Fast Maximum Likelihood for Blind Identification of Multiple FIR Channels

Yingbo Hua

Department of Electrical and Electronic Engineering
The University of Melbourne
Parkville, Victoria 3052, Australia
fax: +61-3-344-6678
email: yhua@ee.mu.oz.au

Abstract

We study the estimation of the impulse responses of multiple FIR channels which are driven by a virtually arbitrary unknown input sequence (including nonstationary, non-Gaussian, or/and colored sequences). The channel identifiability conditions are derived using a Fisher information matrix, which is in contrast to several existing approaches which are based on some specific estimation methods. A fast maximum likelihood method is developed, which is a two-step procedure where each step minimizes a quadratic function. The two-step maximum likelihood (TSML) method requires no initial estimates and is asymptotically (high SNR) optimum. Its implementation is discussed. An orthogonal complement system matrix is introduced which lays a foundation for this work.

1. Introduction

Blind identification of multiple FIR channels is an important problem in many applications such as mobile communications, seismic signal analysis, and image restoration. For some applications such as mobile communications, the unknown (unaccessible) input to the FIR channels is known to have certain statistical or/and algebraic characteristics, but for many others, the unknown input could be virtually arbitrary such as nonstationary, non-Gaussian, and colored. In fact, even in mobile communications, when a fast varying channel needs to be identified within a very short period of time, any known statistical characteristics (such as whiteness) of the unknown input becomes useless because the available data sequence is too short. Therefore, under many practical conditions, the input has to be assumed to be virtually arbitrary.

Blind identification of multiple FIR channels with virtually arbitrary input was recently studied by Liu-Xu et al [1], [2]. They developed a matrix which is directly constructed from the observation data in such a way that the least principal eigenvector of the matrix is equal to the exact channel impulse response (with an unknown scale) when the noise is absent. They also studied the channel identifiability conditions based on this matrix. For short observation data, Liu-Xu et al' s method performs significantly better than the second order statistics (SOS) based methods developed by Tong et al [3], [4] and Li-Ding [5], and the existing higher order statistics (HOS) based methods such as those by Giannakis-Mendel [6], Porat-Friedlander [7], and Tugnait [8]. This is because both the SOS and HOS methods require infinite number of observation data to yield the exact estimate even when the noise is absent.

In this paper, we present a further investigation of the model proposed by Liu-Xu et al. Instead of relying on an intuitive approach as Liu-Xu et al did (i.e., the exploitation of cross relations between channel outputs), we will study the model in a more formal way. In particular, we will develop the channel identifiability (ID) conditions based on a Fisher information (FI) matrix. Note that the inverse of the FI matrix is the Cramer-Rao lower bound (CRB) on the covariance matrix of any unbiased estimator. The CRB is achievable asymptotically (large SNR in this paper) by the maximum likelihood (ML) estimator. Hence, the channel ID conditions are inherent in the FI matrix. Furthermore, we will develop the ML estimator into a two-step (TS) procedure where each step minimizes a quadratic function. The TSML method is asymptotically efficient, i.e., achieving the CRB. An interesting relationship between the TSML method and Liu-Xu et al' s method will be shown.

Using the ML principle for this blind indentification problem was also recently studied by Shao-Nikias [10] and Slock [11]. In [10], an iterative algorithm was developed based on the estimate-maximize (EM) principle, but it requires a good initial estimate of the channel impulse response. In [11], a concept (iterative quadratic maximum likelihood) similar to that used in this paper was mentioned, but no in-depth study was indicated.

The rest of paper is organized as follows. The channel model is shown in Section 2, where an orthogonal complement system matrix is also introduced which lays a foundation for this work. The FI matrix of the channel model and the channel ID conditions are developed in Section 3. The TSML method is developed in Section 4. An efficient implementation of the TSML method is discussed in Section 5. A simulation example is given in Section 6.

1058-6393/95 $4.00 © 1995 IEEE

2. The channel model and an orthogonal complement matrix

Consider M FIR channels which are driven by an arbitrary input, i.e.,

$$\begin{cases} y_1(k) = s(k) * h_1(k) + w_1(k) \\ y_2(k) = s(k) * h_2(k) + w_2(k) \\ \dots \\ y_M(k) = s(k) * h_M(k) + w_M(k) \end{cases} \tag{2.1}$$

where $\{y_i(k), i=1,...,M\}$ are the channel outputs, s(k) the input, $\{h_i(k), i=1,...,M\}$ the finite impulse responses (FIR) of the channels, and $\{w_i(k), i=1,...,M\}$ the noise. Assume that the maximum order of the M channels is L, and the number of available output samples of each channel is N. "*" denotes convolution.

For mathematical convenience, we put the channel outputs in vector form as follows. Define the observation vector **y**:

$$\mathbf{y} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \dots \quad \dots \quad \mathbf{y}_M]^T \tag{2.2}$$

where $\mathbf{y}_i$ is the observation vector from the i^{th} channel:

$$\mathbf{y}_i = [y_i(0) \quad y_i(1) \quad \dots \quad \dots \quad y_i(N-1)]^T \tag{2.3}$$

and ^{T} denotes the transpose. Then,

$$\mathbf{y} = \mathbf{H}_M \mathbf{s} + \mathbf{w} \tag{2.4}$$

where **w** is the noise vector defined in the same way as **y**, and **s** is the input vector, i.e.,

$$\mathbf{s} = [s(-L) \quad s(-L+1) \quad \cdots \quad s(N-1)]^T \tag{2.5}$$

and $\mathbf{H}_M$ is the system matrix defined as:

$$\mathbf{H}_M = \begin{bmatrix} \mathbf{H}_{(1)} \\ \mathbf{H}_{(2)} \\ \vdots \\ \mathbf{H}_{(M)} \end{bmatrix} \tag{2.6}$$

where $\mathbf{H}_{(i)}$ is the i^{th} channel matrix (N×(N+L)):

$$\mathbf{H}_{(i)} = \begin{bmatrix} h_i(L) & \cdots & h_i(0) & & \\ & \ddots & & \ddots & \\ & & h_i(L) & \cdots & h_i(0) \end{bmatrix} \tag{2.7}$$

The system matrix $\mathbf{H}_M$ is a generalized Sylvester matrix which has a number of important properties as shown in [12]. In the following, we mention one of them. Define

$$\mathbf{G}_2^H = [-\overline{\mathbf{H}}_{(2)} \quad \overline{\mathbf{H}}_{(1)}]$$

$$\mathbf{G}_M^H = \left[\begin{array}{ccc|c} & \mathbf{G}_{M-1}^H & & \mathbf{0} \\ \hline -\overline{\mathbf{H}}_{(M)} & & & -\overline{\mathbf{H}}_{(1)} \\ & \ddots & & \vdots \\ & & -\overline{\mathbf{H}}_{(M)} & -\overline{\mathbf{H}}_{(M)} \end{array}\right] . \tag{2.8}$$

where $\overline{\mathbf{H}}_{(i)}$ is the top-left (N-L)×N submatrix of $\mathbf{H}_{(i)}$.

Theorem 1(shown in [12]):

Provided that all channels do not share a common zero and N≥2L (or N≥(L+1) for two channels case), an *orthogonal complement* matrix of the system matrix $\mathbf{H}_M$ is $\mathbf{G}_M$, i.e.,

$$\mathbf{P}_G + \mathbf{P}_H = \mathbf{I}$$

where $\mathbf{P}_G$ and $\mathbf{P}_H$ denote the projection matrices onto range($\mathbf{G}_M$) and range($\mathbf{H}_M$), respectively.

The orthogonal complement matrix $\mathbf{G}_M$ is not a full column rank matrix (i.e., $\mathbf{G}_M$ is "fat"). Searching for a "leaner" orthogonal complement matrix which is constructed *linearly* from the observation with a *known linear* and *constant* transformation is still a challenge. (Deleting any columns of $\mathbf{G}_M$ does not seem to work). Furthermore, whether or not $\mathbf{G}_M$ is already the "leanest" one can get under the proper constraint is unknown. Nevertheless, the matrix $\mathbf{G}_M$ is very useful as shown next.

3. The Fisher information matrix

Assuming that the channel noise is white Gaussian and the input sequence is unknown and deterministic, the channel output vector **y** is then Gaussian distributed with the mean vector $\mathbf{m}=\mathbf{H}_M\mathbf{s}$ and the covariance matrix $\mathbf{R}_y=\sigma^2\mathbf{I}$. Then it can be shown [12] that the Fisher information (FI) matrix is

$$\mathbf{F}=\frac{2}{\sigma^2}\begin{bmatrix}\mathrm{Re}(\mathbf{F}_c) & -\mathrm{Im}(\mathbf{F}_c)\\ \mathrm{Im}(\mathbf{F}_c) & \mathrm{Re}(\mathbf{F}_c)\end{bmatrix} \tag{3.1}$$

where $\mathbf{F_C}$ is a (ML+M+N+L)×(ML+M+N+L) *complex* FI matrix defined as:

$$\mathbf{F}_c=\mathbf{Q}^T\mathbf{Q} \tag{3.2}$$

with

$$\mathbf{Q}=\begin{bmatrix}\mathbf{I}_M\otimes\mathbf{S} & \mathbf{H}_M\end{bmatrix} \tag{3.3}$$

and

$$\mathbf{S}=\begin{bmatrix} s(0) & s(-1) & \dots & s(-L)\\ s(1) & s(0) & \dots & s(-L+1)\\ \vdots & \vdots & \vdots & \vdots\\ s(N-1) & s(N-2) & \dots & s(N-L-1)\end{bmatrix} \tag{3.4}$$

and $\mathbf{I_M}$ is the M×M identity matrix, and ⊗ denotes the Kronecker product. In fact, $\mathbf{I_M}\otimes\mathbf{S}$=diag(S,...,S). The (i, j)th element of $\mathbf{F_C}$ corresponds to the ith and jth elements of the complex parameter vector defined as:

$$\mathbf{a}_c=\begin{bmatrix}\alpha_{c,1}\\ \alpha_{c,2}\\ \vdots\\ \alpha_{c,ML+M+N+L}\end{bmatrix}=\begin{bmatrix}\mathbf{h}\\ \mathbf{s}\end{bmatrix} \tag{3.5}$$

It is easy to show that if $\mathbf{F_C}^{-1}$ exists, then the CRB is

$$\mathbf{F}^{-1}=\frac{\sigma^2}{2}\begin{bmatrix}\mathrm{Re}(\mathbf{F}_c^{-1}) & -\mathrm{Im}(\mathbf{F}_c^{-1})\\ \mathrm{Im}(\mathbf{F}_c^{-1}) & \mathrm{Re}(\mathbf{F}_c^{-1})\end{bmatrix} \tag{3.6}$$

Theorem 2 (shown in [12]):

(a) rank($\mathbf{F_C}$) < dim($\mathbf{F_C}$) = ML+M+N+L.

(b) rank($\mathbf{F_C}$) < dim($\mathbf{F_C}$) - 1 if s(k), k=-L,...,N-1, has no more than L+1 modes, or the M channels share one or more common zeros.

(c) rank($\mathbf{F_C}$) = dim($\mathbf{F_C}$) - 1 if s(k), k=-L,...,N-1, has 2L+1 or more modes, the M channels do not share a common zero, and N≥3L+1.

(d) Let $\mathbf{F}_{C,i}$ be $\mathbf{F_C}$ with its ith row and ith column deleted. Provided that the conditions of (c) are true, then rank($\mathbf{F}_{C,i}$) = dim($\mathbf{F}_{C,i}$) if $\alpha_{C,i} \neq 0$, and rank($\mathbf{F}_{C,i}$) = dim($\mathbf{F}_{C,i}$) -1 if $\alpha_{C,i} = 0$.

The fact that $\mathbf{F_C}$ is always rank deficient (Part (a)) means that the CRB $\mathbf{F_C}^{-1}$ does not exist and one can not find all the parameters in the system even when the noise approaches to zero. However, the equation rank($\mathbf{F_C}$) = dim($\mathbf{F_C}$) - m (Parts (b) and (c)) means that when a set of m complex parameters are known or fixed, one can find all the other parameters. Part (d) means that under the condition of (c) and when any nonzero complex parameter is fixed, all other parameters in the system can be found. The conditions under which the channel is identifiable up a complex constant are commonly referred to as the ID conditions. Part (c) provides a sufficient ID condition, and Part (b) implies two necessary ID conditions. Clearly, the sufficient and necessary ID condition is that rank($\mathbf{F_C}$) = dim($\mathbf{F_C}$)-1, or equivalently, Q has a one-dimensional null space. The above channel ID conditions are identical to those shown by Liu-Xu et al [1]-[2] based on a matrix equation. Also note that the finite sequence s(k), k=-L,...,N-1, can have at most N+L independent modes. The third part of the condition required in Part (c), N≥3L+1, is also implied by the first part, N+L≥2L+1.

4. The Fast Maximum Likelihood

In the sequel, we will use **H** for $\mathbf{H_M}$, and **G** for $\mathbf{G_M}$. Assuming that the channel noise is white Gaussian, it is easy to show that the maximum likelihood (ML) estimation of **h** is obtained by

$$\max_{\mathbf{h}} \mathbf{y}\mathbf{P}_H\mathbf{y}^H \tag{4.1}$$

or equivalently,

$$\min_{\mathbf{h}} \mathbf{y}\mathbf{P}_G\mathbf{y}^H \tag{4.2}$$

Once the ML estimate of **h** is available, the ML estimate of s is then given by $(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{y}$.

It is clear that the projection matrix $\mathbf{P_H}=\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$ is invariant to any complex constant scale on **H** or **h** and hence (4.1) or (4.2) does not have an absolutely unique solution. This was also predicted by the FI matrix. In the sequel, we will assume that the sufficient ID condition as stated in Theorem 2 is satisfied so that the channels are identifiable up to an arbitrary complex constant. For the ML approach to yield an unique solution, we can subject (4.1) or (4.2) to that a nonzero element of **h** is fixed to be a constant. Alternatively, we can simply add to (4.1) or (4.2) that $||\mathbf{h}||=1$. This will leads to a solution which has an arbitrary factor of unit amplitude.

A fast ML estimation can now be developed as follows. We begin by writing

$$\mathbf{y}^H\mathbf{P}_G\mathbf{y}=\mathbf{y}^H\mathbf{G}(\mathbf{G}^H\mathbf{G})^{+}\mathbf{G}^H\mathbf{y} \tag{4.3}$$

where $^+$ denotes the pseudoinverse. One can verify that

$$\mathbf{G}^H \mathbf{y} = \mathbf{Y}\mathbf{h} \tag{4.4}$$

where

$$\mathbf{Y} = \mathbf{Y}_M = \left[\begin{array}{ccc|c} & \mathbf{Y}_{M-1} & & \mathbf{0} \\ \hline \overline{\mathbf{Y}}_{(M)} & & & -\overline{\mathbf{Y}}_{(1)} \\ & \ddots & & \vdots \\ & & \overline{\mathbf{Y}}_{(M)} & -\overline{\mathbf{Y}}_{(M-1)} \end{array}\right] \tag{4.5}$$

$$\overline{\mathbf{Y}}_{(i)} = \begin{bmatrix} y_i(L) & \cdots & y_i(0) \\ \vdots & \vdots & \vdots \\ y_i(N-1) & \cdots & y_i(N-L-1) \end{bmatrix} \tag{4.6}$$

Hence, the ML is to minimize

$$\mathbf{h}^H \mathbf{Y}^H (\mathbf{G}^H \mathbf{G})^+ \mathbf{Y}\mathbf{h} \tag{4.7}.$$

This expression suggests a two-step estimation procedure as shown below:

The TSML Method:

Step 1: Minimize $\mathbf{h}^H(\mathbf{Y}^H\mathbf{Y})\mathbf{h}$ with $||\mathbf{h}||=1$ to yield $\mathbf{h}_c$.

Step 2: Minimize $\mathbf{h}^H(\mathbf{Y}^H(\mathbf{G}_c{}^H\mathbf{G}_c)^+\mathbf{Y})\mathbf{h}$ with $||\mathbf{h}||=1$ to yield $\mathbf{h}_e$, where $\mathbf{G}_c$ is constructed from $\mathbf{h}_c$.

It is interesting to observe that Step 1 of the TSML method coincides with the technique developed by Liu-Xu et al [1]-[2]. It can be shown that Step 1 yields the exact estimate in the absence of noise (i.e., Step 1 is consistent). It can also be shown [12] that Step 2 yields a statistically efficient estimate when $\mathbf{G}_c$ is constructed from a consistent estimate of $\mathbf{h}$.

5. Implementation of the TSML Method

The above formulation of the TS-ML method is given at a fairly high level. The computational efficiency of the TSML method still largely depends on how it is implemented. Note that the matrices G and Y are sparse, and the pseudoinverse $(G^HG)^+$ (with a known rank) needs to be computed. Following a detailed analysis as shown in [12], an efficient implementation (using the URV decomposition [9]) of the TSML method requires $O\{NM^2L\}$ flops for Step 1 and $O\{N^2M^3L\}$ flops for Step 2 assuming N>>L>>1 and M>>1.

6. Simulations

While the TSML method is applicable to a wide range of situations, we consider a (simulated) mobile communication channel for the illustration purpose. Let the observation y(k) consist of the fractionally sampled output of the channel, the unknown input s(i) be a white binary sequence (+1 and -1), the total number of baud intervals observed be N=30, the fractional-sampling rate (i.e., the sampling rate over the baud rate) be M=3, the length of the channel impulse response h(k) be M(L+1)=18, and the impulse response be constructed as

h(k) = g(k) - 0.8 g(k-2) + 0.4 g(k-9)

where g(k) is generated from the well-known raised-cosine-spectrum pulse with the roll-off factor equal to 1, which is windowed (truncated), delayed and sampled. h(k) is shown in Figure 1.

Note that the channel output samples y(k), k = 0, 1, ..., NM-1, and the channel impulse response h(k), k = 0, 1, ..., M(L+1)-1, relate to the multichannel notations $y_i(k)$ and $h_i(k)$ as follows:

$y_i(k) = y(kM+i-1)$, k = 0, 1, ..., N-1; i = 1, 2, ..., M;
$h_i(k) = h(kM+i-1)$, k = 0, 1, ..., L; i = 1, 2, ..., M.

Let the variance of the additive white Gaussian noise be denoted by $\sigma_w{}^2$, and the variance of the input sequence by $\sigma_s{}^2$ ($\sigma_s{}^2$ = 1 for the binary sequence). Define the signal-to-noise ratio as:

$$SNR = 10\log_{10}\left(\frac{E\{\|\mathbf{y}\|^2\}}{E\{\|\mathbf{w}\|^2\}}\right).$$

This can be shown to be

$$SNR = 20\log_{10}\left(\frac{\|\mathbf{h}\|\sigma_s}{\sqrt{M}\sigma_w}\right).$$

For SNR = 44.9 dB (i.e., σ_w=0.005), 75 independent runs are shown in Figures 2(a) and 1(b) for the first step and the second step of the TSML method, respectively. In the simulation, both the input sequence and the noise sequence were independently chosen at each run. All estimates of h(k) are normalized in such a way that at the peak value position of h(k), they have the same value as h(k).

Define the normalized-root-mean-square-error of the estimated channel impuse responses as

$$NRMSE = \frac{1}{\|\mathbf{h}\|}\sqrt{\frac{1}{RUNS}\sum_{i=1}^{RUNS}\left\|\hat{\mathbf{h}}_i - \mathbf{h}\right\|^2}$$

where $\hat{\mathbf{h}}_i$ denotes the ith run estimate of $\mathbf{h}$. For SNR = 44.9dB, NRMSE = 0.1256 for the first step, and NRMSE = 0.0977 for the second step.

7. Conclusions

We have analyzed the Fisher information matrix of a multiple-channel model where the input is unknown deterministic (which can be a realization of a nonstationary, non-Gaussian, non-white, or virtually any random process). The channel identifiability conditions have been derived based on the Fisher information matrix. These conditions have been further shown to be equivalent to those developed by Liu-Xu et al based on a linear equation. We have also developed a fast maximum likelihood method based on this channel model. This fast ML method consists of two steps where each step minimizes a quadratic function, hence called TSML method. The first step of the TSML method, which turns out to be the same as the method developed by Liu-Xu et al, is consistent. The second step of the TSML method yields an asymptotically efficient estimate of the channel impulse response. A flop count of the TSML method based on an efficient implementation has been provided.

8. References

[1] G. Xu, H. Liu, L. Tong, and T. Kailath, "A deterministic approach to blind channel identification," submitted to *IEEE Trans. on Signal Processing*.

[2] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind channel identification of multi-channel FIR systems," *Proc of IEEE ICASSP'94*, Vol. 4, pp. 581-584, Adelaide, Australia, April 1994.

[3] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. Info. Theory*, Vol. 40, No. 2, March 1994. A short version in *Proc. of 25th Asilomar Conf. Signal, Systems and Computers*, Nov. 1991.

[4] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: a frequency-domain approach," submitted to *IEEE Trans. Info. Theory*.

[5] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics," *Proc. of IEEE ICASSP'93*, pp.81-84, Minneapolis, MN, April 1993.

[6] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 3, pp. 360-377, March 1989.

[7] B. Porat and B. Friedlander, "Blind equalization of digital communication channel using higher-order moments," *IEEE Transactions on Signal Processing*, Vol. 39, No. 2, pp. 522-526, Feb. 1991.

[8] J. K. Tugnait, "Fractionally spaced blind equalization and estimation of FIR channels," *International Conference on Communications*, pp.428-432, Geneva, Swizerland, June 1993.

[9] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Transactions on Signal Processing*, pp.1535-1514, June 1992.

[10] M. Shao and C. L. Nikias, "An ML/MMSE estimation approach to blind equalization," *Proc of IEEE ICASSP'94*, Vol. 4, pp. 569-572, Adelaide, Australia, April 1994.

[11] D. T. M. Slock, "Blind fractionally-space equalization, perfect-reconstruction filter bank and multichannel linear prediction," *Proc. of IEEE ICASSP'94*, Vol. 4, pp. 585-588, Adelaide, Australia, April 1994.

[12] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," submitted to *IEEE Transactions on Signal Processing*.

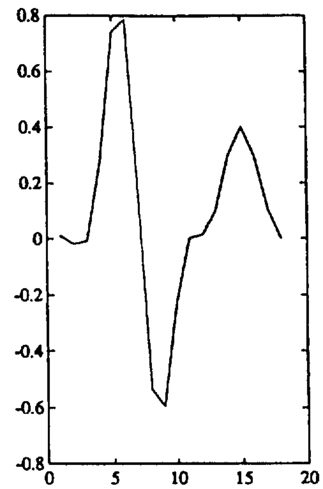

Figure 1

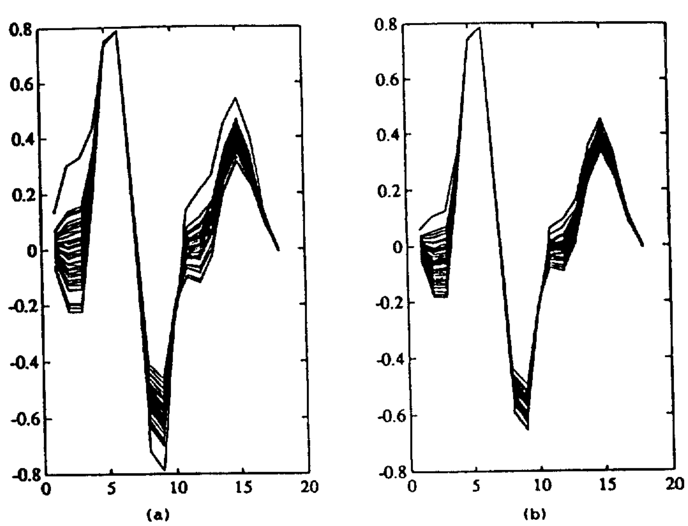

Figure 2