TU Delft

Delft University of Technology

Input selection in N2SID using group lasso regularization

Klingspor, M.; Hansson, A; Löfberg, J.; Verhaegen, Michel

DOI
10.1016/j.ifacol.2017.08.1472

Publication date
2017

Document Version
Final published version

Published in
IFAC-PapersOnLine

Citation (APA)
Klingspor, M., Hansson, A., Löfberg, J., & Verhaegen, M. (2017). Input selection in N2SID using group lasso regularization. In D. Dochain, D. Henrion, & D. Peaucelle (Eds.), *IFAC-PapersOnLine: Proceedings 20th IFAC World Congress* (Vol. 50-1, pp. 9474-9479). (IFAC-PapersOnline; Vol. 50, No. 1). Elsevier. https://doi.org/10.1016/j.ifacol.2017.08.1472

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 50-1 (2017) 9474–9479

Input selection in N2SID using group lasso regularization

Måns Klingspor, Anders Hansson, Johan Löfberg *, Michel Verhaegen **

* *Division of Automatic Control, Linköping University, Linköping, Sweden (e-mail: {mans.klingspor, anders.hansson}@liu.se, johanl@isy.liu.se).*

** *Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands (e-mail: m.verhaegen@tudelft.nl).*

Abstract: Input selection is an important and oftentimes difficult challenge in system identification. In order to achieve less complex models, irrelevant inputs should be methodically and correctly discarded before or under the estimation process. In this paper we introduce a novel method of input selection that is carried out as a natural extension in a subspace method. We show that the method robustly and accurately performs input selection at various noise levels and that it provides good model estimates.

© 2017, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Input selection, System identification, State-space models, N2SID, Subspace methods, Signal-to-noise ratio

1. INTRODUCTION

One of the challenges in system identification is input selection. Being subject to a number of potential inputs, it is desirable to only include relevant inputs to avoid an overly complex model. This task may be troublesome due to an excessive number of inputs and noisy measurements, making input selection a research topic of interest in the system identification community, see for example Van de Wal and De Jager [2001] and Rojas, Tóth and Hjalmarsson [2014]. Thus, there are several methods available for input selection which we will briefly discuss before introducing our method.

Perhaps most popular and widely referenced are various methods of Adaptive Neuro Fuzzy Inference System (ANFIS) which is an artificial neural network method. ANFIS may be used for input selection for both linear and nonlinear systems, see Jang [1993]. As for applications, ANFIS is for example used for input selection in problems related to identifying the most significant input parameters for predicting global solar radiation shown by Mohammadi et al. [2001]. Furthermore, despite its popularity, one has to keep in mind that optimizing neural networks are non-convex problems and implementation can be difficult. Also, as ANFIS is solely used for the input selection problem, if a model is desired one has to consult some suitable method for identification.

Other, perhaps simpler, methods for input selection include Partial Linear Correleation (PLC) and Partial Mutual Information (PMI) which are model-free techniques, see Tran et al. [2015]. These model-free approaches rely on the statistical relationship between the various inputs and outputs using linear and non-linear correleation. Simply, inputs that in a statistical sense highly correlate with any of the outputs should be regarded as relevant inputs. On the contrary, inputs that do not correlate with any of the outputs should not be considered relevant for the system and should thus be discarded from the modeling process.

Another closely related method that is worth mentioning are nearest correlation spectral clustering in combination with a group lasso, see Fujiwara and Kano [2015]. This method is used in applications related to soft sensors and the group lasso regularization is the same technique our method employs. Worth mentioning is also Relative Gain Array (RGA) methods, see Kadhim et al. [2014].

In this paper, we introduce an extension to the Nuclear Norm Subspace Identification (N2SID) framework proposed by Verhaegen and Hansson [2015]. As a natural extension to N2SID, we introduce a novel method where input selection is directly incorporated into the N2SID framework. This extension does not interfere with the convex property of the N2SID problem and neither its property to be recast as a *Semi-Definite Programming* (SDP) problem. Thus, computational complexity remains virtually unchanged with this input selection feature extension. Furthermore, as we show in the paper, the estimated and input selected models that our modified N2SID method provides are excellent, and the input selection works accurately.

1.1 Notation

For simplicity and readability, we introduce a notation for submatrices. Given $X \in \mathbb{R}^{m\times n}$ and integers $1 \leq i \leq j \leq m$, $1 \leq k \leq l \leq n$, then $X_{i:j,k:l}$ is the submatrix of X with rows from i to j and columns k to l. For further readability,

[1] Support from the Swedish Research Council under contract No E05946CI is gratefully acknowledged.

2405-8963 © 2017, 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.2017.08.1472

if the number of rows in the submatrix is equal to m, then $X_{1:m,k:l}$ is written simply as $X_{:,k:l}$. The same simplification applies to columns.

Some special norms are used in this article which include the nuclear norm and the $\mathcal{H}_2$ norm. The nuclear norm is denoted $\|\cdot\|_\star$ and the $\mathcal{H}_2$ norm is denoted $\|\mathcal{S}\|_{\mathcal{H}_2}$ where $\mathcal{S}$ is some state-space model.

Finally, $L = \text{logspace}\,(a, b, n)$ defines an ordered set $L = \{l_1, ..., l_n\}$ where $l_1 = 10^a$, $l_n = 10^b$ and all the elements are logarithmically spaced (compare with the MATLAB function `logspace`).

2. PRELIMINARIES

2.1 State-space representation and input selection

A common problem in system identification is to identify a state-space model of a linear, time-invariant system with multiple inputs and multiple outputs. A general discrete-time state-space representation is given by

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) \\ \quad y(t) = Cx(t) + Du(t) \end{cases} \tag{1}$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^q$, $u(t) \in \mathbb{R}^p$. Explicitly,

$$x(t) = [x_1(t) \ \dots \ x_n(t)]^T \tag{2}$$

$$u(t) = [u_1(t) \ \dots \ u_p(t)]^T \tag{3}$$

$$y(t) = [y_1(t) \ \dots \ y_q(t)]^T \tag{4}$$

where (2) is the state vector, (3) is the input vector and (4) is the output vector. In $u(t)$ and $y(t)$, each component represents an input and output, respectively.

Provided input and output measurement data, given by $u(1), ..., u(N)$ and $y(1), ..., y(N)$, the challenge is to estimate the matrices A, B, C, D in (1) so that they fit the measurement data. However, all inputs might not actually affect the system, and this will correspond to zero columns in the B and D matrices. These possibly redundant inputs leads us to the following definition:

Definition 1. Suppose u_k, $k \in \{1, .., p\}$ is a component of the input vector $u(t)$ in the state-space system in (1). The component u_k is said to be a *non-significant input* if $B_{:,k} = \mathbf{0}$, $D_{:,k} = \mathbf{0}$. Otherwise, the component is said to be a *significant input*.

If an input is non-significant it is clear from the definition that the input has no effect on the state equations (through B) nor directly on the output (through D). Thus, it is unnecessary to include this input component in our state-space model. This is the very core of input selection, where only significant inputs should be included in the final model.

Remark 2. In practice, the condition in (1) that $B_{:,k} = \mathbf{0}$, $D_{:,k} = \mathbf{0}$ for a non-significant input is replaced with $|B_{1,k}| \leq \varepsilon$, ..., $|B_{n,k}| \leq \varepsilon$, $|D_{1,k}| \leq \varepsilon$,...,$|D_{q,k}| \leq \varepsilon$ for some tolerance $\varepsilon > 0$.

One might model the system with all inputs treated as significant. Then, hopefully, the estimations of B and D will tell which inputs are significant and which can be discarded. However, the measurements $u(1), ..., u(N)$ and $y(1), ..., y(N)$ often contain noise and the ideal framework in (1) is not suitable. In this paper, we restrict ourselves to the case where the input is known and the output measurements contain white noise. Then the system description is given by:

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) + Ke(t) \\ \quad y(t) = Cx(t) + Du(t) + e(t) \end{cases} \tag{5}$$

where $e(t) \sim \mathcal{N}(0, \sigma)$ with the same dimension as $y(t)$. Because of the presence of the noise in our system and thus in the output measurements, it will be transferred to the estimations of B and D, and hence make the detection of zero-columns non-trivial.

2.2 State-space identification with N2SID

Before introducing the method of input selection, we will introduce and discuss *Nuclear Norm Subspace Identification* (*N2SID*) briefly. For a more thorough review regarding this particular subspace method, we refer to Verhaegen and Hansson [2015]. The reason why we use N2SID is because it is a convex problem which can be written as a *Semi-Definite Programming* (SDP) problem. Also, N2SID has shown to work well for relatively short batches of measurement data. Before we go into more details about the N2SID problem some definitions are required.

The model in (5) may be represented in its observer form

$$\begin{cases} x(t+1) = (A - KC)x(t) + (B - KD)u(t) + Ky(t) \\ \quad y(t) = Cx(t) + Du(t) + e(t) \end{cases} \tag{6}$$

where we may define $\mathcal{A} = A - KC$ and $\mathcal{B} = B - KD$ for a slightly more compact notation. This yields

$$\begin{cases} x(t+1) = \mathcal{A}x(t) + \mathcal{B}u(t) + Ky(t) \\ \quad y(t) = Cx(t) + Du(t) + e(t) \end{cases} \tag{7}$$

which will be used for constructing the data equation. As before, assume that we have input and output measurements $u(1), ..., u(N)$ and $y(1), ..., y(N)$. Let $s > n$ and define the block Hankel matrix U^s for the input $u(t)$ as

$$U^s = \begin{bmatrix} u(1) & u(2) & \dots & u(N-s+1) \\ u(2) & u(3) & & \vdots \\ \vdots & & \ddots & \\ u(s) & u(s+1) & \dots & u(N) \end{bmatrix}. \tag{8}$$

In the very same manner as (8), define block Hankel matrices Y^s and E^s for $y(t)$ and $e(t)$, respectively. Furthermore, define the block Toeplitz matrix $T^{u,s}$ from the quadruple of system matrices $\{\mathcal{A}, \mathcal{B}, C, D\}$ as

$$T^{u,s} = \begin{bmatrix} D & 0 & \dots & 0 \\ C\mathcal{B} & D & & 0 \\ \vdots & & \ddots & \\ C\mathcal{A}^{s-2}\mathcal{B} & & \dots & D \end{bmatrix} \tag{9}$$

and similarly, define $T^{y,s}$ from the quadruple $\{\mathcal{A}, K, C, 0\}$. Also, define the extended observability matrix, $\mathcal{O}^s$:

$$\mathcal{O}^s = \begin{bmatrix} C^T & \mathcal{A}C^T & \dots & \left(\mathcal{A}^T\right)^{s-1} C^T \end{bmatrix}^T. \tag{10}$$

Finally, let the state sequence be stored as:

$$X = [x(1) \ x(2) \ \dots \ x(N-s+1)] \tag{11}$$

With these definitions, the data equation is given by:

$$Y^s = \mathcal{O}^s X + T^{u,s}U^s + T^{y,s}Y^s + E^s \tag{12}$$

This data equation can be slightly reformulated. Let $\hat{y}(t) = y(t) - e(k)$ and $\hat{Y}^s$ be a Hankel matrix defined in the same way as Y^s for $y(t)$. Furthermore, let $\mathcal{T}^{p,m}$ denote the class of lower triangular block-Toeplitz matrices with block entries $p \times m$ matrices. Also, let $\mathcal{H}^p$ denote the class of block-Hankel matrices with block entries of p column vectors. With these definitions, we may now provide the final expression of the N2SID problem and for more details regarding this, we refer to Verhaegen and Hansson [2015].

This is a convex problem that includes a rank penalty (the nuclear norm) and a penalty on the sample average. Note that the penalty on the sample average may be adjusted with the regularization parameter $\lambda_1 \geq 0$.

$$\begin{aligned}\min_{\hat{Y}^s \in \mathcal{H}^p, \hat{T}^{u,s} \in \mathcal{T}^{p,m}, \hat{T}^{y,s} \in \mathcal{T}^{p,p}} & \|\hat{Y}^s - \hat{T}^{u,s}U^s - \hat{T}^{y,s}Y^s\|_\star + \\ + \lambda_1 \sum_{k=1}^{N} & \|y(k) - \hat{y}(k)\|_2^2\end{aligned} \tag{13}$$

To conclude, with (13) the "best" estimation in the N2SID sense is now well defined.

3. GROUP LASSO REGULARIZATION

The main idea behind our input selection method is adopted from group lasso proposed by Yuan and Lin [2006] and further discussed by Friedman, Hastie and Tibshirani [2010]. Assuming that the variables in some optimization problem have a natural grouping, group lasso imposes regularization on the groups and helps to achieve sparse solutions at group level.

For our problem in (13), the matrix variable $\hat{T}^{u,s}$ is related to the B and D matrices of the system through (9). Looking at (9), and in particular

$$T^{u,s}_{:,1:p} = \begin{bmatrix} D \\ C\mathcal{B} \\ \vdots \\ C\mathcal{A}^{s-2}\mathcal{B} \end{bmatrix} = \begin{bmatrix} D \\ C(B - KD) \\ \vdots \\ C\mathcal{A}^{s-2}(B - KD) \end{bmatrix} \tag{14}$$

we note that $T^{u,s}_{:,1:p}$ uniquely describes $T^{u,s}$ because of the Toeplitz structure. Furthermore, it is not hard to see that a column k in $T^{u,s}_{:,1:p}$ is related to respective column k in B and D and thus related to the component $u_k(t)$ in $u(t)$. Consequently, it makes sense to penalize each column in $\hat{T}^{u,s}_{:,1:p}$ in order to achieve zero columns in the solution of said matrix. These zero columns in $\hat{T}^{u,s}_{:,1:p}$ will then transfer to corresponding columns in the estimates of B and D, and the components that are non-significant inputs become easily distinguishable.

With this background, a slight modification of the problem in (13) is proposed with the introduction of a group lasso term where each column in $\hat{T}^{u,s}_{:,1:p}$ is penalized.

$$\begin{aligned}\min_{\hat{Y}^s \in \mathcal{H}^p, \hat{T}^{u,s} \in \mathcal{T}^{p,m}, \hat{T}^{y,s} \in \mathcal{T}^{p,p}} & \|\hat{Y}^s - \hat{T}^{u,s}U^s - \hat{T}^{y,s}Y^s\|_\star + \\ + \lambda_1 \sum_{k=1}^{N} & \|y(k) - \hat{y}(k)\|_2^2 + \lambda_2 \sum_{k=1}^{p} \|\hat{T}^{u,s}_{:,k}\|_2\end{aligned} \tag{15}$$

Note that a regularization parameter $\lambda_2 \geq 0$ is introduced for the group lasso penalty.

Remark 3. In the group lasso setup proposed by Friedman, Hastie and Tibshirani [2010], there is a unique regularization parameter for each group penalty dependent on the number of elements in the group. For (15) this would be superfluous since all groups have the same number of elements and hence λ_2 is sufficient.

4. METHOD

With the theory established, we may now introduce our method of input selection. The assumption is that we have some linear system as in (5) with n states, inputs $u_1(t), ..., u_p(t)$ and outputs $y_1(t), ...y_q(t)$. We assume that there are measurement data of both the inputs and outputs, $u_1(1), ..., u_p(1), ..., u_1(N), ..., u_p(N)$ and $y_1(1), ..., y_q(1), ..., y_1(N), ..., y_q(N)$.

i. For each $t = 1, ..., N$, stack the measurement data as

$$u(t) = \begin{bmatrix} u_1(t) & \dots & u_p(t) \end{bmatrix}^T \tag{16}$$

$$y(t) = \begin{bmatrix} y_1(t) & \dots & y_q(t) \end{bmatrix}^T. \tag{17}$$

ii. From the measurement data in (16) and (17), form the Hankel matrices U^s and Y^s as according to (8) for some $s > n$.

iii. Define variables $\hat{Y}^s$, $\hat{T}^{u,s}$ and $\hat{T}^{y,s}$ according to the discussion. Note that the variable $\hat{y}(t)$ is extracted from $\hat{Y}^s$.

iv. Solve the problem in (15), that is

$$\begin{aligned}\min_{\hat{Y}^s \in \mathcal{H}^p, \hat{T}^{u,s} \in \mathcal{T}^{p,m}, \hat{T}^{y,s} \in \mathcal{T}^{p,p}} & \|\hat{Y}^s - \hat{T}^{u,s}U^s - \hat{T}^{y,s}Y^s\|_\star + \\ + \lambda_1 \sum_{k=1}^{N} & \|y(k) - \hat{y}(k)\|_2^2 + \lambda_2 \sum_{k=1}^{p} \|\hat{T}^{u,s}_{:,k}\|_2\end{aligned} \tag{18}$$

for some suitable λ_1, λ_2.

v. Let $\hat{T}^{u,s}$ be the optimal solution to problem above. Given a tolerance $\varepsilon > 0$, an input u_k, $k \in \{1, ..., p\}$ is regarded as non-significant if it holds that $|\hat{T}^{u,s}_{1,k}| \leq \varepsilon$, ..., $|\hat{T}^{u,s}_{n,k}| \leq \varepsilon$. Otherwise, the input is significant.

vi. With p_s significant inputs, let $\mathbb{S} = \{k_1, ..., k_{p_s}\}$ be a ordered subset of the indices of the significant inputs. Then for each $k_i \in \mathbb{S}$, define a reduced Toeplitz matrix $\hat{T}^{u,s,\mathrm{GL}}$ given by

$$\hat{T}^{u,s,\mathrm{GL}}_{:,k_i+mp} = \hat{T}^{u,s}_{:,k_i+mp} \tag{19}$$

for $i = 1, ..., p_s$ and $m = 0, ..., s-1$.

vii. Define $u^{\mathrm{GL}}(t)$ as

$$u^{\mathrm{GL}}(t) = \begin{bmatrix} u_{k_1}(t) & \dots & u_{k_{p_s}}(t) \end{bmatrix}^T \tag{20}$$

for each $t = 1, ..., N$. This is a stacking of the input measurements as in (16) but only including the significant inputs. The superscript GL stands for Group Lasso.

viii. From the reduced input measurement data $u^{\mathrm{GL}}(t)$, construct the Hankel matrix $U^{s,\mathrm{GL}}$ according to (8) for the same s as before.
ix. With $U^{s,\mathrm{GL}}$ and $\hat{T}^{u,s,\mathrm{GL}}$ (instead of U^s and $\hat{T}^{u,s}$), use N2SID software to extract estimations $\hat{A}, \hat{B}, \hat{C}, \hat{D}$ yielding the estimated system

$$\begin{cases} x(t+1) = \hat{A}x(t) + \hat{B}u^{\mathrm{GL}}(t) \\ y(t) = \hat{C}x(t) + \hat{D}u^{\mathrm{GL}}(t) \end{cases} \tag{21}$$

where $u^{\mathrm{GL}}(t) \in \mathbb{R}^{p_s}$ is given by

$$u^{\mathrm{GL}}(t) = \begin{bmatrix} u_{k_1}(t) & \dots & u_{k_{ps}}(t) \end{bmatrix}^T. \tag{22}$$

5. VALIDATION STUDY

5.1 Preliminaries

Simulations have been carried out in MATLAB using Mosek 7.1 as an SDP solver and N2SID software for estimating the system matrices, see Verhaegen and Hansson [2016].

For quantitative results, systems are randomly generated in MATLAB using the function `drss` which generates random, stable, discrete-time state-space models. Since the generated systems are on the form

$$\begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{23}$$

and the desired form is (5), some adjustments have to be made. As $e(t)$ may be treated as an input, a system with $p+1$ inputs is generated using `drss`. This yields $p+1$ columns in B and D respectively, and we assign $K = B_{:,p+1}$ and $D_{:,p+1} = \mathbf{1}$. Then, to avoid clumsy notation we redefine $B = B_{:,1:p}$ and $D = D_{:,1:p}$. Finally, given p_s significant inputs we assign $B_{:,p_s+1:p} = \mathbf{0}$ and $D_{:,p_s+1:p} = \mathbf{0}$. Thus, a random system of the form in (5) with p_s significant inputs has been achieved.

Since it makes sense to use Signal-to-Noise Ratio (SNR) for comparison, we must make sure that the amplification between the "regular" input channels and the output is the same as from the noise channel to the output. To achieve this, consider two systems where the matrices are defined as in the discussion above.

$$\mathcal{S}_u : \begin{cases} x(t+1) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{24}$$

$$\mathcal{S}_e : \begin{cases} x(t+1) = Ax(t) + Ke(t) \\ y(t) = Cx(t) + e(t) \end{cases} \tag{25}$$

With $\|\mathcal{S}_u\|_{\mathcal{H}_2}$ and $\|\mathcal{S}_e\|_{\mathcal{H}_2}$, we can define the system that will be simulated:

$$\mathcal{S} : \begin{cases} x(t+1) = Ax(t) + 1/\|\mathcal{S}_u\|_{\mathcal{H}_2} \cdot Bu(t) \\ \qquad\qquad + 1/\|\mathcal{S}_e\|_{\mathcal{H}_2} \cdot Ke(t) \\ y(t) = Cx(t) + 1/\|\mathcal{S}_u\|_{\mathcal{H}_2} \cdot Du(t) \\ \qquad\qquad + 1/\|\mathcal{S}_e\|_{\mathcal{H}_2} \cdot e(t) \end{cases} \tag{26}$$

For (26), the input to output channels and noise to output channels are equally normalized, making SNR of the inputs a consistent tool for comparison (see Zhou [1997] for an extensive description of the $\mathcal{H}_2$ norm).

Provided $\mathcal{S}$, the system will be simulated with inputs $u_1, ..., u_p$ as random telegraph signals. That is, $u_k(t) \in \{-1, 1\}$, $k \in \{1, ..., p\}$ randomly for $t = 1, .., N$, see Ljung [1999]. Since $\mathrm{Var}\,(u_k) = 1$ and $\mathrm{Var}\,(e) = \sigma^2$, the SNR between an input u_k, $k \in \{1, ..., p\}$ and the noise is then given by

$$\mathrm{SNR} = \frac{\mathrm{Var}\,(u_k)}{\mathrm{Var}\,(e)} = \frac{1}{\sigma^2}. \tag{27}$$

5.2 Experiments

In order to vindicate our new method of input selection, the following will be investigated:

i. **Input selection:** How well does our method perform the input selection? Are all non-significant inputs detected? Are all significant inputs correctly detected as such?
ii. **Model estimation:** Does our method, with the input selection, also provide good model estimations?

The first question will be investigated by implementing our input selection method on batches of measurement data generated by systems $\mathcal{S}_1^{\sigma_1}$, ..., $\mathcal{S}_M^{\sigma_1}$, ..., $\mathcal{S}_1^{\sigma_R}$, ..., $\mathcal{S}_M^{\sigma_R}$ respectively. Here, each $\mathcal{S}_m^{\sigma_r}$, $m \in \{1, ..., M\}$, $\sigma_r \in \{\sigma_1, ..., \sigma_R\}$ is a random system on the same form as $\mathcal{S}$ in (26). Each system $\mathcal{S}_m^{\sigma_r}$ has n states, p inputs, q outputs, noise level σ_r and random number of significant inputs p_s.

The outcome of a single input selection experiment is labeled as one of the following disjoint cases:

i. All significant inputs detected as significant. All non-significant inputs detected as non-significant.
ii. All significant inputs detected as significant. Some non-significant input detected as significant.
iii. Some significant input detected as non-significant.

For an input selection experiment on a series of systems $\mathcal{S}_1^{\sigma_r}$, ..., $\mathcal{S}_M^{\sigma_r}$ with fixed noise level σ_r, the result will be presented by the number of outcomes that we labeled (i), (ii) and (iii) respectively.

The second question will be investigated by checking the quality of the estimated model that is provided by our method. For this purpose, the so called Normalized Root Mean Square Error (NRMSE) fitness value will be used as the tool for determining the quality for an estimated model, see Ljung [1999].

Definition 4. Suppose y^v is output data generated from the system $\mathcal{S}$ with input data u^v. Let $\hat{\mathcal{S}}$ be an estimation of $\mathcal{S}$. If $\hat{y}^v$ is output generated from $\hat{\mathcal{S}}$ with input u^v, then the **NRMSE fitness value** for $\hat{\mathcal{S}}$ is given by

$$\mathrm{FIT}_{\hat{\mathcal{S}}} = 100 \cdot \left(1 - \frac{\|y^v - \hat{y}^v\|_2}{\|y^v - \mathrm{mean}\,(y^v)\|_2}\right) [\%] \tag{28}$$

We will consider batches of measurement data generated by systems $\mathcal{S}_1^{\sigma_1}$, ..., $\mathcal{S}_M^{\sigma_1}$, ..., $\mathcal{S}_1^{\sigma_R}$, ..., $\mathcal{S}_M^{\sigma_R}$ as before. Half of the data for each system is used to perform our method which includes estimation and input selection and the other half is used for validation. Let the respective input selected, estimated systems be given by $\hat{\mathcal{S}}_1^{\sigma_1}$, ..., $\hat{\mathcal{S}}_M^{\sigma_1}$, ..., $\hat{\mathcal{S}}_1^{\sigma_R}$, ..., $\hat{\mathcal{S}}_M^{\sigma_R}$. Then for a series of estimations $\hat{\mathcal{S}}_1^{\sigma_r}$, ..., $\hat{\mathcal{S}}_M^{\sigma_r}$ with fixed noise level σ_r, an average FIT value, $\mathrm{FIT}_{\hat{\mathcal{S}}^{\sigma_r}}$, can be defined as

$$\text{FIT}_{\hat{\mathcal{S}}^{\sigma_r}} = \frac{\text{FIT}_{\hat{\mathcal{S}}_1^{\sigma_r}} + ... + \text{FIT}_{\hat{\mathcal{S}}_M^{\sigma_r}}}{M}. \tag{29}$$

This will help us to see the quality of the estimations on average and how different noise levels affects the NRMSE fitness value.

Finally, the tests were conducted with a constant regularization parameter $\lambda_1 = 10$, $n = 6$ states, $p = 6$ inputs, $q = 1$ output(s), $p_s \in \{1, ..., p\}$ significant inputs (random) and $\sigma_r \in \text{logspace}\,(-3, -1/2, 10)$. The number of tests in each series was $M = 350$.

Remark 5. In the N2SID algorithm, an optimal λ_1 is calculated. Since, however, we are concerned with input selection properties which is related to λ_2, λ_1 remains fixed in the experiments.

5.3 Results

In figures 1, 2 and 3 there are plots of the results of input selection tests. For each figure, a different regularization parameter $\lambda_2 \in \{20, 25, 30\}$ was used. Furthermore, for each fixed λ_2, various threshold levels ε were used as indicated by the title of respective subplot.

As indicated by the results, it is imperative to select λ_2 and ε carefully. In Figure 1, for instance, λ_2 was chosen too small and thus non-significant inputs were not detected properly. This is because the corresponding columns in $\hat{T}^{u,s}$ are not sufficiently close to zero because of a relatively low penalty. While this may be remedied with a bigger threshold, the trade-off is that significant inputs may be incorrectly discarded. On the contrary, if λ_2 is too big, the proportion of wrongfully discarded significant inputs will grow as can be seen in Figure 3. As for Figure 2 where $\lambda_2 = 25$, we are closer to the optimal values of λ_2 and ε for this experimental setup. One should, however, still carefully note the apparent trade-off between the level of threshold ε and discarded significant inputs.

From these test results we draw the conclusion that the method is robust for different SNR levels and accurately performs input selection for a suitable choice of the parameters λ_2 and ε.

Now, as the performance of the method with regards to the input selection property has been established, the quality of the estimated model that the method provides will be addressed. For comparison, for each $\mathcal{S}_m^{\sigma_r}$, the following estimations will be performed:

i. **Oracle estimation:** As an omniscient oracle, we only include the significant inputs in the model estimation. Thus, the input selection is conducted theoretically and then a model estimated using N2SID method without the group lasso penalty. This estimation is in a sense an upper bound for the quality of an estimated model.
ii. **Naive estimation:** Here, both significant and non-significant inputs are included in the estimation. Thus, no input selection is performed and the N2SID method without the group lasso penalty term is used to estimate an model. Contrary to the oracle estimation, this estimation is in a sense a lower bound for the quality of an estimated model.
iii. **Input selection and estimation in one step (our method):** Here, we use our method where input selection and estimation of the model is carried out in the same step.
iv. **Input selection and estimation in two steps:** Here, our method is used in one step solely for input selection. Then, using the inputs that have been detected as significant, we estimate an model from scratch.

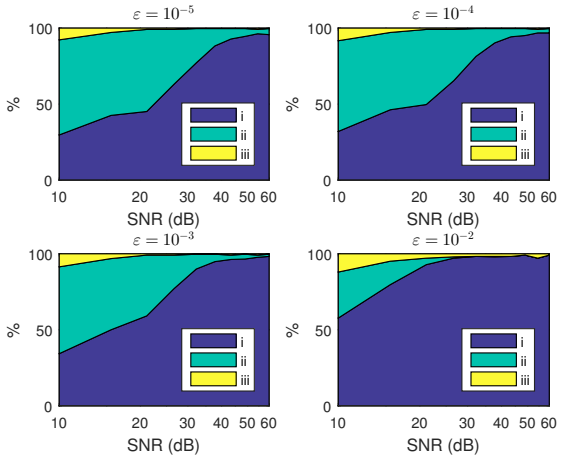

Fig. 1. Input selection test with $\lambda_2 = 20$ and threshold levels given above each single subplot.

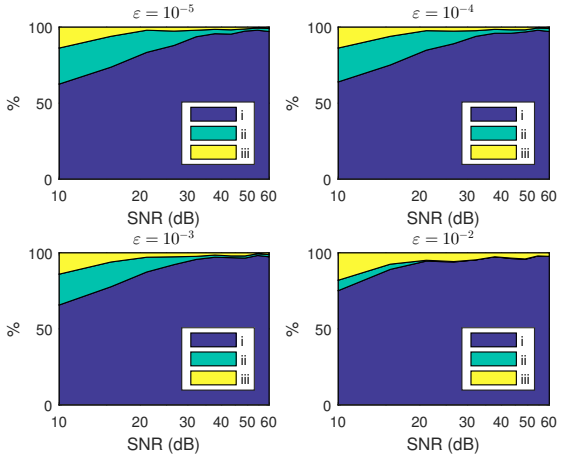

Fig. 2. Input selection test with $\lambda_2 = 25$ and threshold levels given above each single subplot.

In Figure 4 there are plots of the results of a fitness test with $\lambda_2 = 25$ and $\varepsilon = 0.01$. As can be seen in the plot, the oracle estimations are upper bounds and the naive estimations are lower bounds – both in terms of the average FIT. Furthermore, the reward of performing input selection and estimation in two steps is small compared to our method. Thus, we can conclude that our method provides a good model and that input selection and model estimation just as well can be performed in one single step. As also can be seen in the plot, the reward of using our method is larger for lower SNR:s.

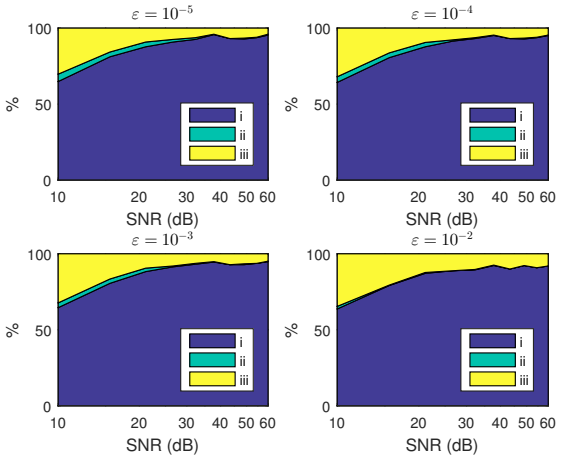

Fig. 3. Input selection test with $\lambda_2 = 30$ and threshold levels given above each single subplot.

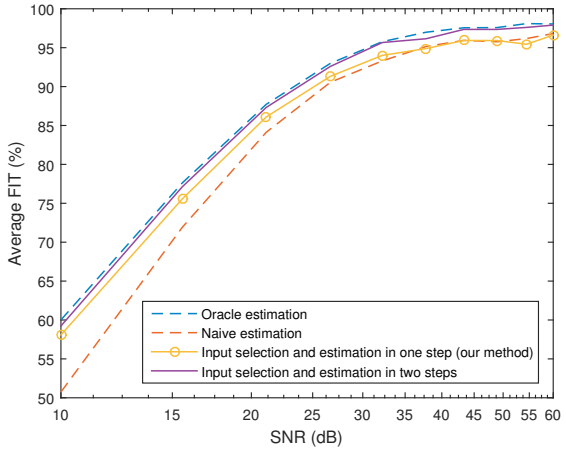

Fig. 4. Average FIT test with $\lambda_2 = 20$ and threshold level $\varepsilon = 0.001$. The upper solid line is the average fit using oracle estimation and the lower dashed line is the average fit using naive estimation. The solid line with circle markers is the average fit using input selection and estimation in one step (our method). The solid line without markers is the average fit using input selection and estimation in two steps.

6. CONCLUSIONS

In this paper, we have introduced a novel method of input selection and estimation in one step for time-discrete state-space models. This method works well and is robust as quantitative results have shown us – both in the input selection feature and in the sense of model fit.

Future work includes a more thorough investigation of the effects of λ_2 and ε and how they should be optimally chosen with regards to the noise levels. Also, software for a more efficient solver should be developed in the future as trials have been carried out in MATLAB with YALMIP.

Lastly, our method should be implemented on real applications where input selection is needed on a data set and compared with other methods of input selection.

REFERENCES

M. van de Wal and B. de Jager. A review of methods for input/output selection *Automatica vol. 37*, 2001.

C. R. Rojas, R. Tóth and H. Hjalmarsson. Sparse Estimation of Polynomial and Rational Dynamical Models *IEEE Transactions on Automatic Control vol. 59*, 2014.

K. Mohammadi et al. Identifying the most significant input parameters for predicting global solar radiation using an ANFIS selection procedure. *Renewable & Sustainable Energy Reviews vol. 63*, 2016.

J. S. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems vol. 23*, 1993.

H. D. Tran, N. Muttil and Perera B.J.C. Selection of significant input variables for time series forecasting. *Environmental Modelling & Software vol. 64*, 2015.

K. Fujiwara and M. Kano. Efficient input variable selection for soft-senor design based on nearest correlation spectral clustering and group Lasso. *ISA Transactions vol. 58*, 2015.

A. M. Kadhim, W. Birk and T. Gustafsson. Relative gain array estimation based on non-parametric frequency domain system identification. *Control Applications (CCA), IEEE Conference on IEEE*, 2014.

M. Verhaegen and A. Hansson. N2SID: Nuclear Norm Subspace Identification. 2015 [Online]. Available: http://arxiv.org/abs/1501.04495.

M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B vol. 68*, 2007.

J. Friedman, T. Hastie and R. Tibshirani. A note on the group Lasso and a sparse group Lasso. 2010 [Online]. Available: http://arxiv. org/pdf/1001.0736.

M. Verhaegen and A. Hansson. N2SID software. 2016 [Online]. Available: http://users.isy.liu.se/en/rt/hansson/n2sid.zip.

K. Zhou. Essentials of robust control. *Pearson Education, London, United Kingdom, 1st edition*, 1997.

L. Ljung. System Identification: Theory for the User. *Prentice-Hall, Upper Saddle River, N.J., 2nd edition*, 1999.

J. Löfberg. YALMIP : A Toolbox for Modeling and Optimization in MATLAB . *In Proceedings of the CACSD Conference, Taipei, Taiwan*, 2004.

MOSEK ApS. The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28), 2015 [Online]. Available: http://docs.mosek.com/7.1.