Adaptive Blind Channel Estimation by Least Squares Smoothing

Qing Zhao and Lang Tong, *Member, IEEE*

***Abstract*— A least squares smoothing (LSS) approach is presented for the blind estimation of single-input multiple-output (SIMO) finite impulse response systems. By exploiting the isomorphic relation between the input and output subspaces, this geometrical approach identifies the channel from a specially formed least squares smoothing error of the channel output. LSS has the finite sample convergence property, i.e., in the absence of noise, the channel is estimated perfectly with only a finite number of data samples. Referred to as the adaptive least squares smoothing (A-LSS) algorithm, the adaptive implementation has a high convergence rate and low computation cost with no matrix operations. A-LSS is order recursive and is implemented in part using a lattice filter. It has the advantage that when the channel order varies, channel estimates can be obtained without structural change of the implementation. For uncorrelated input sequence, the proposed algorithm performs direct deconvolution as a by-product.**

***Index Terms*— Adaptive least squares method, blind channel identification.**

I. INTRODUCTION

BLIND channel equalization has the potential to increase data throughput for communications over time varying channels. To achieve this goal, several requirements must be satisfied. First, blind channel estimation and equalization algorithms must converge quickly. An important property is the so-called *finite sample convergence* property, i.e., the channel can be perfectly estimated using a finite number of samples in the absence of noise. This property is especially critical in packet transmission systems where only a small number of data samples are available for processing. Second, the adaptivity of the algorithm is important in tracking the channel variation and maintaining the communication link. Third, low complexity in both computation and VLSI implementation is desired.

Although many blind channel estimation and equalization algorithms have been proposed in recent years, few can simultaneously satisfy requirements in convergence speed, adaptivity, and complexity. Deterministic batch algorithms such as the subspace (SS) algorithm [14], the cross relation (CR) algorithm[1] [5], [22], the two-step maximum likelihood (TSML) approach [7], the linear prediction-subspace (LP-SS) algorithm [16], and the joint order detection and channel estimation algorithm (J-LSS) [20] converge quickly. Without assuming a specific stochastic model of the input sequence, these methods have the finite sample convergence property. Unfortunately, they suffer from high computation cost, which is usually associated with eigenvalue decomposition, and their adaptive implementations are often cumbersome. On the other hand, recently proposed linear prediction (LP) based algorithms [1], [4] are attractive for their efficient adaptive implementations. The key component of these algorithms is the classical linear predictor that can be implemented recursively both in time and in filter order using, for example, lattice filters. The modular structure of lattice filters makes them good candidates for VLSI implementation. Perhaps the most important shortcoming of these LP-based algorithms is the relatively low convergence speed. Relying on the statistical uncorrelation of the input sequence, these LP-based algorithms fail to have the finite sample convergence property. Consequently, they demand a relatively large sample size for accurate channel estimation, which limits their application in the small data size scenarios.

Aiming to satisfy the three design requirements at the same time, we present in this paper a *least squares smoothing* (LSS) approach to blind channel estimation. Recognizing the isomorphic relation between the input and output subspaces, we first consider estimating the channel from the input *subspace*. By projecting the channel output into a particular input subspace $\mathcal{Z}$, the channel is obtained from the least squares projection error. When $\mathcal{Z}$ is constructed from the channel output by exploiting the isomorphism between the input and output subspaces, this projection leads to least squares smoothing. This geometrical approach to blind channel estimation also provides a simple and unified derivation of different LP-based channel estimators and a clear explanation for the loss of finite-sample convergence property in LP-based approaches.

Based on the LSS approach, a new *adaptive least squares smoothing* (A-LSS) channel estimation algorithm is proposed. Like all deterministic methods, A-LSS preserves the finite sample convergence property with high convergence rate. Similar to linear prediction-based algorithms, the least squares smoothing approach naturally leads to an order and time adaptive implementation that accommodates a wide range of channel variation, both in channel parameters and channel length. In wireless communications, for example, when the channel order suddenly changes due to the addition or loss of reflectors, A-LSS simply selects signals from different parts of

Manuscript received June 2, 1998; revised May 11, 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 Dr. Phillip A. Regalia.

The authors are with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ee.cornell.edu).

Publisher Item Identifier S 1053-587X(99)08320-8.

[1] This is also referred to as the least squares algorithm, or EVAM.

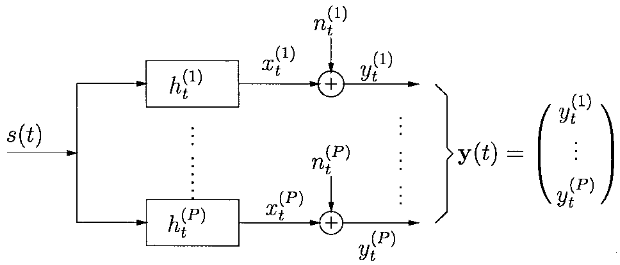

Fig. 1. Single-input P-output linear system.

the filter with neither structural change nor extra computation. Implemented with commonly used basic modules in classical lattice filters, the structure of A-LSS is highly parallel and regular with only scalar operations.

This paper is organized as follows. Section II presents the system model and two assumptions used in this paper. The geometrical approach to linear least squares smoothing channel estimation is introduced in Section III. A batch least squares smoothing (B-LSS) algorithm and its connections with existing linear prediction-based approaches are obtained. In Section IV, we present the data structure for the unknown channel order case followed by the adaptive least squares smoothing channel estimation and order detection algorithm. Simulation results are presented in Section V to demonstrate the convergence and performance of A-LSS in channel estimation and tracking.

II. Problem Statement

A. Notations and Definitions

Notations used in this paper are mostly standard. Uppercase and lowercase bold letters denote matrices and vectors with $(\cdot)^t$, $(\cdot)'$, and $(\cdot)^\dagger$ denoting transpose, Hermitian, and Moore–Penrose pseudoinverse operations, respectively. Given a matrix $\mathbf{A}$, $\mathcal{R}\{\mathbf{A}\}$ ($\mathcal{C}\{\mathbf{A}\}$) is the row (column) space of $\mathbf{A}$. For a matrix $\mathbf{X}$ having the same number of columns as $\mathbf{A}$, $\mathcal{P}_{\mathbf{A}}\{\mathbf{X}\} \triangleq \mathbf{X}\mathbf{A}'(\mathbf{A}\mathbf{A}')^\dagger\mathbf{A}$ ($\mathcal{P}_{\mathbf{A}}^{\perp}\{\mathbf{X}\} \triangleq \mathbf{X} - \mathcal{P}_{\mathbf{A}}\{\mathbf{X}\}$) is the projection of $\mathbf{X}$ into (the orthogonal complement of) the row space of $\mathbf{A}$. We define $\mathcal{E}_{\mathbf{A}}^{\mathbf{X}} \triangleq \mathcal{P}_{\mathbf{A}}^{\perp}\{\mathbf{X}\}$ as the projection error matrix of $\mathbf{X}$ into $\mathcal{R}\{\mathbf{A}\}$. For a set of vectors $\mathbf{x}_1, \cdots, \mathbf{x}_n$, $sp\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ denotes the linear subspace spanned by $\mathbf{x}_1, \cdots, \mathbf{x}_n$. For a vector $\mathbf{x}$ and a linear subspace $\mathcal{X}$, $\hat{\mathbf{x}}_{|\mathcal{X}} \triangleq \arg\min_{\mathbf{z}\in\mathcal{X}} ||\mathbf{x}-\mathbf{z}||^2$ is the orthogonal projection of $\mathbf{x}$ into $\mathcal{X}$, and $\tilde{\mathbf{x}}_{|\mathcal{X}} \triangleq \mathbf{x}-\hat{\mathbf{x}}_{|\mathcal{X}}$ is its projection error. Finally, $||\cdot||$ and $||\cdot||_F$ denote the 2-norm and Frobenius norm, respectively.

B. The Model

The identification and estimation of a single-input P-output linear system shown in Fig. 1 is considered in this paper. The system equation is given by

$$\mathbf{x}(t) = \sum_{i=0}^{L} \mathbf{h}_i s(t-i), \quad \mathbf{y}(t) = \mathbf{x}(t) + \mathbf{n}(t)$$
$$t = 1, 2, \cdots \tag{1}$$

where $\mathbf{x}(t) = [x_t^{(1)}, \cdots, x_t^{(P)}]^t$ is the (noiseless) channel output, $\mathbf{n}(t)$ is the additive noise, $\mathbf{y}(t)$ is the received signal, $\{\mathbf{h}_t\}$ is the channel impulse response, and $\{s(t)\}$ is the input sequence. Given N samples of the system input and output, define the row vector $\mathbf{s}_t$ and the block row vector $\mathbf{x}_t$ as

$$\begin{aligned}\mathbf{s}_t &\triangleq [s(t), s(t+1), \cdots, s(t+N-1)] \\ \mathbf{x}_t &\triangleq [\mathbf{x}(t), \mathbf{x}(t+1), \cdots, \mathbf{x}(t+N-1)].\end{aligned} \tag{2}$$

With $\mathbf{y}_t$ and $\mathbf{n}_t$ similarly defined, we have, from (1)

$$\mathbf{x}_t = \sum_{i=0}^{L} \mathbf{h}_i \mathbf{s}_{t-i}, \quad \mathbf{y}_t = \mathbf{x}_t + \mathbf{n}_t. \tag{3}$$

Our goal is to estimate $\mathbf{h} \triangleq [\mathbf{h}_L', \cdots, \mathbf{h}_0']'$ from $\mathbf{y}_t$.

C. Assumptions and Properties

Two assumptions are made in this paper. The first one is about channel diversity.

- A1: *Channel Diversity*: The subchannel transfer functions do not share common zeros, i.e., $\{h^{(i)}(z)\}$ are co-prime.

A1 is shared by all deterministic blind channel estimation methods. The co-primeness of the subchannel transfer functions ensures that the channel is fully specified (up to a scaling factor) by the noncommon zeros of the channel output. If A1 is not satisfied, although the noncommon zeros of the subchannel transfer functions can still be identified from the output, the common zeros cannot be distinguished from zeros of the input. Therefore, the channel cannot be identified without knowing the input sequence. The following property, which has also been exploited in [13] and [21], reveals the equivalence between the input and output subspaces under the channel diversity assumption. First, define the input and output subspaces spanned by $p > 0$ consecutive row (block row) vectors as

$$\begin{aligned}\mathcal{S}_{t,p} &\triangleq sp\{\mathbf{s}_t, \cdots, \mathbf{s}_{t-p+1}\} \\ \mathcal{X}_{t,p} &\triangleq sp\{\mathbf{x}_t, \cdots, \mathbf{x}_{t-p+1}\}.\end{aligned} \tag{4}$$

Note that for $p < 0$, we similarly define $\mathcal{S}_{t,p}$ and $\mathcal{X}_{t,p}$ as the subspaces spanned by $|p|$ consecutive future data vectors. The following property results directly from A1.

Property 1: Under A1, there exists a (smallest) $w_o \geq (L/P - 1)$ such that for any $w \geq w_0$, we have the isomorphic relation between the input and output subspaces

$$\mathcal{X}_{t,w} = \mathcal{S}_{t,w+L}. \tag{5}$$

Proof: Let

$$\begin{aligned}\mathbf{S}_{t,w} &\triangleq [\mathbf{s}_t^t, \cdots, \mathbf{s}_{t-w+1}^t]^t \\ \mathbf{X}_{t,w} &\triangleq [\mathbf{x}_t^t, \cdots, \mathbf{x}_{t-w+1}^t]^t.\end{aligned} \tag{6}$$

From (3), we have

$$\mathbf{X}_{t,w} = \mathcal{F}_w(\mathbf{h})\mathbf{S}_{t,w+L} \tag{7}$$

where the $wP \times (w+L)$ matrix $\mathcal{F}_w(\mathbf{h})$ is the *filtering matrix* with the following block Toeplitz structure:

$$\mathcal{F}_w(\mathbf{h}) \triangleq \begin{pmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_L & & \\ & \ddots & & \ddots & \\ & & \mathbf{h}_0 & \cdots & \mathbf{h}_L \end{pmatrix}. \tag{8}$$

It has been shown in [9] and [19] that A1 implies the existence of a smallest $w_o \geq (L/P - 1)$ such that for any $w \geq w_o$, $\mathcal{F}_w(\mathbf{h})$ has full column rank. Thus, from (7), we have

$$\mathcal{R}\{\mathbf{X}_{t,\,w}\} = \mathcal{R}\{\mathbf{S}_{t,\,w+L}\} \tag{9}$$

which leads to (5). □□□

Property 1 plays an important role in the smoothing and linear prediction approaches to blind channel estimation. It is this isomorphic relation between the input and output subspaces that makes it possible to identify the channel from the input *subspace* without the direct use of the input sequence.

To ensure channel identifiability, the input sequence must be sufficiently complex in order to excite all modes of the channel. This requirement is imposed on the linear complexity [2] of the sequence.

- A2: *Linear Complexity*: The input sequence $\{s(t)\}$ has linear complexity greater than $L_* = 2w_0 + 2L$.

With the definition of linear complexity[2] [2], A2 implies that for any $N \geq 2L_* - L + 1$, we have

$$\text{rank}\left\{\begin{pmatrix} s(L_* - L + 1) & \cdots & s(N) \\ \vdots & \text{Toeplitz} & \\ s(1-L) & & \end{pmatrix}\right\} = L_* + 1. \tag{10}$$

This implication of A2 will be exploited in Section III as we discuss the necessary number of input symbols for the channel identification by the proposed algorithm. It has been shown in [18] that when $P = 2$, the *necessary and sufficient* condition for the unique identification of the channel and its input is A1 and that $\{s(t)\}$ has linear complexity greater than $2L$. Because both future and past data are required in the smoothing approach, a stronger condition is assumed in A2 in this paper.

III. Least Squares Smoothing—A Geometrical Approach

In this section, we introduce a geometrical approach to linear least squares smoothing channel estimation by exploiting the isomorphic relation between the input and output subspaces. The same approach also leads to a least squares derivation of the LP-based algorithms.

A. Channel Identification from the Input Subspace

The isomorphism between the input and output subspaces given in Property 1 implies that the input subspace can be constructed from the output. The following question arises from this property: *Can we identify the channel from the input subspace?* If so, by constructing the input subspace from the output, the channel is obtained from the output alone. An answer to this question is presented below.

[2]The linear complexity of the vector $\mathbf{v}$ with components $v_0, \cdots, v_{n-1}$ is defined as the smallest value of c for which a recursion of the form $v_i = -\sum_{j=1}^{c} a_j v_{i-j} (i = c, \cdots, n-1)$ exists that will generate $\mathbf{v}$ from its first c components.

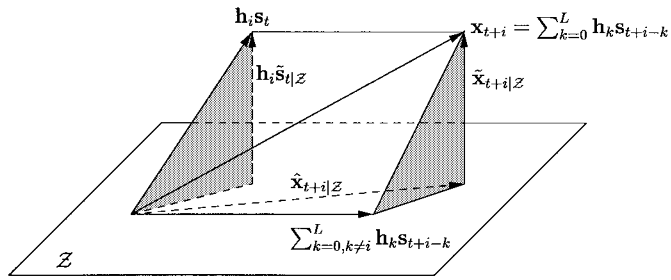

Fig. 2. Projection of $\mathbf{x}_{t+i} (i = 0, \cdots, L)$ onto $\mathcal{Z}$.

Consider $L + 1$ consecutive output block row vectors $\mathbf{x}_{t+L}, \cdots, \mathbf{x}_t$. From (3), we have

$$\begin{aligned} \mathbf{x}_{t+L} &= \mathbf{h}_0\mathbf{s}_{t+L} + \mathbf{h}_1\mathbf{s}_{t+L-1} + \cdots + \underline{\mathbf{h}_L\mathbf{s}_t} \\ \mathbf{x}_{t+L-1} &= \qquad \mathbf{h}_0\mathbf{s}_{t+L-1} + \cdots + \underline{\mathbf{h}_{L-1}\mathbf{s}_t} + \mathbf{h}_L\mathbf{s}_{t-1} \\ &\vdots \qquad\qquad \ddots \qquad\qquad \ddots \\ \mathbf{x}_t &= \qquad\qquad \underline{\mathbf{h}_0\mathbf{s}_t} + \cdots + \mathbf{h}_L\mathbf{s}_{t-L}. \end{aligned} \tag{11}$$

Suppose that we want to identify $\mathbf{h}_L, \cdots, \mathbf{h}_0$ up to a common scaling factor from $\mathbf{x}_{t+L}, \cdots, \mathbf{x}_t$. One way to achieve this is to eliminate in $\mathbf{x}_{t+L}, \cdots, \mathbf{x}_t$ all other terms except the ones associated with $\mathbf{s}_t$. Consider projecting $\mathbf{x}_{t+i}$ $(i = 0, \cdots, L)$ into a "punctured" input subspace $\mathcal{Z}$ that satisfies the following two conditions:

- C1: $\{\mathbf{s}_{t+L}, \cdots, \mathbf{s}_{t+1}, \mathbf{s}_{t-1}, \cdots, \mathbf{s}_{t-L}\} \subset \mathcal{Z}$.
- C2: $\mathbf{s}_t \notin \mathcal{Z}$.

Note that A2 ensures $\mathbf{s}_t \notin sp\{\mathbf{s}_{t+L}, \cdots, \mathbf{s}_{t+1}, \mathbf{s}_{t-1}, \cdots, \mathbf{s}_{t-L}\}$. Thus, this punctured input subspace $\mathcal{Z}$ exists. As illustrated in Fig. 2, $\mathbf{x}_{t+i}$ is the summation of $\mathbf{h}_i\mathbf{s}_t$ (a vector outside $\mathcal{Z}$) and $\sum_{k=0,\,k\neq i}^{L} \mathbf{h}_k\mathbf{s}_{t+i-k}$ (a vector inside $\mathcal{Z}$). The two shaded right triangles immediately suggest

$$\tilde{\mathbf{x}}_{t+i|\mathcal{Z}} = \mathbf{h}_i\tilde{\mathbf{s}}_{t|\mathcal{Z}}. \tag{12}$$

Because $\tilde{\mathbf{s}}_{t|\mathcal{Z}}$ is independent of i, we have the projection error matrix

$$\mathbf{E} \triangleq \begin{pmatrix} \tilde{\mathbf{x}}_{t+L|\mathcal{Z}} \\ \vdots \\ \tilde{\mathbf{x}}_{t|\mathcal{Z}} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_L \\ \vdots \\ \mathbf{h}_0 \end{pmatrix} \tilde{\mathbf{s}}_{t|\mathcal{Z}} = \mathbf{h}\tilde{\mathbf{s}}_{t|\mathcal{Z}}. \tag{13}$$

Note that $\mathbf{E}$ is a rank-1 matrix whose column and row spaces are spanned by $\mathbf{h}$ and $\tilde{\mathbf{s}}_{t|\mathcal{Z}}$, respectively. From $\mathbf{E}$, the channel $\mathbf{h}$ can be directly identified. One approach is the least squares fitting of the column space of $\mathbf{E}$

$$\hat{\mathbf{h}} = \arg \max_{||\mathbf{h}||=1} ||\mathbf{h}'\mathbf{E}||^2. \tag{14}$$

The above optimization can be solved by the singular value decomposition of either $\mathbf{E}$ or the sample covariance of the projection error sequence. A simpler approach is to take the row $(\mathbf{e}_m)$ with the maximum 2-norm in $\mathbf{E}$ as an estimate of $\tilde{\mathbf{s}}_{t|\mathcal{Z}}$. Then, from (13), we have

$$\hat{\mathbf{h}} = \mathbf{E}\mathbf{e}_m' = ||\tilde{\mathbf{s}}_{t|\mathcal{Z}}||^2\mathbf{h}. \tag{15}$$

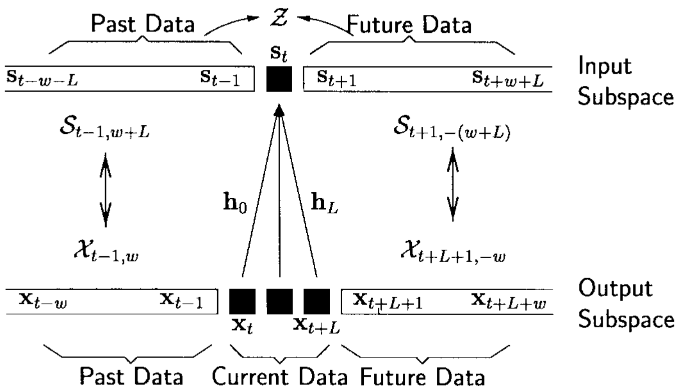

Fig. 3. Isomorphism between input and output subspaces.

To gain a better understanding of this geometrical approach to channel estimation, we make the following remarks.

1) When $\mathbf{s}_t$ is orthogonal to $\mathcal{Z}$, which is asymptotically true for an uncorrelated zero-mean input sequence, the projection error $\tilde{\mathbf{s}}_{t|\mathcal{Z}}$ converges to $\mathbf{s}_t$ (see Fig. 2). In this case, the rank-one decomposition of the projection error matrix $\mathbf{E}$ provides the input sequence $\mathbf{s}_t$ as well as the channel vector $\mathbf{h}$, i.e., this geometrical approach to channel estimation performs direct deconvolution as a by-product.
2) The projections of $\mathbf{x}_{t+i}$ into $\mathcal{Z}$ for different i are independent of each other and can be carried out in parallel. This property is attractive in algorithm implementation.

B. Channel Identification from the Output Subspace—Least Squares Smoothing

1) Construction of the Input Subspace from the Output: It has been shown in Section III-A that the channel can be identified from the projection errors of $\mathbf{x}_{t+L}, \cdots, \mathbf{x}_t$ into the input subspace $\mathcal{Z}$ satisfying C1 and C2. To avoid the direct use of the input sequence, we construct $\mathcal{Z}$ from the channel output by exploiting the isomorphic relation between the input and output subspaces given in Property 1.

Using the definition in (4), conditions C1 and C2 can be written in the form

$$\mathcal{Z} = \mathcal{S}_{t-1,\,p} \cup \mathcal{S}_{t+1,\,-p} \tag{16}$$

for any $p \geq L$. With the isomorphic relation between the input and output subspaces given in (5), we have, for $w = p - L \geq w_0$

$$\mathcal{Z} = \mathcal{S}_{t-1,\,w+L} \cup \mathcal{S}_{t+1,\,-(w+L)} = \mathcal{X}_{t-1,\,w} \cup \mathcal{X}_{t+L+1,\,-w}. \tag{17}$$

The isomorphism between the input and output subspaces is illustrated in Fig. 3. The projection of $\mathbf{x}_{t+L}, \cdots, \mathbf{x}_t$ into $\mathcal{Z}$ is converted into the smoothing [6], [8] of the current data $\mathbf{x}_{t+L}, \cdots, \mathbf{x}_t$ by the past data $\mathbf{x}_{t-1}, \cdots, \mathbf{x}_{t-w}$ and the future data $\mathbf{x}_{t+L+1}, \cdots, \mathbf{x}_{t+L+w}$. The projection error $\mathbf{E}$ used to identify the channel [see (13)] now becomes the smoothing error, which can be obtained from the output alone. Because of the least squares criterion used in the projection, this geometrical approach to blind channel estimation is referred to as the *least squares smoothing* (LSS) algorithm. Here, we introduce two terms: *smoothing window size* and *smoothing order*. The smoothing window size is defined as the number of symbols in the current data, and the smoothing order is the number of symbols in the past data. For the case discussed above, the smoothing window size and the smoothing order are $L+1$ and w, respectively.

The price we paid for avoiding the direct use of input sequence is that more input symbols are required to identify the channel. From Fig. 3, we can see that the projection subspace $\mathcal{Z}$ and the current data span a $(2w+2L+1)$-dimensional input subspace denoted as $\mathcal{V}$, as shown in

$$\mathcal{V} = sp\{\mathbf{s}_{t-w-L}, \mathbf{s}_{t-w-L+1}, \cdots, \mathbf{s}_{t+w+L}\} \tag{18}$$

$$= \mathcal{R}\left\{\begin{pmatrix} s(t+w+L) & s(t+w+L+1) & \cdots & s(N) \\ \vdots & \text{Toeplitz} & & \\ s(t-w-L) & & & \end{pmatrix}\right\}. \tag{19}$$

The construction of $\mathcal{V}$ imposes the full-row-rank condition on the matrix in (19). As a result, we require $4w+4L+1$ input symbols ($4w+3L+1$ observation symbols) to be available and the linear complexity of the input sequence to be greater than $2w+2L$ [see (10)].

2) Batch Least Squares Smoothing Algorithm: We consider here the problem of estimating a channel with order L from a batch of N observation symbols. The data structure is specified, followed by some useful properties. Then, the batch least squares smoothing channel estimation algorithm is presented.

Given $N \geq 4w+3L+1$ observation symbols $\mathbf{y}(t)$, $t = 1, \cdots, N$, for a fixed smoothing order w and the known channel order L, define the overall data matrix as

$$\mathbf{V} \triangleq \left(\begin{array}{cc} \mathbf{y}(2w+L+1) & \cdots \quad \mathbf{y}(N) \\ \vdots & \mathbf{Z}_f \\ \mathbf{y}(w+L+2) & \\ \hline \mathbf{y}(w+L+1) & \cdots \\ \vdots & \mathbf{D} \\ \mathbf{y}(w+1) & \\ \hline \mathbf{y}(w) & \cdots \\ \vdots & \mathbf{Z}_p \\ \mathbf{y}(1) & \end{array}\right) \tag{20}$$

from which we have specified

- the past data matrix $\mathbf{Z}_p \triangleq [\mathbf{y}_w^t, \cdots, \mathbf{y}_1^t]^t$;
- the current data matrix $\mathbf{D} \triangleq [\mathbf{y}_{w+L+1}^t, \cdots, \mathbf{y}_{w+1}^t]^t$;
- the future data matrix $\mathbf{Z}_f \triangleq [\mathbf{y}_{2w+L+1}^t, \cdots, \mathbf{y}_{w+L+2}^t]^t$;
- future-past data matrix $\mathbf{Z} \triangleq \binom{\mathbf{Z}_f}{\mathbf{Z}_p}$.

In the absence of noise, the overall data matrix $\mathbf{V}$ has the following relation with the input signal:

$$\mathbf{V} = \mathcal{F}_{2w+L+1}(\mathbf{h})\begin{pmatrix} s(2w+L+1) & \cdots & s(N) \\ \vdots & \text{Toeplitz} & \\ s(1-L) & & \end{pmatrix}. \tag{21}$$

Least Squares Smoothing Channel Estimation

1. Choose $w \geq w_0 \geq \frac{L}{P-1}$, form data matrices $\mathbf{Z}$ and $\mathbf{D}$.
2. Obtain the orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_{2(w+L)}\}$ that spans the $2(w+L)$-dimensional signal row space of $\mathbf{Z}$.
3. Obtain the projection error of $\mathbf{D}$ into $sp\{\mathbf{u}_1, \cdots, \mathbf{u}_{2(w+L)}\}$:

$$\mathbf{E} \triangleq \mathbf{D} - \mathbf{D}\mathbf{U}'\mathbf{U}, \quad \mathbf{U} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{2(w+L)} \end{pmatrix}.$$

4. Identify the channel from the column space of the projection error $\mathbf{E}$:

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

Fig. 4. Batch LSS algorithm.

As direct consequences of (21) and Property 1, the relation among the data matrices in (20) and various subspaces is summarized in Property 2. The rank conditions given below are useful in finding the least squares approximation of the noisy data matrices.

Property 2: Suppose that the input sequence has linear complexity greater than $2(w+L)$, $w \geq w_0$, $N \geq 4w+3L+1$. We have the following properties in the noiseless case:

1) *Overall Data Matrix* $\mathbf{V}$:

$$\mathcal{R}\{\mathbf{V}\} = \mathcal{X}_{2w+L+1,\, 2w+L+1} = \mathcal{S}_{2w+L+1,\, 2w+2L+1}$$
$$\text{rank}(\mathbf{V}) = 2w + 2L + 1.$$

2) *Past Data Matrix* $\mathbf{Z}_p$:

$$\mathcal{R}\{\mathbf{Z}_p\} = \mathcal{X}_{w,\, w} = \mathcal{S}_{w,\, w+L}$$
$$\text{rank}(\mathbf{Z}_p) = w + L.$$

3) *Future Data Matrix* $\mathbf{Z}_f$:

$$\mathcal{R}\{\mathbf{Z}_f\} = \mathcal{X}_{w+L+2,\, -w} = \mathcal{S}_{w+2,\, -(w+L)}$$
$$\text{rank}(\mathbf{Z}_f) = \text{rank}(\mathbf{Z}_p) = w + L.$$

4) *Future-Past Data Matrix* $\mathbf{Z}$:

$$\mathcal{R}\{\mathbf{Z}\} = \mathcal{X}_{w,\, w} \bigcup \mathcal{X}_{w+L+2,\, -w} = \mathcal{S}_{w,\, w+L} \bigcup \mathcal{S}_{w+2,\, -(w+L)}$$
$$\text{rank}(\mathbf{Z}) = 2(w + L).$$

As stated in the above property, the row vectors in the future-past data matrix $\mathbf{Z}$ span the $2(w+L)$-dimensional projection space $\mathcal{Z}$. Therefore, the projection error $\mathbf{E}$ in (13) can be obtained as the least squares smoothing error of the current data $\mathbf{D}$ by the future-past data $\mathbf{Z}$. Specifically, in the absence of noise, we have

$$\mathbf{E} \triangleq \mathcal{E}_{\mathbf{Z}}^{\mathbf{D}} = \begin{pmatrix} \tilde{\mathbf{y}}_{w+L+1|\mathcal{Z}} \\ \vdots \\ \tilde{\mathbf{y}}_{w+1|\mathcal{Z}} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_L \\ \vdots \\ \mathbf{h}_0 \end{pmatrix} \tilde{\mathbf{s}}_{w+1|\mathcal{Z}} = \mathbf{h}\tilde{\mathbf{s}}_{w+1|\mathcal{Z}} \tag{22}$$

from which the batch least squares smoothing channel estimation algorithm (B-LSS) can be derived directly. One possible implementation is summarized in Fig. 4.

3) Stochastic Equivalence: The least squares smoothing approach presented above is based on the deterministic modeling of the input sequence. In order to obtain the consistent estimate of the channel in the presence of noise, we consider a stochastic equivalence of the LSS approach. After replacing the projection of data vectors $(\mathbf{x}_t, \mathbf{s}_t)$ into a Euclidean space by the projection of corresponding random variables $(\mathbf{x}(t), s(t))$ into a Hilbert space, the geometrical approach presented in Section III-B1 can be adapted to the stochastic framework. Specifically, consider the input sequence $\{s(t)\}$ as a white random process with zero mean and unit variance. Define the output random vectors

$$\begin{aligned} \mathbf{x}_c(t) &\triangleq [\mathbf{x}'(t+L), \cdots, \mathbf{x}'(t)]' \\ \mathbf{x}_{fp}(t) &\triangleq [\mathbf{x}'(t+L+w), \cdots, \mathbf{x}'(t+L+1), \mathbf{x}'(t-1) \\ &\qquad \cdots, \mathbf{x}'(t-w)]'. \end{aligned} \tag{23}$$

The error sequence $\mathbf{e}(t)$ of the linear minimum mean square error (MMSE) estimator of $\mathbf{x}_c(t)$ based on $\mathbf{x}_{fp}(t)$ is then given by

$$\mathbf{e}(t) = \mathbf{x}_c(t) - \mathbf{R}_{c,\, fp}\mathbf{R}_{fp}^{\dagger}\mathbf{x}_{fp}(t) \tag{24}$$

where $\mathbf{R}_{fp}$ and $\mathbf{R}_{c,\, fp}$ are the covariance matrices defined as

$$\mathbf{R}_{fp} \triangleq E\{\mathbf{x}_{fp}(t)\mathbf{x}_{fp}'(t)\}, \, \mathbf{R}_{c,\, fp} \triangleq E\{\mathbf{x}_c(t)\mathbf{x}_{fp}'(t)\}. \tag{25}$$

Based on the principle of orthogonality, the error sequence $\mathbf{e}(t)$ can be considered as the projection error of $\mathbf{x}_c(t)$ into the Hilbert space spanned by the random variables in $\mathbf{x}_{fp}(t)$. Similarly to (13), we then have, for the white input sequence $\{s(t)\}$,

$$\mathbf{e}(t) = \mathbf{h}s(t). \tag{26}$$

With $\mathbf{R}_c$ and $\mathbf{R}_e$ denoting the covariance of $\mathbf{x}_c(t)$ and $\mathbf{e}(t)$, respectively, we have, from (24) and (26)

$$\mathbf{R}_e = \mathbf{R}_c - \mathbf{R}_{c,\, fp}\mathbf{R}_{fp}^{\dagger}\mathbf{R}_{c,\, fp}' = \mathbf{h}\mathbf{h}'. \tag{27}$$

This implies that $\mathbf{h}$ can be directly obtained from $\mathbf{R}_e$. For a noise sequence independent of $\{s(t)\}$ and with known second-order statistics, the covariance matrices $\mathbf{R}_c$, $\mathbf{R}_{fp}$, and $\mathbf{R}_{c,\, fp}$ can be estimated consistently from the observation. The consistent estimate of the channel can then be obtained.

C. Connections with Linear Prediction-Based Approaches

The main difference between the LP-based approaches and the LSS approach is the definition of the subspace $\mathcal{Z}$ into which the projection is made. Naturally, LP-based approaches include only the past data in the projection space. We now draw connections with existing LP-based methods by using the same geometrical approach to rederive them under the deterministic model of the input sequence. We comment that the LP-based algorithms presented here are not identical to the stochastic versions presented in the literature, although they are closely related. LP-based channel estimators can be classified into one-step and multistep linear prediction algorithms.

1) One-Step LP Algorithm: We consider here the LP-based approach proposed by Abed-Meraim *et al.* in [1] and [3] to which we refer as the linear prediction-least squares (LP-LS) algorithm. Instead of projecting $\mathbf{x}_{t+i}$ into $\mathcal{Z} = \mathcal{X}_{t-1,\,w} \bigcup \mathcal{X}_{t+L+1,\,-w}$, let us consider projecting $\mathbf{x}_t$ into $\mathcal{X}_{t-1,\,w}$, which contains only the past data. We then have, again using the isomorphic relation $\mathcal{X}_{t-1,\,w} = \mathcal{S}_{t-1,\,w+L}$

$$\tilde{\mathbf{x}}_{t|\mathcal{X}_{t-1,\,w}} = \mathbf{h}_0 \tilde{\mathbf{s}}_{t|\mathcal{S}_{t-1,\,w+L}} \tag{28}$$

from which $\tilde{\mathbf{s}}_{t|\mathcal{S}_{t-1,\,w+L}}$ can be obtained directly. By treating $\tilde{\mathbf{s}}_{t|\mathcal{S}_{t-1,\,w+L}}$ as an estimate of $\mathbf{s}_t$, the problem becomes conventional channel estimation with known channel input. The least squares criterion can then be employed to estimate the whole channel, which is the reason that we refer to this one-step LP algorithm as LP-LS. Note that if $\mathbf{s}_t \perp \mathcal{S}_{t-1,\,w+L}$, which is asymptotically true for uncorrelated zero-mean sequence, we have $\tilde{\mathbf{s}}_{t|\mathcal{S}_{t-1,\,w+L}} = \mathbf{s}_t$ (see also Fig. 2). Hence, this approach provides consistent estimates for white inputs. Unfortunately, for a finite sample size, $\tilde{\mathbf{s}}_{t|\mathcal{S}_{t-1,\,w+L}} \neq \mathbf{s}_t$, which causes the loss of the finite sample convergence property in the LP-LS approach.

3) Multistep LP Algorithm: In contrast to the LP-LS approach, the multistep LP approach proposed by Gesbert and Duhamel [4] works simultaneously on predictions of several future data samples. By projecting $\mathbf{x}_{t+i}$ into $\mathcal{X}_{t-1,\,w}$, we have the prediction error equations

$$\tilde{\mathbf{x}}_{t+i|\mathcal{X}_{t-1,\,w}} = \mathbf{h}_0 \tilde{\mathbf{s}}_{t+i|\mathcal{S}_{t-1,\,w+L}} + \cdots + \mathbf{h}_i \tilde{\mathbf{s}}_{t|\mathcal{S}_{t-1,\,w+L}} \quad i = 0, 1, \cdots, L. \tag{29}$$

For the uncorrelated zero-mean input sequence, we have, asymptotically

$$\tilde{\mathbf{x}}_{t+i|\mathcal{X}_{t-1,\,w}} \overset{N\to\infty}{=} \mathbf{h}_0 \mathbf{s}_{t+i} + \cdots + \mathbf{h}_i \mathbf{s}_t \triangleq \mathbf{e}_i(t). \tag{30}$$

Interestingly, the above equation was also used by Slock and Papadias in their extension of one-step prediction to n-step prediction [17]. However, these equations were not exploited jointly in [17]. Gesbert and Duhamel treated the above as a triangular system and constructed the error differentials

$$\mathbf{E} \triangleq \begin{pmatrix} \mathbf{e}_L(t) - \mathbf{e}_{L-1}(t+1) \\ \vdots \\ \mathbf{e}_1(t) - \mathbf{e}_0(t+1) \\ \mathbf{e}_0(t) \end{pmatrix} = \mathbf{h}\mathbf{s}_t. \tag{31}$$

The identification of $\mathbf{h}$ from $\mathbf{E}$ becomes a familiar problem. Note that unlike the one-step approach, the entire channel parameter is identified at once.

The multistep LP approach again relies on the uncorrelateness of the input sequence, which, for the same reason as in the LP-LS approach, causes the loss of finite-sample convergence property.

IV. Adaptive Least Squares Smoothing Channel Estimation

In this section, we develop an adaptive LSS algorithm with unknown channel order. The data structure for a variable smoothing window size is first specified. The adaptive least squares smoothing channel estimation and order detection algorithm is then presented.

A. Data Structure and Order Detection

In contrast with the data structure given in (20) where the smoothing window size is fixed, the data structure for an arbitrary smoothing window size is now considered. Suppose that an upper-bound L_u of the channel order is known. For a fixed smoothing order w and a variable smoothing window size l, define the overall data matrix as

$$\mathbf{V}_l \triangleq \left(\begin{array}{cc} \mathbf{y}(2w+L_u+1) & \cdots \quad \mathbf{y}(N) \\ \vdots & \mathbf{Z}_f(l) \\ \mathbf{y}(w+l+1) & \\ \hline \mathbf{y}(w+l) & \cdots \\ \vdots & \mathbf{D}(l) \\ \mathbf{y}(w+1) & \\ \hline \mathbf{y}(w) & \cdots \\ \vdots & \mathbf{Z}_p \\ \mathbf{y}(1) & \end{array}\right) \tag{32}$$

$$\mathbf{Z}(l) \triangleq \begin{pmatrix} \mathbf{Z}_f(l) \\ \mathbf{Z}_p \end{pmatrix}, \quad \mathbf{Z}_c \triangleq \begin{pmatrix} \mathbf{Z}_f(l) \\ \mathbf{D}(l) \end{pmatrix} \tag{33}$$

from which we have defined the future data matrix $\mathbf{Z}_f(l)$, the current data matrix $\mathbf{D}(l)$, and the past data matrix $\mathbf{Z}_p$. The future-past data matrix $\mathbf{Z}(l)$ and the future-current data matrix $\mathbf{Z}_c$ are also defined. Compared with the ones defined in (20), we can see that except for $\mathbf{Z}_p$ and $\mathbf{Z}_c$, the data matrices defined here are functions of the variable smoothing window size l. We emphasize that in this particular data structure, $\mathbf{Z}_c$ and $\mathbf{Z}_p$ are independent of l. This property leads to a convenient implementation of A-LSS, which will be discussed in Section IV-B.

Similar to the case presented in Section III-B2, when $N \geq 4w + 2L_u + L + 1$ and $\mathbf{s}_t$ has linear complexity greater than $2w + L_u + L$, in the noiseless case, we have the following relation among the matrices defined in (32) and the various spaces:

$$\begin{aligned} \mathcal{V}_l &\triangleq \mathcal{R}\{\mathbf{V}_l\} = \mathcal{X}_{2w+L_u+1,\,2w+L_u+1} \\ &= \mathcal{S}_{2w+L_u+1,\,2w+L_u+L+1} \end{aligned} \tag{34}$$

$$\begin{aligned} \mathcal{Z}_l &\triangleq \mathcal{R}\{\mathbf{Z}(l)\} = \mathcal{X}_{w,\,w} \bigcup \mathcal{X}_{w+l+1,\,-(w+L_u-l+1)} \\ &= \mathcal{S}_{w,\,w+L} \bigcup \mathcal{S}_{w+l-L+1,\,-(w+L_u+L-l+1)}. \end{aligned} \tag{35}$$

In parallel to the known channel order case in (22), we consider channel identification by a rank-1 decomposition of $\mathbf{E}(l) \triangleq \mathcal{E}_{\mathbf{Z}(l)}^{\mathbf{D}(l)}$. Since

$$\mathcal{V}_l = \mathcal{Z}_l \cup \mathcal{R}\{\mathbf{D}(l)\} = \mathcal{Z}_l + \mathcal{R}\{\mathbf{E}(l)\} \tag{36}$$

the rank-1 condition on $\mathbf{E}(l)$ requires that the orthogonal complement of $\mathcal{Z}_l$ in $\mathcal{V}_l$ has the dimension 1. Comparing the input subspaces in (34) and (35), we can see that this requirement on $\mathcal{Z}_l$ can only be met when $l = L + 1$. When $l < L + 1$, $\mathcal{Z}_l$ equals $\mathcal{V}_l$; there is no projection (smoothing)

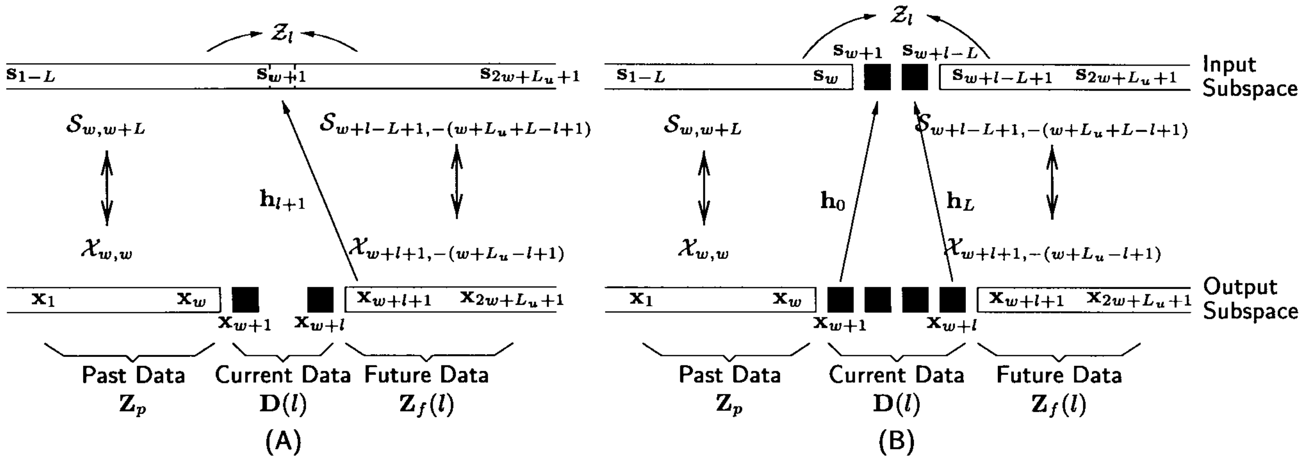

Fig. 5. Isomorphism between input and output subspaces for $l \neq L + 1$.

error. The isomorphism between the input and output spaces in this case is shown in Fig. 5(a). The case when $l > L + 1$ is shown in Fig. 5(b). Here, $\mathbf{s}_{w+1}, \cdots, \mathbf{s}_{w+l-L}$ are not in $\mathcal{Z}_l$. Since each of these input vectors not lying in $\mathcal{Z}_l$ contributes to the smoothing error, we have the formulation of $\mathbf{E}(l)$, as given in the following theorem.

Theorem 1: Let $\mathbf{E}(l) \triangleq \mathcal{E}_{\mathbf{Z}(l)}^{\mathbf{D}(l)}$ be the least squares smoothing error. With the data structure specified in (32), we have, in the absence of noise

$$\mathbf{E}(l) = \begin{cases} \mathbf{0} & l < L+1 \\ \mathbf{h}\tilde{\mathbf{s}}_{w+1|\mathcal{Z}_l} & l = L+1 \\ \underbrace{\begin{pmatrix} \mathbf{h}_L & & \\ \vdots & \ddots & \\ \mathbf{h}_0 & & \mathbf{h}_L \\ & \ddots & \vdots \\ & & \mathbf{h}_0 \end{pmatrix}}_{l-L \text{ columns}} \begin{pmatrix} \tilde{\mathbf{s}}_{w+l-L|\mathcal{Z}_l} \\ \vdots \\ \tilde{\mathbf{s}}_{w+1|\mathcal{Z}_l} \end{pmatrix} & L+1 < l \leq L_u + 1. \end{cases} \tag{37}$$

Theorem 1 holds the key to the adaptive least squares smoothing channel estimation and order detection algorithm. For order detection, we consider the energy $\varepsilon(l)$ of the smoothing error defined as $\varepsilon(l) \triangleq ||\mathbf{E}(l)||_F^2$. Theorem 1 implies that in the noiseless case, $\varepsilon(l)$ is zero when $l < L + 1$. When $l = L + 1$, $\varepsilon(l)$ jumps to a value related to the power of the channel and it increases (asymptotically) with l, as shown in Fig. 6. Hence, the channel order L can be detected by varying l from 1 to $L_u + 1$ and comparing the energy of the smoothing error at each l with a certain threshold.

B. Adaptive Least Squares Smoothing

1) Key Components: When the channel order is unknown, there are three key components involved in the A-LSS approach to channel estimation. First, the smoothing errors $\mathbf{E}(l)$ at each smoothing window size $l = 1, \cdots, L_u + 1$ need to be calculated. Then, based on Theorem 1, the energy of these smoothing errors is compared with a threshold to detect the channel order. Finally, with the detected channel order $\hat{L}$, the channel is obtained from $\mathbf{E}(\hat{L} + 1)$.

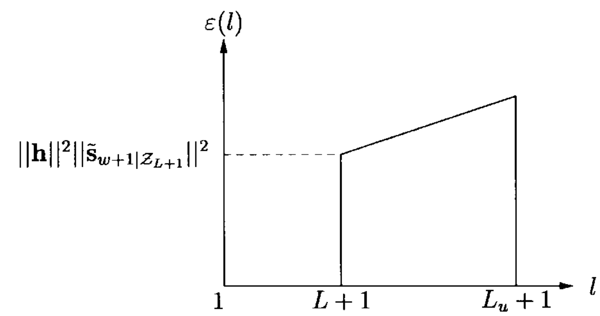

Fig. 6. Energy of the smoothing error at different smoothing window size l.

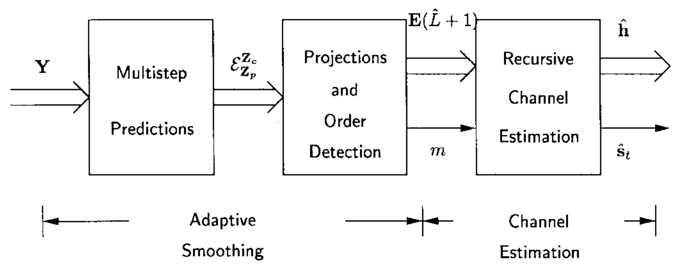

Fig. 7. Main structure of A-LSS.

Our goal here is to develop an algorithm to implement the above three components while simultaneously satisfying the requirements in convergence speed, adaptivity, and complexity. Since the LSS approach has the finite sample convergence property and adaptivity is a characteristic feature of linear least squares based algorithms, we next concentrate on the efficient implementation of A-LSS.

The main computation cost of the A-LSS algorithm comes from the calculation of the smoothing errors $\mathbf{E}(l)$ at all l. Thus, it is desirable to obtain the smoothing errors efficiently, which can be achieved by

a) calculating $\mathbf{E}(l)$ recursively;
b) decomposing the smoothing into multistep predictions followed by linear projections;

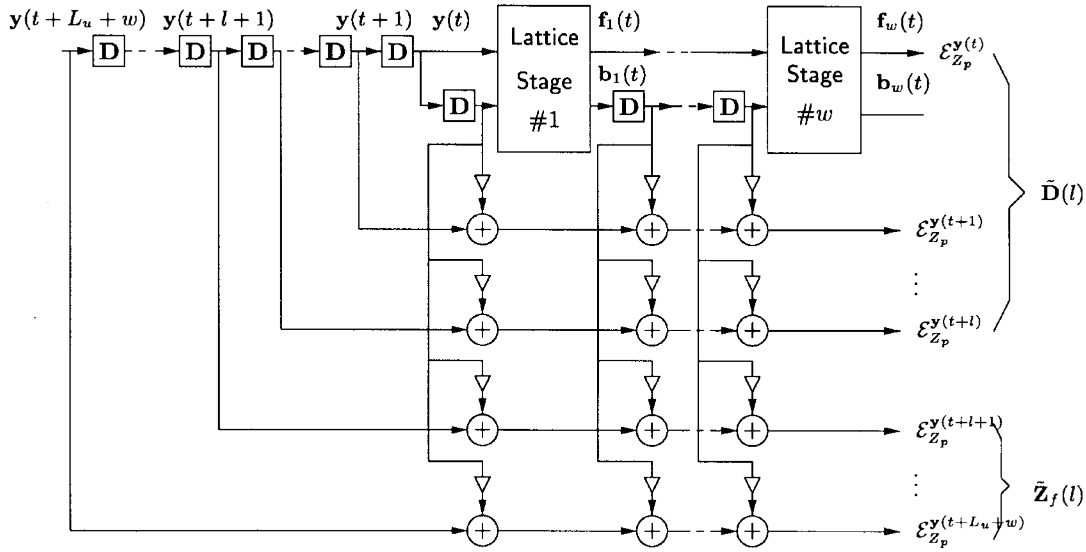

Fig. 8. Lattice filter for multistep prediction.

c) sharing the same multistep predictions among the calculation of the smoothing errors at all l.

Implementing smoothing by multistep predictions enables the use of lattice filters that are modular, computationally efficient, and robust to round-off errors. Details are presented as follows.

2) Key Decomposition: In A-LSS, smoothing is decomposed into multistep predictions and linear projections based on the following basic lemma.

Lemma 1: For a matrix $\mathbf{U}$ and a partitioned matrix $\mathbf{Q} = \binom{\mathbf{A}}{\mathbf{B}}$ with $\mathcal{E}_{\mathbf{B}}^{\mathbf{A}} \neq \mathbf{0}$, the least squares projection error of $\mathbf{U}$ into $\mathcal{R}\{\mathbf{Q}\}$ is given by

$$\mathcal{E}_{\mathbf{Q}}^{\mathbf{U}} = \mathcal{E}_{\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}}^{\mathcal{E}_{\mathbf{B}}^{\mathbf{U}}}. \tag{38}$$

In words, the projection error of $\mathbf{U}$ into $\mathcal{R}\{\mathbf{Q}\}$ is equal to the projection error of $\mathcal{E}_{\mathbf{B}}^{\mathbf{U}}$ into $\mathcal{R}\{\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}\}$, where $\mathcal{E}_{\mathbf{B}}^{\mathbf{U}}$ and $\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}$ are projection errors of $\mathbf{U}$ and $\mathbf{A}$ into $\mathcal{R}\{\mathbf{B}\}$, respectively.

Proof: In $\mathcal{R}\{\mathbf{Q}\}$, $\mathcal{R}\{\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}\}$ is the orthogonal complement of $\mathcal{R}\{\mathbf{B}\}$. Therefore

$$\mathcal{P}_{\mathbf{Q}}^{\mathbf{U}} = \mathcal{P}_{\mathbf{B}}^{\mathbf{U}} + \mathcal{P}_{\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}}^{\mathbf{U}} \tag{39}$$

which leads to

$$\mathcal{E}_{\mathbf{Q}}^{\mathbf{U}} \triangleq \mathbf{U} - \mathcal{P}_{\mathbf{Q}}^{\mathbf{U}} = \mathcal{E}_{\mathbf{B}}^{\mathbf{U}} - \mathcal{P}_{\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}}^{\mathbf{U}} = \mathcal{E}_{\mathbf{B}}^{\mathbf{U}} - \mathcal{P}_{\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}}^{\mathcal{E}_{\mathbf{B}}^{\mathbf{U}}} = \mathcal{E}_{\mathcal{E}_{\mathbf{B}}^{\mathbf{A}}}^{\mathcal{E}_{\mathbf{B}}^{\mathbf{U}}}. \tag{40}$$

□□□

Based on the partition of $\mathbf{Z}(l) \triangleq \binom{\mathbf{Z}_f(l)}{\mathbf{Z}_p}$ given in (33), we apply Lemma 1 to the smoothing error $\mathbf{E}(l)$

$$\mathbf{E}(l) \triangleq \mathcal{E}_{\mathbf{Z}(l)}^{\mathbf{D}(l)} = \mathcal{E}_{\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{Z}_f(l)}}^{\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{D}(l)}} = \mathcal{E}_{\tilde{\mathbf{Z}}_f(l)}^{\tilde{\mathbf{D}}(l)} \tag{41}$$

where $\tilde{\mathbf{D}}(l)$ and $\tilde{\mathbf{Z}}_f(l)$ are the multistep prediction errors of $\mathbf{D}(l)$ and $\mathbf{Z}_f(l)$ by the past data $\mathbf{Z}_p$, respectively. The smoothing error $\mathbf{E}(l)$ is then obtained by projecting $\tilde{\mathbf{D}}(l)$ into the row space of $\tilde{\mathbf{Z}}_f(l)$, which implies that smoothing is decomposed into multistep predictions and linear projections. Although (41) appears to imply that separate linear predictors are required for different l, it turns out that with the special data structure defined in (32), only a single predictor is necessary. This is because from (32), we have

$$\begin{pmatrix} \tilde{\mathbf{Z}}_f(l) \\ \tilde{\mathbf{D}}(l) \end{pmatrix} = \mathcal{E}_{\mathbf{Z}_p}^{\mathbf{Z}_c} \tag{42}$$

where both $\mathbf{Z}_c$ and $\mathbf{Z}_p$ are independent of l.

3) Main Structure: Based on the decomposition presented above, A-LSS identifies the channel in three steps, as shown in Fig. 7. Specifically, the multistep predictor generates the prediction errors $\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{Z}_c} \triangleq \binom{\tilde{\mathbf{Z}}_f(l)}{\tilde{\mathbf{D}}(l)}$, which are independent of l. In the second step, the smoothing errors at all possible l are obtained recursively, and the channel order is detected. Finally, with the detected channel order $\hat{L}$, the channel is estimated from the smoothing error $\mathbf{E}(\hat{L}+1)$ with the approach given in (15) to meet the requirement in adaptivity. The implementation of these three steps is discussed below.

a) Multistep predictions: A lattice filter shown in Fig. 8 is used to obtain the prediction errors of the future and current data $\mathbf{Z}_c$ by the past data $\mathbf{Z}_p$. Various adaptive algorithms for lattice filters can be applied here; see [10], [12], and [15]. Besides the appealing properties mentioned in Section IV-B1, the order-recursive property of lattice filters can be very useful in saving computation cost for the unknown channel-order case. Specifically, when the channel order L is unknown, the smoothing order w (also the prediction order here) has to be determined based on the upper-bound L_u. For example, when $P = 2$, we choose $w = L_u$. If L_u is a poor bound of the channel order, w may be much larger than necessary, which leads to a higher computation cost. Because of the order-recursive property of the lattice filter, for a detected channel order $\hat{L} < w$, only the first $2\hat{L}+1$ prediction errors $(\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t)}, \cdots, \mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+2\hat{L})})$ at the $\hat{L}$th lattice stage are required. This implies that the computation involved in the latter stages and the $2(L_u - \hat{L})$ joint estimation $(\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+2\hat{L}+1)}, \cdots, \mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+2L_u)})$ is saved with no structural change.

b) Projections and order detection: From the prediction errors $\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{Z}_c}$, the smoothing errors $\mathbf{E}(l)$ at each smoothing

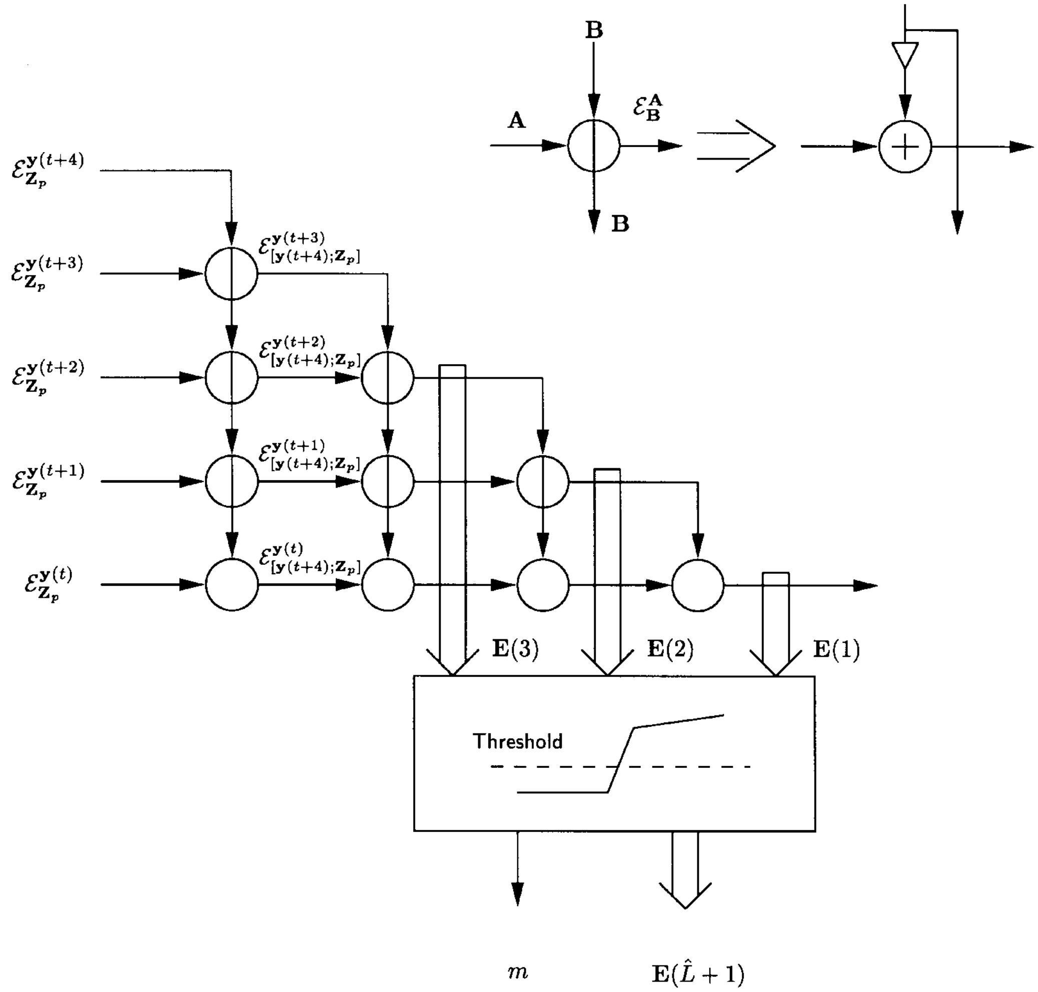

Fig. 9. Recursive projection and order detection.

window size $l = 1, \cdots, L_u + 1$ are obtained in this step for the order detection and channel estimation.

An approach to obtaining $\mathbf{E}(l)$ recursively (both in l and in time) by applying the recursive modified Gram–Schmidt (RMGS) algorithm [11] is presented with a simple example shown in Fig. 9. Suppose that the channel order is upper bounded by $L_u = 2$, and we choose the smoothing order $w = 2$. From the prediction errors $\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+4)}, \cdots, \mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t)}$ obtained by the multistep predictor, the smoothing errors $\mathbf{E}(l)$ at $l = 1, 2, 3$ are calculated recursively. As shown in Fig. 9, by applying Lemma 1 first to $\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+4)}$ and $\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+i)} (i = 0, \cdots, 3)$, we have

$$\mathcal{E}_{\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+4)}}^{\mathcal{E}_{\mathbf{Z}_p}^{\mathbf{y}(t+i)}} = \mathcal{E}_{[\mathbf{y}(t+4); \mathbf{Z}_p]}^{\mathbf{y}(t+i)} \tag{43}$$

where $[\mathbf{y}(t+4); \mathbf{Z}_p]$ denotes the partitioned matrix $\binom{\mathbf{y}(t+4)}{\mathbf{Z}_p}$. Applying Lemma 1 next to $\mathcal{E}_{[\mathbf{y}(t+4); \mathbf{Z}_p]}^{\mathbf{y}(t+3)}$ and $\mathcal{E}_{[\mathbf{y}(t+4); \mathbf{Z}_p]}^{\mathbf{y}(t+i)}$ $(i = 0, \cdots, 2)$, we have

$$\mathcal{E}_{\mathcal{E}_{[\mathbf{y}(t+4); \mathbf{Z}_p]}^{\mathbf{y}(t+3)}}^{\mathcal{E}_{[\mathbf{y}(t+4); \mathbf{Z}_p]}^{\mathbf{y}(t+i)}} = \mathcal{E}_{[\mathbf{y}(t+4); \mathbf{y}(t+3); \mathbf{Z}_p]}^{\mathbf{y}(t+i)} = \mathcal{E}_{\mathbf{Z}(3)}^{\mathbf{y}(t+i)} \tag{44}$$

which is the ith row block of $\mathbf{E}(3)$. Similarly, $\mathbf{E}(2)$ and $\mathbf{E}(1)$ can be obtained. In general, we have the following recursion in l:

$$\mathbf{E}(l) \triangleq \mathcal{E}_{\tilde{\mathbf{Z}}_f(l)}^{\tilde{\mathbf{D}}(l)} = \mathcal{E}_{\mathbf{E}_1(l+1)}^{\mathbf{E}_2(l+1)} \tag{45}$$

where $\mathbf{E}(l+1)$ has been partitioned into $\binom{\mathbf{E}_1(l+1)}{\mathbf{E}_2(l+1)}$ with $\mathbf{E}_1(l+1)$ defined as the first row block in $\mathbf{E}(l+1)$. This recursion in l arises naturally from the data structure given in (32).

Once all $\mathbf{E}(l)$ are obtained, the channel order is detected by an energy detector. Since the energy of $\mathbf{E}(\hat{L}+1)$ is calculated for the channel order detection, the index (m) of the row with the maximum 2-norm in $\mathbf{E}(\hat{L}+1)$ [see (15)] can be easily obtained. This information will be used in the next step to estimate the channel from $\mathbf{E}(\hat{L}+1)$.

Finally, we want to emphasize that the structure shown in Fig. 9 is particularly attractive for time-varying channels. For example, when the true channel order changes from 2 to 1, the energy detector switches from $\mathbf{E}(3)$ to $\mathbf{E}(2)$. The channel with the new order 1 can be identified without structural change or extra computation. In addition, the modular structure shown in Fig. 9 is suitable for implementation using systolic array processors.

c) Recursive channel estimation: With m obtained as the by-product of the energy detector in the second step, the

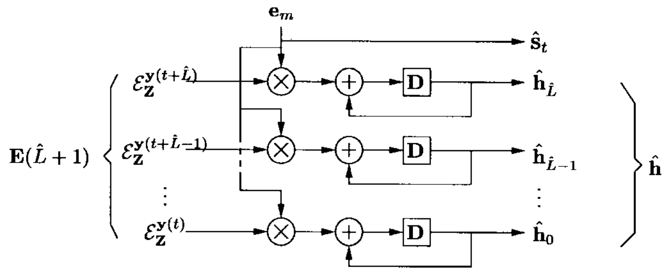

Fig. 10. Recursive channel estimation.

TABLE I
ALGORITHMS COMPARED IN THE SIMULATION AND THEIR CHARACTERISTICS

Abbreviation	Name	Finite-sample convergence property?	Order and time recursive property?
SS [4]	The Subspace Algorithm	Yes	No
CR [5], [3]	The Cross Relation Algorithm	Yes	No
LP-LS [1]	Linear Prediction-Least Squares Algorithm	No	Yes
MSP [2]	Multistep Linear Prediction Algorithm	No	Yes
A-LSS	Adaptive Least Squares Smoothing Algorithm	Yes	Yes

channel is estimated from the smoothing error $\mathbf{E}(\hat{L}+1)$ by the approach given in (15). In addition, $\mathbf{e}_m$ is an asymptotical estimate of the uncorrelated input sequence. The structure of this recursive channel estimator is illustrated in Fig. 10.

V. SIMULATION EXAMPLES

A. Algorithm Characteristics and Performance Measure

Simulation studies of the proposed A-LSS algorithm as it is compared with existing techniques listed in Table I are presented in this section. We remark that only A-LSS has easy adaptive implementations while still preserving the finite sample convergence property.

Algorithms were compared by Monte Carlo simulation using the normalized root mean square error (NRMSE) as a performance measure. Specifically, NRMSE was defined as[3]

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

where $\hat{\mathbf{h}}^{(m)}$ was the estimated channel from the mth trial, and $\mathbf{h}_*$ was the true channel. Noise samples were generated from i.i.d. zero mean Gaussian random sequences, and the signal-to-noise ratio (SNR) was defined as

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

The input to a single-input and 2-output linear FIR system was generated from an i.i.d. binary phase shift keying (BPSK) sequence.

B. A Second-Order Channel

A second-order channel first used by Hua in [7] is considered here. There are several reasons to consider this channel. First, this channel allows us to study the location of zeros and how they affect the performance. Second, Hua showed that the CR algorithm [22] and SS algorithm [14], along with the two-step maximum likelihood (TSML) algorithm [7], approach to the Cramér–Rao lower bound.[4] By comparing with CR(SS), we can evaluate the efficiency of the algorithms listed in Table I.

The second-order channel used by Hua in [7] is specified by a pair of roots on the unit circle. The channel impulse response is given by

$$h_i(z) = (1-e^{j\theta_i}z^{-1})(1-e^{-j\theta_i}z^{-1}), \qquad i=1,2 \quad (48)$$

where θ_i are the angular positions of zeros on the unit circle. The relation between the zeros of the two subchannels is specified by $\theta_2 = \theta_1 + \delta$, where δ is the angular distance between the zeros of the two subchannels (common zeros occur when δ is zero).

C. Performance and Robustness Comparison

The left side of Fig. 11 shows the performance of algorithms listed in Table I against SNR. A well-conditioned channel was used with $\delta = \pi$, $\theta_1 = \pi/10$, as in [7]. All the deterministic algorithms (CR, SS, B-LSS, and A-LSS) had comparable performance. In fact, it was shown in [7] that CR and SS approached the CRB even at a very low SNR for this channel. The LP-based algorithms (MSP and LP-LS) leveled off as SNR $\rightarrow \infty$ because of the loss of finite sample convergence property. We noticed that B-LSS had the same performance as CR (SS), but there was a gap between the performance of A-LSS and B-LSS at low SNR. This gap was caused by the prewindowing problem in A-LSS; several symbols were discarded in the transient stage. Thus, the number of the effective symbols used by A-LSS was less than that used by the batch algorithms, which led to the performance degradation. This performance gap will diminish as SNR $\rightarrow \infty$ or more observation samples become available.

An ill-conditioned channel was used to compare the robustness of different algorithms with respect to the loss of channel diversity. With $\delta = \pi/100$, $\theta_1 = 0$, zeros of the two subchannels were very close to each other. In fact, the condition number of the filtering matrix was around 1.6×10^4 in this case. The right side Fig. 11 shows that CR and SS performed rather poorly for this ill-conditioned channel. Evidently, B-LSS and A-LSS performed considerably better than all other algorithms. This improvement in robustness is probably because the input subspace may still be well approximated by the output subspace when the channel diversity assumption does not hold.

D. Tracking of the Channel Order and Parameters

In this simulation, A-LSS was applied to a case when both channel order and channel parameters had a sudden change. The initial channel was given in (48) with $\delta = \pi$, $\theta_1 = \pi/10$. At time $n = 151$, both the channel order and channel parameters were changed by adding zeros $Z_1 = 1$, $Z_2 = -1$

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

[4]In [7], the CRB was derived based on a normalization that differs from the one used in this paper.

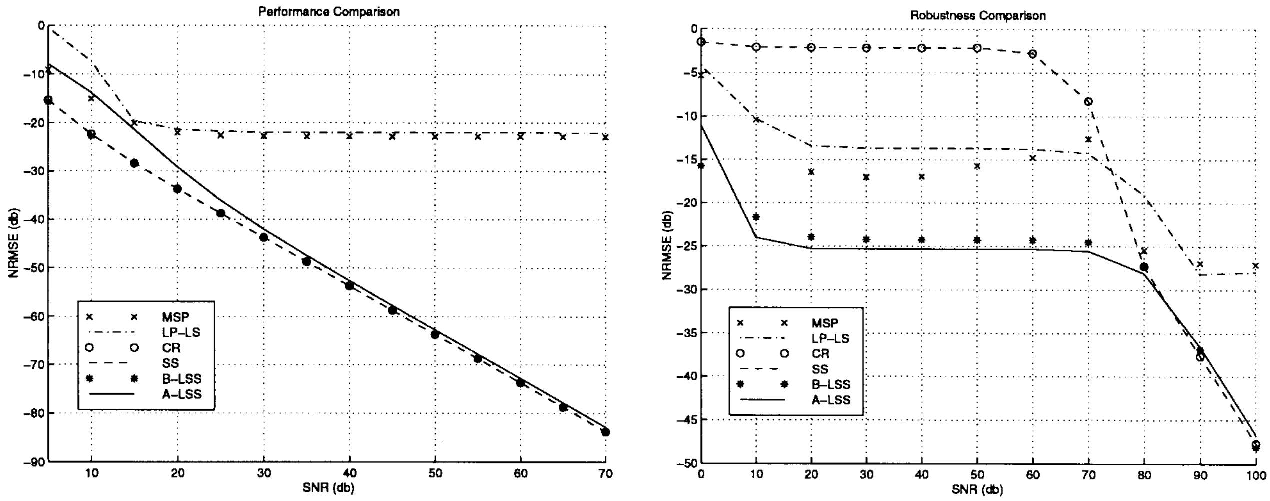

Fig. 11. Performance and robustness comparison (100 Monte Carlo runs, 200 input symbols).

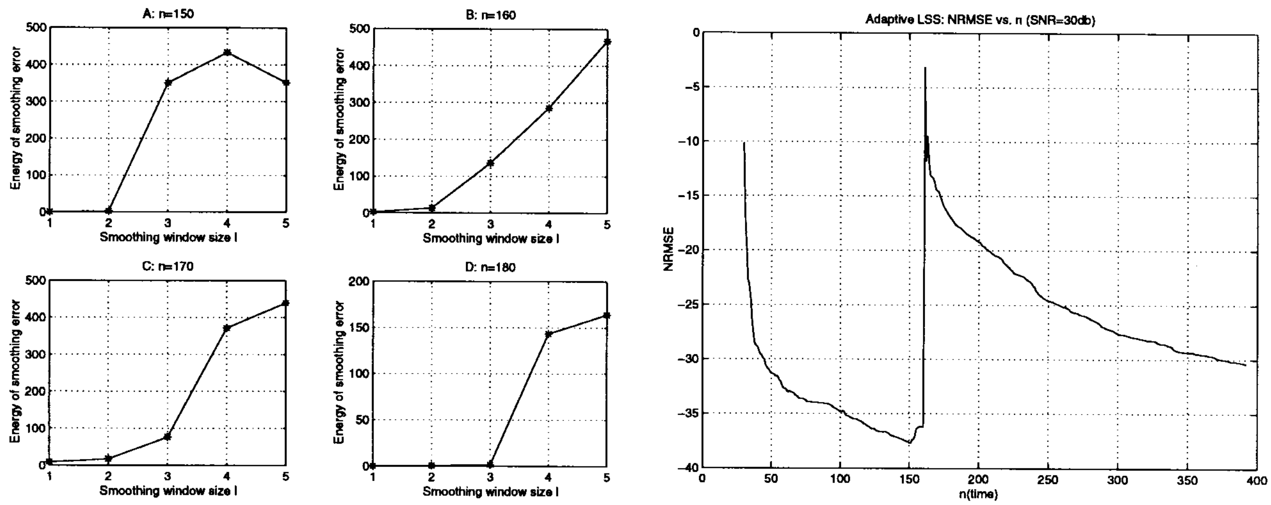

Fig. 12. Channel order and parameter tracking performance (SNR $= 30$ dB).

to the two subchannels, respectively. The left side of Fig. 12 shows the energy of the smoothing error before and after the channel variation (the energy of the smoothing error was calculated every 10 symbols). From Fig. 12(a), we can see that before the channel changed, the energy of the smoothing error was around zero at $l < L+1 = 3$ and was relatively large at $l = L+1 = 3$, as predicted by Theorem 1. Fig. 12(b)–(d) are the snapshots of the smoothing error energy at $n =$ 160, 170, and 180, respectively. We can see that the energy of $\mathbf{E}(3)$ decreased to zero within 30 samples. At $n =$ 180, the new channel order can be accurately detected. Consequently, the channel is estimated from $\mathbf{E}(4)$ instead of $\mathbf{E}(3)$. Neither structural change nor extra computation is involved in this process. The NRMSE convergence of A-LSS is shown in the right side of Fig. 12, where A-LSS tracked the channel order and parameters nicely.

D. Order Detection for Multipath Channels

In order to evaluate the applicability of the proposed order detection and channel estimation algorithm, we present here the simulation study of the performance of A-LSS with a multipath channel. The channel we used is a fifth-order 4-ray raised-cosine channel as shown in the top left of Fig. 13 with the even and odd samples corresponding to the two subchannels. The upperbound of the channel order was assumed to be 6. The energy of the smoothing error at different smoothing window size is shown in the top right of Fig. 13. We can see that the energy of the smoothing error jumped to a large value at the smoothing window size $l = 2$. Based on this fact, the energy detector estimated the channel order as 1. In the bottom left of Fig. 13, we also plotted the relative increment of the smoothing error energy at two consecutive smoothing window sizes, and we chose maximizing the relative increment as an alternative criterion for the order detection. The same estimated channel order $\hat{L} = 1$ was obtained. As shown in the bottom right of Fig. 13, among all the possible channel orders (from 0 to L_u), A-LSS gave the best estimate of the channel at the detected order. In fact, as pointed out in [20], it is perhaps not wise to estimate the small head and tail taps in the multipath channel. Instead, it is better to find the channel order as well as its impulse response that matches the

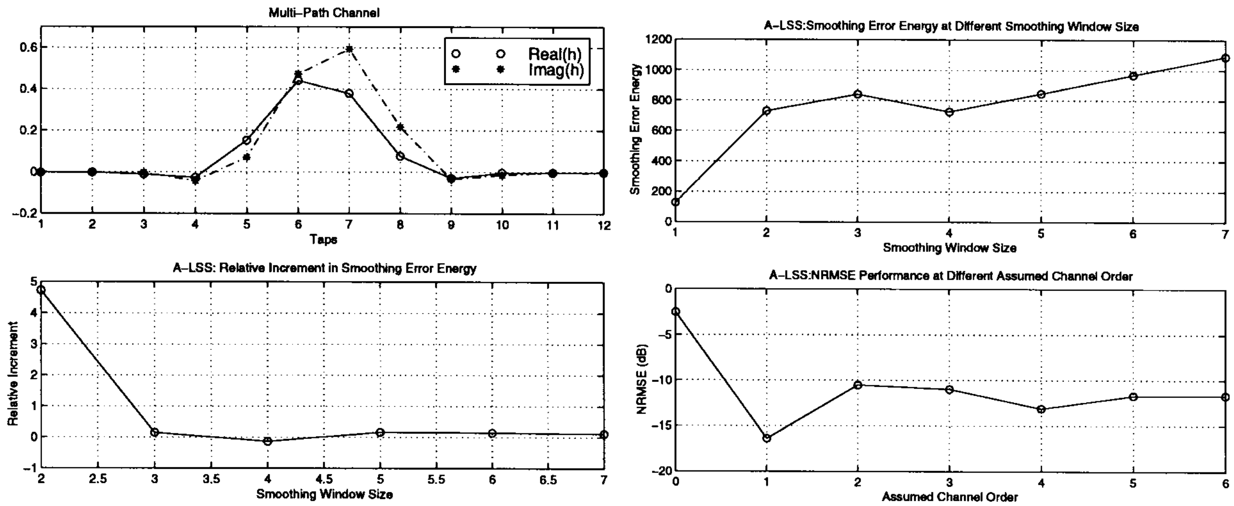

Fig. 13. Order detection and channel estimation with multipath channel (SNR $= 30$ dB, 100 Monte Carlo runs, 200 input symbols).

data in some optimal way. In the case here, with the channel order detected as 1, A-LSS captured the four major taps of the channel impulse response while ignoring the small head and tail taps.

VI. Conclusion

A least squares smoothing (LSS) approach to blind channel estimation based on the isomorphic relation between the input and output subspaces is presented. This approach converts the channel estimation into a linear least squares smoothing problem that can be solved (order and time) recursively. Since the input sequence is modeled as a deterministic signal, this approach preserves the finite-sample convergence property not present in LP-based approaches.

Based on the LSS approach, an adaptive channel estimation (A-LSS) algorithm has been proposed. Compared with the batch algorithms such as SS and CR, the A-LSS algorithm is efficient in both computation and VLSI implementation. Because smoothing is computationally more expensive than prediction, A-LSS has higher complexity than the LP-LS and the MSP approaches, which is a price we paid for the finite sample convergence property. The A-LSS algorithm does not assume the channel order as *a priori* information. The order detector and the recursive property enable A-LSS to detect and track the channel order without structural change or extra computation. Although separate order detection techniques can be applied to the deterministic batch algorithms such as SS and CR, their ability to track the channel order variation demands high implementation cost.

The future work involves the theoretical determination of the threshold for the channel order detection. Maximizing the relative increment in the smoothing error energy at two consecutive smoothing window sizes may be considered as an alternative criterion.

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] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1996.
[7] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.
[8] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1993.
[9] S. Y. Kung, T. Kailath, and M. Morf, "A generalized resultant matrix for polynomial matrices," in *Proc. IEEE Conf. Decision Contr.*, 1976, pp. 892–895.
[10] H. Lev-Ari, "Modular architectures for adaptive multichannel lattice algorithms," *IEEE Trans. Signal Processing*, vol. 35, pp. 543–552, Apr. 1987.
[11] F. Ling, D. Manolakis, and J. G. Proakis, "A recursive modified Gram–Schmidt algorithm for least-squares estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 829–836, Aug. 1986.
[12] F. Ling and J. G. Proakis, "A generalized multichannel least squares lattice algorithm based on sequential processing stages," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 381–390, Apr. 1984.
[13] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714–2723, Nov. 1995.
[14] 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.
[15] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Mag.*, vol. 11, pp. 18–60, July 1994.
[16] D. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filterbanks, and multilinear prediction," in *Proc. ICASSP Conf.*, Adelaide, Australia, Apr. 1994.
[17] D. Slock and C. B. Papadias, "Further results on blind identification and equalization of multiple FIR channels," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Detroit, MI, Apr. 1995, pp. 1964–1967.
[18] L. Tong and J. Bao, "Equalizations in wireless ATM," in *Proc. 1997 Allerton Conf. Commun., Contr. Comput.*, Urbana, IL, Oct. 1997, pp. 64–73.
[19] 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.

[20] L. Tong and Q. Zhao, "Joint order detection and channel estimation by least squares smoothing," *IEEE Trans. Signal Processing*, vol. 47, pp. 2345–2355, Sept. 1999.

[21] A. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 173–190, Jan. 1997.

[22] 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.

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 School of Electrical Engineering, Cornell University, Ithaca, NY.

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

Lang Tong (S'87–M'91), for a photograph and biography, see this issue, p. 2999.