IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 August 2023, accepted 24 September 2023, date of publication 4 October 2023, date of current version 19 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3321794

RESEARCH ARTICLE

Relative Entropy (RE)-Based LTI System Modeling Equipped With Simultaneous Time Delay Estimation and Online Modeling

MAHDI SHAMSI, (Member, IEEE), AND SOOSAN BEHESHTI, (Senior Member, IEEE)
Department of Electrical, Computer, and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada
Corresponding author: Soosan Beheshti (soosan@torontomu.ca)

ABSTRACT This paper proposes a novel and efficient method of impulse response modeling in presence of input and noisy output of a linear time-invariant (LTI) system. The approach utilizes Relative Entropy (RE) to provide the impulse response estimate of the system with an optimum length as well as an optimum time delay. Classical methods for this system modeling use two separate steps for the time delay estimation and for the impulse response length selection. Existing time delay methods focus on various proposed criteria, while the existing order selection approaches choose the optimum impulse response length based on their own criteria that are different from the time delay approaches. The strength of the proposed RE based method is in using only ''one'' criterion, the RE based criterion, to estimate both the time delay and the impulse response length simultaneously. The desired RE is the Kullback-Leilber divergence of the estimated distribution from its unknown true distribution. A unique probabilistic validation approach estimates this unavailable desired relative entropy and minimizes this criterion to provide the impulse response estimate. In addition, estimation of the noise variance, when the Signal to Noise Ratio (SNR) is unknown, is concurrent and is based on optimizing the same RE based criterion. This work elaborates the critical role of the data length and the SNR in data based LTI system modeling. The approach is also extended for online impulse response estimation. The proposed online method reduces computational complexity of the offline model estimation upon the arrival of a new sample. The introduced efficient stopping criterion for the online approach is extremely valuable in practical applications. Simulation results depict superiority of the RE based approach as a time delay estimator or as an order selection approach compared to the conventional methods. They also illustrate precision and efficiency of the proposed method compared to the state of the art methods in simultaneous time delay estimation and order selection. Not only RE based method outperforms the competing approaches, but also is shown to be more robust to the variations of the SNR.

INDEX TERMS Relative entropy, impulse response estimation, LTI system, time delay estimation, order selection, online modeling.

I. INTRODUCTION

Linear Time-Invariant (LTI) systems characterize a wide range of dynamics around us. Modeling impulse response of these systems using a finite length input and noisy output is the focus of this work. In practical applications, due to the uncertainty caused by noisy observations, issues such as underparamatrization or overparameterization of the impulse response estimate cause very challenging problems. Most existing estimators use the mean square error (MSE) to find the parameter estimates in this setting. If the data have a length of N, it is known that up to N coefficients of the impulse response can be estimated from the available data [1]. However, the main challenge in this scenario is which subspace of these coefficients should be chosen for the MSE estimate. In this paper we rely on Relative Entropy (RE) to find the optimum estimate of the impulse response. Relative Entropy (also known as Kullback-Leibler divergence) is a statistical distance that measures the difference between two probability distributions. Relative entropy is a divergence in terms of information geometry [2] and has a wide range of applications from

The associate editor coordinating the review of this manuscript and approving it for publication was Mouquan Shen.

the theory of information theory to signal processing [3], control systems [4], [5], sensor networks [6], cryptography [7], machine learning [8], [9], [10] and physics [11]. If the relative entropy between two probability distributions is zero, then these two distributions have identical quantities of information [12]. In this system modeling context, the relative entropy can quantify the similarity between the unavailable model and the approximated model. The proposed RE based method provides the impulse response coefficients estimate and includes the optimum estimate of the associated time delay as well as the optimum estimate of the impulse response length.

Order selection or estimation of impulse response length is an important task for the purpose of overestimation avoidance [1] and model selection methods are critical in a wide range of applications and in areas such as brain source localization [13], [14], wireless sensor networks [15] and machine learning [16]. While the existing order selection methods concentrate on estimating the length of the impulse response, they do not estimate the time delay of the system. On the other hand, time delay estimation methods such as the Cumulative Sum (CUSUM) method [14], [17], the frequency domain-based method [18] and parametric methods (MATLAB®*delayest*) use different criteria to estimate the time delay only [14]. Note that time delay itself is an ubiquitous physical phenomenon that often occurs in communication systems [19], [20], power systems [21], biological systems [22], [23], transportation systems, mechatronic systems [24] and industrial processes such as chemical processing systems [25]. Inaccurate or improper time delay estimation can cause deterioration in stability and performance in control process. Estimating the time delay is necessary in system modeling, identification algorithms and control systems [26], [27].

The motivation of this work is based on the similarity of the problem in both time delay estimation or order selection in the presence of persistently exciting input and noisy output. The main challenge in both scenarios is caused by the existence of the additive noise in the output. So while approaches in both the time delay estimation and the order selection start with the least square estimate of the impulse response, the main question is how to denoise this resulted impulse response. In the absence of the additive noise, the least square impulse response estimate can provide the exact time delay and will have the exact desired length. Therefore, in this work we concentrate on introducing and utilizing a criterion that extracts the optimally denoised version of the impulse response that can consequently provide the time delay and order of the impulse response simultaneously. The proposed RE based method provides the optimum impulse response with the optimum length and delay all at once by minimizing the desired relative entropy criterion. Through a unique approach probabilistic bounds on the Relative Entropy are calculated which enables this optimization procedure. The powerful relative entropy criterion even enables the method to simultaneously estimate the variance of the output noise when it is unknown. In this procedure, estimation of the reconstruction error, which is the mean squared error (MSE) between the estimated output and the unavailable noise-free output, is required. Preliminary work for the estimation of this error is available in [1], [28]. The probabilistic bounds on reconstruction error have been shown to be useful and efficient in a wide range of applications such as blind source separation [29], brain source localization [30], compressed sensing [31] and the number of source signal estimation [32]. These bounds are shown to be also important and relevant in the calculation of the desired RE criterion.

In many real-world applications of impulse response estimation such as sensor validation [33] and power systems [34], it is desirable to estimate the impulse response of the system, online. For example, power systems must respond quickly to power outages to avoid generators damage, under-frequency load-shedding, and fault cascades [34]. Therefore, online monitoring of power systems is a crucial element in the stability of power networks. In general, in online applications, the dimension of the data increases rapidly as time grows. Consequently the computational complexity, and therefore its cost, increases accordingly [35], [36] and re-estimating the optimal impulse response with arrival of the new data is not efficient. To address this problem we equip the proposed RE based method with a recursive online estimation strategy. The method updates the time delay and model order estimates with respect to the new received data. As a result, the complexity order of the impulse response estimation decreases from $O(N^3)$ to $O(N^2)$. Furthermore, a novel efficient stopping criterion is introduced for this online system modeling. The stopping criterion refers to the condition that must be met in order to terminate the execution of the algorithms. Therefore, it is an important factor in the efficiency of online estimation methods [37]. It is important to note that this online estimation, empowered with the consistent stopping criterion, has a great potential as a reliable method for modeling slowly varying LTI systems. In addition, it is worth mentioning that while this work concentrates on the case of single-input single-output system, the approach is easily expandable and applicable for multi-input and multi-output system.

The paper is organized as follows: Section II states the problem and Section III introduces the mathematical foundation and notations used in this paper. Relative Entropy (RE)-based system modeling is presented in Section IV. Section V introduces the online RE-based system modeling. The simulation results are provided in Section VI and Section VII is the conclusion.

II. PROBLEM STATEMENT

Consider a causal linear time-invariant (LTI), single-input single-output system with the following noise-free output

$$\bar{y}(n) = \sum_{i=0}^{\infty} \bar{\theta}(n) u(n-i) \tag{1}$$

where $u(n)$ is the input at time n, and $\bar{\theta}(n)$ is the unknown impulse response of the system which can be represented in the following column vector format

$$\begin{aligned}\bar{\theta} &= [\bar{\theta}(0),\ \cdots,\ \bar{\theta}(\bar{d}),\ \ldots] \\ &= [\underbrace{0,\ \ldots,\ 0}_{\bar{d}},\ \bar{\theta}(\bar{d}),\ \bar{\theta}(\bar{d}+1),\ \ldots] \end{aligned} \tag{2}$$

where $\bar{d}$ is the true unknown delay of the system. The observed noisy output of the system is

$$y(n) = \bar{y}(n) + \omega(n) \tag{3}$$

where $\omega(n)$ is the additive white Gaussian noise with zero mean and variance of σ_ω^2. Given the following input-output data with length N

$$\begin{aligned} u^N &= [u(0),\ u(1), \ldots, u(N-1)] \\ y^N &= [y(0), y(1), \ldots, y(N-1)], \end{aligned} \tag{4}$$

the goal is to estimate the impulse response and the corresponding time delay, $\bar{d}$, in (2). In this estimation, the optimum length of the impulse response has to be simultaneously provided. Note that this problem setting is different from problems such as echo cancellation in which no impulse response is involved and the signal delay is estimated based on comparing two noisy observations [38].

III. IMPULSE RESPONSE MEAN SQUARE ERROR (MSE) ESTIMATE

Before proposing the method, the notations are established. Here, the true unavailable time delay of the impulse response is $\bar{d}$ and if the filter is finite length, the unavailable true length is $\bar{m}$. First, the MSE estimate of the impulse response for a range of impulse response length m where $1 \le m \le M$ and for delay d, $1 \le d \le m$ is calculated. The following will establish notations for the MSE estimate of the impulse response.

A. IMPULSE RESPONSE ESTIMATE FOR TIME DELAY d

It is known that given the input-output data of length N, at most the first N coefficients of the impulse response can be estimated. In practical application, the length of the estimated impulse response is set to M, which can be smaller than N to avoid excess noise fitting or due to partial information about the system structure. Consequently, for any chosen value of M, such that $M \le N$, it is desired to find the estimate of the first M coefficients of the unknown impulse response in (2):

$$\bar{\theta} = [\underbrace{0,\ \ldots,\ 0}_{\bar{d}},\ \underbrace{\bar{\theta}(\bar{d}),\ \bar{\theta}(\bar{d}+1),\ \ldots\ \bar{\theta}(M-1)}_{\bar{\theta}_{\bar{d},M}}] \tag{5}$$

where

$$\bar{\theta}_{\bar{d},M} = [\bar{\theta}(\bar{d}),\ \bar{\theta}(\bar{d}+1),\ \ldots\ \bar{\theta}(M-1)] \tag{6}$$

denotes the values of $d+1$st to Mth elements of the impulse response.

Lets denote impulse response coefficients (IRC) of length M with a time delay of d $(0 \le d \le M)$ as follows

$$\theta = [\underbrace{0,\ \ldots,\ 0}_{d},\ \underbrace{\theta(d),\ \theta(d+1),\ \ldots\ \theta(M-1)}_{\theta_{d,M}}] \tag{7}$$

B. IMPULSE RESPONSE ESTIMATE FOR TIME DELAY d AND LENGTH m

In real applications the additive noise may corrupt the impulse response estimate of length M. This noise overfitting can happen when the true impulse response is of finite length less than M, or when it has infinite length but small values of the tail coefficients are comparable with the noise standard deviation and their estimates are very much corrupted. In this scenario we consider class of impulse responses with the following structure $(d < m \le M)$, by generalizing the structure of impulse response coefficients in (7) with the new variable m:

$$\theta = [\underbrace{0,\ \ldots,\ 0}_{d\ zeros},\ \underbrace{\theta(d),\ \ldots,\ \theta(m-1)}_{\theta_{d,m}},\ \underbrace{0,\ \ldots,\ 0}_{M-m\ zeros}] \tag{8}$$

This can generalize the representation of the true impulse response parameter in (5) to the following

$$\bar{\theta} = [\underbrace{0,\ \ldots,\ 0}_{\bar{d}\ zeros},\ \underbrace{\bar{\theta}(\bar{d}),\ \ldots,\ \theta(\bar{m}-1)}_{\bar{\theta}_{\bar{d},\bar{m}}},\ \underbrace{0,\ \ldots,\ 0}_{M-\bar{m}\ zeros}] \tag{9}$$

Note that $\bar{m}$ can be finite and less than M or can be the same as M, especially in the cases where the true length of the system is larger than M or even infinity.

Here, the Topelitz matrix generated by the input is:

$$\begin{aligned} A_{0,N} &= \begin{bmatrix} u(0) & 0 & \ldots & 0 \\ u(1) & u(0) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u(N-1) & u(N-2) & \ldots & u(0) \end{bmatrix} \\ &= [A_{0,d}\ \ A_{d,m}\ \ A_{m,M}\ \ A_{M,N}] \end{aligned} \tag{10}$$

where $A_{a,b}$ is a matrix with columns $a+1$st to the bth column of the Toeplitz matrix. Based on (9) and (10), the noise-free output of the system (1) is:

$$\begin{aligned} \bar{y}^N &= [A_{0,\bar{d}}\ \ A_{\bar{d},\bar{m}}\ \ A_{\bar{m},M}] \begin{bmatrix} 0_{\bar{d}\times 1} \\ \bar{\theta}_{\bar{d},\bar{m}} \\ 0_{(M-\bar{m})\times 1} \end{bmatrix} \qquad (11) \\ &= A_{0,\bar{d}} 0_{\bar{d}\times 1} + A_{\bar{d},\bar{m}} \bar{\theta}_{\bar{d},\bar{m}} + A_{\bar{m},M} 0_{(M-\bar{m})\times 1} \\ &= 0 \ + \ \underbrace{A_{\bar{d},\bar{m}} \bar{\theta}_{\bar{d},\bar{m}}}_{delay\ \bar{d}\ and\ length\ \bar{m}} \ + \ 0 \qquad (12) \end{aligned}$$

while the observed noisy data (3) is:

$$\begin{aligned} y^N &= [A_{0,d}\ \ A_{d,m}\ \ A_{m,M}] \begin{bmatrix} \Delta_{0,d} \\ \theta_{d,m} \\ \Delta_{m,M} \end{bmatrix} + \omega^N \qquad (13) \\ &= A_{0,d}\Delta_{0,d} + \underbrace{A_{d,m}\bar{\theta}_{d,m}}_{delay\ d\ and\ length\ m} + A_{m,M}\Delta_{m,M} + \omega^N \end{aligned}$$

$$(14)$$

The possible unmodeled coefficients of the true parameter are denoted by Δ as follows

$$\Delta_{a,b} = \bar{\theta}_{a,b} \tag{15}$$

Impulse response parameters with nonzero values starting at delay d and with the length of m can generate outputs with the following structure

$$y^N_{d,m} = [A_{0,d} \;\; A_{d,m} \;\; A_{m,M}] \begin{bmatrix} 0 \\ \theta_{d,m} \\ 0 \end{bmatrix} + \omega^N \tag{16}$$

$$= A_{0,d} \times 0_{d\times 1} + \underbrace{A_{d,m}\theta_{d,m}}_{\text{delay } d \text{ and length } m} + A_{m,M} \times 0_{(M-m)\times 1} + \omega^N \tag{17}$$

The MSE estimate of θ using the available data with the above structure is as follows

$$\hat{\theta}_{d,m} = \arg\min_{\theta_{d,m}} ||y^N - A_{d,m}\theta||_2^2 = (A^T_{d,m}A_{d,m})^{-1}A^T_{d,m}y^N \tag{18}$$

Using the structure in (8) this estimate can equivalently be represented as

$$\hat{\theta} = [\underbrace{0, \ldots, 0}_{d}, \hat{\theta}_{d,m}, \underbrace{0, \ldots, 0}_{M-m}] \tag{19}$$

The estimate of the observed output, according to this estimated IRC is:

$$\hat{y}^N_{d,m} = A_{d,m}\hat{\theta}_{d,m} \tag{20}$$

C. SUMMARY OF NOTATIONS AND THE FOLLOWING DESIRED QUESTION

y^N, u^N in (4): The input and output with a fixed length N.

θ in (8), $\bar{\theta}$ in (9): While θ represents the impulse response parameters if there is a bar over the θ, $\bar{\theta}$ is the true unavailable parameter.

$\hat{\theta}$ in (19), $\hat{\theta}_{d,m}$ in (18): If there is a hat over θ, $\hat{\theta}$ is an estimate of the impulse response. In addition $\theta_{d,m}$ are values of the θ from $d+1$ to m itself with a length of $m-d$. The MSE estimate of θ for each delay d and with total length of m is $\hat{\theta}_{d,m}$.

$\hat{y}^N_{d,m}$ in (20), $\bar{y}^N$ in (12): The parameter estimate $\hat{\theta}_{d,m}$ consequently produces an estimate of the output that is denoted by $\hat{y}^N_{d,m}$. While $\bar{y}^N$ is the noise-free output of the system, $\hat{y}^N_{d,m}$ is the estimate of this value using the available noisy data.

The goal is to compare $\hat{\theta}_{d,m}$s' estimates of the system, for a range of delay d and a range of impulse response length m, and choose the one that optimally represents the unknown system. In the following section the proposed relative entropy criterion is calculated for the purpose of this comparison.

IV. RELATIVE ENTROPY (RE)-BASED LTI MODELING

The proposed method in this section determines the optimum impulse response estimate based on relative entropy criterion. The true unknown impulse response has an unknown delay of $\bar{d}$ and an unknown length of $\bar{m}$ (can be infinity). From (1) and (3) the probability distribution function of the observed output given the true parameter $\bar{\theta}$ and the input u^N is:

$$f_y(y^N; \bar{\theta}, u^N) = \frac{1}{(\sqrt{2\pi\sigma_\omega^2})^N} e^{-\frac{||y^N - \bar{y}^N||_2^2}{2\sigma_\omega^2}} \tag{21}$$

which is a Gaussian distribution with mean $\bar{y}$ in (3). On the other hand, the probability distribution function of the output given the estimated parameter $\hat{\theta}_{d,m}$, with delay d and length m (18), and input u^N is:

$$g_y(y^N; \hat{\theta}_{d,m}, u^N) = \frac{1}{(\sqrt{2\pi\sigma_\omega^2})^N} e^{-\frac{||y^N - \hat{y}^N_{d,m}||_2^2}{2\sigma_\omega^2}} \tag{22}$$

where $\hat{y}^N_{d,m}$ is generated by the estimated $\hat{\theta}_{d,m}$ in (20). This is a Gaussian distribution with mean $\hat{y}^N_{d,m}$. It is known that the relative entropy between two multi-variate normal distributions $f(y^N)$, with mean μ_1 and covariance matrix Σ_1, and $g(y^N)$, with mean μ_2 and covariance matrix Σ_2, is [39], [40]:

$$D(f||g) = \frac{1}{2}(\log\frac{|\Sigma_2|}{|\Sigma_1|} - N + \text{tr}\{\Sigma_2^{-1}\Sigma_1\} + (\mu_2 - \mu_1)^T\Sigma_2^{-1}(\mu_2 - \mu_1)) \tag{23}$$

The relative entropy between the true distribution of the output in (21) and the estimated distribution in (22) (equivalently denoted by $g_{d,m}$) based on (23) is:

$$D(f||g_{d,m}) = \frac{1}{2}(N\frac{z_{d,m}}{\sigma_\omega^2}) \tag{24}$$

where $z_{d,m}$ is the distance between the true *unavailable* noise-free output $\bar{y}^N$ and the estimated output $\hat{y}_{d,m}$ and is denoted as the **Reconstruction Error**.

$$z_{d,m} = \frac{1}{N}||\bar{y}^N - \hat{y}^N_{d,m}||_2^2 \tag{25}$$

The aim is to compare the relative entropy of the impulse response estimates of different delay and length and to choose the optimum estimate $\hat{\theta}_{d^*,m^*}$ with the optimum delay d^* and the optimum length m^* that minimizes this criterion.

As (24) shows, minimizing the relative entropy is equivalent to minimizing the reconstruction error $z_{d,m}$. Although the true noise-free output is not available, in the following section probabilistic bounds on the reconstruction error are provided by using the connection between the mean and variance of this random variable and the available **output error**, denoted as follows

$$x_{d,m} = \frac{1}{N}||y^N - \hat{y}^N_{d,m}||_2^2 \tag{26}$$

For each m and d a sample of this random variable is available. This sample will help provide probabilistic bounds on $z_{d,m}$.

A. PROBABILISTIC ESTIMATION OF RECONSTRUCTION ERROR

Mean and variance of the output error and mean and variance of the reconstruction error are provided in the following Lemma.

Lemma 1: The output error, $x_{d,m}$, defined in (26), is a sample of Chi-square random variable, $X_{d,m}$, and the reconstruction error $z_{d,m}$, defined in (25), is a sample of Chi-squared random variable $Z_{d,m}$ with the following expectations and variances

$$E(X_{d,m}) = (1 - \frac{m-d}{N})\sigma_\omega^2 + \Delta_{d,m} \tag{27}$$

$$var(X_{d,m}) = \frac{2}{N}(1 - \frac{m-d}{N})(\sigma_\omega^2)^2 + \frac{4\sigma_\omega^2}{N}\Delta_{d,m} \tag{28}$$

$$E(Z_{d,m}) = \frac{m-d}{N}\sigma_\omega^2 + \Delta_{d,m} \tag{29}$$

$$var(Z_{d,m}) = \frac{2(m-d)}{N^2}(\sigma_\omega^2)^2 \tag{30}$$

where σ_ω^2 is the additive noise variance in (3) and $\Delta_{d,m} = \frac{1}{N}||G_{d,m}F_{d,m}||_2^2$, with $G_{d,m}$ and $F_{d,m}$ defined as follows.

$$G_{d,m} = I - A_{d,m}(A_{d,m}^T A_{d,m})^{-1}A_{d,m}^T \tag{31}$$

$$F_{d,m} = \begin{bmatrix} A_{0,d} & A_{m,M} \end{bmatrix} \begin{bmatrix} \Delta_{0,d} \\ \Delta_{m,M} \end{bmatrix} \tag{32}$$

Proof: In Appendix A.

Consequently, for each m and d, calculating the expected value and the mean of the reconstruction error requires knowledge of the noise variance and the additional term $\Delta_{d,m} = \frac{1}{N}||G_{d,m}F_{d,m}||_2^2$. Note that $\Delta_{d,m}$ can be zero or non-zero due to the possible unmodeled dynamics in the subspace with delay d and length m. Using the available sample of the output error, the following theorem provides probabilistic worst-case bounds on the reconstruction error.

Theorem 1: Using the available calculated output error $x_{d,m}$, in (26), the upper bound and the lower bound of the reconstruction error, in (25)), with confidence probability $Q(\beta)$ and validation probability $Q(\alpha)$ are:

$$\overline{z_{d,m}} = U_{d,m} + \frac{m-d}{N}\sigma_\omega^2 + \beta\frac{\sqrt{2(m-d)}\sigma_\omega^2}{N} \tag{33}$$

$$\underline{z_{d,m}} = \max\{0, L_{d,m} + \frac{m-d}{N}\sigma_\omega^2 - \beta\frac{\sqrt{2(m-d)}\sigma_\omega^2}{N}\} \tag{34}$$

where

$$U_{d,m} = x_{d,m} - c_{d,m} + \frac{2\alpha^2\sigma_\omega^2}{N} + \kappa_{d,m}(\alpha) \tag{35}$$

$$L_{d,m} = x_{d,m} - c_{d,m} + \frac{2\alpha^2\sigma_\omega^2}{N} - \kappa_{d,m}(\alpha) \tag{36}$$

and

$$\kappa_{d,m}(\alpha) = 2\alpha\frac{\sigma_\omega}{\sqrt{N}}\sqrt{\frac{\alpha^2\sigma_\omega^2}{N} + x_{d,m} - \frac{1}{2}c_{d,m}} \tag{37}$$

$$c_{d,m} = (1 - \frac{m-d}{N})\sigma_\omega^2. \tag{38}$$

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

Proof: In Appendix B.

The upper bound of the reconstruction error is the worst-case probabilistic upper bound that can provide the optimum value of the delay and length with respect to the relative entropy criterion in (24)

$$(d^*, m^*) = \arg\min_{d,m}\overline{D(f||g_{d,m})} = \arg\min_{d,m}\frac{1}{2}(N\frac{\overline{z_{d,m}}}{\sigma_\omega^2}) \tag{39}$$

where $\overline{D(f||g_{d,m})}$ is the calculated probabilistic worst case of the desired relative entropy. In the above calculation, the noise variance is assumed to be known. In the following section, the relative entropy is optimized for the case of unknown noise variance.

B. RELATIVE ENTROPY CRITERION WITH UNKNOWN NOISE VARIANCE

Unknown noise variance σ_ω^2 in the desired RE criterion in (24) can be estimated along with the unknown time delay $\bar{d}$ and impulse response length $\bar{m}$. A range of possible noise variance values between σ_{min} and σ_{max}, is considered $\sigma \in [\sigma_{min}, \cdots, \sigma_{max}]$. In this case, the true probability distribution of the observed data in (21) can be one of the following distributions

$$f(y^N; \bar{\theta}, u^N, \sigma) = \frac{1}{(\sqrt{2\pi\sigma^2})^N}e^{-\frac{||y^N - \bar{y}^N||_2^2}{2\sigma^2}} \tag{40}$$

while the distribution of the data based on the estimated output is:

$$g(y^N; \hat{\theta}_{d,m}, u^N, \sigma) = \frac{1}{(\sqrt{2\pi\sigma^2})^N}e^{-\frac{||y^N - \hat{y}_{d,m}^N||_2^2}{2\sigma^2}} \tag{41}$$

The upperbound of the reconstruction error in (33) can be calculated for each of these noise variances denoted by $\overline{z_{d,m,\sigma}}$. Therefore, the relative entropy associated to these σs can be calculated as follows (24).

$$\overline{D(f(\sigma)||g_{d,m}(\sigma))} = \frac{1}{2}(N\frac{\overline{z_{d,m,\sigma}}}{\sigma^2}) \tag{42}$$

Minimizing the relative entropy estimate between $f(\sigma)$ and $g_{d,m}(\sigma)$ determines the optimum values d^*, m^* and the optimum noise variance σ_ω^{*2}

$$(d^*, m^*, \sigma_\omega^{*2}) = \arg\min_{d,m,\sigma}\overline{D(f(\sigma)||g_{d,m}(\sigma))} \tag{43}$$

Algorithm 1 shows the complete pseudo code of the proposed RE-based impulse response estimation method. Note that the calculation of each of the IR estimates $\hat{\theta}_{d,m}^N$ is of order $O(N^3)$ due to the use of the Toeplitz matrix and the use of inverse matrices. On the other hand, the process of RE minimization will not add any computational complexities of the order of the data length as the range of possible delay and possible impulse response length are chosen as finite values.

Algorithm 1 RE Based IR Estimation With Optimum Delay d^* and Length m^* and Noise Variance σ_ω^{*2}

Input: Input and output data $x^N = [x_1, x_2, \ldots, x_N]$ and $y^N = [y_1, y_2, \ldots, y_N]$, range for noise variance $\sigma \in [\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_K]$, validation parameter α and confidence parameter β, maximum length of the IRCs M

Output: Estimated time delay d^*, length of the IRCs m^* and estimated noise variance σ_ω^*

for $i = 1; i \leq K; i_{++}$ **do**
  $\sigma = \sigma_i$
  for $(m = 1; m \leq M; m_{++})$ **do**
    for $(d = 1; d \leq m; m_{++})$ **do**
      Estimate the IRCs $\hat{\theta}_{d,m}$ based on (18)
      Calculate the output error, $x_{d,m}$ from (26)
      Calculate the upper bound of reconstruction error $\overline{z_{d,m,\sigma}}$ from (34) to (35)
      if $\overline{z_{d,m,\sigma}} \in \mathbb{R}^+$ **then**
        continue the algorithm
      else
        go to next value of σ
      end if
    end for
  end for
end for
if All the calculated $\overline{z_{d,m}} \in \mathbb{R}^+$ **then**
  Estimate the optimum time delay d^* and length of IRCs m^* and noise variance σ_ω^{*2} (43):
  $(d^*, m^*, \sigma_\omega^{*2}) = \arg\min_{d,m,\sigma} \overline{D_{d,m,\sigma}(f(\sigma)||g_{d,m}(\sigma))}$
end if

V. ONLINE RE BASED IMPULSE RESPONSE (IR) ESTIMATION WITH OPTIMUM DELAY AND OPTIMUM LENGTH

In some practical applications, the received observations are sequential. Each time a new data point arrives and the data set is updated. This additional information should then be used to recompute the parameter estimates of the system and update the existing parameter estimates with respect to this new received data point. Recomputing the impulse response estimate from scratch is costly and requires exponential computing time and memory. Therefore, for online practical applications, it is desirable to update the existing parameter estimate recursively. Let us denote the least square estimate of the impulse response of the system after receiving N observations by $\hat{\theta}^N$. The updated impulse response estimate after receiving the $(N+1)^{th}$ observation (19) is denoted by $\hat{\theta}_{d,m}^{N+1}$.

$$\hat{\theta}_{d,m}^N \rightarrow \hat{\theta}_{d,m}^{N+1} \tag{44}$$

and consequently the output error (26) as a function of N is updated to:

$$x_{d,m}^{N+1} = \frac{1}{N+1}||y^{N+1} - \hat{y}_{d,m}^{N+1}||_2^2 \tag{45}$$

which will lead to an update of the upperbound on the reconstruction error that itself leads to updates of the estimated delay and estimated impulse response length.

$$d_N^*, m_N^* \rightarrow d_{N+1}^*, m_{N+1}^* \tag{46}$$

The following subsection provides details of the online recursive procedure for the proposed RE optimization method. In addition, note that for practical online applications, a stopping criterion is essential and valuable. Next, we propose an stopping criterion for the procedure as N grows. The stopping criterion is beneficial and crucial in practical applications and will also help expand the application of the RE based method for modeling systems with coefficients that are gradually changing with the time.

A. UPDATING PROCEDURE

The parameter estimate (18) using the data length of $N+1$ is:

$$\begin{aligned}\hat{\theta}_{d,m}^{N+1} &= \arg\min_{\theta \in d,m} ||y^{N+1} - A_{d,m}^{N+1}\theta||_2^2 \\ &= ((A_{d,m}^{N+1})^T A_{d,m}^{N+1})^{-1}(A_{d,m}^{N+1})^T y^{N+1}\end{aligned} \tag{47}$$

where $A_{d,m}^N$ is the columns of the Toeplitz matrix in (10) from the index $d+1$ to the index m. If the data length increases by one, a new column and a new row are added to the matrix in (10) and consequently with the $N+1$th data we have:

$$A_{d,m}^{N+1} = \begin{bmatrix} A_{d,m}^N \\ (B_{d,m}^N)^T \end{bmatrix} \tag{48}$$

where

$$(B_{d,m}^N)^T = [u((N+1)-(d+1)), \cdots, u((N+1)-m)] \tag{49}$$

The following Lemma uses the available $\hat{\theta}_{d,m}^N$ and $u(N+1)$ and $y(N+1)$ to provide $\hat{\theta}_{d,m}^{N+1}$.

Lemma 3: Recursive method of updating $\hat{\theta}_{d,m}^N \rightarrow \hat{\theta}_{d,m}^{N+1}$ is as follows.

$$\begin{aligned}\hat{\theta}_{d,m}^{N+1} &= (K_{d,m}^N)^{-1} - \frac{1}{1+tr(\gamma_{d,m}^N)} C_{d,m}^N (C_{d,m}^N)^T \\ &\quad \times ((A_{d,m}^N)^T y^N + B_{d,m} y(N+1)) \\ &= (I - \frac{1}{1+tr(\gamma_{d,m}^N)}\gamma_{d,m}^N)\hat{\theta}_{d,m}^N \\ &\quad + (I + \gamma_{d,m}^N) C_{d,m}^N y(N+1)\end{aligned}$$

where

$$C_{d,m}^N = (K_{d,m}^N)^{-1} B_{d,m}^N, \quad \gamma_{d,m}^N = (K_{d,m}^N)^{-1} B_{d,m}^N (B_{d,m}^N)^T \tag{50}$$

and tr(a) is the trace of matrix a.

As a result of this recursive calculation, the complexity order of non recursive calculation of $\hat{\theta}_{d,m}^{N+1}$ which is of order $O(N^3)$ will be reduced to $O(N^2)$.

Proof: In Appendix C.

This parameter update generates an updated data error which can be used in calculation of the reconstruction error

upperbound in (34) to provide the updated delay and length of the parameter estimate based on (24)

$$(d^*_{N+1}, m^*_{N+1}) = \arg\min_{d,m} \overline{D_{N+1}(f||g_{d,m})} \tag{51}$$

and chooses the optimum parameter estimates.

$$(\hat{\theta}^*)^{(N+1)} = \hat{\theta}^{N+1}_{d^*_{N+1}, m^*_{N+1}} \tag{52}$$

B. STOPPING CRITERION

For the stopping criterion, the following desired signal to noise (SNR) ratio is logical:

$$\frac{\text{Best Estimate of Output}}{\text{Output Estimation Error}} = \frac{||\hat{y}^N_{d^*,m^*}||_2^2}{||\bar{y}^N - \hat{y}^N_{d^*,m^*}||_2^2} = \frac{\frac{1}{N}||\hat{y}^N_{d,m}||_2^2}{z^N_{d^*,m^*}} \tag{53}$$

The desired SNR in dB is in the following form, and the stopping criterion can be defined based on a desired lower bound on this SNR, with a chosen ϵ:

$$\text{SNR} = 10\log(\frac{\frac{1}{N}||\hat{y}^N_{d,m}||_2^2}{z^N_{d^*,m^*}}), \quad \text{SNR} > 10\log(\frac{1}{\epsilon}) \tag{54}$$

To utilize such stopping criterion, we can use $\overline{z_{d^*,m^*}}$ in (34) which is our probabilistic worst-case estimate of the unavailable z_{d^*,m^*}. Consequently, the desired stopping criterion chooses the smallest value of N such that the following is satisfied.

$$\frac{\overline{z^N_{d^*,m^*}}}{\frac{1}{N}||\hat{y}^N_{d,m}||_2^2} < \epsilon \tag{55}$$

VI. SIMULATIONS

The performance of the Relative Entropy (RE)-based system modeling is analyzed and compared with other approaches in two scenarios for Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) modeling. Consider the following two time-delayed systems:

System I: A lowpass FIR filter with 20 kHz pass band-edge frequency, 96 kHz sampling frequency, 0.01 dB peak-to-peak ripple and 80 dB stop band attenuation with time delay $\bar{d} = 7$ and length of $\bar{m} = 69$.[1]

System II: Non-minimum phase IIR system with the following impulse response [26],

$$\bar{\theta}(n) = 0.2545(0.9094)^n - 0.3316(0.8146)^n, \quad n \geq 0 \tag{56}$$

time delayed by $\bar{d} = 11$.

Both system's inputs are independent identically distributed (IID) Bernoulli sequences of ± 1 with data length of $N = 1000$.

[1] https://www.mathworks.com/help/dsp/ref/dsp.lowpassfilter-system-object.html

A. TIME DELAY ESTIMATION ANALYSIS

For time delay estimation, the performance of the RE-based method is compared with three classical methods, CUSUM [26], Frequency-domain and MATLAB *delayest*. Delayest method requires the order (number of poles) of the system, and this number is by default set to two. In addition, most classical time-delay estimators are based on thresholding. CUSUM is one of the most used thresholding approaches for time delay estimation, and its time delay estimation is highly sensitive to the user selected parameters. For CUSUM approach we use the threshold parameters suggested in [26]. However, this method is based on choosing a maximum value in the frequency domain of the estimated impulse response. This value is highly affected by the noise level even as large as 20dB in the case of System I. It is worth mentioning that what is shown here as the Frequency-domain method is our improved version compared to its conventional version [26]. In the improved version, the threshold is further optimized by trial and error. For the RE-based method, parameters α and β in confidence and validation probabilities are set to 4 as discussed in [1], and therefore, these probabilities are approximately 0.999. Note that the method is robust to changing these values within a wide sufficient and necessary range that is a function of the data length [1].

Figure 1 shows the first 90 coefficients of the two impulse responses in the form of $\bar{\theta}$, (blue signal), as well as the results of the RE-based approach for the SNR of 15dB. The black signals in the figure show the first 90 coefficients of the least squares coefficients estimates of these impulse responses from zero to 1000, $\hat{\theta}_{0,N}$, As the data length is 1000, this estimate has 1000 coefficients, and as the figure shows, these are noisy estimates of the impulse responses. The red signal is $\hat{\theta}_{d^*,m^*}$, which is the RE based impulse response estimate. As the figure shows while $\hat{\theta}_{0,N}$ fits the additive noise, the optimum d^* and m^* for the FIR system are $d^* = 7$ and $m^* = 69$ which are the same as the true delay and length ($\bar{d} = 7, \bar{m} = 69$). For the IIR system, the delay is estimated as $d^* = 11$ which is the true unknown time delay. Note that although in this case the impulse response is infinite, ($\bar{m} = \infty$), the method recognizes that an estimate with length 78 ($m^* = 78$) is the best estimate in this SNR and sets the rest of coefficients to zero to minimize the desired least square of RE. Figure 2 shows the average time delay estimate in 100 trials as a function of SNR in dB, where SNR is the ratio of the power of the unavailable noiseless data $\bar{y}$ and the noise variance σ^2_ω in (3). Table 1 shows these estimated time delay values as well as their associated standard deviation (SD) in the same 100 trials. As Figure 2 and Table 1 show, the RE-based algorithm outperforms the other three methods. For the FIR case, the time delay estimate is larger than the true time delay for low SNRs. This is expected, as the small values of the first coefficients are comparable with the noise SD for the low SNRs. At around SNR of 15 dB, the first coefficient is comparable with the noise standard deviation, and therefore, the method starts choosing the delay as 7 which is the true delay, and the standard deviation of

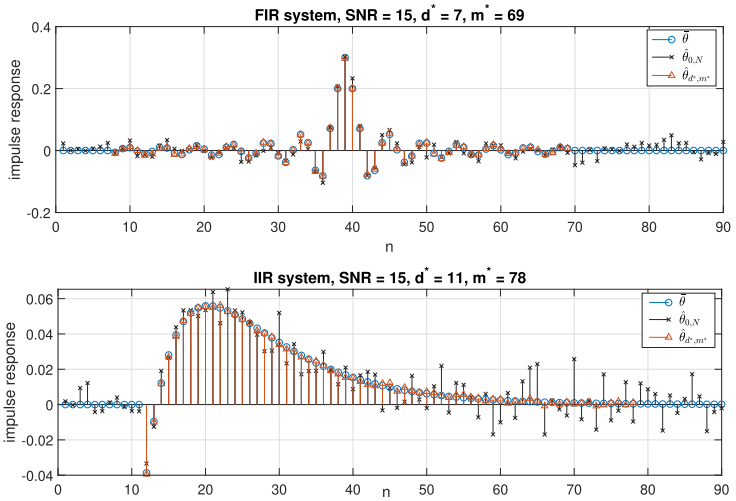

FIGURE 1. True impulse response coefficients $\bar{\theta}$, Estimated minimum MSE impulse response of length 1000, $\hat{\theta}_{0,N}$, Estimated RE based impulse response $\hat{\theta}_{d^*,m^*}$.

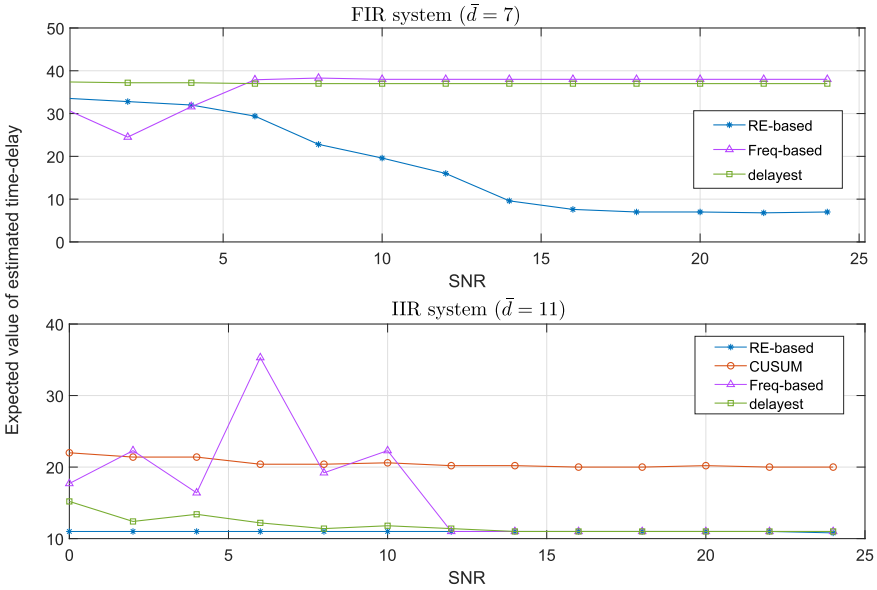

FIGURE 2. Estimated time delay with data length of $N = 1000$ for 0 ≤SNR≤ 24dB (averaged over 100 trials).

error goes to zero as SNR grows. As the figure shows and the table confirms, the RE based method is the only one that estimates the correct time delay as the SNRs grows. For the IIR system on the other hand, as shown in Figure 1, the first coefficient of the impulse response has a large absolute value, so we expect that an efficient delay estimator chooses the correct time delay even for the low SNRs. As the table and the figure depict, the RE-based method chooses the correct delay and outperforms the other methods. While delayest is the next method that converges for higher SNRs, the Freq-based method also converges for an even higher range of SNRs.

Figure 3 shows the Root Mean Square Error (RMSE) in the time delay estimation for the two systems at SNR of 10 dB and as the data length grows from 100 to 1000. In each case, the delay is randomly generated between 1 and 20 (with a uniform distribution), and the RMSE is averaged over 100 runs. As the figure illustrates, RE based method outperforms the other approaches as its error goes to zero as the data length gets around 600 for System I and around 300 for System II. While delayest has the next acceptable performance, for none of these methods the RMSE approaches zero in this range of data length.

B. IMPULSE RESPONSE LENGTH ESTIMATION ANALYSIS

Estimated coefficients with length m for a range of SNRs with the RE-based method are provided in Figure 4 for the FIR system ($\bar{m} = 62$). The method is compared with well-known and most used order selection methods AIC and BIC. Note that it is shown in [1] that AIC and BIC (equivalently two stage MDL) model order selection methods are special cases of the reconstruction error based approach. As the existing

TABLE 1. Mean and standard deviation (SD) of the estimated time delay for 100 trials.

	FIR System I , ($\bar{d} = 7$)								IIR System, ($\bar{d} = 11$)							
SNR	RE Based		CUSUM		Freq-based		Delayest		RE based		CUSUM		Freq-based		Delayest	
(dB)	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
0	**29.81**	**0.72**	80.64	0.93	35.53	2.43	37.58	2.45	**11.0**	**0.0**	15.43	0.93	22.68	12.43	14.67	2.45
2	**28.72**	**0.69**	80.29	0.71	29.10	2.57	37.53	1.61	**11.0**	**0.0**	15.06	0.71	24.67	12.57	13.15	1.61
4	**23.14**	**0.71**	80.16	0.58	32.47	1.82	37.43	1.06	**11.0**	**0.0**	14.73	0.58	21.99	11.82	12.74	1.06
6	**21.52**	**0.63**	80.14	0.54	36.44	1.17	37.39	1.04	**11.0**	**0.0**	14.7	0.54	18.21	11.17	12.55	1.04
8	**18.78**	**0.57**	80.01	0.49	36.99	0.89	37.21	0.94	**11.0**	**0.0**	14.61	0.49	17.49	9.89	12.53	0.94
10	**11.77**	**0.53**	80.0	0.49	37.17	0.91	37.11	1.00	**11.0**	**0.0**	14.39	0.49	16.55	8.91	11.94	1.00
12	**7.34**	**0.41**	80.0	0.46	37.93	0.29	37.10	0.86	**11.0**	**0.0**	14.29	0.47	12.62	5.29	11.48	0.86
14	**7.2**	**0.37**	80.0	0.46	37	0.13	37	0.13	**11.0**	**0.0**	14.3	0.46	11.53	2.54	**11.0**	0.0
16	**7.1**	**0.18**	80.0	0.39	37.85	0.10	37.08	0.06	**11.0**	**0.0**	14.18	0.39	11.24	1.48	**11.0**	0.0
18	**7.0**	**0.0**	80.0	0.33	37	0.05	37.04	0.04	**11.0**	**0.0**	14.12	0.32	11.01	0.05	**11.0**	0.0
20	**7.0**	**0.0**	80.0	0.24	37.48	0.23	37	0.0	**11.0**	**0.0**	14.06	0.24	**11.0**	0.0	**11.0**	0.0
22	**7.0**	**0.0**	80.0	0.1	38	0.01	37	0.0	**11.0**	**0.0**	14.01	0.1	**11.0**	0.0	**11.0**	0.0
24	**7.0**	**0.0**	80.0	0.0	38	0.0	37	0.0	**11.0**	**0.0**	14.0	0.0	**11.0**	0.0	**11.0**	0.0

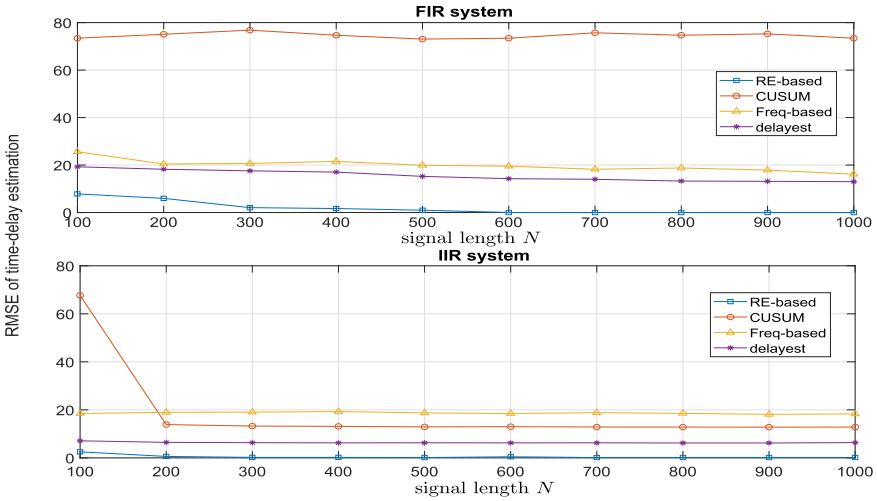

FIGURE 3. Delay RMSE for the FIR system and the IIR system as the data length N grows, averaged over 100 runs, for randomly generated delays between 1 and 20, for SNR=10db.

order selection methods cannot estimate the time delay, we set the time delay to zero for the purpose of comparison. Figure 4 shows the simulation results. As the figure shows, for the RE based method, as the SNR grows from zero dB, the chosen impulse response length is growing from 46 towards the true length of 62 which is chosen after 15 dB. As expected, the figure shows that the AIC method overestimates, while BIC method under estimates the impulse response length. Figure 5 shows the RMSE of the estimated impulse response for both FIR and IIR systems in the order selection setting and as the SNR grows. As the figure indicates, for the FIR system, not only the RMSE of the RE based is minimum of the all methods, but also it approaches zero with correct order selection for SNR $>$ 15 dB. While the AIC error seems to be nonzero even for SNR of 30dB, the BIC approach converges at this SNR which is still much higher than 15dB, that is the convergence point of the proposed RE based method. Note that this simple example shows the important role of the SNR in the choice of coefficient lengths and confirms that while the true length of the FIR is 62, it is more efficient to choose fewer coefficients for lower SNRs and not to fit the additive noise. In other words, what is known as optimum length *estimation* for the impulse response is less important than the optimum length *selection*. As the figure shows, RMSE of order selection for the IIR system is also minimum for the RE based method.

C. SIMULTANEOUS TIME DELAY ESTIMATION AND IMPULSE RESPONSE LENGTH ESTIMATION ANALYSIS

In previous sections we compared the RE based method with existing well-known time delay estimators and order selection methods. To our knowledge, the proposed RE based method is the only approach that is capable of simultaneous estimation of both the time delay and the impulse response length. Table 2 shows the estimated time delay and impulse response length, as well as the respective RMSE between the true coefficients, $\bar{\theta}$, and the estimated coefficients, $\hat{\theta}_{d^*,m^*}$, when the additive noise variance is unknown. The table shows

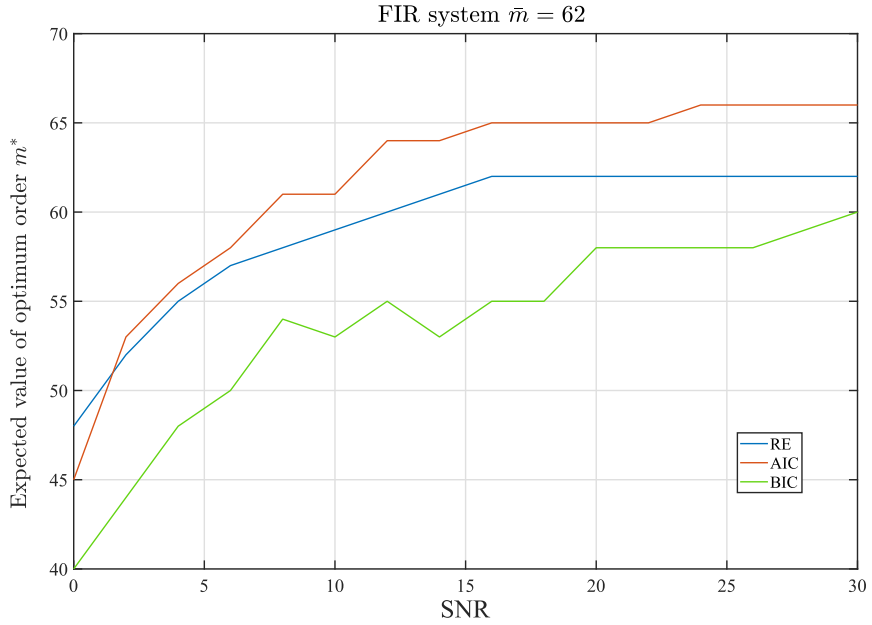

FIGURE 4. Estimated Impulse Response length with observed data length of $N = 1000$ for $0 \leq SNR \leq 30$dB (averaged over 100 trials).

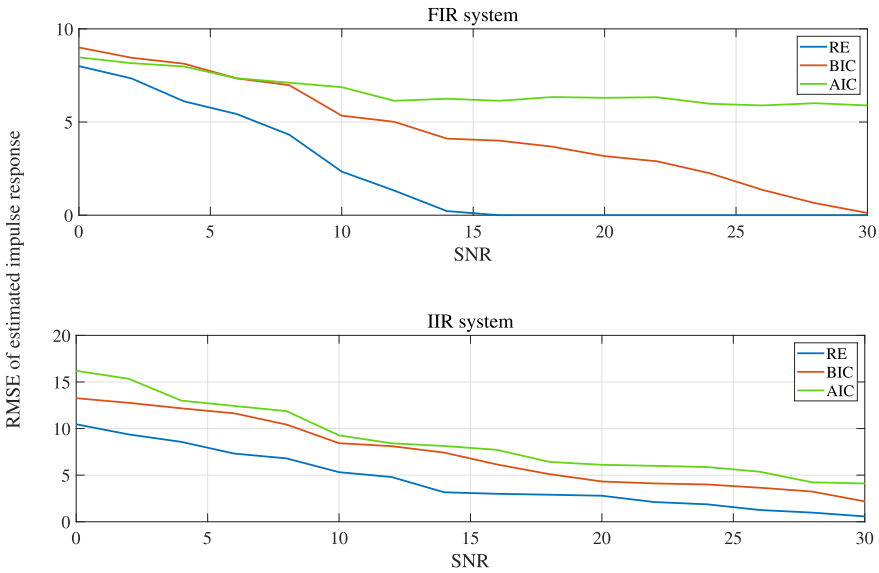

FIGURE 5. RMSE of the estimated impulse response, $N = 1000$ for $0 \leq SNR \leq 30$dB for FIR and IIR systems averaged over 100 trials.

the results averaged over 100 trials for a range of SNR. As the table indicates for the FIR system, as the SNR grows, the estimated time delay and impulse response length converge to the true values. For the IIR system, the optimum time delay is estimated correctly for this range of SNR while the estimated optimal impulse response length increases as the SNR grows. This is a rational expectation as the higher the SNR the more valuable are the tails of the least squares estimate of the impulse response, and therefore the method recognizes to choose more of these estimated coefficients. As the table shows, the RMSE in both scenarios become smaller as the SNR grows and is almost zero for SNRs higher than 20dB for the FIR system.

The theory of simultaneous noise variance and system modeling method is explained in Section IV-B. In the following we illustrate the RE based approach with an example and numbers from the IIR system. For simplicity and without loss of generality for this illustration, lets assume that the time delay is zero and therefore the optimal model order is 48 while the unknown true SNR is 10 dB. Figure 6 shows a typical behavior of the calculated relative entropy in (42) for a range of SNRs and a range of impulse response

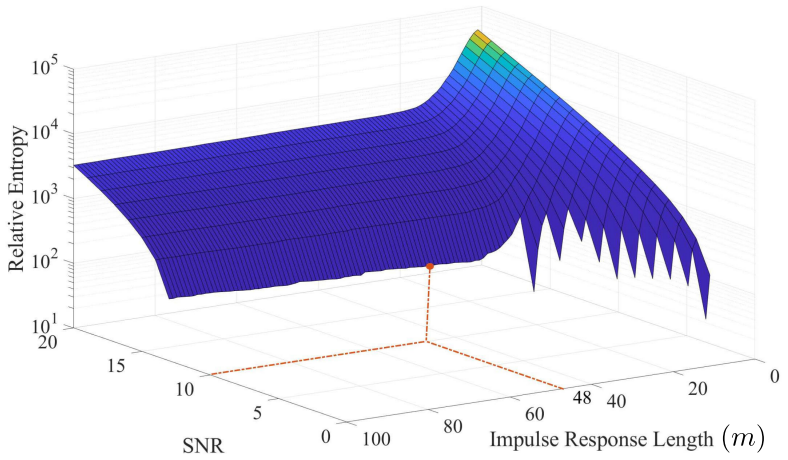

FIGURE 6. Relative entropy in (42) for a range of SNRs form 0 dB to 20 dB and for different impulse response lengths.

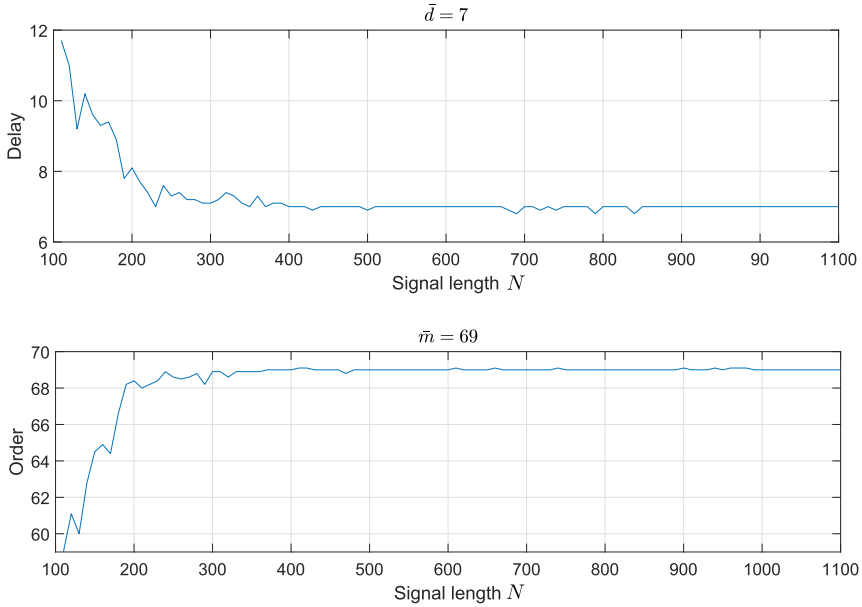

FIGURE 7. Estimated online time delay and impulse response length using the online RE based method for System I, SNR=15dB.

lengths m. The optimum SNR and the optimum m are chosen simultaneously based on the relative entropy minimization. As shown in the figure, the optimum SNR in this example is 10dB which is the true unknown SNR and the optimum m in this example is 48. Note that in the presence of a time delay, this procedure is generalized by adding a range of possible time delays as well for a simultaneous minimization of the calculated RE.

D. ONLINE MODELING AND EFFECTIVE ROLE OF THE STOPPING CRITERION

Figure 7 shows the result of the online time delay and impulse response length estimation for System I when the SNR is 15 dB, averaged over 300 trials. As the figure shows after about 400 data samples (N = 400), the online modeling converges to the true values of delay and impulse response length, which are 7 and 69. It is important to mention that in this scenario the method can be powered by the stopping criterion in practical applications, as Figure 8 illustrates. As indicated in the figures, tdhe stopping criterion in (55) with value $\epsilon = 0.1$ occurs at $N = 80$. This stopping criterion is for when the SNR of the desired acceptable error is set to 10 dB in (54). On the other hand, if the desired acceptable SNR by the user is set higher to the value of 20 dB, then the stopping criterion ϵ is 0.01 and the algorithm automatically halts at $N = 366$. In this scenario, acceptable error SNR

TABLE 2. Estimated time delay, estimated impulse response length, and the corresponding RMSE for the FIR and IIR systems for a range of SNRs (averaged over 100 trials).

	FIR system ($\bar{d}=7, \bar{m}=69$)			IIR system ($\bar{d}=11$)		
SNR (dB)	d^*	m^*	RMSE	d^*	m^*	RMSE
0	29.81	54.23	8.13	11	43.26	13.25
2	28.72	57.16	7.26	11	45.18	12.64
4	23.14	62.72	6.07	11	50.31	11.89
6	21.52	63.88	5.84	11	52.68	11.60
8	18.78	64.47	5.21	11	54.45	10.85
10	11.77	66.19	4.23	11	59.73	10.06
12	7.34	66.93	3.63	11	63.54	8.52
14	7.2	68.58	1.23	11	65.73	7.48
16	7.1	69.21	0.41	11	70.20	6.16
18	7	69	0.07	11	70.32	4.82
20	7	69	0.0	11	70.89	3.02
22	7	69	0.0	11	71.43	1.32
24	7	69	0.0	11	71.68	0.86
26	7	69	0.0	11	71.56	0.24

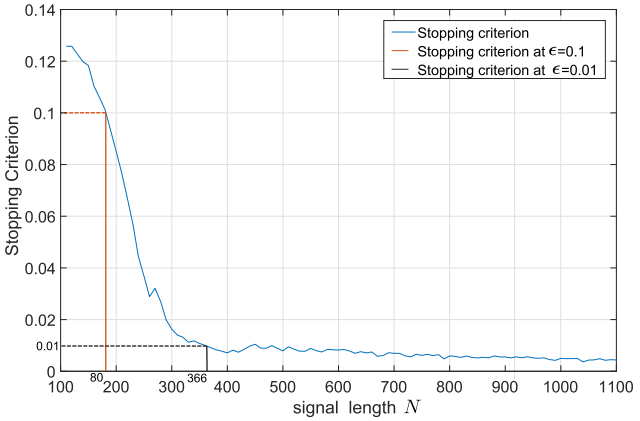

FIGURE 8. Stopping criterion in online modeling of System I as the data length grows.

of 20 dB waits longer and chooses the true time delay and impulse response length shown in Figure 7.

VII. CONCLUSION

A method of impulse response estimation for LTI systems, based on the theory of relative entropy (RE), is proposed. In this RE based approach time delay, impulse response length, and optimally denoised coefficients of the LTI system are estimated simultaneously. Minimizing the estimate of the relative entropy between the estimated models and the true model can also provide the noise variance estimate simultaneously for when the noise variance is also unavailable. Furthermore, the extension of the proposed method for online impulse response estimations has been shown to reduce the computational complexity of the estimation process. The proposed practical and efficient stopping criterion for this online LTI impulse response estimation enables the method to be used in a wide range of applications, including slowly varying LTI systems. Comparison of the proposed method with the existing time delay and order selection approaches illustrates superiority and precision of the method for a wide range of SNRs for finite and infinite impulse responses.

APPENDIX A PROOF OF LEMMA 1

Considering (12) and (20), the reconstruction error in (25) is

$$z_{d,m} = \frac{1}{N}||G_{d,m}F_{d,m} + H_{d,m}\omega^N||_2^2 \tag{57}$$

where w^N is the additive noise vector in (3) and

$$F_{d,m} = \begin{bmatrix} A_{0,d} & A_{m,M} \end{bmatrix} \begin{bmatrix} \Delta_{0,d} \\ \Delta_{m,M} \end{bmatrix} \tag{58}$$

where $\Delta_{a,b}$ is defined in (15) and the matrices

$$G_{d,m} = I - A_{d,m}(A_{d,m}^T A_{d,m})^{-1}A_{d,m}^T \tag{59}$$

$$H_{d,m} = A_{d,m}(A_{d,m}^T A_{d,m})^{-1}A_{d,m}^T \tag{60}$$

are both projection matrices. Since these projection matrices are orthogonal, the inner product of $G_{d,m}F_{d,m}$ and $H_{d,m}w^N$ is zero, and (57) is

$$z_{d,m} = \underbrace{\frac{1}{N}||G_{d,m}F_{d,m}||_2^2}_{\Delta_{d,m}} + \frac{1}{N}||H_{d,m}\omega^N||_2^2 \tag{61}$$

Therefore, while the first term in (61) is $\Delta_{d,m}$, the second term $\frac{1}{N}||H_{d,m}\omega^N||_2^2$ is the sum of $m-d$ zero-mean Gaussian random variables and is a non-central chi-squared random variable. Consequently, the reconstruction error $z_{d,m}$ is a sample of a chi-squared random variable $Z_{d,m}$ of order $m-d$ and the expected value and variance of $Z_{d,m}$ are ([1], [32])

$$E(Z_{d,m}) = \frac{m-d}{N}\sigma_\omega^2 + \frac{1}{N}||G_{d,m}F_{d,m}||_2^2 \tag{62}$$

$$var(Z_{d,m}) = \frac{2(m-d)}{N^2}(\sigma_\omega^2)^2 \tag{63}$$

Taking into account the structure of the output error in (26), it is also the sum of $N-(m-d)$ squares of Gaussian random variables and it can be shown that $x_{d,m}$ is also a sample of Chi-square random variable of order $N-(m-d)$ with the following expectation and variance [1]:

$$E(X_{d,m}) = (1 - \frac{m-d}{N})\sigma_\omega^2 + \frac{1}{N}||G_{d,m}F_{d,m}||_2^2 \tag{64}$$

$$var(X_{d,m}) = \frac{2}{N}(1 - \frac{m-d}{N})(\sigma_\omega^2)^2 + \frac{4\sigma_\omega^2}{N^2}||G_{d,m}F_{d,m}||_2^2 \tag{65}$$

APPENDIX B PROOF OF THEOREM 1

Considering the Chebyshev's Inequality [41], we have

$$Pr\left\{|X_{d,m} - E(X_{d,m})| \le \alpha\sqrt{var(X_{d,m})}\right\} > 1 - \frac{1}{\alpha^2}$$

where the expectation and variance of $X_{d,m}$ are in (64) and (65) and α denotes the validation parameter. The output error can be calculated by using the observed data, and thus one sample of this random variable is available. Using these

values of mean and variance, the Chebyshev's Inequality (66) provides probabilistic bounds on $\Delta_{d,m} = \frac{1}{N}||G_{d,m}F_{d,m}||_2^2$. The resulting inequality to find the upper bound is

$$E(X_{d,m}) - \alpha\sqrt{var(X_{d,m})} \leq x_{d,m} \tag{66}$$

The following variables are defined for simplicity:

$$c_{d,m} = (1 - \frac{m-d}{N})\sigma_\omega^2 \tag{67}$$

$$\upsilon_{d,m} = 2(1 - \frac{m-d}{N})(\frac{\sigma_\omega^4}{N}) \tag{68}$$

where σ_w denotes the standard deviation of the noise. Note that if the validation parameter α is chosen such that $(x_{d,m} - m_w) \leq -\alpha\sqrt{\upsilon_{d,m}}$ then no $\Delta_{d,m}$ term can satisfy the inequality in (66). Therefore, the value of α must be chosen large enough, such that

$$\alpha > \frac{N}{\sqrt{2(N-(m-d))}}\left(\frac{N-(m-d)}{N} - \frac{x_{d,m}}{\sigma_\omega^2}\right) \tag{69}$$

When solving $(x_{d,m} - m_w) > -\alpha\sqrt{\upsilon_{d,m}}$, the upper bound of $\Delta_{d,m}$ is the largest root of the following.

$$\left(\Delta_{d,m} - (x_{d,m} - c_{d,m})\right)^2 = \alpha^2\left(\upsilon_{d,m} + \frac{4m_\omega\Delta_{d,m}}{N-(m-d)}\right) \tag{70}$$

which is

$$U_{d,m} = x_{d,m} - c_{d,m} + \frac{2\alpha^2\sigma_\omega^2}{N} + \kappa_{d,m}(\alpha) \tag{71}$$

where $\kappa_{d,m}(\alpha)$ is defined as

$$\kappa_{d,m}(\alpha) = 2\alpha\frac{\sigma_\omega}{\sqrt{N}}\sqrt{\frac{\alpha^2\sigma_\omega^2}{N} + x_{d,m} - \frac{1}{2}c_{d,m}} \tag{72}$$

To calculate the lower bound for $\Delta_{d,m}$, the following inequality is the result of the Chebyshev inequality:

$$x_{d,m} \leq E(X_{d,m}) + \alpha\sqrt{var(X_{d,m})} \tag{73}$$

which is

$$L_{d,m} = x_{d,m} - c_{d,m} + \frac{2\alpha^2\sigma_\omega^2}{N} - \kappa_{d,m}(\alpha) \tag{74}$$

To find the upper and lower bounds of the reconstruction error, the Chebyshev inequality [41] is used:

$$Pr\left\{|Z_{d,m} - E(Z_{d,m})| \leq \beta\sqrt{var(Z_{d,m})}\right\} > 1 - \frac{1}{\beta^2}$$

where β denotes the confidence parameter. Taking into account the expectation and variance of the random variable $Z_{d,m}$ in (62) and (63) and the lower bound and upper bound calculated for $\Delta_{d,m}$ based on the observed output error in (71) and (74), the upper bound and lower bound for the reconstruction error can be calculated as

$$\overline{z_{d,m}} = 2\frac{m-d}{N}\sigma_\omega^2 + \Delta_{d,m} + \beta var(Z_{d,m}) \tag{75}$$

$$\underline{z_{d,m}} = \max\{0, 2\frac{m-d}{N}\sigma_\omega^2 + \Delta_{d,m} - \beta var(Z_{d,m})\} \tag{76}$$

Using the upper bound of $\Delta_{d,m}$ in (71) in the upper bound of the reconstruction error in (75) and the lower bound in (74) in the lower bound of the reconstruction error in (76) provides the probabilistic worst-case bounds for the reconstruction error.

Note that if the order of the Chi-squared random variable is large enough (usually more than 10 is enough), it can be estimated with a Gaussian distribution. Therefore, when $m-d$ is large enough, the Chi-square distribution of $Z_{d,m}$ and $X_{d,m}$ can be estimated with Gaussian distributions. As a result, the Chebyshev inequality becomes an equality through the law of large numbers, and considering $Q(\alpha) = \int_{-\alpha}^{\alpha}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$ for the Gaussian equality, we have [1], [32]

$$Pr\{|X_{d,m} - E(X_{d,m})| \leq \alpha\sqrt{var(X_{d,m})}\} = Q(\alpha). \tag{77}$$

$$Pr\{|Z_{d,m} - E(Z_{d,m})| \leq \beta\sqrt{var(Z_{d,m})}\} = Q(\beta). \tag{78}$$

It is worth mentioning that the above analysis holds for even when the additive noise has a distribution other than the Gaussian distribution due to the Central Limit Theorem. As a results, in this scenario both reconstruction error and output error are estimated with the above Gaussian distributions and the theorem results still hold.

APPENDIX C PROOF OF LEMMA 3

Matrix A^{N+1} in (48) has two components as follows:

$$A_{d,m}^{N+1} = \begin{bmatrix} A_{d,m}^N \\ (B_{d,m}^N)^T \end{bmatrix} \tag{79}$$

where

$$(B_{d,m}^N)^T = [u((N+1)-(d+1)), \cdots, u((N+1)-m)] \tag{80}$$

This will update the parameter estimate in (47) as follows:

$$\hat{\theta}_{d,m}^{N+1} = (K_{d,m}^{N+1})^{-1}(K_{Sd,m}^N + B_{d,m}B_{d,m}^T)^T y^{N+1} \tag{81}$$

$$= (K_{d,m}^{N+1})^{-1}((A_{d,m}^N)^T y^N + B_{d,m}y(N+1)) \tag{82}$$

where

$$K_{d,m}^N = (A_{d,m}^N)^T A_{d,m}^N \tag{83}$$

and

$$K_{d,m}^{N+1} = K_{d,m}^N + B_{d,m}B_{d,m}^T. \tag{84}$$

On the other hand, the inverse of $K_{d,m}^{N+1}$ is [42]:

$$(K_{d,m}^{N+1})^{-1} = (K_{d,m}^N + B_{d,m}^N(B_{d,m}^N)^T)^{-1} \tag{85}$$

$$= (K_{d,m}^N)^{-1} - \frac{1}{1+tr(\gamma_{d,m}^N)}C_{d,m}^N(C_{d,m}^N)^T) \tag{86}$$

where

$$C_{d,m}^N = (K_{d,m}^N)^{-1}B_{d,m}^N \tag{87}$$

$$\gamma_{d,m}^N = (K_{d,m}^N)^{-1}B_{d,m}^N(B_{d,m}^N)^T \tag{88}$$

and tr(a) is a trace of matrix a. Note that while $C_{d,m}^N$ is a vector of length N, $\gamma_{d,m}^N$ is a $N \times N$ matrix. In addition, it is proven in

[42] that, in this form of inverse calculation, the denominator $1+tr(\gamma_{d,m}^N)$ never becomes zero, and therefore this value can always be updated.

Using the inverse update in (86), (82) is as follows:

$$\begin{aligned}\hat{\theta}_{d,m}^{N+1} &= (K_{d,m}^N)^{-1} - \frac{1}{1+tr(\gamma_{d,m}^N)} C_{d,m}^N (C_{d,m}^N)^T \\ &\quad \times ((A_{d,m}^N)^T y^N + B_{d,m} y(N+1)) \\ &= (I - \frac{1}{1+tr(\gamma_{d,m}^N)} \gamma_{d,m}^N) \hat{\theta}_{d,m}^N \\ &\quad + (I+\gamma_{d,m}^N) C_{d,m}^N y(N+1) \end{aligned} \tag{89}$$

Since $\gamma_{d,m}^N$ is a $N \times N$ matrix (calculated in (87)) and $C_{d,m}^N$ is a vector of length N (calculated in (87)), the complexity order of the recursive calculation of $\hat{\theta}_{d,m}^{N+1}$ with respect to (89) is of order $O(N^2)$.

REFERENCES

[1] S. Beheshti and M. A. Dahleh, "Noisy data and impulse response estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 510–521, Feb. 2010.
[2] S. Amari, *Information Geometry and Its Applications*, vol. 194. Cham, Switzerland: Springer, 2016.
[3] J. Lai and J. J. Ford, "Relative entropy rate based multiple hidden Markov model approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 165–174, Jan. 2010.
[4] N. Zhuo-Yun, Z. Yi-Min, W. Qing-Guo, L. Rui-Juan, and X. Lei-Jun, "Fractional-order PID controller design for time-delay systems based on modified Bode's ideal transfer function," *IEEE Access*, vol. 8, pp. 103500–103510, 2020.
[5] Z. Yan, X. Liu, J. Zhou, and D. Wu, "Coordinated target tracking strategy for multiple unmanned underwater vehicles with time delays," *IEEE Access*, vol. 6, pp. 10348–10357, 2018.
[6] M. Tang, Y. Rong, J. Zhou, and X. R. Li, "Information geometric approach to multisensor estimation fusion," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 279–292, Jan. 2019.
[7] R. Bose, *Information Theory, Coding and Cryptography*. New York, NY, USA: McGraw-Hill, 2008.
[8] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* Cambridge, U.K.: Cambridge Univ. Press, 2003.
[9] A.-K. Seghouane and M. Bekara, "A small sample model selection criterion based on Kullback's symmetric divergence," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3314–3323, Dec. 2004.
[10] W. Wang, B. Zhang, D. Wang, Y. Jiang, S. Qin, and L. Xue, "Anomaly detection based on probability density function with Kullback–Leibler divergence," *Signal Process.*, vol. 126, pp. 12–17, Sep. 2016.
[11] A. Lesne, "Shannon entropy: A rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics," *Math. Struct. Comput. Sci.*, vol. 24, no. 3, Jun. 2014, Art. no. e240311.
[12] O. Techakesari and J. J. Ford, "Relative entropy rate based model selection for linear hybrid system filters of uncertain nonlinear systems," *Signal Process.*, vol. 93, no. 1, pp. 12–22, Jan. 2013.
[13] M. A. Jatoi, N. Kamel, A. S. Malik, and I. Faye, "EEG based brain source localization comparison of sLORETA and eLORETA," *Australas. Phys. Eng. Sci. Med.*, vol. 37, no. 4, pp. 713–721, Dec. 2014.
[14] S. Bjorklund and L. Ljung, "A review of time-delay estimation techniques," in *Proc. 42nd IEEE Int. Conf. Decis. Control*, Dec. 2003, pp. 2502–2507.
[15] Y.-A. Le Borgne, S. Santini, and G. Bontempi, "Adaptive model selection for time series prediction in wireless sensor networks," *Signal Process.*, vol. 87, no. 12, pp. 3010–3020, Dec. 2007.
[16] P. L. Bartlett, S. Boucheron, and G. Lugosi, "Model selection and error estimation," *Mach. Learn.*, vol. 48, no. 1, pp. 85–113, 2002.
[17] F. Gustafsson, *Adaptive Filtering and Change Detection*, vol. 1. Hoboken, NJ, USA: Wiley, 2000.
[18] J. Falk, P. Händel, and M. Jansson, "Direction finding for electronic warfare systems using the phase of the cross spectral density," in *RadioVetenskap och Kommunikation (RVK)*, Stockholm, Sweden, Jun. 2002, pp. 264–268.
[19] R. Carrasco-Alvarez, R. Parra-Michel, A. G. Orozco-Lugo, and J. K. Tugnait, "Time-varying channel estimation using two-dimensional channel orthogonalization and superimposed training," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4439–4443, Aug. 2012.
[20] J. A. del Peral-Rosado, J. A. López-Salcedo, F. Zanier, and G. Seco-Granados, "Position accuracy of joint time-delay and channel estimators in LTE networks," *IEEE Access*, vol. 6, pp. 25185–25199, 2018.
[21] F. Milan and M. Anghel, "Impact of time delays on power system stability," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 4, pp. 889–900, Apr. 2012.
[22] C. M. Ionescu, R. Hodrea, and R. De Keyser, "Variable time-delay estimation for anesthesia control during intensive care," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 363–369, Feb. 2011.
[23] P. Laguna, A. Garde, B. F. Giraldo, O. Meste, R. Jané, and L. Sörnmo, "Eigenvalue-based time delay estimation of repetitive biomedical signals," *Digit. Signal Process.*, vol. 75, pp. 107–119, Apr. 2018.
[24] J. Kim, H. Choi, and J. Kim, "A robust motion control with antiwindup scheme for electromagnetic actuated microrobot using time-delay estimation," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 3, pp. 1096–1105, Jun. 2019.
[25] S. Mehrkanoon, Y. A. W. Shardt, J. A. K. Suykens, and S. X. Ding, "Estimating the unknown time delay in chemical processes," *Eng. Appl. Artif. Intell.*, vol. 55, pp. 219–230, Oct. 2016.
[26] V. A. O. Alves, R. J. Correa de Godoy, and C. Garcia, "Optimal time delay estimation for system identification," in *Proc. Amer. Control Conf.*, Jun. 2013, pp. 95–100.
[27] J.-P. Richard, "Time-delay systems: An overview of some recent advances and open problems," *Automatica*, vol. 39, no. 10, pp. 1667–1694, Oct. 2003.
[28] S. Beheshti and M. A. Dahleh, "A new information-theoretic approach to signal denoising and best basis selection," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3613–3624, Oct. 2005.
[29] E. Naghsh, M. Danesh, and S. Beheshti, "Unified left eigenvector (ULEV) for blind source separation," *Electron. Lett.*, vol. 58, no. 1, pp. 41–43, Jan. 2022.
[30] Y. Sadat-Nejad and S. Beheshti, "Efficient high resolution sLORETA in brain source localization," *J. Neural Eng.*, vol. 18, no. 1, Feb. 2021, Art. no. 016013.
[31] T. Yousefi Rezaii, S. Beheshti, M. Shamsi, and S. Eftekharifar, "ECG signal compression and denoising via optimum sparsity order selection in compressed sensing framework," *Biomed. Signal Process. Control*, vol. 41, pp. 161–171, Mar. 2018.
[32] S. Beheshti and S. Sedghizadeh, "Number of source signal estimation by the mean squared eigenvalue error," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5694–5704, Nov. 2018.
[33] K. S. Brooks and M. Bauer, "Sensor validation and reconstruction: Experiences with commercial technology," *Control Eng. Pract.*, vol. 77, pp. 28–40, Aug. 2018.
[34] E. M. Carlini, F. Del Pizzo, G. M. Giannuzzi, D. Lauria, F. Mottola, and C. Pisani, "Online analysis and prediction of the inertia in power systems with renewable power generation based on a minimum variance harmonic finite impulse response filter," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107042.
[35] L. Lin, L. Yi Wang, and G. Zames, "Time complexity and model complexity of fast identification of continuous-time LTI systems," *IEEE Trans. Autom. Control*, vol. 44, no. 10, pp. 1814–1828, Oct. 1999.
[36] T. Moon and T. Weissman, "Universal FIR MMSE filtering," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1068–1083, Mar. 2009.
[37] S. Asefi, S. Parsegov, and E. Gryazina, "Distributed state estimation: A novel stopping criterion," 2020, *arXiv:2012.00647*.
[38] H. Cao, H. C. So, and Y. T. Chan, "Optimum time delay estimation for complex-valued stationary signals," *Signal Process.*, vol. 131, pp. 434–440, Feb. 2017.
[39] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 4, Apr. 2007, pp. IV-317–IV-320.

[40] A. Dytso, M. Fauß, A. M. Zoubir, and H. V. Poor, "MMSE bounds for additive noise channels under Kullback–Leibler divergence constraints on the input distribution," *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6352–6367, Dec. 2019.
[41] N. I. Achieser, *Theory Approximation*. Chelmsford, MA, USA: Courier Corporation, 2013.
[42] K. S. Miller, "On the inverse of the sum of matrices," *Math. Mag.*, vol. 54, no. 2, pp. 67–72, 1981.

MAHDI SHAMSI (Member, IEEE) received the B.Sc. degree in biomedical engineering, with a focus on bioelectric from the University of Tabriz and the Ph.D. degree from the Electrical and Computer Engineering Department, Toronto Metropolitan University. His research centered on compressed sensing and statistical signal processing with the University of Tabriz. His current research interests include the intersection of statistical learning theory, machine learning, hyperparameter optimization, graph signal processing, and graph neural networks.

SOOSAN BEHESHTI (Senior Member, IEEE) received the B.S. degree from Isfahan University of Technology, Isfahan, Iran, and the M.S. and Ph.D. degreesfrom the Massachusetts Institute of Technology (MIT), Cambridge, in 1996 and 2002 respectively, all in electrical engineering. From September 2002 to June 2005, she was a Postdoctoral Associate and a Lecturer at MIT. Since July 2005, she has been with the Department of Electricaland Computer Engineering, Toronto Metropolitan University (formerly Ryerson), where she is currently a professor. She is also the Director of Signal and Information Processing (SIP) Laboratory,Toronto Metropolitan University. Her research interests include statistical inference and signal processing, machine learning, and data denoising and preprocessing.