BLIND CHANNEL ESTIMATION BY LEAST SQUARES SMOOTHING

Lang Tong and Qing Zhao
Department of Electrical & Systems Engineering
The University of Connecticut Storrs, U-157, CT 06269-3157
Tel: (860) 486-2192. Email: ltong@eng2.uconn.edu

Abstract

A linear least squares smoothing approach is proposed for the blind channel estimation. It is shown that the single-input multiple-output moving average process has the property that the error sequence of the least squares smoother, under certain conditions, uniquely determines the channel impulse response. The relationship among the dimension of the observation space, channel order and smoothing delay is presented. A new algorithm for channel estimation based on the least squares smoothing is developed. The proposed approach has the finite-sample convergence property in the absence of the channel noise. It also has a structure suitable for recursive implementations.

1. 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. Among blind channel estimation techniques developed recently [8], those based on the so-called deterministic models have clear advantage in the speed of convergence. Without assuming specific stochastic models 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 multichannel structure first exploited in [9]. Existing algorithms with this attractive feature include the subspace algorithm (SS) [5], the least squares (LS) algorithm [11], and the two-step maximum likelihood (TSML) approach [4].

Perhaps equally important in blind channel equalization are the adaptivity and the simplicity of the implementation. Unfortunately, most existing deterministic algorithms are developed for batch processing and their adaptive implementations are often cumbersome. In this regard, stochastic algorithms [6, 1, 3] based on the linear prediction (LP) interpretation of the multichannel model, originally exploited by Slock [6], have the potential of having effective adaptive realization. While not derived from the LP framework, the outer product decomposition algorithm (OPDA) [2] is closely related to the multistep prediction (MSP) approach of Gesbert and Duhamel [3] for they have the same identification equation.

The contribution of this paper is twofold. First, we present a linear smoothing interpretation of the multichannel moving average processes. While similar in spirit to the LP approach, the linear smoothing framework, first proposed in [12], allows the deterministic formulation of the problem whereas the assumption of uncorrelated input sequence is crucial in the LP framework. Key properties of the optimal linear smoothing are presented. Second, a least squares smoothing approach is developed that has the finite-sample convergence property in the absence of noise. The proposed algorithm performs better than stochastic methods and is comparable with other deterministic approaches. Using orthogonal projections, the proposed approach is well suited for both order- and time-recursive implementations.

This work was supported in part by the National Science Foundation under Contract NCR-9321813 and by the Office of Naval Research under Contract N00014-96-1-0895.

2. THE MODEL

Considered in this paper is the single input P-output linear channel model given by

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

where $\mathbf{x}_t$ is the channel output, $\mathbf{y}_t$ is the received signal, $\{\mathbf{h}_t\}$ is the channel impulse response, s_t is the input sequence. The matrix representation of the above channel is obtained by considering the vector of W samples of the observation $\mathbf{y}_W(t) \triangleq [\mathbf{y}_t^t, \cdots, \mathbf{y}_{t-W+1}^t]^t$. With $\mathbf{x}_W(t)$ and $\mathbf{n}_W(t)$ similarly defined, we have

$$\mathbf{x}_W(t) = \mathcal{F}_W(\mathbf{h})\mathbf{s}(t), \quad \mathbf{y}_W(t) = \mathbf{x}_W(t) + \mathbf{n}_W(t), \qquad (2)$$

where $\mathcal{F}_W(\mathbf{h}) \in \mathcal{C}^{WP\times(L+W)}$ is the filtering matrix

$$\mathcal{F}_W(\mathbf{h}) \triangleq \begin{pmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_L & & \\ & \ddots & \cdots & \ddots & \\ & & \mathbf{h}_0 & \cdots & \mathbf{h}_L \end{pmatrix} = [\mathbf{f}_1, \cdots \mathbf{f}_{W+L}] \tag{3}$$

We shall make the following assumptions:

A1 $\mathcal{F}_W(\mathbf{h})$ has full column rank for $W \geq L+1$.

A2 $\{s_k\}$ has the linear complexity* greater than $6L$, *i.e.*,

$$rank \begin{pmatrix} s_{5L+1} & \cdots & s_N \\ \vdots & \text{Toeplitz} & \\ s_{-L} & & \end{pmatrix} = 6L+1$$

Remarks: The requirement of linear complexity is stronger than necessary. It is shown in [7] that the necessary and sufficient condition is $2L+1$ when $P=2$.

3. THE LEAST SQUARES SMOOTHING

3.1. A Special Case

To demonstrate the basic ideas of using linear smoothing for channel estimation, we consider a simple example with $L=2$ and $W=L+1$. Here we assume first the channel order is known and there is no noise. From (2), we have

$$\mathbf{x}_3(t) = \begin{pmatrix} \mathbf{h}_0 & \mathbf{h}_1 & \boxed{\mathbf{h}_2} & & \\ & \mathbf{h}_0 & \boxed{\mathbf{h}_1} & \mathbf{h}_2 & \\ & & \boxed{\mathbf{h}_0} & \mathbf{h}_1 & \mathbf{h}_2 \end{pmatrix} \begin{pmatrix} s_t \\ s_{t-1} \\ \boxed{s_{t-2}} \\ s_{t-3} \\ s_{t-4} \end{pmatrix} \tag{4}$$

Under (A1), there exist $\mathbf{g}$ and $\mathbf{b}$ such that

$$s_t = \mathbf{g}^H \mathbf{x}(t) = \mathbf{b}^H \mathbf{x}(t+4). \tag{5}$$

Substituting the above into (4), we have, $\forall t$,

$$\begin{aligned} \mathbf{x}_3(t) &= \mathbf{h}s_{t-2} + \mathbf{A}_1 \begin{pmatrix} \mathbf{x}_{t+4} \\ \vdots \\ \mathbf{x}_{t+1} \end{pmatrix} + \mathbf{A}_2 \begin{pmatrix} \mathbf{x}_{t-3} \\ \vdots \\ \mathbf{x}_{t-6} \end{pmatrix} \\ &= \mathbf{h}s_{t-2} + \mathbf{A}\mathbf{z}(t), \end{aligned} \tag{6}$$

where $\mathbf{A}$ is a constant matrix and $\mathbf{z}(t)$ contains future and past data. Collecting all data samples, we have

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{E}, \tag{7}$$

where $\mathbf{E} \triangleq \mathbf{h}[s_5, \cdots s_{N-6}]$,

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_7 & \cdots & \mathbf{x}_{N-4} \\ \vdots & \text{Block} & \\ \mathbf{x}_5 & \text{Toeplitz} & \end{pmatrix}, \mathbf{Z} \triangleq \begin{pmatrix} \mathbf{x}_{11} & \cdots & \mathbf{x}_N \\ \vdots & \text{Block} & \\ \mathbf{x}_8 & \text{Toeplitz} & \\ \hline \mathbf{x}_4 & \cdots & \mathbf{x}_{N-5} \\ \vdots & \text{Block} & \\ \mathbf{x}_1 & \text{Toeplitz} & \end{pmatrix}. \tag{8}$$

*The linear complexity of sequence $\{s_k\}_{k=1}^n$ is defined as the smallest value of c for which $s_i = -\sum_{j=1}^c \lambda_j s_{i-j}, \;\; i = c, \cdots, n$.

The key observation is that both $\mathbf{X}$ and $\mathbf{Z}$ are made of received data, and the rank 1 matrix $\mathbf{E}$ has column space spanned by the channel vector $\mathbf{h}$. If $\mathbf{E}$ can be computed from $\mathbf{X}$ and $\mathbf{Z}$, then $\mathbf{h}$ can be easily obtained. This is possible when $\mathbf{E}$ is orthogonal to $\mathbf{Z}$, in which case $\mathbf{E}$ is the least squares estimation error of $\mathbf{X}$ by $\mathbf{Z}$. The orthogonal property is immediate in the *stochastic models* when the input sequence is uncorrelated. In the deterministic framework, such an orthogonal property unfortunately does not hold for $\mathbf{E}$, which underscores the difference between the two approaches.

3.2. The General Formulation

We present next the general formulation of the smoothing approach. Specifically, we consider the smoothing estimate of $\mathbf{x}_W(t)$ using $\mathbf{z}_d(t) \triangleq \begin{pmatrix} \mathbf{z}_d^f(t;d) \\ \mathbf{z}_d^p(t;d) \end{pmatrix}$ where the future and past observations are contained, respectively, in

$$\mathbf{z}_d^f(t) \triangleq \begin{pmatrix} \mathbf{x}_{t+d+W-1} \\ \vdots \\ \mathbf{x}_{t+1} \end{pmatrix}, \mathbf{z}_d^p(t) \triangleq \begin{pmatrix} \mathbf{x}_{t-d-1} \\ \vdots \\ \mathbf{x}_{t-d-2W+2} \end{pmatrix}. \tag{9}$$

Following the special case described above, define

$$\begin{aligned} \mathbf{X} &\triangleq [\mathbf{x}_W(d+2W-1), \cdots, \mathbf{x}_W(N-d-W+1)] \\ &= \begin{pmatrix} \mathbf{x}_{d+2W-1} & \cdots & \mathbf{x}_{N-d-W+1} \\ \vdots & \text{Block} & \\ \mathbf{x}_{d+W-1} & \text{Toeplitz} & \end{pmatrix}, \end{aligned} \tag{10}$$

$$\mathbf{Z}_d \triangleq [\mathbf{z}_d(d+2W-1), \cdots, \mathbf{z}_d(N-d-W+1)] \tag{11}$$

$$= \begin{pmatrix} \mathbf{x}_{2d+3W-2} & \cdots & \mathbf{x}_N \\ \vdots & \text{Block} & \\ \mathbf{x}_{d+2W} & \text{Toeplitz} & \\ \hline \mathbf{x}_{2W-2} & \cdots & \mathbf{x}_{N-2d-W} \\ \vdots & \text{Block} & \\ \mathbf{x}_1 & \text{Toeplitz} & \end{pmatrix}, \tag{12}$$

$$\mathbf{E}_d \triangleq \begin{cases} \mathbf{0} & d < L \\ \underbrace{\begin{pmatrix} \mathbf{h}_L & & \\ \vdots & \ddots & \\ \mathbf{h}_0 & & \mathbf{h}_L \\ & \ddots & \vdots \\ & & \mathbf{h}_0 \end{pmatrix}}_{\mathcal{H}_{L,d}(\mathbf{h}) \in \mathcal{C}^{PW\times(d-L+1)}} \mathbf{S}_d & \mathrm{L} \leq \mathrm{d} \leq \mathrm{W}-1 \end{cases} \tag{13}$$

$$\mathbf{S}_d \triangleq \begin{pmatrix} s_{\mathrm{d+2W-L-1}} & \cdots & s_{\mathrm{N-d-W-L+1}} \\ \vdots & \text{Toeplitz} & \\ s_{\mathrm{2W-1}} & & \end{pmatrix} \tag{14}$$

We now present the main result in the least squares smoothing of $\mathbf{x}_W(t)$.

Theorem 1 *Under assumptions (A1-A2) with* $\mathbf{X}$, $\mathbf{Z}_d$ *and* $\mathbf{E}_d$ *defined in (10-13), there exists a matrix* $\mathbf{A}$ *such that*

$$\mathbf{X} = \mathbf{E}_d + \mathbf{A}\mathbf{Z}_d, \tag{15}$$

with the following properties:

1. *orthogonal projection:*

$$rang(\mathcal{P}^{\perp}_{\mathbf{Z}_d}\mathbf{X}) = rang(\mathbf{E}_d); \tag{16}$$

2. *the rank condition:*

$$rank\{\mathbf{Z}_d\} = \begin{cases} 3W + L + 2d - 2 & d < L \\ 3W + 2L + d - 3 & L \leq d \leq W - 1 \end{cases} \tag{17}$$

Remarks:

- The orthogonal projection condition implies that the range space of $\mathbf{E}_d$ is uniquely determined from the observation. This is illustrated in Figures 1. Further, if $d \geq L$, then matrix $\mathcal{P}^{\perp}_{\mathbf{Z}_d}\mathbf{X}$ uniquely determines $\mathbf{h}$.
- If L is known and $d = L$, as in the special case discussed earlier, $\mathcal{P}^{\perp}_{\mathbf{Z}_d}\mathbf{X}$ is a rank-one matrix and $\mathbf{h}$ correspond to the first left singular vector.
- When L is unknown but its upper bound is known, then we can choose $d \geq L$ and find the best fit in the least squares sense between the range space of $\mathbf{E}_d$ and $\mathcal{H}_{L,d}(\mathbf{h})$. Otherwise, it is more attractive to recursively (in d) evaluate the error until it exceeds certain threshold. Notice the structure of $\mathbf{Z}_d$ where only the block corresponding to the future data varies with d. By applying QR decomposition recursively, an order recursion scheme can be easily implemented. Details of these operations can be found in [10].
- The rank condition in (17) is useful in two ways. First, it can be used to determine the dimension of $Row(\mathbf{Z}_d)$. With $\mathbf{Z}_d$ singular in general, the rank condition is especially useful in determining $Row(\mathbf{Z}_d)$ in the presence of noise. Second, the rank of $\mathbf{Z}_d$ increases with d by 2 when $d < L$ and by 1 once $d \geq L$. This property can be used in order-recursive implementation.

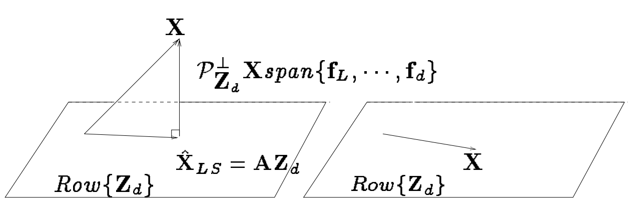

Figure 1: The least squares smoothing. Left: $d \geq L$. Right: $d < L$.

3.3. The Algorithm

So far we have not considered the effect of channel noise. To develop a practical channel estimator, noise must be compensated. Let $\mathbf{X}$ and $\mathbf{Z}_d$ be perturbed to $\mathbf{Y}$ and $\tilde{\mathbf{Z}}_d$, respectively. In light of the rank condition in (17), the effect of noise can be minimized by using the least squares estimate of $\mathbf{Z}$ via the following optimization

$$\hat{\mathbf{Z}}_d = \arg \min_{rank(\mathbf{Z})=r} ||\mathbf{Z} - \tilde{\mathbf{Z}}||_2^2 \tag{18}$$

where r is obtained from (17). The solution of this optimization can be obtained from SVD. Projecting $\mathbf{Y}$ onto $\hat{\mathbf{Z}}_d$, the rank-one approximation of the projection gives the channel estimate. The following algorithm implements this approach.

The LSS Algorithm

1. Form $\mathbf{Y}$, and $\tilde{\mathbf{Z}}_d$.
2. Obtain the orthogonal basis $\mathbf{Q}$ that spans the row space of $\hat{\mathbf{Z}}_d$.
3. Compute

$$\tilde{\mathbf{E}}_d = \mathbf{Y} - \mathbf{Y}\mathbf{Q}^H\mathbf{Q} \overset{\text{SVD}}{=} \mathbf{U}_E\mathbf{D}_E\mathbf{V}_E^H. \tag{19}$$

4. $\hat{\mathbf{h}} = \mathbf{U}_E(:, 1)$

Remarks:

- One of the attractive features of this algorithm is that both order and time recursive implementation is possible [13].
- The above algorithm also provides possibility of direct symbol recovery from $\mathbf{E}_d$.
- An alternative implementation that requires less storage can be derived by replacing (19) using the sample covariance of $\mathbf{Y}$, $\tilde{\mathbf{Z}}_d$ and $\tilde{\mathbf{E}}_d$. Although such an implementation resembles the (stochastic) mean square error smoothing (MSS) approach in [12], the LSS and MSS are fundamentally different in that the latter requires uncorrelated input sequence. The covariance structure from the statistical properties of the input sequence is exploited in [12] whereas no such structure should be imposed in the least squares formulation.

4. SIMULATION

To evaluate the performance of the proposed algorithm, we compared LSS with several existing deterministic schemes including the least squares algorithms [11], the subspace algorithm [5] and TSML [4]. We have also compared with recent stochastic methods based on multistep linear prediction (MSP) [3] and smoothing [12]. The performance of these algorithms is measured by their estimated mean squared error of the normalized channel ($||\mathbf{h}|| = 1$)

$$\hat{MSE} \triangleq \frac{1}{M}\sum_i ||\mathbf{h} - \hat{\mathbf{h}}^{(i)}||^2. \quad (20)$$

The signal-to-noise ratio (SNR) is defined by $SNR = 10\log 10 \frac{||\mathbf{h}||^2\sigma_s^2}{P\sigma_n^2}$.

Figure 2 shows the MSE performance vs. SNR. All deterministic methods are comparable in this case and they all approach the Cramér-Rao bound at high SNR. Comparing with statistical approaches, it is clear that deterministic approaches are efficient at high SNR. In contrast, statistical methods are limited by the number of samples used in the covariance estimation, which explains the "flooring" effect when $SNR \to \infty$.

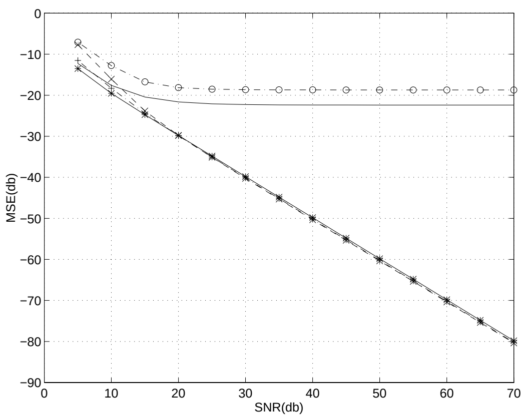

Figure 2: Channel in [4]: $*$:LSS, $+$:LS/SS, $\times$:TSML. $-$: MSS, o: MSP. 50 Monte Carlo Runs. 100 input symbols

REFERENCES

[1] K. Abed-Meraim, E. Moulines, and P. Loubaton. "Prediction error method for second-order blind identification". *IEEE Trans. Signal Processing*, SP-45(3):694–705, March 1997.

[2] Z. Ding. "An outer-product decomposition algorithm for multichannel blind identification". In *Proc. 8th IEEE Signal Processing Workshop on Statistical and Array Signal Processing*, Corfu, Greece, June 1996.

[3] D. Gesbert and P. Duhamel. "Robust blind identification and equalization based on multi-step predictors". In *Proc. IEEE Intl. Conf. Acoust. Speech, Sig. Proc.*, volume 5, pages 2621–24, Munich, Germany, April 1997.

[4] Y. Hua. "Fast maximum likelihood for blind identification of multiple FIR channels". *IEEE Trans. Signal Processing*, SP-44(3):661–672, March 1996.

[5] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue. "Subspace-Methods for the Blind Identification of Multichannel FIR Filters". *IEEE Trans. SP*, SP-43(2):516–525, Feb. 1995.

[6] D. Slock. "Blind Fractionally-Spaced Equalization, Perfect Reconstruction Filterbanks, and Multilinear Prediction". In *Proc. ICASSP'94 Conf.*, Adelaide, Australia, April 1994.

[7] L. Tong and J. Bao. "Equalizations in Wireless ATM". In *Proc. 1997 Allerton Conference on Communications, Control, and Computing*, Urbana, IL, 1997.

[8] L. Tong and S. Perreau. "Multichannel blind channel estimation: From subspace to maximum likelihood methods". *to appear in IEEE Proceedings, 1998.*

[9] L. Tong, G. Xu, and T. Kailath. "Blind identification and equalization based on second-order statistics: A time domain approach". *IEEE Trans. Information Theory*, 40(2):340–349, March 1994.

[10] L. Tong and Q. Zhao. "Linear Smoothing Approach to Blind Channel Estimation and Deconvolution". *to be submitted.*

[11] G. Xu, H. Liu, L. Tong, and T. Kailath. "A Least-Squares Approach to Blind Channel Identification". *IEEE Trans. Signal Processing*, SP-43(12):2982–2993, December 1995.

[12] H. Zeng, S. Perreau, and L. Tong. "Blind Channel Identification using Second Order Statistics: A Smoothing Approach". *submitted to IEEE Signal Processing Letters*, August, 1997.

[13] Q. Zhao and Lang Tong. "An Adaptive Linear Smoothing Approach to Blind Channel Estimation and Deconvolution". *to be submitted.*