From data to diagnosis and control using generalized orthonormal basis filters. Part I: Development of state observers

Sachin C. Patwardhan a, Sirish L. Shah b,*

a Department of Chemical Engineering, Indian Institute of Technology, Bombay, Powai, Mumbai 400 076, India
b Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada T6G 2G6

Abstract

This work is aimed at the development of a state observer (steady state Kalman filter) for a multivariable system with unknown time delays, which is subjected to unmeasured disturbances. To design such a filter, we explore the feasibility of capturing system dynamics using generalized orthonormal basis filters (GOBF). A two step identification procedure is proposed by exploiting the fact that the GOBF based models have output error structure. The deterministic component of the model is identified in the first step and used to compute a residual signal. In the second step, a filter that whitens the residuals is estimated using GOBF and combined with the deterministic component. A minimal order state realization of the innovation form of the state model is then generated from this high order model using realization based sub-space based state space (4SID) identification algorithm. When time delays are not known a-priori, the similarity between GOBF and Pade approximation is used to estimate time delay matrix directly from multivariate data. The efficacy of the proposed modeling technique is demonstrated by carrying out simulation studies on the benchmark Shell control problem and experimental evaluation on a stirred tank heater (STH) system. From the analysis of simulation and experimental results, it can be inferred that the proposed approach produces fairly accurate estimates of the time delay matrix and the deterministic and stochastic components of the dynamic model.

Keywords: Orthonormal basis filters; Time delay estimation; Output error models; Unmeasured disturbance modeling; State observers

1. Introduction

Model predictive control (MPC) has been widely used in the process industry over the last two decades for controlling key unit operations in chemical plants. As a consequence, there has been significant research activity in the process control community with the aim of improving the analysis and synthesis of MPC controllers [1]. The focus of research in the last decade in the area of linear model predictive control (MPC) has been mainly on the development of controller synthesis schemes based on stochastic state space models. Model based diagnosis of process faults is another area of research that is receiving increasing attention in the field of process control [2]. An important class of methods which appear quite promising for on-line fault diagnosis is based on state observers or Kalman filters [3]. Thus, the development of reliable stochastic models for capturing unexplained dynamics of a process appears to be the key to success of these advanced control and on-line diagnosis schemes.

Given the state and measurement noise characteristics, there are well-established procedures to design a predictive control or a fault diagnosis scheme based
on an optimal state observer (Kalman filter). However, in most of process control situations such stochastic models with known state and measurement noise characteristics are seldom available. One remedy to this problem is to develop stochastic models for unmeasured disturbances (or the state observers) directly from the input–output data. From a practical viewpoint, this approach can considerably simplify the implementation of control and diagnosis schemes based on stochastic models.

There are several methods available in the literature for developing stochastic time series models [4,5]. However, it should be noted that most of the conventional methods used for development of time series models assume that the dead time (or time delay) between each input–output pair is known a priori. Since inclusion of time delays as an extra parameter to be estimated makes the estimation problem highly nonlinear and difficult to solve, the dead time is typically estimated a priori using other techniques before developing a time series model. Time delay estimation methods based on analysis of open loop step response behavior have been widely reported in literature [6,7]. These methods can be applied to systems with reasonably fast dynamics where the measured or unmeasured disturbances can be maintained at their nominal levels during the step test. The other class of commonly used methods is based on correlation analysis and attempts have been made to account for effects of unmeasured disturbances while estimating dead time by this approach [8]. A major difficulty with most of the available approaches is that, as proposed, these methods are applicable to dead time estimation for the SISO case and their extension to multivariable case is not obvious. In many practical situations, even if one manipulated input is perturbed at a time, it is not possible to keep measured and unmeasured disturbances constant during the identification tests conducted on the plant. In fact, it may even become necessary to change other manipulated inputs during the identification test to counter the effect of measured or unmeasured disturbances. Also, if the effect of unmeasured disturbances is not accounted properly, the estimates of dead time generated using conventional methods can be erroneous and the resulting time series model will have biased parameters. Thus, prior to development of time series or innovations form of state space models, it is necessary to develop methods for estimation of dead time when multiple inputs and/or disturbances are changing simultaneously.

In recent years there has been growing interest in the use of orthonormal basis filters (OBF) for representing process dynamics [9–12,17,18]. The orthogonal filter approximations provide a simple and elegant method of representing open loop stable systems. In fact, these models can be looked upon as a compact (parsimonious in parameters) representations of convolution type models, which have been widely used in MPC schemes. Advantages of using orthonormal basis filters for process modelling can be summarized as follows:

- If some a priori information about system dynamics is available, then the resulting parameter estimation problem can be solved analytically using linear regression.
- OBF based models have output error structure, which has an advantage that the deterministic component of the model can be estimated consistently whenever the noise is uncorrelated with the inputs.
- Orthonormal functions can represent signals which exhibit long time delays because of their similarity to the Padé approximation. Thus, development of OBF based models do not need any a priori knowledge about system time delays.

If we work with state space realizations of orthonormal filters, modeling of SISO as well as MIMO systems can be carried out with equal ease. Most of the applications of this modeling technique reported in process control literature are, however, for SISO systems with deterministic inputs.

This work is aimed at development of a parameter estimation strategy for multivariable system whereby the estimates of dead time and unmeasured disturbance models are generated from multivariate data [13]. We explore the feasibility of capturing deterministic and unmeasured disturbance dynamics of multivariable systems using generalized orthonormal basis filters (GOBF) proposed by Van den Hof et al. [9] and Ninness and Gustafsson [11]. It is assumed that the dynamics of the system under consideration can be adequately captured by a Box-Jenkins type time series model. To begin with, the deterministic component of the model is identified by assuming output error structure. This step ensures that the resulting model has good long range prediction capabilities. The filter that whitens the residuals (i.e. the unmeasured disturbance model) is identified in the next step and combined with the deterministic component. The innovations form of state space model resulting from the above approach is typically of very high order. A minimal order state representation of the innovation from of state model is then generated from this high order model using realization based 4SID (Sub-space based State Space System Identification) algorithm proposed by Kung [14]. When time delays are not known a priori, we propose to carry out the identification exercise in two phases. The similarity of GOBF to Padé approximation is exploited in the first step to develop a model that approximates time delays as non-minimum phase zeros. Estimates of time delay in each input output pair can then be generated from the analysis of step responses of the deterministic part of the resulting model.
In the second phase, the time delay estimates are further used to refine the model by choosing GOBF basis filters with poles at the origin.

Thus, the proposed method generates a minimal order state observer that can be directly used for model predictive control or model based fault diagnosis. When compared to direct 4SID methods (i.e. N4SID, MOESP or CVA), the main advantage of our approach is that we use the prediction error method to develop deterministic and stochastic components of the model. The prediction error method, though involves nonlinear optimization, has well known statistical properties and is known to perform better than the subspace methods in many situations [15]. However, the main difficulty with using PEM for developing a general multivariable state space model using conventional canonical parameterization is that the number of parameters to be estimated becomes large and this renders the numerical optimization problem to calculate the optimal PEM estimates impractical [15]. We partly overcome this difficulty by parameterizing state space models using GOBF and the resulting NLP is of much smaller dimension. In fact, if GOBF poles are fixed based on some a priori knowledge about the system, the resulting optimization problem can be solved analytically by linear least square. The efficacy of the proposed modeling technique is demonstrated by carrying out modeling studies on

- Shell Benchmark Control Problem (a heavy oil fractionator system).
- Stirred Tank Heater (STH) system (laboratory scale experimental setup).

To begin with, we briefly review some recent results on orthonormal basis filters available in the literature. Procedures for model identification and time delay estimation based on orthogonal basis filters is outlined in the Section 3. Section 4 follows with model development case studies carried out using simulation and experimental data. The main conclusions reached through data analysis are presented in Section 5.

2. Approximation using orthonormal basis filters

Following [12], let \( \mathcal{T} \) denote the unit circle \( \{z:|z|=1\} \) and \( \mathcal{E} \) denote exterior of unit disc \( \{z:|z|>1\} \). We consider the Hardy space \( \mathcal{H}_2 \), of square integrable functions on \( \mathcal{T} \) and analytic in \( \mathcal{E} \). The corresponding inner product of \( F_1(z), F_2(z) \in \mathcal{H}_2 \) is denoted by

\[
\langle F_1(z), F_2(z) \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} F_1(e^{i\omega})^* F_2(e^{i\omega}) \, d\omega
\]

where \( * \) denotes the complex conjugate and \( T \) denotes sampling interval. The norm of \( F(z) \) is defined as \( ||F(z)|| = \sqrt{\langle F(z), F(z) \rangle} \). Also, two transfer functions are called orthonormal if they satisfy \( \langle F_1(z), F_2(z) \rangle = 0 \) and \( ||F_1(z)|| = ||F_2(z)|| = 1 \).

Consider a SISO system represented by a strictly proper stable transfer function \( G(z) \in \mathcal{H}_2 \) and

\[
\vartheta(z) = G(z)\varphi(z)
\]

where \( \varphi(z) \) represents input and \( \vartheta(z) \) represents the model output. Let \( \{F_i(z)\} \) for \( i = 1, 2, \ldots \) be an orthonormal basis for \( \mathcal{H}_2 \) such that

\[
\langle F_i(z), F_j(z) \rangle = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}
\]

Then, there exists a unique generalized Fourier series expansion of \( G(z) \) such that

\[
G(z) = \sum_{i=1}^{\infty} c_i F_i(z)
\]

where \( \{c_i\} \) represent Fourier coefficients defined as \( c_i = \langle G(z), F_i(z) \rangle \). Thus, given a linear time invariant system \( G(z) \), the \( n \)th order finite expansion model that approximates \( G(z) \) best in a \( \mathcal{H}_2 \) sense is given by

\[
G_n(z) = \sum_{i=1}^{n} c_i F_i(z)
\]

There are several possible ways to select basis filters \( F_i(z) \)

- Classical Finite Impulse Response (FIR) model
- Laguerre filter [16]:
- Kautz filter [17]:

\[
F_{2n-1}(z,b,c) = \frac{\sqrt{1-c^2}(z-b)}{z^2+b(c-1)z-c} \times \left[ \frac{-cz^2+b(c-1)z+1}{z^2+b(c-1)z-c} \right]^{i-1}
\]

\[
F_2(z,b,c) = \frac{\sqrt{(1-c^2)(1-b^2)}}{z^2+b(c-1)z-c} \times \left[ \frac{-cz^2+b(c-1)z+1}{z^2+b(c-1)z-c} \right]^{i-1}
\]

with \(-1 < b < 1 \) and \(-1 < c < 1 \).

- Generalized orthonormal basis filters (GOBF): Heuberger et al. [10] showed that an orthonormal basis with repeated and fixed set of poles can be generated via balanced realization of all pass filters. Thus, given
any scalar inner transfer function $g(s)$ with McMillan degree $n_d$ and a minimal balanced realization $(A, B, C, D)$, the sequence of rational functions

$$F_j(z) = e_j^T \left[ \left( zI - A \right)^{-1} B g_j^{-1}(z) \right]$$

$$j = 1, 2, \ldots, n_d; \ i = 1, 2, \ldots$$

(10)

forms an orthonormal basis for $\mathcal{H}_2$. The basis functions can be selected as 

Laguerre or Kautz filters. Also, one of the limitations of a better choice of orthonormal basis filters than Laguerre under consideration has scattered poles, GOBF is generated. If the system model is generated. If the system is an orthonormal basis filters by showing that

$$F_j(z, \xi) = \frac{\sqrt{1 - |\xi_j|^2}}{z - \xi_j} \prod_{i=1}^{n_d} \frac{1 - \xi_i^2 z}{z - \xi_i}$$

(11)

forms a complete orthogonal set in $\mathcal{H}_2$, where $\xi \equiv \{\xi_i; i = 1, 2, \ldots\}$ is an arbitrary sequence of poles inside the unit circle appearing in complex conjugate pairs. Note that GOBF proposed by Heuberger et al. [10] can be generated by using (11) with the additional assumption that poles of $F_j(z)$ are cyclically repeated from set $\{\xi_i; i = 1, 2, \ldots, n_h\}$.

The main advantage of using orthogonal basis filters other than $\{z^{-\tau}\}$ in the classical FIR models is that the transfer function $G(z)$ can be approximated by only a small number of coefficients in the expansion, i.e. a parsimonious in parameters model is generated. If the system under consideration has scattered poles, GOBF is a better choice of orthonormal basis filters than Laguerre or Kautz filters. Also, one of the limitation of models based on Laguerre or Kautz filters is that the time delays are approximated using non-minimum phase zeros. This difficulty can be alleviated by employing GOBF with some of the poles placed at the origin. Thus, for a SISO system with dead time equal to $d$ samples, the basis functions can be selected as

$$F_j(z) = z^{-j} \quad \text{for } j = 1, 2, \ldots, d$$

(12)

$$F_{j+d}(z) = \frac{\sqrt{1 - |\xi_j|^2}}{(z - \xi_j)} \prod_{i=1}^{n_d} \frac{1 - \xi_i^2 z}{(z - \xi_i)} z^{-d}$$

for $j = 1, 2, \ldots, N$

(13)

These filters represent generalization of the Markov–Laguerre filters proposed by Finn et al. [18] and are referred to as Markov-OBF filters in the rest of the text.

While dealing with the parameter identification and control problems, it is often convenient to work with state space realizations of GOBF models. Consider a SISO model of the form

$$\vartheta(z) = \sum_{i=1}^{n_d} c_i F_i(z, \xi) v(z)$$

(14)

Defining a state vector $x(k, \xi) \in \mathbb{R}^n$ as

$$x(k) = [x_1(z, \xi)v(k) \quad x_2(z, \xi)v(k) \quad \ldots \quad x_n(z, \xi)v(k)]^T$$

(15)

where $\xi = [\xi_1 \quad \xi_2 \quad \ldots \quad \xi_j]^T$ represents the vector of GOBF poles, a state realization of the form

$$x(k+1) = \Psi(\xi)x(k) + A(\xi)v(k)$$

(16)

$$\vartheta(k) = \theta^T x(k)$$

(17)

can be developed for the model given by Eq. (14) as discussed in the Appendix A. Here, the elements of the $\theta$ vector are the Fourier coefficients, i.e.

$$\theta = [c_1 \quad c_2 \quad \ldots \quad c_n]^T$$

(18)

### 3. Model identification

In this section, we propose a new sequential approach to the development of general time series models for asymptotically stable systems with unknown time delays from open loop data. Ljung [4] has shown that the parameters of an FIR models can be estimated consistently under open loop condition (i.e. when input sequence is independent of unmeasured disturbances), since the regressor vector contains only the input terms. We exploit the fact that GOBF models are similar to FIR models and their model coefficients can also be estimated consistently in the presence of colored noise, provided that the input signal is sufficiently exciting [9,12]. To begin with we review the procedure available in the literature [9,11,12] for developing deterministic component of the model using GOBF. In the system identification literature, noise modeling is carried out mainly for generating unbiased estimates of the deterministic component of the model. Since such a model can be generated using GOBF parameterization, not much attention has been paid to the modeling of resulting residuals in the GOBF related literature. In our work, we proceed with the modeling of residuals as we intend to use the stochastic component later for predictive control and fault diagnosis. The idea of using GOBF parameterization of state space models is then applied to carry out modeling of residuals by a novel approach. When time delays are unknown, we propose a new procedure to estimate time delay matrix directly from the multivariate input–output data used for identification.

We also propose a new method for determination of GOBF poles, when no a priori knowledge is available about the system poles. A minimal order state realization is then developed using Kung’s realization based 4SID method [14,15].

Consider a SISO linear, time invariant discrete time system modelled using a general Box-Jenkins type model

$$y(k) = G(z)u(k) + H(z)e(k)$$

(19)

where $G(z) \in \mathcal{H}_2$, $H(z)$ is a stable rational monic transfer function and $e(k)$ represents a zero mean Gaussian white noise sequence. We propose to carry out system
identification in two steps. In the first step, we assume an output error (OE) model structure, i.e.
\[ y(k) = G(z)u(k) + v(k) \]  
and generate an estimate \( \hat{G}(z) \) of the deterministic component by parameterizing \( G(z) \) using GOBF. In the next step, a filter that whitens the estimated residuals
\[ \hat{v}(k) = y(k) - \hat{G}(z)u(k) \]  
is identified, again using GOBF based parameterization. The details of the identification procedure are presented in the subsequent subsections.

### 3.1. Modeling the deterministic component \([9,11,12]\)

The model given by (19) can be rearranged as
\[ y(k) = G(z)u(k) + v(k) \]  
where \( v(k) \) represents zero mean colored noise signal. Now, define a one step ahead prediction error as
\[ \hat{v}(k, \xi_u, \theta_u) = y(k) - \sum_{j=1}^{n} c_{u,j}[F_j(z, \xi_u)u(k)] \]  
where \( x_u(k) \in R^n \) represents the state vector and \( \theta_u \) represents the parameter vector as defined by Eqs. (15) and (18), respectively. The least square estimate of the parameter vector \( \theta_u \) can be obtained by solving the following minimization problem:
\[ \hat{\theta}_u(\xi_u) = \arg \min_{\theta_u} \frac{1}{N_s} \sum_{k=1}^{N_s} \hat{v}(k, \theta_u)^2 \]  
where \( \xi_u \) represents a vector of GOBF poles. Given a set of GOBF poles \( \xi_u \), the above minimization problem can be solved analytically using the following simple linear regression scheme:
\[ \hat{\theta}_u(\xi_u) = (R_u)^{-1} E(x_u(k)y(k)) \]  
\[ R_u = \left[ E\left(x_u(k)x_u(k)^T\right) \right] \]  
where \( E(\cdot) \) represents expected value operator as defined by Ljung [4]. Assume that the true process evolves according to
\[ y(k) = x_u(k)^T \theta_u(\xi_u) + \sum_{j=n+1}^{\infty} c_{u,j}[F_j(z, \xi_u)u(k)] + v(k) \]  
where \( \theta_u(\xi_u) \) represents the vector of true parameters. When the data length \( N_s \) is sufficiently large, we have
\[ \hat{\theta}_u(\xi_u) = \theta_u(\xi_u) \]  
\[ + (R_u)^{-1} \left\{ \sum_{j=n+1}^{\infty} c_{u,j}E[\{F_j(z, \xi_u)u(k)\}x_u(k)] \right. \]  
\[ \left. + E[\{v(k)\}x_u(k)] \right\} \]  
where \( \hat{\theta}_u(\xi_u) \) represents the estimated model parameters. When the disturbance \( v(k) \) is uncorrelated with the input sequence (i.e. under open loop conditions), it can be easily shown that
\[ E[\{x_u(k)v(k)\}] = E[F_1(z, \xi_u)u(k)v(k) \ldots F_n(z, \xi_u)u(k)v(k)]^T \]  
\[ = 0 \]  
for any choice of orthogonal filters \( F_j(z, \xi_u) \). Thus, it follows that:
\[ E\left( \hat{\theta}_u(\xi_u) - \theta_u(\xi_u) \right) \]  
\[ = (R_u)^{-1} \sum_{j=n+1}^{\infty} c_{u,j} E[\{F_j(z, \xi_u)u(k)\}x_u(k)] \]  
If the input signal is white, then it follows that \( E[\{x_u(k)F_j(z, \xi_u)u(k)\}] = 0 \) and \( \theta_u(\xi_u) \) can be estimated consistently even when the tail remains unmodeled. However, bias will result in rest of the cases. For a non-white quasi-stationary input signal with spectral density \( \phi_u(w) \), the bias in the estimated coefficients for a given model order \( n \) is bounded by [12]
\[ \left\| \hat{\theta}_u(\xi_u) - \theta_u(\xi_u) \right\|_2 \leq \left\| (R_u)^{-1} \right\|_2 \max_{w} \phi_u(w) \| \theta_{u,e} \|_2 \]  
where \( \theta_{u,e} = [c_{u,n+1} \ c_{u,n+2} \ldots]^T \). Note that the bias in estimated parameters can be made arbitrarily small in the region of interest by choosing model order \( n \) sufficiently large and appropriately designing the input perturbations.

The parameter estimation procedure outlined above can be easily extended to a MIMO \((r \times m)\) system by formulating \( r \) MISO identification problems. Given a set of GOBF poles, the regressor vector for \( i \)th output can be expressed as
\[ X_u(k) = \left[ x_{u1}(k) \ x_{u2}(k) \ldots \ x_{um}(k) \right]^T \]  
as defined in Appendix A and the resulting \( r \) linear regression problem can be solved analytically to generate consistent estimates of the deterministic component of the model.

### 3.2. Unmeasured disturbance modeling

The next step is to estimate a model for the unmeasured disturbances. The \( \hat{G}(z) \) estimated as above can be used to compute the residual signal as follows:
\[ \hat{v}(k) = y(k) - \hat{G}(z)u(k) \]  
and used to develop filter \( \hat{H}(z) \) that whitens residuals
\[ \hat{v}(k) \]  
\[ \hat{v}(k) = \hat{H}(z)e(k) \]
One possibility is to parameterize $\hat{H}(z)$ using GOBF. However, this approach renders the parameter estimation problem highly nonlinear. Alternatively, the above MA type model can be rearranged as an AR type model as follows [4,5]:

$$\hat{v}(k) = [I - \hat{H}(z)^{-1}]\hat{v}(k) + e(k)$$

$$= \hat{W}_e(z)\hat{v}(k) + e(k)$$

and $\hat{W}_e(z)$ can be parameterized using GOBF [17]. This approach considerably simplifies the parameter estimation problem. However, if we choose to work with transfer functions (i.e. in the domain of rational polynomials of $z$), then, recovering $\hat{H}(z)$ from $\hat{W}_e(z)$ can become computationally difficult task particularly for multivariable AR models. This difficulty can be alleviated by formulating a state realizations of $\hat{W}_e(z)$ parameterized using GOBF as

$$x_e(k + 1) = \Phi_e x_e(k) + K_e e(k)$$

such that

$$\hat{W}_e(z) = \theta^T_e [I - \Phi_e]^{-1} K_e$$

where $K_e$ represents the steady state Kalman gain. A simple way to recover $\hat{H}(z)$ from above state space model is to rearrange the state evolution equation as

$$x_e(k + 1) = \Phi_e x_e(k) + K_e e(k)$$

which is to rearrange the state evolution equation as

$$\hat{H}(z) = \theta^T_e [I - \Phi_e]^{-1} K_e + I$$

Thus, the parameter estimation problem for the AR type model given by Eq. (34) can be formulated as

$$\hat{\theta}_e(\xi_e) = \arg\min_{\theta_e} \frac{1}{N}\sum_{k=1}^{N} e(k, \xi_e, \theta_e)^2$$

subject to

$$e(k, \xi_e, \theta_e) = \hat{v}(k) - \sum_{j=1}^{u} c_{e,j} [F_j(z, \xi_e)\hat{v}(k)]$$

$$= \hat{v}(k) - x_e(k, \xi_e)^T \theta_e$$

subject to

$$a_{ij} = \max \{\Delta y_{i,j}(k) : k = 1, 2, \ldots, N_{p}\}$$

where $\Delta = 1 - z^{-1}$ and equation of the tangent is given as

$$y_{i,j} = a_{ij} t + b_{ij}$$

The parameter estimation procedure outlined above can be easily extended to a MIMO $(r \times m)$ system by formulating either $r$ SISO or $r$ MISO identification problems. Given a set of GOBF poles, the regressor vector for $i$th output for the MISO case can be expressed as (see Appendix A)

$$X_{x,i}(k) = [x_{x,1}(k)^T \quad x_{x,2}(k)^T \quad \ldots \quad x_{x,p}(k)^T]^T$$

and the resulting $r$ linear regression problem can be solved analytically to generate a MIMO unmeasured disturbance model.

### 3.3. Estimation of time delays

As stated earlier, most of the conventional methods for system identification require a priori estimates of time delays between each input output channels. The information regarding time delays has to be generated using separate procedures. The GOBF based modeling scheme, on the other hand, can be effectively used to estimate time delays due to similarity of GOBF based models to Padé approximations. When the time delays are unknown, the modeling exercise can be carried out as follows:

**Step 1**: To begin with, the time delays in all input–output channels are assumed to be equal to zero and a model is identified using GOBF given by (11) such that none of the filter poles are at the origin. In this step the time delays get approximated by non-minimum phase zeros and the corresponding step responses show inverse responses.

**Step 2**: The time delay between each input–output pair can be estimated using well known graphical method based on the point of inflection. By this approach, a tangent is drawn at the point of inflection of the step response curve and the intersection of the tangent line with time axis is taken as estimate of time delay [19]. Note that, since a GOBF based model has already been developed in step 1, the noise free step responses can be generated using this model and the points of inflection for each step response curve can be located accurately. Let $\{y_{i,j}(k) : k = 1, 2, \ldots, N_{p}\}$ represents first $N_{p}$ points in the step response between $i$th output and $j$th input generated by introducing unit magnitude step input in the $j$th input such that $y_{i,j}(\infty) \geq 0$. Then, the slope of the tangent at the point of inflection is given as:

$$a_{ij} = \max \{\Delta y_{i,j}(k) : k = 1, 2, \ldots, N_{p}\}$$

where $\Delta = 1 - z^{-1}$ and equation of the tangent is given as

$$y_{i,j} = a_{ij} t + b_{ij}$$

$$b_{ij} = y_{i,j}(k_{m}) - a_{ij} k_{m}$$
where $t$ represents scaled time (with respect to sampling interval) and $k_{m,n} \leq N_{a}$ represents the time instant at which injection occurs in the step response curve. The estimate of the time delay between the $i$th output and the $j$th input can be computed by finding the intersection of this tangent with time axis as

$$\tau_{dij} = T \left[ \frac{b_{ij}}{a_{ij}} - 1 \right]$$

provided $\left( \frac{b_{ij}}{a_{ij}} \right) \geq 1$. Here, $T$ represents sampling interval. Note that a unit delay due to discretization has been subtracted from $\left( \frac{b_{ij}}{a_{ij}} \right)$ to generate the estimate of time delay. The accuracy of the time delay estimate generated by this approach will be governed by the choice of the sampling interval $T$. Note that the above procedure requires introduction of negative unit input step change when the process gain is negative.

Step 3: The time delay estimates generated in step 2 can be used to develop a Markov-GOBF basis of type (12) and (13) and further used to re-identify another GOBF based model for the system.

Note that the procedure outlined above can be easily automated and does not require any graphical construction to estimate the point of inflection.

3.4. Estimation of GOBF Poles

The key step in the development of the GOBF models is the selection of filter poles and number of basis filters $n$ necessary to develop a reasonably good approximation of the system dynamics. In principle, an orthonormal basis for $\mathcal{H}_2$ can be constructed by selecting GOBF poles in an arbitrary manner. However, using an arbitrary set of basis filters may require large number of terms in the Fourier series and the resulting model is no longer parsimonious in parameters. Van den Hof et al. [12] have shown that, for a SISO system with poles $\{ p_j^b : |p_j^b| < 1 \text{ for } j = 1, 2, \ldots, n_0 \}$ the rate of convergence of coefficients $\{ c_1, c_2, \ldots \}$ is determined by the slowest magnitude eigenvalue

$$\rho = \max_j \left| \frac{p_j^b - \xi_j}{1 - \xi_j p_j^b} \right|$$

Thus, if we choose basis functions for which $\rho$ is small, a good approximation can be generated with smaller number of coefficients. This essentially reduces to selecting a basis with poles that closely match the dominant poles of the system to be approximated, as suggested by Walhberg [17]. In many practical situations there is some a priori knowledge about the dominant system time constants, which can be effectively used to achieve increased rate of convergence of the expansion coefficients. This allows incorporation of a priori knowledge about the system dynamics in the model. Note that the bias in the estimated parameters can be significantly reduced if the estimated expansion has an increasing decay rate [12]. An added advantage of choosing filter poles based on the a priori knowledge is that the resulting parameter estimation can be solved analytically.

When a priori knowledge about the dominant system time constants is not available, the parameter estimation problem can be formulated as two nested optimization problems. For example, in the case of unmeasured disturbance modeling, the optimization problem can be re-formulated as follows:

$$\left( \hat{\xi}_v, \hat{\theta}_v \right) = \arg \min_{\xi, \theta} \left[ \sum_{k=1}^{N_v} e(k, \xi, \theta)^2 \right]$$

subject to

$$\hat{\theta}_v(\xi) = (R_v)^{-1} \mathbf{X}(\mathbf{x}(k) \tilde{v}(k))$$

$$|\hat{\xi}_v| < 1 \text{ for } i = 1, \ldots, l$$

This leads to a constrained nonlinear optimization problem which can be solved using any standard nonlinear optimization method. The above nonlinear optimization problem has, in general, multiple solutions. In fact, since the GOBF generated using any choice of filter poles forms a orthonormal basis for $\mathcal{H}_2$, any local feasible solution is an acceptable solution to the optimization problem.

Note that the parameter estimation procedure outlined above also generates estimates of innovation sequence $\{ e(k) \in \mathbb{R}^N : k = 1, 2, \ldots, N \}$. This innovation sequence can be used to estimate the covariance matrix of innovation sequence $\hat{P}_e$ as follows:

$$\hat{P}_e = \frac{1}{N_v} \sum_{k=1}^{N_v} e(k) e(k)^T$$

The proposed decomposition strategy to estimate only the GOBF poles by nonlinear iterative search and the GOBF expansion coefficients analytically is based on the fact that every guess of GOBF poles $\{ \xi_j \}$ generates a valid orthonormal basis for the Hardy space $\mathcal{H}_2$. This scheme has definite numerical advantages over the simultaneous nonlinear iterative search in the space of all parameters, i.e. $\{ \hat{\xi}_v, \hat{\theta}_v \}$. The main difficulties with the later scheme are large dimension of the resulting optimization problem and generation of consistent initial guesses for the GOBF expansion coefficients. The proposed scheme exploits the fact that GOBF expansion coefficients $\{ \hat{\theta}_v \}$ are, in fact, dependent on the choice of GOBF poles $\{ \hat{\xi}_v \}$, i.e. the decision variables for the outer nonlinear optimization problem. Also, given a valid basis orthonormal basis, there is global minimum for the resulting optimization problem and the expansion coefficients can be computed uniquely. As a consequence, the proposed two step decomposition of optimization problem results in the dimensionality reduction.
and alleviates difficulties associated with initialization of GOBF expansion coefficients without compromising on the optimality of the solution.

**Remark.** When the system dynamics is modelled with GOBF having only real poles, it is convenient to restrict the search in the region \(0 < \xi_i < 1\). Also, numerical conditioning of the optimization problem can be improved if the search is performed with respect to an equivalent continuous pole \(a_i\) such that \(\xi_i = \exp(a_iT)\) where \(T\) represents sampling interval. The constraint on continuous pole \(a_i\) equivalent to \(0 < \xi_i < 1\) can be expressed as \(a_i < 0\).

**Remark.** For modeling a MIMO system, it is convenient to formulate the parameter estimation problem based on the regressor vectors computed either using MISO or MIMO non-minimal state realizations given in Appendix A. However, in order to compute the state evolution while formulating parameter estimation problem, the initial state has to be specified. If the initial state is estimated together with the parameter vector, then it can result in a nonlinear programming problem. However, since we assume that \(G(z), H(z) \in \mathcal{H}_2\), this difficulty can be avoided by starting with zero initial state and formulating the loss function for \(k > k_0\).

### 3.5. Development of minimal order state realization

Using the sequential parameter estimation approach outlined above, we get two stacked up MIMO non-minimal state space realizations (see Appendix for details)

- **Model for deterministic component**
  \[
  X_u(k+1) = \Phi_u X_u(k) + \Gamma_u u(k) \quad (55)
  \]
  \[
  y(k) = \Theta_u X_u(k) + v(k) \quad (56)
  \]
  where \(X_u \in \mathbb{R}^{N_u}\).

- **Unmeasured disturbance model**
  \[
  X_v(k+1) = \Phi_v X_v(k) + K_v e(k) \quad (57)
  \]
  \[
  v(k) = \Theta_v X_v(k) + e(k) \quad (58)
  \]
  where \(X_v \in \mathbb{R}^{N_v}\).

Defining an augmented state vector
\[
X_a(k) = \begin{bmatrix} X_u(k) & X_v(k) \end{bmatrix}^T \in \mathbb{R}^N
\]
where \(N = N_u + N_v\), an innovation form of state space realization can be obtained as
\[
X_a(k+1) = \Phi_a X_a(k) + \Gamma_a u(k) + K_v e(k) \quad (60)
\]
\[
y(k) = C_a X_a(k) + e(k) \quad (61)
\]

where \(e(k)\) is the state disturbance. The resulting state space model is of very high order and non-minimal representation of the system dynamics. Application development based on such large dimensional state space models can lead to increased computational complexity. Thus, it is natural to seek a minimal state space realization of the form
\[
X(k+1) = \Phi X(k) + \Gamma u(k) + K e(k) \quad (62)
\]
\[
y(k) = C X(k) + e(k) \quad (63)
\]

for the identified state space model given by Eqs. (60) and (61). Alternatively, deterministic and stochastic components of the models can be reduced separately and the combined to arrive at a low order state space model of the form (62) and (63). There are two approaches available in the literature for construction of a minimal realization:

- Classical Kung’s realization based 4-SID method [14,15]: This approach is based on Ho and Kalman algorithm and constructs a minimal realization from the block Henkel matrix generated using the Markov parameter estimates.

- Minimal partial realization from orthonormal basis expansion [21]: By this approach, \(G(z)\) is expressed in terms of Hambo transform signals and generalized Markov parameter estimates of the Hambo transform operator representation of \(G(z)\) are constructed directly from the estimated GOBF expansion coefficients. The block Henkel matrix constructed from these generalized Markov parameters is then used to develop a minimal order realization in a manner similar to Kung’s method. This approach can be viewed as a generalization of the classical approach.

In the present work, we choose to follow the classical approach for developing a minimal realization. The details of Kung’s realization algorithm can be found in Viberg [15]. In order to arrive at a minimal realization, we define an index \(\rho(j)\)
\[
\rho(j) = \frac{\sigma_1 + \sigma_2 + \ldots + \sigma_j}{\sum_{i=1}^{N_H} \sigma_i}
\]
where \(N_H\) represents the total number of singular values of the estimated Henkel matrix \(H\) of block size \(x \times x\)
\[
H = \begin{bmatrix}
H_1 & H_2 & \ldots & H_x \\
H_{2} & H_{3} & \ldots & H_{x+1} \\
\vdots & \vdots & \ddots & \vdots \\
H_{\beta} & H_{\beta+1} & \ldots & \ldots
\end{bmatrix}
\]
Here \( \{H_i; i = 1, 2, \ldots \} \) represent impulse response coefficient matrices of dimension \((r \times m)\). Such a Hankel matrix can be generated using extended observability and extended controllability matrices for the non-minimal system such as Eqs. (60) and (61). The order \( x \) and \( \beta \) are selected to exceed the largest expected system order. The number of significant singular values \( n \) can be selected equal to the maximum value of \( j \) such that \( \rho(j) \leq x \). Here, \( x \) is a number, typically between 0.99 and 1, decided by the user. Thus, the state space model (62) and (63), together with the innovation covariance matrix \( \hat{V}_e \) represents a minimal (or a low) order state observer for the system under consideration. It should be noted that the resulting minimal order state space model is observable and controllable by construction.

4. Simulation and experimental case studies

The efficacy of the proposed modeling technique is demonstrated by carrying out modeling studies on

(a) Shell Control Problem (a heavy oil fractionator system) [20],
(b) Stirred Tank Heater (STH).

The model proposed for capturing dynamics of these systems is of the form

\[
y(z) = G(z)u(z) + H(z)e(z)
\]  

(64)

All the modeling exercises are carried out in two steps. In the first step, no knowledge of time delays is assumed and a model is developed using GOBF filters without any poles at the origin. This GOBF based model is used for the estimation of time delay matrix. The estimated time delay matrix is then used for the development of an observer based on the Markov-GOBF. When it is desired to estimate GOBF filter parameters using nonlinear optimization, the resulting constrained optimization problem is solved using the Optimization Toolbox of MATLAB. The large order state observer resulting from Markov-OBF based modeling is then reduced to generate a minimal order state space model using realization based 4SID method with \( z = 0.9999 \).

As a part of model validation exercise, the unmeasured disturbance spectrum is estimated from the measured data and is then compared with the spectrum of the unmeasured disturbance model (residuals). The data necessary for estimating the disturbance spectrum is generated from the process by holding all the manipulated inputs constant at \( u(k) = 0 \) and measuring the resulting outputs \( \Xi_v \equiv \{y(k) = v(k); k = 1, 2, \ldots \} \). The spectrum of the unmeasured disturbance model (residuals) is computed as follows:

\[
\Phi_e(w) = \hat{H}(jw)\hat{V}_e \hat{H}(jw)^T
\]

(65)

where \( \hat{H}(jw) \) represents \( i \)th row of matrix \( \hat{H}(jw) \) and \( V_e \) represents estimated covariance of the innovation sequence. The spectral density plots are reported with respect to normalized frequency \((w/w_N)\), where \( w_N = \pi/T \) represents the Nyquist frequency for the system under consideration. The data set \( \Xi_v \) is also used for estimation of the signal to noise ratio matrix \( \hat{S}_{NR} \) for the identification data, where \( (i,j) \)th element of the matrix \( S_{NR} \) is computed as

\[
(\hat{S}_{NR})_{ij} = \frac{\sigma_{ij}^2}{\sigma_{ii} \sigma_{jj}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{ij}(w)dw
\]

(66)

where \( \sigma_{ij}^2 \) represents variance of sequence \( \{v(k)\} \) in the set \( \Xi_v \).

4.1. Shell Control Problem

The Shell Control Problem is a benchmark problem proposed at the Shell Process Control Workshop [20] and involves control of a heavy oil fractionator system characterized by large time delays in each input output pair. The heavy oil fractionator has three product draws, three side circulating loops and a gaseous feed stream. The system consists of seven measured outputs, three manipulated inputs and two unmeasured disturbances. Product specifications for top and side draws are determined by economic considerations. There is no product specification on bottom draw, however, there is an operating constraint on the bottom reflux temperature. Top draw, side draw and bottoms reflux duty can be used as manipulated variables to control the column while heat duties on the two side loops (upper reflux duty and intermediate reflux duty) act as unmeasured disturbances to the column.

Since the controlled outputs of interest are top end point, side end point and bottoms reflux temperature, in this work we consider a subsystem consisting of only these three outputs. The continuous time transfer function model for this sub-system is as follows:

\[
\hat{Y}(s) = G_p(s)u(s) + G_d(s)d(s)
\]

(67)

where

\[
G_p(s) = \begin{bmatrix}
4.05 e^{-27s} & 1.77 e^{-28s} & 5.88 e^{-27s} \\
5.3 e^{-18s} & 5.72 e^{-14s} & 6.9 e^{-15s} \\
4.38 e^{-40s} & 4.42 e^{-22s} & 7.2 e^{-19s}
\end{bmatrix}
\]

(68)

\[
G_d(s) = \begin{bmatrix}
1.2 e^{-27s} & 1.44 e^{-27s} \\
1.52 e^{-15s} & 1.83 e^{-15s} \\
1.14 e^{-27s} & 1.26 e^{-327s}
\end{bmatrix}
\]

(69)

where the time constants and time delays are reported in minutes. Note that all the inputs (including unmeasured
disturbances) are scaled between ±0.5. In the present work, the process dynamics are simulated under following assumptions:

- Manipulated inputs are piecewise constant.
- Disturbances entering the plant can be adequately represented using piecewise constant functions.

Under these assumptions, a discrete dynamic model of the form

\[
\hat{y}(z) = G_p(z)u(z) + G_d(z)d(z)
\]  
(70)

is developed with sampling time \(T\) equal to 2 min. A minimal order state space realization of (70) of the form

\[
X(k + 1) = AX(k) + B_1u(k) + B_2d(k)
\]  
(71)

\[
\hat{y}(k) = CX(k)
\]  
(72)

with 51 state variables is used for simulation of process behavior. The stationary unmeasured disturbances \(d(z)\) are assumed to be generated by the following stochastic process:

\[
x_w(k + 1) = A_w x_w(k) + B_w w(k)
\]  
(73)

\[
d(k) = C_w x_w(k) + D_w w(k)d
\]

where \(A_w = C_w = 0.95I\) and \(B_w = D_w = I\)

or equivalently by

\[
d(z) = \begin{bmatrix}
\frac{z}{z-0.95} & 0 \\
0 & \frac{z}{z-0.95}
\end{bmatrix} w(z)
\]  
(75)

where \(w \in \mathbb{R}^2\) is a zero mean normally distributed white noise process with \(\sigma_{w1} = \sigma_{w2} = 0.0075\). In addition, the measured outputs are assumed to be corrupted with measurement noise

\[
y(k) = \hat{y}(k) + v(k)
\]  
(76)

where \(v \in \mathbb{R}^3\) represents zero mean normally distributed white noise process with \(\sigma_{v1} = 0.005\) for \(i = 1,2,3\). In order to carry out system identification, a low frequency (in the range \([0 \ 0.01\omega_N]\) where \(\omega_N = \pi/T\) represents Nyquist frequency) random binary signal with amplitude 0.075 (generated using the ‘idinput’ function in System Identification Toolbox of MATLAB) were simultaneously introduced in all the manipulated inputs and 2000 data points were collected. The estimates of signal to noise ratios between each input and disturbance in each output are as follows:

\[
\hat{S}_{NR} = \begin{bmatrix}
6.69 & 6.14 & 6.86 \\
2.97 & 2.73 & 3.05 \\
6.57 & 6.03 & 6.74
\end{bmatrix}
\]  
(77)

The transfer function matrix \(\hat{H}(z)\) in model (64) is assumed to be a full \(3 \times 3\) matrix with non-zero off-diagonal entries.

The first step in the identification exercise is the estimation of time delays. To begin with, let us assume that we know a priori that the dominant system time constants of the system lie between 20 and 50 min. Such an assumption is reasonable in most practical situations as approximate knowledge of time constants is necessary even for the selection of sampling interval. Thus, we select GOBF poles for all the individual SISO sub-models as

\[
\zeta_{u1} = \exp(-T/20) \quad \text{and} \quad \zeta_{u2} = \exp(-T/50)
\]

Since the GOBF filter poles are selected based on a priori knowledge about the system, the resulting parameter estimation problems can be solved by ordinary linear least squares. In order to generate good approximation of the system dynamics, eight basis filters generated by repeating above poles four times are used for modelling each input–output pair, i.e. each SISO sub-model has eight states. Fig. 1 compares the initial part of the process and model step responses. As evident from inverse responses in this figure, the time delays are approximated as non-minimum phase zeros outside the unit circle. The estimated time delay matrix based on the analysis of the step responses of the model using the procedure outlined in the previous section is as follows:

\[
\tau_\text{d(estimated)} = \begin{bmatrix}
27 & 27 & 26 \\
18 & 13 & 14 \\
19 & 20 & 0
\end{bmatrix}; \quad \tau_\text{d(true)} = \begin{bmatrix}
27 & 28 & 27 \\
18 & 14 & 15 \\
20 & 22 & 0
\end{bmatrix}
\]

Note that the estimates of time delay (in minutes, rounded off to nearest integer) are fairly close to true time delays and the error is of the order of at most a unit sampling period.

In the next step, the time delay estimates are used to introduce zeros at origin in the GOBF model and the model is re-estimated. The number of GOBF poles is fixed to two and only the first two basis filters generated by these poles are used in each SISO model. Unlike the first step where a priori knowledge is used to select GOBF poles, the nonlinear optimization procedure outlined in Section 3.4 is used in the second step to estimate optimum values of GOBF poles. The corresponding optimum values of poles for deterministic and stochastic components of the model are reported in Tables 1 and 2, respectively.

The minimal realizations of deterministic and stochastic models are developed separately and combined to generate an innovations form of the state space model with 49 states. The estimated innovation covariance matrix is as follows:

\[
\hat{P}_e = 10^{-4} \times \begin{bmatrix}
0.3597 & -0.001 & -0.0075 \\
-0.001 & 0.4555 & 0.0153 \\
-0.0075 & 0.0153 & 0.4144
\end{bmatrix}
\]
Figs. 2 and 3 compare the step and frequency responses of the minimal order model with that of the process, respectively. Note that the model behavior in both the time and frequency domains closely matches that of the plant. Fig. 4 shows that disturbance model spectrum matches closely with the estimated disturbance spectrum.

4.2. Stirred tank heater (STH)

The laboratory scale experimental setup consists of a stirred tank heater in which a hot water stream and a cold water stream are continuously mixed and heated as shown in Fig. 5. The heat necessary to raise the temperature of the water in the tank is supplied by a steam coil, which remains completely immersed in water during the experimental runs. Sensors are available to measure the water level and temperature in the tank and the hot water flow rates. The cold water flow, steam flow and hot water flow rates can be manipulated through the control computer using real-time MATLAB and SIMULINK. The level loop is assumed to be closed a priori as shown in Fig. 5. The level is controlled using the following PI controller:

\[
g_{c1}(s) = \frac{3}{1 + 0.1s} \]

which manipulates the inlet cold water flow rate. The hot water inflow \(F_{H}\) is treated as an unmeasured disturbance while carrying out system identification. Note that the hot water inflow \(F_{H}\) is measured and can be changed from the control computer by changing the valve position of control valve CV-3 on the hot water line (see Fig. 5). A PI controller

\[
g_{c2}(s) = 0.1 \left(1 + \frac{1}{7.5s}\right) \]

is used to control the hot water flow to the system as shown in Figs. 5. The unmeasured disturbance signal is generated by changing the setpoint to the hot water flow loop, which is fed as a piecewise constant input signal. The setpoint to the hot water flow control loop was generated as follows:

Disturbance Model: \(\delta F_{H,sp}(k) = \frac{0.3}{1 - 0.95s^{-1}} v(k)\)  

where \(\delta F_{H,sp}(k) = F_{H,sp}(k) - F_{H,sp}\) (in mA) represents setpoint to the hot water flow loop and \(v(k)\) represents zero mean Gaussian white noise sequence with \(\sigma_v = 1\). The steady state operating setpoint for hot water flow is fixed at \(F_{H,sp} = 7\) mA. The sampling time is chosen as 1 s. The

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Shell Control Problem—optimum GOBF poles for deterministic model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_1)</td>
<td>(0.956, 0.948)</td>
</tr>
<tr>
<td>(y_2)</td>
<td>(0.956, 0.939)</td>
</tr>
<tr>
<td>(y_3)</td>
<td>(0.9438, 0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Shell Control Problem—optimum GOBF poles for stochastic model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{y}_1)</td>
<td>(0.648, 0.999)</td>
</tr>
<tr>
<td>(\hat{y}_2)</td>
<td>(0.368, 0.063)</td>
</tr>
<tr>
<td>(\hat{y}_3)</td>
<td>(0.935, 0.999)</td>
</tr>
</tbody>
</table>
relationship between the measured outputs in mA and their corresponding values in engineering units developed using experimental data [22] are as follows:

\[ h(\text{cm}) = 2.3371h(\text{mA}) - 7.5204 \]  
\[ T(\text{°C}) = 6.2825T(\text{mA}) - 23.9084 \]  

The relationships between (a) valve position v/s hot water flow rate and (b) steam valve position v/s estimated enthalpy obtained from experimental studies [22] are shown in Fig. 6.

Thus, the $2 \times 2$ open loop system in system identification exercises consists of water level ($h$) and temperature ($T$) as measured outputs and level loop setpoint ($h_s$) and steam valve position ($S_v$) as manipulated inputs. The hot water inflow ($F_H$) is treated as the unmeasured disturbance during the identification exercise. The steady state operating point of the process is chosen as

\[ \bar{h} = 12 \text{ mA}; \quad \bar{T} = 11 \text{ mA} \]

The corresponding steady state inputs are

\[ \bar{h}_s = 12 \text{ mA}; \quad \bar{S}_v = 9.8 \text{ mA} \]
Note that the steady state steam flow input necessary to maintain the system at the desired operating temperature changes from time to time depending on the cold water and hot water temperatures and steam header pressure. The values reported above are steady state values at the time of conducting identification experiments.

In the discussion that follows, even though we report the signal to noise ratio matrices, a word of caution is necessary while interpreting these quantities. Note that the experimental setup under consideration exhibits mildly nonlinear dynamics. Also, the unmeasured disturbances (hot water flow, hot water inlet temperature, steam header pressure and cold water inlet temperature) do not enter the system as simple linear additive disturbances. As a consequence, while modeling the system using a linear model of the form (64), the residual signal

\[ \hat{v}(k) = y(k) - \hat{G}(z)u(k) \]  

contains contributions due to (a) approximation errors arising from linear model assumption (b) effects of non-additive disturbances. The estimates of matrix \( S_{\text{NR}} \), on the other hand, are based data set \( \Xi_p \) generated by holding manipulated inputs \( u(k) = \theta \). This data contains only the effects of one component of \( \hat{v}(k) \), i.e. effects of non-additive disturbances. Thus, the resulting matrix \( \hat{S}_{\text{NR}} \) cannot be interpreted in the same way as in the case of a truly linear system. Since approximation errors arising from linear model assumption depend on estimate of \( \hat{G}(z) \), an estimate of apparent signal to noise ratio based on residual \( \hat{v}(k) \) can serve as a better quantitative measure of difficulties involved in the parameter estimation problem. Thus, given an estimate \( \hat{G}(z) \) of
of MATLAB. The signal to noise ratios are estimated as
\[ G(z) \] defined as signal to residual ratio \((\hat{S}_{RR})\) based on the spectrum of \(\hat{e}(k)\). The \((i,j)\)th element of matrix \((\hat{S}_{RR})\) is computed as
\[
(\hat{S}_{RR})_{ij} = \frac{\sigma_{\hat{e}_i}^2}{\sigma_{\hat{e}_j}^2} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{\hat{e}_i}(w) dw \int_{-\pi}^{\pi} \Phi_{\hat{e}_j}(w) dw
\]  
\quad(84)

The manipulated level setpoint and steam valve position were simultaneously perturbed with random binary inputs of amplitude 0.5 mA and 1 mA, respectively, and in frequency bands \([0 \ 0.05\omega_N]\) and \([0 \ 0.025\omega_N]\), respectively. These inputs signals were generated using the ‘idinput’ function in System Identification Toolbox of MATLAB. The signal to noise ratios are estimated as
\[
\hat{S}_{NR} \approx \begin{bmatrix} 23.6 & - \\ 5.5 & 22 \end{bmatrix}
\]  
\quad(85)

The model proposed for capturing system dynamics is of the form
\[
\begin{bmatrix} \delta h(k) \\ \delta T(k) \end{bmatrix} = \begin{bmatrix} g_{11}(z) & 0 \\ g_{21}(z) & g_{21}(z) \end{bmatrix} \begin{bmatrix} \delta h_1(k) \\ \delta h_2(k) \end{bmatrix} + \begin{bmatrix} h_{11}(z) & h_{12}(z) \\ h_{21}(z) & h_{22}(z) \end{bmatrix} \begin{bmatrix} e_1(k) \\ e_2(k) \end{bmatrix}
\]  
\quad(86)

where \(\delta h(k), \delta T(k), \delta h_1(k)\) and \(\delta h_2(k)\) represent deviation level, deviation temperature, deviation level setpoint and deviation steam valve position, respectively.

When the proposed modeling strategy was used for system identification, the estimated time delay matrix in the first phase of parameter estimation is as follows:
\[
\tau_d(\text{estimated}) = \begin{bmatrix} 2 & 0 \\ 5 & 5 \end{bmatrix}
\]

While developing an observer based on Markov-BOBF filters in the second phase of identification, the delay free part of each \(g_{ij}(z)\) and \(h_{ij}(z)\) was approximated using four basis filters. The filter parameters are estimated using the nonlinear optimization procedure and the resulting optimum values of the estimated GOBF filter parameter for deterministic and stochastic components are listed in Tables 3 and 4, respectively.

The resulting Markov-BOF based state space models have 44 states while the 4SID based minimal realization has 18 states for both the models. The estimated innovation covariance matrix is as follows:
\[
\hat{\Phi} = 10^{-3} \times \begin{bmatrix} 0.3831 & 0.0142 \\ 0.0142 & 0.3260 \end{bmatrix}
\]

Fig. 7 compares infinite horizon model predictions (generated without using the state feedback correction \(K_e(k)\) term in predictions) with that of the measured outputs for validation data set. The corresponding variation of manipulated inputs and unmeasured disturbances is reported in Fig. 8. Fig. 7 shows that the identified deterministic component of the model tracks the system output quite accurately over first 800 samples, which is significantly larger than the prediction horizon used in predictive control. Fig. 9 compares the spectrum of the model residuals with that of the unmeasured disturbance spectrum estimated from observed data. The signal to residual ratio matrix \((\hat{S}_{RR})\) computed using estimated \(\hat{G}(z)\) is
\[
\hat{S}_{RR} \approx \begin{bmatrix} 22.7 & - \\ 7.1 & 28.3 \end{bmatrix}
\]

which is not significantly different from estimated \(\hat{S}_{NR}\) and the resulting parameter estimation problem is well conditioned. However, the unmeasured disturbance model spectrum does not match with the disturbance spectrum based on set \(\Xi_c\). This could be attributed to the fact that the process exhibits mildly nonlinear dynamics. In the model development exercise, the plant-model mismatch arising from linear model assumption is lumped together with the unmeasured disturbances. Moreover the effect of unmeasured distur-

---

**Table 3**

<table>
<thead>
<tr>
<th>Stirred tank heater: optimum Markov-BOBF parameters for deterministic component</th>
<th>(h_i)</th>
<th>(S_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(h)</td>
<td>0.854, 0.781</td>
<td>-</td>
</tr>
<tr>
<td>(T)</td>
<td>0.867, 0.934</td>
<td>0.937</td>
</tr>
</tbody>
</table>

**Table 4**

<table>
<thead>
<tr>
<th>Stirred tank heater: optimum Markov-BOBF parameters for stochastic component</th>
<th>(\hat{\epsilon}_1)</th>
<th>(\hat{\epsilon}_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\epsilon}_1)</td>
<td>0.706</td>
<td>0.677</td>
</tr>
<tr>
<td>(\hat{\epsilon}_2)</td>
<td>0.695</td>
<td>0.444</td>
</tr>
</tbody>
</table>
bance on the system outputs is non-additive. These two factors contribute to the large mismatch in the spectrum of unmeasured disturbances and the spectrum of unmeasured disturbance model. Nevertheless, the identified model gives fairly accurate representation of the system dynamics in the neighborhood of the chosen operating point.

5. Conclusions

This work is aimed at the development of a state observer (steady state Kalman filter) for a multivariable system with unknown time delays, which is subjected to unmeasured disturbances. We explore the feasibility of capturing system dynamics using generalized orthonormal basis filters (GOBF). A two step identification procedure is proposed by exploiting the fact that the GOBF based models are similar to FIR models and therefore consistent estimates of deterministic component can be obtained even in presence of colored noise. The deterministic component of the model is identified in the first step and used to compute a residual signal. In the second step, a filter that whitens the residuals is estimated using GOBF. The deterministic and the stochastic components of the model are combined to arrive at a high order innovations form of state space model. A minimal order state realization of the innovations form of the state model is then generated from this high order model using realization based 4SID algorithm. When time delays are not known a priori, the similarity between GOBF and Padé approximation is used to develop a model that approximates time delays as non-minimum phase zeros. Estimates of time delay between each input–output pair can then be generated from the analysis of the step responses of the resulting model. These time delay estimates are further used to re-identify the state estimator by employing basis filters having poles at the origin. The efficacy of the proposed modeling technique is demonstrated by carrying out simulation studies on the benchmark Shell control problem and experimental evaluation on a stirred tank heater (STH) system. From the analysis of simulation and experimental results, it can be inferred that the proposed approach produces fairly accurate estimates of the time delay matrix and the deterministic and stochastic components of the dynamic model. Thus, the proposed
method generates an optimal state observer that can be directly used in a model predictive control formulation or a model based fault diagnosis scheme.

Acknowledgment

The authors would like to acknowledge the financial support through NSERC-Matrikon-ASRA Industrial Research Chair Program at University of Alberta, Canada for carrying out this work.

Appendix A. Non-minimal state realization of gobf model

A.1. SISO model

Consider a SISO linear system modelled using a GOBF network and represented as

\[ \vartheta(k) = \left( \sum_{i=1}^{n} c_i F_i(z) \right) u(k) \]  
(A.1)

where \( u(k) \) represents model input, \( \vartheta(k) \) represents model output and

\[ F_i(z) = \frac{\beta_i}{(z - \xi_i)} \prod_{j=1}^{i-1} \frac{(1 - \xi_j z)}{(z - \xi_j)} \]  
(A.2)

\[ \beta_i = \sqrt{(1 - |\xi_i|^2)} \]

Here, we illustrate the procedure to develop a state realization when all poles, \( \{\xi_i\} \), are real. Defining a state vector

\[ x(k) = [x_1(k) \ x_2(k) \ \ldots \ x_n(k)]^T \]  
(A.3)

where \( x_i(k) = F_i(z)u(k) \) represent output from \( i \)th order GOBF at the \( k \)th sampling instant, a discrete state space realization of the OBF network can obtained as

\[ x(k+1) = \Psi(\xi)x(k) + A(\xi)v(k) \]  
(A.4)

\[ \vartheta(k) = \theta^T x(k) \]  
(A.5)

where \( \xi = [\xi_1 \ \xi_2 \ \ldots \ \xi_n]^T \) represents the vector of filter parameter, \( \Psi(\xi) \) is an \( n \times n \) lower triangular matrix computed as

\[ \Psi(\xi) = [\Xi_1(\xi)]^{-1} \Xi_2(\xi) \]

\[ \Xi_1(\xi) = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
\xi_1 \left( \frac{n}{n-1} \right) & 1 & 0 & 0 & 0 \\
0 & \xi_2 \left( \frac{n}{n-1} \right) & 1 & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \xi_{n-1} \left( \frac{n}{n-1} \right) & 1
\end{bmatrix} \]

\[ \Xi_2(\xi) = \begin{bmatrix}
\xi_1 & 0 & 0 & 0 & 0 \\
\left( \frac{n}{n-1} \right) \xi_2 & 0 & 0 & 0 \\
0 & \left( \frac{n}{n-1} \right) \xi_3 & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \left( \frac{n}{n-1} \right) \xi_n
\end{bmatrix} \]

\[ \Gamma(\xi) \] is an \( n \) dimensional vector computed as

\[ A(\xi) = [\Xi_1(\xi)]^{-1} [1 \ 0 \ 0 \ \ldots \ 0]^T \]  
and

\[ \theta = [c_1 \ c_2 \ \ldots \ \ c_n]^T \]  
(A.6)

where elements of \( \theta \) are Fourier coefficients.

When system time delay is known a priori it is more convenient to choose the following subset of basis filters:

\[ F_i(z) = \frac{\beta_i}{(z - \xi_i)} \prod_{j=1}^{i-1} \frac{(1 - \xi_j z)}{(z - \xi_j)} z^{-d} \]  
for \( k = 1, 2, \ldots, n \)  
(A.7)

where \( d \) represents time delay. The resulting model can be written as

\[ \vartheta(k) = \left( \sum_{i=1}^{n} c_i F_i(z) \right) [z^{-d}u(k)] \]  
(A.8)

A state realization for above model can be constructed as follows:

\[ x(k+1) = \Psi_d(\xi)x(k) + A_du(k) \]  
(A.9)

\[ \vartheta(k) = \theta^T x(k) \]  
(A.10)

\[ \Psi_d(\xi) = \begin{bmatrix}
\Psi(\xi) & A(\xi) & 0 & \ldots & 0 \\
0 & \ldots & 1 & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1 \\
0 & 0 & \ldots & 0 & 0
\end{bmatrix} \]

\[ \theta_d = \begin{bmatrix}
\theta \0
\end{bmatrix} \]

where matrices \( \Psi(\xi) \), \( A(\xi) \) and \( \theta \) are matrices computed for delay free part.

A.2. MIMO Model

Consider a \( r \times m \) MIMO system and let the state realization of SISO model relating \( i \)th output with \( j \)th input be represented as

\[ x_{ij}(k+1) = \Psi_{ij} x_{ij}(k) + A_{ij} v_{ij}(k) \]  
(A.11)

\[ \vartheta_{ij}(k) = \theta_{ij} x_{ij}(k) \]  
(A.12)

where \( \Psi_{ij} = \Psi(\xi_{ij}) \) and \( A_{ij} = A(\xi_{ij}) \) are computed as described in the above section, \( \xi_{ij} \) represents the vector
of GOBF poles and $x_{ij} \in \mathbb{R}^m$ represent the state vector associated with the state realization of the SISO model. Defining an augmented state vector $X(k)$ and input vector $v(k)$ as

$$X_i(k) = \begin{bmatrix} x_{11}^T(k) & x_{12}^T(k) & \cdots & x_{1m}^T(k) \\ \end{bmatrix}^T$$  \hspace{1cm} (A.13)

$$v(k) = [v_1(k) \ldots v_m(k)]^T$$  \hspace{1cm} (A.14)

and $N_i = \sum_{j=1}^m n_{ij}$, a non-minimal MISO state realization can be constructed as

$$X_i(k+1) = \Psi_i X_i(k) + A_i v(k)$$  \hspace{1cm} (A.15)

$$\theta_i(k) = \Theta_i X_i(k)$$  \hspace{1cm} (A.16)

where

$$\Psi_i = \text{block diag}[\Psi_i \Psi_2 \ldots \Psi_m]_{N_i \times N_i}$$

$$A_i = \text{block diag}[A_{i1} A_{i2} \ldots A_{im}]_{N_i \times m}$$

$$\theta_i = \begin{bmatrix} \theta_{i1}^T & \theta_{i2}^T & \cdots & \theta_{im}^T \end{bmatrix}_{1 \times N_i}$$

for $i = 1, 2, \ldots, r$. The above set of $r$ MISO models can be further combined to formulate a non-minimal MIMO state realization of the form

$$X(k+1) = \Psi X(k) + A v(k)$$  \hspace{1cm} (A.17)

$$\theta(k) = \Theta X(k)$$  \hspace{1cm} (A.18)

where

$$X(k) = \begin{bmatrix} x_1^T(k) & x_2^T(k) & \cdots & x_m^T(k) \end{bmatrix}_{N \times 1}$$

$$\Psi = \text{block diag}[\Psi_1 \Psi_2 \ldots \Psi_m]_{N \times N}$$

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mm} \end{bmatrix}_{N \times m}$$

$$\Theta = \begin{bmatrix} \theta_1^T & \theta_2^T & \cdots & \theta_m^T \end{bmatrix}_{1 \times N}$$

$$N = \sum_{i=1}^r N_i$$

References


