Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 52-28 (2019) 1–6

An Atomic Norm Minimization Framework for Identification of Parameter Varying Nonlinear ARX Models

Rajiv Singh * **Mario Sznaier** ** **Lennart Ljung** ***

* *The MathWorks, Inc., 1 Apple Hill Drive, Natick, MA 01760 USA (e-mail: rsingh@mathworks.com)*
** *ECE Dept., Northeastern University, Boston, MA 02115 USA (e-mail: msznaier@coe.neu.edu)*
*** *Div. of Automatic Control, Linköping University, Sweden (e-mail: ljung@isy.liu.se)*

Abstract: We propose a generalization of the popular nonlinear ARX model structure by treating its parameters as varying over time. The parameters are considered generated by linear filters operating on the model's regressors. The filters are expressed as a sum of atoms that are either sum of damped exponentials and sinusoids, or sinusoids with time varying frequencies. This form allows us to enforce stability of the parameter evolution as well as leverage the atomic norm minimization framework for inducing sparsity. It also facilitates easy incorporation of smoothness related priors that that making it possible to treat these models as nonlinear extensions of the familiar LPV models.

Keywords: LPV, NARX, system identification, atomic norm, Frank-Wolfe, regularization, stable spline kernel, RKHS

1. INTRODUCTION

A popular mathematical representation of nonlinear dynamic systems is an auto-regressive model, where the model's states are defined by a finite number of past inputs and outputs.

$$\begin{aligned}\varphi(t) &= [y(t-1),\ldots,y(t-na)),u(t-1),\ldots,u(t-nb)]^T\\ y(t) &= f(\varphi(t),u(t),\boldsymbol{\theta})+e(t) \qquad (1)\end{aligned}$$

$y(t)$ is the model output, $u(t)$ is the input, $\varphi(t)$ is the state vector and $\boldsymbol{\theta}$ denotes the parameter vector. If the function $f(.)$ is linear, we get the familiar ARX model of a linear system. While keeping $f(.)$ linear, a convenient yet flexible way of incorporating nonlinearity is to generalize the regressor set to contain nonlinear functions of model's states and inputs, for example $R(t) = [1, u(t-1)^2, |y(t-1)|, u(t) * y(t-3)]^T$. The nonlinear ARX model then takes a linear-in-regressor form:

$$y(t) = \boldsymbol{\theta}^T R(t) + e(t) \qquad (2)$$

Under a prediction error minimization objective, the estimation problem is linear in $\boldsymbol{\theta}$. For allowing even more flexibility into the model structure, we have two choices:

- Use a nonlinear function for $f(.)$, for example an expansion using sigmoid functions or wavelets as basis (Sjoberg et al. (1995); Zhang (1997)).
- Allow the parameters $\boldsymbol{\theta}$ to vary with time while keeping $f(.)$ linear. Specifically, following a linear parameter varying (LPV) inspired parameterization, represent $\boldsymbol{\theta}$ as a function of system inputs and states. (Singh et al. (2018)).

In this paper we propose using the second approach - approximate the nonlinear dynamics using a parameter-varying model structure. In this approach, $\boldsymbol{\theta}$ changes can be thought of as describing the slower dynamics related to operating point movements while the faster changes in the vicinity of a fixed operating point are described by a $\mathbb{R}^{Nr} \to \mathbb{R}$ transformation of $R(t)$. Note that $R(t)$ can still be nonlinear functions of input-output variables. In an LPV perspective, $\boldsymbol{\theta}$ is considered a function of *scheduling variables*, which are a low-dimensional but possibly nonlinear projection of the inputs and/or states. Our approach of modeling $\boldsymbol{\theta}$ as a function of model's regressors is an effort to deduce the scheduling function directly from data.

We present an efficient approach of constructing parameter varying models under sparsity and smoothness constraints on parameter dynamics. Singh et al. (2018) considered a similar problem with two major differences:

1. In the earlier work, a linear, parameter-varying ARX structure was considered, where the parameters were described using a MIMO linear transfer function that used measured I/O signals as inputs. By contrast, this paper considers a more general modeling framework where the regressors are allowed to be nonlinear. The parameters dependence on regressors is expressed using a transfer function described by a weighted sum of atoms plus an offset; see section 2.2. This representation is stable by construction. Thus for constant input and output values (equilibrium state), it leads to constant values of parameters.

10.1016/j.ifacol.2019.12.342

(2) The minimal order model was obtained by minimizing the rank of a (projected) parameter Hankel matrix. Here, we use an atomic norm minimization framework which makes it easier to sparsify the dynamic dependence of individual parameter dynamics. The corresponding algorithm also works significantly faster while handling larger number of data samples.

The paper is arranged as follows. In section 2 we present the model structure including the parameterization of the $\boldsymbol{\theta}$ dynamics. We also describe the desired constraints on the model structure. In section 3, the identification problem is formally stated. In section 4, an atomic norm minimization based approach to parameter estimation is described. Section 5 illustrates the proposed solution approach on a practical problem. Finally section 6 summarizes our conclusions.

2. PRELIMINARIES

2.1 Model Structure

Consider the parameter-varying extension of the linear-in-regressor nonlinear ARX model:

$$y(t) = \boldsymbol{\theta}(t)^T R(t) + e(t) \tag{3}$$

The number of model regressors is N_r so $\dim(\boldsymbol{\theta}) = N_r$. A constant signal ($= 1$) is often included in the regressor set $R(t)$ to account for output offsets. $\theta(t)$ is a linear function of the model inputs and outputs, $\boldsymbol{\theta}(t) = L(\mathbf{U}(t))$, where $L(.)$ is a multivariate linear filter and $\mathbf{U}(t)$ denotes the set $[u(t), y(t-1)]^T$. The choice of inputs to model the $\boldsymbol{\theta}(t)$ dynamics is mainly for convenience; we also found this to be sufficient for problems described in this paper. However, $\mathbf{U}(t)$ can be generalized to include any possible, linear or nonlinear, regressors.

The proposed model structure allows us to treat the multiple time-scale nature of the parameter-varying systems more readily; we can impose different types of restrictions on the choices of $\boldsymbol{\theta}$ (such as slow variation, smoothness) and $R(t)$ (such as prediction ability in the vicinity of an equilibrium).

2.2 Choice of Atoms

We analyze two different atomic representations for L. In both cases the basic idea is to decompose the impulse response of L into a set of "simpler" components with desirable properties such as smoothness and stability.

Rational Forms of Damped Exponentials and Sinusoids For a given parameter-regressor pair (m, n), a partial fraction expansion of L_{mn} is used so that its impulse response is a sum of damped exponentials and sinusoids. The advantage of this representation is that it yields a stable parameterization and is amenable to introduction of parsimony under an atomic norm minimization framework (Shah et al. (2012)). Specifically, the linear filter is expressed as a linear combination of first and second order strictly proper transfer functions:

$$L_{mn}(z) = \frac{B(z)}{A(z)} = \sum_{i=1}^{N_1} \frac{c_i\lambda_i}{z+p_i} + \sum_{i=1}^{N_2} \frac{c_i\lambda_i}{z^2+2b_i z+a_i^2}$$
$$p_i, a_i \in \mathbb{D}_\rho,\ c_i, b_i \in \mathbb{R},\ |a_i| > |b_i|$$

$\mathbb{D}_\rho$ denotes the origin centered closed disc in $\mathbb{C}$, with radius ρ. The constituting transfer functions are referred to as *atoms*. The constant λ_i are chosen such that the Hankel matrix of size N associated with the impulse response of each atom has nuclear norm equal to one. See Yilmaz et al. (2018) for the exact definition of these atoms.

The set of atoms, $\mathcal{A}_R$ (subscript "R" for "rational"), has the following properties:

- Every proper rational transfer function with poles in $\mathbb{D}_\rho$ can be approximated as a real linear combination of atoms in the set (Shah et al. (2012)).
- Each atom in the set has a transfer function with real coefficients, hence it has a purely real impulse response.

Sinusoids of Time Varying Frequencies Instead of building up from a set of atoms in Hilbert space, one can instead first design a symmetric, positive definite kernel that has desirable properties such as smoothness and exponential decay (Pillonetto et al. (2014)). Such kernels can be parameterized and trained using data, for example, by marginal likelihood maximization. Each such kernel defines a unique reproducing kernel Hilbert space (Moore-Aronszajn theorem) and further provides a basis for this space (Mercer's theorem). An example is the first order stable spline kernel (Pillonetto et al. (2015)). This kernel leads to a basis composed of decaying sinusoids parameterized by a single constant $0 \le \alpha \le 1$, which measures the distance from instability.

$$\mathcal{A}_\alpha = \left\{ \sin\left(\frac{\pi\alpha^t}{2}\right), \sin\left(\frac{3\pi\alpha^t}{2}\right), \sin\left(\frac{5\pi\alpha^t}{2}\right), \dots \right\} \tag{4}$$

We can treat these as first-order stable spline atomic set. Unlike the atoms $\mathcal{A}_R$ corresponding to the damped exponentials and sinusoids of section 2.2.1, the spline atoms $\mathcal{A}_\alpha$ do no yield to a rational representation.

2.3 $\boldsymbol{\theta}(t)$ Dynamics

Let $h(t)$ denote the impulse response of one atom from either the set $\mathcal{A}_R$ or $\mathcal{A}_\alpha$. Then the equation for one element of $\boldsymbol{\theta}$ vector takes the form:

$$\boldsymbol{\theta}_k(t) = \bar{\boldsymbol{\theta}}_k + (T_u)_t h_k^u + (T_y)_t h_k^y,\ \ k = 1, \dots, N_r \tag{5}$$

where h_k^u is the impulse response of the transfer function between $u(t)$ and parameter $\boldsymbol{\theta}_k(t)$. Similarly, h_k^y is the $y(t-1) \to \boldsymbol{\theta}_k(t)$ impulse response. $\bar{\boldsymbol{\theta}}_k$ is a constant scalar so that $\boldsymbol{\theta}_k(t) = \bar{\boldsymbol{\theta}}_k$ in the time-invariant case, that is, a fixed parameter NARX model. $(T_u)_t$ and $(T_y)_t$ are the t^{th} rows of the Toeplitz matrices associated with $u(t)$ and $y(t-1)$ respectively. Each impulse response is then considered as a weighted sum of atoms. This yields:

$$\boldsymbol{\theta}_k(t) = \bar{\boldsymbol{\theta}}_k + (T_u)_t \sum_i^{N_k^u} c_{i,k}^u \mathcal{A}_{i,k}^u + (T_y)_t \sum_i^{N_k^y} c_{i,k}^y \mathcal{A}_{i,k}^y,$$
$$k = 1, \dots, N_r \tag{6}$$

where $h_k^* = \left(\sum_i^{N_k^*} c_{i,k}^* \mathcal{A}_{i,k}^* \right)$ is the (partial) impulse response expressed as sum of atoms, and "$*$" denotes either "u" or "y".

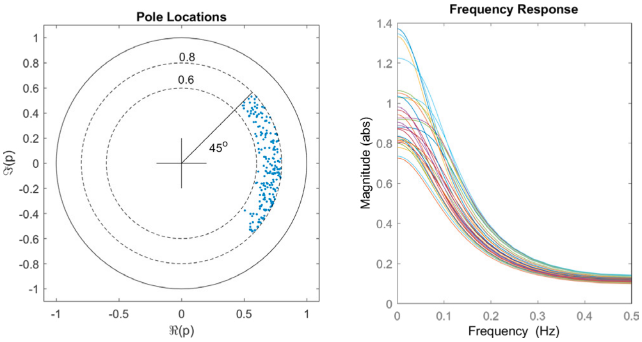

Fig. 1. Rational atoms confined to a sector. Top: Pole locations by random sampling; Bottom: Frequency response of a subset of chosen atoms.

2.4 Priors and Constraints

Parsimony The parsimony requirement can potentially have different elements. We consider two of them in this paper:

(1) *Low order dynamic dependence*: The number of atoms constituting the impulse response from $R_j(t)$ to θ_k are as few as possible (make N_k^u, N_k^y small). This is the main parsimony requirement considered in this paper.

(2) *θ-input sparsity*: We have used only two inputs, $u(t)$ and $y(t-1)$ for modeling the dynamics of each $\theta_k(t)$. As discussed before, the input set can be enriched to include many more regressors. If the inclusion in done in a black-box manner (that is, not guided by physical reasons), it would also be useful to enforce sparsity in this set of inputs.

Smoothness It is also common to consider the parameter variation to be smooth and slow relative to the system dynamics.

- In case of atoms of rational form of atoms, the smoothness requirement can be met by limiting the bandwidth of the filter L to a small value. We consider atoms to be restricted to a sector of the unit circle, bounded in radius and angle; see Figure 1 where the radius $\in [0.6, 0.9]$ and angle between $\pm 45°$. This choice leads to filters with band-limited frequency responses as shown in figure 1.
- In case of stable spline atoms, the smoothness is guaranteed by construction, since they are based on a smooth kernel. The hyper-parameter α controls the decay rate. For low bandwidth, we found it useful to use $\alpha > 0.6$. Model generalizability is also aided by limiting the number of atoms used. See figure 2 for frequency response of these atoms.

3. PROBLEM STATEMENT

The identification problem can be stated as follows:

Problem 1. Given:

- N samples of input output data measured at a constant sampling frequency, $z(t) = \{y(t), u(t),\ \ t = 1, \ldots, N\}$
- A finite set of model regressors $R(t)$, $\dim(R) = N_r$

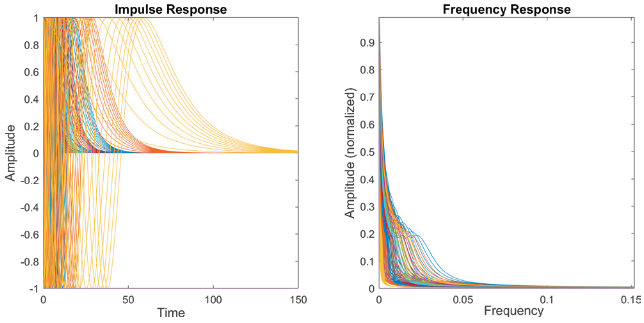

Fig. 2. SS kernel induced atoms: impulse and frequency responses. $0.5 < \alpha \leq 1$, first 11 atoms.

- A priori bounds on the location of atoms (either a sector or a limit on α)

find the most parsimonious model that explains the observed data in the sense that prediction error to measured output is minimized.

Note that the model class is a nonlinear structure $\hat{y}(t) = L(\mathbf{U}(\mathbf{t}))R(t)$ where $L(\mathbf{U}(\mathbf{t}))$ is a signal constructed by a constant linear system $L(.)$ driven by "signals" $u(t), y(t-1)$. The goal is to determine this linear system $L(.)$ so that $\hat{y}(t))$ becomes close to the measured $y(t)$ under the constraints placed on $L(.)$.

4. SOLUTION TECHNIQUE

We propose an atomic norm minimization formulation wherein the prediction error is minimized subject to minimal cardinality of each group of atoms.

$$\min_{c_k^u, c_k^y} \sum_{k}^{N_r} \|c_k^u\|_0 + \|c_k^y\|_0$$

subject to

$$\|y(t) - \hat{y}(t)\|_{\ell_2}^2 \leq \eta, \text{where}$$

$$\hat{y}(t) = \sum_{k}^{N_r} \Bigg(\bar{\theta}_k + (T_u)_t \sum_{i}^{N_k^u} c_{i,k}^u \mathcal{A}_{i,k}^u + (T_y)_t \sum_{i}^{N_k^y} c_{i,k}^y \mathcal{A}_{i,k}^y \Bigg) R_k(t)$$

$$\textit{plus additional/optional constraints} \qquad (7)$$

$\|.\|_0$ denotes the cardinality of a vector. The 2-norm constraint enforces fidelity to the data in the minimum 1-step prediction error sense. By this formulation, we aim to minimize all cardinalities - for each θ_k and from each input $u(t)$ or $y(t-1)$. In the limit case of a parameter not varying with time, its group cardinality is zero. Then that parameter achieves a constant value equal to the constant portion $\bar{\theta}$ of (5). The stated minimization objective satisfies the first parsimony constraint of low order dynamic dependence, while the second requirement is guaranteed by construction (fixed choices of inputs to each $\boldsymbol{\theta}_k(t)$).

4.1 Convex Relaxation and Additional Priors

Here we discuss the convex relaxations of non-convex problem (7). Consider fixed length impulse responses of the atoms. Let $\mathbf{h}_k^{*,N}$ be the first N terms of the impulse response of the transfer function between a certain input ($*$ denotes either u or y), and a certain parameter θ_k. The atomic norm associated with these atoms is defined as:

$$\left\|\mathbf{h}_k^{*,N}\right\|_{\mathcal{A}} \doteq \left\{ \inf \sum_{a \in \mathcal{A}(\mathbb{D}_\rho)} |c_a| \colon \mathbf{h}_k^{*,N} = \sum_{a \in \mathcal{A}(\mathbb{D}_\rho)} c_a \Upsilon^N\{a\} \right\} \tag{8}$$

where Υ^N denotes the truncated N-length impulse response vector of a discrete transfer function. The atomic norm definition leads to the following convex relaxation of problem (7):

$$\min_{h_k^{u,N}, h_k^{y,N}} \|y(t) - \hat{y}(t)\|_{\ell_2}^2, \quad \text{subject to :}$$
$$\sum_k^{N_r} (\|\mathbf{h}_k^{u,N}\|_{\mathcal{A}} + \|\mathbf{h}_k^{y,N}\|_{\mathcal{A}}) \leq \tau \tag{9}$$

Expressing the impulse responses as sum of atoms and using the definition of the atomic norm, the optimization problem becomes:

$$\min_{\mathbf{c}} \|y(t) - \hat{y}(t)\|^2, \quad \text{subject to :}$$
$$\sum_k^{N_r} \|\mathbf{c}_k\|_1 \leq \tau \tag{10}$$

where:

$$\hat{y}(t) = \sum_k^{N_r} \left(\bar{\theta}_k + (T_u)_t \sum_i^{N_k^u} c_{i,k}^u \mathcal{A}_{i,k}^u + (T_y)_t \sum_i^{N_k^y} c_{i,k}^y \mathcal{A}_{i,k}^y \right) R_k(t)$$

We make the following observations about problem (10):

(1) Low complexity is promoted by constraining the optimal solution to be inside the τ-scaled atomic norm ball ($\sum_k^{N_r} \|\mathbf{c}_k\|_1 \leq \tau$) (Chandrasekaran et al. (2012)). A small value of τ would move the model towards a constant-coefficient one. Too large a value would lead to over-fitting and the results may not generalize well.

(2) This formulation is very similar to the familiar $L1$-penalty based *LASSO* formulation where the minimization objective is:

$$\min_{\mathbf{c}} \|y(t) - \hat{y}(t)\|^2 + \tau \sum_k^{N_r} \|\mathbf{c}_k\|_1 \tag{11}$$

We chose to treat the atomic norm as a hard constraint since it allows us to formulate a Frank-Wolfe type algorithm for efficiently solving for the unknowns (see Algorithm 1 in Section 4.2). This algorithm is significantly faster than the traditional L1 solvers. Note also that τ value can often encode our prior knowledge about the order of the θ dynamics. For this reason too it is not desirable to treat τ as an arbitrary trade-off hyper-parameter between the estimation MSE and model sparsity.

(3) The a priori information about the stability margin of the unknown system, or other information about the poles of the identified system, is implicitly incorporated in the choice of the atomic set. In particular, the smoothness requirement is met by choices described in Section 2.4.2.

(4) Since the noise sequence is assumed to be bounded, the system to be identified can be approximated to arbitrary precision as described above with a *finite* atomic norm.

4.2 Minimizing the Objective

We use a randomized version of Frank-Wolfe algorithm, proposed in Yilmaz et al. (2018), for finding sparse solutions to optimization problem (10). This algorithm is summarized in 1.

Algorithm 1 Randomized algorithm to minimize a convex function f over the τ-scaled atomic norm ball

```
1: $\mathbf{x_0} \leftarrow \tau\Upsilon^N\{a_0(z)\}$ for arbitrary $a_0(z) \in \mathcal{A}$        ▷ Init.
2: for $i = 0, 1, \dots, i_{max}$ do
3:     Select $N_i$ elements $\{S_i\}$ in the atom set $\mathcal{A}$
4:     $\mathbf{a_i} \leftarrow \Upsilon^N\{\text{argmin}_{a(z)\in\mathcal{A}\{S_i\}} \langle \nabla f(x_i), \Upsilon^N\{a(z)\}\rangle\}$
5:     $\alpha_i \leftarrow \text{argmin}_{\alpha\in[0,1]} f(\mathbf{x_i} + \alpha[\tau\mathbf{a_i} - \mathbf{x_i}])$
6:     $\mathbf{x_{i+1}} \leftarrow \mathbf{x_i} + \alpha_i[\tau\mathbf{a_i} - \mathbf{x_i}]$
7:     Compute $\bar{\theta}_i$ by backcasting.
8: end for
```

$\mathbf{x}$ is the impulse response of one atom group. There are $2 \times N_r$ atom groups, one group for each element of $\boldsymbol{\theta}$ and from each input u and y. In step 1, N random atoms are picked and scaled by τ to serve as the initial solution for each group. Note that the initial solution belongs to the boundary of the feasible set. Then in step 3, a fixed number N_i of random atoms are selected from the atomic set $\mathcal{A}$, for each of the atom groups; usually $N_i = N$.

- For the rational atoms ($\mathcal{A} = \mathcal{A}_R$), the atoms correspond to poles $\mathbf{p}$ uniformly distributed over the sector $\mathbb{D}_\rho,\ \rho \geq \rho_{min},\ \angle\mathbf{p} \leq \phi_{max}$.
- For the stable spline atoms ($\mathcal{A} = \mathcal{A}_\alpha$), the atoms are impulse responses $\sin((2k+1)\pi\alpha^t/2)$ with k distributed uniformly over $[0, k_{max}]$.

For step 4, the gradient for a particular element with respect to a given input (denoted by *) is calculated as $\nabla f(x_k) = -(R_k \odot \mathbf{T}_*)^T\mathbf{E}$, where $\mathbf{E}$ is the prediction error vector $\mathbf{y} - \hat{\mathbf{y}}$ and $\odot$ denotes Hadamard product. Steps 4 and 5 are implemented in a coordinate descent fashion: they are performed individually for each input ($* = u, y$) and each element of θ value while holding all other atom groups to their values from earlier iteration.

Backcasting for $\bar{\theta}$ In the i^{th} iteration, after the impulse response vectors $h_k^{*,N}$ are updated (step 6), a linear least squares operation is performed to minimize the prediction error, using $\bar{\theta}$ as the only free variable. Note that $\bar{\theta}$ is a vector with N_r components. This calculation is carried out once per iteration.

5. EXAMPLES

5.1 Regularized LTI Identification

To get a feel for identification under an atomic norm framework, consider first the problem of identification of a SISO LTI system subject to poor excitation. The objective is to recover the true transfer function, verified by

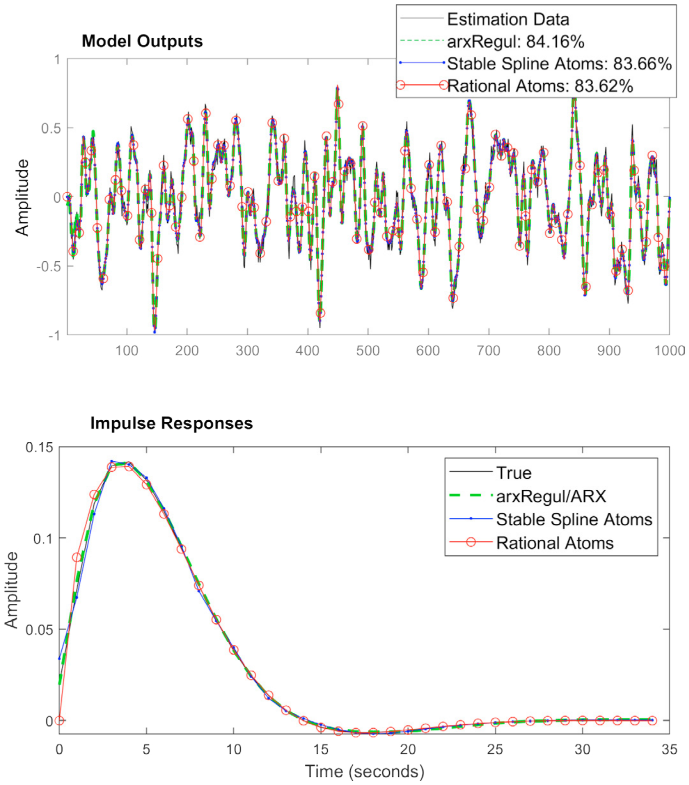

Fig. 3. Linear identification under atomic norm minimization framework. Top: Model outputs to low-pass filtered random input signal. Bottom: Retrieved impulse responses.

matching its impulse response with that of the estimated model. If using a long FIR form of the identified model, this problem is solved efficiently by using a stable spline kernel based quadratic regularizer, for example, using the `arxRegul, arx` functions of MATLAB® System Identification Toolbox™. Under an atomic representation, the objective would be to match the measured response using as few atoms as possible. This leads to the following formulation:

$$\min_{\mathbf{g}} \quad \|\mathbf{T_u g} - \mathbf{y}\|^2_{\ell_2} \tag{12}$$

$$\text{subject to} \quad \|\mathbf{g}\|_{\mathcal{A}} \leq \tau \tag{13}$$

where $\mathbf{g}$ is considered a weighted sum of the atoms; see equation (8) for a definition of the atomic norm. The example system considered is:

$$G(z) = \frac{0.02008 + 0.04017z^{-1} + 0.02008z^{-2}}{1 - 1.561z^{-1} + 0.6414z^{-2}} \tag{14}$$

This system is excited using a low-pass filtered white noise as input. The response is corrupted with a small additive disturbance. 10 random atoms are chosen in every iteration uniformly inside the unit circle with maximum radius of 0.999; complex poled are used in conjugate pairs. $\tau = 0.89$ was used as sparsity constraint with rational atoms and $\alpha = 0.88$ with stable spline atoms. The fit to the input-output data and the retrieved impulse responses using the two types of atoms are shown in figure 3. As seen, the atomic norm minimization framework delivers results virtually identical to those obtained by regularized estimation using `arxRegul, arx` functions. The goodness of fit is based on the normalized root mean square error (NRMSE) metric, expressed in percentage (fit = (1-NRMSE)*100). Furthermore, both impulse responses can be described by second order systems as verified by computing the Hankel singular values. Figure 4 shows the effect of varying τ on the fit quality and model sparsity. $\tau = 0.1$ delivers a first order model that is (obviously) biased. $\tau = 10$ overfits as seen by wiggles in the impulse response. $\tau = 1$ is close to being the optimal value which leads to good fit and second order model.

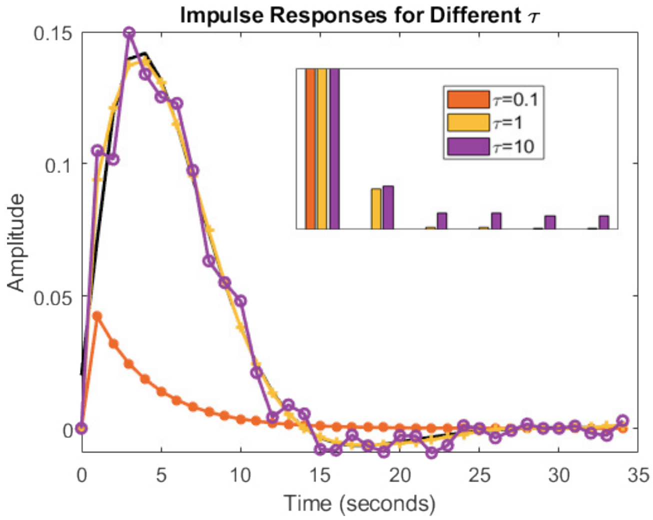

Fig. 4. Estimated impulse responses using rational atoms for 3 different values of τ: 0.1, 1 and 10. The significant singular values (normalized) for the 3 choices are shown in the bar plot.

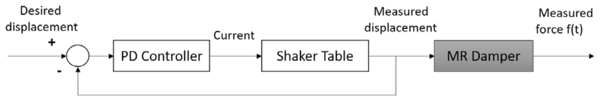

Fig. 5. Experimental set-up diagram for MR damper.

5.2 *Magneto-Rheological Fluid Damper*

We now apply the parameter-varying NARX structure to the black-box modeling of the dynamic behavior of a magneto-rheological (MR) fluid damper. MR fluid dampers are semi-active control devices used for reducing vibrations in dynamic structures. MR fluids, whose viscosities depend on the input voltage/current of the device, provide controllable damping forces. The experiment (Wang et al. (2009)) consisted of fixing a damper at one end to the ground and connecting the other end to a shaker table, as shown in diagram 5.

The voltage of the damper was set to 1.25 V. The damping force $f(t)$ was sampled every 0.005s. The displacement was sampled every 0.001s, which was then used to estimate the velocity $v(t)$ at the sampling period of 0.005s. 1000 samples of the input $v(t)$ (cm/s) and output $f(t)$ (N) were used for estimation and another 1500 were reserved for validation.

The regressor set used was $R(t) = [f(t-1), v(t-1), \ldots, v(t-4), f(t-1)^2, v(t-1)^2, \ldots, v(t-4)^2, 1(t)]^T$. The NARX coefficients were modeled as a low-pass filtered input $v(t)$, that is, $\theta_k(t) = G_k u(t), k = 1, \ldots, 11$, where G_k

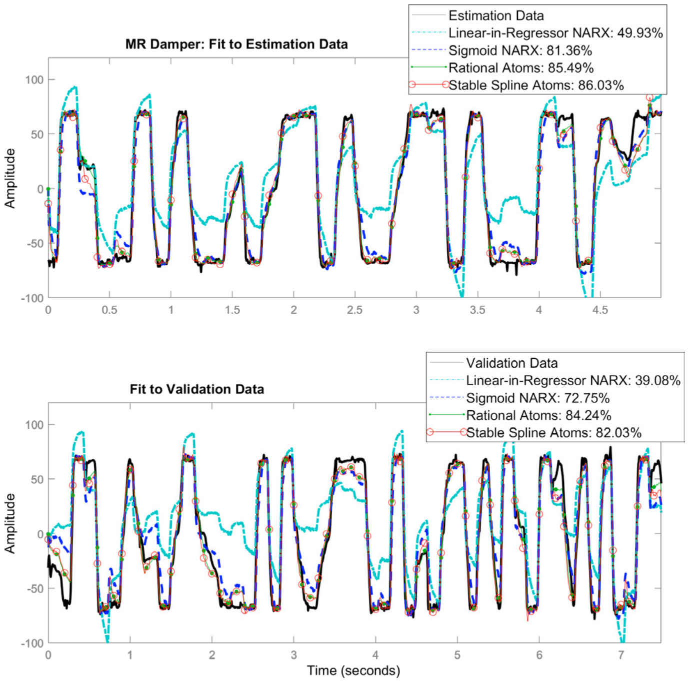

Fig. 6. MR Damper: simulation results for the linear-in-regressor model, Sigmoid NARX and the two parameter-varying NARX models. Top: compared to estimation data. Bottom: compared to validation data.

are low-pass linear filters. $1(t)$ refers to a vector of ones that is used to account for offsets. We identify 4 models:

(1) *Linear-in-Regressor NARX*: A weighted sum of the regressors $R(t)$.
(2) *Sigmoid NARX*: Constant-coefficient NARX model using a sigmoid network nonlinearity (function $f(.)$ of Equation (1)) and regressors $R(t)$.
(3) *Rational Atoms*: A parameter-varying NARX model with regressors $R(t)$ and poles confined to $0.2 - 0.6$ radius, $\pm\pi/16$ rad angles. Atomic norm threshold $\tau = 5$ was used. The value of τ was derived by cross validation tests on a validation dataset.
(4) *Stable Spline Atoms*: Same as (3) above but using stable spline induced atoms, with $\alpha = 0.6$, again found by cross validation.

An analysis of parameter trajectories can help further simplify the model structure. Comparison of the parameters' static portion ($\bar{\boldsymbol{\theta}}$) to the dynamic portion ($(T_u)_t h_k^u + (T_y)_t h_k^y$) revealed that parameters 1:5 out of the original 11 were essentially constants. Hence their dynamic contributions could be set to zero. For the remaining parameters the dynamic contribution from $y(t-1)$ is small and can also be removed. The final model has the first 5 parameters as constants while the rest (6:11) showing dynamic dependence of orders 2 or 3 on only the input signal $u(t)$.

The results are shown in figure 6. As seen, the parameter-varying NARX models validate better than the constant-parameter NARX model. The constant-parameter NARX model employs a sigmoid nonlinear function giving it a flavor of 1-hidden layer neural network. Configuration of such models - choice of nonlinearity, initialization and configuration of nonlinear function properties (such as the number of units) is difficult. The resulting estimation problem is also non-convex making it significantly more difficult to train than the proposed structure.

6. CONCLUSIONS

Atomic representation offers a flexible framework for identifying parameter-varying models where the nature of scheduling is not known in advance. This formulation leads to algorithms that are both faster and memory efficient compared to the nuclear norm minimization formulation presented in Singh et al. (2018). This approach guarantees stability of parameter dynamics while making it easier to introduce sparsity and smoothness related priors. We presented two different choices of atoms that perform equally well in capturing parameter dependence on system's states and inputs. As such this framework can be extended to any choice of atoms, such as those induced by various kernels in RKHS.

Practically, two datasets are needed to tune the kernel parameters - atomic norm bound τ and pole sectors for $\mathcal{A}_R$, sinusoidal frequency parameter α for $\mathcal{A}_\alpha$) and identify the model.

REFERENCES

Chandrasekaran, V., Recht, B., Parrilo, P.A., and Willsky, A.S. (2012). The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6), 805–849.

Pillonetto, G., Chen, T., Chiuso, A., Nicolao, G., and Ljung, L. (2015). Regularized linear system identification using atomic, nuclear and kernel-based norms: the role of the stability constraint. *Automatica*.

Pillonetto, G., Dinuzzo, F., Chen, T., Nicolao, G.D., and Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3), 657–682.

Shah, P., Bhaskar, B., Tang, G., and Recht, B. (2012). Linear system identification via atomic norm regularization. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 6265–6270.

Singh, R., Sznaier, M., and Ljung, L. (2018). A rank minimization formulation for identification of linear parameter varying models. *IFAC-PapersOnLine*, 51(26), 74 – 80. 2nd IFAC Workshop on Linear Parameter Varying Systems LPVS 2018.

Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.Y., Hjalmarsson, H., and Judisky, A. (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12), 1691–1724.

Wang, J., Sano, A., Chen, T., and Huang, B. (2009). Identification of hammerstein systems without explicit parameterisation of non-linearity. *International Journal of Control*, 82(5), 937–952. doi: 10.1080/00207170802382376.

Yilmaz, B., Bekiroglu, K., Lagoa, C., and Sznaier, M. (2018). A randomized algorithm for parsimonious model identification. *IEEE Transactions on Automatic Control*, 63(2), 532–539.

Zhang, Q. (1997). Using wavelet network in nonparametric estimation. *IEEE Transactions on Neural Networks*, 8(2), 227–236.