

# LTI Systems, Additive Noise, and Order Estimation

Soosan Beheshti, Munther A. Dahleh  
 Laboratory for Information and Decision Systems  
 Department of Electrical Engineering and Computer Science  
 Massachusetts Institute of Technology  
 Cambridge, MA 02139, USA

**Abstract**—This paper presents an important application of a novel information theoretic order estimation method, minimum description complexity (MDC). The selection of optimum number of poles and zeros in identification of LTI systems based on observed data is accomplished by MDC. The comparison of MDC with important existing order estimation methods, MDL and AIC, is provided.

## I. INTRODUCTION

In parametric order estimation problem the main goal is to provide a proper parametric model representation by using the finite observed noisy data which is generated by the true model. In this estimation, competing model sets of possibly different order are considered. The goal is to choose a model set and a member of that model set as the “best” representative of the correct model set and the true model. The competing sets might not include the true mode. Hence, the tradeoff between noise fitting and under-modelling plays an important role in order estimation.

Minimum description complexity (MDC) is a new method of order estimation which is proposed in [7]. The approach is based on comparison of a defined distance measure, description complexity. The method is comparable with the existing order estimation methods such as Akaike information criterion (AIC) [1] and two-stage minimum description length (MDL) [3]. MDC can be implemented in various order estimation problems. For example, it is comparable with the existing methods in signal denoising [6]. Also, the method can be used in order estimation with any orthonormal basis. In this paper, we examine one important application of MDC which is order estimation in identification of LTI systems. Order estimation for LTI systems is a special case of order estimation of linear models with additive noise. In this case, both AIC and MDL coincide with special cases of MDC. In this scenario, the order denotes the number of poles and zeros of the LTI system. We compare MDC with AIC and MDL, and discuss some properties of the methods in this application such as importance of consistency.

## II. PROBLEM STATEMENT

We consider the class of stable, causal, single-input/single-output, linear time-invariant, discrete-time systems. Input and output of the system are related as follows

$$\sum_{i=0}^{M_1} b_i^* y[n-i] = \sum_{i=0}^{M_2} a_i^* u[n-i] + w[n], \quad (1)$$

where  $b_0^* = 1$  and  $w[n]$  is additive white Gaussian noise (AWGN) with zero-mean and variance  $\sigma_w^2$ . The additive noise is independent of the input. The input is a quasi-stationary signal [2]. Finite length data, input  $u^N = [u_1, \dots, u_N]^T$  and output  $y^N = [y_1, \dots, y_N]^T$ , is available. The following equation is another representation of (1)

$$By = Au + w. \quad (2)$$

The goal is to estimate the true parameter  $\theta^* = [a_0^*, a_1^*, \dots, a_{M_2}^*, b_1^*, \dots, b_{M_1}^*]^T$ , given the observed data. Note that with the data of finite length  $N$ , the maximum value for  $M_1$  and  $M_2$  is  $N$ . The prior assumption is that  $\theta^*$  is a member of a compact set  $S_{(M_1, M_2)}$ , which is a subset of  $R^{M_1+M_2}$ .

## III. ORDER ESTIMATION METHODS

Consider a subset of  $S_{(M_1, M_2)}$  which is denoted by  $S_{(m_1, m_2)}$  with the following specification. The subset includes parameters of  $S_{(M_1, M_2)}$  in the form

$$\theta = [a_0, a_1, \dots, a_{m_1}, 0, \dots, b_1, \dots, b_{m_1}, 0, \dots, 0]^T. \quad (3)$$

In order estimation methods, first the estimate of the true model in subsets of different order is provided. In this paper the estimation method is ML estimate.

In subset  $S_{(m_1, m_2)}$  the maximum likelihood (ML) estimate of  $\theta^*$  is

$$\hat{\theta}(y^N, u^N, S_{(m_1, m_2)}) = \arg \min_{\theta \in S_{(m_1, m_2)}} f(y^N; \theta, u^N) \quad (4)$$

where  $f(Y^N; \theta, u^N)$  is the probability distribution function (pdf) of the output given the parameter  $\theta$  is generated the data. Here,  $Y^N$  denotes the random process and  $y^N$  is a sample of this random process. To simplify notations, the ML estimate is also denoted by  $\hat{\theta}(S_{(m_1, m_2)})$ .

The next step in order estimation is comparison of the estimates in subsets of different order. In this paper, we consider the nested subsets of the form  $S_{(m_1, m_2)}$ ,  $1 \leq m_1 \leq M_1$  and  $1 \leq m_2 \leq M_2$ . By the comparison, the order estimation method chooses a subset and the estimate in that subset as the best representatives of the true model set and the true model. The method proposes a criterion for comparison of the subsets. Akaike information criterion (AIC) is an estimate of the mean of Kullback-Leibler information distance of

the true density  $f(Y^N; \theta^*, u^N)$ , and the estimated density  $f(Y^N; \hat{\theta}(S_{(m_1, m_2)}, u^N))$ . The criterion in this approach is

$$\text{AIC}(y^N, S_{(m_1, m_2)}) = -\frac{1}{N} \log f(y^N; \hat{\theta}(S_{(m_1, m_2)}, u^N)) + \frac{m}{N}. \quad (5)$$

Another important order estimation method is two-stage minimum description length (MDL) which is defined in [3]. This approach suggests choosing a model set and the ML estimate in that model set which provides the minimum code length. The codelength corresponds to a prefix code which is defined based on Shannon coding. The provided criterion in this approach is

$$\text{MDL}(y^N, S_{(m_1, m_2)}) = -\frac{1}{N} \log f(y^N; \hat{\theta}(S_{(m_1, m_2)}, u^N)) + m \frac{\log N}{2N}. \quad (6)$$

Bayesian information criterion (BIC) approach assumes a prior probability for the competing model sets and suggests to select the model which yields the maximum posterior probability [4]. BIC's criterion is the same as the two-stage MDL in (6).

It is very important to note that the closed forms for AIC in (5) and MDL and BIC in (6) are provided with the following two assumptions

$$\theta^* \in S_{(m_1 + m_2)} \text{ and } m_1 + m_2 \ll N. \quad (7)$$

However, in application these assumptions can't be checked a priori for all the competing subsets.

One important problem here is calculation of these criteria for the subsets which do not include  $\theta^*$ . No method of validation of the prior assumptions in (7) for all the subsets exists. Hence, since the closed forms obtained in (5) and (6) are only functions of  $y^N$ ,  $\hat{\theta}(S_{(m_1, m_2)})$ ,  $m_1$ ,  $m_2$ , and  $N$ , the same closed forms are used for *all* the subsets.

#### A. Minimum Description Complexity (MDC)

A new method of subset selection and quality evaluation is introduced in [7]. The method assumes that the parametric pdf  $f(y^N; \theta, u^N)$  is a continuous function of  $Y^N$ ,  $\theta$ . Unlike the existing methods, this approach does not use the prior assumption in (7).

Here we describe the method briefly. Define a positive cost function  $V(\theta, y^N)$  on  $S_{(M_1, M_2)}$  and  $Y^N$  for which  $E_{\theta_1} \frac{1}{N} V(\theta_2, Y^N) \geq 0$  is a finite number and

$$E_{\theta_1} \frac{1}{N} V(\theta_2, Y^N) \geq E_{\theta_1} \frac{1}{N} V(\theta_1, Y^N) \quad (8)$$

for any  $\theta_1$  and any  $\theta_2$  in  $S_{(M_1, M_2)}$ , with the equality only for when  $\theta_1 = \theta_2$ . Choose the cost function such that it is a continuous function of both  $\theta$  and  $y^N$ .

*Definition:* The description complexity of  $Y^N$  with parameter  $\theta_1$ , when the data is generated by  $\theta$ , is defined by

$$\text{DC}_N(\theta, \theta_1) \equiv E_{\theta} \frac{1}{N} V(\theta_1, Y^N). \quad (9)$$

For any element of  $S_{(M_1, M_2)}$ , define  $\bar{\theta}_{S_{(m_1, m_2)}}$  in set  $S_{(m_1, m_2)}$  as

$$\bar{\theta}(S_{(m_1, m_2)}) = \arg \min_{\theta_1 \in S_{(m_1, m_2)}} E_{\theta} \frac{1}{N} V(\theta_1, Y^N). \quad (10)$$

In this process the class of estimators are chosen such that for any  $\theta \in S_{(M_1, M_2)}$

$$E_{\theta} \left( \hat{\theta}(Y^N, u^N, S_{(m_1 + m_2)}) \right) = \bar{\theta}(S_{(m_1 + m_2)}) \quad (11)$$

Note that this condition implies that if  $\theta$  is an element of  $S_{(m_1, m_2)}$ , the estimator in that subset is unbiased.

Given a set of observed data, the goal is to estimate  $\text{DC}(\theta^*, \hat{\theta}(S_{(m_1, m_2)}))$  in each subset and choose the subset which minimizes this criterion. The challenge is to use the observed data, and therefore  $V(y^N, \hat{\theta}(S_{(m_1, m_2)}))$ , in each subset to provide bounds on the criterion, without the prior assumption in (7). With validation probability  $p$  bounds on this criterion is provided in [7]

$$L_{S_{(m_1, m_2)}} \leq \text{DC}(\theta^*, \hat{\theta}(S_{(m_1, m_2)})) \leq U_{S_{(m_1, m_2)}} \quad (12)$$

where both the upper and lower bounds are functions of  $p$ ,  $V(y^N, \hat{\theta}(S_{(m_1, m_2)}))$ , and the probability distribution of  $V(Y^N, \hat{\theta}(S_{(m_1, m_2)}))$ . Details of the calculation of the bounds on the desired description complexity is in [7]. For comparison of the subsets the obtained upper bound is compared and the subset for which this bound is minimized is chosen.

The behavior of lower and upper bounds as a function of  $N$  and as the length of data is growing is studied in [7]. In each subset, the relation between the estimator  $\hat{\theta}(S_{(m_1, m_2)})$  and  $\bar{\theta}^*(S_{(m_1, m_2)})$  depend on the cost function, the probability distribution family and the prior assumption on the order of  $S_{(M_1, M_2)}$ .

#### IV. APPLICATION OF AIC, MDL, AND MDC FOR THE CONSIDERED MODEL CLASS

In this section the order estimation criteria for the model class in (1) is provided. The probability distribution of output for each  $\theta$  is

$$f(y^N; \theta, u^N) = \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{\|y^N - \bar{y}^N(\theta)\|_2^2}{2N\sigma_w^2}} \quad (13)$$

where  $\bar{y}^N(\theta)$  is the mean of  $Y^N$  when the data is generated by  $\theta$  in (1). From (4), the ML estimator of  $\theta^*$  in subset  $S_{(m_1, m_2)}$  is

$$\hat{\theta}(S_{(m_1, m_2)}) = \arg \min_{\theta \in S_{(m_1, m_2)}} \|y^N - \bar{y}^N(\theta)\|^2 \quad (14)$$

For calculation of AIC and MDL, the distribution in (13) is substituted in (5) and (6).

1) *Calculation of MDC*: We use the following cost function

$$V(\theta, y^N) = -\log f(y^N; \theta, u^N) \quad (15)$$

which satisfies the condition in (8). In this case

$$\text{DC}_N(\theta, \theta) = \log \sqrt{2\pi\sigma_w^2} + \frac{1}{2} \quad (16)$$

is the differential entropy which is a fixed number for all elements of  $S_{(m_1, m_2)}$ . Therefore, comparison of MDC is same as comparison of the Kullback-Leibler distance of  $\theta^*$  and  $\hat{\theta}_{S_{(m_1, m_2)}}$ . Note that this is the same criterion that AIC estimates. However, since MDC does not implement the prior assumptions in (7), the approach is different from AIC.

It is important to note that for this class of problems the ML estimator satisfies the necessary condition in (11).

The upper and lower bounds on MDC are provided probabilistically in two steps:

*step 1*: The observed information  $V(\hat{\theta}(S_{(m_1, m_2)}), y^N)$  is a sample of a Chi-square distribution. The expected value and variance of this random variable are given in Appendix A. Because of the structure of this random variable we can probabilistically validate  $\text{DC}(\theta^*, \hat{\theta}^*(S_{(m_1, m_2)}))$ . In this step, the description complexity,  $\text{DC}(\theta^*, \hat{\theta}^*(S_{(m_1, m_2)}))$  is validated for each subset with probability  $p_1$ .

*step 2*: Next is to estimate  $\text{DC}(\theta^*, \hat{\theta}(S_{(m_1, m_2)}))$ , which itself is a sample of a Chi-square random variable. The expected value and variance of this random variable are given in Appendix A. Because of the structure of this random variable, the validation in Step 1 is enough to provide probabilistic bound on this random variable. With confidence probability  $p$  and validation probability  $p_1$  upper and lower bounds for  $\text{DC}(\theta^*, \hat{\theta}(S_{(m_1, m_2)}))$  are provided.

Note that calculation of this criterion is not as straightforward as calculation of AIC in (5) and MDL in (6). This is due to the fact that the prior restricted assumptions in (7) are not used. For subsets that  $m_1 + m_2$  is large enough and  $M_1 + M_2 - (m_1 + m_2)$  is also Large enough, the two Chi-square distributions can be estimated with Gaussian distributions by the use of central limit theorem. The provided bounds with confidence probability  $p = Q(\beta_N)$  and validation probability  $p_1 = Q(\alpha_N)$  are provided in [7]<sup>1</sup>. The method of approach in this calculation is similar to what has been introduced in [5]. In this case, provided that  $\alpha_N \geq \sqrt{\frac{N}{2}}(1 - \frac{x}{(1 - \frac{x}{N})\sigma_w^2})$ , the upperbound on MDC which is used as the comparison criterion is

$$\begin{aligned} \text{MDC}(y^N, S_{(m_1, m_2)}) &= g(y^N, S_{(m_1, m_2)}) + \frac{m}{N}\sigma^2 \\ &+ \beta \sqrt{\frac{4m}{N} + g(y^N, S_{(m_1, m_2)})} \quad (17) \end{aligned}$$

$${}^1 Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-\frac{u^2}{2}} du$$

where

$$g(y^N, S_{(m_1, m_2)}) = x - m_w + \frac{2\alpha^2\sigma_w^2}{N} + K(\alpha). \quad (18)$$

and  $m_w = (1 - \frac{m_1 + m_2}{N})\sigma_w^2$  and

$$K(\alpha) = 2\alpha \frac{\sigma_w}{\sqrt{N}} \sqrt{\frac{\alpha^2\sigma_w^2}{N} + x - \frac{1}{2}m_w}. \quad (19)$$

where  $x$  is defined as follows

$$x = V(\hat{\theta}_{S_{(m_1, m_2)}}, y^N) \quad (20)$$

and is calculated by using the observed data.

Proper conditions on  $\alpha_N$  and  $\beta_N$  for when the length of data is growing are provided in [5], [7]. In the subset  $S_{(m_1, m_2)}$ , as the length of the data grows, if the following conditions are satisfied

$$\lim_{N \rightarrow \infty} \alpha_N = \infty, \quad \lim_{N \rightarrow \infty} \beta_N = \infty, \quad (21)$$

$$\lim_{N \rightarrow \infty} \frac{\alpha_N}{\sqrt{N}} = 0, \quad \lim_{N \rightarrow \infty} \frac{\beta_N}{N} = 0, \quad (22)$$

the upper and lower bounds on MDC approach each other and we have

$$\hat{\theta}(S_{(m_1, m_2)}) \rightarrow \bar{\theta}^*(S_{(m_1, m_2)}). \quad (23)$$

Also, with these conditions the validation and confidence probabilities approach one, which guarantees the consistency of MDC.

It is important to mention that for this model class both AIC in (5) and MDL in (6) coincide with special cases of MDC with special choices of  $\alpha_N$  and  $\beta_N$ . When  $\alpha_N = \beta_N = 0$ , MDC is the same as AIC. With  $\alpha_N = 0$  and  $\beta_N = \sqrt{m_1 + m_2} \log(N)$ , MDC is the same as two-stage MDL [7].

## V. IDENTIFICATION AND ORDER ESTIMATION FOR LTI SYSTEMS IN APPLICATION

In this section, we implement the methods for order estimation of LTI systems with input-output relation given in (1). We start with the least complicated model class, FIR models. The presence of the additive noise plays an important role in the following model classification and order estimation.

### A. Finite Length Impulse Response (FIR) Models

Consider a subclass of models in (1) such that

$$y[n] = \sum_{i=0}^{M_2} a_i^* u[n-i] + w[n]. \quad (24)$$

This is the class for which  $M_1 = 0$ . To implement AIC and MDL based on the required conditions in (7),  $M_2$  has to be much smaller than  $N$ ,  $M_2 \ll N$ . Since for calculation of these criteria only the asymptotic results are used, the notion of very small,  $\ll$ , for finite  $N$  is not well defined. For these conventional order estimation approaches the only important

method of quality evaluation is to check the behavior of the estimator as length of data grows, i.e., to check whether the order estimate is consistent. A consistent method chooses the correct model set which includes  $\theta^*$  and has minimum order asymptotically.

On the other hand, MDC does not impose any constraint on  $M_2$ . In this approach the finite number  $M_2$  can be the same as the length of data  $N$ . The consistency of MDC is guaranteed with proper choice of  $p$  and  $p_1$ . Also, the provided upper bound and lower bound on MDC for different  $p$  and  $p_1$  (validation and confidence probabilities) can be used for evaluation of the quality of the estimates.

We demonstrate the performance of the methods with the following example. Consider an FIR filter of length 30, for which

$$a_i^* = .3(.5)^{i-1} + 3(i-1)(.8)^{i-1}, 0 \leq i \leq 30. \quad (25)$$

In this simulation, the input is an independent identically distributed (IID) Bernoulli sequence of  $\pm 1$ . In the first simulation, the length of data is  $N = 200$  and the signal to noise ratio (SNR) is 10dB. The optimum  $m_2$  provided by the order estimation methods for 50 trials are shown in Figure (1). For MDC the validation and confidence probabilities are  $p = p_1 = .9991$ , i.e., for when the Gaussian estimates of the Chi-square distributions are used and we have  $\alpha = \beta = 2 \log(\log(N))$ . As the figure shows, AIC overestimates the

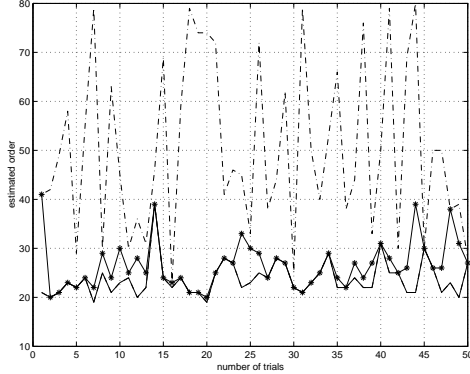


Fig. 1. Optimum order  $m_2^*$  for 50 trials with  $N = 200$ , SNR=10dB. AIC: Dashed line '-'. MDL: Solid line. MDC: Solid line with star '\*',  $p = p_1 = 0.9991$ .

length of the filter. The average of  $\bar{m}_2^*$ s in this 50 trials is

$$\begin{aligned} \text{AIC} : \bar{m}_2^* &= 50, & \text{MDL} : \bar{m}_2^* &= 23 \\ \text{MDC} : \bar{m}_2^* &= 27 \end{aligned} \quad (26)$$

On average MDC's estimate is closer to 30. Figure (2) shows another example with a higher SNR (SNR=15db) and a longer data ( $N = 300$ ). In this case

$$\begin{aligned} \text{AIC} : \bar{m}_2^* &= 52, & \text{MDL} : \bar{m}_2^* &= 27 \\ \text{MDC} : \bar{m}_2^* &= 29 \end{aligned} \quad (27)$$

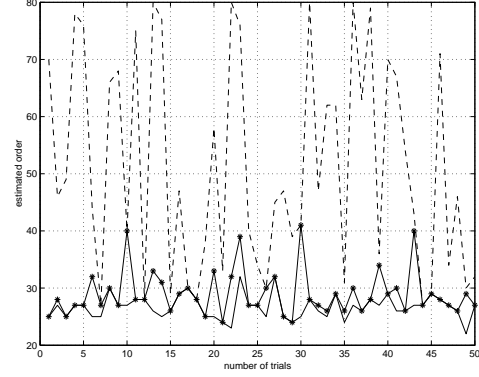


Fig. 2. Optimum order  $m_2^*$  for 50 trials with  $N=300$ , SNR=15dB. AIC: Dashed line '-'. MDL: Solid line. MDC: Solid line with star '\*',  $p = p_1 = 0.9995$ .

The optimum order choice of all the methods is a function of  $N$ ,  $M_2$  and the noise variance ( or equivalently SNR). However, only for MDC, a method of quality evaluation of the estimates for *finite* length data is available through calculation of both upper and lower bounds on the desired criterion.

## B. Infinite Length Impulse Response (IIR) Models

Consider a subclass of models in (1) such that

$$y[n] = \sum_{i=0}^{\infty} a_i^* u[n-i] + w[n]. \quad (28)$$

This is the class for which  $M_1 = 0$  and the prior knowledge is that  $M_2$  in (24) is not finite, but the impulse response is an  $l_1$  sequence. Although AIC and MDL are implemented for this case, only a fast decay rate on the impulse response compare to the data length provides a proper situation similar to the condition  $M_2 \ll N$ , for large enough  $N$ . Note that in this case with a finite length data, a model with  $M_2 = N$  is enough to generate the output. Similar to the FIR case, MDC can assume that  $M_2 = N$ .

In the following simulation, the impulse response is

$$a_i^* = .3(.5)^{i-1} + 3(i-1)(.8)^{i-1}, 0 \leq i. \quad (29)$$

This is a stable system with 2 poles. With a fixed SNR, Both MDL and AIC choose larger and larger orders as the length of the data grows. Note that consistency of MDL guarantees the choice of larger and larger orders as the length grows since the true filter here is not finite length. However, in practical problems we might need to provide an FIR estimate of the IIR system even as the length of data grows. This issue is raised in problems such as in blind channel identification [9]. For these cases, with proper choice of  $p$  and  $p_1$ , MDC provides an FIR estimate for the system. MDC thresholding is another method which can be used in this scenario [7].

1) *Additive Output Noise*: Colored noise in the estimation causes nonlinearity in the estimation error. To avoid this problem we suggest representing a class of models with colored noise with a subclass of IIR systems. Consider a class of models with the following structure

$$\sum_{i=0}^{M_1} b_i^* (y[n-i] + w[n-i]) = \sum_{i=0}^{M_2} a_i^* u[n-i], \quad (30)$$

where  $b_0^* = 0$  and  $M_1$  and  $M_2$  are finite and  $w[n]$  is AWGN. The input-output relation can be represented also as follows

$$B(y+w) = Au. \quad (31)$$

For this model class the noise is colored. The input-output relationship is also in the form

$$y = \frac{A}{B}u + w. \quad (32)$$

For this new representation  $M_1 = 0$ , and  $M_2$  is not finite. Therefore, this is a subclass of the IIR models which were discussed previously.

### C. Auto-Regression(ARX) Models

Consider the model class in (1) or equivalently

$$By = Au + w. \quad (33)$$

In this case, the order estimation method estimates the number of poles and zeros of the system.

An example of order estimation is provided in the following simulation. The simulated model has the following zero and poles

$$\text{zeros} : [.5 \quad -.8 \quad -.3 + .5j \quad -.3 - .5j \quad 2 \quad 5] \quad (34)$$

$$\text{poles} : [.2 + .8j \quad .2 - .8j \quad .3 \quad -.2 \quad -.6 \quad .7 \\ .35 \quad -.1 - .1j \quad -.1 + .1j] \quad (35)$$

Therefore, the correct order is for a subset with 7 poles,  $m_1^* = 7$  and 9 zeros,  $m_2^* = 9$ . Similar to the previous simulations, the unit power input is IID. The length of the data is  $N = 300$ , and the noise variance is  $\sigma_w = .1$ . For 30 trials the average estimated order with AIC and MDL are

$$\text{AIC} : \bar{m}_1^* = 10, \quad \bar{m}_2^* = 10 \quad (36)$$

$$\text{MDL} : \bar{m}_1^* = 5, \quad \bar{m}_2^* = 7 \quad (37)$$

MDC estimate with  $p = p_1 = .9995$  is also calculated. The average estimated order of the 30 trials is

$$\text{MDC} : \bar{m}_1^* = 9, \quad \bar{m}_2^* = 7. \quad (38)$$

As the length of data is increased to  $N = 1000$  in all the methods the variance of optimum order in 30 trials becomes smaller. AIC is over estimating both the number of zeros and poles and MDL still is under estimating and chooses

$$\text{MDL} : \bar{m}_1^* = 7, \quad \bar{m}_2^* = 7. \quad (39)$$

while MDC still selects the correct order.

In this simulation for finite  $N$ , the performance of MDC is better than MDL. As it was mentioned before, MDL is the same as MDC when  $\alpha_N = 0$  and  $\beta_N = \sqrt{m_1 + m_2} \log(N)$ . MDL is under-modeling compare to the MDC we used. In application of MDC,  $\alpha_N$  and  $\beta_N$  are not functions of the order. For the purpose of comparison, these numbers have to be the same for all the subsets. MDL provides a lower bound for MDC with  $\alpha_N = 0$  and  $\beta_N = \log(N)$ . This argument explains the MDL under-modeling for when the length of data is finite and is not large *enough*.

1) *Additive Colored Noise*: For the discussed model class in (1) we assume that the additive noise is white. Here, we consider the model class which has a colored additive noise. Similar approach to what is argued for the IIR systems and models with colored noise in Section V-B.1 can be considered for this model sets. Consider the following class of models

$$\sum_{i=0}^{M_1} b_i y[n-i] = \sum_{i=0}^{M_2} a_i u[n-i] + \sum_{i=0}^{M_3} c_i w[n-i], \quad (40)$$

where  $b_0 = 1$ , or equivalently

$$\frac{B}{C}y = \frac{A}{C}x + w \quad (41)$$

This model is a subclass of models with infinite number of zero and poles, i.e., the ARX models with white noise when  $M_1$  and  $M_2$  are not finite. With the additional assumption that  $C$ 's zeros are inside the unit circle, the system model is stable and the identification and order estimation procedure is similarly extended from the case with finite number of zero and poles for this model class. Similar to IIR models, in this case consistency is not the helpful property of the order estimator. It is important to provide a method of quality evaluation of the estimated order and therefore, MDC is the preferred.

### D. Unknown Noise Variance

The noise variance in this paper is assumed to be known. In AIC and MDL unknown noise variance is estimated in each subset *separately*. The calculation of the variance in each subset can be done by estimation of the cross-correlation and the use of singular value decomposition (SVD) [8]. However, in MDC for comparison of the model sets the *same* estimate of variance for all the subsets is used. [6] has the details on simultaneous order estimation and noise variance estimation with MDC.

## VI. CONCLUSION

In this paper, an efficient method of order estimation for LTI systems is provided. It is assumed that the additive noise is Gaussian. However, by using the law of large numbers, the order estimation method can be implemented for a larger set of probabilistic model classes with additive non-Gaussian noise. Also, the white additive noise can be generalized to colored additive noise. This extends the application of the

new method for order estimation of ARX models to order estimation of ARMAX models.

Some critical prior assumptions which are used in calculating AIC and MDL can not be validated. Also, the closed forms in these methods are provided for large enough data length. For these methods the important property to check is consistency. It is known that AIC is not consistent and over-estimates the order. On the other hand, as it is shown in the examples, the consistent MDL approach is under-estimating the true order in some cases. In this paper it is shown that the consistency of MDC depends on the choice of the validation and confidence probabilities. In practical problems, more than consistency of a method is needed. It is required to provide a method of quality evaluation of the order estimation for a given large, but finite, length of data. MDC provides a method of probabilistic validation of the bounds on the desired criterion. Unlike the conventional methods, the variances of the involved random variables in calculation of the criterion, which are nonzero due to the finiteness of the length of data, are not ignored. The quality of the estimated order depends on the structure of the model class and validation and confidence probability. The consistency of the method, as the length of data grows, is guaranteed by proper choice of validation and confidence probabilities. With these properties and in comparison with the existing methods, MDC promises to be the prominent candidate for order estimation of LTI systems.

#### VII. APPENDIX A

The mean and variance of Chi-square random variable  $V(\hat{\theta}(S_{(m_1, m_2)}), Y^N)$  are

$$E_{\theta}(V(\hat{\theta}(S_{(m_1, m_2)}), Y^N)) = DC_N(\theta, \bar{\theta}(S_{(m_1, m_2)})) - \frac{m_1 + m_2}{2N} \quad (42)$$

$$\begin{aligned} \text{var}\left(V(\hat{\theta}(S_{(m_1, m_2)}), Y^N)\right) &= \frac{N - (m_1 + m_2)}{2N^2} \\ &+ \frac{1}{N} (DC_N(\theta, \bar{\theta}(S_{(m_1, m_2)})) - DC_N(\theta, \theta)). \end{aligned} \quad (43)$$

For the first step, consider the subsets in  $Y^N$  which are defined by each  $\theta$  as follows. The set is a ball whose center is the mean of  $V(\hat{\theta}(S_{(m_1, m_2)}), Y^N)$  and probability of the set is  $p_1$ . Validate  $\theta$ s for which the observed sample  $V(\hat{\theta}(S_{(m_1, m_2)}), y^N)$  is in this set defined by  $\theta$ .

The desired criterion  $DC(\theta, \hat{\theta}_{S_{(m_1, m_2)}})$  is a random variable with the following mean and variance

$$E_{\theta} DC_N(\theta, \hat{\theta}(S_{(m_1, m_2)})) = DC_N(\theta, \bar{\theta}(S_{(m_1, m_2)})) + \frac{m_1 + m_2}{2N} \quad (44)$$

$$\text{var}_{\theta} DC_N(\theta, \hat{\theta}(S_{(m_1, m_2)})) = \frac{m_1 + m_2}{2N^2}. \quad (45)$$

Therefore, with probability  $p$ , similar to the validation step, upper and lower bounds on this random variable based on

the validated bounds on  $DC_N(\theta, \bar{\theta}(S_{(m_1, m_2)}))$  are provided

#### VIII. REFERENCES

- [1] H. Akaike. "A New Look at the Statistical Model Identification". *IEEE Trans. Automat. Contr.*, vol.AC-19, pp.716-723, 1974.
- [2] L. Ljung. *System Identification: Theory for the User*. NJ: Prentice-Hall, 1998.
- [3] Y. Barron, J. Rissanen, Bin Yu. "The Minimum Description Length Principle in Coding and Modeling". *IEEE Trans. Inform. theory*, vol 44, pp.2743-2760, Oct. 1998.
- [4] G. Schwarz. "Estimating The Dimension of a Model". *The Annals of Statistics*, vol.6, pp.461-464, 1978.
- [5] S. Beheshti. and M.A. Dahleh. "On Model Quality Evaluation of Stable LTI Systems". *Proceeding of the 39th IEEE Conference on Decision and Control* pp.2716-2721, 2000.
- [6] S. Beheshti and M.A. Dahleh. "Noise Variance and Signal Denosing". *International Conference on Acoustics, Speech, and Signal Processing(ICASSP)* April 2003.
- [7] S. Beheshti and M.A. Dahleh. "A New Information Theoretic Approach to Order Estimation Problem". *13th IFAC Symposium on System Identification*, August 2003.
- [8] M. Wax and T. Kailath. "Detection of Signals by Information Theoretic Criteria". *IEEE Trans. Acoust., Speech, Signal Processing*, vol.ASS P-33, pp.387-392, 1985.
- [9] A.P. Liavas, P.A. Regalia, and J. Delmas. "Blind channel Approximation: Effective Channel Order Estimation". *IEEE Trans. Signal Processing*, vol.47, pp.3336-3344, 1999.