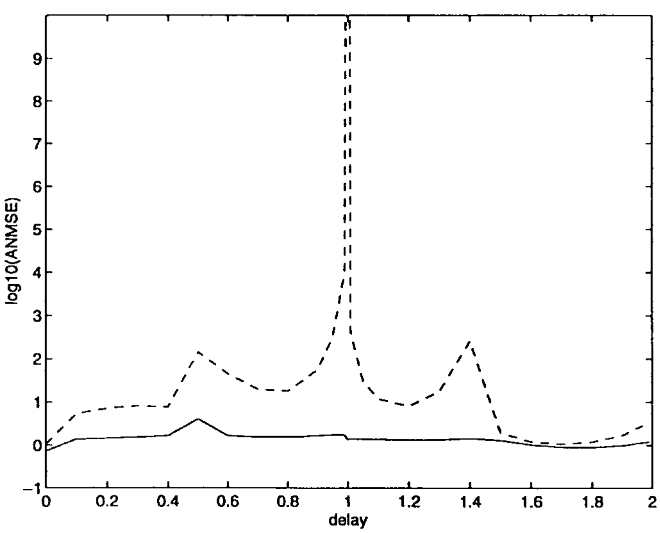

Fig. 4. Comparison the LS/SS and the optimal estimator: Dashed line: ANMSE_{LS}. Solid line: $\text{ANMSE}(g_{\text{opt}})$. SNR $= 35$ dB.

asymptotic efficiencies below 10% for some channels. Note that LS/SS method is derived without the knowledge of the source correlation structure, and the achievable ANMSE is derived under the assumption that the source is uncorrelated. The significant gap between the achievable ANMSE and that of the LS/SS estimators shows that the knowledge of the source correlation structure is of great value. In [18] and [19], we propose a class of new algorithms that can improve the performance of the LS/SS approaches.

VI. Conclusion

In this paper, we investigated the asymptotic performance of blind channel estimators using the second-order statistics. An achievable lower bound of the asymptotic normalized mean-square error (ANMSE) was derived to serve as a benchmark.

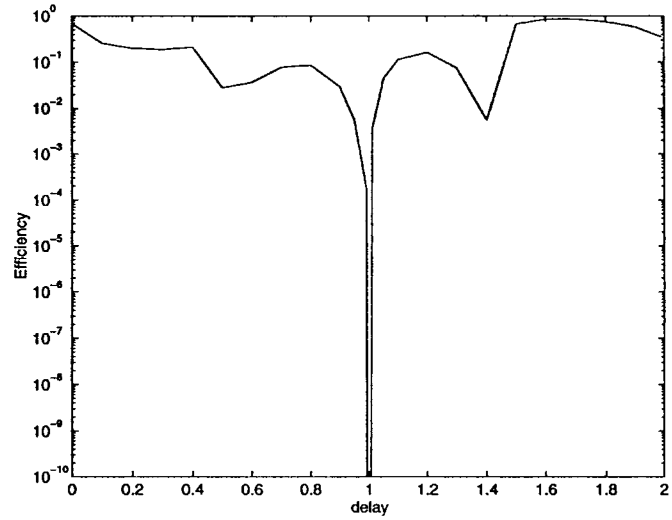

Fig. 5. Asymptotic efficiency of LS/SS estimator. SNR $= 35$ dB.

When a channel is $T/2$-sampled, the ANMSE of the least-squares estimator (or the subspace method with the same data window) was also derived. It is shown that the performance of the least-squares estimator degrades when the channel is close to unidentifiable, whereas the asymptotically best consistent estimator may perform well even when the channel does not have common conjugate reciprocal zeros. According to the analysis of relative efficiency of the least-squares estimator, significant potential improvement over the eigenspace-based algorithms is possible when the source correlation structures are exploited.

Appendix A
Proof of Theorem 1

The first inequality is a straightforward application of standard results (see [4], [5], and [10]). To prove the second inequality, we have

$$\begin{aligned}\text{ANMSE} &\geq \text{tr}((\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1})\\ &= \sum_i \lambda_i((\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1})\\ &= \sum_i \frac{1}{\lambda_i(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\\ &\geq \frac{1}{\lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}. \end{aligned} \tag{38}$$

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be the left and right singular vectors of $\boldsymbol{S}(\boldsymbol{h})$ associated with $\lambda_{\min}(\boldsymbol{S}(\boldsymbol{h}))$, respectively. We have

$$\begin{aligned}&\lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))\\ &\quad\leq \boldsymbol{v}^T\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h})\boldsymbol{v}\\ &\quad\leq \lambda_{\min}(\boldsymbol{S}(\boldsymbol{h}))^2(\boldsymbol{u}^H\Sigma^{-1}(\boldsymbol{h})\boldsymbol{u})\\ &\quad\leq \lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))\lambda_{\max}(\Sigma^{-1}(\boldsymbol{h})). \end{aligned} \tag{39}$$

We now claim

$$\Sigma(\boldsymbol{h}) \geq \sigma^4\boldsymbol{I}. \tag{40}$$

The proof of the above claim that involves deriving an expression of $\Sigma(\boldsymbol{h})$ is given in Appendix B. Applying (40)

$$\begin{aligned}\text{ANMSE} &\geq \frac{1}{\lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))\lambda_{\max}(\Sigma^{-1}(\boldsymbol{h}))}\\ &= \frac{\lambda_{\min}(\Sigma(\boldsymbol{h}))}{\lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\\ &\geq \frac{\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}{\lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\frac{\sigma^4}{(\|\boldsymbol{h}\|^2/M)^2}\\ &\quad\cdot\frac{(\|\boldsymbol{h}\|^2/M)^2}{\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\\ &= \frac{\kappa^2(\boldsymbol{S}(\boldsymbol{h}))}{\text{SNR}^2}\frac{(\|\boldsymbol{h}\|^2/M)^2}{\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}.\end{aligned}\tag{41}$$

Next, we show that $\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))$ is upper bounded by some quantity depending only on M, L, and $\|\boldsymbol{h}\|^2$. This allows us to conclude that the last term of (41) is lower bounded by the quantity depends only on M, L. (Recall that $\|\boldsymbol{h}\| = 1$.) From the definition of the Jacobian matrix

$$\boldsymbol{S}(\boldsymbol{h}) = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{h}} = \frac{\partial\begin{pmatrix}\vdots\\ r_{ij}(u)\\ \vdots\end{pmatrix}}{\partial \boldsymbol{h}} = \begin{pmatrix}\vdots\\ \left(\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}}\right)^T\\ \vdots\end{pmatrix}\tag{42}$$

we have

$$\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h})) \leq \text{tr}(\boldsymbol{S}(\boldsymbol{h})^T\boldsymbol{S}(\boldsymbol{h})) = \sum_{i,j,u}\left\|\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}}\right\|^2.\tag{43}$$

Since

$$\left\|\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}}\right\|^2 = \begin{cases}\left\|\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}_i}\right\|^2 + \left\|\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}_j}\right\|^2, & i \neq j\\ \left\|2\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}_i}\right\|^2, & i = j\end{cases}\tag{44}$$

then

$$\begin{aligned}&\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))\\ &\quad\leq \sum_{i,j,u}2\left(\left\|\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}_i}\right\|^2 + \left\|\frac{\partial r_{ij}(u)}{\partial \boldsymbol{h}_j}\right\|^2\right)\\ &\quad\leq \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{u=0}^{L}2(\|\boldsymbol{h}_i\|^2 + \|\boldsymbol{h}_j\|^2)\\ &\quad= \sum_{i=1}^{M}\sum_{j=1}^{M}2(L+1)(\|\boldsymbol{h}_i\|^2 + \|\boldsymbol{h}_j\|^2)\\ &\quad= 2(L+1)\sum_{i=1}^{M}(M\|\boldsymbol{h}_i\|^2 + \|\boldsymbol{h}\|^2)\\ &\quad\leq 4(L+1)M\|\boldsymbol{h}\|^2.\end{aligned}\tag{45}$$

Hence

$$\frac{(\|\boldsymbol{h}\|^2/M)^2}{\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))} \geq \frac{1}{4(L+1)M^3} \triangleq K_*.\tag{46}$$

Substituting the above into (41), we have now completed the proof. □

Appendix B
$\Sigma(\boldsymbol{h})$ and Its Lower Bound

Denote $y_t^{(i)}$ as the signal (noise-free) output corresponding to $x_t^{(i)}$, i.e.,

$$y_t^{(i)} = \sum_k s_k h_{t-k}^{(i)}\tag{47}$$

$$x_t^{(i)} = y_t^{(i)} + n_t^{(i)}.\tag{48}$$

Then, the estimated correlation of the channel output is given by

$$\hat{r}_{ij}(u) = \frac{1}{N_s}\sum_{t=0}^{N_s-1}x_t^{(i)}x_{t+u}^{(j)}\tag{49}$$

$$\begin{aligned}&= \underbrace{\frac{1}{N_s}\sum_{t=0}^{N_s-1}y_t^{(i)}y_{t+u}^{(j)}}_{\hat{r}_{ij}^{(yy)}(u)} + \underbrace{\frac{1}{N_s}\sum_{t=0}^{N_s-1}y_t^{(i)}n_{t+u}^{(j)}}_{\hat{r}_{ij}^{(yn)}(u)}\\ &\quad+ \underbrace{\frac{1}{N_s}\sum_{t=0}^{N_s-1}n_t^{(i)}y_{t+u}^{(j)}}_{\hat{r}_{ij}^{(ny)}(u)} + \underbrace{\frac{1}{N_s}\sum_{t=0}^{N_s-1}n_t^{(i)}n_{t+u}^{(j)}}_{\hat{r}_{ij}^{(nn)}(u)}.\end{aligned}\tag{50}$$

Similar to the definition of $\hat{\boldsymbol{r}}$ [see (12)], we define vectors of $\hat{\boldsymbol{r}}^{(yy)}, \hat{\boldsymbol{r}}^{(nn)}, \hat{\boldsymbol{r}}^{(ny)}$ and $\hat{\boldsymbol{r}}^{(yn)}$. Thus, we can write the estimated correlations in the vector form as

$$\hat{\boldsymbol{r}} = \hat{\boldsymbol{r}}^{(yy)} + \hat{\boldsymbol{r}}^{(yn)} + \hat{\boldsymbol{r}}^{(ny)} + \hat{\boldsymbol{r}}^{(nn)}.\tag{51}$$

The correlations are given by

$$\begin{aligned}\underbrace{E\{x_t^{(i)}x_{t+u}^{(j)}\}}_{r_{ij}(u)} &= \underbrace{E\{y_t^{(i)}y_{t+u}^{(j)}\}}_{r_{ij}^{(yy)}(u)} + \underbrace{E\{y_t^{(i)}n_{t+u}^{(j)}\}}_{r_{ij}^{(yn)}(u)}\\ &\quad+ \underbrace{E\{n_t^{(i)}y_{t+u}^{(j)}\}}_{r_{ij}^{(ny)}(u)} + \underbrace{E\{n_t^{(i)}n_{t+u}^{(j)}\}}_{r_{ij}^{(nn)}(u)}\end{aligned}\tag{52}$$

where

$$r_{ij}^{(yy)}(u) = \sum_n h_n^{(i)}h_{n+u}^{(j)}E\{s_t^2\}\tag{53}$$

$$r_{ij}^{(nn)}(u) = \sigma^2\delta_{ij}\delta_u\tag{54}$$

$$r_{ij}^{(yn)}(u) = r_{ij}^{(ny)}(u) = 0.\tag{55}$$

In addition, similar to the definition of $\boldsymbol{r}$ [see (12)], we define vectors of $\boldsymbol{r}^{(yy)}, \boldsymbol{r}^{(nn)}, \boldsymbol{r}^{(ny)}$ and $\boldsymbol{r}^{(yn)}$. Therefore, the mean of the estimated correlations is given by

$$\begin{aligned}\underbrace{E\{\hat{\boldsymbol{r}}\}}_{\boldsymbol{r}} &= \underbrace{E\{\hat{\boldsymbol{r}}^{(yy)}\}}_{\boldsymbol{r}^{(yy)}} + \underbrace{E\{\hat{\boldsymbol{r}}^{(yn)}\}}_{\boldsymbol{r}^{(yn)}}\\ &\quad+ \underbrace{E\{\hat{\boldsymbol{r}}^{(ny)}\}}_{\boldsymbol{r}^{(ny)}} + \underbrace{E\{\hat{\boldsymbol{r}}^{(nn)}\}}_{\boldsymbol{r}^{(nn)}}\\ &= \underbrace{E\{\hat{\boldsymbol{r}}^{(yy)}\}}_{\boldsymbol{r}^{(yy)}} + \underbrace{E\{\hat{\boldsymbol{r}}^{(nn)}\}}_{\boldsymbol{r}^{(nn)}}.\end{aligned}\tag{56}$$

The *normalized asymptotic covariance matrix* is given by

$$\mathcal{AC}(\hat{\boldsymbol{r}}) = \lim_{N_s\to\infty} N_s E\{(\hat{\boldsymbol{r}}-\boldsymbol{r})(\hat{\boldsymbol{r}}-\boldsymbol{r})^T\} \tag{57}$$

$$= \mathcal{AC}(\hat{\boldsymbol{r}}^{(yy)}) + \mathcal{AC}(\hat{\boldsymbol{r}}^{(yn)}+\hat{\boldsymbol{r}}^{(ny)}) + \mathcal{AC}(\hat{\boldsymbol{r}}^{(nn)}). \tag{58}$$

Now, we derive each matrix in (58). First, we derive entries of $\mathcal{AC}(\hat{\boldsymbol{r}}^{(yy)})$.

$$\begin{aligned}\mathcal{AC}&(\hat{r}_{ij}^{(yy)}(u), \hat{r}_{kl}^{(yy)}(v))\\ &= \lim_{N_s\to\infty} N_s \Bigg(\frac{1}{N_s^2}\sum_{t=0}^{N_s-1}\sum_{t'=0}^{N_s-1}\\ &\quad\cdot E\{y_t^{(i)}y_{t+u}^{(j)}y_{t'}^{(k)}y_{t'+v}^{(l)}\} - r_{ij}^{(yy)}(u)r_{kl}^{(yy)}(v)\Bigg)\\ &= \lim_{N_s\to\infty} N_s\Bigg(\frac{1}{N_s^2}\sum_{t',t}\sum_{n_i,n_j,n_k,n_l}\\ &\quad\cdot E\{h_{n_i}^{(i)}s_{t-n_i}h_{n_j}^{(j)}s_{t+u-n_j}h_{n_k}^{(k)}s_{t'-n_k}h_{n_l}^{(l)}s_{t'+v-n_l}\}\\ &\quad - r_{ij}^{(yy)}(u)r_{kl}^{(yy)}(v)\Bigg)\\ &= \lim_{N_s\to\infty} N_s\Bigg(\frac{1}{N_s^2}\sum_{t',t}\sum_{n_i,n_j,n_k,n_l}\\ &\quad\cdot h_{n_i}^{(i)}h_{n_j}^{(j)}h_{n_k}^{(k)}h_{n_l}^{(l)}E\{s_{t-n_i}s_{t+u-n_j}s_{t'-n_k}s_{t'+v-n_l}\}\\ &\quad - r_{ij}^{(yy)}(u)r_{kl}^{(yy)}(v)\Bigg).\end{aligned} \tag{59}$$

Since BPSK signal $\{s_t\}$ is i.i.d.

$$\begin{aligned}E\{&s_{t-n_i}s_{t+u-n_j}s_{t'-n_k}s_{t'+v-n_l}\}\\ &= (E\{s_t^2\})^2[\delta(u-n_j+n_i)\delta(v-n_l+n_k)\\ &\quad+\delta(t'-t-n_k+n_i)\delta(t'-t+v-u-n_l+n_j)\\ &\quad+\delta(t'-t+v-n_l+n_i)\delta(t'-t-u-n_k+n_j)\\ &\quad+\delta(u-n_j+n_i)\delta(v-n_l+n_k)\\ &\quad\cdot\delta(t'-t-n_k+n_i)\kappa_4(s)]\end{aligned} \tag{60}$$

where

$$\kappa_4(s) = \frac{E\{s_t^4\} - 3(E\{s_t^2\})^2}{(E\{s_t^2\})^2}. \tag{61}$$

Hence, we have

$$\begin{aligned}\mathcal{AC}&(\hat{r}_{ij}^{(yy)}(u), \hat{r}_{kl}^{(yy)}(v))\\ &= \sum_\tau [r_{ik}^{(yy)}(\tau)r_{jl}^{(yy)}(\tau+v-u)\\ &\quad + r_{il}^{(yy)}(\tau+v)r_{jk}^{(yy)}(\tau-u)]\\ &\quad + \kappa_4(w)r_{ij}^{(yy)}(u)r_{kl}^{(yy)}(v).\end{aligned} \tag{62}$$

Similarly, we have

$$\begin{aligned}\mathcal{AC}&(\hat{r}_{ij}^{(yn)}(u)+\hat{r}_{ij}^{(ny)}(u), \hat{r}_{kl}^{(yn)}(v)+\hat{r}_{kl}^{(ny)}(v))\\ &= \sigma^2 r_{ik}^{(yy)}(u-v)\delta_{jl} + \sigma^2 r_{jl}^{(yy)}(v-u)\delta_{ik}\\ &\quad + \sigma^2 r_{jk}^{(yy)}(-u-v)\delta_{il} + \sigma^2 r_{il}^{(yy)}(u+v)\delta_{jk}.\end{aligned} \tag{63}$$

For the last term in (58), since $\{n_t^{(i)}\}$ is white Gaussian, we have

$$\begin{aligned}\mathcal{AC}&(\hat{r}_{ij}^{(nn)}(u), \hat{r}_{kl}^{(nn)}(v))\\ &= \lim_{N_s\to\infty} N_s\Bigg(\frac{1}{N_s^2}\sum_{t=0}^{N_s-1}\sum_{t'=0}^{N_s-1}\\ &\quad\cdot E\{n_t^{(i)}n_{t+u}^{(j)}n_{t'}^{(k)}n_{t'+v}^{(l)}\} - \sigma^2\delta_{ij}\delta_u\Bigg)\\ &= \sigma^4\delta_{ik}\delta_{jl}\delta_{u-v} + \sigma^4\delta_{il}\delta_{jk}\delta_{u+v}.\end{aligned} \tag{64}$$

Now, we give the lower bound of $\Sigma(\boldsymbol{h})$. According to (58)

$$\Sigma(\boldsymbol{h}) = \mathcal{AC}(\hat{\boldsymbol{r}}) \ge \mathcal{AC}(\hat{\boldsymbol{r}}^{(nn)}) \tag{65}$$

where each entry of $\mathcal{AC}(\hat{\boldsymbol{r}}^{(nn)})$ is given by (64). Since $\hat{r}_{ij}^{(nn)}(u)$ is defined at $u = 0, \cdots, L$ when $i \le j$ and at $u = 1, \cdots, L$ when $i > j$, it is easy to verify that

$$\begin{aligned}\delta_{ik}&\delta_{jl}\delta_{u-v} + \delta_{il}\delta_{jk}\delta_{u+v}\\ &= \begin{cases} 2, & i=l=j=k, u=v=0\\ 1, & i=l=j=k, u=v\neq 0\\ 1, & i=k, j=l, i\neq j, u=v\\ 0, & \text{otherwise.}\end{cases}\end{aligned} \tag{66}$$

Therefore, $\mathcal{AC}(\hat{\boldsymbol{r}}^{(nn)})$ is a diagonal matrix. Hence, we have

$$\Sigma(\boldsymbol{h}) \ge \sigma^4\boldsymbol{I}. \tag{67}$$

Appendix C
Proof of Theorem 2

The proof of Theorem 2 depends on the following useful Lemma.

Lemma 1: Jacobian $\boldsymbol{S}_{ij}(\boldsymbol{h}), i \neq j$ is singular iff there exist two vectors $\boldsymbol{v}_i = [v_0^{(i)}, \cdots, v_{-L}^{(i)}]^T$ and $\boldsymbol{v}_j = [v_0^{(j)}, \cdots, v_{-L}^{(j)}]^T$ such that

$$h^{(i)}(z)v^{(i)}(z) + h^{(i)}(1/z)v^{(i)}(1/z) = 0 \tag{68}$$

$$h^{(j)}(z)v^{(j)}(z) + h^{(j)}(1/z)v^{(j)}(1/z) = 0 \tag{69}$$

$$h^{(i)}(z)v^{(j)}(z) + h^{(j)}(1/z)v^{(i)}(1/z) = 0 \tag{70}$$

where $h^{(i)}(z), h^{(j)}(z), v^{(i)}(z), v^{(j)}(z)$ are z transforms of $\boldsymbol{h}_i, \boldsymbol{h}_j, \boldsymbol{v}_i, \boldsymbol{v}_j$, respectively.

Proof: Without loss of generality, we assume that $i < j$. With the definition of $\boldsymbol{S}_{ij}(\boldsymbol{h})$ in (20), $\boldsymbol{S}_{ij}(\boldsymbol{h})$ is singular iff there exist two vectors $\boldsymbol{v}_i = [v_0^{(i)}, \cdots, v_{-L}^{(i)}]^T, \boldsymbol{v}_j = [v_0^{(j)}, \cdots, v_{-L}^{(j)}]^T$ such that

$$\boldsymbol{S}_{ij}(\boldsymbol{h})\begin{pmatrix}\boldsymbol{v}_i\\ \boldsymbol{v}_j\end{pmatrix} = \begin{pmatrix}\boldsymbol{F}_i+\boldsymbol{G}_i & \boldsymbol{0}\\ \boldsymbol{F}_j & \boldsymbol{G}_i\\ \boldsymbol{G}_j & \boldsymbol{F}_i\\ \boldsymbol{0} & \boldsymbol{F}_j+\boldsymbol{G}_j\end{pmatrix}\begin{pmatrix}\boldsymbol{v}_i\\ \boldsymbol{v}_j\end{pmatrix} = \boldsymbol{0}. \tag{71}$$

Define the cross-correlation function between $\boldsymbol{h}_i$ and $\boldsymbol{v}_j$ by

$$\rho_{ij}(\tau) \triangleq \sum_{k=-L}^{0} v_k^{(j)}h_{\tau-k}^{(i)}, \qquad \tau = -L, \cdots, L \tag{72}$$

where we assume $h_\tau^{(i)} = 0$ when $\tau < 0$. Then

$$\boldsymbol{F}_i \boldsymbol{v}_j = \begin{pmatrix} h_0^{(i)} & \cdots & h_L^{(i)} \\ \vdots & \iddots & \\ h_L^{(i)} & & \end{pmatrix} \begin{pmatrix} v_0^{(j)} \\ \vdots \\ v_{-L}^{(j)} \end{pmatrix} = \begin{pmatrix} \rho_{ij}(0) \\ \vdots \\ \rho_{ij}(L) \end{pmatrix} \tag{73}$$

$$\boldsymbol{G}_i \boldsymbol{v}_j = \begin{pmatrix} h_0^{(i)} & \cdots & h_L^{(i)} \\ & \ddots & \vdots \\ & & h_0^{(i)} \end{pmatrix} \begin{pmatrix} v_0^{(j)} \\ \vdots \\ v_{-L}^{(j)} \end{pmatrix} = \begin{pmatrix} \rho_{ij}(0) \\ \vdots \\ \rho_{ij}(-L) \end{pmatrix}. \tag{74}$$

Hence, (71) can be rewritten as

$$\rho_{ii}(\tau) + \rho_{ii}(-\tau) = 0, \qquad \tau = 0, \cdots, L \tag{75}$$

$$\rho_{ji}(\tau) + \rho_{ij}(-\tau) = 0, \qquad \tau = 0, \cdots, L \tag{76}$$

$$\rho_{ij}(\tau) + \rho_{ji}(-\tau) = 0, \qquad \tau = 1, \cdots, L \tag{77}$$

$$\rho_{jj}(\tau) + \rho_{jj}(-\tau) = 0, \qquad \tau = 0, \cdots, L. \tag{78}$$

Taking z transforms, we have (68)–(70). □

Now we will give the proof of Theorem 2 for the real case where the conjugate reciprocal zero is equivalent to reciprocal zero. The proof for the general complex case can be found in [21].

Sufficient Condition

First, consider the case where $\{h^{(i)}(z)\}$ share common reciprocal zero $z_0 = 1$.

$$h^{(i)}(z) = g^{(i)}(z)(1 - z^{-1}), \qquad i = 1, \cdots, M. \tag{79}$$

Let

$$v^{(i)}(1/z) = g^{(i)}(z)(1 + z^{-1}), \qquad i = 1, \cdots, M. \tag{80}$$

It can be easily verified that $\{v^{(i)}(1/z), v^{(j)}(1/z)\}$ are polynomials satisfying the condition of Lemma 1. The case when $h^{(i)}(z)$ and $h^{(j)}(z)$ share common reciprocal zero -1 is similarly proven.

Second, $\{h^{(i)}(z)\}$ share common reciprocal zero $c \neq \pm 1$. In this case

$$h^{(i)}(z) = g^{(i)}(z)(1 - cz^{-1})(1 - c^{-1}z^{-1}) \quad i = 1, \cdots, M. \tag{81}$$

A set of $\{v^{(i)}(1/z)\}$ satisfying the condition of Lemma 1 is given by

$$v^{(i)}(1/z) = g^{(i)}(z)(1 - z^{-1})(1 + z^{-1}). \tag{82}$$

Hence

$$\boldsymbol{S}_{ij}(\boldsymbol{h}) \begin{pmatrix} \boldsymbol{v}_i \\ \boldsymbol{v}_j \end{pmatrix} = \boldsymbol{0}, \quad \forall i \neq j. \tag{83}$$

According to (21), we have

$$\boldsymbol{S}(\boldsymbol{h}) \begin{pmatrix} \boldsymbol{v}_i \\ \vdots \\ \boldsymbol{v}_M \end{pmatrix} = \boldsymbol{0}. \tag{84}$$

Necessary Condition

First, let L_i be the order of polynomial $h^{(i)}(1/z) = h_0^{(i)} + h_1^{(i)} z + \cdots + h_{L_i}^{(i)} z^{L_i}$. It can be seen that $h^{(i)}(z) z^{L_i}$ is also a polynomial. Let

$$\tilde{h}^{(i)}\left(\frac{1}{z}\right) \triangleq \text{GCD}\{h^{(i)}\left(\frac{1}{z}\right), h^{(i)}(z) z^{L_i}\} \tag{85}$$

where GCD denotes the *greatest common divisor*. Define

$$\overline{h}^{(i)}\left(\frac{1}{z}\right) \triangleq h^{(i)}\left(\frac{1}{z}\right) / \tilde{h}^{(i)}\left(\frac{1}{z}\right). \tag{86}$$

Thus, the polynomial $h^{(i)}(1/z)$ can be decomposed into

$$h^{(i)}\left(\frac{1}{z}\right) = \overline{h}^{(i)}\left(\frac{1}{z}\right) \tilde{h}^{(i)}\left(\frac{1}{z}\right), \qquad i = 1, \cdots, M. \tag{87}$$

If $h^{(i)}(z)$ has reciprocal zero, the $\tilde{h}^{(i)}(1/z)$ is not a constant. Thus, to prove the necessary condition of the theorem, we need to show that if $\boldsymbol{S}(\boldsymbol{h})$ is singular

$$\text{GCD}\left(\tilde{h}^{(1)}\left(\frac{1}{z}\right), \cdots, \tilde{h}^{(M)}\left(\frac{1}{z}\right)\right) \neq \text{constant}. \tag{88}$$

Second, from (21), we conclude that there exist $\boldsymbol{v}_i, \cdots, \boldsymbol{v}_M$ such that

$$\boldsymbol{S}_{ij} \begin{pmatrix} \boldsymbol{v}_i \\ \boldsymbol{v}_j \end{pmatrix} = \boldsymbol{0}, \quad \forall j \neq i. \tag{89}$$

According to Lemma 1, there exist $v^{(i)}(1/z)$ and $v^{(j)}(1/z)$ satisfying (68)–(70). Multiplying both sides of (68) by $z^{L_i}/\tilde{h}^{(i)}(1/z)$, we have

$$\frac{h^{(i)}(z) z^{L_i}}{\tilde{h}^{(i)}(1/z)} v^{(i)}(z) = -\underbrace{\frac{h^{(i)}\left(\frac{1}{z}\right)}{\tilde{h}^{(i)}(1/z)}}_{\overline{h}^{(i)}\left(\frac{1}{z}\right)} z^{L_i} v^{(i)}(1/z). \tag{90}$$

By the definition of $\tilde{h}^{(i)}(1/z)$ in (85), the two polynomials $h^{(i)}(z) z^{L_i}/\tilde{h}^{(i)}(1/z)$ and $\overline{h}^{(i)}(1/z)$ are coprime. Thus, $\overline{h}^{(i)}(1/z)$ is a factor of $v^{(i)}(z)$, and we have the following decomposition:

$$v^{(i)}(z) = \overline{h}^{(i)}(1/z) \tilde{v}^{(i)}(z) \tag{91}$$

where $\tilde{v}^{(i)}(z)$ is a polynomial.

Finally, Substituting (91) and (87) into (68) and (70), after some manipulations, we have

$$z^{\tilde{L}_0} \tilde{h}^{(i)}(z) \tilde{v}^{(j)}(z) = -\tilde{h}^{(j)}(1/z) \tilde{v}^{(i)}(1/z) z^{\tilde{L}_0} \quad j = 1, \cdots, M \tag{92}$$

$$z^{\tilde{L}_i} \tilde{h}^{(i)}(z) \tilde{v}^{(i)}(z) = -\tilde{h}^{(i)}(1/z) z^{\tilde{L}_i} \tilde{v}^{(i)}(1/z) \tag{93}$$

where $\tilde{L}_0 = \max\{\tilde{L}_i\}$. From (92), $z^{\tilde{L}_0} \tilde{h}^{(i)}(z)$ is a factor of $\tilde{h}^{(j)}(1/z) \tilde{v}^{(i)}(1/z) z^{\tilde{L}_0}$ for all $j = 1, \cdots, M$. Thus, $z^{\tilde{L}_0} \tilde{h}^{(i)}(z)$ is a factor of

$$\tilde{v}^{(i)}(1/z) z^{\tilde{L}_0} \text{GCD}(\tilde{h}^{(1)}(1/z), \cdots, \tilde{h}^{(M)}(1/z)). \tag{94}$$

Now, we prove that $z^{\tilde{L}_0} \tilde{h}^{(i)}(z)$ is not a factor of $\tilde{v}^{(i)}(1/z) z^{\tilde{L}_0}$, and consequently, $\text{GCD}(\tilde{h}^{(1)}(1/z), \cdots, \tilde{h}^{(M)}(1/z))$ is not a constant. Since $\tilde{h}^{(i)}(1/z)$ in (93) only has reciprocal zeros, then $z^{\tilde{L}_i} \tilde{h}^{(i)}(z) \propto \tilde{h}^{(i)}(1/z)$. From (93), we have $\tilde{v}^{(i)}(z) \propto$

$z^{\tilde{L}_i}\tilde{v}^{(i)}(1/z)$. Thus, $z^{\tilde{L}_i}\tilde{v}^{(i)}(1/z)$ is a polynomial with order less than or equal to $\tilde{L}_i$. Therefore, if $z^{\tilde{L}_i}\tilde{h}^{(i)}(z)$ is a factor of $z^{\tilde{L}_i}\tilde{v}^{(i)}(1/z)$, then $z^{\tilde{L}_i}\tilde{h}^{(i)}(z) = \alpha z^{\tilde{L}_i}\tilde{v}^{(i)}(1/z)$, where α is a constant. Substituting this into (93), we have a contradiction $\tilde{h}^{(i)}(z)\tilde{h}^{(i)}(1/z) = -\tilde{h}^{(i)}(1/z)\tilde{h}^{(i)}(z)$. Hence, $z^{\tilde{L}_i}\tilde{h}^{(i)}(z)$ is NOT a factor of $z^{\tilde{L}_i}\tilde{v}^{(i)}(1/z)$. Consequently, $z^{\tilde{L}_0}\tilde{h}^{(i)}(z)$ is not a factor of $\tilde{v}^{(i)}(1/z)z^{\tilde{L}_0}$. □

Appendix D
Proof of Theorem 3

The LS estimator is given by

$$\begin{aligned}\hat{\boldsymbol{h}}_{LS} &= \arg\min_{\|\hat{\boldsymbol{h}}\|=1} \hat{\boldsymbol{h}}^H(\boldsymbol{T}\hat{\boldsymbol{R}}^*\boldsymbol{T}^H)\hat{\boldsymbol{h}}\\ &= \arg\min_{\|\hat{\boldsymbol{h}}\|=1} (\boldsymbol{T}^H\hat{\boldsymbol{h}})^H\hat{\boldsymbol{R}}^*(\boldsymbol{T}^H\hat{\boldsymbol{h}})\\ &= (\boldsymbol{T}^H)^{-1}\arg\min_{\|\hat{\boldsymbol{h}}\|=1} \hat{\boldsymbol{h}}^H\hat{\boldsymbol{R}}^*\hat{\boldsymbol{h}} = \boldsymbol{T}\hat{\boldsymbol{g}}^* \end{aligned} \tag{95}$$

where $\hat{\boldsymbol{g}}$ is the eigenvector corresponding to the smallest eigenvalue of $\hat{\boldsymbol{R}}$.

The SVD of the covariance matrix $\boldsymbol{R} \triangleq E\{\boldsymbol{x}(i)\boldsymbol{x}^H(i)\}$ is given by

$$\begin{aligned}\boldsymbol{R} &= \mathcal{H}(\boldsymbol{h})\mathcal{H}(\boldsymbol{h})^H + \sigma^2\boldsymbol{I}\\ &= (\boldsymbol{U}_s\ \ \boldsymbol{g})\begin{pmatrix}\lambda_1^2+\sigma^2 & & & \\ & \ddots & & \\ & & \lambda_{2L+1}^2+\sigma^2 & \\ & & & \sigma^2\end{pmatrix}(\boldsymbol{U}_s\ \ \boldsymbol{g})^H\end{aligned} \tag{96}$$

where $\lambda_{\max} = \lambda_1 > \lambda_2 > \cdots > \lambda_{2L+1} = \lambda_{\min}$ are the singular values of the $\mathcal{H}(\boldsymbol{h})$. Thus, the MSE of the estimator is given by

$$\text{MSE} = E\{\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2\} = E\{\|\boldsymbol{T}\hat{\boldsymbol{g}}^* - \boldsymbol{T}\boldsymbol{g}^*\|^2\} = E\{\|\hat{\boldsymbol{g}} - \boldsymbol{g}\|^2\}. \tag{97}$$

According to [1], the asymptotic normalized covariance of the sample eigenvector is given by

$$\begin{aligned}\lim_{N_s\to\infty} & N_s E\{(\hat{\boldsymbol{g}}-\boldsymbol{g})(\hat{\boldsymbol{g}}-\boldsymbol{g})^H\}\\ &= \sum_{k=1}^{2L+1}\frac{\sigma^2(\lambda_k^2+\sigma^2)}{(\lambda_k^2)^2}\boldsymbol{s}_k\boldsymbol{s}_k^H\end{aligned} \tag{98}$$

where $\boldsymbol{s}_k$ is the kth column vector of $\boldsymbol{U}_s$. Hence

$$\text{ANMSE}_{\text{LS/SS}} = \sum_{k=1}^{2L+1}\frac{\sigma^2(\lambda_k^2+\sigma^2)}{(\lambda_k^2)^2}. \tag{99}$$

For the lower bound, we note that

$$\begin{aligned}\text{ANMSE}_{\text{LS/SS}} &\ge \frac{\sigma^2(\lambda_{\min}^2+\sigma^2)}{(\lambda_{\min}^2)^2} \ge \frac{\sigma^2}{\lambda_{\min}^2}\\ &= (2\sigma^2)\left(\frac{\lambda_{\max}^2}{\lambda_{\min}^2}\right)\left(\frac{1}{2\lambda_{\max}^2}\right)\\ &= \frac{\kappa^2(\mathcal{H}(\boldsymbol{h}))}{\text{SNR}\ 2\lambda_{\max}^2}.\end{aligned} \tag{100}$$

Since

$$\sum_{k=1}^{2L+1}\lambda_k^2 = \text{tr}(\mathcal{H}(\boldsymbol{h})\mathcal{H}(\boldsymbol{h})^H) = L+1 \ge \lambda_{\max}^2 \tag{101}$$

we have

$$\text{ANMSE}_{\text{LS/SS}} \ge K_{\text{LS/SS}}\frac{\kappa^2(\mathcal{H}(\boldsymbol{h}))}{\text{SNR}} \tag{102}$$

where $K_{\text{LS/SS}} = 1/2(L+1)$. □

Appendix E
Proof of Theorem 4

We first derive the upper bound of ANMSE(g_{opt}). For $M = 2$

$$\begin{aligned}\text{ANMSE}(g_{\text{opt}}) &= \text{tr}\{(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1}\}\\ &= \sum_i \lambda_i((\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))^{-1})\\ &\le \frac{2(L+1)}{\lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}.\end{aligned} \tag{103}$$

Since

$$\boldsymbol{S}^T(\boldsymbol{h})\Sigma^{-1}(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}) \ge \lambda_{\min}(\Sigma^{-1}(\boldsymbol{h}))(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h})) \tag{104}$$

then

$$\begin{aligned}\text{ANMSE}(g_{\text{opt}}) &\le \frac{2(L+1)}{\lambda_{\min}(\lambda_{\min}(\Sigma^{-1}(\boldsymbol{h}))\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\\ &= \frac{2(L+1)\lambda_{\max}(\Sigma(\boldsymbol{h}))}{\lambda_{\min}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\\ &= \frac{2(L+1)\lambda_{\max}(\Sigma(\boldsymbol{h}))}{\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\kappa^2(\boldsymbol{S}(\boldsymbol{h})).\end{aligned} \tag{105}$$

Combining the lower bound of $\text{ANMSE}_{\text{LS/SS}}$ in Theorem 3, we have

$$\eta \le \frac{2(L+1)\lambda_{\max}(\Sigma(\boldsymbol{h}))}{\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\kappa^2(\boldsymbol{S}(\boldsymbol{h}))\frac{\text{SNR}}{K_{\text{LS/SS}}\kappa^2(\mathcal{H}(\boldsymbol{h}))} \tag{106}$$

$$\begin{aligned}&= \frac{2(L+1)\lambda_{\max}(\Sigma(\boldsymbol{h}))}{K_{\text{LS/SS}}\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\text{SNR}\frac{\kappa^2(\boldsymbol{S}(\boldsymbol{h}))}{\kappa^2(\mathcal{H}(\boldsymbol{h}))}\\ &\le \frac{2(L+1)\text{tr}(\Sigma(\boldsymbol{h}))}{K_{\text{LS/SS}}\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))}\text{SNR}\frac{\kappa^2(\boldsymbol{S}(\boldsymbol{h}))}{\kappa^2(\mathcal{H}(\boldsymbol{h}))}.\end{aligned} \tag{107}$$

To bound $\text{tr}\{\Sigma(\boldsymbol{h})\}$ in (107), we need expressions (58), (62)–(64) in Appendix A. Each element in $\Sigma(\boldsymbol{h})$ can be expressed in terms of σ^2 and $r_{ij}(\tau)$'s. Since $\|\boldsymbol{h}\|_2 = 1$, $r_{ij}(\tau)$'s are all bounded by some constant. Thus, $\text{tr}\{\Sigma(\boldsymbol{h})\}$ has an upper bound

$$\text{tr}\{\Sigma(\boldsymbol{h})\} \le K_1 + K_2\sigma^2 + K_3\sigma^4 \tag{108}$$

where K_1, K_2, and K_3 are some constants. Similar to (42)–(45), one can show that $\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))$ is lower bounded by a constant. Therefore

$$\frac{2(L+1)\text{tr}(\Sigma(\boldsymbol{h}))}{K_{\text{LS/SS}}\lambda_{\max}(\boldsymbol{S}^T(\boldsymbol{h})\boldsymbol{S}(\boldsymbol{h}))} \le K_\eta \tag{109}$$

where K_η is a constant depending on L and σ^2. The derivation of the exact expression of K_η is not easy and not necessary for the qualitative understanding of the asymptotic efficiency of the LS/SS estimator. Substituting K_η in (107), we have the upper bound of the efficiency for LS/SS estimator. □

ACKNOWLEDGMENT

The authors wish to thank anonymous reviewers for the constructive suggestions.

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 1984.

[2] L. A. Baccala and S. Roy, "A new blind time-domain channel identification method based on cyclostationarity," in *Proc. 26th Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1994.

[3] Z. Ding, "Blind identification and equalization using spectral correlation measurements, Pt. I: A frequency-domain approach," *Cyclostationarity in Communications and Signal Processing*, W. A. Gardner, Ed. New York: IEEE, 1993.

[4] G. B. Giannakis and S. D. Halford, "Asymptotically optimal blind fractionally-spaced channel estimation and performance analysis," submitted for publication.

[5] ——, "Performance analysis of blind equalizers based on cyclostationary statistics," in *Proc. 26th Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1994, pp. 711–716.

[6] M. I. Gürreli and C. L. Nikias, "A new eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," in *Proc. IEEE Intl. Conf. Acoust. Speech, Signal Processing*, vol. IV, Minneapolis, MN, Apr. 1993, pp. 448–451, .

[7] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind identification of multichannel FIR systems," in *Proc. 27th Asilomar Conf. Signals, Syst., Comput.*, Asilomar, CA, Oct. 1993.

[8] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace-methods for the blind identification of multichannel FIR filters," in *Proc. ICASSP'94 Conf.*, Adelaide, Australia, Apr. 1994.

[9] ——, "Subspace-methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[10] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[11] S. V. Schell, D. L. Smith, and S. Roy, "Blind channel identification using subchannel response matching," in *Proc. 26th Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1994.

[12] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multilinear prediction," in *Proc. ICASSP'94 Conf.*, Adelaide, Australia, Apr. 1994.

[13] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization of multipath channels: A frequency domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.

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

[15] ——, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, Mar. 1994.

[16] L. Tong and H. Zeng, "Blind channel identification using cyclic spectra," in *Proc. 26th Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1994, pp. 711–716.

[17] 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.

[18] H. H. Zeng and L. Tong, "Blind channel estimation using the second-order statistics: algorithms," *IEEE Trans. Signal Processing*, this issue, pp. 1919–1930.

[19] ——, "Some new results on blind channel estimation: performance and algorithms," in *Proc. 27th Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 1995.

[20] ——, "Connections between the least-squares and subspace approaches to blind channel estimation," *IEEE Trans. Signal Processing*, vol. 44, pp. 1593–1596, June 1996.

[21] H. H. Zeng, "Blind estimation: performance analysis and new algorithms," Ph.D dissertation, Univ. Connecticut, Storrs, 1997.

Hanks H. Zeng (S'95), for photograph and biography, see this issue, p. 1930.

Lang Tong (S'87–M'91), for photograph and biography, see this issue, p. 1930.