Proceedings of the 39th IEEE
Conference on Decision and Control
Sydney, Australia • December, 2000

On Model Quality Evaluation of Stable LTI Systems

Soosan Beheshti, Munther A. Dahleh
MIT
soosan@mit.edu,dahleh@lids.mit.edu

Abstract

The problem of quantifying the error in estimation of low-complexity models for stable linear time-invariant(LTI) systems is investigated. We elaborate on the advantages of implementing a new method for order selection of the model class.

1 Introduction

Robust control synthesis requires not only a nominal model, but also bounds on the accuracy (or quality) of the model. This has motivated a number of researchers in control community to come up with a proper quantified error model. The existing methods, such as conventional or set-membership identification methods, have their own strengths and drawbacks in quantifying the impulse response error of LTI systems [4]. In this paper we present a method which considers a similar prior assumption about the additive noise that is used in conventional system identification methods [3]. However, unlike the conventional method, the new method distinguishes between the effects of noise and the unmodeled portion of the impulse response in quantifying the impulse response error.

A fundamental question of system identification is how to choose a finite order subspace and search for the "closest" element of this subspace to the true impulse response of the system using finite noisy data [10]. When the only source of noise is an additive output noise and the measure of closeness is in the form of l_2 norm, the least-square algorithm can be used to estimate the model [3]. The objective of this paper is to quantify the error between such an estimate of the impulse response and the true impulse response and to compare the estimates in subspaces of different orders. As the length of the data grows in each subspace, the estimate approaches a fixed point in the subspace. Here we consider two errors: *impulse response error* defined as the distance between the true system and the estimate in the subspace, and *subspace impulse response error*, the distance of the estimate to the projection of the true system (the closest element of the subspace to the true system).

Venkatesh [8] designs an input that guarantees convergence of the *subspace impulse response error* to zero as the length of the data grows. He uses a bound on the norm of unmodeled dynamics to derive an upper bound for the rate of convergence of *subspace impulse response error*. In conventional system identification the trade off between bias and variance error represents the *subspace impulse response error*. Our goal is to use the data to find a tight upper bound on the norm of the unmodeled dynamics for each subspace of different dimensions, i.e., to find bounds on both *impulse error* and *subspace impulse response error*.

The comparison of the impulse errors of subspaces with different dimensions can be used as a tool in the order selection problem. We discuss the conventional methods of order estimation, AIC [1], MDL [6], BIC [7], and mention the benefits of implementing the proposed method in the process of order estimation.

2 Problem Statement

We consider causal, single-input/single-output, linear time-invariant, discrete-time systems with bounded power/power gain (bounded H_∞ norm). Input and output of the system are related as follows

$$y_n = \sum_{i=1}^{n} h_i u_{n-i+1} + w_n, \qquad (1)$$

where $h = [h_1, \cdots]^T$ is the impulse response of the system and $w = [w_1, \cdots]^T$ is the additive white, zero-mean random vector. Each w_i has variance σ_w^2, and is independent of the input. The input is assumed to be a quasi-stationary signal [3].

Finite length data, input $[u_1, \cdots, u_N]$, and output $[y_1, \cdots, y_N]$, is available and $u_i = 0$ for $i \leq 0$. There is no assumption on the length of the impulse response. However, only the first N elements of h, h^N, relate the N points of the input and output.

Consider 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 for h_m^N in S_m, $\hat{h}_m^N$, is obtained using the conventional least-square method (ML estimation). *Subspace impulse response error*(S.I.R.E) is the l_2 norm of the distance between h_m^N and $\hat{h}_m^N$ in S_m. *Impulse response error*(I.R.E) is the l_2 norm of the distance between h^N and $\hat{h}_m^N$. Our objective is to estimate the variance and expected value of both S.I.R.E and I.R.E for any subspace S_m when $m << N$.

Model Quality Evaluation The method we propose can be used for quality evaluation of the estimate of impulse response in subspace S_m. For each subspace S_m there exist functions of α, u, y, m, and N, $L^1_{S_m}(\alpha, u, y, m, N)$, $L^2_{S_m}(\alpha, u, y, m, N)$, $U^1_{S_m}(\alpha, u, y, m, N)$, and

0-7803-6638-7/00$10.00 © 2000 IEEE

$U^2_{S_m}(\alpha, u, y, m, N)$, such that with probability greater than $Q^2(\alpha)$, $(Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-\frac{u^2}{2}} du)$

$$L^1_{S_m} \leq ||h^N_m - \hat{h}^N_m||^2 \leq U^1_{S_m}, L^2_{S_m} \leq ||h^N - \hat{h}^N_m||^2 \leq U^2_{S_m}.$$

To estimate the order of the system we suggest to pick m^* such that

$$m^* = \arg\min_m U^2_{S_m}(\alpha, u, y, m, N). \quad (2)$$

We show that α can be a function of N such that $\lim_{N\to\infty} \alpha_N = \infty$, so as N grows, $Q(\alpha_N)$ goes to one.

3 Impulse Response Error

The following method can be used for any subspace of R^N, S_m. However, for simplicity of presentation let S_m be a subspace which includes the first m taps of the impulse response. Form (1) the input-output relationship for the finite data is as follows

$$\begin{aligned} Y^N &= U(N)h^N + w^N \\ &= \begin{bmatrix} A_m(N) & B_m(N) \end{bmatrix} \begin{bmatrix} h^N_m \\ \Delta^N_m \end{bmatrix} + w^N \quad (3) \end{aligned}$$

where $Y^N = [y_1, \cdots, y_N]^T$, $U(N)$ is $N \times N$ a Toeplitz matrix generated by the input, $h^N = [h^N_m; \Delta^N_m]$, $h^N_m = [h_1, \cdots, h_m]^T$, $\Delta^N_m = [h_{m+1}, \cdots, h_N]^T$, $A_m(N)$ is a $N \times m$ matrix, $B_m(N)$ is a $N \times N - m$ matrix and w^N is the white additive noise. The least-square method is used to find the estimate of the first m taps of the impulse response h^N_m

$$\hat{h}^N_m = ((A_m(N))^T A_m(N))^{-1}((A_m(N))^T Y^N_1). \quad (4)$$

(From here we drop N from w^N) Then the S.I.R.E and the I.R.E are

$$||\hat{h}^N_m - h^N_m||^2_2 = \underbrace{(B_m\Delta^N_m)^T C_m B_m \Delta^N_m}_{unmodeled} + \underbrace{w^T C_m w}_{noise} + \underbrace{2w^T C_m B_m \Delta^N_m}_{cross\ term} \quad (5)$$

$$||\hat{h}^N_m - h^N||^2_2 = ||\hat{h}^N_m - h^N_m||^2_2 + ||\Delta^N_m||^2. \quad (6)$$

where $C_m = A_m(A^T_m A_m)^{-1}(A^T_m A_m)^{-1}A^T_m$ (From here we drop N from $A_m(N)$, $B_m(N)$).

We use the stochastic properties of the noise to provide soft bounds for the noise related parts of the S.I.R.E and I.R.E. Then we use the output error to estimate the components of S.I.R.E and I.R.E which are caused by the unmodeled part, Δ^N_m.

Asymptotic Behavior As N goes to infinity, the terms which are noise dependent asymptotically go to zero ($\lim_{N\to\infty} 2w^T C_m B_m \Delta^N_m = 0$, $\lim_{N\to\infty} w^T C_m w = 0$) [11],

$$\lim_{N\to\infty} ||\hat{h}^N_m - h^N_m||^2_2 = (\Delta^\infty_m)^T(\lim_{N\to\infty} B^T_m C_m B_m)\Delta^\infty_m \quad (7)$$

$$\lim_{N\to\infty} ||\hat{h}^N_m - h||^2_2 = \lim_{N\to\infty} ||\hat{h}^N_m - h^N_m||^2_2 + ||\Delta^\infty_m||^2. \quad (8)$$

The second component of the impulse response error in (8) is the unmodeled dynamics of the system. The first component, however, is a function of the input and the unmodeled dynamic. Since the input is quasi-stationary $\lim \frac{1}{N} A^T_m B_m$ and $\lim \frac{1}{N} A^T_m A_m$ exist, $B^T_m C_m B_m$ has a limit. If the input is such that $\lim \frac{1}{N} A^T_m B_m$ goes to zero, then this term vanishes as N goes to infinity, which implies that the subspace impulse error(7) asymptotically goes to zero. If $\lim \frac{1}{N} A^T_m B_m$ does not go to zero asymptotically, there is a fixed bias in the subspace impulse response error (7) as N goes to infinity.

3.1 Output Error

The output error is the l_2 norm of the distance between the output and the estimated output obtained by the estimated impulse response in S_m

$$\frac{1}{N}||Y - \hat{Y}^m_1||^2_2 = \underbrace{\frac{1}{N}(B_m\Delta^N_m)^T G_m B_m \Delta^N_m}_{unmodeled} + \underbrace{\frac{1}{N} w^T G_m w}_{noise} + \underbrace{\frac{2}{N} w^T G_m B_m \Delta^N_m}_{cross\ term} \quad (9)$$

where $G_m = (I - A_m(A^T_m A_m)^{-1}A^T_m)$.

In absence of an additive noise, the S.I.R.E, the I.R.E (5),(6) and the output error are decreasing functions of m. Assume that there exists M such that $h_M \neq 0$, $h_i = 0$, $i > M$. If $M < N$, then all errors are none zero for $m < M$ and zero for $m \geq M$. If $M \geq N$, then all errors are decreasing functions of m. In this case using the output error, which is available, is enough for "comparing" the distance of the impulse response of the system from the subspace estimates of different order. Therefore to find the model set with minimum order m^* which minimizes the impulse error, we can use the output error. If the output error is non-zero for all m, then $m^* = N$, otherwise, the smallest m for which output error is zero is $m^* = M$.

In presence of an additive noise, the output error is a decreasing function of m. However, regardless of the amount of M, which can be less than N or greater than N, the impulse response error is minimized at some point m^*. And m^*, which is less than or equal to M, might be less than or equal to N. In the next section we elaborate the relationship between the output error and the impulse response error in presence of the additive noise. There exist methods which use the output error and add an extra term, which is a function of m, N and σ_w, to select the optimum order for the system such as AIC, MDL, BIC. We will compare our method to these conventional methods later.

Note that in off-line identification the expected value and variance of the subspace estimate, $\hat{h}^N_m$, can be calculated approximately by repeating the experiment, using the same input(similar to Monte Carlo simulations). Although, in this case, we obtain estimates for the bias and variance error of the subspace estimate,

this method can not provide any estimate of I.R.E [5]. Our goal is to estimate both S.I.R.E and I.R.E.

4 Bounds on The Impulse Response Error

In this section we find the estimate of variance and expected value of S.I.R.E and I.R.E in presence of an additive white noise. The additive white noise is such that

$$E(w_i w_j) = \sigma_w^2, \text{ if } i = j \quad E(w_i w_j) = 0,\ i \neq j \quad (10)$$

and $E(w_i^4)$ is bounded. From (5)and (6) we have

$$E(||\hat{h}_m^N - h_m^N||^2) = tr(C_m)\sigma_w^2 + m_c \quad (11)$$
$$E(||\hat{h}_m^N - h^N)||^2) = E(||\hat{h}_m^N - h_m^N)||^2) + ||\Delta_m^N||^2 \quad (12)$$

where $m_c = (B_m\Delta_m^N)^T C_m B_m \Delta_m^N$ and the variance of both errors is

$$\text{var}(||\hat{h}_m^N - h^N)||^2) = \text{var}(||\hat{h}_m^N - h_m^N)||^2) = \quad (13)$$
$$\text{var}(w^T C_m w) + 4(B_m\Delta_m^N)^T C_m^2 B_m \Delta_m^N \sigma_w^2$$

(Note that $E(w^T C_m w)(2w^T C_m B_m \Delta_m^N) = 0$). We can calculate the two noise related elements of both expected value and variance of the S.I.R.E and the I.R.E, $tr(C_m)\sigma_w^2$ and $\text{var}(w^T C_m w)$, using only the input and statistics of the noise.

Next we want to estimate the unmodeled related components of the expected value and variance of impulse response errors in (11), (12), and (13), $(\Delta_m^N)^T B_m^T C_m B_m \Delta_m^N$, $(\Delta_m^N)^T B_m^T C_m^2 B_m \Delta_m^N$, and $||\Delta_m^N||^2$. In the following three steps we provide probabilistic bounds on the S.I.R.E and the I.R.E using the output error.

4.1 Step One: Using The Output Error

The output error (9) is a random variable for which we have

$$E(\frac{1}{N}||Y - \hat{Y}_1^m||_2^2) = \frac{1}{N}tr(G_m)\sigma_w^2 + gg_m \quad (14)$$

where

$$gg_m = \frac{1}{N}(B_m\Delta_m^N)^T G_m B_m \Delta_m^N. \quad (15)$$

The variance of the output error is

$$\text{var}(\frac{1}{N}||Y - \hat{Y}_1^m||_2^2) = \text{var}(\frac{1}{N}w^T G_m w) + \frac{4\sigma_w^2}{N} gg_m \quad (16)$$

(Note that since G_m is a projection matrix $G_m^2 = G_m$). This random variable, $\frac{1}{N}||Y - \hat{Y}_1^m||_2^2$, is almost a Gaussian random variable when N is large (using Central Limit Theorem). By using the observed output error we can find bounds on the unmodeled part of the output error, gg_m.

If a Gaussian random variable X_m has mean m_X and variance σ_X^2 then

$$Prob(m_X - \alpha\sigma_X < X_m < m_X + \alpha\sigma_X) = Q(\alpha) \quad (17)$$

where $Q(\alpha) = \int_{-\alpha}^{\alpha} \frac{1}{2\pi} e^{-x^2/2} dx$.

Using the input and the output, we can calculate $X_m = \frac{1}{N}||Y - \hat{Y}_1^m||_2^2 - \frac{1}{N}tr(G_m)\sigma_w^2$. Here X_m is a random variable with $E(X_m) = gg_m$. Variance of X_m is the same as variance of $\frac{1}{N}||Y - \hat{Y}_1^m||_2^2$ in (16). Therefore with probability $Q(\alpha)$ for X_m we have

$$gg_m - \alpha\sqrt{Z_m} < X_m < gg_m + \alpha\sqrt{Z_m} \quad (18)$$

where $Z_m = 4\sigma_w^2 \frac{gg_m}{N} + v_m$ and $v_m = \text{var}(\frac{1}{N}w^T G_m w)$. This gives an upper and lower bound for gg_m, $Lg_m \leq gg_m \leq Ug_m$ where

$$Ug_m = max\{0, \frac{\alpha^2 a^2}{2} + X_m + \alpha a\sqrt{J_m}\} \quad (19)$$

$$Lg_m = max\{0, \frac{\alpha^2 a^2}{2} + X_m - \alpha a\sqrt{J_m}\} \quad (20)$$

and $J_m = \frac{\alpha^2 a^2}{4} + \frac{v_m}{a^2} + X_m$, $a = \frac{2\sigma_w}{\sqrt{N}}$.

4.2 Step Two: Bounds on The Unmodeled Parts of S.I.R.E and I.R.E

To find the expected value and variance of S.I.R.E and I.R.E use the upper and lower bounds on gg_m in (15) which are calculated in previous step and find upper and lower bounds for terms which are caused by unmodeled dynamics, $m_c = (\Delta_m^N)^T B_m^T C_m B_m \Delta_m^N$, $v_c = 4\sigma_w^2 (B_m\Delta_m^N)^T C_m^2 B_m \Delta_m^N$, $||\Delta_m^N||^2$in (11),(12),(13). Therefore with probability $Q(\alpha)$

$$\begin{aligned} Lm_c &\leq m_c \leq Um_c,\ Lv_c \leq v_c \leq Uv_c, \\ Lm_\Delta &\leq ||\Delta_m^N||^2 \leq Um_\Delta \end{aligned} \quad (21)$$

Note that this step is a deterministic procedure which solves a constrained optimization problem.

4.3 Step Three: Upper and Lower Bounds for S.I.R.E and I.R.E

Here we use the results of the previous steps and conclude with the following theorem [11] :

Theorem 1 For $z_1 = ||\hat{h}_m^N - h_m^N||^2 - tr(C_m)\sigma_w^2$ with probability larger than $Q(\alpha)Q(\beta)$ we have

$$max\{-tr(C_m)\sigma_w^2, Lm_c - \beta\sqrt{Uv_c + \text{var}(w^T C_m w)}\}$$
$$\leq z_1 \leq Um_c + \beta\sqrt{Uv_c + \text{var}(w^T C_m w)} \quad (22)$$

For $z_2 = ||\hat{h}_m^N - h^N||^2 - tr(C_m)\sigma_w^2$ we can use the same boundary by replacing Lm_c and Um_c in (22) with $Lm_c + Lm_\Delta$ and $Um_c + Um_\Delta$ respectively.

Asymptotic Behavior In [11] we show that the noise related part of the output error, $v_m = \frac{1}{N}w^T G_m w$ in (9), is such that $v_m \leq \frac{l}{N}$ where l is a fixed finite number. Therefore $\alpha\sqrt{v_m} \leq \frac{\alpha\sqrt{l}}{\sqrt{N}}$ and by choosing α as a function of N, α_N, such that $\lim \frac{\alpha_N}{\sqrt{N}} = 0$, the upper bound and lower bounds of gg_m in (19) and (20) approach each other as N goes to infinity. In order for these upper and lower bound to hold with probability

which goes to one as N grows, we can choose α_N such that $\lim_{N\to\infty}\alpha_N = \infty$, therefore $Q(\alpha_N) \to 1$. One candidate for α_N for example is $log(N)$.
While as N grows, with proper choice of α_N, the lower bound and upper bound of gg_m (15) in (20),(19) approach each other, the upper and lower bounds of unmodeled related terms of S.I.R.E and I.R.E in (21) might not approach each other as N grows. Note that since the input is quasi-stationary the limits exist and are finite numbers. The asymptotic behavior of these bounds depends on the structure of the input. For example if the input is independent identically distributed random variable the upper bound and lower bounds, bounds of the unmodeled part of S.I.R.E and I.R.E in second step, also converge to each other as N grows.
To guarantee that the bounds we find in step three are also valid with probability which goes to one as N grows, we pick β in (22) as a function of N such that $lim_{N\to\infty}\beta_N = \infty$. In order to have finite values for the upper and lower bounds of z_1 and z_2 the term $\beta\sqrt{Uv_c + \text{var}(w^T C_m w)}$ in (22) has to be finite for all N. In [11] we show that $Uv_c + \text{var}(w^T C_m w) \leq \frac{l_2}{N}$, where l_2 is a finite number. Therefore as long as the rate of growth of β_N is such that $\lim_{N\to\infty}\frac{\beta_N}{\sqrt{N}} = 0$, the upper bound and lower bounds in (22) are only functions of bounds of the mean m_c and unmodeled dynamic $||\Delta_m^N||^2$ in (21). Note that in section 2, the bounds for the model quality evaluation is given by probability $Q^2(\alpha)$, there we have considered a special case $\alpha = \beta$.

5 Deterministic Noise

Set description for the additive noise and worse case gains are the tools which enable the robust control designer to integrate both disturbances and other descriptions of uncertainty in system models. Most works in this field use a very conservative description for additive noise, i.e., bounded power signals, signals with bounded L_2 or L_∞ norm. However, in practical problems the additive noise has more properties, i.e., it is uncorrelated with the input and therefore belongs to a subset of these sets. One method of presenting such noise is to restrict the set with additional constraints on the correlation of the noise with the input or with itself. Paganini introduces such set descriptions in [5]. Also [8] defines the noise with a set description which has some correlation constraints.
To make a bridge between set description and stochastic additive noise, similar to the idea presented in [5], we can check the richness of a set with the probability that a stochastic noise is a member of such set. For example let's assume that the additive noise in (1) belongs to the following set, W,

$$W = \{v| \ |\sum_{i=1}^{N}\frac{v_i a_i}{N}| \leq \frac{\alpha}{\sqrt{N}}\sqrt{R_o^v(0)}\sqrt{R_o^a(0)}, \quad (23)$$

$$|R_a^v(\tau)| \leq \frac{\alpha}{\sqrt{N}}R_o^v(0)\sqrt{R_o^a(0)}, 1 \leq \tau \leq N$$

$$|R_a^v(0) - R_o^v(0)\sum\frac{a_i}{N}| \leq \frac{\alpha\sqrt{R_o^a(0)(R_o^{v^2}(0) - (R_o^v(0))^2)}}{\sqrt{N}}\}$$

where a is a bounded power sequence and $R_a^v(\tau) = \frac{1}{N}\sum_{i=1}^{N} v_i v_{i+\tau} a_i$, $R_o^v(\tau) = \frac{1}{N}\sum_{i=1}^{N} v_i v_{i+\tau}$. Then the method presented in previous section provides the $\sup_{w\in W}$ and $\inf_{w\in W}$ of $||h_m^N - \hat{h}_m^N||^2$ and $||h^N - \hat{h}_m^N||^2$. As N grows, an additive white Gaussian noise is a member of such set with probability $[Q(\alpha)]^{N+2}$. As it is mentioned in previous sections α can be a function of N such that α_N goes to infinity as N grows. Therefore AWGN is a member of such set asymptotically.

6 Independent Identically Distributed Input

Here we assume that the additive noise is white Gaussian noise(AWGN) and the input is a sequence of independent identically distributed(I.I.D) random variables with unit variance and zero mean. An Example of such input is a Bernoulli sequence of ± 1 which is commonly used in communication. We can use the properties of this type of input to show the asymptotic behavior of the procedure introduced in section 4. In [11] we prove the following theorem.
Theorem 2 If the input of the linear system in (3) is I.I.D, then the upper bound and lower bound for subspace impulse response error(5) and impulse response error(6) with probability greater than $Q(\alpha)Q(\beta)Q(\gamma/k)$ $(k = max_j([h_1, \cdots, h_j]_{H_\infty}) + m, 1 \leq j \leq N$, x_{H_∞} is the H_∞ norm of vector x) are

$$max\{0, \frac{m}{N}\sigma^2 + \frac{m}{N}\frac{Lg_m}{1+\frac{\gamma}{\sqrt{N}}} - \frac{\beta\sqrt{m}}{N}\sqrt{P_m}\} \leq$$

$$||\hat{h}_m^N - h_m^N||_2^2 \leq \frac{m}{N}\sigma^2 + \frac{m}{N}\frac{Ug_m}{1-\frac{\gamma}{\sqrt{N}}} + \frac{\beta\sqrt{m}}{N}\sqrt{P_m} \quad (24)$$

where $P_m = 2(\sigma^2)^2 + \frac{mUg_m}{1-\frac{\gamma}{\sqrt{N}}}$, Lg_m and Ug_m are lower and upper bounds on the unmodeled part of the output error in (20),(19) $Ug_m = max\{0, \frac{2\alpha^2\sigma_w^2}{N} + X_m + \frac{2\alpha\sigma_w}{\sqrt{N}}\sqrt{Q_m}\}$, $Lg_m = max\{0, \frac{2\alpha^2\sigma_w^2}{N} + X_m - \frac{2\alpha\sigma_w}{\sqrt{N}}\sqrt{Q_m}\}$, where $Q_m = \frac{\alpha^2\sigma_w^2}{N} + X_m + \frac{1}{2}\sigma_w^2$ and $X_m = \frac{1}{N}||Y - \hat{Y}_1^m||_2^2 - \sigma_w^2(1 - \frac{m}{N})$. The lower bound for impulse error $||\hat{h}_m^N - h^N||_2^2$ is $max\{0, \frac{m}{N}\sigma^2 + (1+\frac{m}{N})\frac{Lg_m}{1+\frac{\gamma}{\sqrt{N}}} - \frac{\beta\sqrt{m}}{N}\sqrt{P_m}\}$ and the upper bound is

$$\frac{m}{N}\sigma^2 + (1+\frac{m}{N})\frac{Ug_m}{1-\frac{\gamma}{\sqrt{N}}} + \frac{\beta\sqrt{m}}{N}\sqrt{P_m}. \quad (25)$$

Rate of Convergence If we choose α, β and γ as functions of N such that $\lim\alpha_N = \infty$, $\lim\beta_N = \infty$ and $\lim\gamma_N = \infty$, then the probability $Q(\alpha)Q(\beta)Q(\gamma/k)$ in theorem 2 goes to one as N goes to infinity. Also if $\lim\frac{\alpha_N}{\sqrt{N}} = 0$, $\lim\frac{\beta_N}{N} = 0$, and $\lim\frac{\gamma_N}{\sqrt{N}} = 0$, then Ug_m and Lg_m both converge to $max\{0, \frac{1}{N}||Y - \hat{Y}_1^m||_2^2 - \sigma_w^2\}$, therefore

$$||\hat{h}_m^N - h_m^N||_2^2 \to 0 \quad (26)$$

$$||\hat{h}_m^N - h^N||_2^2 \to max\{0, \frac{1}{N}||Y - \hat{Y}_1^m||_2^2 - \sigma_w^2\}. \quad (27)$$

The bounds of errors in (24),(25), provide tight estimates for the rate of convergence of the errors to their limits.

7 Results and Simulations

The impulse response in our experiment is $h(n) = .3(.5)^{n-1} + 3(n-1)(.8)^{n-1}$. The input is an I.I.D Bernoulli sequence of ± 1 and the noise is an additive white Gaussian. Figure (1) shows the result of simula-

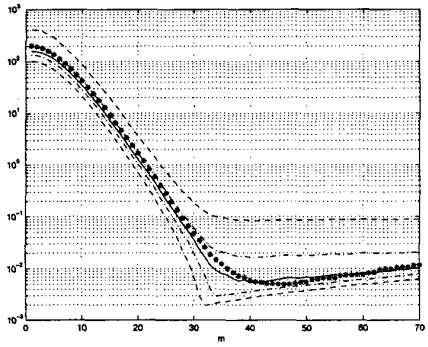

Figure 1: *'*': Impulse response error, $||\hat{h}_1^m - h^N||_2^2$, for $\sigma_w = .2$ and N=400. '-': Bounds calculated for $\alpha^2 = \gamma^2 = \beta = log(N)$, '-.':Bounds calculated for $\alpha^2 = \gamma^2 = \beta = 2$. Solid line: the estimate of $E||\hat{h}_m^N - h^N||_2^2$ (29).*

tion for I.R.E. The bounds on the error are calculated base on upper and lower bounds given in (24), (25). The solid line in both figures is the estimate of expected value of S.I.R.E [11]

$$E(||\hat{h}_m^N - h_m^N)||^2) \approx tr(C_m)\sigma_w^2 + \frac{m}{N}(\max\{0, X_m\}) \quad (28)$$

and expected value of I.R.E $E(||\hat{h}_m^N - h^N)||^2) \approx$

$$tr(C_m)\sigma_w^2 + \max\{0, (1 + \frac{m}{N})X_m\} \quad (29)$$

where $X_m = \frac{1}{N}||Y - \hat{Y}_1^m||_2^2 - \frac{1}{N}tr(G_m)\sigma_w^2$. Figure (2) shows the simulation results for inputs with different lengths and a fixed noise variance, $\sigma_w = .02$. Figure (3) shows the simulation results for S.I.R.E when $N = 400$ for two different noise variances. Finally, figure (4) shows the simulation results for $N = 400$ with two different noise variances, $\sigma_w = .2$, $\sigma_w = .02$.

8 Order Estimation Problem

Assume that the system has a finite impulse response of length M. Conventional order estimation methods attempt to detect M with a given input/output of length N. Akaike suggests to use a cost function which is the information distance of the estimate of the true system to the true system. When the additive white noise is Gaussian, the cost function for (3) is [1]

$$V(\hat{h}_m(N)) = \lim_{N\to\infty} -E\log(\frac{1}{(\sqrt{2\pi}\sigma_w)^N} e^{-\frac{\epsilon_m^N}{2N\sigma_w^2}}). \quad (30)$$

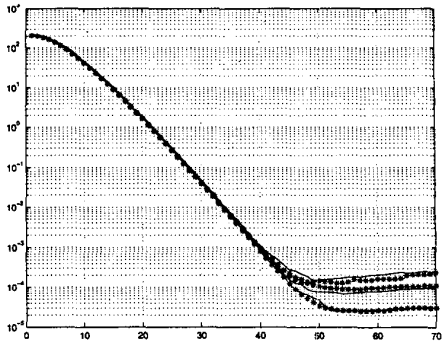

Figure 2: *'*' line: Impulse response error, $||\hat{h}_1^m - h||_2^2$, $\sigma_w = .02$ and N=200,400,1000. Solid line: $E||\hat{h}_1^m - h||_2^2$*

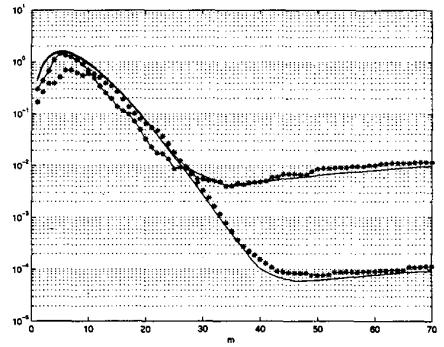

Figure 3: *'*' line: Subspace impulse response error, $||\hat{h}_1^m - h_m||_2^2$, $\sigma_w = .02$, $\sigma_w = .2$, N=400, Solid line: $E||\hat{h}_1^m - h_m||_2^2$.*

where $\epsilon_m^N = ||Y(h) - Y(\hat{h}_m(N))||^2$ is the output error. If the model set includes the true system($m \geq M$), Akaike shows that

$$V(\hat{h}_m(N)) \approx -\log(\frac{1}{(\sqrt{2\pi}\sigma_w)^N} e^{-\frac{\epsilon_m^N}{2N\sigma_w^2}}) + \frac{m}{N} \quad (31)$$

which is Akaike Information Criterion (AIC). Although the expression is not a valid estimate of (30) when $m \leq M$, in AIC method the same expression (31) is used for the model sets of any order.

The idea of minimum description length (MDL) is rooted deeply in information theory. It suggests to pick the model order which minimizes the description length of a given output of the system. Rissanen [6] shows that for model sets of order $m \geq M$, when the additive noise is Gaussian for (3), the MDL criterion is

$$-\log(\frac{1}{(\sqrt{2\pi}\sigma_w)^N} e^{-\frac{||Y(h)-Y(\hat{h}_m(N))||^2}{2N\sigma_w^2}}) + m\frac{\log N}{N}. \quad (32)$$

While BIC uses a different approach to the order estimation problem [7], the criterion in this method is the same as MDL in (32) for when $m \geq M$. Each of the above methods introduces a criterion and calculates the

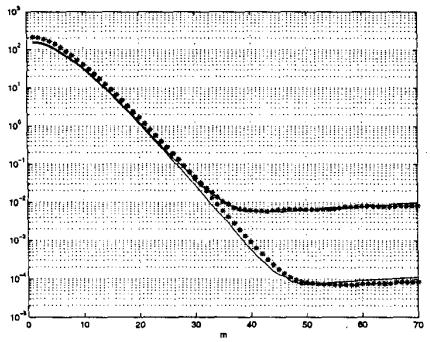

Figure 4: *'*' line: Impulse response error, $||\hat{h}_1^m - h||_2^2$, for $\sigma_w = .02$ and $\sigma_w = .2$ N=400. Solid line: $E||\hat{h}_1^m - h||_2^2$.*

closed form of the criterion for when the true system is an element of the model set, and uses the same closed form for all model sets. To find the closed form of the criterion, in all the mentioned methods, the assumption is that the estimate of the system, $\hat{h}_m$, approaches the true system, h, as N grows, which is valid only if the true system is an element of model set.

We suggest using the impulse response error, $||h-\hat{h}_m^N||^2$, as the model order selection criterion. For example when the input is I.I.D, the estimate of $E||h - \hat{h}_m^N||^2$ in (29) is obtained by using theorem 2 with $\alpha = \beta = \gamma = 0$. When $0 \leq (1 + \frac{m}{N})(\frac{1}{N}||Y - \hat{Y}_1^m||_2^2 - (1 - \frac{m}{N}\sigma_w^2)$, from (29), we have $E(||\hat{h}_m^N - h^N||^2) \approx -\sigma_w^2 + \frac{m}{N}\sigma_w^2 + (1+\frac{m}{N})\frac{1}{N}||Y - \hat{Y}_1^m||^2$. This expression resembles AIC criterion when $m \geq M$ since in this case $\frac{m}{N}(\frac{1}{N}||Y - \hat{Y}_1^m||_2^2 \approx \frac{m}{N}\sigma_w^2$.

In real problems the impulse response usually is not finite. While our method finds the order, m, which minimizes $E||h - \hat{h}(N)||^2$, we could also pick a fixed threshold for acceptable minimum mean-square impulse error. For example let's assume that we ignore any impulse error less than 10^{-3}. As figure 4 shows, for $\sigma_w = .02$, the impulse response error is 10^{-3} at $\bar{m} \approx 40$, while the order, m, which minimizes $||h - \hat{h}_m||^2$ is around 50. In this problem as N goes to infinity, $\bar{m}$ stays around 40. However, the MDL method is not able to use the desired threshold. Since MDL is a consistent method, and the length of the impulse response is infinity, as N goes to infinity, $\hat{m}(MDL)$ also goes to infinity. Same problem raises in order determination for channel identification in communication. [2] shows that as SNR grows, MDL picks higher orders for the impulse response of the channel, while the goal is to only detect the order for which the "significant" part of the impulse response is included.

If the variance of the noise is not known, conventional methods, AIC, MDL, estimate the variance of noise for each model set m "separately" as $\hat{\sigma}_w^2(m) = \frac{1}{N}||Y(h) - Y(\hat{h}_m(N))||^2$ and use it in the proposed criterion. However, in the process of finding the criterion, the variance of the noise is assumed to be fixed, σ_w^2 [3]. In the new method, we suggest to use the estimate of the variance obtained for the highest order, i.e. M_{max}, for all the model sets. Therefore the estimated variance for all model sets is $\hat{\sigma}_w^2 = \frac{1}{N}||Y(h) - Y(\hat{h}_{M_{max}}(N))||^2$.

9 Conclusion

In this paper we suggest a method of quantifying the error of impulse response estimates of LTI systems. The estimation is obtained by parametric least-square identification of the true system in the space of finite impulse responses. We assume that the input is a quasi-stationary signal and for a subset of such input we find a tight upper and lower bound for the impulse response error. We compare the advantages of this method over the conventional order estimation methods. The method can be expanded to order estimation of the impulse response with any general orthonormal basis, for example in the frequency domain estimation[9].

References

[1] H. Akaike. *A New Look at the Statistical Model Identification.* IEEE Trans. on Automatic Control, vol.AC-19, pp.716-723, 1974.

[2] 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.

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

[4] B. Ninness and G.C. Goodwin. *Estimation of Model Quality.* Automatica, vol31, pp.1771-1797, 1995.

[5] F. Paganini. *Set Descriptions of White Noise and Worst Induced Norms.* Proceedings of the 32nd Conference on Decision and Control, pp.3658-3663, December,1993.

[6] J. Rissanen. *Universal Coding, Information, Prediction, and Estimation.* IEEE Trans. on Information Theory, vol.IT30, pp. 629-636, 1984.

[7] G. Schwarz. *Estimating The Dimension of a Model.* The Annals of Statistics, vol.6, pp.461-464, 1978.

[8] S.R. Venkatesh and M.A. Dahleh. *On System-Identification of Complex Processes.* to appear in IEEE Trans. on Automatic Control.

[9] D. Veries and P.M. Van den Hof. *Frequency Domain Identification with Generalized Orthonormal Basis Functions.* IEEE Trans. on Automatic Control, vol.43, pp.656-669, 1998.

[10] G. Zames. *On the Metric Complexity of Causal Linear Systems: Estimates of ϵ-entropy and ϵ-dimension for Continuous Time.* IEEE Trans. A-C, vol.24, pp.222-230, 1979.

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