Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Blind Identification via Lifting ⋆

Henrik Ohlsson *,** **Lillian Ratliff** * **Roy Dong** *
S. Shankar Sastry *

* *Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA (e-mail: ohlsson@eecs.berkeley.edu).*
** *Division of Automatic Control, Department of Electrical Engineering, Linköping University, Sweden.*

Abstract: Blind system identification is known to be an ill-posed problem and without further assumptions, no unique solution is at hand. In this contribution, we are concerned with the task of identifying an ARX model from only output measurements. We phrase this as a constrained rank minimization problem and present a relaxed convex formulation to approximate its solution. To make the problem well posed we assume that the sought input lies in some known linear subspace.

Keywords: System identification; Parameter estimation; Identification algorithms.

1. INTRODUCTION

Consider an auto-regressive exogenous input (ARX) model

$$\begin{aligned} y(t)-a_1y(t-1)-\cdots-a_{n_a}y(t-n_a)& \\ =b_1u(t-n_k-1)+\cdots+b_{n_b}u(t-n_k-n_b)& \end{aligned} \quad (1)$$

with input $u \in \mathbb{R}$ and output $y \in \mathbb{R}$. 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 Ljung [1999]. The common setting is that $\{(y(t),u(t))\}_{t=1}^N$ is given and the summed residuals

$$\sum_{t=n}^{N}\left(y(t)-\sum_{k_1=1}^{n_b}b_{k_1}u(t-k_1-n_k)-\sum_{k_2=1}^{n_a}a_{k_2}y(t-k_2)\right)^2$$

where $n=\max(n_a,n_k+n_b)+1$, is minimized to obtain an estimate for $a_1,\ldots,a_{n_a},b_1,\ldots,b_{n_b}$. This estimate is often referred to as the *least squares* (LS) estimate.

In this paper we study the more complicated problem of estimating an ARX model from solely outputs $\{y(t)\}_{t=1}^N$. This is an ill-posed problem and it is easy to see that under no further assumptions, it would be impossible to uniquely determine $a_1,\ldots,a_{n_a},b_1,\ldots,b_{n_b}$. We will here therefore study this problem under the assumption that the stacked inputs belong to some known subspace. The input could for example be:

- known to change only at a set of discrete times due to a discrete controller and therefore constrained to a subspace of the input space or
- known to be band-limited and therefore well represented by the projection on the first discrete Fourier transform basis vectors. The subspace would then be spanned by the first discrete Fourier transform basis vectors.

It should be noticed that this assumption is not enough to uniquely determine the input or the ARX model. Specifically, we will not be able to decide the input or the ARX coefficients $b_1,\ldots,b_{n_b}$ more than up to a multiplicative scalar. 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 input and the ARX coefficients $b_1,\ldots,b_{n_b}$, further knowledge is needed.

The main contribution of the paper is a novel method for ARX model identification from only output measurements. The method takes the form of a convex optimization problem and gives a computationally flexible framework for handling different types of measurement noises, constraints, *etc.*

2. BACKGROUND

Blind system identification (BSI) has a broad application area and has been applied in fields such as data communications, speech recognition and seismic signal processing, see for instance Abed-Meraim et al. [1997]. Common for the type of modeling problems that BSI has been applied to is 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 therefore a BSI problem, see *e.g.,* Zerva et al. [1999].

⋆ The work presented is supported by the NSF Graduate Research Fellowship under grant DGE 1106400, NSF CPS:Large:ActionWebs award number 0931843, TRUST (Team for Research in Ubiquitous Secure Technology) which receives support from NSF (award number CCF-0424422), and FORCES (Foundations Of Resilient CybEr-physical Systems), the Swedish Research Council in the Linnaeus center CADICS, the European Research Council under the advanced grant LEARN, contract 267381, a postdoctoral grant from the Sweden-America Foundation, donated by ASEA's Fellowship Fund, and by a postdoctoral grant from the Swedish Research Council.

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 Abed-Meraim et al. [1997], Hua [2002], 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. See Abed-Meraim and Moulines [1994]. 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. See for instance Abed-Meraim et al. [2006]. The method proposed in Zerva et al. [2000] works under the assumption that two or more output series are available and that these were generated by the same input. See also Van Vaerenbergh et al. [2013]. The methods proposed in Sato [1975], Tong et al. [1991] assume that the input consists of independent and identically (iid) 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. Sun et al. [1999], Bai et al. [2010], Bai and Fu [2002], Wang et al. [2007, 2009, 2010]. Our approach assumes that the input belongs to some known subspace. Piecewise constant signal can be represented using the subspace assumption used here. However, we note that we are not restricted to piecewise constant signals, and our approach is significantly different. Also, we consider the blind identification of ARX models while the blind identification problem of Hammerstein systems is considered in Sun et al. [1999], Bai et al. [2010], Bai and Fu [2002], Wang et al. [2007, 2009, 2010].

The related problem of blind deconvolution have been studied in a number of contributions. In particular, see the very interesting paper by Ahmed et al. [2012] for a solution where the signals to be recovered are assumed to be in some known subspaces. The development presented in Ahmed et al. [2012] has similarities to the approach presented in this paper and was done in parellel to our work. Note that only FIR models are discussed in Ahmed et al. [2012] and that the analysis does not apply.

3. PROBLEM FORMULATION

Given the sequence of outputs $\{y(t)\}_{t=1}^{N} \in \mathbb{R}$, find an estimate for $a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b} \in \mathbb{R}$ and $u(t) \in \mathbb{R}, t = 1, \ldots, N$, such that

$$\begin{aligned} y(t) - a_1 y(t-1) - \cdots - a_{n_a} y(t-n_a) \\ = b_1 u(t-n_k-1) + \cdots + b_{n_b} u(t-n_k-n_b) + w(t), \end{aligned}$$

for $t = n, \ldots, N$, where $n = \max(n_a, n_k + n_b) + 1$, and $w(t)$, $t = n, \ldots, N$, some unknown zero mean noise. We will for simplicity assume that n_a, n_b, n_k, are known. To make the problem well posed, we will seek an input in a given subspace of $\mathbb{R}^N$.

4. NOTATION AND ASSUMPTIONS

We will use y to denote the output and u the input. We will for simplicity only consider *single input single output* (SISO) systems, however with some extra bookkeeping also MIMO systems could be treated. We will assume that N measurements of y are available and stack them in the vector $\mathbf{y}$, *i.e.,*

$$\mathbf{y} = [y(1) \ \ldots \ y(N)]^{\mathsf{T}}. \tag{2}$$

We also introduce $\mathbf{u}$, η, $\mathbf{a}$ and $\mathbf{b}$ as

$$\mathbf{u} = [u(1) \ \ldots \ u(N)]^{\mathsf{T}}, \tag{3}$$

$$\eta = [\eta(1) \ \ldots \ \eta(N)]^{\mathsf{T}}, \tag{4}$$

$$\mathbf{a} = [a_1 \ \ldots \ a_{n_a}]^{\mathsf{T}}, \tag{5}$$

$$\mathbf{b} = [b_1 \ \ldots \ b_{n_b}]^{\mathsf{T}}. \tag{6}$$

We will use $\mathbf{y}(i)$ to denote the ith element of $\mathbf{y}$. To pick out a subvector of $\mathbf{y}$ consisting of the ith to the jth element we will use the notation $\mathbf{y}(i:j)$ and similarly for picking out a subvector of $\mathbf{u}$, $\mathbf{a}$ and $\mathbf{b}$. To pick out a submatrix consisting of the ith to the jth rows of $\mathbf{X}$ we use the notation $\mathbf{X}(i:j,:)$. We will use normal font to represent scalars and bold for vectors and matrices. $\|\cdot\|_0$ is the zero (quasi) norm which returns the number of nonzero elements of its argument and $\|\cdot\|_*$ the nuclear norm returning the sum of the singular values.

We will assume that it is known that the sought input, $\mathbf{u}$, lies in some known subspace. We can hence write

$$\mathbf{u} = \mathbf{D}\mathbf{x} \tag{7}$$

for some known $N \times m$-matrix $\mathbf{D}$ and an unknown vector $\mathbf{x} \in \mathbb{R}^m$. It is assumed that $m \leq N$.

5. BLIND IDENTIFICATION VIA LIFTING

Consider the noise free setting where $w(t) = 0, t = n, \ldots, N$. We can formulate the problem of finding an input and the ARX coefficients as the feasibility problem

$$\text{find} \quad \begin{matrix} u(t),\ t = 1, \ldots, N, \\ a_1, \ldots, a_{n_a}, b_1, \ldots, b_{n_b} \end{matrix} \tag{8a}$$

$$\begin{aligned} \text{subj. to} \quad & y(t) - \sum_{k_2=1}^{n_a} a_{k_2} y(t-k_2) \\ & = \sum_{k_1=1}^{n_b} b_{k_1} u(t-n_k-k_1), \quad t = n, \ldots, N. \end{aligned} \tag{8b}$$

Note that the $\{a_k\}_{k=1}^{n_a}$, $\{b_k\}_{k=1}^{n_b}$, and $\{u(t)\}_{t=1}^{N}$ are unknown. The problem is therefore non-convex.

Introduce $\mathbf{X} = \mathbf{x}\mathbf{b}^{\mathsf{T}} \in \mathbb{R}^{m \times n_b}$ and note that (7) gives that $\mathbf{D}\mathbf{X} = \mathbf{D}\mathbf{x}\mathbf{b}^{\mathsf{T}} = \mathbf{u}\mathbf{b}^{\mathsf{T}}$. Since $\mathbf{u}\mathbf{b}^{\mathsf{T}}$ contains all products $u(i)b_j$, $i = 1, \ldots, N$, $j = 1, \ldots, n_b$, the sum

$$\sum_{k_1=1}^{n_b} b_{k_1} u(t-n_k-k_1) \tag{9}$$

can be realized by summing appropriate entries of $\mathbf{D}\mathbf{X}$. Problem (8) can now be reformulated as

$$\text{find}_{\mathbf{X},\mathbf{a}} \tag{10a}$$

$$\text{subj. to} \quad y(t) - \sum_{k_2=1}^{n_a} a_{k_2} y(t-k_2) = \sum_{k_1=1}^{n_b} (\mathbf{DX})(t-n_k-k_1,k_1),\ t=n,\dots,N, \tag{10b}$$

$$rank(\mathbf{X}) = 1. \tag{10c}$$

Note that we need to require that $rank(\mathbf{X}) = 1$ to not lose the possibility to decompose $\mathbf{X}$ as $\mathbf{X} = \mathbf{x}\mathbf{b}^\mathsf{T}$. The problem (10) is equivalent to (8) in the following sense. Assume that (10), has a unique solution $\mathbf{X}^*$, then $\mathbf{X}^*$ must satisfy $\mathbf{X}^* = \mathbf{x}^*(\mathbf{b}^*)^\mathsf{T}$, with $\mathbf{x}^*$ and $\mathbf{b}^*$ solving (8). Extracting the rank 1 component of $\mathbf{X}^*$, using *e.g.,* singular value decomposition, we can hence decide both $\mathbf{x}^*$ (and $\mathbf{u}^* = \mathbf{D}\mathbf{x}^*$) and $\mathbf{b}^*$ up to a multiplicative scalar (note that we can never do better with the information at hand, not even if we could solve (8)). The estimates of $\mathbf{a}$ will be identical for both problems (if the estimates are unique).

The technique of introducing the matrix $\mathbf{X}$ to avoid products between $\mathbf{x}$ and $\mathbf{b}$ is well known in optimization and referred to as *lifting* [Shor, 1987, Lovász and Schrijver, 1991, Nesterov, 1998, Goemans and Williamson, 1995].

Problem (10) is a non-convex optimization problem and not easier to solve than (8). To get an optimization problem we can solve, we remove the rank constraint and instead minimize the rank. Since the rank of a matrix is not a convex function, we replace the rank with a convex heuristic. Here we choose the nuclear norm, but other heuristics are also available (see for instance Fazel et al. [2001]). We then obtain the convex program

$$\min_{\mathbf{X},\mathbf{a}} \quad \|\mathbf{X}\|_* \tag{11a}$$

$$\text{subj. to} \quad y(t) - \sum_{k_2=1}^{n_a} a_{k_2} y(t-k_2) = \sum_{k_1=1}^{n_b} (\mathbf{DX})(t-n_k-k_1,k_1),\ t=n,\dots,N, \tag{11b}$$

which we refer to as *blind identification via lifting* (BIL).

Last, in the noisy setting we have to tolerate some nonzero modeling error. If the noise e is known to be bounded, say that $|e(t)| \leq \epsilon$, we suggest to use

$$\min_{\mathbf{X},\mathbf{a},\eta} \quad \|\mathbf{X}\|_* \tag{12a}$$

$$\text{subj. to} \quad y(t) - \sum_{k_2=1}^{n_a} a_{k_2} y(t-k_2) = \sum_{k_1=1}^{n_b} (\mathbf{DX})(t-n_k-k_1,k_1) + \eta(t), \tag{12b}$$

$$|\eta(t)| \leq \epsilon, \quad t=n,\dots,N, \tag{12c}$$

and if the noise is Gaussian,

$$\min_{\mathbf{X},\mathbf{a},\eta} \quad \|\mathbf{X}\|_* + \lambda\|\eta\|_2^2 \tag{13a}$$

$$\text{subj. to} \quad y(t) - \sum_{k_2=1}^{n_a} a_{k_2} y(t-k_2) = \sum_{k_1=1}^{n_b} (\mathbf{DX})(t-n_k-k_1,k_1) + \eta(t), \tag{13b}$$

$$t = n,\dots,N. \tag{13c}$$

In the latter case, we see $\lambda > 0$ as a design parameter and seek the largest λ such that $\mathbf{X}$ is rank 1.

6. ANALYSIS

The number of optimization variables in (8) is essentially $n_a + n_b + m$, under the assumption that $\mathbf{u} = \mathbf{D}\mathbf{x}$. We can hence not expect a reliable identification result from fewer than $n_a + n_b + m$ measurements. One may wonder how many measurements needed. As the constraint (11b) of BIL is linear in $\mathbf{X}$, we have the following result:

Theorem 1. (Guaranteed Recovery using BIL).

Consider the noise free blind ARX identification problem (8) and assume that it has a unique solution (up to a multiplicative scalar). Let the row vector $d_i \in \mathbb{R}^m$ be the i:th row of $\mathbf{D}$. If $\mathbf{A} =$

$$\begin{bmatrix} d_{n-n_k-1} & d_{n-n_k-2} & \cdots & d_{n-n_k-n_b} & y(n-1) & \cdots & y(n-n_a) \\ d_{n-n_k} & d_{n-n_k-1} & \cdots & d_{n-n_k-n_b+1} & y(n) & \cdots & y(n-n_a+1) \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ d_{n-n_k-2+n_b} & \cdots & & d_{n-n_k-1} & y(n+n_b-2) & \cdots & y(n-n_a+n_b-1) \\ \vdots & & & \vdots & \vdots & & \vdots \\ d_{N-n_k-1} & d_{N-n_k-2} & \cdots & d_{N-n_k-n_b} & y(N-1) & \cdots & y(N-n_a) \end{bmatrix}$$

has full column rank, then the ARX model and input solving (8) are recovered, up to a multiplicative scalar, by BIL.

Proof. Problem (8) has a unique solution by assumption. Form $\mathbf{X}^*$ by multiplying the solutions, $\mathbf{x}^*$ and $\mathbf{b}^*$, of (8). That is, $\mathbf{X}^* = \mathbf{x}^*(\mathbf{b}^*)^\mathsf{T}$. Let $a_1^*,\dots,a_{n_a}^*$ denote the values for $a_1,\dots,a_{n_a}$ that solve (8) and define θ^* as

$$\theta^* = \begin{bmatrix} \mathbf{X}^*(:,1)^\mathsf{T} & \mathbf{X}^*(:,2)^\mathsf{T} & \dots & \mathbf{X}^*(:,n_b)^\mathsf{T} & -a_1^* & \dots & -a_{n_a}^* \end{bmatrix}^\mathsf{T}.$$

We must have that

$$[y(n)\ y(n+1)\ \dots\ y(N)]^\mathsf{T} = \mathbf{A}\theta^*. \tag{14}$$

Note that all solutions of (11) also must satisfy (14). Since $\mathbf{A}$ has full column rank the solution of (14) is unique and we must have that BIL gives θ^*. □

Note that if the linear constraints of (11) alone give the solution of BIL, no optimization is necessary. Seeking the matrix $\mathbf{X}$ that gives the minimum nuclear norm is only of interest if we have too few measurements for the constraints to uniquely define the solution but more measurements than $n_a + n_b + m$.

The noisy case is harder to analyze and we leave the analysis as future work.

7. COMPUTING $\lambda^{\min}$

In the noisy version (13) of BIL, the design parameter λ has to be chosen. Since λ regulates the tradeoff between the nuclear norm and the squared norm of the estimated noise

η, it is natural to seek the largest λ such that the estimate $\mathbf{X}$ is rank 1. In seeking this λ, the value for $\lambda^{\min}$ may come in handy. $\lambda^{\min}$ is defined as the largest λ such that $\mathbf{X} = 0$ in BIL. Since the estimate for $\mathbf{X}$ will stay the same for all $\lambda \leq \lambda^{\min}$, we should limit our search of λ to be within $[\lambda^{\min}, \infty]$. One may for example start with $\lambda = \lambda^{\min}$ and then successively increase λ as long as $rank(\mathbf{X}) = 1$.

Theorem 2. (Computing $\lambda^{\min}$).

Consider the optimization problem given in (13). There exists a λ, denoted $\lambda^{\min}$, such that whenever $\lambda \leq \lambda^{\min}$, solving (13) results in $\mathbf{X} = 0$. $\lambda^{\min}$ is given by:

$$1/\lambda^{\min} = \underset{\mathbf{V}\in\mathbb{R}^{m\times n_b}}{\arg\min} \quad \|\mathbf{V}\| \tag{15a}$$

$$\text{subj. to} \quad 0 = \mathbf{V}(i,j) - 2\sum_{t=n}^{N}\Big(y(t) - \sum_{k_2=1}^{n_a}\hat{a}_{k_2}y(t-k_2)\Big)\mathbf{D}(t-n_k-j,i), \tag{15b}$$

$$i = 1,\dots,m,\ j = 1,\dots,n_b, \tag{15c}$$

with

$$\{\hat{a}_1,\dots,\hat{a}_{n_a}\} = \underset{\mathbf{a}}{\arg\min}\sum_{t=n}^{N}\Big(y(t) - \sum_{k_2=1}^{n_a}a_{k_2}y(t-k_2)\Big)^2. \tag{16}$$

Proof. The noisy version of BIL can be rewritten to take the form

$$\min_{\mathbf{X},\mathbf{a}} \quad \|\mathbf{X}\|_* + \lambda\sum_{t=n}^{N}\Big(y(t) - \sum_{k_2=1}^{n_a}a_{k_2}y(t-k_2) - \sum_{k_1=1}^{n_b}(\mathbf{DX})(t-n_k-k_1,k_1)\Big)^2. \tag{17}$$

The nuclear norm is not differentiable and it follows that for $\mathbf{X} = 0$ to be a valid solution, zero needs to be in the subdifferential of the objective with respect to $\mathbf{X}$ evaluated at $\mathbf{X} = 0$ (see *e.g.,* Bertsekas et al. [2003, Prop. 4.7.2]). The subdifferential of the objective of (17) at $\mathbf{X} = 0$ and $\mathbf{a} = \hat{\mathbf{a}}$ with respect to the (i,j)th element of $\mathbf{X}$ can be shown equal to

$$\mathbf{V}(i,j) - 2\lambda\sum_{t=n}^{N}\Big(y(t) - \sum_{k_2=1}^{n_a}\hat{a}_{k_2}y(t-k_2)\Big)\mathbf{D}(t-n_k-j,i). \tag{18}$$

We further have that $\|\mathbf{V}\| \leq 1$ from the subdifferential of the nuclear norm (see for instance Watson [1992] or Recht et al. [2010]). $\|\cdot\|$ is here the operator norm (the largest singular value). To find $\lambda_{\min}$ we now consider the optimization problem

$$\max_{\mathbf{V}\in\mathbb{R}^{m\times n_b},\lambda} \quad \lambda \tag{19a}$$

$$\text{subj. to} \quad 0 = \mathbf{V}(i,j) - 2\lambda\sum_{t=n}^{N}\Big(y(t) - \sum_{k_2=1}^{n_a}\hat{a}_{k_2}y(t-k_2)\Big)\mathbf{D}(t-n_k-j,i), \tag{19b}$$

$$i = 1,\dots,m,\ j = 1,\dots,n_b, \tag{19c}$$

$$\|\mathbf{V}\| \leq 1, \tag{19d}$$

which can be shown equivalent to (15). □

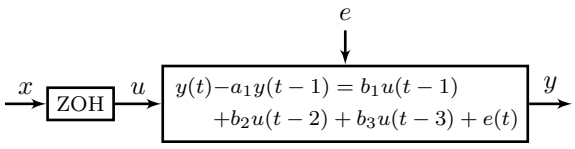
$\lambda^{\min}$ was also numerically verified.

8. SOLUTION ALGORITHMS AND SOFTWARE

Many standard methods of convex optimization can be used to solve problem (11), (12) and (13). Systems such as CVX [Grant and Boyd, 2010, 2008] or YALMIP [Löfberg, 2004] can readily handle the nuclear norm. For large scale problems, the *alternating direction method of multipliers* (ADMM, see *e.g.,* Bertsekas and Tsitsiklis [1997], Boyd et al. [2011]) is an attractive choice and we have previously shown that ADMM can be very efficient on similar problems Ohlsson et al. [2013]. Code for solving (11), (12) and (13) will be made available on `http://www.rt.isy.liu.se/~ohlsson/code.html`

9. NUMERICAL ILLUSTRATION

Consider the system given in the diagram below.



Here the values x were generated by independently sampling from a unit Gaussian and the noise e by independently sampling from a uniform distribution between $-\epsilon/2$ and $\epsilon/2$. The ZOH (zero-order hold) block holds the input to the ARX system constant for 6 consecutive samples. We can therefore express $\mathbf{u}$ in terms of $\mathbf{x}$ as

$$\mathbf{u} = \begin{bmatrix} \mathbb{1}_{6\times1} & \mathbb{0}_{6\times1} & \mathbb{0}_{6\times1} & \dots & \mathbb{0}_{6\times1} \\ \mathbb{0}_{6\times1} & \mathbb{1}_{6\times1} & \mathbb{0}_{6\times1} & \dots & \mathbb{0}_{6\times1} \\ \vdots & & \ddots & & \vdots \\ \mathbb{0}_{6\times1} & \dots & & \mathbb{1}_{6\times1} & \mathbb{0}_{6\times1} \\ \mathbb{0}_{6\times1} & \dots & & \mathbb{0}_{6\times1} & \mathbb{1}_{6\times1} \end{bmatrix}\mathbf{x}. \tag{20}$$

We identify the matrix in the relation between $\mathbf{u}$ and $\mathbf{x}$ as $\mathbf{D}$. The ARX coefficients used were

$$a_1 = -0.3,\ b_3 = 1,\ b_2 = 2,\ b_1 = 3. \tag{21}$$

Figure 1 shows the output y for $\epsilon = 5$.

If the noisy version of BIL (12) is used to estimate $\mathbf{u}$ and an ARX model, we get the input-estimate given in Figure 2 and the ARX coefficients:

$$a_1 = -0.21,\ b_3 = 0.91,\ b_2 = 1.80,\ b_1 = 2.7. \tag{22}$$

It is interesting to notice that if we instead would be given the true input $\mathbf{u}$ and only estimated the ARX coefficients

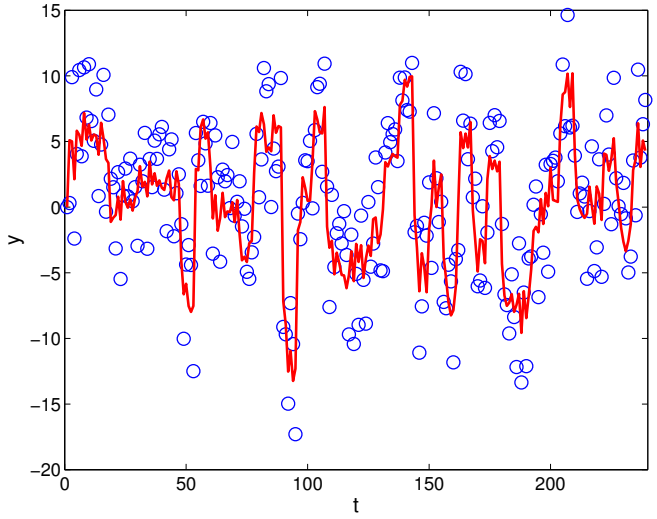


Fig. 1. The noise free (solid line) and noisy outputs (circles).

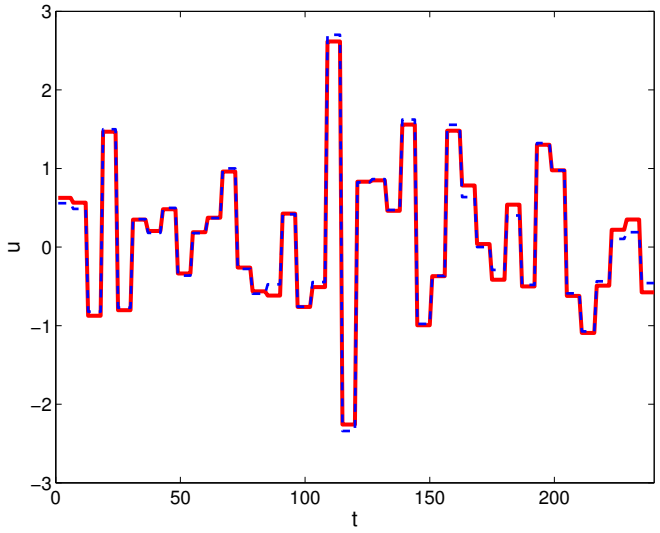


Fig. 2. The estimated (dashed line) and the true input $\mathbf{u}$ (solid line).

by minimizing the squared residuals between the output $\mathbf{y}$ and the predicted output, we would have got the estimates:

$$a_1 = -0.30,\ b_3 = 0.88,\ b_2 = 2.39,\ b_1 = 2.85. \qquad (23)$$

As seen, these estimates are not that much better than what BIL is providing (see (22)). Remember that BIL is only given the y-measurements and not the inputs $\mathbf{u}$. It is therefore quite remarkable that the estimates of BIL is comparable to those given in (23).

To further study the robustness of BIL we carried out a Monte Carlo simulation. In the simulation, the noise level ϵ was varied between 0 and 5. For each noise level, 100 trials were carried out with different noise and input realizations. The true ARX model was kept fixed (the same as above). The results are summarized in Figure 3.

The setup of above example does not give that $\mathbf{A}$ has full column rank. Nevertheless, a perfect result was obtained in the noise free case. It can however be verified that if $\mathbf{D}$ is instead generated by independently sampling each

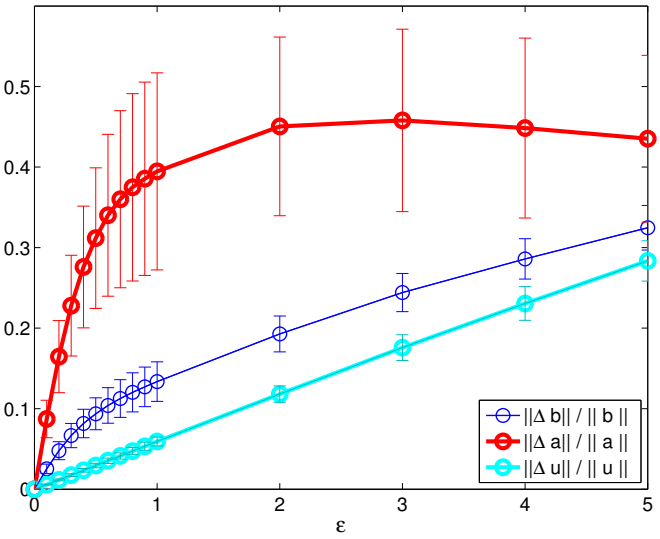


Fig. 3. The relative errors along with their 0.5 standard deviation error bounds for varying noise levels.

element from a unit Gaussian distribution (everything else unchanged), the resulting $\mathbf{A}$ has full column rank.

10. CONCLUSION

This paper presented a novel framework for blind system identification of ARX model. The framework uses the fact that the problem can be rewritten as a rank minimization problem. A convex relaxation is presented to approximate the sought ARX parameters and the unknown inputs.

REFERENCES

K. Abed-Meraim and E. Moulines. A maximum likelihood solution to blind identification of multichannel FIR filters. In *EUSIPCO*, pages 1011–1015, 1994.

K. Abed-Meraim, Wanzhi Qiu, and Y. Hua. Blind system identification. *Proceedings of the IEEE*, 85(8):1310–1322, 1997.

K. Abed-Meraim, P. Loubaton, and E. Moulines. A subspace algorithm for certain blind identification problems. *IEEE Transansactions on Information Theory*, 43(2):499–511, September 2006.

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

E.-W. Bai and Minyue Fu. A blind approach to Hammerstein model identification. *IEEE Transactions on Signal Processing*, 50(7):1610–1619, 2002.

E.-W. Bai, Q. Li, and S. Dasgupta. Blind identifiability of IIR systems. *Automatica*, 38(1):181–184, 2010.

D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.

D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, 2003.

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*, 2011.

M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system

approximation. In *Proceedings of the 2001 American Control Conference*, pages 4734–4739, 2001.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42 (6):1115–1145, November 1995.

M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. D. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. `http://cvxr.com/cvx`, August 2010.

Y. Hua. Blind methods of system identification. *Circuits, Systems and Signal Processing*, 21(1):91–108, 2002. ISSN 0278-081X.

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. URL `http://control.ee.ethz.ch/~joloef/yalmip.php`.

L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 1:166–190, 1991.

Y. Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods & Software*, 9:141–160, 1998.

H. Ohlsson, A. Y. Yang, R. Dong, M. Verhaegen, and S. S. Sastry. Quadratic basis pursuit. *CoRR*, abs/1301.7002, 2013.

B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

Y. Sato. A method of self-recovering equalization for multilevel amplitude-modulation systems. *IEEE Transactions on Communications*, 23(6):679–682, 1975.

N.Z. Shor. Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25:1–11, 1987.

L. Sun, W. Liu, and A. Sano. Identification of a dynamical system with input nonlinearity. *IEE Proceedings – Control Theory and Applications*, 146(1):41–51, 1999.

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*, volume 2, pages 856–860, 1991.

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*, 61(9):2219–2230, 2013.

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*, 80(2): 302–313, 2007.

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

J. Wang, A. Sano, T. Chen, and B. Huang. A blind approach to identification of Hammerstein systems. In F. Giri and E.-W. Bai, editors, *Block-oriented Nonlinear System Identification*, volume 404 of *Lecture Notes in Control and Information Sciences*, pages 293–312. Springer London, 2010.

G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170(0):33–45, 1992. ISSN 0024-3795.

A. Zerva, A.P. Petropulu, and P.-Y. Bard. Blind deconvolution methodology for site-response evaluation exclusively from ground-surface seismic recordings. *Soil Dynamics and Earthquake Engineering*, 18(1):47–57, 1999.

A. Zerva, A.P. 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.