Blind Channel Estimation Using the Second-Order Statistics: Algorithms

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

***Abstract*—Most second-order moment-based blind channel estimators belong to two categories: i) optimal correlation/spectral fitting techniques and ii) eigenstructure-based techniques. These two classes of algorithms have complementary advantages and disadvantages. In this paper, a new optimization criterion referred to as the *joint optimization with subspace constraints* (JOSC) is proposed to unify the two types of approaches. Based on this criterion, a new algorithm is developed to combine the strength of the two classes of blind channel estimators. Among a number of attractive features, the JOSC algorithm does not require the accurate detection of the channel order. When compared with existing eigenstructure-based techniques, the JOSC performs better, especially when the channel is close to being unidentifiable. When compared with correlation/spectral fitting schemes, the JOSC is less affected by the presence of local minima.**

I. Introduction

ESTIMATING transmission channels is important in many communication and signal processing applications. When the input of the channel is not available for processing at the receiver, channel estimation is blind. The class of techniques that exploits either the cyclostationarity of the signal or the single-input multiple-output structure of the channel is of significant current interest. In both cases, it has been shown that the second-order statistics contain sufficient information for the identification and estimation of finite impulse response channels [4], [12], [13].

Many blind channel estimators developed recently belong to the class of moment-based estimators. Although they are not always efficient, such algorithms are often simple and provide good initial estimates. Existing moment-based blind channel estimators can be further classified into two categories: i) *optimal correlation/spectrum fitting* methods [3], [5], [14], and ii) *eigenstructure-based* algorithms [2], [6]–[8], [10]–[12]. In [15] and [16] asymptotic performance analysis of second-order moment-based blind channel estimators is presented. The analysis shows that there is a gap in performance between the best moment-based estimator [5] and some of the eigenstructure-based algorithms [6], [7]. Such a gap is significant when the channel is close to unidentifiable.

Manuscript received July 18, 1995; revised August 8, 1996. This work was supported in part by the National Science Foundation under Contract NCR-9321813, by the Office of Naval Research under Contract N00014-96-1-0895, 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)05776-0.

Existing moment-based blind channel estimators face some of the following major difficulties:

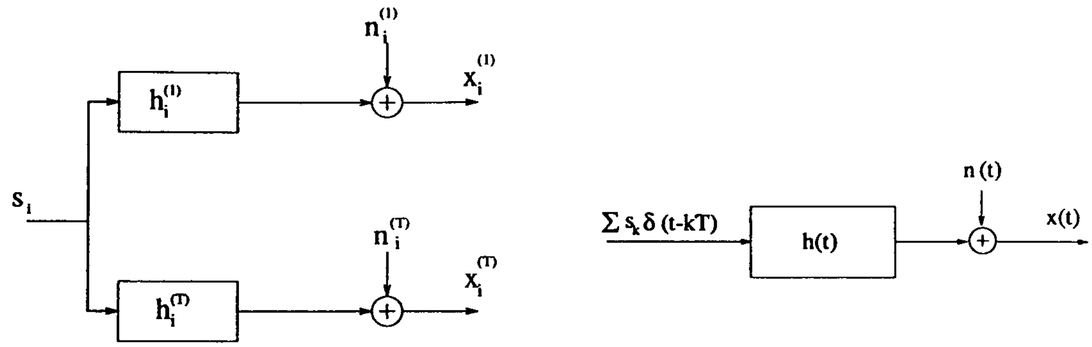
1) *Nonconvex Optimization*: Optimal correlation/spectrum fitting techniques offer superior performance when the channel is close to unidentifiable. Unfortunately, such optimization requires the search of channel parameters in a high-dimensional parameter space. The existence of local minima often renders such techniques ineffective.
2) *Ill-Conditioning of Channels*: Many eigenstructure-based techniques involve the optimization of some quadratic criteria from which the estimators can be obtained in closed form. Unfortunately, when the channel is close to unidentifiable, the performance of eigenstructure-based algorithms degrades drastically (see Section III).
3) *Channel Order Determination*: This issue has so far not been addressed adequately. Many eigenstructure-based algorithms require the accurate detection of the channel order, which is very difficult for bandlimited channels.

Our goal in this paper is to develop a unified approach to the second-order moment-based blind channel estimations. To achieve this goal, we propose a new approach referred to as the *joint optimization with subspace constraints* (JOSC), which enables us to derive most existing moment-based estimators by specifying several key parameters. Based on this criterion, a new algorithm is proposed to combine the strength, in both performance and implementation, of the two classes of moment-based estimators. The JOSC algorithm has the following attractive features:

- The JOSC does not require the accurate determination of the channel order.
- The JOSC is robust with respect to the ill conditioning of the channel.
- The JOSC involves the search of parameters in a low dimensional space.

The price paid by the JOSC is the loss of unbiasedness and some efficiency. When tested for a class of 500 two-ray multipath channels, the proposed algorithm shows considerable improvement over existing techniques.

The organization of this paper is as follows. Problem formulations are given in Section II. In Section III, we discuss the performance and limitations of the two classes of blind estimators. Our goal is to motivate the new optimization criterion that unifies the two approaches. In Section IV, the JOSC criterion is formulated, and existing blind channel estimators are classified. We present the new algorithm in



(a) (b)

Fig. 1. Two equivalent models: (a) Single-input multiple-output model. (b) Single-input single-output model.

Section V. The simulation results are presented in Section VI, where the class of two-ray multipath channels subject to random fadings is considered. We conclude in Section VII by summarizing advantages and disadvantages of the JOSC estimator.

II. Problem Formulation

A. Notations

Notations used in this paper are standard. Upper- and lower-case bold letters denote matrices and vectors, respectively. Other notations are listed as follows.

$(\cdot)^t$	Transpose.
$(\cdot)^*$	Complex conjugate.
$(\cdot)^H$	Hermitian.
$E\{\cdot\}$	Expectation operator.
$\|\|\boldsymbol{h}\|\|$	2-norm.
$\boldsymbol{I}_n n \times n$	Identity matrix.
$\lambda_{\min}(\boldsymbol{A})(\lambda_{\max}(\boldsymbol{A}))$	Minimum (maximum) singular value of $\boldsymbol{A}$.
$\kappa(\boldsymbol{A})$	Condition number of $\boldsymbol{A}$.
tr$\{\boldsymbol{A}\}$	Trace of $\boldsymbol{A}$.
C^n	n-dimensional complex vector space.
diag$(a_1, \cdots, a_n)$	Diagonal matrix with elements $\{a_1, \cdots, a_n\}$ on the main diagonal.

B. Models

Second-order moment-based blind channel estimators are derived from two equivalent models. The discrete-time single-input multiple-output (SIMO) model [see Fig. 1(a)] is given by

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

where

$\{s_i\}$	input sequence;
$\{h_i^{(j)}\}_{i=0}^L$	finite impulse response of the jth subchannel;
$n_i^{(j)}$	noise at the jth receiver;
$x_i^{(j)}$	received signal from the jth receiver.

In the matrix form, we have

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

where

$$\boldsymbol{h}_j = [h_0^{(j)}, \cdots, h_L^{(j)}]^t \tag{3}$$
$$\boldsymbol{h} = [\boldsymbol{h}_1^H, \cdots, \boldsymbol{h}_T^H]^H \in C^{T(L+1)} \tag{4}$$
$$\boldsymbol{s}(i) = [s_i, \cdots, s_{i-L-N}]^t \tag{5}$$
$$\boldsymbol{x}(i) = [x_i^{(1)}, \cdots, x_{i-N}^{(1)}, \cdots, x_i^{(T)}, \cdots, x_{i-N}^{(T)}]^t \tag{6}$$
$$\boldsymbol{n}(i) = [n_i^{(1)}, \cdots, n_{i-N}^{(1)}, \cdots, n_i^{(T)}, \cdots, n_{i-N}^{(T)}]^t \tag{7}$$

where T, N, and L are integers, and the multichannel filtering transform $\mathcal{H}(\boldsymbol{h})$ is defined by

$$\mathcal{H}(\boldsymbol{h}) = [\mathcal{F}^H(\boldsymbol{h}_1) \cdots \mathcal{F}^H(\boldsymbol{h}_T)]^H \tag{8}$$

$$\mathcal{F}(\boldsymbol{h}_i) = \begin{pmatrix} h_0^{(i)} & \cdots & h_L^{(i)} & & \\ & \ddots & & \ddots & \\ & & h_0^{(i)} & \cdots & h_L^{(i)} \end{pmatrix}. \tag{9}$$

Equivalent to the above model is the single-input single-output model [see Fig. 1(b)]

$$x(t) = \sum_{k=-\infty}^{\infty} s_k h(t - kT) + n(t) \tag{10}$$

where

$x(t)$	channel output;
s_k	information sequence;
$n(t)$	noise;
$h(t)$	unknown channel response.

The relation between the two models is given by

$$x_i^{(j)} \triangleq x(iT + (j-1)) \tag{11}$$
$$h_i^{(j)} \triangleq h(iT + (j-1)), \qquad i = 0, \cdots, L \tag{12}$$
$$n_i^{(j)} \triangleq n(iT + (j-1)) \tag{13}$$

for $j = 1, \cdots, T$. We assume throughout this paper that i) the input sequence $\{s_k\}$ is zero mean and $E\{s_k s_l^*\} = \delta(k - l)$, and ii) the noise $n(t)$ is zero-mean and white with variance σ^2.

C. Problem Formulation

We restrict our discussion in this paper to the class of blind channel estimators using only the second-order statistics. As functions of the channel vector $\boldsymbol{h}$, the *correlation vector* $\boldsymbol{r}(\boldsymbol{h})$ and the *covariance matrix* $\boldsymbol{R}(\boldsymbol{h})$ are defined, respectively, by

$$\boldsymbol{r}(\boldsymbol{h}) \triangleq (r_{ij}(\tau)), r_{ij}(\tau) \triangleq E\{x_t^{(i)}(x_{t-\tau}^{(j)})^*\} \tag{14}$$

$$\boldsymbol{R}(\boldsymbol{h}) \triangleq E\{\boldsymbol{x}(i)\boldsymbol{x}(i)^H\} \tag{15}$$

where

$$(r_{ij}(\tau)) \triangleq [\boldsymbol{r}_{11}^t, \cdots, \boldsymbol{r}_{1T}^t, \cdots, \boldsymbol{r}_{T1}^t, \cdots, \boldsymbol{r}_{TT}^t]^t \tag{16}$$

$$\boldsymbol{r}_{ij} \triangleq \begin{cases} [r_{ij}(0), \cdots, r_{ij}(-L)]^t, & i \geq j \\ [r_{ij}(1), \cdots, r_{ij}(-L)]^t, & i < j. \end{cases} \tag{17}$$

Given the *estimated correlation vector* $\hat{\boldsymbol{r}}$ or the *estimated covariance matrix* $\hat{\boldsymbol{R}}$

$$\hat{\boldsymbol{r}} \triangleq (\hat{r}_{ij}(\tau)), \hat{r}_{ij}(u) \triangleq \frac{1}{N_s} \sum_{t=1}^{N_s} x_t^{(i)} x_{t-u}^{(j)} \tag{18}$$

$$\hat{\boldsymbol{R}} \triangleq \frac{1}{N_s} \sum_{i=1}^{N_s} \{\boldsymbol{x}(i)\boldsymbol{x}(i)^H\} \tag{19}$$

the problem is to estimate $\boldsymbol{h}$ using $\hat{\boldsymbol{r}}$ or $\hat{\boldsymbol{R}}$.

III. Blind Channel Estimators

Many moment-based channel estimators belong to two different categories. The first includes those derived by matching the moments or the power spectra in some optimal way. The second exploits the eigenstructures of the second-order moments to obtain closed-form channel identifications. In this section, we discuss the performance and the limitations of these two classes of estimators by examining several key algorithms. Our goal is to motivate the joint optimization approach proposed in the next section by highlighting the differences between the two classes of algorithms. Detailds may be found in [15].

A. Achievable Performance Bound

We begin by posing the following question: *What is the best achievable performance among all consistent moment-based estimators?* To answer this question, one must show that all moment-based estimators perform no better than one particular estimator. Given the estimated second-order moment $\hat{\boldsymbol{r}}$, the performance of an estimator $\hat{\boldsymbol{h}} = g(\hat{\boldsymbol{r}})$ of the channel vector $\boldsymbol{h}_*$ can be measured by the *asymptotic normalized mean square error* (ANMSE)

$$\text{ANMSE}(g) \triangleq \lim_{N_s \to \infty} N_s E\{\|\hat{\boldsymbol{h}} - \boldsymbol{h}_*\|^2\}. \tag{20}$$

Without loss of generality, the channel is assumed to be normalized, i.e., $\|\boldsymbol{h}_*\| = 1$. The answer to the above question is given by the following theorem whose proof is given in [15].

Theorem 1: Under the assumptions that 1) $\sqrt{N_s}(\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h}_*))$ converges to a Gaussian random vector with zero mean and covariance $\Sigma(\boldsymbol{h}_*) > \boldsymbol{0}$, and 2) Jacobian $\boldsymbol{S}(\boldsymbol{h}) \triangleq \partial \boldsymbol{r} / \partial \boldsymbol{h}$ is nonsingular for all $\boldsymbol{h}$ in some open neighborhood $\mathcal{S}$ of $\boldsymbol{h}_*$, then for any estimator $\hat{\boldsymbol{h}} = g(\hat{\boldsymbol{r}})$ such that $g(\boldsymbol{r}(\boldsymbol{h})) = \boldsymbol{h}$ for all $\boldsymbol{h} \in \mathcal{S}$

$$\begin{aligned} \text{ANMSE}(g) &\geq \text{tr}\{(\boldsymbol{S}^T(\boldsymbol{h}_*)\Sigma^{-1}(\boldsymbol{h}_*)\boldsymbol{S}(\boldsymbol{h}_*))^{-1}\} \\ &\geq K\left(\frac{\kappa(\boldsymbol{S}(\boldsymbol{h}_*))}{\text{SNR}}\right)^2 \end{aligned} \tag{21}$$

where K is a constant independent of channel parameters $\text{SNR} \triangleq (1/T\sigma^2)E\{\Sigma_{j=1}^T |x_i^{(j)}|^2\}$. Moreover, there exists an optimal $g_*(\cdot)$ such that $\text{ANMSE}(g_*) = \text{tr}\{(\boldsymbol{S}^T(\boldsymbol{h}_*)\Sigma^{-1}(\boldsymbol{h}_*)\boldsymbol{S}(\boldsymbol{h}_*))^{-1}\}$.

Remarks:

1) $g_*(\cdot)$ is referred to as *asymptotic best consistent* (ABC) estimator [5].
2) The condition that the Jacobian $\boldsymbol{S}(\boldsymbol{h})$ is nonsingular is not obvious. We shall elaborate on this condition below. When $\boldsymbol{S}(\boldsymbol{h}_*)$ is nonsingular, however, the performance of all moment-based algorithms is limited by $\kappa(\boldsymbol{S}(\boldsymbol{h}_*))$.

It is natural to ask the next question: *For which kinds of channels will the moment-based methods be effective?* Specifically, when is $\boldsymbol{S}(\boldsymbol{h}_*)$ singular? The answer to this question is simple.

Theorem 2: $\boldsymbol{S}(\boldsymbol{h}_*)$ is nonsingular if and only if the subchannels $h^{(j)}(z) \triangleq \Sigma_{i=0}^L h_i^{(j)} z^{-i}$ do not share common reciprocal zeros[1].

Proof: See [15].

It is significant that the above condition is different from the identifiability condition. It is well known that the channel is identifiable from the second-order moments if and only if the subchannels do not share common zeros [12]. When the channel identifiability condition is satisfied, the bound given in (21) is achievable. On the other hand, one may ask: *What happens when the identifiability condition is not satisfied?* The answer lies in the specification of the neighborhood $\mathcal{S}$ of $\boldsymbol{h}_*$. When the identifiability condition is violated, there is a finite number of channel vectors that correspond to the same $\boldsymbol{r}(\boldsymbol{h}_*)$. In this case, one must choose $\mathcal{S}$ so that there is an estimator satisfying $g(r(\mathcal{S})) = \mathcal{S}$. Specifying the neighborhood $\mathcal{S}$ is nontrivial. It usually requires additional information about the channel. We shall address this issue later. It is important to note here that, when compared with the identifiability condition, the existence of the asymptotic best consistent estimator requires a much weaker condition: the absence of common *reciprocal* zeros.

B. Correlation/Spectrum Fitting Algorithms

In this section, we present two algorithms derived from optimal fitting of moments or power spectra. The first one is the asymptotic best consistent estimator, presented by Giannakis and Halford [5], that achieves the performance bound given by (21). This estimator, however, is not practical. We present next a suboptimal approach [14] and evaluate its performance.

[1] A zero z_0 of $h(z)$ is a reciprocal zero if $h(1/z_0) = 0$.

1) The Asymptotic Best Consistent (ABC) Estimator: Giannakis and Halford [5] are perhaps the first to derive the optimal estimator for the blind channel estimation using second-order moments. The asymptotic best consistent (ABC) estimator given in [5] is equivalent to

$$\hat{\boldsymbol{h}}_{\text{ABC}} = \arg\min_{\boldsymbol{h}} (\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h}))^H \Sigma^{-1}(\boldsymbol{h})(\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h})) \tag{22}$$

where $\boldsymbol{r}(\boldsymbol{h})$ is computed from (14). $\Sigma(\boldsymbol{h})$ can be computed in closed form. It can be shown [9, pp. 82–85] that the estimator given above is strongly consistent and the ANMSE achieves the bound given in (21).

2) Correlation Fitting (CF) Estimator: In practice, it is difficult to optimize the nonlinear objective function in (22) where it involves the computation of $\Sigma^{-1}(\boldsymbol{h})$. A suboptimal approach may be obtained by fitting correlation or the power spectrum with a fixed weighting (see [14]). For example, one simple criterion may involve fitting the autocorrelations directly:

$$\hat{\boldsymbol{h}}_{\text{CF}} = \arg\min_{\boldsymbol{h}} \|\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h})\|^2. \tag{23}$$

It is again easy to show that $\hat{\boldsymbol{h}}_{\text{CF}}$ is strongly consistent. In Appendix A, we derive the ANMSE of $\hat{\boldsymbol{h}}_{\text{CF}}$

$$\begin{aligned}\text{ANMSE}(\hat{\boldsymbol{h}}_{\text{CF}}) = &\operatorname{tr}\{(\boldsymbol{S}^T(\boldsymbol{h}_*)\boldsymbol{S}(\boldsymbol{h}_*))^{-1}\boldsymbol{S}^T(\boldsymbol{h}_*) \\ &\cdot \Sigma(\boldsymbol{h}_*)\boldsymbol{S}(\boldsymbol{h}_*)(\boldsymbol{S}^T(\boldsymbol{h}_*)\boldsymbol{S}(\boldsymbol{h}_*))^{-1}\}. \end{aligned} \tag{24}$$

Similar to the ABC estimator, the CF estimator is consistent when $\boldsymbol{S}(\boldsymbol{h}_*)$ is nonsingular.

When compared to the eigenstructure-based algorithms, an ABC estimator offers better performance [5]. The performance of the ABC and the CF estimator is determined in part by $\boldsymbol{S}(\boldsymbol{h}_*)$. Thus, they are robust to the ill conditioning of the channels. However, they also share important disadvantages. Since the optimizations in (22) and (23) are both nonlinear, achieving the global minimum is nontrivial. Such difficulties are avoided by eigenstructure-based estimators discussed next.

C. Eigenstructure-Based Algorithms

Eigenstructure-based algorithms provide closed-form identifications of the channel, which has the advantage over the correlation/spectrum fitting algorithms. The question is how much loss in performance is associated with the eigenstructure-based algorithms.

1) The Subspace and Least Squares Estimators: Since the publication of [13], a number of eigenstructure-based algorithms have been proposed that have shown promising performance in various simulation scenarios. The so-called subspace (SS) algorithm [8] and the least-squares (LS) algorithm [6] are representatives of such ideas that exploited the algebraic structure of the channel.

2) The Subspace (SS) Estimator [8]: One form of the SS estimator is given by a quadratic optimization. From (2)

$$\boldsymbol{R}(\boldsymbol{h}) \triangleq E\{\boldsymbol{x}(i)\boldsymbol{x}(i)^H\} = \mathcal{H}(\boldsymbol{h})\boldsymbol{R}_s\mathcal{H}(\boldsymbol{h})^H + \sigma^2\boldsymbol{I}. \tag{25}$$

Assuming $\boldsymbol{R}_s \triangleq E\{\boldsymbol{s}(i)\boldsymbol{s}(i)^H\} > \boldsymbol{0}$, the singular value decomposition (SVD) of $\boldsymbol{R}(\boldsymbol{h})$ has the form

$$\boldsymbol{R}(\boldsymbol{h}) = [\boldsymbol{U}_s\boldsymbol{U}_n]\begin{pmatrix}\Sigma_s & \boldsymbol{0} \\ \boldsymbol{0} & \sigma^2\boldsymbol{I}\end{pmatrix}[\boldsymbol{U}_s\boldsymbol{U}_n]^H. \tag{26}$$

Moulines *et al.* showed [8] that if the subchannels do not share common zeros, $\boldsymbol{h}$ is uniquely determined by the noise subspace $\boldsymbol{U}_n$. With the estimated covariance matrix and its corresponding estimated noise eigenmatrix $\hat{\boldsymbol{U}}_n = [\hat{\boldsymbol{u}}_{n1}, \cdots, \hat{\boldsymbol{u}}_{nl}]$, the SS estimator is given by

$$\begin{aligned}\hat{\boldsymbol{h}}_{\text{SS}} &= \arg\min_{\|\boldsymbol{h}\|=1} \boldsymbol{h}^H \underbrace{\mathcal{H}(\hat{\boldsymbol{U}}_n))\mathcal{H}(\hat{\boldsymbol{U}}_n))^H}_{\hat{\boldsymbol{Q}}_{\text{SS}}}\boldsymbol{h} \\ &= \arg\min_{\|\boldsymbol{h}\|=1} \boldsymbol{h}^H\hat{\boldsymbol{Q}}_{\text{SS}}\boldsymbol{h}\end{aligned} \tag{27}$$

where $\mathcal{H}(\hat{\boldsymbol{U}}_n) \triangleq [\mathcal{H}(\hat{\boldsymbol{u}}_{n1}), \cdots, \mathcal{H}(\hat{\boldsymbol{u}}_{nl})]$. We note here that the channel parameter vector is estimated via a quadratic optimization whose solution can be obtained in closed form by using the SVD.

3) The Least-Squares (LS) Estimator [6]: Derived in a deterministic setting, the LS estimator has the same form as that of the SS estimator. For $T = 2$, the LS estimator is given by

$$\hat{\boldsymbol{h}}_{\text{LS}} = \arg\min_{\|\boldsymbol{h}\|=1} \boldsymbol{h}^H \underbrace{\boldsymbol{T}\hat{\boldsymbol{R}}^*\boldsymbol{T}^H}_{\hat{\boldsymbol{Q}}_{\text{LS}}}\boldsymbol{h} = \arg\min_{\|\boldsymbol{h}\|=1} \boldsymbol{h}^H\hat{\boldsymbol{Q}}_{\text{LS}}\boldsymbol{h} \tag{28}$$

where $\boldsymbol{T}$ is a matrix made of blocks of identity matrices [17]. For $T > 2$, $\hat{\boldsymbol{h}}_{\text{LS}}$ has the similar quadratic form [17]. Again, the channel parameter is obtained by a quadratic optimization.

4) Performance: Although $\hat{\boldsymbol{Q}}_{\text{SS}}$ and $\hat{\boldsymbol{Q}}_{\text{LS}}$ are different, the LS and the SS estimators are identical for the important case when $T = 2$ and $N = L$[16]. In such a case, the performance of the LS and SS estimators is given by [15]

$$\text{ANMSE}_{\text{LS/SS}} = \sum_{k=1}^{2L+1} \frac{\sigma^2(\lambda_k^2 + \sigma^2)}{(\lambda_k^2)^2} \geq K_{\text{LS/SS}}\frac{\kappa^2(\mathcal{H}(\boldsymbol{h}))}{\text{SNR}} \tag{29}$$

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

We make the following observations:

1) The performance of the LS/SS estimators is limited by $\kappa(\mathcal{H}(\boldsymbol{h}_*))$. In contrast, the performance of the ABC estimator is limited by $\kappa(\boldsymbol{S}(\boldsymbol{h}_*))$.
2) When the channel is not identifiable, $\mathcal{H}(\boldsymbol{h}_*)$ is singular [12], and the LS and SS approaches fail. This is not necessarily true for the optimal correlation/spectra fitting approaches.
3) Both LS and SS estimators do not require the knowledge of source covariance as long as $\boldsymbol{R}_s > 0$. On the other hand, optimal correlation/spectra fitting approaches require the knowledge of the source statistics and take advantage of such information.

IV. The Joint Optimization with Subspace Constraints (JOSC)

In the previous section, we discussed the performance of two classes of blind channel estimators. Of particular importance is that correlation/spectra fitting schemes may still work when the channel is close to unidentifiable as long as the parameter space is restricted to the neighborhood of the true channel. The most important disadvantage of this approach is that the optimization is nonlinear. The search for the channel parameters in a high-dimensional parameter space often fails due to the existence of local minima. As shown in the previous section, the main shortcoming of eigenstructure-based algorithms is the significant performance deterioration when the channel is close to being unidentifiable. The question at this point is: *How do we combine the strength of these two classes of approaches?*

The key observation made here is that the two types of optimization criteria are not the same. Optimizing one criterion does not automatically optimize the other. Motivated by this observation, we propose a joint optimization approach that also unifies most existing moment-based blind channel estimators.

A. The New Criterion

Moment-based algorithms are based on either one of the following two optimization criteria:

$$\min_{||\boldsymbol{h}||=1} \boldsymbol{h}^H \hat{\boldsymbol{Q}} \boldsymbol{h} \tag{30}$$

$$\min_{\boldsymbol{h}} ||\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h})||_W^2. \tag{31}$$

Various eigenstructure-based algorithms are derived based on (30) with different $\hat{\boldsymbol{Q}}$'s. The class of optimal correlation/spectral fitting schemes is defined by (31) through the selection of weighting function $\boldsymbol{W}$. Our approach rests on the idea of optimizing (30) and (31) *jointly*. Specifically, the joint optimization can be defined by

$$\min_{\boldsymbol{h}\in\mathcal{S}} J_F(\boldsymbol{h}) \quad \text{subject to} \quad J_E(\boldsymbol{h}) \leq \alpha ||\boldsymbol{h}||^2 \tag{32}$$

where α is a given threshold, and

$$J_E(\boldsymbol{h}) = \boldsymbol{h}^H \hat{\boldsymbol{Q}} \boldsymbol{h} \tag{33}$$

$$J_F(\boldsymbol{h}) = (\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h}))^H \boldsymbol{W} (\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h})) \tag{34}$$

where $\boldsymbol{W}$ is some weighting matrix, and $\mathcal{S}$ is a restricted parameter set.

Remarks:

1) *What is the advantage of including $J_E(\boldsymbol{h})$?* The inclusion of $J_E(\boldsymbol{h})$ enables us to exploit various eigenstructures of the second-order statistics. Indeed, for many channels, the eigenstructure-based algorithms are effective. The constraint involving $J_E(\boldsymbol{h})$ can be satisfied by restricting the search of channel parameters in a subspace specified by the eigenstructures of $\hat{\boldsymbol{Q}}$. More discussions will be given in Section V.
2) *How do we select α?* The selection of α reflects the degree of emphasis on different aspects of the channel. When $\alpha = \lambda_{\min}(\hat{\boldsymbol{Q}})$, the optimization of the new criterion is equivalent to the optimization of $J_E(\boldsymbol{h})$. When $\alpha = \lambda_{\max}(\hat{\boldsymbol{Q}})$, $J_E(\boldsymbol{h})$ becomes ineffective, the optimization of the new criterion is equivalent to the optimization of $J_F(\boldsymbol{h})$, which leads to algorithms including the asymptotic best consistent estimator. What is interesting is the case of $\alpha \in [\lambda_{\min}(\hat{\boldsymbol{Q}}), \lambda_{\max}(\hat{\boldsymbol{Q}})]$, which enables us to tune the algorithm to different channels.
3) *How do we select $\mathcal{S}$?* A key element of the new criterion is the restriction of the parameter to $\mathcal{S}$. This comes from the fact that the asymptotic best consistent estimator is defined only in the neighborhood of the true channel parameter (see Theorem 1). The specification of $\mathcal{S}$ requires additional knowledge of the true channel. In Section V, $\mathcal{S}$ is defined by exploiting the statistical structures of the channel.

TABLE I
Classification of Moment-Based Algorithms by the JOSC Criterion. ABC: Asymptotic Best Consistent Estimator; CSF: Cyclic Spectral Fitting Estimator; CF: Correlation Fitting Estimator; LS: Least-Squares Estimator; SS: Subspace Estimator; ESRM: Extended Subchannel Response Matching Estimator; CSLS: Cyclic Spectra Least-Squares Estimator

Algorithm	W	α	$\mathcal{S}$	$\hat{\mathbf{Q}}$
ABC [5]	$\Sigma^{-1}(\mathbf{h})$	λ_{max}	$C^{T(L+1)}$	-
CSF [14]	$\mathbf{T}_{CSF}$	λ_{max}	$C^{T(L+1)}$	-
CF [14]	$\mathbf{I}$	λ_{max}	$C^{T(L+1)}$	-
LS [6]	-	λ_{min}	$C^{T(L+1)}$	$\mathbf{T}\hat{\mathbf{R}}^*\mathbf{T}^H$
SS [8]	-	λ_{min}	$C^{T(L+1)}$	$\mathcal{H}(\hat{\mathbf{U}}_n)\mathcal{H}(\hat{\mathbf{U}}_n)^H$
ESRM [10]	-	λ_{min}	$\mathcal{S}_{ESRM}$	$\mathbf{T}\hat{\mathbf{R}}^*\mathbf{T}^H$
CSLS [12]	-	λ_{min}	$C^{T(L+1)}$	$\hat{\mathbf{Q}}_{CSLS}$

In summary, the new optimization aims to achieve a compromise between the performance of the estimation and the complexities of the implementation. Although optimization of $J_F(\boldsymbol{h})$ alone may lead to the best consistent estimator when the channel is close to unidentifiable, without the constraint involving $J_E(\boldsymbol{h})$ and the specification of $\mathcal{S}$, it is difficult to achieve the global minimum in the optimization of $J_F(\boldsymbol{h})$.

B. Classification of Existing Blind Channel Estimators

The new criterion proposed in this paper provides a framework from which most existing second-order moment-based estimators can be derived by specifying different $\mathcal{S}, \boldsymbol{Q}, W$, and α. Table I lists several key algorithms proposed recently.

It is obvious that the selections of $\mathcal{S}, \boldsymbol{Q}, W$, and α lead to the ABC, LS, and SS estimators discussed in Section III. The technique of JOSC is the new algorithm presented in Section V. We present next the selections of key parameters for several other algorithms proposed in recent years.

1) Cyclic Spectra Fitting (CSF) Method [14]: The kth *cyclic spectrum* $\Gamma_x^{(k)}(\omega)$ of the output $x(\cdot)$ in (10) is the kth Fourier coefficient of the Fourier series expansion of $r_x(t, \tau)$

with respect to t, where

$$r_x(t,\tau) \triangleq E\{x(t)x^*(t-\tau)\} \tag{35}$$

$$\Gamma_x^{(k)}(\omega) \triangleq \sum_{t=0}^{T-1}\sum_{\tau} r_x(t,\tau)e^{-j\omega}e^{-jk(2\pi/T)t}. \tag{36}$$

It can be shown [12] that the z transform of $\Gamma_x^{(k)}(\omega)$ satisfies

$$\Gamma_x^{(k)}(z) = H(z)H\left(e^{-jk(2\pi/T)}\ \frac{1}{z^*}\right) + T\sigma^2\delta(k) \qquad k = 0,\cdots,T-1 \tag{37}$$

where $H(z)$ is the z transform of $h(t)$. It is easy to verify that the coefficients of $\Gamma_x^{(k)}(z)$ (cyclic correlations) are linear combinations of the correlation vector $\boldsymbol{r}(\boldsymbol{h})$ in (14), i.e., $\boldsymbol{\gamma}(\boldsymbol{h}) = \boldsymbol{A}\boldsymbol{r}(\boldsymbol{h})$, where $\boldsymbol{\gamma}(\boldsymbol{h})$ is the vector of cyclic correlations, and $\boldsymbol{A}$ is a nonsingular matrix. The cyclic spectra fitting (CSF) is obtained by the optimization

$$\begin{aligned}\hat{\boldsymbol{h}}_{\text{CSF}} &= \arg\min_{\boldsymbol{h}} \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{h})\|^2 \\ &= \arg\min_{\boldsymbol{h}} (\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h}))^H \underbrace{\boldsymbol{A}^H\boldsymbol{A}}_{W}(\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{h})).\end{aligned} \tag{38}$$

2) Cyclic Spectra Least-Squares (CSLS) Method [12]: When $\sigma^2 = 0$, it is easy to verify from (37) that

$$\Gamma_x^{(k_1)}(z)H\left(e^{jk_2(2\pi/T)}\ \frac{1}{z^*}\right) - \Gamma_x^{(k_2)}(z)H\left(e^{jk_1(2\pi/T)}\ \frac{1}{z^*}\right) = 0. \tag{39}$$

Recasting the above equation in the matrix form, we have

$$\boldsymbol{G}(k1,k2)\boldsymbol{h} = 0 \tag{40}$$

where $\boldsymbol{G}(k1,k2)$ can be obtained from $r_{ij}(\tau)$ [12]. The cyclic-spectra least-squares method is then obtained by

$$\begin{aligned}\hat{\boldsymbol{h}}_{\text{CSLS}} &= \arg\min_{\|\boldsymbol{h}\|=1} \|\hat{\boldsymbol{G}}(k1,k2)\boldsymbol{h}\|^2 \\ &= \arg\min_{\|\boldsymbol{h}\|=1} \boldsymbol{h}^H \underbrace{\hat{\boldsymbol{G}}(k1,k2)^H\hat{\boldsymbol{G}}(k1,k2)}_{\hat{\boldsymbol{Q}}_{\text{CSLS}}}\boldsymbol{h}.\end{aligned} \tag{41}$$

3) Extended Subchannel Response Matching (ESRM) Method [10]: The extended subchannel response matching (ESRM) was derived by Schell *et al*. The cost function used in the ESRM is the same as that of the LS. The difference is the specification of the parameter set $\mathcal{S}$. Perhaps the most significant contribution of this approach is the exploitation of the fact that in communication applications, the impulse response is not arbitrary. The knowledge of the waveform of the transmitted pulses should be incorporated. Suppose that the channel impulse response satisfies the model

$$h(t) = \sum_{k=1}^{K} \alpha_k p(t-k) \tag{42}$$

where $p(t)$ is the *known* impulse response of the shaping filter. Similar to the definition of $\boldsymbol{h}$, we define $\boldsymbol{p}_i$ as a channel vector for channel response $p(t-i)$. Therefore, the channel vector $\boldsymbol{h}$ is in the space

$$\mathcal{S}_{\text{ESRM}} = \text{span}(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_K). \tag{43}$$

The advantage of incorporating such information is twofold. First, it reduces the dimension of the parameter space from $C^{T(L+1)}$ to C^K. Second, it considerably improves the performance in simulations [10].

V. JOSC Algorithms

In this section, we present a new class of blind channel estimators based on the JOSC. As mentioned earlier, the optimization of the new criterion hinges upon three factors:

i) the specification of $\mathcal{S}$;
ii) the quadratic constraint involving $J_E(\boldsymbol{h})$ by specifying α;
iii) the minimization of $J_F(\boldsymbol{h})$.

We present in this section one special implementation of the JOSC criterion that exploits two subspace structures. The first subspace that defines $\mathcal{S}$ is associated with the principal component structure of the channel. The second subspace is obtained by the constraint involving $J_E(\boldsymbol{h})$. Jointly, these two subspaces form the constraint in the optimization of $J_F(\boldsymbol{h})$. Deriving the optimal estimator from the new criterion given in (32) is nontrivial. Our strategy here is to develop a suboptimal algorithm that is relatively easy to implement.

A. The Principal Component Analysis of Channel Statistics

The first element of the new algorithm involves specifying a neighborhood $\mathcal{S}$ of the channel vector. In general, describing $\mathcal{S}$ requires additional knowledge of the channel. Here, we consider an approach that specifies $\mathcal{S}$ by exploiting the principal components of the channel. This is particularly appropriate in wireless communication.

Wireless channels can be modeled by random parameters. Consider the case when the channel vector $\boldsymbol{h}$ is zero-mean with known covariance matrix

$$\boldsymbol{R}_h \triangleq \{\boldsymbol{h}\boldsymbol{h}^H\}. \tag{44}$$

Let the SVD of $\boldsymbol{R}_h$ be

$$\begin{aligned}\boldsymbol{R}_h &= \boldsymbol{V}\Lambda\boldsymbol{V}^H \\ &= [\boldsymbol{v}_1\cdots\boldsymbol{v}_m]\,\text{diag}(\lambda_1,\cdots,\lambda_m)[\boldsymbol{v}_1\cdots\boldsymbol{v}_m]^H\end{aligned} \tag{45}$$

where $m = T(L+1)$ is the dimension of channel vectors. Hence, $\boldsymbol{h}$ can be expressed as a linear combination of the orthogonal basis $\{\boldsymbol{v}_i\}$ through the principal component vector $\boldsymbol{g} = [g_1,\cdots,g_m]^T$

$$\boldsymbol{h} = \boldsymbol{V}\boldsymbol{g}. \tag{46}$$

One advantage of the principal components analysis is that the best approximation (in the sense of minimum variance) of $\boldsymbol{h}$ by any d-dimensional vector is achieved by taking the linear combination of the first d eigenvectors. The variance of the approximation error is given by $\sigma_e = \Sigma_{i=d+1}^m \lambda_i$. When σ_e is small with some d less than the dimension of $\boldsymbol{h}$, one can approximate $\boldsymbol{h}$ by the first d principal components. The important observation at this point is that the dimension of the channel vector is replaced by d, which can be determined off line as long as $\boldsymbol{R}_h$ is either known or can be estimated. To illustrate this idea, we consider the case of a multiray fading channel.

1) The Principal Component Structure of Multi-ray Fading Channels: A multiray channel model is described by

$$\begin{aligned} h(t) = & a_0 p(t-t_0) + a_1 p(t-t_0-\tau_1) \\ & + \cdots + a_n p(t-t_0-\tau_n) \end{aligned} \tag{47}$$

where

- $p(t)$ *known* shaping filter, which is a continuous function lasting L_pT (integer T is the symbol interval);
- $\{a_i\}$ independent zero-mean complex Gaussian variables;
- t_0 timing uncertainty assumed to be uniformly distributed in $[-(T/2), T/2]$;

and the delays $\{\tau_i\}$ are independent and uniformly distributed in $[0, \tau_{\max}]$ ($\tau_{\max}$ is the maximum delay).

For example, when $p(t)$ is the raised-cosine waveform with roll-off factor 0.25, $L_p = 6, T = 2, \tau_{\max} = 1.5T, n = 2$, the maximum length of $h(t)$ is 16, or the maximum order of channels is 7. The corresponding channel vector $\boldsymbol{h}$ is a 16-dimensional complex vector obtained according to (12). The covariance matrix $\boldsymbol{R}_h$ of $\boldsymbol{h}$ is given by taking expectation over random variables $\{\tau_k\}, \{\alpha_k\}$, and t_0.

$$\boldsymbol{R}_h = E_{\{\tau_k\},\{\alpha_k\},t_0}\{\boldsymbol{h}\boldsymbol{h}^H\}. \tag{48}$$

Computing $\boldsymbol{R}_h$ analytically is difficult and hardly necessary. Using Monte Carlo techniques, $\boldsymbol{R}_h$ can be estimated as accurately as possible. The principal component analysis shows that with only five principal components, i.e., $d = 5$, the relative error is given by

$$\varepsilon \triangleq \sqrt{\frac{\sigma_e}{E\{\|\boldsymbol{h}\|^2\}}} = \sqrt{\frac{\sum_{i=d+1}^{m} \lambda_i}{\sum_{i=1}^{m} \lambda_i}} = 0.0133. \tag{49}$$

With such a small approximation error, it is reasonable to define the parameter subspace $\mathcal{S} = \text{span}\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_d\}$ and to assume that $\boldsymbol{h} = \boldsymbol{B}\boldsymbol{g} \in \mathcal{S}$, where $\boldsymbol{B} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_d]$ and $\boldsymbol{g} = [g_1, \cdots, g_d]^t$. Therefore, the optimization (32) becomes

$$\min_{\boldsymbol{g} \in C^d} J_F(\boldsymbol{B}\boldsymbol{g}) \quad \text{subject to} \quad \boldsymbol{g}^H \underbrace{\boldsymbol{B}^H \hat{\boldsymbol{Q}} \boldsymbol{B}}_{\hat{\boldsymbol{\Phi}}} \boldsymbol{g} \le \alpha \|\boldsymbol{g}\|^2. \tag{50}$$

One of the advantages of the principal component analysis is the reduction of complexity in optimization. In the above example, instead of performing a search in 16-dimensional space, the search of five-dimensional space is sufficient with approximation error known *a priori*. As a result of this restriction in parameter space, the ill effects of local minima and singularity are reduced.

B. The Quadratic Constraint

We make a heuristic argument to further simplify the optimization in (50). To incorporate the quadratic constraint, the JOSC searches the solution in a special subspace that contains most information about $\boldsymbol{g}_*$. This subspace is obtained from the eigenstructure of $\hat{\Phi}$.

Ideally, $\boldsymbol{g}_*$ is in the noise subspace of $\boldsymbol{\Phi} \triangleq \boldsymbol{B}^H \boldsymbol{Q} \boldsymbol{B}$. However, the noise subspace is often not easy to obtain accurately when $\boldsymbol{Q}$ is estimated using a finite number of samples. This is especially the case when the channel is close to unidentifiable and the smallest several eigenvalues of $\hat{\boldsymbol{\Phi}}$ are clustered. In such a case, the eigenvector associated with the smallest eigenvalue can no longer be used to form a reliable channel estimator. Indeed, most eigenstructure-based schemes use only this eigenvector to form the channel estimate, which is the main cause of performance degradation.

In contrast, the JOSC uses the *extended noise subspace* that includes additional eigenvectors of $\hat{\Phi}$ associated with several of the the smallest eigenvalues. The idea is that although the estimate of the eigenvector associated with the smallest eigenvalue may have large perturbation when there is a small separation between the last several eigenvalues, the perturbation of the *subspace* spanned by the last several eigenvectors can be made small. Specifically, the *extended noise subspace* is defined by the singular vectors $\hat{\boldsymbol{u}}_i, i = 1, \cdots, d$ of $\hat{\Phi}$

$$\begin{aligned} \mathcal{S}_n & \triangleq \text{range}(\boldsymbol{C}), \\ \boldsymbol{C} & \triangleq [\hat{\boldsymbol{u}}_{d-k+1}, \cdots, \hat{\boldsymbol{u}}_d], \qquad 1 \le k \le d. \end{aligned} \tag{51}$$

To satisfy the constraint $J_E(\boldsymbol{B}\boldsymbol{g}) \le \alpha$, the parameter k in (51) satisfies

$$\hat{\lambda}_{d-k+1} \le \alpha. \tag{52}$$

Asymptotically, the extended noise subspace $\mathcal{S}_n$ contains the true channel vector. For example, when channel matrix is column rank deficient by k', there are k' repeated smallest singular values, and the true channel vector is in the range of the corresponding singular vectors. From (52) and $\alpha \ge \lambda_{\min}$, the extended noise subspace $\mathcal{S}_n$ will include all these singular vectors. Thus, the true channel is in $\mathcal{S}_n$. For finite data case, the constraint α is chosen so that there is a good separation between $\hat{\lambda}_{d-k}$ and $\hat{\lambda}_{d-k+1}$. In summary, as a suboptimal approach to the optimization proposed in Section IV, a JOSC algorithm optimizes $J_F(\boldsymbol{h})$ subject to two subspace constraints derived from the principal component analysis and the noise subspace of data covariance. The JOSC estimator is given by

$$\hat{\boldsymbol{h}}_{\text{JOSC}} = \boldsymbol{B}\boldsymbol{C} \arg \min_{\boldsymbol{f} \in C^k} \|\hat{\boldsymbol{r}} - \boldsymbol{r}(\boldsymbol{B}\boldsymbol{C}\boldsymbol{f})\|^2. \tag{53}$$

Remarks:

- When noise and input are Gaussian signals, it can be shown that the extended noise subspace contains the most information about the channel. See Appendix B.
- Since the JOSC algorithm (53) imposes two subspace constraints, there is an approximation error that makes the estimator biased. Although the exact bias cannot be

The JOSC Algorithm

Channel Subspace (Off-line)

1. From the channel model, obtain the channel covariance $\mathbf{R}_h$.

2. Compute the SVD of $\mathbf{R}_h$

$$\mathbf{R}_h = \mathbf{V}\Lambda\mathbf{V}^H = [\mathbf{v}_1 \cdots \mathbf{v}_m] diag(\lambda_1, \cdots, \lambda_m)[\mathbf{v}_1 \cdots \mathbf{v}_m]^H.$$

3. Obtain the principal component subspace $\mathcal{S}$ from $\mathbf{R}_h$: given the relative error ε, the dimension d is chosen according to

$$\sqrt{\frac{\sum_{i=d+1}^{m} \lambda_i}{\sum_{i=1}^{m} \lambda_i}} \leq \varepsilon.$$

Let $\mathbf{B} = [\mathbf{v}_1 \cdots \mathbf{v}_d]$, $\mathcal{S}$ is given by

$$\mathcal{S} = range(\mathbf{B}) = span\{\mathbf{v}_1, \cdots, \mathbf{v}_d\}.$$

Channel Estimation (On-line)

1. Estimate the covariance matrix by

$$\hat{\mathbf{R}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \{\mathbf{x}(i)\mathbf{x}(i)^H\}.$$

2. Obtain $\hat{\mathbf{\Phi}}$ from the transform:

$$\hat{\mathbf{\Phi}} = \mathbf{B}^H \mathbf{T} \hat{\mathbf{R}}^* \mathbf{T}^H \mathbf{B}.$$

3. Compute the SVD of $\hat{\mathbf{\Phi}}$

$$\hat{\mathbf{\Phi}} = \hat{\mathbf{U}}\hat{\Lambda}\hat{\mathbf{U}}^H = [\hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_d] diag(\hat{\lambda}_1, \cdots, \hat{\lambda}_d)[\hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_d]^H.$$

4. Obtain the extended noise subspace $\mathcal{S}_n$ from $\hat{\mathbf{U}}$: given the threshold level α, the dimension k of $\mathcal{S}_n$ is chosen according to

$$\hat{\lambda}_{d-k+1} \leq \alpha.$$

Fig. 2. JOSC algorithm.

controlled, it can be reduced by increasing the dimensionality of the two subspaces. A tradeoff needs to be made in practice between the bias of the estimator and the complexity of the optimization.

C. Algorithm Implementation

The JOSC is summarized below and illustrated in Fig. 2. The algorithm involves the specification of channel parameter subspace by the analysis of the principal components, performed off line, and the on-line channel estimation. In the following, we discuss some implementation issues.

1) Dimension of $\mathcal{S}$: The dimension of $\mathcal{S}$ depends on the acceptable approximation error ε. This approximation results in a biased estimation. There is a tradeoff between the bias and the high dimensionality of parameter space. The latter may cause the existence of excessive local minima.

2) Threshold α: The parameter α is used to tune the algorithm to different types of channels. The choice of α or, equivalently, the choice of k in (52), depends on the singular values $\{\hat{\lambda}_1, \cdots, \hat{\lambda}_d\}$ of $\hat{\boldsymbol{\Phi}}$. When the smallest several singular values are clustered, the eigenstructure-based schemes do not perform well. The JOSC uses the extended noise subspace. On the other hand, if there is a good separation between $\hat{\lambda}_d$ and $\hat{\lambda}_{d-1}$, the eigenstructure-based methods have good performance. With $k = 1$, the minimization of $J_F(\boldsymbol{h})$ is inactive.

3) Channel Order: The JOSC does not require the accurate detection of the channel order. To obtain the channel subspace (off line), one needs the maximum length of a class channels. In channel estimation (on line), the dimensionality k of the parameter space is determined by the analysis of the quadratic constraint.

4) Computation Complexity: We briefly derive the computation complexity for the JOSC algorithm in Fig. 2. Only the on-line part is considered. If JOSC uses $\hat{\boldsymbol{Q}}_{\mathrm{LS}}$ in its quadratic cost, then $N = L$. Thus, $\boldsymbol{x}(i)$ is a $T(L+1) \times 1$ vector, and $\boldsymbol{T}^H\boldsymbol{B}$ is a $T(L+1) \times d$ matrix. The first and second steps in Fig. 2 are matrix multiplication and additions, which require $\mathcal{O}(N_sT^2(L+1)^2)$ and $\mathcal{O}(dT^2(L+1)^2)$ flops, respectively. The third step involves the SVD of a $d \times d$ matrix, which has $\mathcal{O}(d^3)$ operations. For each iteration of the gradient search, it needs the computation of the correlation function $\boldsymbol{r}(\boldsymbol{h})$, the error vector and cost, the derivative of $\partial J_F / \partial \boldsymbol{h}$, and the gradient update as

$$r_{ij}(m) = \sum_k h_i(k)h_j^*(k-m) + \sigma^2\delta_{ij}\delta_m \quad i,j = 1, \cdots, T, \quad m = 0, \cdots, L \tag{54}$$

$$e_{i,j}(m) = r_{ij}(m) - \hat{r}_{ij}(m)$$

$$J_F(\boldsymbol{h}) = \sum_{i,j,m} |e_{i,j}(m)|^2 \tag{55}$$

$$\frac{\partial J_F}{\partial h_i^*(l)} = \sum_{j,m} e_{i,j}(m)h_j(l-m) + \sum_m e_{i,i}(m)^* h_i(l+m) \quad i = 1, \cdots, T, \quad l = 0, \cdots, L \tag{56}$$

$$\nabla J_F = (\boldsymbol{BC})^H \left(\frac{\partial J_F}{\partial \boldsymbol{f}^*}\right)^T$$

$$\hat{\boldsymbol{f}}(n+1) = \hat{\boldsymbol{f}}(n) - \mu \nabla J_F. \tag{57}$$

These four steps require $\mathcal{O}(T^2(L+1)), \mathcal{O}(T^2(L+1)), \mathcal{O}(T^2(L+1)^2)$, and $\mathcal{O}(kdT(L+1))$ flops, respectively. Thus, the total flops for one iteration is $\mathcal{O}(T^2(L+1)^2)$ flops. In summary, suppose the JOSC needs N_t iterations, the overall computation load of the JOSC is then given by

$$\begin{aligned} &\mathcal{O}(N_sT^2(L+1)^2) + \mathcal{O}(dT^2(L+1)^2) \\ &\quad + \mathcal{O}(d^3) + \mathcal{O}(N_tT^2(L+1)) \\ &= (N_s + d + N_t)\mathcal{O}(T^2(L+1)^2). \end{aligned} \tag{58}$$

For the SS and the LS methods, a SVD of $\boldsymbol{T}\hat{\boldsymbol{R}}^*\boldsymbol{T}^H$ is used to determine the channel order. The computation of this SVD is the main load for the LS/SS methods, and its complexity is order $T^3(L+1)^3$.

VI. Simulation Results

A. Performance Measure

We evaluated the performance of the JOSC algorithm by Monte Carlo simulation using two examples. For N_m independent trials, the normalized root mean square error (NRMSE) was defined by

$$\mathrm{NRMSE}^2 \triangleq \sqrt{\frac{1}{N_m\|\boldsymbol{h}_*\|^2} \sum_{m=1}^{N_m} \|\hat{\boldsymbol{h}}^{(m)} - \boldsymbol{h}_*\|^2} \tag{59}$$

where $\hat{\boldsymbol{h}}^{(m)}$ was the estimated channel from the mth trial. The signal-to-noise ratio (SNR) was defined and given by

$$\mathrm{SNR} \triangleq \frac{1}{T\sigma^2} E\left\{\sum_{j=1}^{T} |x_i^{(j)}|^2\right\} \tag{60}$$

where σ^2 was the noise variance.

B. Effects of Channel Condition Number

The performance of the JOSC and the LS/SS estimators was compared for the class of two-ray multipath channels

$$h(t) = \alpha_1 p(t) + \alpha_2 p(t - \tau) \tag{61}$$

where $p(t)$ was the continuous-time raised-cosine function with roll-off factor 0.1 and finite support of $5T$ ($T = 2$ was the symbol interval), and $\tau \in [0, 2T)$. The corresponding channel vector $\boldsymbol{h}$ was obtained from (12). In order to reduce the effect of approximation, we chose a small $\varepsilon = 1.25 \times 10^{-4}$. Accordingly, $\mathcal{S}$ was spanned by the 12 vectors obtained from the principal components analysis. With fixed strength $\alpha_1 = 1, \alpha_2 = 0.7$ and varying delay τ, a set of channels was used to evaluate the performance. The input signal was an i.i.d. sequence of $\{\pm 1\}$.

The JOSC estimator assumes only the knowledge of the *maximum* channel order. The LS estimator, on the other hand, requires an accurate detection of the actual channel order. In this simulation, we assume that the actual channel order is known to the LS estimator (the actual performance of the LS/SS estimator is worse than the one shown in this example).

Fig. 3 shows the comparison results. The JOSC algorithm generally performs better than the LS method. The peak of $\mathrm{NRMSE}_{\mathrm{LS}}$ at $\tau = 1T$ is caused by the violation of the identifiability condition. Fig. 4 shows the condition number of the channel matrix $\mathcal{H}(\boldsymbol{h})$. It is evident that the performance of the LS estimator correlates strongly with $\kappa(\mathcal{H}(\boldsymbol{h}))$. The JOSC estimator shows considerable improvement near $\tau = 1T$. The smaller peak of $\mathrm{NRMSE}_{\mathrm{JOSC}}$ is due to the fact that the restriction of the channel parameter space may not always be sufficient to exclude multiple solutions.

[2] The inherent ambiguity was removed before the computation of NRMSE.

TABLE II
EFFECTS OF THE THRESHOLD α. SNR = 25 dB, $N_s = 100, N_m = 50$

	LS	ESRM	JOSC($\alpha = \hat{\lambda}_{min}$)	JOSC($\alpha = 3\hat{\lambda}_{min}$)	JOSC($\alpha = 10\hat{\lambda}_{min}$)
NRMSE	0.8105	0.6496	0.6559	0.2112	0.4659

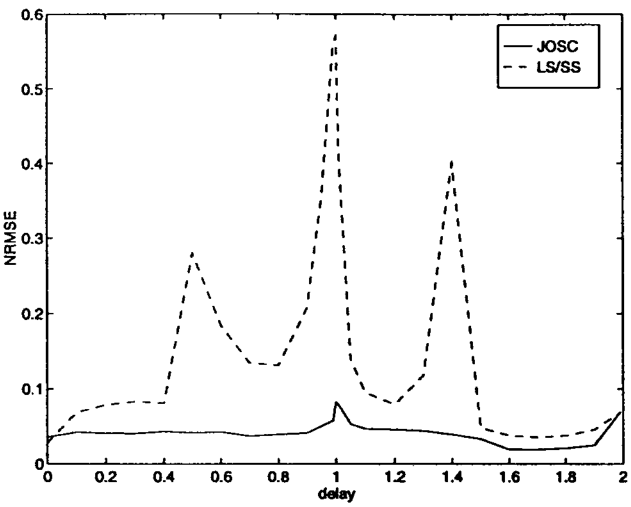


Fig. 3. Two-ray channels: SNR = 35 dB, $N_s = 1000, N_m = 50$.

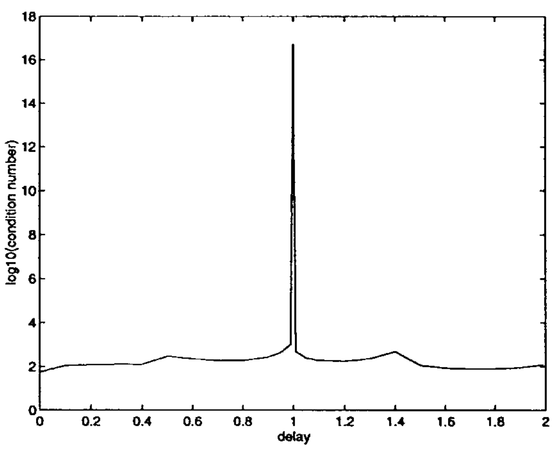


Fig. 4. Two-ray channels: condition numbers of channel matrices.

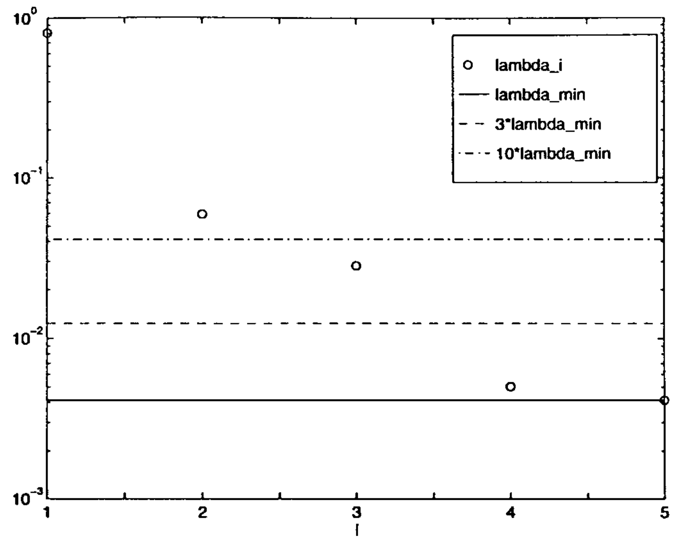


Fig. 5. Singular values and thresholds.

C. Effects of the Threshold

An important parameter in the JOSC algorithm is the threshold α, which determines the subspace for the correlation fitting. When α is small, the dimension of the extended noise subspace is reduced. The correlation fitting search is more restricted. Thus, the eigenstructure-based cost J_E is more effective in the joint optimization. In implementation, the choice of α is according to the singular value distribution as discussed in the previous section.

In this simulation, we study the effects of α. The test channel is two ray with delays $0.13T, 0.89T$. The transmitted signal is 8-PSK, SNR $= 25$ dB, and $N_s = 100$. Fig. 5 shows the singular value distribution and thresholds for a Monte Carlo trial. We choose α as 1, 3, and 10 times $\hat{\lambda}_{\min}$. The corresponding dimensions k are 1, 2, and 3, respectively. Table II shows the NRMSE results for 50 Monte Carlo trials. It appears that $\alpha = 3\hat{\lambda}_{\min}$ $(k = 2)$ is the best. This is consistent with the observation (Fig. 5) that the last two singular values are clustered. For LS, ESRM, and JOSC $(k = 1)$ methods, one singular vector is used to determine the channel; thus, they have large errors.

D. Performance of A Class Multipath Channels

In this experiment, we examined the performance of the JOSC for the channel model given by (47). Five vectors from the principal components analysis were used to construct the $\mathcal{S}$ with relative error $\varepsilon = 0.0133$. In the simulation, 500 randomly generated channels were tested. The performance comparison was made against the LS/SS algorithm [6].

The cumulative percentage of channels versus NRMSE are shown in Figs. 6–8. For most channels, the JOSC algorithm performs better than the LS method. Increasing the number of symbols from 100 to 500, the performance of the JOSC improves as shown in Fig. 9, whereas the LS method does not show visible improvement.

VII. CONCLUSION

In this paper, we studied the two classes of existing moment-based blind channel estimation algorithms. A new optimization criterion, referred to as the joint optimization with subspace constraints (JOSC), was established to form a common framework from which existing algorithms can be derived. The JOSC algorithm aims to combine the strength of the two types of approaches. Through correlation fitting, the JOSC has better performance than the eigenstructure-based methods,

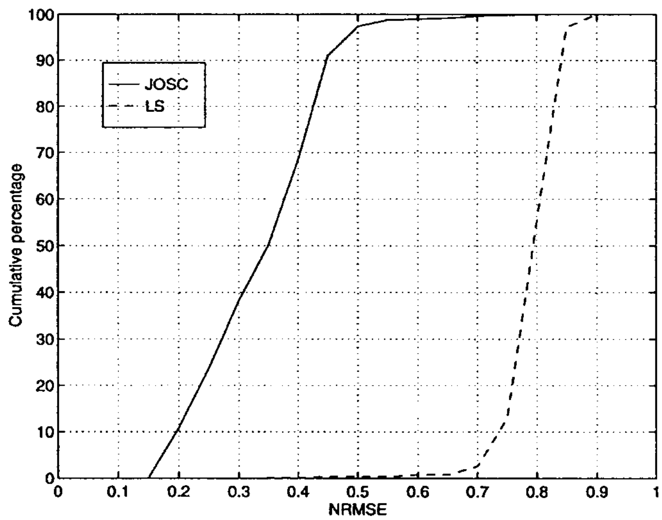


Fig. 6. 500 channels: SNR = 25 dB, $N_s = 100, N_m = 50$.

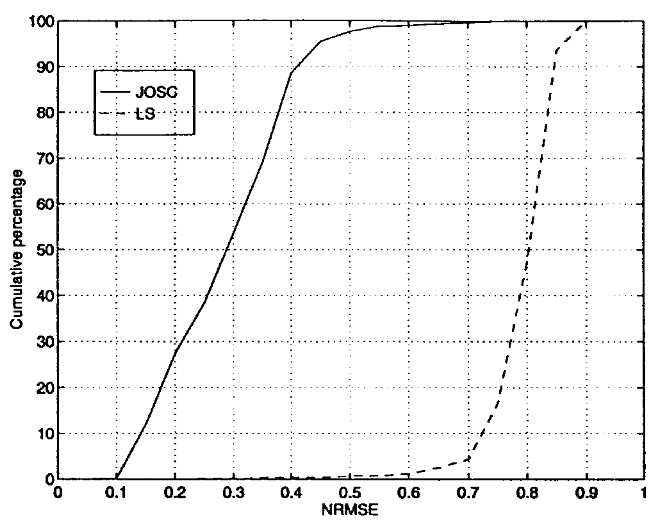


Fig. 7. 500 channels: SNR = 25 dB, $N_s = 200, N_m = 50$.

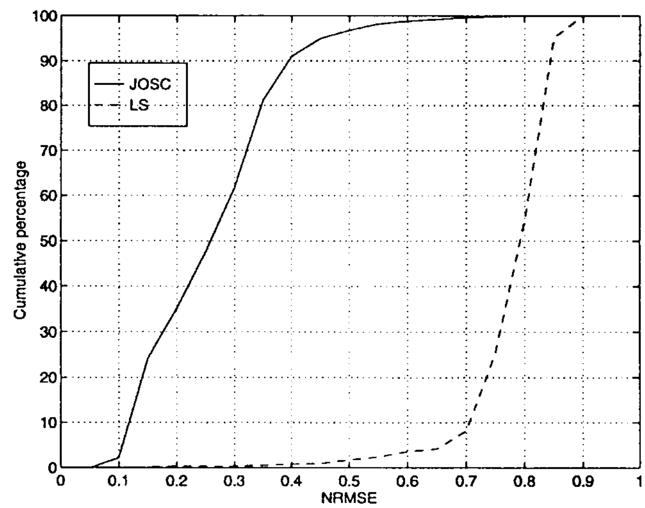


Fig. 8. 500 channels: SNR = 25 dB, $N_s = 500, N_m = 50$.

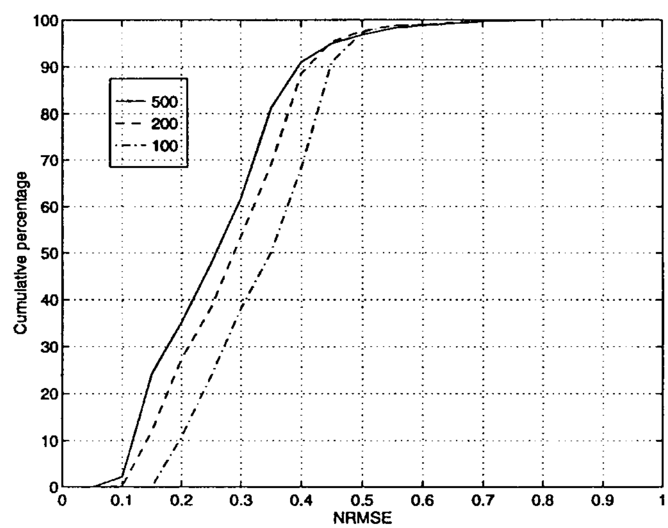


Fig. 9. Performance of the JOSC for 500 channels: SNR = 25 dB, $N_s = 100, 200, 500, N_m = 50$.

especially when the channel is ill conditioned. By exploiting both the statistical and algebraic structure of the channel, the JOSC searches the channel vector in the subspace containing the most information about the channel, which reduces the ill effects of the local minima. By selecting parameter α, the JOSC can be tuned to different channels. One disadvantage of the JOSC is the biasedness of the estimator. Although the simulations demonstrate the superior performance of the JOSC, the performance analysis of this algorithm needs to be investigated in the future.

APPENDIX A
ANMSE FOR CORRELATION FITTING (CF) ESTIMATOR

Consider a cost function of $\boldsymbol{h}$ and $\hat{\boldsymbol{r}}$:

$$J(\boldsymbol{h}, \hat{\boldsymbol{r}}) = (\boldsymbol{r}(\boldsymbol{h}) - \hat{\boldsymbol{r}})^T (\boldsymbol{r}(\boldsymbol{h}) - \hat{\boldsymbol{r}}). \tag{62}$$

The CF estimator is given by

$$\hat{\boldsymbol{h}}_{\text{CF}} = \arg \min_{\boldsymbol{h}} J(\boldsymbol{h}, \hat{\boldsymbol{r}}) \triangleq g(\hat{\boldsymbol{r}}). \tag{63}$$

Under certain mild conditions [9, Th. 3.16], the asymptotic normalized covariance of $\hat{\boldsymbol{h}}_{\text{CF}}$ is given by

$$\mathcal{A} \operatorname{cov}(\hat{\boldsymbol{h}}_{\text{CF}}) \triangleq \lim_{N_s \to \infty} N_s \operatorname{cov}(\hat{\boldsymbol{h}}_{\text{CF}}) = G(\boldsymbol{r}_*) \Sigma(\boldsymbol{h}_*) G(\boldsymbol{r}_*)^T \tag{64}$$

where $\Sigma(\boldsymbol{h}_*)$ is the asymptotic normalized covariance of $\hat{\boldsymbol{r}}$, $G(\hat{\boldsymbol{r}})$ is the Jacobian of $g(\hat{\boldsymbol{r}})$ with respect to $\hat{\boldsymbol{r}}$, and $\boldsymbol{r}_* = \boldsymbol{r}(\boldsymbol{h}_*)$.

Now, we compute Jacobian $G(\boldsymbol{r}_*)$. Note that the nonlinear cost function $J(\boldsymbol{h}, \hat{\boldsymbol{r}})$ satisfies the regularity conditions [9, Lemma 3.2], and Jacobian $G(\boldsymbol{r}_*)$ is given by

$$G(\boldsymbol{r}_*) = -\left(\frac{\partial^2 J(\boldsymbol{h}_*, \boldsymbol{r}_*)}{\partial \boldsymbol{h}^2}\right)^{-1} \frac{\partial^2 J(\boldsymbol{h}_*, \boldsymbol{r}_*)}{\partial \boldsymbol{h} \partial \hat{\boldsymbol{r}}}. \tag{65}$$

Since

$$\frac{\partial J(\boldsymbol{h}, \hat{\boldsymbol{r}})}{\partial \boldsymbol{h}} = \left(\frac{\partial \boldsymbol{r}(\boldsymbol{h})}{\partial \boldsymbol{h}}\right)^T (\boldsymbol{r}(\boldsymbol{h}) - \hat{\boldsymbol{r}}) = \boldsymbol{S}(\boldsymbol{h})^T (\boldsymbol{r}(\boldsymbol{h}) - \hat{\boldsymbol{r}}) \tag{66}$$

where $\boldsymbol{S}(\boldsymbol{h})$ is the Jacobian of $\boldsymbol{r}(\boldsymbol{h})$, and thus

$$\frac{\partial^2 J(\boldsymbol{h}_*, \boldsymbol{r}_*)}{\partial \boldsymbol{h}^2} = \boldsymbol{S}(\boldsymbol{h}_*)^T \boldsymbol{S}(\boldsymbol{h}_*) \tag{67}$$

$$\frac{\partial^2 J(\boldsymbol{h}_*, \boldsymbol{r}_*)}{\partial \boldsymbol{h} \partial \hat{\boldsymbol{r}}} = \boldsymbol{S}(\boldsymbol{h}_*)^T. \tag{68}$$

Substituting the above into (65) and (64), we have

$$\begin{aligned} \text{ANMSE}_{\text{CF}} &= \mathcal{A} \operatorname{cov}(\hat{h}_{\text{CF}}) \\ &= (\boldsymbol{S}^T(\boldsymbol{h}_*)\boldsymbol{S}(\boldsymbol{h}_*))^{-1} \boldsymbol{S}^T(\boldsymbol{h}_*) \Sigma(\boldsymbol{h}_*) \boldsymbol{S}(\boldsymbol{h}_*) \\ &\quad \cdot (\boldsymbol{S}^T(\boldsymbol{h}_*)\boldsymbol{S}(\boldsymbol{h}_*))^{-1}. \end{aligned} \tag{69}$$

Hence, we complete the proof of (24). □

APPENDIX B
PROPERTIES OF THE EXTENDED NOISE SUBSPACE

We show that the extended noise subspace contains the most information about the channel under Gaussian assumption. Specifically, assume that the noise is independent Gaussian and that the input signals $\{s_i\}$ are jointly independent Gaussian and independent of the noise. To study the information property of the extended noise subspace, let us consider the SVD of $\hat{\boldsymbol{\Phi}}$

$$\hat{\boldsymbol{\Phi}} = [\hat{\boldsymbol{u}}_1, \cdots, \hat{\boldsymbol{u}}_d] \operatorname{diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_d)[\hat{\boldsymbol{u}}_1, \cdots, \hat{\boldsymbol{u}}_d]^H. \tag{70}$$

Note that $\boldsymbol{g}_* = \boldsymbol{u}_d$ ($\boldsymbol{u}_i$'s eigenvectors of $\boldsymbol{\Phi}$) can also be represented by the orthogonal basis $\{\hat{\boldsymbol{u}}_i\}_1^d$, i.e.,

$$\boldsymbol{g}_* = \boldsymbol{u}_d = \sum_{i=1}^{d} \alpha_i \hat{\boldsymbol{u}}_i \tag{71}$$

where $\alpha_i = \hat{\boldsymbol{u}}_i^H \boldsymbol{g}_*$ is the projection coefficient. Observe that not all α_i's carry the same amount of information about $\boldsymbol{g}_*$. One can argue, by evaluating the variance of α_i, that α_i carries more information than α_j when $i > j$. According to the asymptotic analysis in [1]

$$\begin{aligned} \mathcal{A} \operatorname{var}(\alpha_i) &\triangleq \lim_{N_s \to \infty} N_s \operatorname{var}(\alpha_i) \\ &= \mathcal{A} \operatorname{var}(\hat{\boldsymbol{u}}_i^H \boldsymbol{g}_*) = \boldsymbol{g}_*^H \mathcal{A} \operatorname{cov}(\hat{\boldsymbol{u}}_i) \boldsymbol{g}_* &(72) \\ &= \boldsymbol{g}_*^H \sum_{k=1, k\neq i}^{d} \frac{\lambda_k \lambda_i}{(\lambda_k - \lambda_i)^2} \boldsymbol{u}_k \boldsymbol{u}_k^H \boldsymbol{g}_* &(73) \\ &= \frac{\lambda_d \lambda_i}{(\lambda_d - \lambda_i)^2}, \qquad i \neq d &(74) \end{aligned}$$

where λ_i's are eigenvalues. Note that $\mathcal{A} \operatorname{var}(\alpha_i)$ decreases as λ_i increases. This implies that the extended noise subspace $\mathcal{S}_n$ defined in (51) contains the most information about the channel. □

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] W. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Mag.*, vol. 8, pp. 14–36, Apr. 1991.

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

[6] 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.*, Oct. 1993.

[7] 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.

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

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

[10] 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.

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

[12] 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.

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

[14] 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.

[15] H. Zeng and L. Tong, "Blind channel estimation using the second-order statistics: asymptotic performance analysis," *IEEE Trans. Signal Processing*, this issue, pp. 2060–2071.

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

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

Hanks H. Zeng (S'95) was born in Beijing, China, in 1965. He received the B.E. degree in electrical engineering and the B.S. degree in applied mathematics from Tsinghua University, Beijing, China, in 1989 and the M.S. degree in acoustics from the Chinese Academy of Science, Beijing, in 1992. He is currently a Ph.D. candidate at the University of Connecticut, Storrs.

His research interests include estimation theory, (blind) equalization, and performance analysis.

Lang Tong (S'87–M'91) received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame, Notre Dame, IN, where he received the *Best Graduating Student Award*.

After being a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, Stanford, CA, he joined the Department of Electrical and Computer Engineering, West Virginia University, Morgantown. Since 1993, he has been with the Department of Electrical and Systems Engineering, The University of Connecticut, Storrs. He also held a Visiting Assistant Professor position at Stanford University in the summer of 1992. His research interests include statistical signal processing, wireless communication, and system theory.

Dr. Tong received *Young Investigator Award* from the Office of Naval Research in 1996 and the *Outstanding Young Author Award* from the IEEE Circuits and Systems Society.