Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-1 (2018) 542–547

Identification of FIR Models for LTI Multiscale Systems using Sparse Optimization Techniques

Vivek S. Pinnamaraju * **Arun K. Tangirala** **

* *Indian Institute of Technology Madras, Chennai - 600036, India (e-mail: vivekshankar1992@gmail.com)*
** *Indian Institute of Technology Madras, Chennai - 600036, India (e-mail: arunkt@gmail.com)*

Abstract: Finite impulse response (FIR) models are very popular in process industries because of their simple model structure, flexibility to explain arbitrary complex stable linear dynamics and finally their ease of implementation in on-line applications. In general, identification of FIR models requires large number of parameters to be estimated. In case of systems with multiple time scales, the length of FIR model structure under conventional uniform sampling becomes arbitrarily high due to simultaneous presence of fast and slow dynamics. This results in more variability in the estimated parameters when the conventional methods such as ordinary least squares are used. In this work, the FIR model estimation problem is formulated as a sparse optimization problem, where the sparse representation of impulse response coefficients for linear-time invariant multiscale systems in the time-frequency domain is exploited in order to explain the overall FIR model effectively with fewer number of coefficients and thereby incurring less variability in the estimated parameters. The effectiveness of proposed methodology is demonstrated by means of simulation case studies.

Keywords: Multiscale, Linear Time-Invariant, FIR Models, System Identification, Parsimony, Sparse Optimization, Wavelets

1. INTRODUCTION

Identification of linear-time-invariant (LTI) systems from input-output data is an important problem in several process operations such as control. There are several linear model structures which can explain LTI systems. These are comprehensively described in Ljung (1999). Among these model structures, FIR models are very popular because of their simple model structures and their flexibility to model linear systems with any complexity. Despite their several advantages, they are non-parsimonious and require large number of parameters to be estimated. Under uniform sampling, this problem becomes very acute in the case of systems with multiple time scales as the model structure becomes quite large in order to simultaneously accommodate both fast and slow dynamics of the process. Existing methods such as ordinary least squares yield large errors in the parameter estimates because of large parameter size. In this work, we exploit the sparse representations of IR coefficients (of multiscale systems) in time-frequency domain and formulate a sparse optimization problem in order to improve the overall estimates of the FIR coefficients for multiscale processes.

One of the widely known method for estimating the FIR coefficients is ordinary least squares (OLS). The estimates of OLS can be further improved by invoking additional constraints on parameters. This is generally known as regularized least squares (RLS). If the constraint is minimization of magnitude of IR coefficients, it is known as ridge regression (RR). The smoothness of the estimated coefficients can be achieved by imposing a constraint that minimizes the distance between the consecutive IR coefficients (RRD). Further, one can incorporate weights for the rate of decay of the change in coefficients. The details of all these methods and their robustness issues are discussed in Dayal and MacGregor (1996). Some of the works utilized the filter bank aspects of wavelets in order to improve the estimates of FIR coefficients (Nikolaou and Vuthandam, 1998; Nounou, 2006).

In Nikolaou and Vuthandam (1998), discrete wavelet transform on FIR model coefficients is performed and a suitable threshold is used in order to discard the insignificant coefficients and finally a few coefficients are retained after compression. By applying inverse DWT, the FIR model coefficients are recovered. The method is shown to provide superior results when compared to the existing methods but only issue seems to be that the parameter estimation and parsimony are not achieved jointly. In Nounou (2006), scaling and wavelet filters are used for decomposing the input-output data at different scales. The FIR models are built between the approximation projections of input-output data at different scales and the optimal scale is chosen where the signal-to-noise ratio (SNR) of the predicted model output is maximum. Despite the advantages of this method such as joint parameter estimation and reduction in effective length of FIR model coefficients, the method considers only the scaled approximations of input-output data at various dyadic levels while

2405-8963 © 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.05.091

the details are completely neglected. In case of systems with multiple time scales, the fast dynamics features are generally present in the detail space of input-output decomposition and application of this method captures only the steady state information of fast subsystem while the dynamic features are completely neglected as they are present in the detail projections of the signal.

In the present work, the FIR model parameter estimation and parsimony are simultaneously obtained by resorting to spare optimization techniques. This is achieved by exploiting the sparse nature of FIR coefficients in the time-frequency domain. The main contributions of present work include recognizing the sparse representation of FIR coefficients in an appropriate domain and formulation of the sparse optimization problem. The paper is organized as follows. The problem of FIR identification is mathematically described in Section 2. The preliminaries required for the work and some background on existing methods for FIR identification are provided in Section 3. The proposed methodology is formulated in Section 4 and is implemented on two simulation case studies, the results of which are discussed in Section 5. Finally, the paper ends with conclusive remarks in Section 6.

2. PROBLEM STATEMENT

Consider a general discrete (discretized based on ZOH) linear time-invariant SISO system as follows:

$$y[k] = G_0(q^{-1})u[k] + e[k] \tag{1}$$

where $u[k], y[k]$ are the input and output measurements respectively at time instant k; $G_0(q^{-1})$ is the transfer function operator and the output is corrupted by the white noise $e[k]$ which is independent of $u[k]$. For stable systems, the transfer function can be equivalently described as

$$G_0(q^{-1}) = \sum_{n=1}^{\infty} g_0[n]q^{-n} \tag{2}$$

In case of stable systems, the impulse response decays to zero in a finite length of time and hence it is enough to truncate the infinite impulse response to certain order M and estimate those finite number of impulse response coefficients. The approximated system's transfer function and the corresponding FIR model are described below. A unit delay is inherently assumed in the transfer function by virtue of ZOH discretization.

$$G(q^{-1}) = \sum_{n=1}^{M} g[n]q^{-n}, \;\; \mathbf{g} = [g[1] \;\; g[2] \;\; \cdots \;\; g[M]] \tag{3}$$

$$y[k] = \sum_{n=1}^{M} g[n]u[k-n] + e[k] \tag{4}$$

When a uniform sampling scheme is adopted, the sampling time has to be chosen based on the fast subsystem. In such a case, the discrete-time slow subsystem poles are very close to the unit circle. This results in large number of IR coefficients as the poles of the slower subsystem requires larger number of time samples in order to decay to zero. As a result, a very large number of impulse response coefficients are required to be estimated in case of systems with multiple time scales unlike the single scale systems where the number of coefficients are relatively very small.

It is well known that larger the parameter size, larger will be the variability in the estimated parameters. Model predictions obtained from such parameters are likely to be very poor. The overall objective of this paper is to simultaneously achieve good parameter estimates and parsimonious representations of FIR models in the case of LTI systems with multiple time scales.

3. PRELIMINARIES

3.1 LTI Multiple-time scale systems

A continuous-time LTI system is said to posses two-time scale behaviour when some of the poles of the transfer function are much smaller compared to the remaining. If the real part of the poles of a continuous-time transfer function are arranged in a descending order as $\lambda_1 \geq ... \geq \lambda_i \geq \lambda_{i+1} \geq ... \geq \lambda_n$ and if there exist $Re(\lambda_i) >> Re(\lambda_{i+1})$, then the subsystem corresponding to first i set of poles evolves much faster than the remaining poles. The equivalence in the case of discrete-time systems is that some set of poles are far from the unit circle in $z-$plane while others are very close to the unit circle.

3.2 Sparse representation of MS FIR coefficients

The impulse response of multiscale systems in general requires very large number of samples in order to decay to zero. The slow evolution of the slow subsystem dynamics is responsible for the large number of FIR coefficients. Despite the large lengths in time domain, the slow subsystem can be sparsely represented in the frequency-domain because of its relatively small bandwidth in comparison to the overall bandwidth. On the other hand, the fast subsystem has sparse representation of FIR coefficients in the time domain as they decay to zero within small time scales but has has very large bandwidths in the frequency domain. FIR coefficients of the overall system which is combination of both slow and fast subsystems is neither sparse in time nor in frequency domain. But the sparsity can be achieved in time-frequency domain or a time-scale domain. For instance, by using time-frequency tools such as wavelets it is possible to represent the FIR coefficients of multi-time scale systems sparsely in the time-scale domain. In order to demonstrate this, we consider the following continuous-time two-time scale system where the fast subsystem evolves 20 times faster than the slower subsystem.

$$G(s) = \frac{10(90s+1)}{(200s+1)(10s+1)} \tag{5}$$

When a sampling interval of $T_s = 1$ sec is used, the ZOH based discretized system is obtained as follows:

$$G_d = \frac{z^{-1}(0.4296 - 0.4248z^{-1})}{1 - 1.9z{-1} + 0.9003z^{-2}} \tag{6}$$

The discrete wavelet transform of impulse response of the above system is shown in Figure 1.

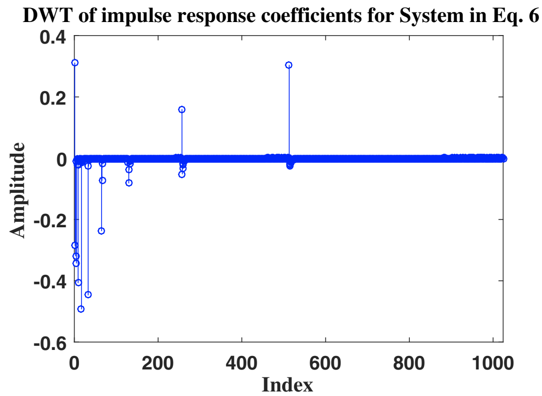

Fig. 1. Sparse FIR coefficients representation using DWT

It can be observed from Figure 1 that the impulse response coefficients can be sparsely represented in the wavelet domain. In this work, this characteristic of impulse response is exploited and a sparse optimization problem is formulated in order to estimate the FIR model parameters of LTI multiscale systems. The wavelet transformation matrix which computes the discrete wavelet transform is briefly described in Section 3.3.

3.3 Wavelet Transform Matrix

The matrix representation of discrete wavelet transform of any signal $\mathbf{z}$ is as follows:

$$\begin{pmatrix} \mathbf{a}_{\mathbf{J}_{dec}} \\ \mathbf{d}_{\mathbf{J}_{dec}} \\ \vdots \\ \mathbf{d}_1 \end{pmatrix} = (\mathbf{W}_J \mathbf{W}_{J-1} \cdots \mathbf{W}_1)\mathbf{z} = \mathbf{W}\mathbf{z} \tag{7}$$

where,

$$\mathbf{W}_k = \begin{bmatrix} \mathbf{W}_1^{N/2^{k-1}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N-N/2^{k-1}} \end{bmatrix} \tag{8}$$

and,

$$\mathbf{W}_1^{N/2^{k-1}} = \begin{bmatrix} \mathbf{W}_{1,A} \\ \hline \mathbf{W}_{1,D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi} \\ R_2\boldsymbol{\phi} \\ \vdots \\ R_{N/2^{k-2}}\boldsymbol{\phi} \\ \hline \boldsymbol{\psi} \\ R_2\boldsymbol{\psi} \\ \vdots \\ R_{N/2^{k-2}}\boldsymbol{\psi} \end{bmatrix}_{N/2^{k-1}\times N/2^{k-1}} \tag{9}$$

Here $\mathbf{W}$ is the wavelet transformation matrix for maximum level of decomposition $\mathbf{J}_{dec}$; $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ are scaling and wavelet filters respectively. The operator R_m generates the time shifted versions of wavelet and scaling filters. $\mathbf{a}_{J_{dec}}$ is the approximation coefficient at the maximum level of decomposition while $\mathbf{d}_n$ denotes the detail coefficients at different levels of decomposition.

3.4 FIR model Identification

The FIR model as shown in (4) is expressed in the regression matrix form (10) as shown below:

$$\underbrace{\begin{bmatrix} y[M+1] \\ y[M+2] \\ \vdots \\ y[N-1] \\ y[N] \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} u[M] & u[M-1] & \cdots & u[1] \\ u[M+1] & u[M] & \cdots & u[2] \\ \vdots & \vdots & \ddots & \vdots \\ u[N-2] & \cdots & \cdots & u[N-M-1] \\ u[N-1] & u[N-2] & \cdots & u[N-M] \end{bmatrix}}_{\mathbf{U}} \underbrace{\begin{bmatrix} g[1] \\ g[2] \\ \vdots \\ g[M] \end{bmatrix}}_{\mathbf{g}} \tag{10}$$

Some of the well known methods for estimation of FIR model parameters are briefly discussed below.

Ordinary Least Squares: The FIR coefficients, $\mathbf{g}$ can be estimated by solving the standard least squares problem as given below:

$$\min_{\mathbf{g}} \ (\mathbf{y}-\mathbf{U}\mathbf{g})^T(\mathbf{y}-\mathbf{U}\mathbf{g}) \tag{11}$$

The solution to the above problem is as follows:

$$\hat{\mathbf{g}}_{OLS} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y} \tag{12}$$

Ridge Regression: A constraint on the magnitude of FIR coefficients is imposed along with minimization of the sum square of the residuals in the case of ridge regression. The objective function for this case is as follows:

$$\min_{\mathbf{g}} \ (\mathbf{y}-\mathbf{U}\mathbf{g})^T(\mathbf{y}-\mathbf{U}\mathbf{g}) + \alpha_1\mathbf{g}^T\mathbf{g} \tag{13}$$

where α is the non-negative parameter which may be optimally tuned based on cross validation. The ridge regression coefficient estimates are given by,

$$\hat{\mathbf{g}}_{RR} = (\mathbf{U}^T\mathbf{U} + \alpha_1\mathbf{I})^{-1}\mathbf{U}^T\mathbf{y} \tag{14}$$

Unlike ordinary least squares, the rigde regression results in biased estimates for the FIR coefficients but with a lower variability.

Regularization with constraints on change in coefficients: In most of the processes, the IR coefficients change in a smooth manner. This can be achieved by penalizing the distance between the consecutive FIR coefficients. The objective function for this case is given below:

$$\min_{\mathbf{g}} \ (\mathbf{y}-\mathbf{U}\mathbf{g})^T(\mathbf{y}-\mathbf{U}\mathbf{g}) + \alpha_2\mathbf{g}^T\mathbf{H}\mathbf{g} \tag{15}$$

The solution for the above formulation is as follows:

$$\hat{\mathbf{g}}_{RRD} = (\mathbf{U}^T\mathbf{U} + \alpha_2\mathbf{H})^{-1}\mathbf{U}^T\mathbf{y} \tag{16}$$

where, $\mathbf{H} = \mathbf{A}^T\mathbf{A}$ with

$$A = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix} \tag{17}$$

FIR coefficients estimated from ARX models: Instead of identifying the FIR coefficients directly, they can be indirectly estimated by first constructing suitable ARX models and then estimating the FIR coefficients. This indirect identification leads to an advantage such as parsimony and hence a lower variability in the estimated parameters.

Unfortunately, in the case of systems with multiple-time scales, identification of parametric models such as ARX results in ill-conditioning at the stage of parameter estimation (Vána and Preisig, 2012) due to presence of slow subsystem poles very close to the unit circle and hence it is advisable not to estimate FIR coefficients using this method.

4. PROBLEM FORMULATION

As discussed in Section 3.2, the FIR coefficients, $\mathbf{g}$ have a sparse representation in the time-frequency domain. Let $\mathbf{W}$ be the discrete wavelet transform matrix that transforms the discrete signal from time domain to time-frequency domain, then the sparse representation of the FIR coefficients can be obtained as follows:

$$\mathbf{x} = \mathbf{W}\mathbf{g} \tag{18}$$

Here $\mathbf{x}$ is a sparse vector with few non-zero coefficients. Now $\mathbf{g}$ can be written in terms of $\mathbf{x}$ as follows:

$$\mathbf{g} = \mathbf{W}^{-1}\mathbf{x} \tag{19}$$

In the case of orthonormal wavelets, $\mathbf{T} = \mathbf{W}^{-1} = \mathbf{W}^T$. Now, (10) can be written as,

$$\mathbf{y} = \mathbf{U}(\mathbf{T}\mathbf{x}) \equiv \mathbf{A}\mathbf{x} \tag{20}$$

The objective is to simultaneously minimize the prediction error along with the sparse constraint on $\mathbf{x}$. The sparsity of $\mathbf{x}$ can be achieved by minimizing $||\mathbf{x}||_0$ but unfortunately l_0-norm minimization is regarded as NP-hard problem which is too complex and impossible to solve. The l_0-norm minimization can be relaxed to l_1-norm minimization in order to simplify the optimization under certain conditions. The least absolute shrinkage and selection operator (LASSO) method proposed by Tibshirani (1996) is used in order to achieve sparsity along with the minimization of prediction error. The formulation is given below:

$$\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1 \tag{21}$$

The sparse vector $\mathbf{x}$ is determined by minimizing the objective function given in (21). An algorithm based on Alternating Direction Method of Multipliers (ADMM) is used for solving the LASSO problem (Boyd et al., 2011). The impulse response coefficients are finally recovered using the following expression:

$$\hat{\mathbf{g}} = \mathbf{W}^{-1}\hat{\mathbf{x}} \tag{22}$$

Now, the effectiveness of proposed methodology is demonstrated by means of case studies in Section 5.

5. SIMULATION EXAMPLES:

In this section, two case studies are considered in order to demonstrate the effectiveness of proposed methodology in identification of FIR models for LTI multiscale systems. In this work, the wavelet transformation matrix is generated using Haar wavelet. The scaling and wavelet filters in the case of Haar are as follows:

$$\boldsymbol{\phi} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \boldsymbol{\psi} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{23}$$

The wavelet transformation matrix is generated using the equations (7), (8) and (9). The sparse optimization problem is formulated as given in (21). The results obtained using the proposed sparse optimization (SPOPT) are compared with existing methods namely ordinary least squares (OLS), ridge regression (RR) and regularized least squares with smoothness constraint on IR coefficients (RRD). The comparison is based on mean square error (MSE) of estimated FIR coefficients with respect to their true values (obtained from noise free case),

$$\text{MSE} = \frac{||\hat{\mathbf{g}} - \mathbf{g}_0||_2^2}{M} \tag{24}$$

The steady state gain estimates for different methods are computed from the estimated FIR coefficients as follows:

$$\hat{K} = \sum_{i}^{M} \hat{g}_i \tag{25}$$

In practice, it is not possible to compare the estimates of FIR coefficients and steady state gain with the true values but are included here in an academic setting in order to assess the goodness of FIR estimates. The penalty parameters for the case of ridge regression (α_1), regularized least squares with constraint on distance between consecutive coefficients (α_2) and the proposed sparse optimization method (λ) are determined based on cross validation minimum mean square error (MMSE). The optimal penalty parameter determination is shown in Figure 3 for the case of SPOPT for a particular noise realization with SNR of 1. The penalty parameters α_1 and α_2 for other methods are also determined in a similar manner. 100 Monte Carlo (MC) simulations are performed for different SNRs ($1000, 100, 10$ and 1) and the average MSE values are reported for each of the methods in Table I(a) and II(a). The average steady state gains are reported in Table I(b) and II(b).

5.1 Case study 1:

The system in (6) is again considered to demonstrate the effectiveness of proposed methodology. A full band PRBS is given as an input to the system. The sampling interval is chosen as 1 sec and the time of simulation is 8000 sec. Initial 6000 samples are used for training purpose while the remaining 2000 are used for validation. The length of FIR model used for this case study is 1024.

Table I(a): MSE values for Case study 1

Method / SNR	**MSE** ($\times 10^{-4}$)			
	1000	100	10	1
OLS	0.002675	0.02674	0.2676	2.6720
RR	0.002675	0.02670	0.2615	2.1600
RRD	0.002661	0.02549	0.2013	1.1030
SPOPT	0.000877	0.00613	0.0414	0.2600

Table I(b): Steady-state gain values for Case study 1

Method / SNR	**Steady state gain** (True value = 10)			
	1000	100	10	1
OLS	9.9670	9.9663	9.9733	9.9701
RR	9.9638	9.9463	9.7761	8.3702
RRD	9.9662	9.9588	9.9241	9.7650
SPOPT	9.9486	9.9016	9.7410	9.1296

5.2 Case study 2:

The proposed algorithm is implemented and tested on a simulation example considered in Vána and Preisig (2012). Unlike case study 1, here the system also contains oscillatory behaviour along with presence of two time scales.

$$G(s) = \frac{0.0068(s+100)(s+1)^3}{(s+0.1)^2(s^2+4s+68)} \quad (26)$$

The poles of the considered system are $\{-0.1, -0.1\}$ for slow dynamics and $\{-2-8i, -2+8i\}$ for fast dynamics. The system is discretized using a sampling time of 0.1 sec. A PRBS input of length 4500 is used to excite the system. The first 3000 observations are used for training purpose while the remaining are used for validation. The length of FIR model used for this case study is 512.

Table II(a): MSE values for Case study 2

Method / SNR	**MSE** ($\times 10^{-6}$)			
	1000	100	10	1
OLS	0.00346	0.03466	0.3471	3.4720
RR	0.00419	0.03527	0.3380	2.7120
RRD	0.00475	0.03558	0.3220	2.3440
SPOPT	0.01928	0.02665	0.0878	0.5100

Table II(b): Steady-state gain values for Case study 2

Method / SNR	**Steady-state gain** (True value = 1)			
	1000	100	10	1
OLS	0.9622	0.9622	0.9629	0.9629
RR	0.9561	0.9543	0.9331	0.7756
RRD	0.9617	0.9616	0.9606	0.9524
SPOPT	0.9539	0.9513	0.9401	0.8918

The optimal value of λ for the expression shown in Eq. (21) is determined based on cross validation minimum mean square error as shown in Figure 3. The average mean square error (MSE) values calculated across 100 different noise realizations are reported in Tables I(a) and II(a). It can be inferred from MSE values that by incorporating additional information that the IR coefficients are sparse in the time-frequency domain, it is possible to improve the FIR estimates to a great extent. The estimated FIR coefficients in the case of SNR = 10 for the case studies 1 and 2 using different methods are shown in Figures 2 and 4 respectively. From these figures, it can be observed that the variability of the parameters in the case of proposed methodology are much lower compared to other methods. Any type of regularization introduces bias in the parameter estimates and the same is observed for the proposed methodology too. This is reflected in relatively large bias in the steady state gain estimates as shown in Tables I(b) and II(b). It is well known that there is always a trade-off between bias and variance. Overall, the method generates superior estimates and this is reflected in the lower values of MSE values for the FIR parameters. For the 2^{nd} case study with SNR = 1, the 95% confidence region (obtained from 100 Monte-Carlo simulations) for the parameter estimates is shown in Figure 5. The narrow confidence region indicates that the method yields good estimates of FIR coefficients even when the SNRs are poor.

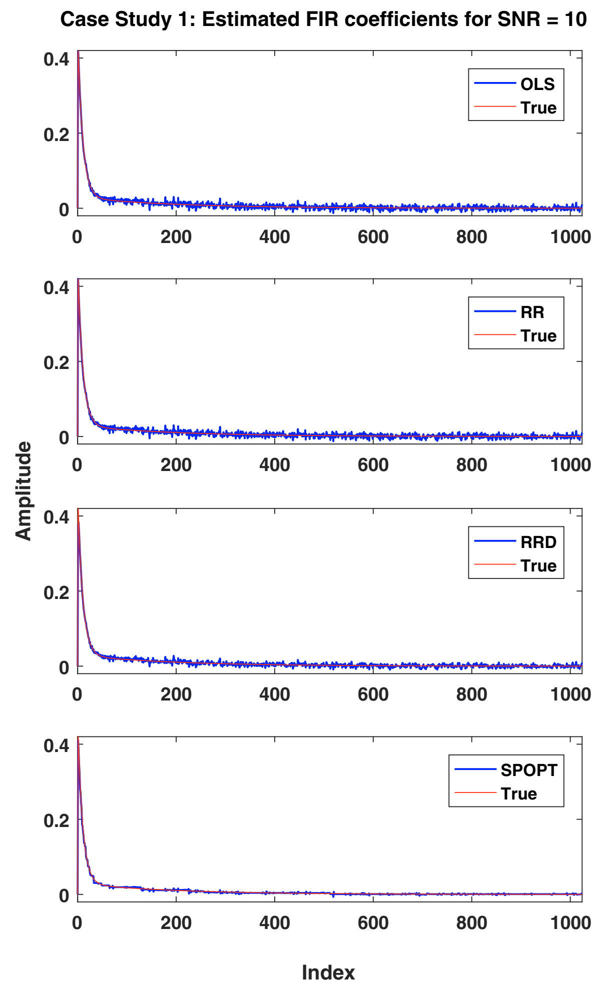

Fig. 2. Estimated FIR coefficients for SNR = 10 using different methods for the case study 1

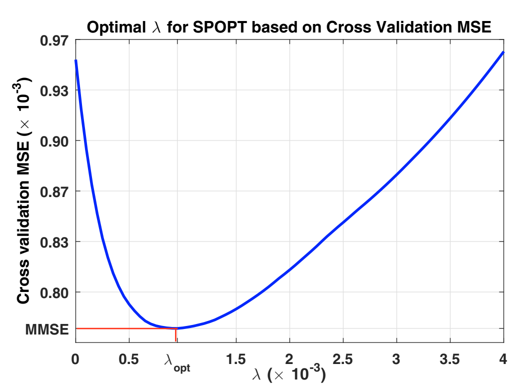

Fig. 3. Choice of optimal λ for SPOPT based on minimum cross validation MSE for case study 2 (for SNR = 10)

6. CONCLUSIONS

In this work, a sparse optimization framework for identification of FIR models has been proposed for LTI systems with multiple time scales. The proposed methodology exploits the sparse representations of IR coefficients in the time-frequency domain. Although, the efficacy of the method is demonstrated on two time scale systems, it is equally applicable for multiple time scale systems and even for single scale band limited systems where the observations are collected at much faster sampling rates. In this work, for the purpose of demonstration, we have used Haar wavelets for computing wavelet transformation matrix. Other wavelets with higher vanishing moments

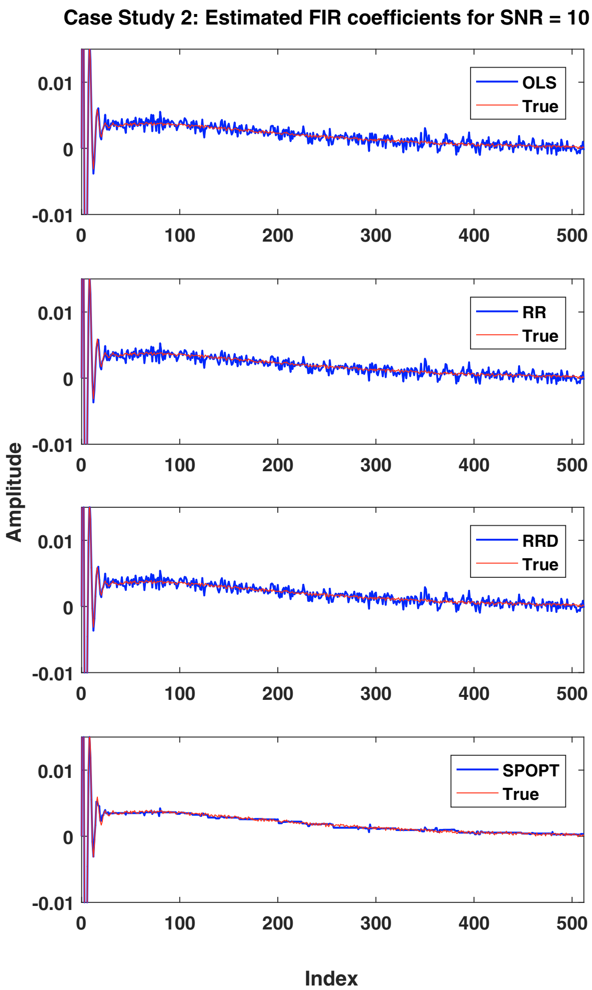

Fig. 4. Estimated FIR coefficients for SNR = 10 using different methods for the case study 2

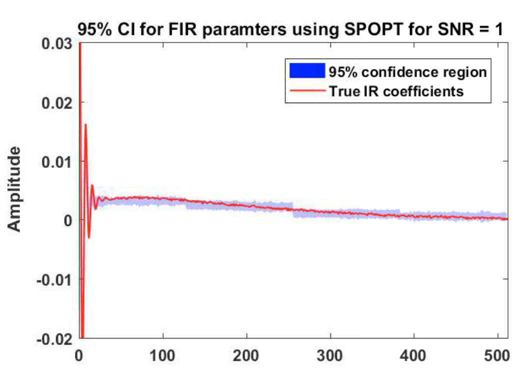

Fig. 5. 95% confidence region for FIR model parameters using SPOPT for case study 2 (SNR = 1)

can also be used in order to improve the smoothness of parameter estimates. Issues such as edge effects caused by DWT and the choice of length of FIR coefficients need to be carefully considered for effective exploitation of proposed methodology.

REFERENCES

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1), 1–122.

Dayal, B.S. and MacGregor, J.F. (1996). Identification of finite impulse response models: methods and robustness issues. *Industrial & engineering chemistry research*, 35(11), 4078–4090.

Ljung, L. (ed.) (1999). *System Identification (2nd Ed.): Theory for the User*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

Nikolaou, M. and Vuthandam, P. (1998). FIR model identification: parsimony through kernel compression with wavelets. *AIChE Journal*, 44(1), 141–150.

Nounou, M.N. (2006). Multiscale finite impulse response modeling. *Engineering Applications of Artificial Intelligence*, 19(3), 289–304.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Vána, Z. and Preisig, H.A. (2012). System identification in frequency domain using wavelets: Conceptual remarks. *Systems and Control Letters*, 61(10), 1041–1051.