Nuclear Norm Minimization for Blind Subspace Identification (N2BSID)

Dexter Scobee, Lillian Ratliff, Roy Dong, Henrik Ohlsson, Michel Verhaegen and S. Shankar Sastry

***Abstract*— In many practical applications of system identification, it is not feasible to measure both the inputs applied to the system as well as the output. In such situations, it is desirable to estimate both the inputs and the dynamics of the system simultaneously; this is known as the blind identification problem. In this paper, we provide a novel extension of subspace methods to the blind identification of multiple-input multiple-output linear systems. We assume that our inputs lie in a known subspace, and we are able to formulate the identification problem as rank constrained optimization, which admits a convex relaxation. We show the efficacy of this formulation with a numerical example.**

I. Introduction

Consider a discrete-time multi-input multi-output (MIMO) state space model

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k) + Du(k)$$

with the state $x(k) \in \mathbb{R}^n$, input $u(k) \in \mathbb{R}^m$ and output $y(k) \in \mathbb{R}^p$. Estimation of this type of model is one of the most common tasks in system identification and a very well studied problem, see for instance [1], [2]. The common setting is that $\{(y(k), u(k))\}_{k=1}^N$ is given and the matrices A, B, C, and D are found by minimizing the prediction error or by using linear algebra transformations and factorizations with structured Hankel matrices constructed from the input-output data. The former family of methods is referred to as prediction error methods (PEM, see for instance [1]) and the later subspace identification (SID, [3]–[9]) methods.

In this paper we study the more complicated problem of estimating a discrete-time MIMO state-space model from solely outputs $\{y(k)\}_{k=1}^N$. This is an ill-posed problem and it is easy to see that under no further assumptions, it would be impossible to uniquely determine A, B, C and D. To form a well posed problem, additional assumptions on the input or the model are needed. We will work under the assumption that each of the components of the input vector lives in some known subspace. This would be true if, for example, the input:

- changes only at a set of discrete times due to a discrete controller,
- is a linear combination of known possible signals, or
- is band-limited and therefore well represented by the projection on the first discrete Fourier transform basis vectors.

It should be noticed that this assumption is still not enough to uniquely determine the inputs or the A, B, C and D matrices. In particular, we will not be able to decide the inputs or the matrix B more than up to a non-singular ambiguity matrix. It should be stressed that this is not a limitation of the method that we propose but an inherent limitation of the system identification problem since the sought quantities always appear as products. To uniquely determine the inputs and the matrix B, further knowledge is needed.

The work presented is supported by the NSF CPS:Large:ActionWebs award number 0931843, TRUST (Team for Research in Ubiquitous Secure Technology) which receives support from NSF (award number CCF-0424422), the ONR under the Embedded Humans MURI (N00014-13-1-0341), and FORCES (Foundations Of Resilient CybErphysical Systems) which receives support from NSF (CNS-1239166). The work of Verhaegen is sponsored by the European Research Council, Advanced Grant Agreement No. 339681

Scobee, Ratliff, Dong, Ohlsson and Sastry are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720. Ohlsson is also with C3 Energy and the Division of Automatic Control, Department of Electrical Engineering, Linköping University, Sweden. Verhaegen is with the Delft Center for Systems and Control, Delft University, Delft, The Netherlands. {dscobee, ratliffl, roydong, ohlsson, sastry}@eecs.berkeley.edu, m.verhaegen@tudelft.nl

II. Background

System identification refers to problems in which some unknowns about a system are recovered, including system parameters and system inputs, from system observations. It is a research area whose tools are applicable in a wide variatey of domains. In the remainder of this section, we discuss existing work in the areas of subspace identification and blind identification.

A. Subspace Identification

The two main thrusts within the system identification literature for identification of state space models are prediction error methods and subspace identification methods. Prediction error methods focus on finding unknowns by optimizing a cost function of given observations. The draw back to such methods is that if the cost function is non-convex, the optimization problem gets stuck in local minima. In an effort to avoid such issues, subspace identification methods came about.

Subspace methods find approximate models as opposed to models that are optimal with respect to some cost [10]. They further allow for exploitation of underlying structure that may exist in a particular system. In particular, subspace identification methods aim to derive a low rank matrix from which key subspaces are identified [10]. A recent advance in subspace identification is replacing the traditional singular–value decomposition (SVD) technique for obtaining the low rank approximation with the nuclear norm [11]–[15].

A new subspace identification algorithm named Nuclear Norm Subspace IDentification (N2SID)—and on which we build our blind subspace identification method—exploits the key structural properties in the data equation used by system identification methods [10]. In particular, the core structural properties are the low rank and the the block-Toeplitz structure of the unknown model dependent matrices in the data equation [10]. The key attribute of this approach is that both of these structural properties are invoked in the first step of the algorithm thereby addressing an issue that arises in subspace identification methods in general. Namely, in typical subspace methods the low rank approximation is not invoked at the first step which, in turn, means that the low rank decomposition does not operate on the raw input-output data.

B. Blind Identification

Extending beyond identification of a system model, we can also consider simultaneously identifying the system model and some unknown inputs. This type of problem is typically referred to as blind system identification (BSI). BSI is a tool with a broad application area; it has been applied in fields such as data communications, speech recognition and seismic signal processing, see for instance [16]. For the type of modeling problems that BSI has been applied to it is common that the input is difficult, costly or impossible to measure. In, for example, exploration seismology, the physical properties of the earth are explored by studying the response of an excitation (often a charge of dynamite). The excitation is often difficult to measure and the modeling problem is therefore a BSI problem, see e.g. [17].

Many methods have been proposed to solve the BSI problem throughout the years. We give a short overview here but refer the interested reader to [16], [18], for a more extensive and complete review.

The maximum likelihood (ML) approach to BSI aims at finding the ML estimate of the model and input. The resulting non-convex optimization problem is often treated by alternating between optimizing with respect to the input and the system model [19]. The channel subspace (CS) methods to BSI indirectly determine the sought finite impulse response (FIR) model by estimating the nullspace of the Sylvester matrix associated with the FIR model to be identified. This is done by an eigen–decomposition of a matrix derived from the outputs [20]. The methods proposed in [21] and [22] work under the assumption that two or more output series are available and that these were generated by the same input. The methods proposed in [23], [24] assume that the input consists of independent and identically distributed random variables and considers the autocorrelation of the output to decide a FIR model and the unknown input.

A number of approaches consider the blind identification problem of Hammerstein systems under the assumption that the input is piecewise constant [25]–[30]. In our recent work [31], we assume that the input belongs to some known subspace. A piecewise constant signal can be represented using the subspace assumption given here. However, we note that we are not restricted to piecewise constant signals, and our approach is significantly different. We also consider the blind identification of autoregressive models with exogenous inputs (ARX models) while the blind identification problem of Hammerstein systems is considered in [25]–[30]. The related problem of blind deconvolution has been studied in a number of contributions. In particular, see the very interesting paper [32] for a solution where the signals to be recovered are assumed to be in some known subspaces. However, only FIR models are considered.

In this paper, we address the problem of BSI by building upon our previous work [31] and the N2SID [10] approach that allows us to exploit key structural properties in the data equation. We call our algorithm Nuclear Norm minimization for Blind Subspace IDentification (N2BSID).

III. Problem Statement and Mathematical Formulation (N2BSID)

Given the sequence of outputs $\{y(k)\}_{k=1}^{N} \in \mathbb{R}^p$, we desire to find an estimate for $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$ and $u(k) \in \mathbb{R}^m, k = 1, \dots, N$, such that

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned} \tag{1}$$

To make the problem well posed, we seek a length N time history of m-dimensional inputs that lies in a given known subspace of $\mathbb{R}^{m\times N}$.

Consider the state space model given by (1) and assume that, for any given time k, the input can be decomposed as

$$u(k) = d(k)z \tag{2}$$

where $z \in \mathbb{R}^l$, $u(k) \in \mathbb{R}^m$, and $d(k) \in \mathbb{R}^{m\times l}$ is a known matrix (refered to as a dictionary). The assumed constraint expressed in (2) restricts the input to a subspace as in [31]. For a problem instance with a sequence of N samples ($k = 1, \cdots, N$), it is taken that $l \leq mN$ (the number of unknowns is reduced by the assumption of (2)) and $l \leq pN$ (more measurements are available than unknowns).

Let $s > n$ be an integer. Define

$$\mathcal{O}_s^T = \left[C^T A^T C^T \cdots A^{T^{s-1}} C^T \right], \tag{3}$$

$$X = \begin{bmatrix} x(1) & x(2) & \cdots & x(N-s+1) \end{bmatrix}, \tag{4}$$

and the $s \times (N-s+1)$ block-Hankel matrices

$$U_s = \begin{bmatrix} u(1) & u(2) & \cdots & u(N-s+1) \\ u(2) & u(3) & & \vdots \\ \vdots & & \ddots & \\ u(s) & u(s+1) & \cdots & u(N) \end{bmatrix}, \tag{5}$$

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

where entries $u(i) \in \mathbb{R}^{m\times 1}$ are column vectors corresponding to system inputs and entries $y(i) \in \mathbb{R}^{p\times 1}$ are column vectors corresponding to system outputs. Further define the $s \times s$ lower-triangular block-Toeplitz matrix

$$T_s = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{s-2}B & \cdots & CB & D \end{bmatrix} \tag{7}$$

Using the above defined notation, we rewrite (1) as

$$Y_s = \mathcal{O}_s X + T_s U_s. \tag{8}$$

which is known as the *data equation*. For simplicity, we introduce the notation for T_s given by

$$T_s = \begin{bmatrix} H_1 & 0 & \cdots & 0 \\ H_2 & H_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ H_s & \cdots & H_2 & H_1 \end{bmatrix}$$

where $H_1 = D$, $H_2 = CB$, etc. With this notation, we write

$$T_s U_s = \begin{bmatrix} H_1 u(1) & \cdots & H_1 u(N-s+1) \\ H_2 u(1) + H_1 u(2) & & \vdots \\ \vdots & \ddots & \\ H_s u(1) + \cdots + H_1 u(s) & \cdots & H_s u(N-s+1) + \cdots + H_1 u(N) \end{bmatrix} \tag{9}$$

By the data equation (given in (8)), all block entries of $T_s U_s$ will have the same dimensions as the block entries of Y_s, which are $p \times 1$ column vectors. For any column vector v, $v = \text{vec}(v) = \text{vec}(v^T)$, which can be seen readily from the definition of vectorization. Therefore, every entry $H_i u(j) = H_i d(j) z$ in $T_s U_s$ will be equal to $\text{vec}\left((H_i d(j) z)^T\right) = \text{vec}\left(z^T d(j)^T H_i^T\right)$. Using the identity

$$\text{vec}(XYZ) = (Z^T \otimes X)\,\text{vec}(Y) \tag{10}$$

where $\otimes$ represents the Kronecker product, we can now write

$$\text{vec}\left(z^T d(j)^T H_i^T\right) = \left(H_i \otimes z^T\right) \text{vec}\left(d(j)^T\right) \tag{11}$$

which implies that we can express $T_s U_s$ as

$$T_s U_s = \begin{bmatrix} H_1 \otimes z^T & 0 & \cdots & 0 \\ H_2 \otimes z^T & H_1 \otimes z^T & & 0 \\ \vdots & & \ddots & \vdots \\ H_s \otimes z^T & H_{s-1} \otimes z^T & \cdots & H_1 \otimes z^T \end{bmatrix} \times \begin{bmatrix} \text{vec}\left(d(1)^T\right) & \text{vec}\left(d(2)^T\right) & \cdots & \text{vec}\left(d(N-s+1)^T\right) \\ \text{vec}\left(d(2)^T\right) & \text{vec}\left(d(3)^T\right) & & \text{vec}\left(d(N-s+2)^T\right) \\ \vdots & & \ddots & \vdots \\ \text{vec}\left(d(s)^T\right) & \text{vec}\left(d(s+1)^T\right) & & \text{vec}\left(d(N)^T\right) \end{bmatrix} \tag{12}$$

We will rename the right-hand side of the equation such that

$$T_s U_s = H(z) D_s. \tag{13}$$

Like T_s, $H(z)$ is an $s \times s$ lower-triangular block Toeplitz matrix, and we denote the block entries of $H(z)$ by

$$H(z) = \begin{bmatrix} \mathcal{H}_1 & 0 & \cdots & 0 \\ \mathcal{H}_2 & \mathcal{H}_1 & & 0 \\ \vdots & & \ddots & \vdots \\ \mathcal{H}_s & \mathcal{H}_{s-1} & & \mathcal{H}_1 \end{bmatrix}$$

We seek to find a minimal representation of the system which can faithfully explain the observed output. Therefore, as in [10], we seek to minimize the rank of $\mathcal{O}_s X_s$, which, by the data equation (given in (8)), is equal to $Y_s - T_s U_s$. We express this goal via the following optimization problem

$$\min_{H(z) \in \mathcal{T}_s} \text{rank}\left(Y_s - H(z) D_s\right) \tag{14}$$

where $\mathcal{T}_s$ denotes the set of $s \times s$ lower-triangular block Toeplitz matrices. Rank minimization is known to be non-convex, so we use an accepted convex relaxation and replace the rank with the nuclear norm, denoted $\|\cdot\|_*$. The relaxed problem is given by

$$\min_{H(z) \in \mathcal{T}_s} \|Y_s - H(z) D_s\|_* \tag{15}$$

The optimization problem formulated in (15) is convex.

There is, however, a constraint on our decision variable $H(z)$. We must retain the possibility that $H(z)$ has the form shown in (12), specifically the fact that each block entry is given by the Kronecker product of a $p \times m$ matrix (H_i) and a $1 \times l$ row vector (z^T). In the case where $m = 1$ (single input), we note that $H_i \otimes z^T = H_i z^T$. Therefore, we can express this requirement as

$$\left(\begin{bmatrix} H_1 \otimes z^T \\ H_2 \otimes z^T \\ \vdots \\ H_s \otimes z^T \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_s \end{bmatrix} z^T \right) \text{is rank 1.} \tag{16}$$

When the rank is 1, we can recover z^T via an SVD and then recover the input via (2). Directly enforcing this rank constraint would again lead to a non-convex optimization problem, so we again apply the convex relaxation of using the nuclear norm to enforce that this matrix be low rank. The full optimization problem then becomes

$$\min_{H(z) \in \mathcal{T}_s} \|Y_s - H(z) D_s\|_* + \lambda \left\| \begin{bmatrix} \mathcal{H}_1 \\ \mathcal{H}_2 \\ \vdots \\ \mathcal{H}_s \end{bmatrix} \right\|_* \tag{17}$$

The optimization problem expressed in (17) seeks a balance between minimizing the two nuclear norms. Therefore, similarly to [10], we can express the full range of potential optimal solutions by varying the weighting coefficient λ.

In the general case of m inputs, we still need to enforce that we can recover z^T. From the structure of $H(z)$, it is possible to derive a matrix $H^*(z)$ such that

$$H^*(z) = \begin{bmatrix} \text{vec}\left(H_1^T\right) \\ \text{vec}\left(H_2^T\right) \\ \vdots \\ \text{vec}\left(H_s^T\right) \end{bmatrix} z^T \tag{18}$$

If $H^*(z)$ has rank 1, then we can recover z^T from an SVD.

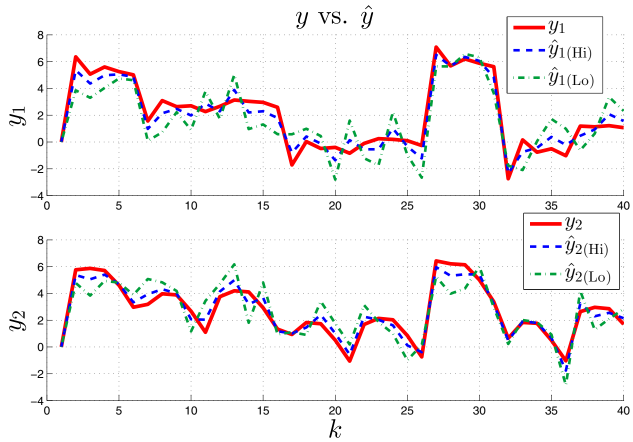

Fig. 1. Time history of true output y vs. noised output $\hat{y}$ for High and Low SNR

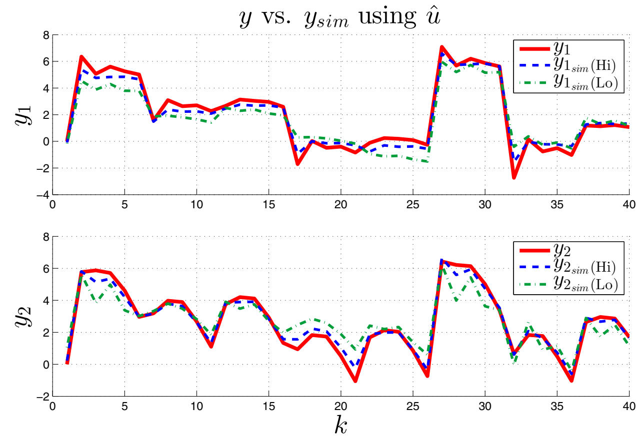

Fig. 2. Time history of true output y vs. simulated output y_{sim} generated using N2BSID recovered input and system matrices for High and Low SNR

Therefore, the full MIMO problem formulation for N2BSID is given by the following convex optimization problem:

$$\min_{H(z)\in\mathcal{T}_s} \|Y_s - H(z)D_s\|_* + \lambda\|H^*(z)\|_* \tag{19}$$

IV. Numerical Example

Consider a linear time-invariant (LTI) system represented by (1) whose system matrices are given by

$$A = \begin{bmatrix} -0.4 & 0.2 \\ 0.1 & -0.3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 2 & 0.5 \\ 2 & 1 & 6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{20}$$

We generate output data from this system using the initial state $x_0 = [0\ 0]^T$. The pre-image of the input, z, is sampled from the standard normal distribution, and the dictionary matrix $d(1:N)$ (the vertical concatenation of all $d(i)$ as in (2)) is chosen such that the first two inputs u_1, u_2 hold constant values over 5-step intervals and the third input u_3 repeats one half period of a sine wave every 5-step inteval (these can be seen along with their estimates from N2BSID in Figure 3). At every time step k, noise is drawn from a uniform distribution between $-\epsilon/2$ and $\epsilon/2$ and added to the output vector $y(k)$ to generate the noisy output $\hat{y}(k)$ which will be used as input to the N2BSID algorithm (Figure 1 shows examples of both y and $\hat{y}$). The algorithm was implemented and tested in MATLAB, making use of the CVX package for specifying and solving convex programs [33]–[35].

In order to characterize the degree to which the output signal is affected by this noise, we analyze the signal-to-noise ratio (SNR) of the data. For this application, the SNR, given in decibels, is defined as

$$\text{SNR} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{21}$$

where P_{signal} is equal to the sum of the squares of each sample of the true output signal, and P_{noise} is the sum of the squares of each sample of the noise that has been added to that output signal.

In order to describe how well a noised signal $\hat{\phi}$ (or a simulated signal based on identified system matrices ϕ_{sim}) tracks the original ϕ, we analyze the root mean square (RMS) error, defined as

$$\varepsilon_{\text{RMS}}(\hat{\phi},\phi) = \sqrt{\frac{\sum_{k=1}^{N}(\hat{\phi}(k)-\phi(k))^2}{N}} \tag{22}$$

The magnitude of this error will depend on the magnitude of the true signal, so we normalize ε_{RMS} by the RMS magnitude of the true signal, given by

$$\mu_{\text{RMS}}(\phi) = \sqrt{\frac{\sum_{k=1}^{N}\phi(k)^2}{N}} \tag{23}$$

Therefore, the normalized RMS error is given by

$$\varepsilon^*(\hat{\phi},\phi) = \frac{\varepsilon_{\text{RMS}}(\hat{\phi},\phi)}{\mu_{\text{RMS}}(\phi)} \tag{24}$$

The data shown in Figures 1, 2, and 3 correspond to performing N2BSID on the system described by (20) under two conditions:

High SNR (Hi):
$\epsilon = 2$, which leads to $\text{SNR}_{y_1} = 13.9$dB and $\text{SNR}_{y_2} = 16.1$dB.

Low SNR (Lo):
$\epsilon = 5$, which leads to $\text{SNR}_{y_1} = 6.0$dB and $\text{SNR}_{y_2} = 8.1$dB.

Under both of these conditions, the system was simulated for 40 time steps ($N = 40$). We selected $\lambda = 0.1$ as the weighting parameter for N2BSID in both cases. Algorithm performance is approximately constant for any choice $\lambda \leq 1$, corresponding to cases where less weight is assigned to the minimization of $\|H^*(z)\|_*$ than to the minimization of $\|Y_s - H(z)D_s\|_*$ (see Figure 5).

For the particular set of noisy outputs shown in Figure 1, the N2BSID method, which made use of the same dictionary matrix as was used in data generation, recovered the

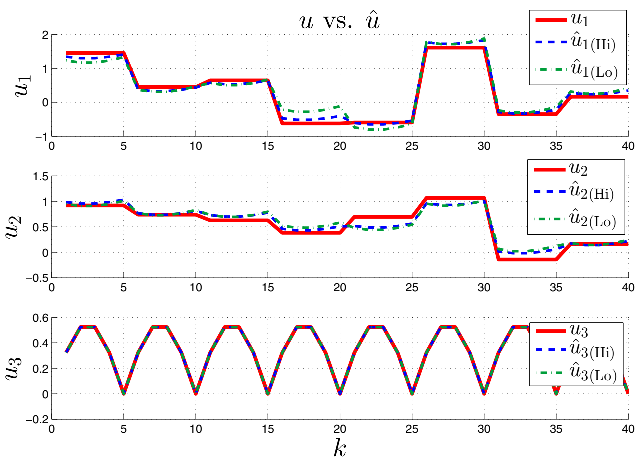

Fig. 3. Time history of true input u vs. N2BSID recovered input $\hat{u}$ for High and Low SNR. For a sense of performance, $\varepsilon^*(\hat{u}_{1(\mathrm{Hi})}, u_1) = 0.125$ and $\varepsilon^*(\hat{u}_{1(\mathrm{Lo})}, u_1) = 0.233$.

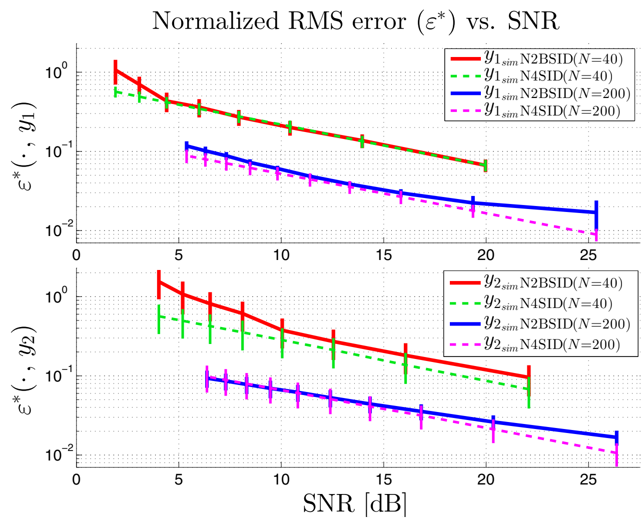

Fig. 4. Variation of normalized RMS error (ε^*) with respect to sample size and SNR, comparing N2BSID with N4SID. Each data point corresponds to 10,000 random trials. Error bars show $\pm$ one standard deviation. Note that the vertical axes are logarithmic.

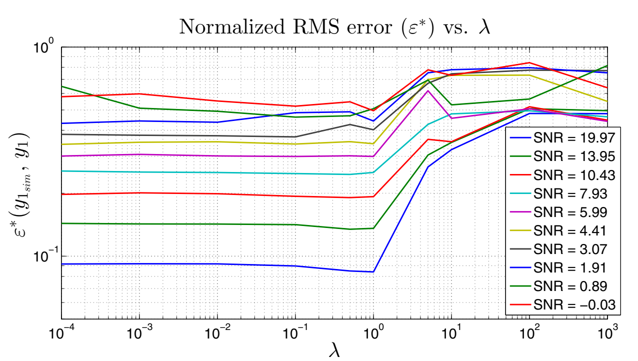

Fig. 5. Variation of normalized RMS error (ε^*) of $y_{1_{sim}}$ (produced by simulation of identified system using $\hat{u}$) with respect to λ for a variety of SNR values. Independent of SNR, ε^* is relatively constant for $\lambda \leq 1$, but ε^* increases for $\lambda > 1$. Note that the axes are logarithmic.

following system matrices for the Hi and Lo cases:

$$A_{\mathrm{Hi}} = \begin{bmatrix} -0.28 & 0.16 \\ 0.11 & -0.42 \end{bmatrix}, \quad B_{\mathrm{Hi}} = \begin{bmatrix} 2.83 & 1.29 & 0.77 \\ 1.68 & 1.78 & 5.91 \end{bmatrix}$$

$$A_{\mathrm{Lo}} = \begin{bmatrix} 0.06 & -0.25 \\ 0.10 & -0.68 \end{bmatrix}, \quad B_{\mathrm{Lo}} = \begin{bmatrix} 2.55 & 1.02 & 2.17 \\ 0.73 & 3.93 & 5.33 \end{bmatrix}$$

Note that, in recovering the system matrices, it was assumed that there was no direct feedthrough from input to output, leading D_{Hi} and D_{Lo} to be taken as zero matrices of appropriate dimension.

To enforce $C_{\mathrm{Hi}} = C_{\mathrm{Lo}} = I_{2\times 2}$ (as in (20)) and maintain consistency among the recovered matrices, we use the matrices returned by the algorithm (A_{out}, B_{out}, and C_{out}), and perform a similarity transformation on the system they describe such that $A_{\mathrm{Hi}} = C_{\mathrm{out}} A_{\mathrm{out}} C_{\mathrm{out}}^{-1}$, $B_{*\mathrm{Hi}} = C_{\mathrm{out}} B_{\mathrm{out}}$, and $C_{\mathrm{Hi}} = C_{\mathrm{out}} C_{\mathrm{out}}^{-1} = I_{2\times 2}$, and similarly for the Lo case. Recall that it is not possible to determine the B matrix and the input signal u beyond a non-singular ambiguity matrix. In order to make comparisons between the true u and the u_{out} produced by N2BSID, we find the least-squares solution P to the equation $u = P u_{\mathrm{out}}$, letting $\hat{u} = P u_{\mathrm{out}}$ and $B_{\mathrm{Hi}} = B_{*\mathrm{Hi}} P^{-1}$ (and similarly for B_{Lo}). The estimated input data $\hat{u}$ is shown in comparison to the true input u in Figure 3. Note that these transformations do not affect the output that will be estimated by the identified system.

The estimated output data y_{sim} in Figure 2 was calculated by simulating the system using the estimated system matrices and the estimated input $\hat{u}$, all products of N2BSID. The only inputs to the N2BSID algorithm were the noisy output signal (see Figure 1) and the dictionary matrix $d(1:N)$, and the algorithm was able to closely reproduce the input signal (see Figure 3), which lead to smoothed estimates of the output compared to the noisy estimates given to the algorithm (see Figure 2). In fact, for the Hi case, the error was reduced from $\varepsilon^*(\hat{y}_1, y_1) = 0.201$ and $\varepsilon^*(\hat{y}_2, y_2) = 0.156$ to $\varepsilon^*(y_{1_{sim}}, y_1) = 0.145$ and $\varepsilon^*(y_{2_{sim}}, y_2) = 0.096$. For the Lo case, the error was reduced from $\varepsilon^*(\hat{y}_1, y_1) = 0.501$ and $\varepsilon^*(\hat{y}_2, y_2) = 0.390$ to $\varepsilon^*(y_{1_{sim}}, y_1) = 0.300$ and $\varepsilon^*(y_{2_{sim}}, y_2) = 0.275$.

In order to illustrate the behavior of N2BSID with respect to noise and sample size, we performed a parametric study, testing the algorithm at a range of SNRs for both $N = 40$ and $N = 200$, using the system descibed by (20) above (the results are summarized in Figure 4). At each SNR level shown, system identification was performed to recover estimates of the true system matrices. Ten thousand random input trajectories were then generated by using the same dictionary defined above, with randomly generated values for z. These random inputs, which are distinct from the input which generated the output signal utilized during the system identification process, are used to drive both the true and the estimated system, allowing us to calculate the value of ε^* by comparing the output signals. Every point in Figure 4 represents the mean value of ε^* over ten thousand trials, and the error bars show plus and minus one standard deviation.

This same analysis was performed using the MATLAB implementation of N4SID (numerical algorithms for subspace state space system identification) for second order systems [9], [33]. It is important to note, however, that N4SID is not a blind method, so it was necessary to provide N4SID with the input signal u in order to identifiy system matrices that could be used for simulation and error analysis. The results for N4SID are shown alongside those for N2BSID in Figure 4.

There are three salient features to the resulting plot. First, we see that, under the same testing conditions, the performance of N2BSID is comparable to that of N4SID. This correspondance indicates that, although N2BSID is a blind method and has no access to the input signal, the knowledge of the subspace in which the input lies (defined by the dictionary) allows N2BSID to acheive results comparable to those of non-blind methods. Second, it can be clearly seen that as sample size increases from $N = 40$ to $N = 200$, the errors in system estimation uniformly decrease. And third, it can be seen that, regardless of sample size, the performance of the algorithm improves with increasing SNR, and the algorithm performance degrades smoothly as SNR decreases.

V. Conclusion

The paper presents a novel extension of subspace methods to the blind identification of MIMO LTI systems. The extension takes the form of a convex optimization problem and we show that it can effectively handle problems where the input is unknown but restricted to some known subspace.

References

[1] L. Ljung, *System Identification — Theory for the User*, 2nd ed. Upper Saddle River, N.J.: Prentice-Hall, 1999.

[2] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach*, 1st ed. New York, NY, USA: Cambridge University Press, 2007.

[3] B. De Moor, M. Moonen, L. Vandenberghe, and J. Vandewalle, "A geometrical approach for the identification of state space models with singular value decomposition," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, 1988, pp. 2244–2247 vol.4.

[4] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle, "On- and off-line identification of linear state space models," *International Journal of Control*, vol. 49, pp. 219–232, 1989.

[5] M. Verhaegen and P. Dewilde, "Subspace model identification part 1. The output-error state-space model identification class of algorithms," *International journal of control*, vol. 56, no. 5, pp. 1187–1210, 1992.

[6] ——, "Subspace model identification part 2. Analysis of the elementary output-error state-space model identification algorithm," *International journal of control*, vol. 56, no. 5, pp. 1211–1241, 1992.

[7] M. Verhaegen, "Subspace model identification part 3. Analysis of the ordinary output-error state-space model identification algorithm," *International Journal of control*, vol. 58, no. 3, pp. 555–586, 1993.

[8] ——, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61 – 74, 1994.

[9] P. Van Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75 – 93, 1994.

[10] M. Verhaegen and A. Hansson, "Nuclear norm subspace identification (N2SID) for short data batches," *arXiv:1401.4273*, 2014.

[11] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2009.

[12] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, PhD thesis, Stanford University, 2002.

[13] K. Mohan and M. Fazel, "Reweighted nuclear norm minimization with application to system identification," in *American Control Conference*. IEEE, 2010, pp. 2953–2959.

[14] Z. Liu, A. Hansson, and L. Vandenberghe, "Nuclear norm system identification with missing inputs and outputs," *Systems & Control Letters*, vol. 62, no. 8, pp. 605–612, 2013.

[15] A. Hansson, Z. Liu, and L. Vandenberghe, "Subspace system identification via weighted nuclear norm optimization," *arXiv:1207.0023*, 2012.

[16] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.

[17] A. Zerva, A. Petropulu, and P.-Y. Bard, "Blind deconvolution methodology for site-response evaluation exclusively from ground-surface seismic recordings," *Soil Dynamics and Earthquake Engineering*, vol. 18, no. 1, pp. 47–57, 1999.

[18] Y. Hua, "Blind methods of system identification," *Circuits, Systems and Signal Processing*, vol. 21, no. 1, pp. 91–108, 2002.

[19] K. Abed-Meraim and E. Moulines, "A maximum likelihood solution to blind identification of multichannel FIR filters," in *EUSIPCO*, 1994, pp. 1011–1015.

[20] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Transansactions on Information Theory*, vol. 43, no. 2, pp. 499–511, Sep. 2006.

[21] A. Zerva, A. Petropulu, and P.-Y. Bard, "Blind system identification," in *Proceedings of the 8th ASCE Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability, PMC2000*, University of Notre Dame, Indiana, USA, 2000.

[22] S. Van Vaerenbergh, J. Vía, and I. Santamaría, "Blind identification of SIMO Wiener systems based on kernel canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2219–2230, 2013.

[23] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Transactions on Communications*, vol. 23, no. 6, pp. 679–682, 1975.

[24] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *1991 Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, vol. 2, 1991, pp. 856–860.

[25] L. Sun, W. Liu, and A. Sano, "Identification of a dynamical system with input nonlinearity," *IEE Proceedings – Control Theory and Applications*, vol. 146, no. 1, pp. 41–51, 1999.

[26] E.-W. Bai, Q. Li, and S. Dasgupta, "Blind identifiability of IIR systems," *Automatica*, vol. 38, no. 1, pp. 181–184, 2010.

[27] E.-W. Bai and M. Fu, "A blind approach to Hammerstein model identification," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1610–1619, 2002.

[28] J. Wang, A. Sano, D. Shook, T. Chen, and B. Huang, "A blind approach to closed-loop identification of Hammerstein systems," *International Journal of Control*, vol. 80, no. 2, pp. 302–313, 2007.

[29] J. Wang, A. Sano, T. Chen, and B. Huang, "Identification of Hammerstein systems without explicit parameterisation of non-linearity," *International Journal of Control*, vol. 82, no. 5, pp. 937–952, 2009.

[30] ——, "A blind approach to identification of Hammerstein systems," in *Block-oriented Nonlinear System Identification*, ser. Lecture Notes in Control and Information Sciences, F. Giri and E.-W. Bai, Eds. Springer London, 2010, vol. 404, pp. 293–312.

[31] H. Ohlsson, L. J. Ratliff, R. Dong, and S. S. Sastry, "Blind identification via lifting," *arXiv:1312.2060*, 2013.

[32] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *CoRR*, vol. abs/1211.5608, 2012.

[33] MATLAB, *version 8.3.0.532 (R2014a)*. Natick, Massachusetts: The MathWorks Inc., 2014.

[34] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[35] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/ boyd/graph_dcp.html.