A LEAST SQUARES COMPONENT NORMALIZATION APPROACH TO BLIND CHANNEL IDENTIFICATION

Carlos Avendano

CIPIC
University of California, Davis CA 95616
Email: carlos@ece.ucdavis.edu

Jacob Benesty and Dennis R. Morgan

Bell Laboratories, Lucent Technologies
700 Mountain Av., Murray Hill, NJ 07974
Email: { jbenesty,drrm }@bell-labs.com

ABSTRACT

We describe a new method for blind system identification that uses the cross relation properties between two or more sensor signals to estimate the impulse responses of the channels. The method performs as well or better than other similar blind identification techniques under noisy and ill-conditioned channel conditions, and is computationally simpler to implement.

1. INTRODUCTION

In most applications, the degradation of the signal caused by a propagation channel is not acceptable. Traditionally, a known pilot or training signal is sent through the channel to identify it and correct or equalize it. However, in many cases, the use of such a training signal is not possible.

Blind identification and equalization techniques that do not require training signals have been proposed. In an early paper, Sato [1] presented such a technique and performed channel equalization directly. Blind equalization of non-minimum phase channels, which cannot be inverted, led to first considering channel identification before performing equalization. Schemes for blind system identification that use higher order statistics have been proposed [2]. Most such methods require long segments of data to identify the channel and therefore are unattractive for fast equalization.

Using only second-order statistics for channel identification was first proposed in [3]. These techniques require observations from two or more sensors or, equivalently, the oversampling of a single observation. Several such algorithms have been reported [4]-[6].

This paper presents a new blind system identification technique that uses only second order statistics. The method proposed exploits the linear relationships between the observations of multiple sensors as in [4], but since it does not require an eigenvalue decomposition, it is less complex and more attractive computationally. More importantly, this new method outperforms and is more robust than previous methods under very noisy and ill-conditioned observations. We demonstrate significantly improved performance, as compared to similar techniques, for a two sensor well-conditioned system, a two sensor ill-conditioned system, and a three sensor system with random channels. The simulation results show a reduction of up to 5 dB in mean squared error for SNR ranging from 0 to 25 dB.

2. PROBLEM STATEMENT

Consider the multi-channel FIR system of Fig.1. The objective is to estimate the channels' responses h_m given a known input signal

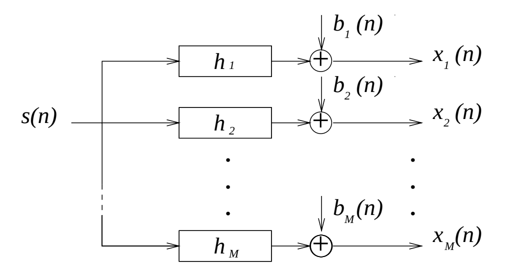

Figure 1: Multi-channel FIR system.

$s(n)$ and M ($M \geq 2$) noisy observations $x_m(n)$. The mth observation sequence $\mathbf{x}_m(n)$ is written as:

$$\mathbf{x}_m(n) = \mathbf{H}_m \mathbf{s}(n) + \mathbf{b}_m(n), \tag{1}$$

where

$$\mathbf{x}_m(n) = \begin{bmatrix} x_m(n) & x_m(n-1) & \cdots & x_m(n-L+1) \end{bmatrix}^T,$$
$$\mathbf{b}_m(n) = \begin{bmatrix} b_m(n) & b_m(n-1) & \cdots & b_m(n-L+1) \end{bmatrix}^T,$$
$$\mathbf{s}(n) = \begin{bmatrix} s(n) & s(n-1) & \cdots & s(n-2L+2) \end{bmatrix}^T,$$

and

$$\mathbf{H}_m = \begin{bmatrix} h_{m,0} & \cdots & h_{m,L-1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & h_{m,0} & \cdots & h_{m,L-1} \end{bmatrix}.$$

The observation noise $b_m(n)$ is assumed to be uncorrelated with the source signal $s(n)$. The parameter matrix $\mathbf{H}_m$ is a $[L, 2L-1]$ matrix constructed from the vector:

$$\mathbf{h}_m(n) = \begin{bmatrix} h_{m,0} & h_{m,1} & \cdots & h_{m,L-1} \end{bmatrix}^T,$$

which is the impulse response corresponding to the mth channel. In the present paper this vector is assumed to be of length L, which without loss of generality we set to the length of the longest channel impulse response.

Now, the global system response can be written as:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{b}(n), \tag{2}$$

with

$$\mathbf{x}(n) = \begin{bmatrix} \mathbf{x}_1^T(n) & \cdots & \mathbf{x}_M^T(n) \end{bmatrix}^T,$$

$$\mathbf{b}(n) = \begin{bmatrix} \mathbf{b}_1^T(n) & \cdots & \mathbf{b}_M^T(n) \end{bmatrix}^T,$$

and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1^T & \cdots & \mathbf{H}_M^T \end{bmatrix}^T.$$

The blind system identification technique presented in this paper belongs to a class of techniques that are based on the correlation between sensor outputs [4]-[6]. This correlation can be expressed as:

$$\mathbf{R}_{xx} = \mathbf{H}\mathbf{R}_{ss}\mathbf{H}^T + \mathbf{R}_{bb}, \tag{3}$$

where

$$\mathbf{R}_{xx} = E\{\mathbf{x}(n)\mathbf{x}^T(n)\},$$
$$\mathbf{R}_{bb} = E\{\mathbf{b}(n)\mathbf{b}^T(n)\},$$
$$\mathbf{R}_{ss} = E\{\mathbf{s}(n)\mathbf{s}^T(n)\},$$

and $E\{\cdot\}$ denotes mathematical expectation. $\mathbf{R}_{xx}$ is the covariance matrix of the observation, and the signal and noise covariance matrices are $\mathbf{R}_{ss}$ and $\mathbf{R}_{bb}$ respectively. Note that $\mathbf{R}_{ss}$ has dimension $[2L-1, 2L-1]$ whereas $\mathbf{R}_{xx}$ and $\mathbf{R}_{bb}$ are $[ML, ML]$.

In the absence of noise, the assumptions under which the parameter matrix $\mathbf{H}$ can be identified (up to a constant scalar) and which are assumed to hold throughout the rest of this paper are the following [3]:

1. The parameter matrix $\mathbf{H}$ is of full column rank. This implies that the channels do not share common zeros.
2. The autocorrelation matrix $\mathbf{R}_{ss}$ is of full rank.

3. THE COMPONENT NORMALIZATION METHOD

Although our interest is in the real situation where noise is present, we will use the noise free condition to develop the new approach which we call Component Normalization. Then we apply it to the noisy situation using a least squares formulation.

3.1. Noise-free Observations

The method is based on the following linear relationships between the sensor outputs [4]:

$$\mathbf{x}_i^T(n)\mathbf{h}_j = \mathbf{x}_j^T(n)\mathbf{h}_i,\ i,j = 1,2,...,M,\ i \neq j. \tag{4}$$

These linear relations follow from the fact that $x_i = s * h_i$, and thus $x_i * h_j = s * h_i * h_j = x_j * h_i$ (the $*$ symbol is the linear convolution operator). If we pre-multiply (4) by $\mathbf{x}_i(n)$ and take the expectation we get a set of $M(M-1)$ relations of the form:

$$\mathbf{R}_{x_i,x_i}\mathbf{h}_j = \mathbf{R}_{x_i,x_j}\mathbf{h}_i,\ i,j = 1,2,...,M,\ i \neq j. \tag{5}$$

The set of equations described by (5) can be arranged in the following form that reduces the number of equations from $M(M-1)$ to M [4]:

$$\mathbf{R}\mathbf{h} = \mathbf{0}, \tag{6}$$

where

$$\mathbf{R} = \begin{bmatrix} \sum_{i\neq 1}\mathbf{R}_{x_i,x_i} & -\mathbf{R}_{x_2,x_1} & \cdots & -\mathbf{R}_{x_M,x_1} \\ -\mathbf{R}_{x_1,x_2} & \sum_{i\neq 2}\mathbf{R}_{x_i,x_i} & \cdots & \\ \vdots & & \ddots & \vdots \\ -\mathbf{R}_{x_1,x_M} & \cdots & & \sum_{i\neq M}\mathbf{R}_{x_i,x_i} \end{bmatrix},$$

and

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1^T & \mathbf{h}_2^T & \cdots & \mathbf{h}_M^T \end{bmatrix}^T.$$

Equation (6) has to be solved for the unknown $\mathbf{h}$. Note that in contrast to the cross-relation (CR) method proposed in [4] we work with the covariance matrix rather than operating directly on the observed data.

Since the vector $\mathbf{h}$ is different from zero (the trivial solution), the $[ML, ML]$ matrix $\mathbf{R}$ is rank deficient. The rank of $\mathbf{R}$ is $ML-1$, which is the condition for unique channel identifiability in [5]. This condition is required for (6) to have a unique solution and is a consequence of the identifiability conditions previously listed. Now we choose an element of the vector $\mathbf{h}$ which is known to be non-zero, say, $h_{m,k} = \alpha$. If we denote the column of $\mathbf{R}$ corresponding to that element as

$$\mathbf{r} = E\left\{ \begin{bmatrix} -\mathbf{x}_m(n)x_1(n-k) \\ \vdots \\ \sum_{i\neq m}\mathbf{x}_i(n)x_i(n-k) \\ \vdots \\ -\mathbf{x}_m(n)x_M(n-k) \end{bmatrix} \right\}, \tag{7}$$

we can write (6) as:

$$\tilde{\mathbf{R}}\tilde{\mathbf{h}} = -\alpha\mathbf{r}, \tag{8}$$

with the following definitions. The matrix $\tilde{\mathbf{R}} =$

$$\begin{bmatrix} \sum_{i\neq 1}\mathbf{R}_{x_i,x_i} & \cdots & -\mathbf{R}_{x_m,\tilde{x}_1} & \cdots & -\mathbf{R}_{x_M,x_1} \\ -\mathbf{R}_{x_1,x_2} & \cdots & -\mathbf{R}_{x_m,\tilde{x}_2} & \cdots & -\mathbf{R}_{x_M,x_2} \\ \vdots\ \ddots & \vdots & \vdots & & \\ -\mathbf{R}_{x_1,x_m} & & \sum_{i\neq m}\mathbf{R}_{x_i,\tilde{x}_i} & \cdots & \cdots \\ \vdots & \cdots & & \ddots & \\ -\mathbf{R}_{x_1,x_M} & & -\mathbf{R}_{x_m,\tilde{x}_M} & \cdots & \sum_{i\neq M}\mathbf{R}_{x_i,x_i} \end{bmatrix}$$

is of size $[ML, ML-1]$, where

$$\mathbf{R}_{x_i,\tilde{x}_j} = E\{\mathbf{x}_i(n)\tilde{\mathbf{x}}_j^T(n)\}$$

is an $[L, L-1]$ matrix and $\tilde{}$ indicates that we have removed the kth element from the corresponding vector. Vector $\tilde{\mathbf{h}}$ is of size $ML-1$ and it is constructed from $\mathbf{h}$ by removing only the kth element from $\mathbf{h}_m$:

$$\tilde{\mathbf{h}} = \begin{bmatrix} \mathbf{h}_1^T & \mathbf{h}_2^T & \cdots & \tilde{\mathbf{h}}_m^T & \cdots & \mathbf{h}_M^T \end{bmatrix}^T. \tag{9}$$

If we arbitrarily set $\alpha = 1$ and solve the set of equations represented by (8) we obtain the *component normalized* impulse responses of the multi-channel FIR system. We see that (8) can be solved analytically by removing any row from $\tilde{\mathbf{R}}$ and inverting the resulting square matrix. Since the system is overdetermined we could use the least squares (LS) pseudoinverse instead, but given that there is no noise, the solution would be the same, i.e. the unique solution (up to the arbitrary scalar factor α).

3.2. Noisy Observations

In the noisy case, the relationships (5) (8) no longer hold and the error term now becomes:

$$\mathbf{e} = \alpha \mathbf{r} + \tilde{\mathbf{R}}\tilde{\mathbf{h}}. \tag{10}$$

Under noisy conditions we minimize (in the LS sense) a cost function that depends on the error (10):

$$\min_{\tilde{\mathbf{h}}} \ \mathbf{e}^T \mathbf{e} \tag{11}$$

which has the following solution:

$$\tilde{\mathbf{h}} = -\alpha(\tilde{\mathbf{R}}^T \tilde{\mathbf{R}})^{-1} \tilde{\mathbf{R}}^T \mathbf{r}. \tag{12}$$

Note that (11) is equivalent to

$$\min_{\mathbf{h}} \ \mathbf{h}^T \mathbf{R}^T \mathbf{R} \mathbf{h} \tag{13}$$

subject to the constraint $h_{m,k} = \alpha$, whereas in the CR method [4] the minimization problem is

$$\min_{\mathbf{h}} \ \mathbf{h}^T \mathbf{R} \mathbf{h} \tag{14}$$

subject to various constraints. For $\| \mathbf{h} \| = 1$, (14) leads to a solution that is the eigenvector of **R** corresponding to its smallest eigenvalue. Another constraint proposed in [4] is $\mathbf{c}^T \mathbf{h} = 1$, where **c** is a constant vector. For the special case $\mathbf{c} = [0, ..., 1/\alpha, ..., 0]^T$ with $1/\alpha$ being the kth element of **c**, this constraint is equivalent to $h_{m,k} = \alpha$ which follows an error criterion similar to our formulation. However, we found solutions based on (14) to be somewhat less robust to noise than our method, especially in high noise conditions and as the number of channels increases; we obtained the best results, in all cases, with the proposed criterion (11).

It may seem somewhat arbitrary to minimize the error in (10) which involves least-squares fitting of statistical quantities that derive from idealized models. However, this approach can be further justified from another viewpoint stemming from linear estimation theory. The presence of noise causes unbalance on the relationships (4). According to the orthogonality principle this error is to be uncorrelated with the data vectors, and its minimization in the least squares sense leads to the CR solution. In a similar way, noise introduces unbalance on the relationships (5), which are used by the Component Normalization method. The error in this case is to be uncorrelated with the cross-correlations of the data vectors. If we explicitly write the error (11) for $M = 2$ we get

$$\mathbf{e} = \alpha E \left\{ \begin{bmatrix} \mathbf{x}_2(n) \\ -\mathbf{x}_1(n) \end{bmatrix} x_2(n-k) \right\} + \tag{15}$$

$$\begin{bmatrix} \mathbf{R}_{x_2,\tilde{x}_2} & -\mathbf{R}_{x_2,x_1} \\ -\mathbf{R}_{x_1,\tilde{x}_2} & \mathbf{R}_{x_1,x_1} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{h}}_1 \\ \mathbf{h}_2 \end{bmatrix}, \tag{16}$$

which we would like to set to zero. Because this is not possible, we then try to at least minimize $\mathbf{e}^T \mathbf{e}$, which is the essence of our error criterion (11). This result applies for the general case $M \geq 2$ as well.

4. SIMULATIONS

In this section we present a set of simulations performed to evaluate the least squares component normalization (LSCN) technique. The simulations are in the context of digital communications where the channel responses are short and the input sequences have flat spectra. We use the two-channel simulation example in [7] to compare our method to the CR method of [4]. Additionally, we evaluate the LSCN and CR methods for the case of three-channel and sensors. Unlike the simulations performed in [8], for three sensors we evaluate the two methods when the unknown channels are longer than in the two-channel case.

4.1. Two Sensors

In [7] the following two-channel FIR system was considered:

$$\mathbf{h}_1 = \begin{bmatrix} 1 & -2\cos(\theta) & 1 \end{bmatrix}^T \quad \text{and}$$
$$\mathbf{h}_2 = \begin{bmatrix} 1 & -2\cos(\theta+\delta) & 1 \end{bmatrix}^T,$$

where the parameter δ controls the angular proximity between the zeros of the channels. The source signal was a binary $(-1, 1)$ sequence with white spectrum and unit power. The signal consisted only of 30 samples and a different realization of noise (independent on each channel) with a given power was added to the channel outputs for each of 100 runs of the simulation to obtain the desired signal to noise ratio (SNR):

$$\mathrm{SNR} = 10 \log_{10} \frac{\sigma_s^2 \, \| \mathbf{h} \|^2}{\mathrm{M}\sigma^2}$$

where σ_s^2 is the signal power (equal to 1 here) and σ^2 is the noise power. Our performance measure is given by:

$$\mathrm{MSE} = 20 \log_{10} \left[\frac{1}{\| \mathbf{h} \|} \sqrt{\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \| \varepsilon^{(\mathrm{i})} \|^2} \right],$$

where $N = 100$ is the number of runs,

$$\varepsilon^{(i)} = \mathbf{h} - \frac{\mathbf{h}^T \hat{\mathbf{h}}^{(i)}}{\hat{\mathbf{h}}^{(i)T} \hat{\mathbf{h}}^{(i)}} \hat{\mathbf{h}}^{(i)}$$

is the normalized projection error, and $\hat{\mathbf{h}}^{(i)}$ is the channel estimate for the ith run. The above error measure reflects the fact that we are only trying to estimate **h** to within an arbitrary scale factor; further discussion on the philosophy of error measures in this context can be found in [9].

For the two channel case we performed the following two simulations:

1. Well-conditioned channels ($\theta = \pi/10$ and $\delta = \pi$):
   Noise was added at different SNR's, from $5 - 50$ dB and the corresponding MSE of the estimates was computed. The two methods, CR and LSCN, were used to estimate the channels.
2. Ill-conditioned channels ($\theta = \pi/10$ and $\delta = \pi/10$):
   The same noise conditions as in the first simulations.

The results of these two cases are shown respectively in Figs. 2 and 3. We can see that the LSCN method outperforms the CR method at high noise levels (low SNR) for an ill condition channel, while the two methods give similar results when the channels are well conditioned.

4.2. Three sensors

In this simulation we were primarily interested in looking at the performance of the LSCN method when the channel length increases. In a preliminary experiment, adding observations from additional sensors was found to give more robust estimates of long channels (128 taps) under noisy conditions. Another advantage of adding sensors is that the first identifiability condition is more likely to hold.

We used three channels whose coefficients were randomly selected. The channels were 15 taps long. The source signal was generated in the same way as in the previous simulations but was longer (120 samples) since the number of parameters to be determine is now larger. We added noise at different levels and computed the MSE over $N = 200$ runs. The channel estimation was done by both the CR and LSCN methods. In Fig.4 we plot the results for those conditions. The advantage of the LSCN over the CR is again clear for a significant range in SNR.

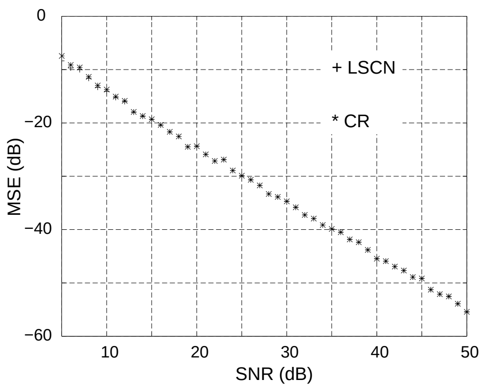

Figure 2: Comparison between LSCN and CR methods for the well-conditioned channel case ($\delta = \pi$).

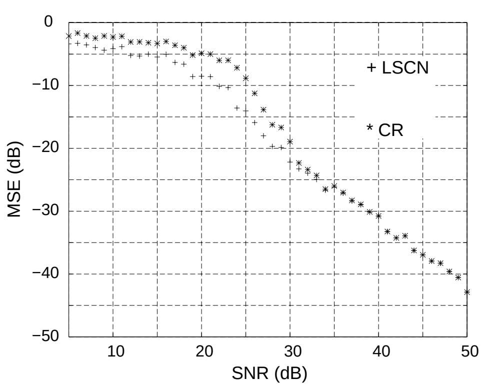

Figure 3: Comparison between LSCN and CR methods for the ill-conditioned channel case ($\delta = \pi/10$).

5. DISCUSSION AND CONCLUSION

For short channels as encountered in digital communications, the new method is relatively simple to implement and performs as well

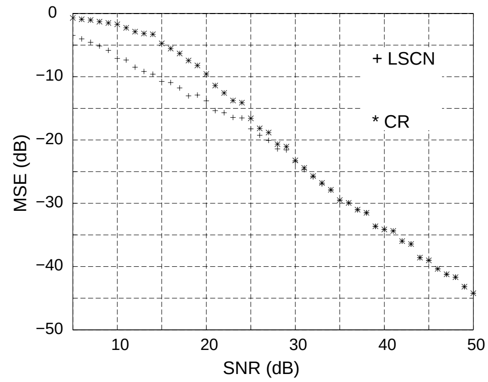

Figure 4: Comparison between LSCN and CR for three random channels.

or better than other recently proposed methods for fast blind identification. In spite of the similarity of the problem formulation with other previous methods (all of which are based on the linear relationships between observations), the error and error minimization criteria proposed are different and lead to more accurate results, especially at low SNR. With the addition of more sensors the LSCN method becomes more robust to noise but at the expense of higher complexity.

6. REFERENCES

[1] Y. Sato, "A method of self-recovering equalization for multilevel amplitude modulation," *IEEE Trans. Commun*, vol. COM-6, pp. 679-682, Jun. 1975.

[2] C. Nikias, "Blind deconvolution using higher-order statistics," in *Proc. 2nd Int. Conf. Higher-Order Stat.* Elsevier, 1992. pp. 49-56.

[3] L. Tong, G. Xu and T. Kailath, "Fast blind equalization via antenna arrays," in *Proc. IEEE ICASSP*, 1993, vol. IV, pp. 272-275.

[4] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind equalization," in *Proc. 27th Asilomar Conf.*, 1993, pp. 751-755.

[5] 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.

[6] E. Moulines, P. Duhamel, J. 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.

[7] W. Qiu and Y. Hua, "Performance comparison of three methods for blind channel identification," in *Proc. IEEE ICASSP*, 1996, pp. 2423-2426.

[8] W. Qiu and Y. Hua, "Performance analysis of the subspace method for blind channel identification," *Signal Processing*, vol. 50, pp. 71-81, 1996.

[9] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Lett.*, vol. 5, pp. 174-176, July 1998.