Available online at www.sciencedirect.com
ScienceDirect
IFAC-PapersOnLine 49-7 (2016) 1002–1007

Parameter Estimation and Model Order Identification of LTI Systems

Santhosh Kumar Varanasi,* Phanindra Jampana*

* *Department of Chemical Engineering, Indian Institute of Technology Hyderabad, Medak, Telangana 502285, India (e-mail: ch13m15p000005@iith.ac.in,pjampana@iith.ac.in).*

Abstract: In this paper, a sparsity seeking optimization method for estimating the parameters along with the order of output error models of single-input and single-output (SISO), linear time invariant (LTI) system is proposed. It is demonstrated with the help of simulations that the proposed algorithm gives accurate parameter and order estimates on a variety of systems.

Keywords: Linear ODE, Parameter estimation, Model order identification, Sparsity seeking optimization, Regularization.

1. INTRODUCTION

It is essential to identify an accurate model before designing and tuning control algorithms. Identification of continuous-time (CT) linear time-invariant (LTI) systems from the input- output data was pursued initially but advancements in digital computers and availability of digital data acquisition schemes interested researchers to move from Continuous time to Discrete time(DT) identification and establish the theory for Discrete time identification (Ljung (1987), Soderstrom et al. (1988)). However identifying a CT model is essential because most of the systems are continuous in nature and estimating a CT model enables a better understanding of the system.

CTI methods are broadly classified into two categories, viz. Indirect method and the Direct method. In the Indirect method, firstly a discrete time model is identified from the sampled data and the identified model is then converted into continuous time. An example for such method is given in (Sinha, 2000). The main challenge in indirect identification is to find an accurate method that converts the DT model to CT. In the direct method, the major challenge lies in the numerical estimation of derivatives. If the derivatives information is readily available, the parameters can be estimated easily using least squares technique.

To estimate the derivatives from data, (Swartz et al., 1975) proposed a method of fitting a piece-wise polynomial, which smoothens the noise in the data. (Varah, 1982) used cubic splines to fit the noisy data to estimate the derivatives. The main problem with above methodologies is that the parameters estimated are biased because of the derivatives estimation based on the noisy data and no regularization term is involved. (Ramsay, 1996) used B-spline as basis functions and proposed Principle differential analysis (PDA) algorithm in which alongside the data fit, a regularization term which is a second order penalty term derived from the differential equation is introduced. (Poyton et al., 2008) demonstrated that if the data is heavily corrupted by noise, the parameter estimates from the PDA are biased because of the penalty term in PDA algorithm forces the second derivatives of the splines to be smaller than the actual second derivatives. Thus (Poyton et al., 2008) introduced a model (ordinary differential equation) based penalty term while fitting the data and proposed an iteratively refined algorithm in which iteration between the data fit and the parameter estimation step is performed until convergence.

On the other hand, compressed sensing, Lasso have become an important tool for estimating sparse vectors from linear system of equations. Recently, (Beck et al., 2013) proposed a general sparsity constrained algorithm for the non-linear case. The algorithm is guaranteed to converge to a co-ordinate wise minima but not the global optima. Being a search based algorithm the final solution depends on the initial parameter guess. The algorithm needs only sparsity of the desired solution as an input and, is advantageous compared to l_1-minimization based algorithms which require l_1 norm of the solution vector. Another advantage is that unlike in the linear case the function need not satisfy stringent conditions such as Restricted Isometry Property (RIP).

In this paper the idea of fitting B-splines to the data and penalizing it with respect to the system (ordinary differential equation) by (Poyton et al., 2008) is coupled with sparsity seeking optimization method of (Beck et al., 2013) to estimate the parameters along with the order of LTI system for both noiseless and noisy data. In the noisy case, the signal to noise ratio (SNR, defined as ratio of root mean square value of noise-free output to that of the noise) is assumed to be in between 5 and 10.

This paper is organized as follows. Section 2 gives the formulation of the problem, introduces the concept of co-ordinate wise minima and briefly discusses B-splines. Section 3 describes the objective function and the optimization procedure. Section 4 provides numerical results whereas Section 5 gives the concluding remarks.

10.1016/j.ifacol.2016.07.333

2. FORMULATION

2.1 Problem statement

Consider a continuous-time (CT) linear time-invariant (LTI) system described by

$$a_n \frac{d^n x}{dt^n} + a_{n-1}\frac{d^{n-1}x}{dt^{n-1}} + \dots + a_1 \frac{dx}{dt} + a_0 x(t) = b_m \frac{d^m u}{dt^m} + b_{m-1}\frac{d^{m-1}u}{dt^m-1} + \dots + u(t) \quad (1)$$

A noise corrupted version of x is measured,

$$y(t_k) = x(t_k) + \epsilon(t_k),\ k = 1, 2, \dots, M$$

where, $\epsilon(t_k)$ is Gaussian noise with mean 0 and variance σ^2, M is the number of samples taken.

The goal of identification is to estimate the parameters $a_n, a_{n-1}, \dots, a_1, a_0, b_m, b_{m-1}, \dots, b_1$ from the input-output data i.e., $\{(u(t_k), y(t_k); k = 1, 2, \dots, M\}$, along with the output order n. Throughout the paper it is assumed that the input is known and the parameters are denoted with the θ vector i.e.,
$\theta = (a_n, a_{n-1}, \dots, a_1, a_0, b_m, b_{m-1}, \dots, b_1)$

2.2 Overview of estimation procedure

As explained earlier, only the output is a measured variable and the derivative values are unmeasured, to estimate the derivatives, the output data is to be fitted first. Thus, two sets of parameters are to be estimated simultaneously. First set, the fitting parameters and second set, the actual parameters i.e., the θ vector. In the first step, alongside the data fit, a regularization term involving the fit to ODE is also included.

In the second step, using the information of noise free output and its derivatives from the first step, the parameter vector is to be estimated. As the order of the system is also an unknown quantity, the parameters of model have to be estimated in such a way that the zero norm of the parameter vector (the number of non-zero entries in a vector) is less than s i.e., $||\theta||_0 \leq s$ where s is sparse index (in the proposed algorithm it is greater than the assumed order of system). Before giving the detailed description of algorithm, some notations and definitions are fixed.

2.3 Notations and Definitions

Notations: For a given vector, $x \in \mathbb{R}^n$, and the function $f(x)$, the support set of x is defined as

$$S(x) \equiv \{i : x_i \neq 0\}$$

and its compliment is given by

$$S_c(x) \equiv \{i : x_i = 0\}$$

The set of vectors which are at most $s-$sparse is denoted by

$$C_s = \{x : ||x||_0 \leq s\}$$

Definitions:

Basic Feasible Vector: A vector, $x^* \in C_s$ is called a *basic feasible* vector of the function $f(x)$, if

1. when $||x^*||_0 < s, \bigtriangledown f(x^*) = 0$
2. when $||x^*||_0 = s, \bigtriangledown_i f(x^*) = 0\ \forall\ i \in S_s(x^*)$

Coordinate-wise Minima: Let x^* be a feasible solution of function $f(x)$, then it is called a Coordinate-wise Minima if

1. $||x^*||_0 < s$ and $i = 1, 2, \dots, n,\ f(x^*) = \min_{t\in\mathbb{R}} f(x^* + te_i)$
2. $||x^*||_0 = s$ and for every $i \in S_s(x^*)$ and $j = 1, 2, \dots, n,\ f(x^*) \leq \min_{t\in\mathbb{R}} f(x^* - x_i^* e_i + te_j)$

In the hierarchy of estimation, any optimal point is a CW-Minima and any CW-Minima is a Basic Feasible vector (Beck et al., 2013). The same authors give an algorithm called the Greedy Sparse Algorithm, which computes a CW-minima.

B-splines and its Construction

A B-spline is a piece wise polynomial of some defined order fitted between the data points. An interesting property of B-splines is that they are local i.e., if a few data points were changed, the splines related to that data point will change and the splines corresponding to unchanged data will remain the same which enables us to get local control.

Let U be a set of $m+1$ non-decreasing sequence of numbers of form u_i, called the knots and P be a vector of control points of the form P_i then the general n^{th} order spline is expressed as a recursive relation of the form

$$N_{i,p}(t) = \frac{t - u_i}{u_{i+p} - u_i} N_{i,p-1}(t) + \frac{u_{i+p+1} - t}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(t)$$

where

$$N_{i,0}(t) = \begin{cases} 1, & \text{if } u_i \leq t < u_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

i indicates the position of the spline and p indicates the order of the spline to fit. and the curve is defined by

$$C(t) = \sum_{i=0}^{L} P_i N_{i,p}(t)$$

where L is the number of control points considered. The Derivative of the B-spline curve is given by the recursive expression

$$\frac{dN_{i,p}}{dt} = \frac{p}{u_{i+p} - u_i} N_{i,p-1} + \frac{p}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}$$

2.4 Detailed Description of Algorithm

As explained in the overview subsection, the overall algorithm comprises of two steps. The first step in which the data is fitted subject to the system and the second step, in which the parameters are estimated with the constraint on the sparsity of the parameter vector.

Expressing mathematically, in the step-1, objective is to minimize

$$H = f + \lambda g \quad (2)$$

with respect to the spline coefficients P_i, where

$$f = ||y(t_k) - \hat{Y}(t_k)||^2$$

$$g = \int_0^T \left[a_n \frac{d^n \hat{Y}}{dt^n} + \dots + a_o \hat{Y} - \left(b_m \frac{d^m u}{dt^m} + \dots + u(t) \right) \right]^2 dt$$

and

$$\hat{Y}(t_k) = \sum_{i=0}^{L} P_i N_{i,p}(t)$$

λ is the weighting factor which plays a major role in estimation of the spline parameters especially in the noisy cases so it is essential to understand the optimal value of λ or to have some control over λ. Selection of optimal value of lambda was given in (Varziri et al., 2008). Applying the same approach, λ_{opt} can be obtained as sum of variance of $(y - \hat{Y})$ and mean of $(y - \hat{Y})^2$. L is the number of knot points to be taken. As spline fit highly depends on the knot point selection and order of the spline to be fitted, the order of the spline is taken as one greater than the order of system to be observed and knot points are selected on a trail and error basis in such a way that there are no sharp peaks while fitting and maximum possible information of data is captured by the fit.

In the second step, the objective function is minimized with respect to the parameters alone and with the spline coefficients that are obtained from the step-1 such that $||\theta||_0 \leq s$ where s can be chosen to be greater than or equal to the order of system, one would like to identify. Thus the objective function in step-2 is

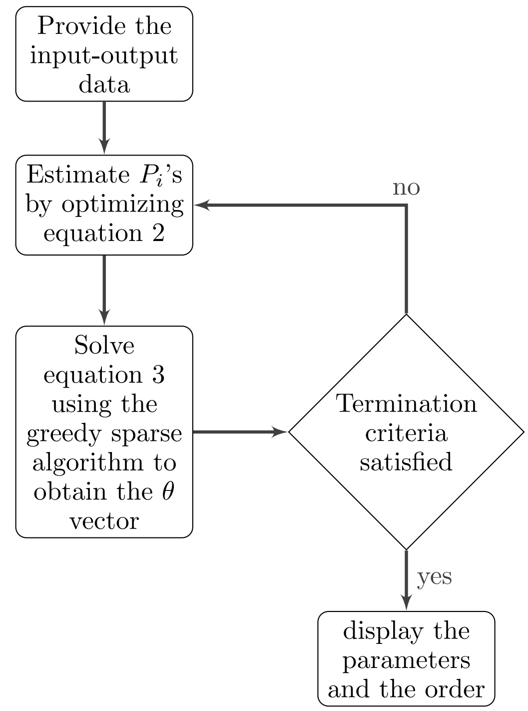
$$\min_{||\theta||_0 < s} H$$

As the function f is independent of parameters in the second step, minimizing the above function is equivalent to

$$\min_{||\theta||_0 < s} g \tag{3}$$

The overall algorithm is as follows. Firstly an initial guess of actual parameters is provided in step-1 and the spline coefficients are obtained. Using the spline coefficients from the step-1, the parameter vector, θ is estimated from step-2. This procedure is repeated until the termination criteria is satisfied (error between the parameter vectors obtained in two successive iterations is small enough or the function value H with estimated spline coefficients and the parameter vector should be minimum). In the proposed algorithm, the termination criteria is selected to be the error between the parameters be less than 0.005.

As the objective function in step-2 is non-linear, one cannot use standard sparsity seeking optimization methods such as LASSO, Basis Pursuit. Thus the Greedy Sparse (GS) algorithm given in (Beck et al., 2013) is used, which needs a bound on the sparsity an the initial guess. Selection of initial values is described in the later part and the overall algorithm is explained here.



The assumptions to apply GS algorithm are

i) The objective function should be continuously differentiable
ii) The objective function is lower bounded

$g(\theta)$ is continuously differentiable and is lower bounded so one can apply GS algorithm which is guaranteed to converge to a CW-minima (Beck et al., 2013). The Greedy Sparse algorithm is as follows.

- Choose $\theta_0 \in C_s$ and $k = 0, 1, \ldots$
- If $||\theta^k||_0 < s$, for every i=1,2,...,n
  $$t_i \in \arg\min_{t \in \mathbb{R}} g(\theta^k + te_i)$$
  $$g_i = \min_{t \in \mathbb{R}} g(\theta^k + te_i)$$
  where e_i is the vector whose i^{th} component is one and remaining all are zeros.
  let $i_k \in \arg \min_{i=1,2,\ldots,n} g_i$ and if $g_{i_k} < g(\theta^k)$
  $$\theta^{k+1} = \theta^k + t_{i_k} e_{i_k}$$
  else return the parameter vector θ and terminate the algorithm.
- If $||\theta^k||_0 = s$ for each $i \in S_s(\theta^k), j = 1, 2, \ldots, n$
  $$t_{i,j} \in \arg\min_{t \in \mathbb{R}} g(\theta^k - \theta_i^k e_i + te_j)$$
  $$g_{i,j} = \min_{t \in \mathbb{R}} g(\theta^k - \theta_i^k e_i + te_j)$$
  $(i_k, j_k) \in \arg\min g_{i,j}$, $i \in S_s(\theta^k), j = 1, 2, \ldots, n$ and
  If $g_{i_k j_k} < g(\theta^k)$;
  $$\theta^{k+1} = \theta^k - \theta_{i_k}^k e_{i_k} + t_{i_k, j_k} e_{i_k}$$
  else return the parameter vector θ and terminate the algorithm.

As Greedy sparse algorithm is a search based algorithm, one needs a good initial guess to start with otherwise algorithm converges to the CW minima which might not be close to the true solution. Thus, a way to obtain an initial guess is proposed. After step-1 in the actual algorithm with lambda a very small quantity (preferably zero for noiseless case and a non-zero value in noisy case this is because λ is chosen based on measurement uncertainties and in the noisy case the measurements are more uncertain), equation 3 is solved for parameter vector θ without sparsity constraint. Thus, the identification problem now becomes a least squares problem and the solution can be easily obtained. After obtaining the solution vector θ, the elements with the negative values are made zero to assure the stability of system and also if any values are close to zero ($\frac{1}{100}^{th}$ of the maximum value of the estimated parameter vector) these too are made zero to ensure proper order estimation. Thus the dense θ vector has now became sparse, which was considered as an initial guess to solve step-2 with sparsity constraint. An important aspect of the aforementioned algorithm is that one can obtain the actual solution even if the actual parameters of the solution become zero in the initial guess, which is not the case with l_1 norm minimization.

In the comparison section, the results obtained by the proposed algorithm are compared with the l_1 norm minimization algorithm (Basis Pursuit Denoising (BPDN) algorithm, where in $||x||_1$ is minimized subject to $||Ax - b||_2 \leq \sigma$, where $\sigma = ||Ax_{ls} - b|| + \gamma$, $\gamma > 0$) as explained in (Jampana et al., 2013).

3. NUMERICAL EXPERIMENTS

The proposed algorithm is experimented on a variety of systems and in all the following experiments, the input is a multi-frequency sine wave with frequency range in between 0.1 and 1.0 and the input order (m) is assumed to be known.

3.1 System-1

Let the system-1 is a first order system as

$$\frac{dx}{dt} + 2x = u(t)$$

The system has been simulated in simulink to obtain $u(t_k)$ and $x(t_k)$ and Gaussian noise $\epsilon(t_k)$ is added to $x(t_k)$ to obtain the output data, $y(t_k)$, which is the output. Input-output data is provided to the algorithm and in the general case one does not know the order, one can specify a high order as n value and estimate the parameters. In the following simulations, the order is varied as $n = 1$(the actual order), $n = 3$ and $n = 5$. For the noisy case, the variances of noise is varied in between 5 and 10 and a histogram of parameters are plotted. The mean along with the standard deviation are reported in the following tabulated results.

Table 1. First order Noiseless Case

Actual parameters	n=1	n=3	n=5
			a5=0
		a3=0	a4=0
a1=1	a1=0.9998	a2=0	a3=0
a0=2	a0=2.0001	a1=0.9621	a2=0
		a0=1.99	a1=0.9983
			a0= 1.997

Table 2. First order Noisy Case with SNR varied in between 5 and 10

Actual	n=1	n=3	n=5
			a5=0
		a3=0	a4=0
a1=1	a1=0.96±0.01	a2=0	a3=0
a0=2	a0=2±0.001	a1=1.02±0.02	a2=0
		a0=2.02±0.06	a1=0.95±0.02
			a0= 1.961±0.04

From the above results, it can be said that for first order systems, even with the higher order assumption of order, the identification method satisfactorily approximates the true system.

3.2 System-2

System-2 is the same first order system with additional input derivative term

$$\frac{dx}{dt} + 2x = \frac{du}{dt} + u(t)$$

The data is obtained using the same procedure mentioned in system-1. The solution from the algorithm is as follows

Table 3. First order Noiseless Case

Actual parameters	n=1	n=3	n=5
			a5=0
		a3=0	a4=0
a1=1	a1=1.05	a2=0	a3=0
a0=2	a0=2.0006	a1=1.026	a2=0
b1=1	b1=1.028	a0=1.999	a1=1.072
		b1=1.019	a0= 1.993
			b1=1.059

Table 4. First order Noisy Case with SNR varied in between 5 and 10

Actual parameters	n=1	n=3	n=5
			a5=0
		a3=0	a4=0
a1=1	a1=1.01±0.03	a2=0	a3=0
a0=2	a0=1.99±0.04	a1=0.99±0.05	a2=0
b1=1	b1=1.07±0.02	a0=1.99±0.02	a1=0.92±0.02
		b1=1.02±0.02	a0= 1.98±0.03
			b1=0.96±0.02

3.3 System-3

Now, consider a second order system

$$\frac{d^2x}{dt^2} + 2.8\frac{dx}{dt} + x = u(t)$$

Applying the identification algorithm after obtaining the input-output data, the results are as follows.

For a second order system also it can be said that the identification method approximates the true system along with correct order estimation.

Table 5. Second order Noiseless case

Actual parameters	n=2	n=5
		a5=0
a2=1	a2=0.9898	a4=0
a1=2.8	a1=2.801	a3=0
a0=1	a0=0.9988	a2=1.001
		a1=2.769
		a0= 1.003

Table 6. Second order Noisy Case with SNR varied from 10 to 5

Actual parameters	n=2	n=5
		a5=0
a2=1	a2=0.93±0.01	a4=0
a1=2.8	a1=2.83±0.04	a3=0
a0=1	a0=0.99±0.001	a2=0.85±0.04
		a1=2.73±0.02
		a0=0.97±0.005

3.4 System-4

A fourth order system of the form

$$\frac{d^4x}{dt^4} + 10\frac{d^3x}{dt^3} + 35\frac{d^2x}{dt^2} + 50\frac{dx}{dt} + 24x = u(t)$$

is considered as our third system and after obtaining the data by the procedure mentioned above, identification algorithm is applied with $n = 6$ and the results are tabulated below.

Table 7. Fourth order Noiseless Case

Actual parameters	n=4	n=6
		a6=0
a4=1	a4=0.9581	a5=0
a3=10	a3=10.02	a4=0.93
a2=35	a2=34.94	a3=9.86
a1=50	a1=50.01	a2=34.75
a0=24	a0=23.99	a1=49.93
		a0=23.982

Table 8. Fourth order Noisy Case with SNR=8

Actual parameters	least squares solution	Initial guess sparse algorithm	final solution
	a6=-3.492	a6=0	a6=0
a4=1	a5=-0.3713	a5=0	a5=0
a3=10	a4=-6.37	a4=0	a4=0.9659
a2=35	a3=10.69	a3=10.69	a3=10.69
a1=50	a2=31.79	a2=31.79	a2=31.79
a0=24	a1=49.62	a1=49.62	a1=49.70
	a0=24.08	a0=24.08	a0=24.08

From Table 8, it can be seen that the initial guess to the sparse algorithm has the a_4 coefficient set to zero, however this is recovered back by the sparse algorithm.

With the same fourth order system, and with first order input derivative term i.e., $\frac{d^4x}{dt^4} + 10\frac{d^3x}{dt^3} + 35\frac{d^2x}{dt^2} + 50\frac{dx}{dt} + 24x = 2\frac{du}{dt} + u(t)$, the solution obtained from the algorithm is $\theta = (0, 0.9746,\ 9.28,\ 33.34,\ 47.34,\ 23.96,\ 1.864)$ with $n = 5$ and $m = 1$.

4. COMPARISON

The proposed method is compared with the method of solving for sparse solutions given in (Jampana et al., 2013), where l_1 norm minimization is done using Basis Pursuit Denoising(BPDN) algorithm. In the following table, $x_{(1s,\gamma)}$ is the result of BPDN algorithm on System-4 where γ is the parameter which is selected on a trail and error basis.

Table 9. Comparison of proposed method with the BPDN algorithm of fourth order system in noiseless case

coefficient	Proposed	$x_{(1s,10)}$	$x_{(1s,5)}$
a6	0	0	0.0603
a5	0	0	0
a4	0.93	0.3155	0.8918
a3	9.86	8.6773	9.3467
a2	34.75	33.6388	34.3501
a1	49.93	49.0486	49.7257
a0	23.982	22.4575	23.0507

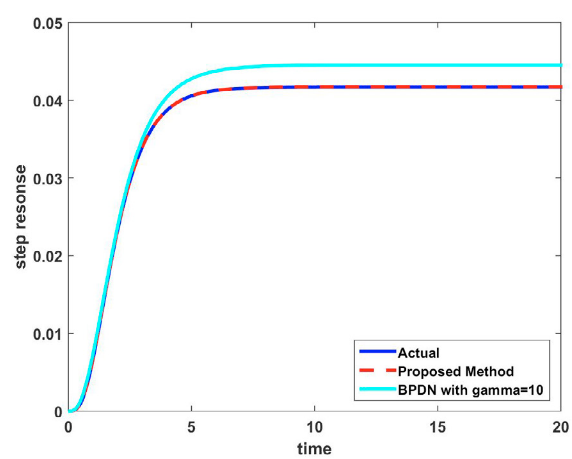


Fig. 1. Comparison of step response of proposed method with the actual system and with BPDN algorithm

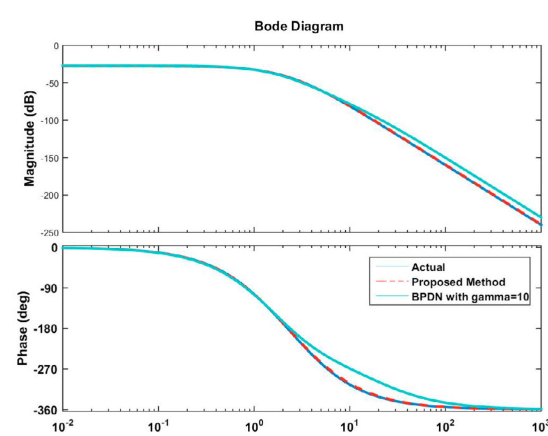


Fig. 2. Comparison of Frequency response of proposed method with the actual system and with BPDN algorithm

1-norm minimization with γ=10 predicted the order correctly but the error between the actual parameters and the parameters estimated are very high when compared with our proposed method and reducing the value of γ results in error in model order estimation as one can check the results given in the table 9 with $\gamma = 5$.

Table 10. Comparison of proposed method with the BPDN algorithm of fourth order system in noisy case

coefficient	Proposed	$x_{(1s,10)}$	$x_{(1s,5)}$
a6	0	0	0.0556
a5	0	0	0
a4	0.9659	0.2894	0.8650
a3	10.69	8.7057	9.3447
a2	31.79	33.6113	34.3223
a1	49.78	49.0695	49.7467
a0	24.08	22.4461	23.0387

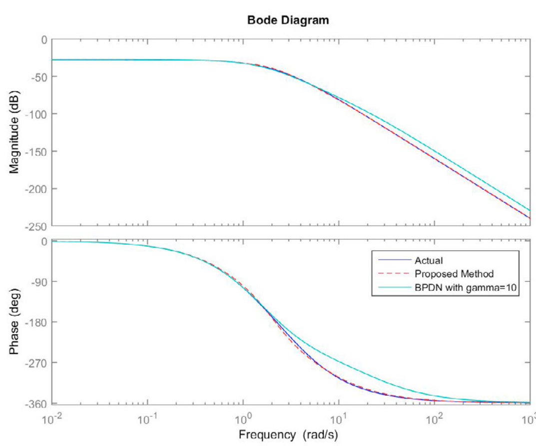


Fig. 3. Comparison of Frequency response of proposed method with the actual system and with BPDN algorithm in noisy case

From Figure 2 and Figure 3 it can be seen that the Bode plot of the proposed method is closer to the original system when compared with the BPDN algorithm.

5. CONCLUSIONS

In this paper, a new algorithm to estimate the parameters as well as order of the system (differential equation) simultaneously is presented. With the help of the simulations, it is successfully shown that the algorithm works well with the aforementioned way of selecting the initial values, and the weighting factor (λ) value. However for higher order systems, the error between the actual and the estimated parameters are considerable and one reason may be error in estimation of derivatives based on the spline fit to the data, which can be reduced upon increasing the number of knot points. If one has a huge data, placing a knot point at each time resolves the issue but at the cost of heavy computation. Thus one has to balance between the number of knots which implicitly influences the results and the computation time. Generalizing the selection of knots will be pursued in the future.

In this paper, the input is assumed to be known. The aforementioned algorithm can be extended to a general case where one does not have any information of input function but only has a set of input data. In such cases, one more additional step of fitting splines to the input data is required and this work will also be pursued in the future.

REFERENCES

Beck, A., and Eldar, Y.C. (2013). Sparsity constrained nonlinear optimization: Optimality conditions and algorithms. *SIAM Journal on Optimization*, volume 23, pages 1480–1509.

Garnier, H., Mensler, M., Richard, A. (2003). Continuous-time model identification from sampled data: implementation issues and performance evaluation. *International Journal of Control*, volume 76, pages 1337–1357.

Jampana, P., Detroja, K. (2013). Continuous Time Identification in Laplace Domain *Dynamics and Control of Process Systems*, volume 10, pages 367–372.

Ljung, L. (1987). *System Identification: Theory for the user.* PTR Prentice-Hall Information and System Sciences Series. Pearson Education Canada.

Poyton, A.A., Varziri, M.S., McAuley, K.B., McLellan,P.J., Ramsay,J.O. (2006). Parameter estimation in continuous-time dynamic models using principal differential analysis. *Computers & Chemical Engineering*, Elsevier, volume 30, pages 698–708.

Ramsay, J.O. (1996). Principal Differential Analysis. *Encyclopedia of Statistical Sciences.*

Sinha, N.K. (2000). Identification of continuous-time systems from samples of input-output data: An introduction. *Sadhana*, volume 25, pages 75–83.

Söderström, T., and Stoica, P. (1988). *System identification.* Prentice-Hall, Inc.

Swartz, J., and Bremermann, H. (1975). Discussion of parameter estimation in biological modelling: algorithms for estimation and evaluation of the estimates. *Journal of Mathematical Biology*, volume 1, pages 241–257.

Unbehauen, H., and Rao, G.P. (1998). A review of identification in continuous-time systems. *Annual reviews in Control*, volume 22, pages 145–171.

Varah, J.M. (1982). A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing*, volume 3, pages 28–46.

Varziri, M.S., Poyton, A.A., McAuley, K.B., McLellan, P.J., Ramsay, J.O. (2008). Selecting optimal weighting factors in iPDA for parameter estimation in continuous-time dynamic models. *Computers & Chemical Engineering*, volume 32, pages 3011–3022.

Young, P. (1981). Parameter estimation for continuous-time modelsa survey. *Automatica*, volume 17, pages 23–29.