Joint Order Detection and Blind Channel Estimation by Least Squares Smoothing

Lang Tong, *Member, IEEE*, and Qing Zhao

***Abstract*— A joint order detection and blind estimation algorithm for single input multiple output channels is proposed. By exploiting the isomorphic relation between the channel input and output subspaces, it is shown that the channel order and channel impulse response are uniquely determined by finite least squares smoothing error sequence in the absence of noise. The proposed subspace algorithm is shown to have marked improvement over existing algorithms in performance and robustness in simulations.**

***Index Terms*— Blind channel identification, least squares method.**

I. Introduction

One of the most important requirements for blind channel estimation and equalization is the speed of convergence. This is especially the case when it is used in packet transmission systems where only a small number of data samples are available for processing. Among blind channel estimation techniques developed recently [12], those based on the so-called deterministic models have a clear advantage in the speed of convergence. Without assuming a specific stochastic model of the input sequence, these "deterministic" techniques are capable of obtaining perfect channel estimation within a *finite* number of samples in the absence of noise. Such a finite-sample convergence property comes mainly from the exploitation of the multichannel structure first used in [13]. Existing algorithms with this attractive feature include the subspace (SS) algorithm [7], the cross relation (CR) (also referred to as the least squares) algorithm [16], the EVAM [5], the two-step maximum likelihood (TSML) approach [6], and the linear prediction-subspace (LP-SS) algorithm proposed by Slock [9].

Existing algorithms with the finite-sample convergence property share a common difficulty: the determination of channel order. While many order detection algorithms can be applied (see, e.g., [15] and references therein), the approach of separate order detection and channel estimation may not be effective, especially when the channel impulse response has small head and tail taps. Addressing this issue, a class of channel estimation algorithms based on the linear prediction (LP) interpretation of multichannel moving-average processes have been proposed first by Slock [9], followed by Slock and Papadias [10], Abed-Meraim *et al.* [1], [3], and more recently, by Gesbert and Duhamel [4]. Although only the upper bound of the channel order is required, these algorithms, with the exception of the (LP-SS) approach [9], which still requires the knowledge of the channel order, suffer considerable performance loss due to the requirement that the input sequence is white. Consequently, these algorithms need a relatively large sample size for accurate channel estimation, which causes the loss of finite-sample convergence property. Further, we may argue that although these algorithms provide a consistent estimate with only the knowledge of the bound of the channel order, overdetermination of the channel order does affect the performance when the sample size is finite.

The contribution of this paper is twofold. First, by exploiting the isomorphic relation between the input and output subspaces, we introduce a geometrical approach to linear least squares smoothing channel estimation that preserves the finite sample convergence property. This geometrical approach provides a simple and unified derivation of different LP-based channel estimators. Second, we develop a joint order detection and channel estimation algorithm that aims to minimize the smoothing error by jointly choosing the channel order and coefficients. When compared with existing approaches, the proposed algorithm provides considerable improvement in convergence over LP-based approaches. There is also marked improvement over CR and SS algorithms in robustness against the loss of channel diversity.

This paper is organized as follows. Section II presents a list of key notations followed by the channel model. Geometrical properties of least squares smoothing based on the isomorphic relation between the output and input subspaces are presented in Section III. In Section IV, we present a general formulation of LSS, data structures used in algorithm development and their properties, and a joint order detection and channel estimation algorithm. Simulation results are presented in Section V, where we compared the proposed algorithm with existing techniques. In conclusion, we comment on the strength and weakness of the proposed approach.

II. The Model and Preliminaries

A. Notations

Notations used in this paper are mostly standard. Signals are discrete-time and complex in general. We use $x(z)$ to denote the $\mathcal{Z}$-transform of signal x_t, and $x_t * y_t$ stands for the convolution of x_t and y_t. Upper- and lower-case bold

Manuscript received February 10, 1998; revised January 19, 1999. This work was supported in part by the National Science Foundation under Contract CCR-9804019 and by the Office of Naval Research under Contract N00014-96-1-0895. The associate editor coordinating the review of this paper and approving it for publication was Prof. Victor A. N. Barroso.

The authors are with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA.

Publisher Item Identifier S 1053-587X(99)06760-4.

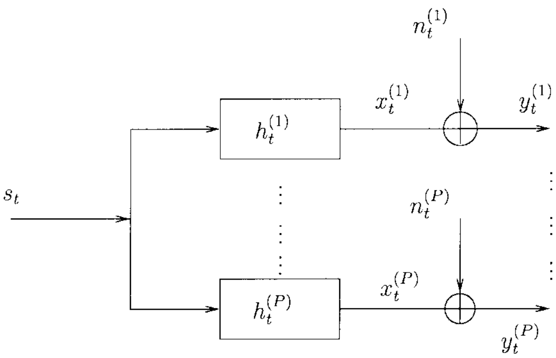

Fig. 1. Single-input multiple-output linear system.

letters denote matrices and vectors, respectively. $(\cdot)^t$ and $(\cdot)'$ are transpose and Hermitian operations. Matrix $\mathbf{0}_{m\times n}$ stands for the $m\times n$ zero matrix. Given a matrix $\boldsymbol{A}$, $\mathcal{R}\{\boldsymbol{A}\}$ ($\mathcal{C}\{\boldsymbol{A}\}$) is the Row (column) space of matrix $\boldsymbol{A}$. For a matrix $\boldsymbol{X}$ having the same number of columns as $\boldsymbol{A}$, $\mathcal{P}_{\boldsymbol{A}}\{\boldsymbol{X}\}$ ($\mathcal{P}_{\boldsymbol{A}}^{\perp}\{\boldsymbol{X}\}$) is the projection of $\boldsymbol{X}$ onto (orthogonal complement of) the row space of $\boldsymbol{A}$. For a set of vectors $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$, $sp\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$ denotes the linear subspace spanned by $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$. $\|\cdot\|$ denotes the 2-norm. Finally, $\mathcal{V}_n(x_1,\cdots,x_m)$ denotes the $n\times m$ Vandermonde matrix specified by $\{x_1,\cdots,x_m\}$

$$\mathcal{V}_n(x_1,\cdots,x_m)\triangleq\begin{pmatrix}1&\cdots&1\\x_1&\cdots&x_m\\\vdots&\cdots&\vdots\\x_1^{n-1}&\cdots&x_m^{n-1}\end{pmatrix}. \qquad (1)$$

B. The System Model

The identification and estimation of a single input P-output linear system (channel), shown in Fig. 1, using only the observation data is considered in this paper. The system is described by

$$\boldsymbol{x}(t)=\sum_{i=0}^{L}\boldsymbol{h}_i s_{t-i}$$
$$\boldsymbol{y}(t)=\boldsymbol{x}(t)+\boldsymbol{n}(t),\qquad t=1,2,\cdots,N \qquad (2)$$

where $\boldsymbol{x}(t)=[x_t^{(1)},\cdots,x_t^{(P)}]^t$ is the (noiseless) channel output, $\boldsymbol{n}(t)$ is the additive noise, $\boldsymbol{y}(t)$ is the received signal, $\{\boldsymbol{h}_t=[h_t^{(1)},\cdots,h_t^{(P)}]^t\}$ is the channel impulse response, and s_t is the input sequence. Consider a block of w samples of the observation in (2), and let $\boldsymbol{y}_w(t)\triangleq[\boldsymbol{y}^t(t),\cdots,\boldsymbol{y}^t(t-w+1)]^t$. With $\boldsymbol{x}_w(t)$, $\boldsymbol{s}_{L+w}(t)$, and $\boldsymbol{n}_w(t)$ similarly defined, we have

$$\boldsymbol{x}_w(t)=\mathcal{F}_w(\boldsymbol{h})\boldsymbol{s}_{L+w}(t),\quad \boldsymbol{y}_w(t)=\boldsymbol{x}_w(t)+\boldsymbol{n}_w(t) \qquad (3)$$

where the $wP\times(w+L)$ complex matrix $\mathcal{F}_w(\boldsymbol{h})$ is the so-called filtering matrix

$$\mathcal{F}_w(\boldsymbol{h})\triangleq\begin{pmatrix}\boldsymbol{h}_0&\cdots&\boldsymbol{h}_L&&\\&\ddots&\cdots&\ddots&\\&&\boldsymbol{h}_0&\cdots&\boldsymbol{h}_L\end{pmatrix}. \qquad (4)$$

Our goal is to estimate $\boldsymbol{h}\triangleq[\boldsymbol{h}_L',\cdots,\boldsymbol{h}_0']'$ from $\boldsymbol{y}_w(t), t=1,\cdots,N$. All signals are deterministic, although most results can be generalized to statistical models of the input and noise.

The following two assumptions (one on the system, the other on the input sequence) are made throughout the paper.

A1) There exists a (smallest) w_o such that the filtering matrix $\mathcal{F}_{w_o}(\boldsymbol{h})$ has full column rank.

A2) The input sequence s_t has linear complexity [2] greater than $L_*=2w_0+2L$, i.e.,

$$\text{rank}\left\{\begin{pmatrix}s_{L_*-L+1}&\cdots&s_N\\\vdots&\text{Toeplitz}&\\s_{1-L}&&\end{pmatrix}\right\}=L_*+1. \qquad (5)$$

Assumption A1, which was first exploited in [13], is necessary for all methods based on (general) *deterministic* modeling of the input sequence. Specifically, if *A1)* is not satisfied, there exists a different $\{\tilde{\boldsymbol{h}}_t,\tilde{s}_t\}$ such that $\boldsymbol{h}_t*s_t=\tilde{\boldsymbol{h}}_t*\tilde{s}_t$, i.e., two channels and their inputs produce the same noiseless observation and are unidentifiable from the observation. Implications of *A1)* are summarized below.

Property 1: Under *A1)*, we have the following:

P1.1) The subchannel transfer functions do not share common zeros, i.e., $\{h_i(z)\}$ are co-prime.

P1.2) $\mathcal{F}_w(\boldsymbol{h})$ has full column rank for all $w\geq w_o$.

P1.3) If $P=2$, then $w_o=L$. In general, $w_o\leq L$.

Assumption *A2)* ensures that the input sequence is sufficiently complex to excite the channel, and it is related to the persistent excitation condition. The minimum required for the smoothing technique presented in this paper is assumed here. Larger complexity may be necessary, depending on the implementations, which will be pointed out later in our discussion. We note here that *A2)* is stronger than necessary. It is shown in [11] that when $P=2$, the *necessary and sufficient* condition for the unique identification of the channel and its input is *A1)* and that the input sequence s_t has linear complexity greater than $2L$. The reason that a stronger condition is required due largely to the smoothing approach that requires both future and past data.

III. Geometrical Properties of Least Squares Smoothing

The essential idea behind the linear prediction and smoothing approaches to channel estimation rests on the isomorphic relationship between the output and the input subspaces. It is this isomorphic relation of the two spaces that allows us to avoid the direct use of input sequence, using instead the input subspace that can be obtained from (noiseless) observation. Our presentation relies heavily on geometrical intuition. It is therefore necessary to begin with precise definitions of relevant variables and spaces.

A. Key Variables, Spaces, and Isomorphic Relations

From (2), let $\boldsymbol{s}_t$ be the row vector of input symbols and $\boldsymbol{x}_t$ be the data matrix of the noiseless observation defined, respectively, by

$$\boldsymbol{s}_t\triangleq[s_t,s_{t+1},\cdots],\quad \boldsymbol{x}_t\triangleq[\boldsymbol{x}(t),\boldsymbol{x}(t+1),\cdots]. \qquad (6)$$

Consider next subspaces spanned by p ($p > 0$) consecutive vectors

$$\begin{aligned}\mathcal{S}_{t,p} &\triangleq sp\{\boldsymbol{s}_t, \cdots, \boldsymbol{s}_{t-p+1}\}\\ &= \mathcal{R}\left\{\begin{pmatrix} s_t & s_{t+1} & \cdots \\ \vdots & \text{Toeplitz} & \\ s_{t-p+1} & & \end{pmatrix}\right\}\\ &= \mathcal{R}\{[\boldsymbol{s}_p(t), \boldsymbol{s}_p(t+1), \cdots]\}\end{aligned} \tag{7}$$

$$\begin{aligned}\mathcal{X}_{t,p} &\triangleq sp\{\boldsymbol{x}_t, \cdots, \boldsymbol{x}_{t-p+1}\}\\ &= \mathcal{R}\left\{\begin{pmatrix} \boldsymbol{x}(t) & \boldsymbol{x}(t+1) & \cdots \\ \vdots & \text{Block} & \\ \boldsymbol{x}(t-p+1) & \text{Toeplitz} & \end{pmatrix}\right\}\\ &= \mathcal{R}\{[\boldsymbol{x}_p(t), \boldsymbol{x}_p(t+1), \cdots]\}.\end{aligned} \tag{8}$$

The above definition also applies to $p < 0$, in which case, we have the span of $|p|$ future data vectors. It is also useful to note that

$$\mathcal{S}_{t,p} = \mathcal{S}_{t-p+1,-p}, \quad \mathcal{X}_{t,p} = \mathcal{X}_{t-p+1,-p}. \tag{9}$$

Given a linear subspace $\mathcal{S}$, the orthogonal projection $\hat{\boldsymbol{s}}_t$ of $\boldsymbol{s}_t$ and its projection error $\tilde{\boldsymbol{s}}_t$ are defined by

$$\hat{\boldsymbol{s}}_{t|\mathcal{S}} \triangleq \arg\min_{\boldsymbol{z}\in\mathcal{S}} \|\boldsymbol{s}_t - \boldsymbol{z}\|^2, \tilde{\boldsymbol{s}}_{t|\mathcal{S}} \triangleq \boldsymbol{s}_t - \hat{\boldsymbol{s}}_{t|\mathcal{S}}. \tag{10}$$

Similarly, the (row-wise) projection $\hat{\boldsymbol{x}}_{t|\mathcal{X}}$ of $\boldsymbol{x}_t$ onto a linear subspace $\mathcal{X}$ is a matrix whose rows are projections of $\boldsymbol{x}_t$ onto $\mathcal{X}$.

Playing a critical role in the smoothing as well as LP-based approaches is the equivalence between the input and output spaces as a result of *A1)* and *P1.2)*. Specifically, we have the following.

Properties 2: Under *A1)*, for $w \geq w_o$, $\mathcal{X}_{t,w} = \mathcal{S}_{t,L+w}$, i.e., $\mathcal{X}_{t,w}$ is isomorphic to $\mathcal{S}_{t,L+w}$ with isomorphism $\mathcal{F}_w(\boldsymbol{h})$.

In general, $\mathcal{X}_{t,w} \subseteq \mathcal{S}_{t,L+w}$ for any w. This implies that given a fixed observation window w, the input space $\mathcal{S}_{t,L+w}$ may not be "seen" completely from the output space $\mathcal{X}_{t,w}$. On the other hand, with *A1)*, all the information of the input space is contained in the output space $\mathcal{X}_{t,w}$ when subchannels do not have common zeros [*P1.1)*] and w is chosen large enough. Such equivalence enables us to replace the direct use of input sequence by the use of observation in channel estimation. Interestingly, when *A1)* does not hold, $\mathcal{X}_{t,w}$ may still be a good approximation of $\mathcal{S}_{t,L+w}$, which is one of the reasons that the algorithm proposed here offers considerable improvement in robustness over existing methods such as the subspace algorithm (SS) [7] and cross relation (CR) algorithm [16].

B. Least Squares Smoothing—The Basic Idea

The isomorphism between the input space $\mathcal{S}_{t,w}$ and the output spaces $\mathcal{X}_{t,w}$ leads to the following question: *Can the channel be identified from $\mathcal{S}_{t,w}$ without the direct use of the input sequence s_t and how?* Without going into implementation details, we explain in this section how this can be achieved by properly constructing subspaces that contain both past and future data, namely, by smoothing. LSS algorithms and their implementation issues are presented in Section IV.

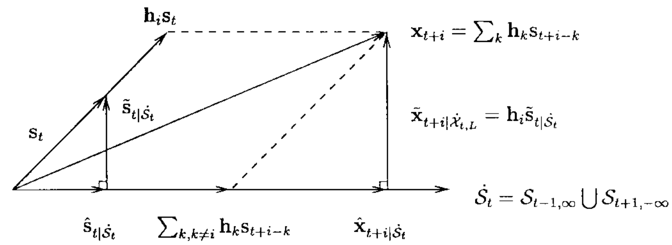

Fig. 2. Projection of $\boldsymbol{x}_{t+i}$ onto $\dot{\mathcal{S}}_t$.

1) The Use of Input Subspaces: From (2) and (6), we have

$$\boldsymbol{x}_t = \boldsymbol{h}_0\boldsymbol{s}_t + \boldsymbol{h}_1\boldsymbol{s}_{t-1} + \cdots + \boldsymbol{h}_L\boldsymbol{s}_{t-L}. \tag{11}$$

To avoid cumbersome boundary problems, we assume for the moment the data size is infinity, i.e., $N = \infty$. Let $\dot{\mathcal{S}}_t$ be the subspace that includes all future and past input data except $\boldsymbol{s}_t$, i.e.,

$$\dot{\mathcal{S}}_t \triangleq sp\{\cdots, \boldsymbol{s}_{t-1}\} \cup sp\{\boldsymbol{s}_{t+1}, \cdots\} = \mathcal{S}_{t-1,\infty} \cup \mathcal{S}_{t+1,-\infty}. \tag{12}$$

Projecting $\boldsymbol{x}_{t+i}$ onto $\dot{\mathcal{S}}_t$ with only $\boldsymbol{h}_i\boldsymbol{s}_t$ not contained in $\dot{\mathcal{S}}_t$, we have

$$\hat{\boldsymbol{x}}_{t+i|\dot{\mathcal{S}}_t} = \boldsymbol{h}_i\hat{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_t} + \sum_{k,k\neq i}\boldsymbol{h}_k\boldsymbol{s}_{t+i-k} \tag{13}$$

$$\tilde{\boldsymbol{x}}_{t+i|\dot{\mathcal{X}}_{t,L}} = \boldsymbol{x}_{t+i} - \hat{\boldsymbol{x}}_{t+i} = \boldsymbol{h}_i\tilde{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_t}. \tag{14}$$

The above process is illustrated in Fig. 2. The similarity of two right triangles immediately suggests (14).

Note that the projection error $\tilde{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_t}$ of $\boldsymbol{s}_t$ is independent of i. Consequently, we have

$$\boldsymbol{E} \triangleq \begin{pmatrix}\tilde{\boldsymbol{x}}_{t+L|\dot{\mathcal{S}}_t}\\ \vdots \\ \tilde{\boldsymbol{x}}_{t|\dot{\mathcal{S}}_t}\end{pmatrix} = \boldsymbol{h}\tilde{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_t}. \tag{15}$$

From $\boldsymbol{E}$, there are several ways of finding $\boldsymbol{h}$ up to a scaling factor, and they have different performance when there is noise and when other implementation issues are considered. We remark that because subspaces are invariant with respect to scaling, the identification of $\boldsymbol{h}$ up to a scaling factor using only the input *subspace* is the best we should expect. One approach is the least squares fitting of the column space of $\boldsymbol{E}$:

$$\hat{\boldsymbol{h}} = \arg\max_{\|\boldsymbol{h}\|=1} \|\boldsymbol{h}'\boldsymbol{E}\|^2. \tag{16}$$

The above optimization can be obtained by the singular value decomposition of either $\boldsymbol{E}$ or the sample covariance of the projection error sequence $\hat{\boldsymbol{R}}_E \triangleq (1/M)\boldsymbol{E}\boldsymbol{E}'$, where M is the number of columns in $\boldsymbol{E}$.

There is an interesting connection with the conventional LS approach when the input sequence is known. Indeed, were the input sequence available, the LS channel estimate would have been

$$\begin{aligned}\hat{\boldsymbol{h}}_{LS} &= \boldsymbol{x}_t\boldsymbol{S}'(\boldsymbol{S}\boldsymbol{S}')^{-1} = \frac{1}{\|\tilde{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_t}\|^2}\boldsymbol{E}\tilde{\boldsymbol{s}}'_{t|\dot{\mathcal{S}}_t}\\ \boldsymbol{S} &= \begin{pmatrix}\boldsymbol{s}_t\\ \vdots\\ \boldsymbol{s}_{t-L}\end{pmatrix}\end{aligned} \tag{17}$$

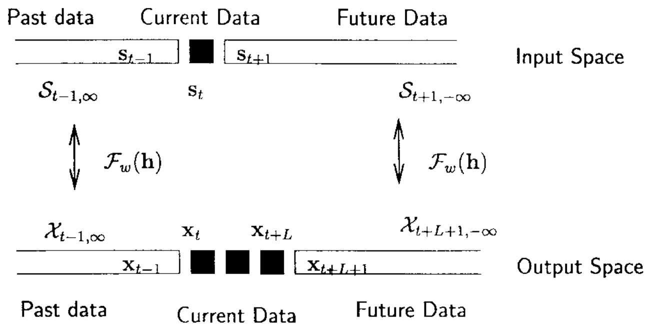

Fig. 3. Isomorphism between input and output subspaces.

which is the same as that in (16). The second equality is obtained by using the formula of inverting matrices with subblocks [8, p. 413, A20.].

2) Use of Output Spaces-Least Squares Smoothing: The identification procedure presented above requires the projection of $\boldsymbol{x}_t$ onto the input space $\dot{\mathcal{S}}_t$. Because $\mathcal{X}_{t,w}$ is isomorphic to $\mathcal{S}_{t,w}$, the application of the above approach using only the received signal requires only a careful construction of the output subspace that is isomorphic to $\dot{\mathcal{S}}_t$. Specifically, using Property 2 and (9), we have

$$\mathcal{S}_{t,\infty} = \mathcal{X}_{t,\infty},\ \mathcal{S}_{t+1,-\infty} = \mathcal{X}_{t+L+1,-\infty} \tag{18}$$

$$\dot{\mathcal{X}}_{t,L} \triangleq \mathcal{X}_{t-1,\infty} \cup \mathcal{X}_{t+L+1,-\infty} = \dot{\mathcal{S}}_t. \tag{19}$$

The isomorphism between the input and output subspaces is illustrated in Fig. 3. The idea of smoothing arises naturally as the projection of $\boldsymbol{x}_t$ onto $\dot{\mathcal{S}}_t$ in (14) is equivalent to the projection of $\boldsymbol{x}_t$ onto the output subspace spanned by all the past data $\mathcal{X}_{t-1,\infty}$ and future data $\mathcal{X}_{t+L+1,-\infty}$. From (15), we have the identification equation of the LSS approach [14]

$$\boldsymbol{E} \triangleq \begin{pmatrix} \tilde{\boldsymbol{x}}_{t+L|\dot{\mathcal{X}}_t} \\ \vdots \\ \tilde{\boldsymbol{x}}_{t|\dot{\mathcal{X}}_t} \end{pmatrix} = \boldsymbol{h}\tilde{\boldsymbol{s}}_{t|\dot{\mathcal{X}}_t} \tag{20}$$

where the left-hand side can be obtained from the observation alone.

IV. Algorithm and Implementations

In this section, we provide more details about the LSS approach including its properties and implementations. We begin with a general formulation of LSS that forms the basis of our approach. Data structures of the LSS approach are specified along with their properties. We then derive a joint order detection and channel estimation algorithm and discuss its implementations.

A. General Formulation of LSS

The projection space $\dot{\mathcal{X}}_{t,L}$ defined in (19) requires the knowledge of channel order. We consider here a more general formulation of the problem by defining slightly different projection spaces that enable us to deal with practical issues such as finite sample size and unknown channel order. Instead of using the projection space given in (19), consider the smoothing of $l+1$ observations $\boldsymbol{x}_{t+i}, i = 0, \cdots, l$ by forward and backward predictors of order $w \geq w_0$. The projection space is given by

$$\dot{\mathcal{X}}_{t,l} \triangleq \mathcal{X}_{t-1,w} \cup \mathcal{X}_{t+l+1,-w} = \mathcal{X}_{t-1,w} \cup \mathcal{X}_{t+l+w,w}. \tag{21}$$

We notice that $\dot{\mathcal{X}}_{t,l}$ is essentially the same as that defined in (19), except that we treat l as a variable not necessarily equal to the channel order L. Because of the isomorphic relation between the output and input spaces, we have, using (9)

$$\dot{\mathcal{X}}_{t,l} = \mathcal{S}_{t-1,L+w} \cup \mathcal{S}_{t+l+w,L+w} \triangleq \dot{\mathcal{S}}_{t,l}. \tag{22}$$

Therefore

$$\dot{\mathcal{S}}_{t,l} = \begin{cases} sp\{\boldsymbol{s}_{t-L-w}, \cdots, \boldsymbol{s}_t, \cdots, \boldsymbol{s}_{t+l+w}\} \\ \quad l < L \\ sp\{\boldsymbol{s}_{t-L-w}, \cdots, \boldsymbol{s}_{t-1}\} \cup sp\{\boldsymbol{s}_{t+l-L+1}, \cdots, \boldsymbol{s}_{t+l+w}\} \\ \quad l \geq L. \end{cases} \tag{23}$$

Projecting $\boldsymbol{x}_{t+i}, i = 0, \cdots, l$ onto $\dot{\mathcal{X}}_{t,l} = \dot{\mathcal{S}}_{t,l}$, we have the following result as a generalization of (15).

Theorem 1: Let the forward and backward predictor order $w \geq w_o$. Let $\dot{\mathcal{X}}_{t,l}$ be defined in (22), and let $\boldsymbol{E}_{t,l}$ be the projection error matrix defined by

$$\boldsymbol{E}_{t,l} \triangleq \begin{pmatrix} \tilde{\boldsymbol{x}}_{t+l|\dot{\mathcal{X}}_{t,l}} \\ \vdots \\ \tilde{\boldsymbol{x}}_{t|\dot{\mathcal{X}}_{t,l}} \end{pmatrix}. \tag{24}$$

Then

$$\boldsymbol{E}_{t,l} = \begin{cases} \boldsymbol{0} & l < L \\ \mathcal{H}_l(\boldsymbol{h}) \begin{pmatrix} \tilde{\boldsymbol{s}}_{t+l-L|\dot{\mathcal{S}}_{t,l}} \\ \vdots \\ \tilde{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_{t,l}} \end{pmatrix} & L \leq l \end{cases}$$

$$\mathcal{H}_l(\boldsymbol{h}) \triangleq \underbrace{\begin{pmatrix} \boldsymbol{h}_L & & \\ \vdots & \ddots & \\ \boldsymbol{h}_0 & & \boldsymbol{h}_L \\ & \ddots & \vdots \\ & & \boldsymbol{h}_0 \end{pmatrix}}_{l-L+1 \text{columns}}. \tag{25}$$

Further, if $\{s_t\}$ has linear complexity greater than $2w+l+L$ and $l \geq L$

$$\mathcal{C}\{\boldsymbol{E}_{w+1,l}\} = \mathcal{C}\{\mathcal{H}_l(\boldsymbol{h})\} \tag{26}$$

Proof: From (11), (21), (23), we have, for $0 \leq i \leq l$

$$\tilde{\boldsymbol{x}}_{t+i|\dot{\mathcal{X}}_{t,l}} = \tilde{\boldsymbol{x}}_{t+i|\dot{\mathcal{S}}_{t,l}} \tag{27}$$

$$= \begin{cases} \boldsymbol{0} & l < L \\ \displaystyle\sum_{k=i-l+L}^{i} \boldsymbol{h}_k \tilde{\boldsymbol{s}}_{t+i-k|\dot{\mathcal{S}}_{t,l}} & l \geq L \cdot \end{cases} \tag{28}$$

With $\boldsymbol{h}_k = 0$ for all $k < 0$ and $k > L$, the above equation leads to (25).

To prove (26), we need to show that the projection error matrix of the input sequence in (25) has full row rank. Consider the Toeplitz matrix

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{s}_{2w+l+1} & \cdots & s_N \\ \vdots & \boldsymbol{A} & \\ \boldsymbol{s}_{w+l-L+2} & \cdots & \\ \hline \boldsymbol{s}_{w+l-L+1} & \cdots & \\ \vdots & \boldsymbol{B} & \\ \boldsymbol{s}_{w+1} & \cdots & \\ \hline \boldsymbol{s}_{w} & \cdots & \\ \vdots & \boldsymbol{C} & \\ \boldsymbol{s}_{1-L} & & \end{pmatrix}. \tag{29}$$

With $t = w+1$ in (25), we note that

$$\begin{pmatrix} \tilde{\boldsymbol{s}}_{w+l-L+1|\dot{\mathcal{S}}_{w+1,l}} \\ \vdots \\ \tilde{\boldsymbol{s}}_{w+1|\dot{\mathcal{S}}_{w+1,l}} \end{pmatrix} = \mathcal{P}^{\perp}_{\boldsymbol{D}}\{\boldsymbol{B}\}, \boldsymbol{D} \triangleq \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{C} \end{pmatrix}. \tag{30}$$

When $\{s_t\}$ has linear complexity greater than $2w+l+L$, $\boldsymbol{S}$ has full row rank, which implies that $\mathcal{P}^{\perp}_{\boldsymbol{D}}\{\boldsymbol{B}\}$ has full row rank. We now have (26). □□□

The above result holds the key to our approach, especially when the channel order L is unknown. When the smoothing window size l is too small, the smoothing error contains no information about the channel because all input sequences are in the output space. When $l = L$, we have the case described in Section III-B, where the channel vector spans the column space of $\boldsymbol{E}_{t,l}$. When the window size l is greater than the channel order L, the projection space misses more than $\boldsymbol{s}_t$, which complicates the channel identification. Nonetheless, it is shown in Section IV-C that the column space of $\boldsymbol{E}_{t,l}$ still uniquely determines the channel vector, which, along with another useful property of $\boldsymbol{E}_{t,l}$, forms the basis of a joint order-detection and channel estimation algorithm that requires only the upper bound of the channel order.

B. Data Structures

We consider now the problem of estimating the channel using only a finite number of received signal samples $\boldsymbol{y}(t), t = 1, \cdots, N$. For a fixed predictor size $w \geq w_o$ and smoothing window $l \geq 0$, define the overall data matrix

$$\boldsymbol{Z}_{w,l} \triangleq \begin{pmatrix} \boldsymbol{y}(2w+l+1) & \cdots & \boldsymbol{y}(N) \\ \vdots & \boldsymbol{F}_{w,l} & \\ \boldsymbol{y}(w+l+2) & & \\ \hline \boldsymbol{y}(w+l+1) & \cdots & \\ \vdots & \boldsymbol{Y}_{w,l} & \\ \boldsymbol{y}(w+1) & & \\ \hline \boldsymbol{y}(w) & \cdots & \\ \vdots & \boldsymbol{P}_{w,l} & \\ \boldsymbol{y}(1) & & \end{pmatrix} \tag{31}$$

from which we have defined the "current" data matrix $\boldsymbol{Y}_{w,l}$, the "past" data matrix $\boldsymbol{P}_{w,l}$, and the "future" data matrix $\boldsymbol{F}_{w,l}$.

Denote the "future-past" data matrix $\boldsymbol{D}_{w,l}$ as

$$\begin{aligned} \boldsymbol{P}_{w,l} &\triangleq [\boldsymbol{y}_w(w), \cdots, \boldsymbol{y}_w(N-w-l-1)] \\ \boldsymbol{F}_{w,l} &\triangleq [\boldsymbol{y}_w(2w+l+1), \cdots, \boldsymbol{y}_w(N)] \\ \boldsymbol{Y}_{w,l} &\triangleq [\boldsymbol{y}_w(w+l+1), \cdots, \boldsymbol{y}_w(N-w)] \end{aligned} \tag{32}$$

$$\boldsymbol{D}_{w,l} \triangleq \begin{pmatrix} \boldsymbol{F}_{w,l} \\ \boldsymbol{P}_{w,l} \end{pmatrix}. \tag{33}$$

To see the relation between these data matrices and various spaces, we summarize their properties. The rank conditions given below are useful in dealing with noise by finding the least squares approximation of the noisy data matrix.

Property 3: Suppose that the input sequence has linear complexity greater than $2w+l+L+1$ and there is no noise. For $w \geq w_o$, we have the following properties.

P3.1) Data Matrix $\boldsymbol{Z}_{w,l}$:

$$\text{rank}(\boldsymbol{Z}_{w,l}) = 2w+l+L+1. \tag{34}$$

P3.2) Past Data Matrix $\boldsymbol{P}_{w,l}$:

$$\mathcal{R}\{\boldsymbol{P}_{w,l}\} = \mathcal{X}_{w,w} = \mathcal{S}_{w,w+L} \tag{35}$$
$$\text{rank}(\boldsymbol{P}_{w,l}) = L+w. \tag{36}$$

P3.3) Future Data Matrix $\boldsymbol{F}_{w,l}$

$$\mathcal{R}\{\boldsymbol{F}_{w,l}\} = \mathcal{X}_{2w+l+1,w} = \mathcal{S}_{2w+l+1,w+L} \tag{37}$$
$$\text{rank}(\boldsymbol{F}_{w,l}) = \text{rank}(\boldsymbol{P}_{w,l}) = L+w. \tag{38}$$

P3.4) Projection Data Matrix $\boldsymbol{D}_{w,l}$

$$\mathcal{R}\{\boldsymbol{D}_{w,l}\} = \mathcal{S}_{w,w+L} \cup \mathcal{S}_{w+l-L+1,-w-L} = \dot{\mathcal{S}}_{w+1,l} \tag{39}$$
$$\text{rank}(\boldsymbol{D}_{w,l}) = \begin{cases} 2w+l+L+1, & l < L \\ 2w+2L, & L \leq l \leq w. \end{cases} \tag{40}$$

Proof: See the Appendix.

C. J-LSS: Joint Order Detection and Channel Estimation via LSS

If the channel order is known or can be detected, channel estimation by LSS can be derived directly from (20). This approach and its adaptive implementations are explored in [14], [18], and [17]. Here, we describe a *joint order detection and channel estimation* approach based on Theorem 1 and the data structure defined above assuming only that an upper bound of channel order is available.

The idea here is to fit the smoothing error matrix $\boldsymbol{E}_{w,l}$ by jointly choosing both the channel order and the channel impulse response. With fixed l as the upper bound of the true channel order L, recall Theorem 1 for the case when $l > L$. Consider the smoothing error matrix $\boldsymbol{E}_{l,l} = \mathcal{P}^{\perp}_{\boldsymbol{D}_{l,l}}\{\boldsymbol{Y}_{l,l}\}$ obtained from projecting $\boldsymbol{Y}_{l,l}$ onto the row space of $\boldsymbol{D}_{l,l}$. We now have from (26), when there is no noise

$$\mathcal{C}\{\boldsymbol{E}_{l,l}\} = \mathcal{C}\{\mathcal{H}_l(\boldsymbol{h})\}. \tag{41}$$

Letting $\boldsymbol{Q} = [\boldsymbol{Q}_0, \cdots, \boldsymbol{Q}_l]$ be the matrix whose row vectors are orthogonal to the range space of $\boldsymbol{E}_{l,l}$, we then have

$$[\boldsymbol{Q}_0, \cdots, \boldsymbol{Q}_l]\boldsymbol{E}_{l,l} = 0$$

which implies

$$\underbrace{\begin{pmatrix} \boldsymbol{Q}_0 & \cdots & \boldsymbol{Q}_L \\ & \text{Block} & \vdots \\ & \text{Hankel} & \boldsymbol{Q}_l \end{pmatrix}}_{\mathcal{T}_L(\boldsymbol{Q})} \begin{pmatrix} \boldsymbol{h}_L \\ \vdots \\ \boldsymbol{h}_0 \end{pmatrix} = \mathcal{T}_L(\boldsymbol{Q})\boldsymbol{h} = \boldsymbol{0}. \tag{42}$$

In other words, the channel coefficients satisfy a homogeneous linear equation. What remains to be answered is whether the solution is unique up to a scaling factor. The proposed joint order detection and channel estimation algorithm is motivated by the following Theorem.

Theorem 2: Assume that there is no noise, and the input sequence has linear complexity greater than $3l + L$. Assume also that the channels do not share common zeros. Let $\boldsymbol{E}_{l,l} = \mathcal{P}^{\perp}_{\boldsymbol{D}_{l,l}}\{\boldsymbol{Y}_{l,l}\}$ be the projection error matrix, and let $\hat{\boldsymbol{R}}_E \triangleq (1/N - 3l)\boldsymbol{E}_{l,l}\boldsymbol{E}'_{l,l}$ be the sample covariance matrix of the smoothing error sequence. Let the rows of $\boldsymbol{Q}$ be the singular vectors associated with $(P(l+1) - l + k - 1)$ smallest singular values of $\hat{\boldsymbol{R}}_E$, and let $\boldsymbol{Q}$ be partitioned by $(P(l+1)-l+k-1)\times P$ submatrices $\boldsymbol{Q} \triangleq [\boldsymbol{Q}_0, \cdots, \boldsymbol{Q}_l]$. Define

$$\mathcal{T}_k(\boldsymbol{Q}) \triangleq \begin{pmatrix} \boldsymbol{Q}_0 & \cdots & \boldsymbol{Q}_k \\ & \text{Block} & \vdots \\ & \text{Hankel} & \boldsymbol{Q}_l \end{pmatrix}. \tag{43}$$

Then, the homogeneous linear equation

$$\mathcal{T}_k(\boldsymbol{Q})\boldsymbol{z} = \boldsymbol{0} \tag{44}$$

has the unique nontrivial solution $\boldsymbol{z} = \alpha\boldsymbol{h}$ when $k = L$ and trivial solutions otherwise.

Proof: See the Appendix.

We note that the above result does not apply to the subspace algorithm. When $k = L$, (44) defines a channel estimator that bears some similarity to the subspace algorithm used by Moulines *et al.* [7]. In both cases, the so-called noise subspace is used in constructing a homogeneous linear equation of which the channel vector is a unique solution. However, there are several important differences. First, the filtering matrix $\mathcal{F}_w(\boldsymbol{h})$ used in the subspace approach is different from the smoothing error matrix $\boldsymbol{E}_{l,l}$. Maybe more importantly, the homogeneous equation used in the subspace algorithm has nontrivial solutions when the estimated channel order is larger than the true channel order, which is the reason that the joint order detection and channel estimation approach does not apply to the subspace algorithm directly.

It is perhaps surprising that when $k \neq L$, (44) has only the trivial solution. Intuitively, we can argue as follows. When the channel order is overdetermined, i.e., $k > L$, in constructing $\boldsymbol{Q}$, we must include eigenvectors that are in the range space of $\mathcal{H}_l(\boldsymbol{h})$, which leads to inconsistency of $\boldsymbol{Q}\boldsymbol{E}_{l,l} = \boldsymbol{0}$. (Note that such inconsistency does not occur for the subspace algorithm when the channel order is overdetermined.) For a generically chosen channel, $\mathcal{T}_k(\boldsymbol{Q})$ has full column rank. On the other hand, when the channel order is underestimated $k < L$, there are an insufficient number of parameters to specify the null space of $\hat{\boldsymbol{R}}_E$.

Theorem 2 enables us to define the following joint channel order detection and estimation criterion:

$$\{\hat{L}, \hat{\boldsymbol{h}}\} = \arg\min_{k, \|\boldsymbol{h}\|=1} \|\mathcal{T}_k(\boldsymbol{Q})\boldsymbol{h}\|^2. \tag{45}$$

The above optimization has a closed-form solution involving the singular vector associated with the smallest singular value. The joint order detection and channel estimation approach, which is referred to as J-LSS, is summarized in Fig. 4.

There are many ways of implementing the algorithm outlined in Fig. 4. We discuss here several implementation issues that are likely to affect the performance.

The Smoothing Window Size l: It is clearly possible to implement the algorithm with variable smoothing window size. For simplicity, we considered the fixed window size case. Although not necessary for $P > 2$, the smoothing window size l, in theory, upper bounds the channel order. In practice, channel order is perhaps fictitious, and we can always argue that l can never upper bound the "true" channel order. Fortunately, when $\boldsymbol{h}_k \neq 0$ for $k > l$, the performance is not drastically affected as long as these "spill-out" coefficients are sufficiently small. In the simulation example shown in Section V, the robustness of J-LSS with respect to the underestimation of channel order is clearly demonstrated. In such a case, the finite-sample convergence property is lost as in all other algorithms.

Order Selection for the Predictors:

In selecting the order w for the forward and backward predictors, we should observe the following factors. First, for fixed data length, large w implies a fewer number of columns in data matrices. This corresponds to smaller sample size in least squares problems. In this regard, it is desirable to choose w as small as possible, which is the reason why we have considered $w = l$ in the algorithm. Certainly, if $P > 2$, a smaller w can be choosen. On the other hand, larger w may provide a certain degree of robustness, especially when subchannels have zeros approximately common near the unit circle. It is clearly possible to vary the predictor size w with k.

V. Simulation Examples

A. Algorithm Characteristics and Performance Measure

Simulation studies of the proposed LSS algorithms as they are compared with existing techniques listed in Table I are presentd in this section. We remark that only J-LSS does not require the knowledge of channel order while still preserving the finite sample convergence property.

Algorithms are compared by Monte Carlo simulation using the normalized root mean square error (NRMSE) as a performance measure. Specifically, NRMSE is defined by

$$\text{NRMSE}^1 \triangleq \sqrt{\frac{1}{N_m\|\boldsymbol{h}_*\|^2} \sum_{m=1}^{N_m} \|\hat{\boldsymbol{h}}^{(m)} - \boldsymbol{h}_*\|^2} \tag{46}$$

[1] The inherent ambiguity was removed before the computation of NRMSE. This includes scaling $\hat{\boldsymbol{h}}^{(\mathrm{m})}$ and adjusting delays by adding zeros to either $\hat{\boldsymbol{h}}^{(\mathrm{m})}$ or $\boldsymbol{h}_*$.

J-LSS Order Detection and Channel Estimation

1. Choose $l > L$ and form data matrices $\mathbf{Y}_{l,l}$ and $\mathbf{D}_{l,l}$.
2. Obtain the $4l$ orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_{4l}\}$ that spans the row space of $\mathbf{D}_{l,l}$.
3. Obtain the projection error of $\mathbf{Y}_{l,l}$ onto $sp\{\mathbf{u}_1, \cdots, \mathbf{u}_{4l}\}$:

$$\mathbf{E}_{l,l} \triangleq \mathbf{Y}_{l,l} - \mathbf{Y}_{l,l}\mathbf{U}'\mathbf{U}, \quad \mathbf{U} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{4l} \end{pmatrix} \tag{66}$$

4. For each $1 \leq k < l$, treated as the estimated channel order, let $\mathbf{Q} = [\mathbf{Q}_0, \cdots, \mathbf{Q}_l]$ be matrix whose rows are the last $P(l+1) - l + k - 1)$ left singular vectors of $\mathbf{E}_{l,l}$, or equivalently, the sample covariance of the smoothing error sequence. Form

$$\mathcal{T}_k(\mathbf{Q}) \triangleq \begin{pmatrix} \mathbf{Q}_0 & \cdots & \mathbf{Q}_k \\ & \text{Block} & \vdots \\ & \text{Hankel} & \mathbf{Q}_l \end{pmatrix}. \tag{67}$$

5. Joint Order Detection and Channel estimation:

$$\{\hat{L}, \hat{\mathbf{h}}\} = \arg \min_{k, ||\mathbf{h}||=1} ||\mathcal{T}_k(\mathbf{Q})\mathbf{h}||^2 \tag{68}$$

Fig. 4. J-LSS algorithm.

TABLE I
LIST OF ALGORITHMS COMPARED IN THE SIMULATION AND THEIR CHARACTERISTICS

Abbreviation	Name	Convergence	Order Required?
SS [7]	The Subspace Algorithm	finite	Yes
CR [16], [5]	The Cross Relation Algorithm†	finite	Yes
LP-SS[9]	Linear Prediction-Subspace Algorithm	finite	Yes
LP-LS [1]	Linear Prediction-Least Squares Algorithm	infinite	No
MSP [4]	Multistep Linear Prediction Algorithm	infinite	No
LSS [14]	Least Squares Smoothing Algorithm	finite	Yes
J-LSS	Joint Order Detection and Channel Estimation by Least Squares Smoothing.	finite	No.

where $\hat{\boldsymbol{h}}^{(m)}$ was the estimated channel from the mth trial. Noise samples are generated from i.i.d. zero mean Gaussian random sequence, and the signal-to-noise ratio (SNR) was defined and given by

$$\text{SNR} \triangleq \frac{1}{P\sigma^2} E\left\{\sum_{j=1}^{P} |x_i^{(j)}|^2\right\} \tag{47}$$

where σ^2 was the noise variance. The input sequence to the channel is an i.i.d. quadrature phase shift keying (QPSK) complex sequence.

B. Performance Comparison: A Multipath Channel

Fig. 5 shows the NRMSE performance comparison using a four-ray multipath channel generated from the raised-cosine pulse. The T/2-sampled channel parameters are given in Table II with even and odd samples corresponding to the two subchannels. This channel has severe intersymbol interference. It is also close to violate the identifiability condition in the sense that the filtering matrix $\mathcal{F}_w(\boldsymbol{h})$ has condition number around 3×10^3. We have also performed simulation comparisons for well-conditioned channels [18]. The performance of J-LSS and SS are comparable in those cases.

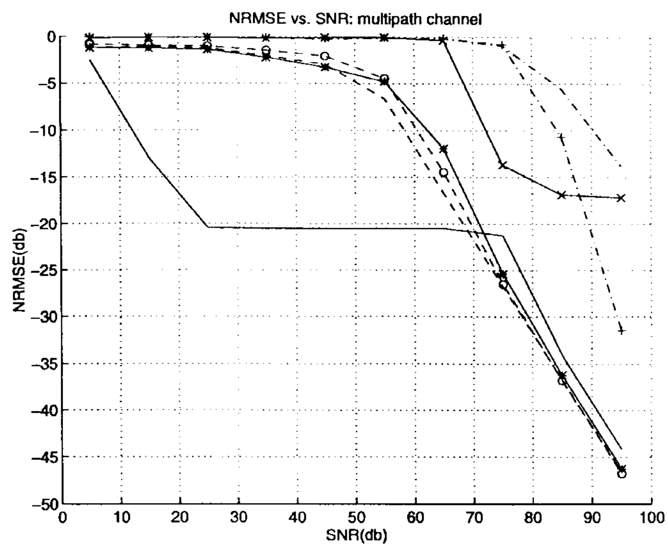

Fig. 5. NRMSE performance comparison for the multipath channel. One hundred Monte-Carlo runs. One hundred input symbols. Legend: SS: ‘’; CR: ‘o’; LP-SS: ‘+’; LP-LS: ‘−.’; MSP: ‘×’; LSS: ‘*’; J-LSS: ‘−’.

TABLE II
MULTIPATH CHANNEL

t	1	2	3	4
h_t	$-0.0031 - j0.0017$	$-0.0016 - j0.0047$	$-0.0109 - j0.0025$	$-0.0263 - j0.0433$
t	5	6	7	8
h_t	$0.1522 + j0.0705$	$0.4409 + j0.4736$	$0.3789 + j0.5930$	$0.0766 + j0.2168$
t	9	10	11	12
h_t	$-0.0301 - j0.0348$	$-0.0042 - j0.0154$	$-0.0032 - j0.0017$	$-0.0017 - j0.0044$

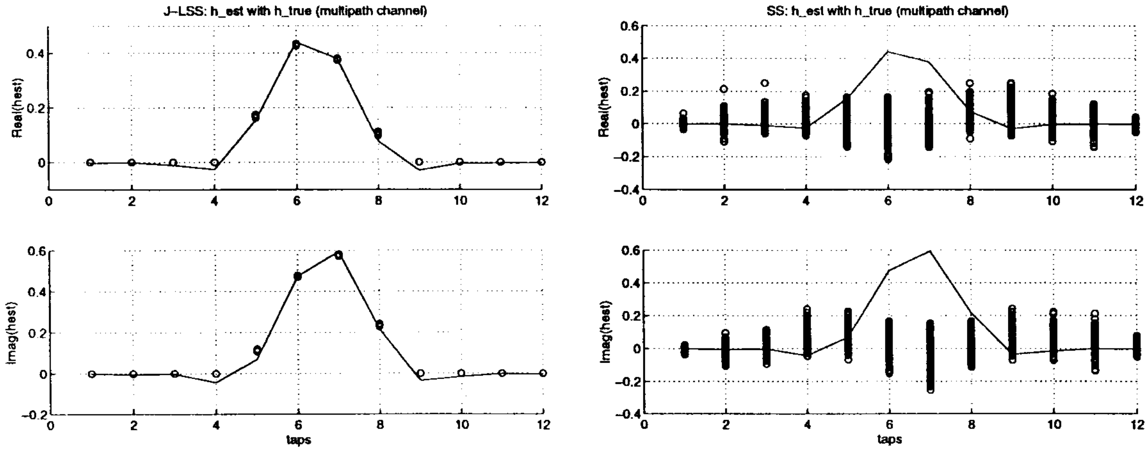

Fig. 6. Scatter-plots of 100 estimates at SNR = 30 dB. Solid lines: true channel. Left: J-LSS estimates. Right: SS estimates.

Observations and Discussions:

- J-LSS performs considerably better than the rest of the algorithms although its behavior is somewhat peculiar. Because the multipath channel has small head and tail taps, correct channel order detection is difficult. Consequently, J-LSS almost always underdetermines the channel order during the SNR range from 20–80 dB. On the other hand, it is perhaps not wise to estimate these small head and tail taps. Instead, it is better, as J-LSS apparently aims to do, to find the channel order as well as its impulse response that matches the data in some optimal way. Fig. 6 shows the scatterplot at 30 dB SNR of the magnitude response of the J-LSS and the SS algorithms. In this case, the J-LSS algorithm has detected channel order $\hat{L} = 1$ (rather than the true channel order $L = 5$). As we can see, the J-LSS algorithm captures the four major taps of the channel impulse response. In contrast, when the true channel order is used in the SS algorithm, the performance of the estimator is rather poor.
- It appears that CR, SS, and LSS perform comparably. Indeed, all of them assume knowledge of the channel order, and all have the finite sample convergence property, although this shows up only at relatively high SNR. From the implementation point of view, the advantage goes to the LSS algorithm that has a recursive implementation both in time and in channel order [17], [18].
- It is interesting to observe that when the channel order is correctly detected at high SNR, J-LSS is slightly worse than CR, SS, and LSS, although the difference eventually disappears as SNR $\rightarrow \infty$. This is due to the selection of $l > L$ in J-LSS, which reduces the effective sample size in the estimation.
- MSP performs better than LP-SS and LP-LS because it estimates the channel in a single step, whereas LP-SS and LP-LS estimate $\boldsymbol{h}_0$ first. For this multipath channel, the estimate of $\boldsymbol{h}_0$ is rather poor. MSP and LP-LS levels off as SNR $\rightarrow \infty$ because of the loss of finite sample convergence. The floor reduces as the number of samples increases. Note also that LP-SS does indicate finite sample convergence, although its breaking point occurs about 20 dB higher than that of CR, SS, and LSS.

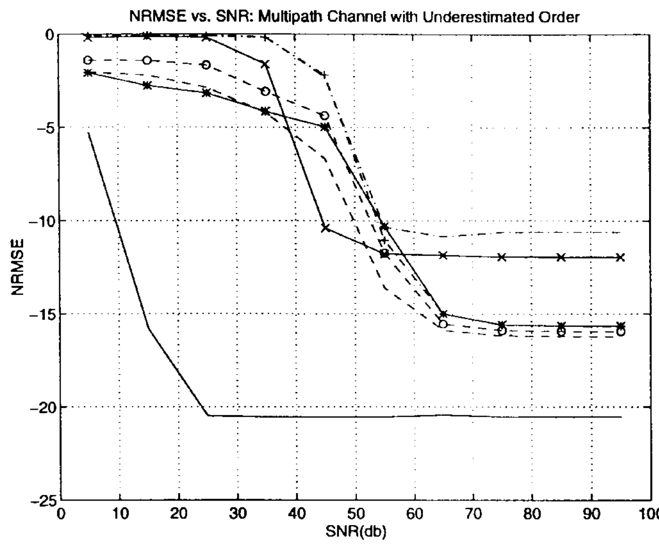

Fig. 7. NRMSE performance comparison for the multipath channel. One hundred Monte Carlo Runs. One hundred input symbols. Channel order under-determined by 1. Legend: SS: '−−'; CR: 'o'; LP-SS: '+'; LP-LS: '−.'; MSP: '×'; LSS: '*'; J-LSS: '−'.

- In deriving the algorithm, we have assumed that the smoothing window l is greater than the channel order. When this is not true, it is interesting to test the robustness of J-LSS. Fig. 7 shows the performance of these algorithms when the channel order is underestimated. In this simulation, the upper bound on the channel order used in J-LSS and the channel order used in all other algorithms are underestimated by one. We see that J-LSS performs better than all other algorithms throughout the entire SNR range. The flooring effect of all algorithms comes from the underdetermination of the channel order.

VI. CONCLUSION

We have presented a geometrical approach to the least squares smoothing algorithm for the blind estimation of multichannel finite-impulse response channels. The main idea arises from the isomorphic relationship between the input and output spaces, which serves as the basis of smoothing and linear prediction-based algorithms. The LSS approach to channel estimation preserves the finite sample convergence property critical to short data sample applications. The main attraction of the joint channel order detection and channel estimation algorithm is that it does not require knowledge of channel order and, at the same time, preserves the finite sample convergence property. There are, of course, several weakness of J-LSS. It requires a number of eigendecompositions that can be computationally expensive (it costs about L_u times more than the subspace algorithms where L_u is the upper bound of the channel order). For certain channels, the joint order detection and channel estimation approach may not be as effective as one that detect the channel order first and implement CR, SS, and especially LSS. In [17] and [18], we explore this strategy by developing time and order recursive schemes based on LSS.

APPENDIX

Proof of Property 3: Denote

$$\boldsymbol{S}_{k,t} \triangleq \begin{pmatrix} \boldsymbol{s}_t \\ \vdots \\ \boldsymbol{s}_{t-k+1} \end{pmatrix} = \begin{pmatrix} s_t & \cdots & s_{N-2w-l-1+t} \\ \vdots & \text{Toeplitz} & \\ s_{t-k+1} & & \end{pmatrix}. \quad (48)$$

When s_k has linear complexity greater than $2w+l+L$,

$$\text{rank}\{\boldsymbol{S}_{2w+l+1+L,2w+l+1}\} = 2w+l+L+1. \quad (49)$$

From (3), when there is no noise, we have

$$\boldsymbol{Z}_{w,l} = \mathcal{F}_{2w+l+1}(\boldsymbol{h})\boldsymbol{S}_{2w+l+1+L,2w+l+1} \quad (50)$$

$$\boldsymbol{F}_{w,l} = \mathcal{F}_w(\boldsymbol{h})\boldsymbol{S}_{w+L,2w+l+1}, \boldsymbol{P}_{w,l} = \mathcal{F}_w(\boldsymbol{h})\boldsymbol{S}_{w+L,w}. \quad (51)$$

Because of *P1.2)* and (49), we have (34)–(38).

To prove *P3.4)*, we note that

$$\boldsymbol{D}_{w,l} = \underbrace{\begin{pmatrix} \mathcal{F}_w(\boldsymbol{h}) & \boldsymbol{0}_{wP\times(w+L)} \\ \boldsymbol{0}_{wP\times(w+L)} & \mathcal{F}_w(\boldsymbol{h}) \end{pmatrix}}_{\boldsymbol{H}} \begin{pmatrix} \boldsymbol{S}_{w+L,2w+l+1} \\ \boldsymbol{S}_{w+L,w} \end{pmatrix}. \quad (52)$$

Under *A1)*, $\boldsymbol{H}$ has full column rank, and hence

$$\mathcal{R}\{\boldsymbol{D}_{w,l}\} = \mathcal{R}\left\{\begin{pmatrix} \boldsymbol{S}_{w+L,2w+l+1} \\ \boldsymbol{S}_{w+L,2w+l+1} \end{pmatrix}\right\}. \quad (53)$$

Because of (49), we have (40). □□□

Proof of Theorem 2: When channels do not share common zeros, it is sufficient to consider the case for $P=2$. Let

$$\boldsymbol{h}(z) = \begin{pmatrix} f(z) \\ g(z) \end{pmatrix}$$

where the two subchannels are $f(z) = \Sigma_{k=0}^{L} f_i z^{-k}, g(z) = \Sigma_{k=0}^{L} g_i z^{-k}$. Define

$$\mathcal{H}_l(\boldsymbol{f}) \triangleq \underbrace{\begin{pmatrix} f_0 & & \\ \vdots & \ddots & \\ f_L & & f_0 \\ & \ddots & \vdots \\ & & f_L \end{pmatrix}}_{l-L+1\text{columns}}, \boldsymbol{H}_{l,L} \triangleq \begin{pmatrix} \mathcal{H}_l(\boldsymbol{f}) \\ \mathcal{H}_l(\boldsymbol{g}) \end{pmatrix}. \quad (54)$$

For convenience, by rearranging rows of $\boldsymbol{E}_{l,l}$ we have, for $l > L$

$$\tilde{\boldsymbol{E}}_{l,l} = \begin{pmatrix} \mathcal{E}_{l,L}(\boldsymbol{f}) \\ \mathcal{E}_{l,L}(\boldsymbol{g}) \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_{t,l}} \\ \vdots \\ \tilde{\boldsymbol{s}}_{t+l-L|\dot{\mathcal{S}}_{t,l}} \end{pmatrix} = \boldsymbol{H}_{l,L} \begin{pmatrix} \tilde{\boldsymbol{s}}_{t|\dot{\mathcal{S}}_{t,l}} \\ \vdots \\ \tilde{\boldsymbol{s}}_{t+l-L|\dot{\mathcal{S}}_{t,l}} \end{pmatrix}. \quad (55)$$

We now consider the equivalent problem of finding $\boldsymbol{h}$ from the column space of $\boldsymbol{H}_{l,L}$.

Define

$$q(x) \triangleq f_0 + \cdots + f_L x^L + x^{l+1}(g_0 + \cdots + g_L x^L) \quad (56)$$

and let $\{z_1, \cdots, z_{l+L+1}\}$ be the $l+L+1$ distinct roots. Then, columns of the Vandermonde matrix $\mathcal{V}_{2l+2}(z_1^*, \cdots, z_{l+L+1}^*)$ form the orthogonal complement of the column space of $\boldsymbol{H}_{l,L}$. We now consider three separate cases:

I) $k = L$;
II) $L < k < l$;
III) $k < L$.

Case I–$k = L$: In this case, the full null space of $\boldsymbol{H}'_{l,L}$ is used. Constructing matrix $\boldsymbol{Q}$ from singular vectors is equivalent to that from $\mathcal{V}_{2l+2}(z_1^*, \cdots, z_{l+L+1}^*)$, whose columns span the null space of $\boldsymbol{H}'_{l,L}$. Therefore, solving $\boldsymbol{h}$ from (44) is equivalent to solving

$$\mathcal{V}_{2l+2}(z_1^*, \cdots, z_{l+L+1}^*)' \begin{pmatrix} \mathcal{H}_l(\boldsymbol{f}) \\ \mathcal{H}_l(\boldsymbol{g}) \end{pmatrix} = \boldsymbol{0} \quad (57)$$

which, after removing redundant equations, leads to solving the homogeneous equation (58), shown at the top of the next page. Because the roots $\{z_i\}$ are distinct, the solution of the above is unique.

Case II: In this case, matrix $\boldsymbol{Q}$ is constructed from the entire null space of $\boldsymbol{H}'_{l,L}$ along with $k-L$ eigenvectors in the range space of $\boldsymbol{H}_{l,L}$. In other words

$$\mathcal{T}_k(\boldsymbol{Q}) = \boldsymbol{G}\begin{pmatrix} \boldsymbol{T}_{l,k} \\ \boldsymbol{V} \end{pmatrix} \quad (59)$$

where

- $\boldsymbol{G}$ full-rank matrix;
- $\boldsymbol{T}_{k,l}$ corresponds to the matrix constructed from columns of $\mathcal{V}_{2l+2}(z_1^*, \cdots, z_{l+L+1}^*)$
- $\boldsymbol{V}$ matrix associated with the $(k-L)$ eigenvectors in the range space of $\boldsymbol{H}_{l,L}$.

$$\begin{pmatrix} 1 & z_1 & \cdots & z_1^L & z_1^{l+1} & z_1^{l+2} & \cdots & z_1^{l+L+1} \\ \vdots & & & & & & & \vdots \\ 1 & z_{l+L+1} & \cdots & z_{l+L+1}^L & z_{l+L+1}^{l+1} & z_{l+L+1}^{l+2} & \cdots & z_{l+L+1}^{l+L+1} \end{pmatrix} \begin{pmatrix} f_0 \\ \vdots \\ f_L \\ g_0 \\ \vdots \\ g_L \end{pmatrix} = \mathbf{0}. \tag{58}$$

$$\boldsymbol{T}_{l,k} \triangleq \begin{pmatrix} 1 & z_1 & \cdots & z_1^k & z_1^{l+1} & z_1^{l+2} & \cdots & z_1^{l+k+1} \\ \vdots & & & & & & & \vdots \\ 1 & z_{l+L+1} & \cdots & z_{l+L+1}^k & z_{l+L+1}^{l+1} & z_{l+L+1}^{l+2} & \cdots & z_{l+L+1}^{k+l+1} \end{pmatrix}. \tag{60}$$

$$[\boldsymbol{Q}_0, \cdots, \boldsymbol{Q}_k] = \boldsymbol{G} \begin{pmatrix} 1 & z_1 & \cdots & z_1^k & z_1^{l+1} & z_1^{l+2} & \cdots & z_1^{l+k+1} \\ \vdots & & & & & & & \vdots \\ 1 & z_{l+L+1} & \cdots & z_{l+L+1}^k & z_{l+L+1}^{l+1} & z_{l+L+1}^{l+2} & \cdots & z_{l+L+1}^{k+l+1} \end{pmatrix} = \boldsymbol{G}\boldsymbol{T}_{l,k}. \tag{64}$$

To show that (44) has only the trivial solution, it is sufficient to show that

$$\text{rank}\left\{ \begin{pmatrix} \boldsymbol{T}_{l,k} \\ \boldsymbol{V} \end{pmatrix} \right\} = 2k + 2.$$

Treating k as the estimated channel order, $\boldsymbol{T}_{l,k}$ has the form of (60), shown at the top of the page. We note that $\mathcal{V}_{l+L+1}(z_1^*, \cdots, z_{l+L+1}^*)$ has full column rank. Consequently, the first $k + L + 1$ columns of $\boldsymbol{T}_{l,k}$ are linearly independent. Because $\{z_i\}$ are distinct roots of $q(x)$, all other columns of $\boldsymbol{T}_{l,k}$ are linearly dependent on the first $k + L + 1$ columns

$$\text{rank}\{\boldsymbol{T}_{l,k}\} = k + 1 + L. \tag{61}$$

Next, when $k > L$, $k - L$ vectors from the range space of $\boldsymbol{H}_{l,L}$ are used to form $\boldsymbol{Q}$, where each vector introduces $l - k + 1$ rows in the $(k - L)(l - k + 1) \times 2(k + 1)$ matrix $\boldsymbol{V}$. For generic channels, these vectors are linearly independent among themselves and linearly independent to rows in $\boldsymbol{T}_{l,k}$, i.e.,

$$\begin{aligned} \text{rank}&\left\{ \begin{pmatrix} \boldsymbol{T}_{l,k} \\ \boldsymbol{V} \end{pmatrix} \right\} \\ &= \min\{2(k+1), k + 1 + L + (k - L)(l - k + 1)\} \\ &= 2k + 2. \end{aligned} \tag{62}$$

Therefore, $\text{rank}\{\mathcal{T}_k(\boldsymbol{Q})\} = 2k + 2$; hence, (44) has only the trivial solution.

Case III—$k < L$: In this case, again treating k as an estimated channel order, $l + k + 1$ null space vectors will be used in forming $\boldsymbol{Q}$. Since $\mathcal{V}_{2l+2}(z_1^*, z_2^*, \cdots, z_{l+L+1}^*)$ forms the orthogonal complement of the range space of $\boldsymbol{H}_{l,L}$, we have

$$\boldsymbol{Q} = \boldsymbol{G}\mathcal{V}_{2l+2}(z_1^*, z_2^*, \cdots, z_{l+L+1}^*)' \tag{63}$$

where $\boldsymbol{G}$ is a $(l + k + 1) \times (l + L + 1)$ matrix with full row rank. Since the singular vectors used to form $\boldsymbol{Q}$ are associated with the repeated zero singular value, $\boldsymbol{G}$ can be considered to be a randomly generated $(l+k+1) \times (l+L+1)$ matrix. Further, we have (64), shown at the top of the page. Because $k < L$, $\text{rank}\{\boldsymbol{T}_{l,k}\} = 2k+2$. For randomly generated $(l + k + 1) \times (l + L + 1)$ matrix $\boldsymbol{G}'$

$$\text{rank}\{[\boldsymbol{Q}_0, \cdots, \boldsymbol{Q}_k]\} = 2k + 2. \tag{65}$$

Therefore, (44) has only the trivial solution. □ □ □

REFERENCES

[1] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.

[2] R. E. Blahut, *Algebraic Methods for Signal Processing and Communications Coding*. New York: Springer-Verlag, 1992.

[3] A. K. Meraim *et al.*, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP Conf.*, Detroit, MI, May 1995, vol. 3, pp. 1960–1963.

[4] D. Gesbert and P. Duhamel, "Robust blind identification and equalization based on multi-step predictors," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, Munich Germany, Apr. 1997, vol. 5, pp. 2621–2624.

[5] M. L. Gürelli and C. L. Nikias, "EVAM: An eigenvector-based deconvolution of input colored signals," *IEEE Trans. Signal Processing*, vol. 43, pp. 134–149, Jan. 1995.

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

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

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

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

[10] D. Slock and C. B. Papadias, "Further results on blind identification and equalization of multiple FIR channels," in *IEEE Proc. Int. Conf. Acoust. Speech Signal Process.*, Detroit, MI, Apr. 1995, pp. 1964–1967.
[11] L. Tong and J. Bao, "Equalizations in wireless ATM," in *Proc. 1997 Allerton Conf. Commun., Contr. Comput.*, Urbana, IL, Oct. 1997, pp. 64–73.
[12] L. Tong and S. Perreau, "Multichannel blind channel estimation: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.
[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, pp. 340–349, Mar. 1994.
[14] L. Tong and Q. Zhao, "Blind channel estimation by least squares smoothing," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Seattle, WA, 1998, vol. IV, pp. 2121–2124.
[15] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.
[16] 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.
[17] Q. Zhao and L. Tong, "Adaptive blind channel estimation by least squares smoothing," *IEEE Trans. Signal Processing*, to be published.
[18] ——, "Adaptive blind channel estimation by least squares smoothing for CDMA," in *Proc. SPIE*, 1998.

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.

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, and was with the University of Connecticut, Storrs. Since the fall of 1998, he has been with the School of Electrical Engineering, Cornell University, Ithaca, NY, where he is an Associate Professor. 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.

Qing Zhao received the B.S. degree in electrical engineering in 1994 from Sichuan University, Chengdu, China, and the M.S. degree in 1997 from Fudan University, Shanghai, China. She is now pursuing the Ph.D. degree at the Department of Electrical Engineering, Cornell University, Ithaca, NY.

Her current research interests include wireless communications and array signal processing.