# ON CHOICE OF IMPULSE RESPONSE LENGTH IN CHANNEL IDENTIFICAITON

*Soosan Beheshti, Munther A. Dahleh*

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
soosan@mit.edu,dahleh@mit.edu

## ABSTRACT

A new approach to the problem of estimating length of a channel impulse response is presented. Unlike the information theoretic approaches , the new method examines the unmodeling effect explicitly and independent of the additive noise effect.

## 1. INTRODUCTION

Information theoretic criteria are commonly employed for estimation of the length of the impulse response of a channel using finite noisy data. The strong assumption in estimating these criteria is that the true impulse response is an element of at least one of the competing model sets. In practical problems the length of the impulse response is not finite. In such cases the information theoretic criteria point to model sets of higher and higher order as the length of the data becomes larger and/or the signal to noise ratio(SNR) grows. It is expected that with higher SNR and longer data, the significant part of the impulse response can be captured by estimates in model sets with much smaller order than the solution given by these criteria [1].

To overcome the shortcomings of the information theoretic criteria, we propose a new approach based on a novel method of quality evaluation of the impulse response estimate in the competing model sets [7]. The method attempts to simply estimate the $l_2$ norm of the impulse response error in each model set. In statistical approaches, the estimate of such criterion obtained by considering the unmodeled part of the impulse response as a part of the additive noise effects. Unlike such approaches, the unmodeling is explicit in the problem formulation for the new approach. Using the information embedded in the output error in each model set, an estimate of the unmodeled effects in the impulse response estimate is provided. The proposed method corresponds to the compelling problem of quantification of the estimation error in system identification [3].

## 2. PROBLEM STATEMENT

Consider a single-input/k-output channel for which the input and output of the sub-channel $j$, at time $n$, are related as follows

$$y_n(j) = \sum_{i=1}^{n} h_i(j) x_{n-i+1} + w_n(j), \tag{1}$$

where $h(j) = [h_1(j), \cdots]^T$ is the impulse response of the $j$th channel, and $w_n(j)$ is the additive white Gaussian noise(AWGN) with zero mean and variance $\sigma_w^2$, independent of the input. The input is independent samples of a binary sequence, $\pm 1$, with Bernoulli distribution. Finite length data, input $X^N = [x_1, \cdots, x_N]$, and outputs $Y^N(j) = [y_1(j), \cdots, y_N(j)]$, $1 \leq j \leq k$, is available. Choose a proper length for the impulse response estimate.

## 3. CHANNEL ORDER ESTIMATION: INFORMATION THEORETIC CRITERIA

The information theoretic criteria for model selection, introduced by Akaike, Schwarz and Rissanen, address the following problem. Given a finite set of observation, in our problem $X^N$ and $Y^N$, and a family of models, which are parameterized by elements of a set $\Theta$, and a family of probability densities $f(Y^N|\Theta, X^N)$, select the model that best fits the data [5].

Akaike information criterion(AIC) is the estimate of the mean of Kulback-Liebler distance of the true density $f(Y^N|\theta, X^N)$, and the estimated density $f(Y^N|\hat{\theta}_m^N, X^N)$ in each model set of order $m$, $S_m$. The AIC estimate is given by

$$\text{AIC}(S_m) = -\frac{1}{N} \log f(Y^N|\hat{\theta}_m^N, X^N) + \frac{m}{N} \tag{2}$$

where $\hat{\theta}_m^N$ is the maximum likelihood estimate of the parameter in $S_m$. The approach suggests to select the model set which minimizes AIC. Based on information

0-7803-7080-5/01/$10.00 

theoretic argument, each model set can be used to encode the observed data. Rissanen proposes to select the model which yields the minimum code length, defined by minimum description length(MDL),

$$\text{MDL}(S_m) = -\frac{1}{N}\log f(Y^N|\hat{\theta}_m^N, X^N) + m\frac{\log N}{2N}. \quad (3)$$

Bayesian information criterion(BIC) is proposed by Schwarz. In this method Shcwarz assumes a prior probability for the competing model sets and suggests to select the model that yields the maximum posterior probability. The result in this approach is similar to MDL criterion in (3).

The information theoretic criteria for the problem in section 2 is computed as follows. Consider a model set $S_m$, a subspace of order $m$ in $R^N$. In this case $\theta$ is the impulse response of the system $h$, and each estimate in subspace $S_m$ is the least-square solution $\hat{\theta}_m^N = \hat{h}_m^N$. For each model set $S_m$, we have

$$\text{AIC}(S_m) = -\log\left(\frac{1}{\sqrt{2\pi}\sigma_w}e^{-\frac{\|Y-\hat{Y}_m^N\|_2^2}{2N\sigma_w^2}}\right) + \frac{m}{N}, \quad (4)$$

$$\text{MDL}(S_m) = -\log\left(\frac{1}{\sqrt{2\pi}\sigma_w}e^{-\frac{\|Y-\hat{Y}_m^N\|_2^2}{2N\sigma_w^2}}\right) + m\frac{\log N}{2N}, \quad (5)$$

where $\|Y - \hat{Y}_m^N\|_2$ is the $l_2$ norm of $Y - \hat{Y}_m^N$, and $\hat{Y}_m^N$ is the output estimate in subspace $S_m$.

## 4. A NEW METHOD FOR CHANNEL ORDER ESTIMATION

Information theoretic methods attempt to "determine" the length of the channel impulse response. In most practical problems, the impulse response does not have a finite length. At the receiver we require to detect the minimum number of taps of the impulse response which represents the "significant part" of the impulse response of the channel. Implementing the information theoretic method in this situation, provides an estimate for the length of the impulse response which is very sensitive to the variation in signal to noise ratio(SNR) [1] and in the length of the data [1]. To overcome such problems, we propose a new method of selecting the length of the impulse response based on the $l_2$ norm of the error in the impulse response estimation.

The input-output relationship for each channel in (1) can be written as

$$Y^N = A_m^N h_m^N + B_m^N \Delta_m^N + w^N, \quad (6)$$

where $h_m^N = [h_1, \cdots, h_m]^T$, $\Delta_m^N = [h_{m+1}, \cdots, h_N]^T$ and $A_m^N$, $B_m^N$ are functions of the input. Consider the subspace $S_m$ of order $m$ in space $R^N$. The projection of $h^N$ on $S_m$ is $h_m^N$. Given the finite length input and output of the system, an estimate of the first $m$ taps of the impulse response of the system in $S_m$, $\hat{h}_m^N$, is obtained by using the conventional least-square method, therefore $\hat{h}_m^N = ((A_m^N)^T A_m^N)^{-1}(A_m^N)^T Y^N$. *Subspace impulse response error*(SIRE) is the square of $l_2$ norm of the distance between $h_m^N$ and $\hat{h}_m^N$ in $S_m$, $\|\hat{h}_m^N - h_m^N\|_2^2$. *Impulse response error*(IRE) is the square of the $l_2$ norm of the distance between $h^N$ and $\hat{h}_m^N$, $\|\hat{h}_m^N - h^N\|_2^2$. In [4] we estimate the variance and expected value of both SIRE and IRE for any subspace $S_m$ when $m << N$.

We apply the output estimate in each subspace $S_m$, $\hat{Y}_N^m = A_m^N \hat{h}_m^N$, and the output error $\frac{1}{N}\|Y - \hat{Y}_m^N\|_2^2$ to estimate lower and upper bound for the IRE [4].

In each subspace $S_m$, with probability greater than $Q(\alpha)Q(\beta)Q(\gamma/p)$, IRE is bounded with $L_{S_m}$ and $U_{S_m}$, which are functions of $u, y, m, N$ and $\alpha, \gamma, \beta$.[2]

$$L_{S_m} \le \|h - h_m^N\|_2^2 \le U_{S_m}. \quad (7)$$

Note that the computed upper and lower bound converge to each other for particular choise of $\alpha, \beta, \gamma$ as $N$ grows. We suggest to select the length of the impulse response, $m^*$, such that $m^* = \arg\min_m U_{S_m}$. In single-input/multi-output case we can compute $m^*(k)$ for each channel separately, or choose one length, $m^*$, for all the sub-channels such that the summation of all the upper bounds for all IREs is minimized. The upper and lower bounds in (7) are as follows [4].

$$L_{S_m} = max\{0, \frac{m}{N}\sigma^2 + (1+\frac{m}{N})\frac{Lg_m}{1+\frac{\gamma}{\sqrt{N}}} - \frac{\beta\sqrt{m}}{N}\sqrt{P_m}\}, \quad (8)$$

$$U_{S_m} = \frac{m}{N}\sigma^2 + (1+\frac{m}{N})\frac{Ug_m}{1-\frac{\gamma}{\sqrt{N}}} + \frac{\beta\sqrt{m}}{N}\sqrt{P_m}. \quad (9)$$

where $P_m = 2(\sigma^2)^2 + \frac{mUg_m}{1-\frac{\gamma}{\sqrt{N}}}$, and

$$Ug_m = max\{0, \frac{2\alpha^2\sigma_w^2}{N} + X_m + \frac{2\alpha\sigma_w}{\sqrt{N}}\sqrt{Q_m}\}, \quad (10)$$

$$Lg_m = max\{0, \frac{2\alpha^2\sigma_w^2}{N} + X_m - \frac{2\alpha\sigma_w}{\sqrt{N}}\sqrt{Q_m}\}, \quad (11)$$

where $Q_m = \frac{\alpha^2\sigma_w^2}{N} + X_m + \frac{1}{2}\sigma_w^2$ and $X_m = \frac{1}{N}\|Y - \hat{Y}_m^N\|_2^2 - \sigma_w^2(1-\frac{m}{N})$. If we choose $\alpha$, $\beta$ and $\gamma$ to be functions of $N$ such that,

$$\lim_{N\to\infty}\alpha_N = \infty, \lim_{N\to\infty}\beta_N = \infty, \lim_{N\to\infty}\gamma_N = \infty, \quad (12)$$

[1] $SNR = 10\log_{10}\frac{\|Y^N\|_2^2}{N\sigma_w^2}$

[2] $p = max_j([h_i, \cdots, h_j]_{H_\infty}) + m$, $1 \le j \le N$, where $x_{H_\infty}$, the $H_\infty$ norm of a vector $x$, is $\max_v \|\text{conv}(v,x)\|_2$ where $v$ is a vector with unit $l_2$ norm and $\text{con}(v,x)$ is the convolution of $v$ and $x$.

then the probability $Q(\alpha)Q(\beta)Q(\gamma/p)$ goes to one as $N$ goes to infinity. The tightness of the upper and lower bound depends on the choice of $\alpha, \beta$ and $\gamma$. If

$$\lim_{N\to\infty} \frac{\alpha_N}{\sqrt{N}} = 0, \lim_{N\to\infty} \frac{\beta_N}{N} = 0, \lim_{N\to\infty} \frac{\gamma_N}{\sqrt{N}} = 0, \quad (13)$$

then both $Ug_m$ and $Lg_m$ in (10), (11) converge to $max\{0, \frac{1}{N}||Y - \hat{Y}_1^m||_2^2 - \sigma_w^2\}$ as $N$ grows [7]. Therefore

$$||\hat{h}_m^N - h^N||_2^2 \to 0 \quad (14)$$

$$||\hat{h}_m^N - h^N||_2^2 \to max\{0, \frac{1}{N}||Y - \hat{Y}_m^N||_2^2 - \sigma_w^2\}. \quad (15)$$

Note that we can consider subspaces of form $S_{(m_1,m_2)}$, which are of order $m_1 - m_2 + 1$ and are model sets representing the impulse response taps of the channel from tap $m_1$ to tap $m_2$. Implementing the proposed criterion for such model sets enables us to estimate the delay of the system. The expansion of the new approach to the multi-input/multi-output case is straightforward [7].

## 5. COMPARISON OF THE NEW CRITERION WITH INFORMATION THEORETIC CRITERIA

As we mentioned in previous sections, all the discussed approaches, use the least-square estimate of the true impulse response in each subspace $S_m$. Therefore, for each subspace $S_m$, the relation between SIRE and IRE is as follows

$$||h - \hat{h}_m^N||_2^2 = ||h_m^N - \hat{h}_m^N||_2^2 + ||\Delta_m^N||^2, \quad (16)$$

where $\Delta_m^N = [h_{m+1}, \cdots, h_N]^T$ is the unmodeled part. As $N$ goes to infinity, the SIRE converges to zero and the IRE converges to $\lim ||\Delta_m^N||_2^2 = ||\Delta_m^\infty||_2^2$. If the length of the impulse response is finite, $M$, then for any subspace $S_m$, $m \geq M$, SIRE and IRE are the same. The estimate of the information theoretic criteria, expressions in (4) and (5), are obtained with the strong assumption that IRE goes to zero as $N$ grows. Such assumption is valid only if $||\Delta_m^N||^2$ is small enough to be ignored. If the true impulse response, $h$, is an element of $S_m$, AIC and MDL are expressions in (4) and (5). Otherwise an estimate for AIC and MDL are not computable. In order to use (4) and (5) as AIC and MDL for when $h$ is not an element of the model set, the effect of the unmodeled part of the impulse response, $||\Delta_m^N||_2^2$, is treated as a part of the additive white noise effects. With this assumption the expression in (4) and (5) is estimated using only the statistics of the input [5]. With such inaccurate assumption the variance of noise varies for different subspaces. The obtained criteria is widely used for order estimation in blind channel identification. The new approach, however, applies the output error, to separate the effects of the noise versus the effects of the unmodeled part in estimation of IRE. For that reason, we require a training signal.

**The Additive Noise Variance:** In practical problems the variance of the additive noise is usually unknown. In the information theoretic approach, the parameter in each subspace $S_{m+1}$ is $\theta_{m+1} = (h_m, \sigma_m^2)$. Here $\sigma_m^2$ is the variance of the noise, to be estimated for each model set, $S_m$, separately. For each subspace, $\hat{\theta}_m^N$ is the maximum likelihood estimate of the parameter. First $\hat{h}_m^N$ is computed by minimizing the output error and then the variance is estimated by $\hat{\sigma}_m^2 = \frac{1}{N}||Y - \hat{Y}_m^N||^2$. The AIC and MDL in this case are given by [2].

$$\text{AIC} \approx (1 + \frac{2m}{N})\frac{||Y - \hat{Y}_m^N||^2}{N}, \quad (17)$$

$$\text{MDL} \approx (1 + m\frac{\log N}{N})\frac{||Y - \hat{Y}_m^N||^2}{N}. \quad (18)$$

In the proposed approach, we first estimate the variance of the noise which is the same for all subspaces. We suggest using the estimate of the variance obtained for the model set with highest order in the estimation, i.e. $M_{max}$. Therefore, the estimated variance for all model sets is $\hat{\sigma}_w^2 = \frac{1}{N}||Y - \hat{Y}_{M_{max}}^N||_N^2$.

### 5.1. Consistency Issues

When the length of the impulse response, M, is finite the information theoretic methods of order estimation focus on estimating $M$. One method to compare these different criteria is to check whether they are consistent, i.e., as $N \to \infty$, the selected order, $\hat{m}$, approaches $M$ with probability one. Consequently, we can consider the question of finding $f(N)$ for which the following criterion

$$\frac{||Y - \hat{Y}_m^N||^2}{N} + m\sigma_w^2 f(N), \quad (19)$$

is consistent. It has been shown that AIC, for which $f(N) = \frac{2}{N}$, is not consistent. Whereas, in MDL and BIC, $f(N) = \frac{\log(N)}{N}$ makes the criterion consistent. Hannan suggests another consistent method by using $f(N) = \frac{\log\log N}{N}$, which decreases faster than $f(N) = \frac{\log(N)}{N}$ in MDL as $N$ grows [6].

In the new approach, the choice of $\alpha, \beta, \gamma$ play a major role for checking the consistency. If they satisfy conditions in (12), (13), then the consistency of the method is guaranteed. For example, if $\alpha = \beta = \gamma = 0$ in (8), (9), and $0 \leq (1+\frac{m}{N})(\frac{1}{N}||Y - \hat{Y}_m^N||_2^2 - (1 - \frac{m}{N}\sigma_w^2)$, then $E(||\hat{h}_m^N - h^N||^2) \approx -\sigma_w^2 + \frac{m}{N}\sigma_w^2 + (1 + \frac{m}{N})\frac{1}{N}||Y -$

$\hat{Y}_m^N||^2$. This expression resembles AIC when $m \geq M$, since in this case $\frac{m}{N}(\frac{1}{N}||Y - \hat{Y}_m^N||_2^2 \approx \frac{m}{N}\sigma_w^2$. On the other hand by choosing $\alpha^2 = \beta = \gamma^2 = \log\log N$, both conditions of (12), (13) are satisfied. Note that in this case the provided upperbound in 7 is similar to the criterion proposed by Hannan.

When the length of the channel impulse response is not finite, the consistent methods point to a higher and higher length for the impulse response estimate as $N$ grows. With the new proposed method we can avoid such problem by using a threshold for the impulse response error. If a threshold $\epsilon$ is assumed for the minimum acceptable IRE, then we choose the smallest $m$ for which $U_{S_m} \leq \epsilon$.

## 6. SIMULATION RESULTS

We use the microwave radio channel, *chan*10.*mat*, which is found at http://spib.rice.edu/spib/microwave.html. Figure (1) shows the real part of the first 60 taps of the

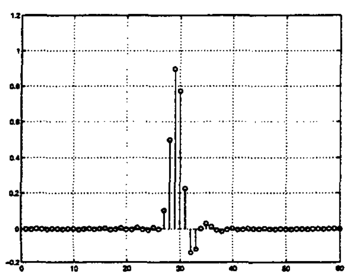

Figure 1: *'*': The real part of the first 60 taps of a microwave radio channel impulse response.*

impulse response. The simulation result for $N = 300$ and SNR=30db is as follows: $\hat{m}$(AIC)=41, $\hat{m}$(MDL)=38. The new proposed criterion selects $m^* = 41$. Figure(2) shows the upper and lower bound on IRE for $N$=800, SNR=90db. The bounds on the error are from (8), (9) with $\alpha = \sqrt{\beta} = \gamma = \log\log N$. The solid line is the estimate of expected value of I.R.E, $E(||\hat{h}_m^N - h^N)||^2)$, when $\alpha = \beta = \gamma = 0$. In this case all the methods select an impulse response length which is larger than 130. With higher SNR and/or longer data sample, all the methods choose a larger and larger length for the impulse response estimate. However, if we choose a threshold for the acceptable IRE to be $10^{-3}$, the new criterion selects $m^* = 35$. With this threshold $m^* \leq 35$ when SNR grows and/or the length of data gets larger. Counting for the delay of the system, with the same threshold, the proposed method chooses the 10 taps of the impulse response estimate from 27 to 36 for modeling the channel.

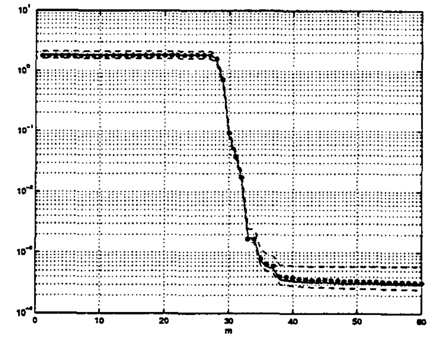

Figure 2: *'*' line: Impulse response error, $||\hat{h}_m^N - h||_2^2$, for SNR=90db, N=800. '-.': Upperbound and lowerbound of IRE. Solid line: Estimate of $E||\hat{h}_m^N - h||_2^2$.*

## 7. CONCLUSION

We presented a new approach to the estimation of the length of a channel impulse response. We elaborated on the advantages of implementing this method over the available information theoretic solutions.

## 8. REFERENCES


[1] A.P. Liavas, P.A. Regalia, and J. Delmas. "Blind channel approximation: effective channel order estimation," *IEEE Trans. on Signal Processing*, vol.47, pp.3336-3344, 1999.

[2] L. Ljung. *System Identification: Theory for the User*. NJ: Prentice-Hall, 1998.

[3] B. Ninness and G.C. Goodwin. "Estimation of model quality," *Automatica*, vol31, pp.1771-1797, 1995.

[4] S. Beheshti and M.A. Dahleh. "On model quality evaluation of stable LIT systems," to appear in *39th IEEE Conference on Decision and Control*, Sydney, December 2000.

[5] M. Wax and T. Kailath. "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing,* vol.ASSP-33,pp.387-392,Apr 1985.

[6] E. Hannan. "The estimation of the order of an ARMA process," *The Annals of Statistics,* vol.8 No.5, pp.1071-1081, 1980.

[7] S. Beheshti. *On Quality Evaluation of Linear Systems* . LIDS Technical Report-P-2483, MIT, September 2000.