N2SID

Nuclear norm subspace identification of innovation models

Verhaegen, Michel; Hansson, A

DOI
10.1016/j.automatica.2016.05.021

Publication date
2016

Document Version
Accepted author manuscript

Published in
Automatica

N2SID: Nuclear Norm Subspace Identification of Innovation Models ⋆

Michel Verhaegen [a] Anders Hansson [b]

[a] *Delft Center for Systems and Control*
Delft University
Delft, The Netherlands

[b] *Division of Automatic Control*
Linköping University
Linköping, Sweden

Abstract

The identification of multivariable state space models in innovation form is solved in a subspace identification framework using convex nuclear norm optimization. The convex optimization approach allows to include constraints on the unknown matrices in the data-equation characterizing subspace identification methods, such as the lower triangular block-Toeplitz of weighting matrices constructed from the Markov parameters of the unknown observer. The classical use of instrumental variables to remove the influence of the innovation term on the data equation in subspace identification is avoided. The avoidance of the instrumental variable projection step has the potential to improve the accuracy of the estimated model predictions, especially for short data length sequences.

Key words: Subspace system identification, Optimization, Structural Constraints, Innovation State Space Models

1 Introduction

Subspace IDentification (SID) methods for the identification of Linear Time-Invariant (LTI) state space models as developed originally in [20, 10, 21] derive *approximate* models rather than models that are "optimal" with respect to a goodness of fit criterion defined in terms of the weighted norm of the difference between the measured output and the model predicted output. The approximation is based on linear algebra transformations and factorizations of structured Hankel matrices constructed from the input-output data that are related via the so-called *data equation* [24]. All existing SID methods aim to derive a low rank matrix from which key subspaces, hence the name subspace identification, are derived [28]. The low rank approximation is in general done using a Singular Value Decomposition (SVD).

A number of recent developments have been made to integrate the low rank approximation step in SID with a goodness of fit into a single multi-criteria convex optimization problem. These contributions were inspired by the work in [3] to approximate a constraint on the rank of a matrix by minimizing its nuclear norm. It resulted into a number of improvements to the low rank approximation step over the classically used SVD in SID, [12, 13, 16, 5, 7, 11, 18, 19].

When considering identifying innovation state space models, a common approach is to make use of instrumental variables [7] . It is well known that the projection operation related to the use of instrumental variables may result into a degradation of the accuracy of the estimated quantities.

In this paper we present a new SID method for identifying multivariable state space models in innovation form within the framework of nuclear norm optimization. The new SID method avoids the use of instrumental variables. The method is a convex relaxation of Pareto opti-

⋆ Part of the research was done while the first author was a Visiting Professor at the Division of Automatic Control, Department of Electrical Engineering, Linköping University, Sweden. This work was partially supported by the European Research Council Advanced Grant Agreement No. 339681. A preliminary version of the paper was presented at the 19th IFAC World Congress in South Africa, 2014. Corresponding Author.
m.verhaegen@tudelft.nl.

mization in which structural constraints are imposed on the unknowns in the data equation, such as their block-Toeplitz matrix structure. This Pareto optimization approach allows to make a trade-off between a Prediction Error type of optimality criteria, that is minimizing the (co-)variance of the one-step ahead prediction of a linear Kalman filter type observer, on one hand, and finding an observer of lowest complexity, i.e. of lowest model order, on the other hand. The key result is that the structural Toeplitz constraint is sufficient to find the minimal observer realization when the optimal one-step ahead prediction of the output is known. The incentive to estimate a Kalman filter type of observer also justifies the constraint to attempt to minimize the variance of the one-step ahead prediction error.

It is interesting to note that this key result stipulates precise conditions on the persistancy of excitation of the input (in open-loop experiments). For many instrumental variable based SID methods it is still an open question what the persistency of excitation condition is on a generic input sequence to guarantee the algorithm to work for finite data length samples or to be consistent [9].

The convex relaxation of the new SID approach is denoted by Nuclear Norm Subspace IDentification (N2SID). The foundations for N2SID were presented in [22]. There the resulting optimization problem was solved using a Semi-Definite Programming (SDP) solver after a reformulation of the problem into an equivalent SDP problem. In addition to this the problem formulation was approximated in order to obtain a problem of manageable size for current SDP solvers. To speed up the calculations or to be able to treat longer data sequences, a new implementation in the ADMM framework has been developed. Due to the restriction of this paper, we refer the reader to [23] for a description of this new implementation. In [23] the interested reader can also find experimental results from a comparison study of N2SID with two other SID methods, N4SID and the recent Nuclear Norm based SID method presented in [11] and with PEM. This comparison study makes use of real-life data batches of the DaSIy library [2].

The paper is organized as follows. In Section 2 the identification problem for identifying a multi-variable state space model in a subspace context while taking a prediction error cost function into consideration is presented. The data equation and preliminaries on the assumptions made in the analysis and description of the subspace identification method are presented in Section 3. The multi-criteria optimization problem, the analysis of the uniqueness of solution and its convex relaxation are presented in Section 4. Finally, we end this paper with some concluding remarks.

1.1 Notations

We introduce the Matlab-like notation that for a vector or matrix $X \in \mathbb{R}^{M\times N}$ $\left(\mathbb{C}^{M\times N}\right)$ it holds that $X(m:n, p:q)$ is the sub-matrix of X with rows m through n and columns p through q.

2 The Subspace Identification Problem

In system identification a challenging problem is to identify Linear Time Invariant (LTI) systems with multiple inputs and multiple outputs using short length data sequences. Taking process and measurement noise into consideration, a general state space model for LTI systems can be given in so-called innovation form, [24]:

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

with $x(k) \in \mathbb{R}^n, y(k) \in \mathbb{R}^p, u(k) \in \mathbb{R}^m$ and $e(k)$ a zero-mean white noise sequence. Since we are interested in short data sets no requirement on consistency is included in the following problem formulation.

Problem Formulation: Given the input-ouput (i/o) data batches $\{u(k), y(k)\}_{k=1}^{N}$, with $N > n$ and assumed to be retrieved from an identification experiment with a system belonging to the class of LTI systems as represented by (1), the problem is to determine approximate system matrices $(\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{D}, \hat{K}_T)$ that define the $\hat{n}$-th order observer of *"low"* complexity:

$$\begin{cases} \hat{x}_T(k+1) \;=\; \hat{A}_T\hat{x}_T(k) + \hat{B}_T u_v(k) + \hat{K}_T\Big(y_v(k) - \hat{C}_T\hat{x}_T\Big) \\ \quad \hat{y}_v(k) \;=\; \hat{C}_T\hat{x}_T(k) + \hat{D}u_v(k) \end{cases} \tag{2}$$

such that the approximated output $\hat{y}_v(k)$ is *"close"* to the measured output $y_v(k)$ of the validation pair $\{u_v(k), y_v(k)\}_{k=1}^{N_v}$ as expressed by a small value of the cost function,

$$\frac{1}{N_v}\sum_{k=1}^{N_v} \|y_v(k) - \hat{y}_v(k)\|_2^2. \tag{3}$$

The quantitative notions like *"low"* and *"close approximation"* will be made more precise later on. The solution to this problem is provided under two assumptions. The first is listed here, the second at the end of Section 3.

Assumption A.1. The pair (A, C) is observable and the pair $(A, \begin{bmatrix} B & K \end{bmatrix})$ is reachable.

A key starting point in the formulation of subspace methods is the relation between structured Hankel matrices

constructed from the i/o data. This relationship will as defined in [24] be called the data equation. It will be presented in the next section.

3 The Data Equation, its structure and Preliminaries

Let the LTI model (1) be represented in its so-called observer form:

$$\begin{cases} x(k+1) = (A-KC)x(k) + (B-KD)u(k) + Ky(k) \\ y(k) = Cx(k) + Du(k) + e(k) \end{cases} \tag{4}$$

We will denote this model compactly as:

$$\begin{cases} x(k+1) = \mathcal{A}x(k) + \mathcal{B}u(k) + Ky(k) \\ y(k) = Cx(k) + Du(k) + e(k) \end{cases} \tag{5}$$

with $\mathcal{A}$ the observer system matrix $(A-KC)$ and $\mathcal{B}$ equal to $(B-KD)$. Though this property will not be used in the sequel, the matrix $\mathcal{A}$ can be assumed to be asymptotically stable.

For the construction of the data equation, we store the measured i/o data in block-Hankel matrices. For fixed N assumed to be larger then the order n of the underlying system, the definition of the number of block-rows fully defines the size of these Hankel matrices. Let this dimensioning parameter be denoted by s, then the Hankel matrix of the input is defined as:

$$U_{s,N} = \begin{bmatrix} u(1) & u(2) & \cdots & u(N-s+1) \\ u(2) & u(3) & & \vdots \\ \vdots & & \ddots & \\ u(s) & u(s+1) & \cdots & u(N) \end{bmatrix}. \tag{6}$$

The Hankel matrices from the output $y(k)$ and the innovation $e(k)$ are defined similarly and denoted by $Y_{s,N}$ and $E_{s,N}$, respectively. The relationship between these Hankel matrices, that readily follows from the linear model equations in (5), require the definition of the following *structured* matrices. First we define the extended observability matrix $\mathcal{O}_s$:

$$\mathcal{O}_s^T = \begin{bmatrix} C^T & \mathcal{A}^T C^T & \cdots & \mathcal{A}^{T\,s-1} C^T \end{bmatrix}. \tag{7}$$

Second, we define a Toeplitz matrix from the quadruple of systems matrices $\{\mathcal{A}, \mathcal{B}, C, D\}$ as:

$$T_{u,s} = \begin{bmatrix} D & 0 & \cdots & 0 \\ C\mathcal{B} & D & & 0 \\ \vdots & & \ddots & \\ C\mathcal{A}^{s-2}\mathcal{B} & & \cdots & D \end{bmatrix} \tag{8}$$

and in the same way we define a Toeplitz matrix $T_{y,s}$ from the quadruple $\{\mathcal{A}, K, C, 0\}$. Finally, let the state sequence be stored as:

$$X_N = \begin{bmatrix} x(1) & x(2) & \cdots & x(N-s+1) \end{bmatrix}. \tag{9}$$

Then the data equation compactly reads:

$$Y_{s,N} = \mathcal{O}_s X_N + T_{u,s} U_{s,N} + T_{y,s} Y_{s,N} + E_{s,N}. \tag{10}$$

This equation is a simple linear matrix equation that highlights the challenges in subspace identification, which is to approximate from the given Hankel matrices $Y_{s,N}$ and $U_{s,N}$ the column space of the observability matrix and/or that of the state sequence of the observer (5).

The equation is highly structured. In this paper we focus on the following key structural properties of the unknown matrices in (10):

(1) The matrix product $\mathcal{O}_s X_N$ is *low rank* when $s > n$.
(2) The matrices $T_{u,s}$ and $T_{y,s}$ are block-Toeplitz.
(3) The matrix $E_{s,N}$ is block-Hankel.

The interesting observation is that these three structural properties can all be considered in the multi-criteria optimization problem while preserving convexity. This is demonstrated in Section 4.

The analysis in Section 4 requires the following preliminaries.

Definition 1 *[24]: A signal $u(k) \in \mathbb{R}^m$ is persistently exciting of order s if and only if there exists an integer N such that the matrix $U_{s,N}$ has full row rank.*

Lemma 2 *[9]: Consider the state space model in innovation form* (1) *and let all stochastic signals be stationary and ergodic, let Assumption A.1 be satisfied and let the input $u(k)$ be quasi-stationary [15] and persistently exciting of order $s+n$, then:*

$$\lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} X_N \\ U_{s,N} \end{bmatrix} \begin{bmatrix} X_N^T & U_{s,N}^T \end{bmatrix} > 0$$

Corollary 3 *Let the conditions stipulated in Lemma 2 hold, and let $u(k)$ be statistically independent from the innovation sequence $e(\ell)$ for all k, ℓ, then,*

$$\lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} X_N \\ U_{s,N} \\ Y_{s,N} \end{bmatrix} \begin{bmatrix} X_N^T & U_{s,N}^T & Y_{s,N}^T \end{bmatrix} > 0$$

Proof: Since $e(k)$ is white noise, it follows that $\mathbb{E}[x(k)e(\ell)^T] = 0$ (with $\mathbb{E}$ denoting the expectation operator), for $\ell \geq k$. This in combination with the independency between $u(k)$ and $e(\ell)$, the white noise property of $e(k)$ and the ergodicity or the quasi-stationarity of the signals yields,

$$\lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} X_N \\ U_{s,N} \\ E_{s,N} \end{bmatrix} \begin{bmatrix} X_N^T & U_{s,N}^T & E_{s,N}^T \end{bmatrix} > 0 \qquad (11)$$

Consider model (1) and let the block-Toeplitz matrices $T'_{u,s}$ and $T_{e,s}$ be defined as the Toeplitz matrix $T_{u,s}$ in (8) but from the quadruples (A, B, C, D) and (A, K, C, I), respectively. Let $O_s^T = \begin{bmatrix} C^T & A^T C^T & \cdots & A^{T\,s-1} C^T \end{bmatrix}$. Then we can state the following alternative data equation:

$$Y_{s,N} = O_s X_N + T'_{u,s} U_{s,N} + T_{e,s} E_{s,N}$$

By this data equation, we have that,

$$\begin{bmatrix} X_N \\ U_{s,N} \\ Y_{s,N} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ O_s & T'_{u,s} & T_{e,s} \end{bmatrix} \begin{bmatrix} X_N \\ U_{s,N} \\ E_{s,N} \end{bmatrix}$$

The results follows from (11) and the fact that the matrix $T_{e,s}$ is square and invertible. □

Based on this result the following assumption is stipulated.

Assumption A.2. Consider the model (5), then there exists an integer N such that the compound matrix,

$$\begin{bmatrix} X_N \\ U_{s,N} \\ Y_{s,N} \end{bmatrix}$$

has full row rank.

4 N2SID

4.1 Pareto optimal Subspace Identification

When assuming the optimal observer given, the quantity $\hat{y}(k)$ is the minimum variance prediction of the output and equal to $y(k) - e(k)$. Let the Hankel matrix $\hat{Y}_{s,N}$ be defined from this sequence $\hat{y}(k)$ as we defined $Y_{s,N}$ from $y(k)$. Then the data equation (10) can be reformulated into:

$$\hat{Y}_{s,N} = \mathcal{O}_s X_N + T_{u,s} U_{s,N} + T_{y,s} Y_{s,N}. \qquad (12)$$

Let $\mathcal{T}_{p,m}$ denote the class of lower triangular block-Toeplitz matrices with block entries $p \times m$ matrices and let $\mathcal{H}_p$ denote the class of block-Hankel matrices with block entries of p column vectors. Then the three key structural properties listed in Section 3 are taken into account in an optimization problem seeking a trade-off between the following cost functions,

$$\begin{aligned} &\operatorname{rank}\Big(\Gamma_{s,N} - \Theta_u U_{s,N} - \Theta_y Y_{s,N}\Big) \\ &\text{and} \quad \operatorname{Tr}\mathbb{E}\left[\Big(y(k) - \gamma(k)\Big)\Big(y(k) - \gamma(k)\Big)^T\right] \end{aligned} \qquad (13)$$

Here $\mathbb{E}$ denotes the expectation operator. The optimization variables in the cost function are $\Gamma_{s,N}$ with entries $\gamma(k)$ and Θ_u, Θ_y. The matrix $\Gamma_{s,N} \in \mathcal{H}_p$ is the (block-) Hankel matrix approximating the Hankel matrix $\hat{Y}_{s,N}$ and constructed from the approximation of the one-step ahead prediction of the output denoted by $\gamma(k)$ in the same way $\hat{Y}_{s,N}$ was constructed from $\hat{y}(k)$. Further, we have the following constraints on the matrices $\Theta_u \in \mathcal{T}_{p,m}$ and $\Theta_y \in \mathcal{T}_{p,p}$.

An optimal trade-off between the above two cost functions is called a Pareto optimal solution. Moreover, the Pareto optimization problem is not tractable. For that purpose we will develop in the next subsection a *convex relaxation* of the cost functions.

Before stating this convex relaxation an analysis is made on additional constraints that can be imposed on the block-Toeplitz matrices Θ_u and Θ_y and/or under what conditions their block-Toeplitz structure is sufficient to find a unique solution.

4.2 Additional structure in the block-Toeplitz matrices $T_{u,s}$ and $T_{y,.s}$

In this section we analyse the additional structure present in the block-Toeplitz matrices $T_{u,s}$ and $T_{y,.s}$ as well as the conditions under which the block-Toeplitz structure is sufficient to find the system matrices $(\mathcal{A}_T, \mathcal{B}_T, C_T, D, K_T)$. These conditions are summarized in Theorem 1 of this paper. First, we state in Lemma 4,

the additional structure in the block-Toeplitz matrices $T_{u,s}$ and $T_{y,.s}$.

Lemma 4 *Let $s > n$. Then we can partition the block-Toeplitz matrices $T_{u,s}$ and $T_{y,.s}$, defined in the data equation* (10) *as,*

$$T_{u,s} = \left[\begin{array}{c|c} T_{u,n} & 0 \\ \hline H_{u,s-n} \mid T_{u,s-n} \end{array}\right] \tag{14}$$

and likewise for the matrix $T_{y,s}$. Here the matrices $H_{u,s-n}$ and $H_{y,s-n}$ can be decomposed as,

$$\left[\begin{array}{c|c} H_{u,s-n} & H_{y,s-n} \end{array}\right] = \begin{bmatrix} C \\ C\mathcal{A} \\ \vdots \\ C\mathcal{A}^{s-n-1} \end{bmatrix} \left[\begin{array}{ccc|ccc} \mathcal{A}^{n-1}\mathcal{B} & \cdots & \mathcal{B} & \mathcal{A}^{n-1}K & \cdots & K \end{array}\right] \tag{15}$$

Proof: Follows by construction. □

Remark 5 *Lemma 4 can be used to impose an additional constraint on the block-Toeplitz matrices Θ_u and Θ_y. If we partition these block-Toeplitz matrices conformal their counterparts $T_{u,s}$ and $T_{y,s}$ as highlighted in Lemma 4, as follows,*

$$\Theta_{u,s} = \left[\begin{array}{c|c} \Theta_{u,n} & 0 \\ \hline H^{\Theta}_{u,s-n} \mid \Theta_{u,s-n} \end{array}\right]$$

(likewise for $\Theta_{y,s}$), then for the case $s \geq 2n$ we can impose the following additional constraint,

$$\text{rank}\Big(\begin{bmatrix} H^{\Theta}_{u,s-n} & H^{\Theta}_{y,s-n} \end{bmatrix} \Big) = n$$

The additional constraint highlighted in Remark 5 can be reformulated, as done e.g. in [18, 19], as a rank minimization constraint, that can be relaxed to a convex constraint using the nuclear norm. However we seek to avoid imposing this additional constraint in order to minimize the number of regularization parameters. The basis herefore is provided in the next theorem.

Theorem 1 *Consider the observer in* (1) *with $x(k) \in \mathbb{R}^n$ and consider the rank optimization problem in Eq.* (13) *only with $\Gamma_{s,N}$ fixed to $\hat{Y}_{s,N}$, let $s > n$ and let Assumptions A.1 and A.2 be satisfied, Then,*

$$\min_{\Theta_u \in \mathcal{T}_{p,m}, \Theta_y \in \mathcal{T}_{p,p}} \text{rank}\Big(\hat{Y}_{s,N} - \Theta_u U_{s,N} - \Theta_y Y_{s,N}\Big) = n$$

Further the arguments optimizing the above optimization problem, denoted as $\hat{\Theta}_u, \hat{\Theta}_y$ are unique and equal to,

$$\hat{\Theta}_u = T_{u,s} \quad \hat{\Theta}_y = T_{y,s}$$

with $T_{u,s}, T_{y,s}$ the true underlying block-Toeplitz matrices in the data equation(10).

Proof: Let $\delta_u \in \mathcal{T}_{p,m}, \delta_y \in \mathcal{T}_{p,p}$, then,

$$\begin{aligned} \hat{Y}_{s,N} - \Theta_u U_{s,N} - \Theta_y Y_{s,N} &= \hat{Y}_{s,N} - (T_{u,s} + \delta_u) U_{s,N} - \\ &\quad (T_{y,s} + \delta_y) Y_{s,N} \\ &= \mathcal{O}_s X_N - \delta_u U_{s,N} - \delta_y Y_{s,N} \end{aligned}$$

Therefore,

$$\text{rank}\Big(\hat{Y}_{s,N} - \Theta_u U_{s,N} - \Theta_y Y_{s,N}\Big) =$$

$$\text{rank}\Big(\mathcal{O}_s X_N - \delta_u U_{s,N} - \delta_y Y_{s,N}\Big)$$

Application of Sylvester's inequality [24] and under Assumption A.2, we further have,

$$\text{rank}\Big(\hat{Y}_{s,N} - \Theta_u U_{s,N} - \Theta_y Y_{s,N}\Big) = \text{rank}\Big(\begin{bmatrix} \mathcal{O}_s & \delta_u & \delta_y \end{bmatrix} \Big) \tag{16}$$

First notice that under Assumption A.1 the rank of this matrix is n for $\delta_u = 0$ and $\delta_y = 0$. Since the $\text{rank}\Big(\begin{bmatrix} \mathcal{O}_s & \delta_u & \delta_y \end{bmatrix} \Big) \geq \text{rank}\ \Big(\mathcal{O}_s\Big)$ for all δ_u, δ_y, we have that n is the minimal value of the rank in (16).

It will now be shown that this minimal value of the rank, can only be reached for both δ_u and δ_y equal to zero.

For that purpose, let $t = \{t_i \in \mathbb{R}^{p \times (m+p)}\}_{i=1}^{s}$ be a sequence of arbitrary matrices that define the lower triangular block-Toeplitz matrix $\Delta^s(t)$ as:

$$\Delta^s(t) = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ t_2 & t_1 & & \vdots \\ \vdots & & \ddots & \vdots \\ t_s & t_{s-1} & \cdots & t_1 \end{bmatrix} \in \mathbb{R}^{sp \times s(m+p)}$$

The columns of the compound matrix $\begin{bmatrix} \delta_u & \delta_y \end{bmatrix}$ in (16) can always be permuted into a matrix of the form $\Delta^s(t)$ and since column permutations do not change the rank of a matrix we have that,

$$\text{rank}\Big(\begin{bmatrix} \mathcal{O}_s & \delta_u & \delta_y \end{bmatrix} \Big) = \text{rank}\Big(\begin{bmatrix} \mathcal{O}_s & \Delta^s(t) \end{bmatrix} \Big)$$

Now we show that the following condition

$$\text{rank}\Big(\begin{bmatrix} \mathcal{O}_s & \Delta^s(t) \end{bmatrix} \Big) = n$$

implies that $\Delta^s(t)$ has to be zero. In order for the above rank constraint to hold we need $\Delta^s(t)$ to be of the following form:

$$\begin{bmatrix} t_1 & 0 & \cdots & 0 & 0 \\ t_2 & t_1 & & 0 & 0 \\ \vdots & & \ddots & & \\ t_s & t_{s-1} & \cdots & t_2 & t_1 \end{bmatrix} = \mathcal{O}_s \begin{bmatrix} q_1 & q_2 & \cdots & q_{s-1} & q_s \end{bmatrix} \qquad (17)$$

The fact that $s > n$, we have that $\text{rank}\Big(\mathcal{O}_{s-1}\Big) = n$ and therefore we can deduce from the first $p(s-1)$ rows of the last $p+m$ columns in the expression (17) that,

$$q_s = 0 \Rightarrow t_1 = 0$$

Using this result, and the Toeplitz structure of $\Delta^s(t)$, we can in the same way conclude from the first $p(s-1)$ rows and from the columns $(s-2)(m+p)+1$ to $(s-1)(m+p)$ in (17) that,

$$q_{s-1} = 0 \Rightarrow t_2 = 0 \quad \text{etc.}$$

Hence there cannot be a $\Delta^s(t)$ with the given Toeplitz structure that is different from zero such that $\text{rank}\Big(\begin{bmatrix} \mathcal{O}_s & \Delta^s(t) \end{bmatrix} \Big) = n$. Hence the minimal value of the rank of the matrix $\begin{bmatrix} \mathcal{O}_s & \delta_u & \delta_y \end{bmatrix}$ in (16) w.r.t. δ_u, δ_y yields zero value of both. This concludes the proof. □

4.3 A convex relaxation

A convex relaxation of the NP hard problem formulation in (13) will now be developed. The original problem is reformulated in two ways. First, the rank operator is substituted by the nuclear norm. The nuclear norm of a matrix X denoted by $\|X\|_\star$ is defined as the sum of the singular values of the matrix X. It is also known as the trace norm, the Ky Fan norm or the Schatten norm, [14]. This is known to be a good approximation of the rank operator when it is to be minimized, [4, 3]. Second, the minimum variance criterion is substituted by the following sample average $\frac{1}{N}\sum_{k=1}^{N} \|y(k) - \gamma(k)\|_2^2$.
By introducing a scalarization parameter $\lambda \in [0, \infty)$, which can be interpreted as a regularization parameter, all Pareto optimal solutions of the convex reformulation of the N2SID problem can be obtained by solving:

$$\begin{aligned} &\min_{\Gamma_{s,N} \in \mathcal{H}_p, \Theta_{u,s} \in \mathcal{T}_{p,m}, \Theta_{y,s} \in \mathcal{T}_{p,p}} \|\Gamma_{s,N} - \Theta_{u,s}U_{s,N} - \Theta_{y,s}Y_{s,N}\|_\star \\ &+ \tfrac{\lambda}{N}\sum_{k=1}^{N} \|y(k) - \gamma(k)\|_2^2 \end{aligned} \qquad (18)$$

for all values of $\lambda \in [0, \infty)$.

Remark 4 *The method can be extended to other related identification problems. As a first example we consider the identification problem of an innovation model with absence of a measurable input, is to consider the following convex relaxed problem formulation:*

$$\begin{aligned} &\min_{\Gamma_{s,N} \in \mathcal{H}_p, \Theta_{y,s} \in \mathcal{T}_{p,p}} \|\Gamma_{s,N} - \Theta_{y,s}Y_{s,N}\|_\star + \\ &\tfrac{\lambda}{N}\sum_{k=1}^{N} \|y(k) - \gamma(k)\|_2^2. \end{aligned} \qquad (19)$$

A second example is the case of the Kalman gain K in (1) to be zero. In that case the parameter (matrix) Θ_y in (13) is simply set to zero. Such a constraint, if valid, will again improve the results.

It is well-known that the problem (18) can be recast as a Semi-Definite Programming (SDP) problem, [4, 3], and hence it can be solved in polynomial time with standard SDP solvers. The reformulation, however, introduces additional matrix variables of dimension $N \times N$. Therefore limiting the size of problems possible to solve when using current SDP solvers [22]. In [23] an alternative exact method using ADMM is presented that was inspired by its successful application in [11].

4.4 Calculation of the system matrices

The convex-optimization problem (18) yields the estimates of the quantities $\Gamma_{s,N}, \Theta_{u,s}$ and $\Theta_{y,s}$. Since the outcome depends on the regularization parameter λ, let us denote these estimates as $\hat{\Gamma}_{s,N}(\lambda), \hat{\Theta}_{u,s}(\lambda)$ and $\hat{\Theta}_{y,s}(\lambda)$ respectively. The determination of the system matrices starts with an SVD of the "low rank" approximated matrix as follows:

$$\begin{aligned} &\hat{\Gamma}_{s,N}(\lambda) - \hat{\Theta}_{u,s}(\lambda)U_{s,N} - \hat{\Theta}_{y,s}(\lambda)Y_{s,N} = \\ &\qquad \Big[\, U_{\hat{n}}(\lambda) \mid \star \,\Big] \left[\begin{array}{c|c} \Sigma_{\hat{n}}(\lambda) & 0 \\ \hline 0 & \star \end{array}\right] \left[\begin{array}{c} V_{\hat{n}}^T(\lambda) \\ \hline \star \end{array}\right] \end{aligned} \qquad (20)$$

where $\hat{n}$ is an integer denoting the $\hat{n}$ largest singular values and the notation $\star$ denotes a compatible matrix not of interest here. The selection of $\hat{n}$ is outlined in the algorithmic description given next.

The algorithm requires in addition to the input-output data sequences the user to specify the parameter s to fix the number of block rows in the block-Hankel matrices $U_{s,N}$ and $Y_{s,N}$ and an interval for the parameter λ denoted by $\Lambda = [\lambda_{\min}, \lambda_{\max}]$. As for the implementation described in [11] the identification data set could be partitioned in two parts to avoid overfitting. The first part is referred to as the `ide-1` part of the identification data set and the remaining part of the identification data set is referred to as the `ide-2` part. This splitting of the data set was recommended in [11] to avoid overfitting.

The `N2SID algorithm` is summarized as follows. We start with gridding the interval $\Lambda = [\lambda_{\min}, \lambda_{\max}]$ into L different points, e.g. using the Matlab notation $\Lambda = \texttt{logspace}\Big(\log(\lambda_{\min}), \log(\lambda_{\max}), L\Big)$. The set Λ is defined by $\Lambda = \{\lambda_i\}_{i=1}^{L}$.

For each $\lambda_i \in \Lambda$ we do the following (leaving out for the sake of brevity the dependency of the estimated system matrices on this parameter λ_i):

(1) *Order Selection:* Solve (18) using the data set `ide-1` and compute the SVD as in (20). Based on the singular values of this SVD the model order $\hat{n}$ is selected. This can be done manually by the user or automatically. Such automatic selection can be done as in the `N4SID` implementation in [15] as highlighted in [11]: order the singular values in (20) in descending order, then select that index of the singular value that in logarithm is closest to the logarithmic mean of the maximum and minimum singular values in (20).

(2) *Computing the pair $(\hat{\mathcal{A}}_T, \hat{C}_T)$:* From the SVD in (20), and the selected model order $\hat{n}$, the pair is derived from the matrix $U_{\hat{n}}(\lambda_i)$ as done in classical SID methods by considering $U_{\hat{n}}(\lambda_i)$ to be an approximation of the extended observability matrix $\mathcal{O}_s$, see e.g. [24].

(3) *Computing the Kalman gain $\hat{K}_T$:* With $U_{\hat{n}}(\lambda_i)$ and the estimated matrix $\hat{\Theta}_{y,s}(\lambda_i)$ we exploit that the latter matrix approximates the block-Toeplitz matrix $T_{y,s}$ to estimate the observer gain via the solution of a standard linear least squares problem. This is seen as follows. Let us assume the block Toeplitz matrix $T_{y,s}$ be given and denoted explicitly as,

$$T_{y,s} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ CK & 0 & & 0 & 0 \\ C\mathcal{A}K & CK & & 0 & 0 \\ \vdots & & & \ddots & \\ C\mathcal{A}^{s-2}K & & \cdots & CK & 0 \end{bmatrix}$$

If we know the matrix $\mathcal{O}_s$, we can write the following set of equations,

$$\mathcal{O}_s(1:(s-1)p,:)K = T_{y,s}(p+1:ps, 1:p)$$

Let us now denote the first $(s-1)p$ rows of the matrix $U_{\hat{n}}(\lambda_i)$ by $\hat{\mathcal{O}}_{s-1,T}$ and let us denote the submatrix of the matrix $\hat{\Theta}_{y,s}(\lambda_i)$ from rows $p+1$ to row ps and from column 1 to p by $\hat{T}_{y,s}(p+1:ps, 1:p)$, then we can estimate K_T from:

$$\min_{K_T} \|\hat{\mathcal{O}}_{s-1,T} K_T - \hat{T}_{y,s}(p+1:ps, 1:p)\|^2 \qquad (21)$$

The estimate of the observer gain is used to estimate the system matrix A_T as:

$$\hat{A}_T = \hat{\mathcal{A}}_T + \hat{K}_T \hat{C}_T \qquad (22)$$

(4) *Computing the pair $(\hat{\mathcal{B}}_T, \hat{D})$ and the initial conditions:* Let the approximation of the observer be denoted as:

$$\begin{aligned} \hat{x}_T(k+1) &= \hat{\mathcal{A}}_T \hat{x}_T(k) + \hat{\mathcal{B}}_T u(k) + \hat{K}_T y(k) \\ \hat{y}(k) &= \hat{C}_T \hat{x}_T(k) + \hat{D}u(k) \end{aligned} \qquad (23)$$

Then the estimation of the pair $\hat{\mathcal{B}}_T, \hat{D}$ and the initial conditions of the above observer can again be done via a linear least squares problem as outlined in [24] by minimizing the RMS value of the prediction error $y(k) - \hat{y}(k)$ obtained from the identification data in `ide-1`. The estimated input matrix $\hat{B}_T$ is then determined as:

$$\hat{B}_T = \hat{\mathcal{B}}_T + \hat{K}_T D \qquad (24)$$

(5) *Evaluating Goodness of fit:* Using the computed quadruple of system matrices $\{\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{D}\}$ and the validation data in `ide-2` of length $N_{\text{ide}-2}$ we calculate the simulated output $\hat{y}(k, \lambda_i)$ as,

$$\begin{aligned} \hat{x}_T(k+1) &= \hat{A}_T \hat{x}_T(k) + \hat{B}_T u(k) \\ \hat{y}(k, \lambda_i) &= \hat{C}_T \hat{x}_T(k) + \hat{D}u(k) \end{aligned} \qquad (25)$$

and evaluate the cost function $J(\lambda_i) = \sum_{k=1}^{N_{\text{ide}-2}} \|y(k) - \hat{y}(k, \lambda_i)\|_2^2$.

The selected output of the `N2SID` algorithms is that model that corresponds to the optimal λ_{opt} given as:

$$\lambda_{\text{opt}} = \min_{\lambda \in \Lambda} J(\lambda)$$

It should be remarked that other variants to compute the system matrices are possible as is done for classical subspace identification methods. For the sake of brevity we refer the interested reader to [23]. Here the interested reader can also find an efficient implementation in the ADMM framework and a more elaborate validation study comparing the new N2SID method with other existing subspace identification methods as well as the Prediction Error method.

5 Concluding Remarks

Subspace identification of multivariable state space innovation models is revisited in this paper in the scope of nuclear norm optimization methods and using the observer form. A new subspace identification method is presented, referred to as N2SID. N2SID is the first subspace identification method that addresses the identification of innovation state space models *without* the use of instrumental variables (IVs). The avoidance of using IVs leads to a number of improvements. First as shown in the experimental study in [23], it leads to improved results in identifying innovation models when compared to existing SID methods, like N4SID and the recent Nuclear Norm based SID methods presented in [11] and with the Prediction Error Method (PEM) [15]. This improvement especially holds for small length data batches, i.e. when the number of samples is only a small multiple of the order of the underlying system. Second, as illustrated by Theorem 1, the methodology presented enables to provide insight on the necessary conditions of persistency of

excitation of the input on the existance of a unique solution. Finally, the new N2SID methodology will enable to address other interesting identification problems in a subspace identification framework, such as the identification of distributed systems as shown in [25, 26, 27].

Acknowledgement: The authors kindly acknowledge Mr. Baptiste Sinquin from the Ecole Centrale Lyon for his help in a preliminary matlab comparison study with N2SID during his internship at the Delft Center for Systems and Control under the supervision of Prof. M. Verhaegen. Also the discussions with Dr. Chengpu Yu of the Delft Center for Systems and Control on the topic of Theorem 1 are very much appreciated.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1-122, 2011. Michael Jordan, Editor in Chief.

[2] B. De Moor, P. De Gersem, B. De Schutter, and W. Favoreel. DAISY: A database for the identification of systems. *Journal A: Special Issue on CACSD (Computer Aided COntrol System Design)*, 38(3):4–5, Sep. 1997.

[3] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, 2002.

[4] M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, pages 4734–4739, 2001.

[5] M. Fazel, T. K. Pong, D. Sun, and P. Tseng. Hankel matrix rank minimization with applications to system identification and realization. *SIAM Journal on Matrix Analysis and Applications*, 34(3):946–977, 2013.

[6] G. H. Golub and Ch. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, 1996. Third Edition.

[7] A. Hansson, Z. Liu, and L. Vandenberghe. Subspace system identification via weighted nuclear norm optimization. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 3439–3444, 2012.

[8] B. S. He, H. Yang, and S. L. Wang. Alternating direction method with self adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 106(2):337-356, 2000.

[9] M. Jansson and B. Wahlberg. On consistency of subspace methods for system identification. *Automatica*, 34(12):1507-1519, 1998.

[10] W.E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *Proceedings of the 29th IEEE Conference on Decision and Control, 1990*, pages 596–604 vol.2, 1990.

[11] Z. Liu, A. Hansson, and L. Vandenberghe. Nuclear norm system identification with missing inputs and outputs. *Systems & Control Letters*, 62:605–612, 2013.

[12] Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256, 2009.

[13] Z. Liu and L. Vandenberghe. Semidefinite programming methods for system realization and identification. In *Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, pages 4676–4681, 2009.

[14] Z. Liu and L. Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256, 2010.

[15] L. Ljung. *System Identification Toolbox for use with* MATLAB. *Version 7*. The MathWorks, Inc, Natick, MA, 7th edition, 2007.

[16] K. Mohan and M. Fazel. Reweighted nuclear norm minimization with application to system identification. In *Proceedings of American Control Conference*, pages 2953–2959, 2010.

[17] T. Roh and L. Vandenberghe. Discrete transforms, semi-definite programming, sum-of-squares representations of nonnegative polynomials. *SIAM Journal on Optimization*, 16(4):939–964, 2006.

[18] R.S. Smith. Frequency domain subspace identification using nuclear norm minimization and hankel matrix realizations. *IEEE-TAC*, 59(11):2886 – 2896, 2012.

[19] R.S. Smith. Nuclear norm minimization methods for frequency domain subspace identification. In *Proceedings of the American Control Conference*, pages 2689 – 2694, 2012.

[20] P. Van Overschee and B. De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75 – 93, 1994.

[21] M. Verhaegen. Identification of the deterministic part of mimo state space models given in innovations form from input-output data. *Automatica*, 30(1):61 – 74, 1994.

[22] M. Verhaegen and A. Hansson. Nuclear norm subspace identification (N2SID) for short data batches. In *Proceedings of the 19th World Congress IFAC*, pages 9528–9533, Cape Town, South Africa, 2014.

[23] M. Verhaegen and A. Hansson. N2SID: Nuclear norm subspace identification. *CoRR*, abs/1401.4273, 2015.

[24] M. Verhaegen and V. Verdult. *Filtering and Identification: A Least Squares Approach*. Cambridge University Press, 2007.

[25] C. Yu and M. Verhaegen. *Local Subspace Identification of Distributed Homogeneous Systems with General Interconnection Patterns*. IFAC Symposium SYSID, Beijing, pp. 585–589, 2015.

[26] C. Yu, M. Verhaegen and A. Hansson. *Subspace*

Identification of Local 1D Homogeneous Systems. IFAC Symposium SYSID, Beijing, pp. 886–890, 2015.

[27] M. Verhaegen. *System Identification for High Resolution Optical Imaging. Plenary Talk*, IFAC Symposium SYSID, Beijing, 2015.

[28] M. Verhaegen. *Subspace Techniques in System Identification*, In. *Encyclopedia of Systems and Control*, Edt. John Baillieul and Tariq Samad, Springer-Verlag London 2015.