Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-7 (2016) 1014–1019

Identification of Linear Dynamic Systems using Dynamic Iterative Principal Component Analysis ⋆

Deepak Maurya * **Arun K. Tangirala** ** **Shankar Narasimhan** **

* *Department of Electrical Engineering, IIT Madras, Chennai, India (e-mail: ee11b109@ee.iitm.ac.in).*
** *Department of Chemical Engineering, IIT Madras, Chennai, India (e-mail: arunkt@iitm.ac.in, naras@iitm.ac.in).*

Abstract: The paper is concerned with identifying models from data that have errors in both outputs and inputs, popularly known as the errors-in-variables (EIV) problem. The total least squares formulation of the problem is known to offer a few well-known solutions. In this work, we present a novel and systematic approach to the identification of linear dynamic models for the EIV case in the principal component analysis (PCA) framework. A methodology for the systematic recovery of the process model, including the determination of order and delay, using what we term as *dynamic, iterative* PCA is presented. The core step consists of determining the structure of the constraint matrix by a systematic exploitation of the stacking and PCA order, input-output partitioning of the constraint matrix and an appropriate rotation. Optimal estimates of the (input-output) noise covariance matrices are also obtained. The proposed method can be applied to a broad class of linear processes including the case of unequal and unknown error variances. Simulation results are presented to demonstrate the effectiveness and consistency of the proposed method.

Keywords: identification; principal component analysis; errors-in-variables; multivariable

1. INTRODUCTION

System identification of linear dynamic systems has been a subject of study for several decades with a rich literature at the user's disposition (Ljung, 1999), naturally owing to its crucial role in all applications of process systems engineering. A larger proportion of the literature has concerned with the traditional case of deterministic (error-free) inputs. A realistic problem that is, however, relatively less-evolved is that of the errors-in-variables (EIV) case where the inputs are also known with errors.

Among the different approaches that exist for the EIV problem, the total least squares (TLS) formulation offers a few efficient solutions. One of the early techniques involved finding solutions to the EIV (linear) regression problem in the TLS framework (Huffel and Vandewalle, 1991). Instrumental variable techniques and their generalizations offer a different approach to the EIV regression problem through the use of instruments (Soderstrom, 2007). Multi-input, multi-output (MIMO) systems are effectively handled by the state-space modelling paradigm and the use of subspace identification techniques. The subspace error-in-variables (Chou and Verhaegen, 1997) formulation is a member of this class of methods. On the other hand, multivariate statistical data analysis techniques have been proved to be highly successful in identifying linear relationships between variables from measurements, especially when the causal variables are collinear. Principal component analysis (PCA) is one of the most widely used and effective technique for solving the TLS problem, albeit under certain assumptions. A key factor attributed to the success of PCA (and its variants) methods is its simplicity, numerical robustness and versatility. Applications of PCA span a diverse set including identification, dimensionality reduction, process monitoring (fault detection and diagnosis), feature extraction and as a pre-processing tool for other identification methodologies. The use of dynamic PCA methods in errors-in-variables subspace identification (Li and Qin, 2001) is an example of the blend of PCA and standard subspace methods in identification.

A distinguishing factor in the modelling philosophy of PCA-based approach vis-a-vis standard approaches (e.g., regression, subspace) is that it does not *a priori* distinguish between dependent and independent variables, whereas the latter approaches observe this distinction right upfront. Further, linear models are expressed as *constraints* (among variables) in the PCA framework. These constraints are identified by either a singular value decomposition (SVD) of the data matrix or the eigenvalue analysis of the sample covariance matrix. The distinction between dependent and independent variables is invoked in the post-analysis stage, but only when required by the application. For instance, fault detection applications do not require this distinction since they largely work with deviations from constraints. In contrast, fault diagnosis, identification or prediction applications invoke this distinction and accordingly partition the constraint matrix to extract the (causal) input-output model.

The generic formulation of PCA is suited for identifying steady-state or instantaneous linear relationships among variables. Dynamic PCA (DPCA) (Ku et al., 1995) was introduced as a natural extension of this formulation to

⋆ Corresponding author: Arun K. Tangirala, email: arunkt@iitm.ac.in

10.1016/j.ifacol.2016.07.335

suit identification of dynamic relationships by way of applying PCA on a matrix of lagged variables. While this idea has been extensively applied in process monitoring and to a certain extent in regression, a rigorous method of recovering the dynamic model for the error-free variables from the results of DPCA is still elusive. Specifically, a theoretically sound way of determining the order, dynamic relationships and the noise covariance matrix begs attention. Two related works merit mention in this regard. The first one pertains to the last principal component analysis (LPCA). technique (Huang, 2001), which is based on the DPCA idea. It recovers the dynamic model from the last eigenvector of the sample covariance matrix. However, it is severely limited by the assumption of exactly one constraint among the lagged variables, which arises only when the user has an accurate knowledge of the process order (of dynamics) - an assumption that is largely unrealistic. Further it assumes that the errors in the input and output have equal variances, which is again quite restrictive. The second work is that of Vijaysai et al. (2005) who develop what they term as the *generalized augmented* PC-based regression, largely motivated by the limitation of multivariate statistical techniques such as partial least squares and principal component regressors in solving the EIV problem as well as LPCA in handling multiple linear relationships and unequal variances. The idea therein is to scale the data with the square root of the inverse of noise covariance matrix that is assumed to be known. However, in almost all realistic situations, the availability of this information is largely confined to academic presentations.

In this work, we present a systematic identification of linear dynamic processes from input-output data using dynamic, iterative PCA (IPCA). The steady-state IPCA was proposed by Narasimhan and Shah (2008) to overcome the shortcomings of PCA in identifying the number of steady-state relationships when the errors in variables have unequal variances, and are possibly spatially correlated. The main strategy in IPCA is the scaling of data with the square root of inverse of noise covariance matrix followed by an application of the ordinary PCA. In this respect, IPCA is to PCA as to what weighted LS method is to ordinary LS in standard linear regression. The noise covariance matrix and the linear relations are estimated iteratively, and hence the name. Section 2.2 reviews the technical details of IPCA.

The success of the proposed method rests on two key steps: (i) the relation between the process order, the number of constraints identified by the PCA and the maximum lag that is used for stacking the input and output variables, and (ii) the rotation of constraint matrices for the dependent (output) and independent (input) variables. The first one is easy to derive, as we show later in this work. Rotation of constraint matrices are derived by optimally solving a set of overdetermined equations. Finally, the noise covariance matrix estimate is obtained naturally as an outcome of the IPCA algorithm. A minor contribution of this work is also a modified IPCA algorithm that relaxes the identifiability constraints in the original version. In general, the proposed method also offers provision for incorporating prior knowledge of any or all of delay, order, input dynamics and noise covariance matrix, if available. It must be noted that the development in this work is confined to the class of open-loop systems.

The rest of the paper is organized as follows. Section 2 reviews the basic ideas underpinning PCA and IPCA in solving the EIV identification or regression problem. Further, a review of the DPCA with an example that highlights the limitations of technique is also included. In Section 3 we present the main contribution of this work, which is a systematic method for identifying the dynamic model and its accompaniments from data. Simulation results are discussed in Section 4, wherein Monte-Carlo simulations are presented to study the goodness of estimates and to illustrate the impact of maximum lag on model quality. The paper ends with a few concluding remarks in 5.

2. FOUNDATIONS

We begin with a review of the use of PCA in the identification of linear static models for the EIV case, followed by a brief exposition of the iterative PCA for the same purpose, however for a broader class of problems.

It is useful to first introduce the generic problem that is addressed in the PCA and IPCA literature. The set up and statement is as follows.

A matrix $\mathbf{X} \in \mathbb{R}^{N\times M}$, constructed by arranging N observations of M variables[1], i.e.,

$$\mathbf{X} = [\mathbf{x}[0]\ \mathbf{x}[1]\ \cdots\ \mathbf{x}[N-1]]^T, \quad \mathbf{x}[k] = [x_1[k]\ \cdots\ x_M[k]]^T \tag{1}$$

is constrained by $d < M$ linear relations, i.e., $\{x_i[k]\}_{i=1}^{M}$ fall out of a deterministic (noise-free) linear process. Mathematically,

$$\mathbf{A}_0\mathbf{x}[k] = 0, \qquad \text{where } \mathbf{A}_0 \in \mathbb{R}^{d\times M} \tag{2a}$$

The goal is to identify the row dimension d and the constraint matrix $\mathbf{A}_0$ from $\mathbf{X}$.

The EIV identification problem is that of identifying the constraints in (2) and thereby the regression model later in (A.4) from *noisy measurements* of $\mathbf{X}$,

$$\mathbf{z}[k] = \mathbf{x}[k] + \mathbf{e}[k] \tag{3}$$

$$\mathbf{Z} = \mathbf{X} + \mathbf{E} \tag{4}$$

where $\mathbf{e}[k]$ is a vector of white-noise errors with noise covariance $\Sigma_{\mathbf{e}}$.

In the following sections we review the mechanics of PCA and IPCA in solving the above problems, particularly the EIV problem.

2.1 PCA for steady-state identification

Principal component analysis is more than a century-old multivariate statistical analysis tool (Pearson, 1902) that searches for correlations among the columns of a matrix through a search for the zero singular values or the zero eigenvalues of the sample covariance matrix. For the *noise-free* case, the number of zero singular values of $\mathbf{X}$ / eigenvalues of $\mathbf{S_x} \triangleq \dfrac{1}{N}\mathbf{X}^T\mathbf{X}$ and the associated singular vectors / eigenvectors provide the dimensionality of and a basis for the constraint matrix $\mathbf{A}$, respectively. A short review of the use of PCA is identification is provided in Appendix A.

In the EIV case, the PCA of $\mathbf{Z}$, i.e., eigenvalue analysis of the sample covariance matrix $\mathbf{S_z} \triangleq \dfrac{1}{N}\mathbf{Z}^T\mathbf{Z}$ does not result in any zero eigenvalue since there exists no pair of columns of $\mathbf{Z}$ that are linearly related. Moreover, the impact of noise

[1] Another convention to set up $\mathbf{X}$ is as an $M\times N$ matrix, for example in chemometrics,

variance on the eigenvalues of $\mathbf{S_z}$ is too complicated to be understood. Consequently, in general, it is not possible to accurately determine the size of $\mathbf{A}$ in (2) and hence the matrix as well.

Furthermore, even when the dimension d is known, PCA cannot theoretically recover the constraint matrix consistently in general. Despite this fact, one finds in literature several heuristic methods of identifying the dimension d and the constraint matrix $\mathbf{A}$. The only exception is the case of spatially uncorrelated errors in variables with equal variances, i.e., $\Sigma_e = \sigma_e^2 \mathbf{I}_{M\times M}$. In such a situation, the last d eigenvalues (assuming that the signal-to-noise ratio (SNR) > 1) are all equal to σ_e^2, i.e., $\lambda_{M-d+1} = \cdots = \lambda_M = \sigma_e^2$ and PCA recovers an optimal and consistent estimate of $\mathbf{A}_0$, still differing by a non-singular factor.

For the general scenario, i.e., non-diagonal $\Sigma_\mathbf{e}$, the iterative PCA, briefly reviewed in the following section, overcomes several of the above shortcomings in a theoretically sound manner.

2.2 IPCA for linear identification

The iterative PCA (IPCA) introduced by Narasimhan and Shah (2008) not only produces a consistent estimate of the constraint matrix in (2) for a given order, but also correctly identifies d when it is not known a priori. The key idea is to scale the data matrix with $\Sigma_\mathbf{e}^{-1/2}$ so that the problem is transformed to the ordinary PCA on the scaled data. There is, however, a difference. Unlike in the standard case, the dimension of $\mathbf{A}_0$ is determined by the number of *unity-valued eigenvalues* of the sample covariance matrix of the *scaled* data. The theoretical basis for the foregoing properties of IPCA is briefly reviewed below. For full technical details, the reader is referred to Narasimhan and Shah (2008).

Starting from (3) and assuming "open-loop" conditions, i.e., errors are uncorrelated with the variables in $\mathbf{x}$,

$$\Sigma_\mathbf{z} = \Sigma_\mathbf{x} + \Sigma_\mathbf{e} \tag{5}$$

where $\Sigma_\mathbf{x} = \lim_{N\to\infty} \mathbf{S_x}$ exists under the quasi-stationarity assumption (Ljung, 1999) on $\mathbf{z}[k]$.

Introduce the scaled data $\mathbf{z}_s = \Sigma_\mathbf{e}^{-1/2}\mathbf{z}[k]$ and similarly for the deterministic component. Then, the covariance equation above in (5) for the scaled data takes the form

$$\Sigma_{\mathbf{z}_s} = \Sigma_{\mathbf{x}_s} + \mathbf{I} \tag{6}$$

Eigenvalue analysis of the LHS matrix, by virtue of the eigenvalue shift theorem, results in

$$\lambda(\Sigma_{\mathbf{z}_s}) = \lambda(\Sigma_{\mathbf{x}_s}) + 1 \tag{7}$$

Consequently, *all zero-valued eigenvalues of* $\Sigma_{\mathbf{x}_s}$ *map to unity-valued eigenvalues of* $\Sigma_{\mathbf{z}_s}$, the size of which determines the number of linear relations d This applies to sample covariance matrices $\mathbf{X}$ and $\mathbf{Z}$ as well. Observe that scaling preserves the rank of $\Sigma_{\mathbf{x}_s}$. Thus, IPCA offers a theoretically sound way of determining the number of linear relations, d, and hence a basis for $\mathbf{A}_0$ as well from noisy data.

Estimation of Error Covariance Matrix Denote the estimate of $\mathbf{A}_0$ at the i^{th} iteration by $\hat{A}^{(i)}$ (the subscript is dropped since only a basis can be determined). Subsequently, one generates constrained residuals as

$$\mathbf{r}_i[k] = \hat{A}^i\mathbf{z}[k] = \hat{A}^i\mathbf{x}[k] + \hat{A}^i\mathbf{e}[k] \tag{8}$$

If the estimate $\hat{A}^{(i)}$ is indeed a basis for $\mathbf{A}_0$, then $\hat{A}^{(i)} = \mathbf{TA}_0$. Then, the first term on the RHS vanishes and the constrained residuals are independent and normally distributed with zero mean and covariance matrix $\Sigma_{\mathbf{r}_i}$.

$$\Sigma_{\mathbf{r}_i} = \hat{A^{(i)}}\Sigma_\mathbf{e}(\hat{A^{(i)}})^T \tag{9}$$

Consequently, an estimate $\Sigma_\mathbf{e}$ is obtained by solving the likelihood problem

$$\min_{\Sigma_\mathbf{e}} \quad N\log|\hat{A^{(i)}}\Sigma_\mathbf{e}(\hat{A^{(i)}})^T| + \sum_{i=1}^{N}(\mathbf{r}_i^T[k](\hat{A^{(i)}}\Sigma_\mathbf{e}(\hat{A^{(i)}})^T)^{-1}\mathbf{r}_i[k]) \tag{10}$$

Identifiability constraints exist since we are essentially recovering an $M\times M$ error covariance matrix from $d\times d$ residual covariance matrix using the relation in (9). Since $\Sigma_{\mathbf{r}_i}$ is symmetric, we have $d(d+1)/2$ unique equations. For a diagonal $\Sigma_\mathbf{e}$, the constraint is therefore

$$\frac{d(d+1)}{2} \geq P \tag{11}$$

where P is the number of elements of $\Sigma_\mathbf{e}$ that need to be estimated. For a diagonal $\Sigma_\mathbf{e}$, $P = M$.

The idea is to iterate between estimates of $\mathbf{A}$ and $\Sigma_\mathbf{e}$. An initial estimate of constraint matrix $\hat{A}^0$ is usually generated by PCA for a guess of d, the number of linear relations. If the guess d is correct, the converged estimate of $\Sigma_\mathbf{e}$ should result in exactly d unity-valued eigenvalues. A mismatch indicates an incorrect guess. Usually one chooses a guess of d that satisfies the identifiability constraint in (11) and proceeds as required.

3. PROPOSED METHODOLOGY

We open this section with a short review of dynamic PCA as proposed by Ku et al. (1995) by means of a motivating example that also highlights the shortcomings of this method. Subsequently, the proposed method is elucidated.

The generic problem of interest is that of identifying a MIMO system from measured data. We restrict ourselves, however, in this article to a single-input single-output (SISO) system.

Consider the class of parametric deterministic SISO linear time-invariant dynamic input($u^\star$)-output($y^\star$) systems described by

$$y^\star[k] + \sum_{i=1}^{n_y} a_i y^\star[k-i] = \sum_{j=D}^{n_u} b_j u^\star[k-i] \tag{12}$$

where n_y and n_u are output and input order, respectively and D is the input-output delay. The EIV identification problem is that of estimating the coefficients $\{a_i\}_{i=1}^{n_y}$, $\{b_j\}_{j=D}^{n_u}$ from measurements of $y^\star[k]$ and $u^\star[k]$, denoted by $y[k]$ and $u[k]$, respectively.

3.1 Dynamic PCA and its shortcomings

Dynamic PCA attempts to solve the above identification problem by constructing a matrix of lagged measurements. To illustrate the basic idea, consider a unit-delay, second-order, deterministic SISO system

$$y^\star[k] + 0.4y^\star[k-1] + 0.6y^\star[k-2] = 1.2u^\star[k-1] \tag{13}$$

Assume that N measurements $\{y[k]\}$ and $\{u[k]\}$ are available.

DPCA performs a PCA of the lagged data matrix, i.e., whose columns are the stacked measurements $\{y[k]\}$

through $\{y[k-L_y]\}$ and $\{u[k]\}$ through $\{u[k-L_u]\}$, where L_y and L_u are user-specified maximum lags for the output and input, respectively. A general practice is to let $L_y = L_u = L$, a sufficiently large value.

For the system in (13), we assume that the order is known, i.e., $L = 2$ and construct the k^{th} row of $\mathbf{Z}$ as

$$\begin{aligned}\mathbf{z}[k] &= [y[k]\ y[k-1]\ y[k-2]\ u[k]\ u[k-1]\ u[k-2]]^T \\ \mathbf{Z} &= [\mathbf{z}[3]\ \mathbf{z}[4]\ \cdots\ \mathbf{z}[N]]^T \end{aligned} \tag{14}$$

From (13), there exists only one relation between the underlying deterministic variables, with the true constraint:

$$A_0 = [1 \quad 0.4 \quad 0.6 \quad 0 \quad -1.2 \quad 0] \tag{15}$$

When the data is generated with errors of equal variances, i.e., $\sigma_{e_y}^2 = \sigma_{e_u}^2 = 0.09032$, DPCA of $\mathbf{Z}$ in (14) identifies the constraint correctly. Firstly, the eigenvalues of $\mathbf{S_z}$ are

$$\Lambda = \text{diag}([4.1028 \quad 3.4908 \quad 1.6919 \quad 1.1134 \quad 0.5024 \quad 0.1035])$$

The last eigenvalue is, theoretically, an estimate of σ_e^2 and in close agreement with the value used in simulation. The corresponding eigenvector, after normalizing the first coefficient to unity, is

$$\hat{A} = [1 \quad 0.4204 \quad 0.6052 \quad -0.0028 \quad -1.2025 \quad -0.0377]$$

which is in close agreement with the true value in (15). The truly zero-valued coefficients on $u[k]$ and $u[k-3]$ are non-zero, but negligibly small, due to finite-sample errors.

The realistic situation is that (i) the order is seldom known accurately and (ii) the noise variances are in general unequal and unknown. Unfortunately, DPCA fails to identify the model in these situations and there exist no known single method that overcomes both these shortcomings. To illustrate the effect of the second factor, which is more severe among the two, we generate data by setting $\sigma_{e_y}^2 = 0.2365$ and $\sigma_{e_u}^2 = 0.098$ (such that the SNR is 10). Assuming once again that the order is known, eigenvalue of $\mathbf{S_z}$ is performed to yield,

$$\Lambda = \text{diag}([4.3376 \quad 3.706 \quad 1.6914 \quad 1.0846 \quad 0.3956 \quad 0.1583])$$

Since it is known that a single constraint exists, we turn to the eigenvector corresponding to the last eigenvalue

$$\hat{A} = [1 \quad 0.2916 \quad 0.5542 \quad 0.0570 \quad -1.2913 \quad 0.1787]$$

which is, evidently, a *biased* estimate of the true constraint in (15).

Difficulties compound, of course, when the order is unknown, in which case the maximum lag L is set to a sufficiently high value. Consequently, multiple linear relations exist among the corresponding deterministic columns of $\mathbf{Z}$. For instance, it is easy to verify that choosing $L = 3$ results in $d = 2$ constraints. There exists no systematic way of recovering the model from these excess constraints, even if variances were equal.

The proposed methodology addresses both the foregoing issues in the IPCA framework.

3.2 Dynamic IPCA for EIV identification

To begin with, the basic idea underlying the proposed method is explained with a symbolic example of a first-order, unit-delay system.

Assume that we are interested in identifying a first-order, unit delay SISO system:

$$y^\star[k] + a_1 y^\star[k-1] = b_1^\star u[k-1] \tag{16}$$

Assume that the order is unknown and that the maximum lag is set to $L = 2$. Two linear relations would exist for the error-free variables with such a choice, i.e., $d = 2$. Ideally the objective is to recover the true constraint matrix,

$$\mathbf{A}_0 = \begin{bmatrix} 1 & a_1 & 0 & 0 & b_1 & 0 \\ 0 & 1 & a_1 & 0 & 0 & b_1 \end{bmatrix} \tag{17}$$

assuming that the *error-free inputs are not bound by any linear relation*, an assumption that is fairly unrestrictive.

Employing the IPCA for this problem will result in two unity eigenvalues, which correspond to the two relations that exist in the deterministic part of $\mathbf{Z}$, one for $y[k]$ in terms of $y[k-1]$ and $u[k-1]$ and the other for $y[k-1]$ in terms of $y[k-2]$ and $u[k-2]$. However, the constraint matrix obtained from DIPCA is a rotated version of the true $\mathbf{A}_0$. The key result in this paper is the development of a method to rotate back the identified constraint matrix so as to get the true $\mathbf{A}_0$.

The first step is that of *process order determination.* From the number of identified linear relations and the fact that only one relation should be expected (since the system is a single-output and no input relations exist), it is easy to derive the process order as $n_y = L-(d-1) = 2-(2-1) = 1$. This is the key step.

Once the order is determined, the remaining step is to *reconstruct the true constraint.* For this purpose, one could go back and re-assemble $\mathbf{Z}$ with $L = 1$. However, it is not recommended to do so due to the unequal variances of errors, which calls for an iterative estimation of $\Sigma_{\mathbf{e}}$ and $\mathbf{A}_0$. From Section 2.2, it is clear that estimation of $\Sigma_{\mathbf{e}}$ is bound by the identifiability constraint in (11) that calls for a "large enough" d, which translates to a large enough L. For instance, in this example, it is required to estimate $2(L+1) = 6$ diagonal unknowns in $\Sigma_{\mathbf{e}}$, of which strictly speaking there are only *two* distinct elements, σ_y^2 and σ_u^2. Therefore, L should be large enough so that $d(d+1)/2 \geq 2$, yielding a minimum of $d = 2$. In fact, this is the requirement for all SISO systems, i.e., L should be chosen sufficiently large that there are at least two linear relations in the deterministic part of $\mathbf{Z}$. This is *not necessarily* a limitation of the proposed method, but in general any other PCA-based technique since at least two equations are required to estimate two unknowns.

An important remark and a modification of the IPCA algorithm with respect to the estimation of $\Sigma_{\mathbf{e}}$ for the dynamic EIV identification problem is discussed in the generic approach shortly.

With an accurate estimate of the order, the next step is that of partitioning the constraint matrices corresponding to dependent (lagged outputs) and independent (lagged inputs) variables as $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_I$, respectively. Subsequently, the structure of $\hat{\mathbf{A}}_D$ is constructed, i.e., the locations of zero- and unity-valued elements are identified. For the example under discussion, the system has already been determined to be first-order. Therefore, $\hat{\mathbf{A}}_D$ should possess the structure

$$\mathbf{A}_D^\star = \begin{bmatrix} 1 & \alpha_1 & 0 \\ 0 & 1 & \alpha_2 \end{bmatrix} \tag{18}$$

where we have used the fact that leading coefficients of difference equation models are always unity and exploited the shift property of the rows containing shifted difference equations. Furthermore, the constants $\alpha_1, \alpha_2 \in \mathbb{R}$ should also be identical, i.e., $\alpha_1 = \alpha_2$.

As a penultimate step, a rotation of the matrix $\hat{\mathbf{A}}_D$ to yield this structure is determined so that

$$\text{structure}(\mathbf{R}_D\hat{\mathbf{A}}_D) = \text{structure}(\mathbf{A}_D^\star) \tag{19}$$

This rotation matrix may be determined by solving an exact set of equations so that only the zero- and unity-valued entries of the rotated and $\mathbf{A}_D^\star$ match, which we term as the first approach. Alternatively, an overdetermined set of equations may be solved by also forcing the equality of non-zero coefficients in the shifted rows. Intuitively, the second approach can be expected to yield more efficient estimates of the model coefficients. Regardless of the approach, the same rotation matrix is applied to both the sub-blocks, $\hat{\mathbf{A}}_D$ and $\hat{\mathbf{A}}_I$ of the constraint matrix. For the example under discussion, the rotation matrix contains $2\times 2 = 4$ unknowns. With the first approach, we would set up four equations to obtain an exact solution, while with the second approach, five overdetermined equations would be used in estimating these elements.

Finally, the difference equation model is obtained by averaging the rows of the respective rotated matrices, $\mathbf{A}_D^\star$ and $\hat{\mathbf{A}}_I^\star$. Table 1 summarizes the algorithm to recover the true matrix $\mathbf{A}_0$

Remark: Essentially, we have solved a structured TLS problem, where the structure (for the constraint matrix) is also derived from the data using IPCA.

Table 1. Proposed DIPCA algorithm

(1) Construct the data matrix $\mathbf{Z}$ by choosing a sufficiently large L.
(2) Identify the number of linear relations d and the constraint matrix $\hat{\mathbf{A}}$ underpinning the error-free part of $\mathbf{Z}$ through IPCA. Also obtain the estimate of $\Sigma_\mathbf{e}$.
(3) Determine the process order using the relation $d = L - \max(n_y, n_u) + 1$
(4) Partition the constraint matrix into sub-matrices corresponding to dependent (output) and independent (input) variables as follows

$$\hat{\mathbf{A}} = \begin{bmatrix}\hat{\mathbf{A}}_D & \hat{\mathbf{A}}_I\end{bmatrix} \quad \hat{\mathbf{A}}_D \in \mathbb{R}^{d\times(L+1)},\ \hat{\mathbf{A}}_I \in \mathbb{R}^{d\times(L+1)} \tag{20}$$

so that

$$\hat{\mathbf{A}}_D \begin{bmatrix} y[k] \\ \vdots \\ y[k-L]\end{bmatrix} = -\hat{\mathbf{A}}_I \begin{bmatrix} u[k] \\ \vdots \\ u[k-L]\end{bmatrix} \tag{21}$$

(5) Using information in Step 3 and the shift property of lagged constraints, determine the location of unity- and zero-valued entries in $\mathbf{A}_0$. Denote this structured matrix as $\mathbf{A}_D^\star$
(6) Determine the rotation matrix $\mathbf{R}_D$ so that $\mathbf{R}_D\hat{\mathbf{A}}_D$ structurally matches $\mathbf{A}_D^\star$, preferably using the overdetermined approach.
(7) Finally, obtain the difference equation model by averaging the rows of $\mathbf{R}_D\hat{\mathbf{A}}_D$ and $\mathbf{R}_D\hat{\mathbf{A}}_I$.

Remark: The structure and rotation matrix determination in Steps 5 and 6, respectively, of the proposed algorithm are equipped to accommodate any prior information on the input order, delay, etc.

Estimating $\Sigma_\mathbf{e}$ using constrained optimization The generic formulation of IPCA, as presented in Section 2.2, assumes that all elements of $\Sigma_\mathbf{e}$ are distinct. However, when applied to the dynamic SISO EIV identification problem, it is known a priori that for the matrix $\mathbf{Z}$ constructed from L lagged input and output measurements, only two of the $2(L+1)$ elements of $\Sigma_\mathbf{e}$ are distinct. Consequently, the lower limit of the identifiability constraint in (11) is always two. Furthermore, the optimization problem for estimating $\Sigma_\mathbf{e}$ in (10) is modified to reflect this constraint as follows:

$$\min_{\Sigma_\mathbf{e}} \quad N\log|\hat{A^{(i)}}\Sigma_\mathbf{e}(\hat{A^{(i)}})^T| + \sum_{i=1}^{N}(\mathbf{r}_i^T[k](\hat{A^{(i)}}\Sigma_\mathbf{e}(\hat{A^{(i)}})^T)^{-1}\mathbf{r}_i[k]) \tag{22}$$

$$\text{subject to} \quad \text{diag}(\Sigma_\mathbf{e}) = \begin{bmatrix}\sigma_y^2\mathbf{I}_{L+1} & \sigma_u^2\mathbf{I}_{L+1}\end{bmatrix} \tag{23}$$

where $\mathbf{I}_{L+1}$ is the $(L+1)\times(L+1)$ identity matrix.

The next section presents results from simulation studies to demonstrate the effectiveness of the proposed algorithm.

4. SIMULATION RESULTS

The case study pertains to the second-order example in (13). $N = 1000$ observations of the data were generated with unequal error variances of $\sigma_y^2 = 0.2406$ and $\sigma_u^2 = 0.091$ so as to achieve an SNR of 10.

The data is run through the proposed algorithm in Table 1 with $L = 3$. Two unity eigenvalues were identified by the IPCA algorithm, as reported below.

$$\Lambda = \text{diag}([11.6 \quad 11 \quad 8.9 \quad 8.09 \quad 4.7 \quad 1.92 \quad 1.0037 \quad 0.9963])$$

Further, $\hat{\Sigma}_\mathbf{e} = \text{diag}([0.2363 \quad 0.1087])$, which closely matches with the true value.

From the relation given in Step 3, the order is estimated to be $n_y = 3 - 2 + 1 = 2$, thus identifying the order correctly.

The 2×8 constraint matrix is estimated and partitioned into dependent and independent matrices. The structure of $\mathbf{A}_D^\star$ is subsequently determined to be,

$$\mathbf{A}_D^\star = \begin{bmatrix} 1 & \times & \times & 0 \\ 0 & 1 & \times & \times \end{bmatrix} \tag{24}$$

To determine the rotation matrix, both approaches are implemented. The first (exact) approach yields

$$\mathbf{R}_D = \begin{bmatrix} -3.240 & -22.7 \\ -22.721 & -3.564 \end{bmatrix} \mathbf{R}_D\mathbf{A}_D = \begin{bmatrix} 1 & 0.420 & 0.572 & 0 \\ 0 & 1 & 0.404 & 0.595 \end{bmatrix}$$

$$\mathbf{R}_D\mathbf{A}_I = \begin{bmatrix} 0.054 & 1.231 & -0.019 & -0.056 \\ 0.046 & 0.053 & 1.2321 & -0.038 \end{bmatrix}$$

From the above, the estimated difference equation model is obtained using the last step of Table 1:

$$y[k] + 0.412y[k-1] + 0.583y[k-2] = 0.053u[k] + 1.232u[k-1] - 0.029u[k-2] - 0.056u[k-3]$$

Using the second approach (over-determined LS),

$$\mathbf{R}_D = \begin{bmatrix} -3.068 & -22.798 \\ -22.529 & -3.716 \end{bmatrix} \mathbf{R}_D\mathbf{A}_D = \begin{bmatrix} 1 & 0.412 & 0.567 & -0.005 \\ 0.008 & 1 & 0.407 & 0.592 \end{bmatrix}$$

$$\mathbf{R}_D\mathbf{A}_I = \begin{bmatrix} 0.053 & 1.231 & -0.029 & -0.056 \\ 0.047 & 0.063 & 1.228 & -0.038 \end{bmatrix}$$

yielding the difference equation as

$$y[k] + 0.409y[k-1] + 0.581y[k-2] - 0.005y[k-3] = 0.058u[k] + 1.229u[k-1] - 0.034u[k-2] + 0.056u[k-3]$$

The estimates obtained with both approaches are satisfactorily close to the true values used in simulation.

In order to demonstrate that the extra coefficients appearing in both model estimates are negligibly small and also to show that the remaining coefficients are unbiased and significant, Monte Carlo simulations were carried out. Results from 200 runs for lag order 3 are presented in Table 2. Results suggest that both approaches yield estimates with similar error characteristics, although it

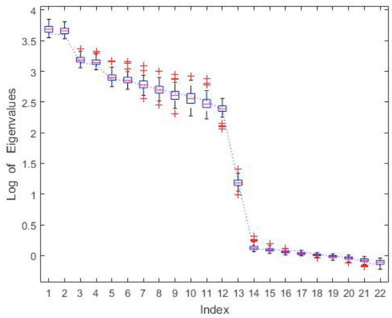

Fig. 1. Plot of $\log_2 \lambda$ with error bounds

must be said that Approach 2 is capable of accommodating prior information (of input order, delay) in a simple way.

The results for estimates of noise variances are reported in Table 3. From the results, it can be safely said that the noise variance estimates are unbiased and of low error.

Monte Carlo Simulation results of 200 runs for lag order $L = 10$ are reported in Table 4 and Table 5. From these results, it is clear that stacking in excess yields improved estimates of parameters.

Figure 1 shows the eigenvalues obtained from IPCA with the error bounds computed from MC simulations. The expected number of unit eigenvalues, nine in this case, remain within bounds.

5. CONCLUSIONS

In this paper, we have presented a systematic method for recovering the dynamic model of a linear multivariable open-loop process from measurements of inputs and outputs in the EIV case. The method is based on an extension of the iterative PCA, which was originally formulated for static systems. Using the ideas of dynamic PCA and a systematic exploitation of the information given by IPCA, we developed a method for determining the structure of the true constraint matrices. The result is a consistent and efficient method for discovering the dynamic model purely from data with minimal user intervention. Future work consists of extending the proposed method to closed-loop multivariable systems.

Table 2. MC simulation results for L = 3

		Approach 1		Approach 2	
Para.	True Value	mean	std. dev.	mean	std. dev.
a_1	0.4	0.4029	0.0192	0.4028	0.0193
a_2	0.6	0.6001	0.0175	0.60	0.0187
a_3	0	0*	0*	-0.0002	0.0026
b_0	0	0.0018	0.0228	0.0021	0.0230
b_1	1.2	1.2031	0.0390	1.2030	0.0386
b_2	0	0.0041	0.0374	0.0037	0.0378
b_3	0	-0.0033	0.0446	-0.0033	0.0446

Table 3. μ and σ of $\hat{\sigma}^2_{e_i}$ for lag $L = 3$

Parameters	True Value	mean	std. deviation
σ_y^2	0.2406	0.2456	0.0405
σ_u^2	0.0910	0.0979	0.0280

Table 4. μ and σ of $\hat{\sigma}^2_{e_i}$ for lag $L = 10$

Parameters	True Value	mean	std. deviation
σ_y^2	0.2406	0.2402	0.0273
σ_u^2	0.0910	0.0979	0.0154

Table 5. μ and σ of estimates for $L = 10$

		Approach 1		Approach 2	
Parameters	True Value	mean	std. dev.	mean	std. dev.
a_1	0.4	0.3979	0.0456	0.3990	0.0251
a_2	0.6	0.6049	0.0356	0.6015	0.0221
b_0	0	0.0014	0.0225	0.0015	0.0225
b_1	1.2	1.2022	0.0295	1.2022	0.0294

REFERENCES

Chou, C. and Verhaegen, M. (1997). Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, 33(10), 1857–1869.

Huang, B. (2001). Process identification based on last principal component analysis. *Journal of Process Control*, 11, 19–33.

Huffel, S.V. and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM, Philadelphia, PA, USA.

Ku, W., Storer, R., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30, 179–196.

Li, W. and Qin, S.J. (2001). Consistent dynamic PCA based on errors-in-variables subspace identification. *Journal of Process Control*, 11(6), 661–676.

Ljung, L. (1999). *System Identification - A Theory for the User*. Prentice Hall International, Upper Saddle River, NJ, USA.

Narasimhan, S. and Shah, S. (2008). Model identification and error covariance matrix estimation from noisy data using PCA. *Control Engineering Practice*, 16, 146–155.

Pearson, K. (1902). On the systematic fitting of curves to observations and measurements, part 1. *Biometrika*, 1, 265–303.

Soderstrom, T. (2007). Errors-in-variables methods in system identification. *Automatica*, 43(6), 939–958.

Vijaysai, P., Gudi, R., and Lakshminarayanan, S. (2005). Errors-in-variables-based modeling using augmented principal components. *Industrial & Engineering Chemistry Research*, 44, 368–380.

Appendix A. PCA FOR STATIC IDENTIFICATION

A basis for the constraint matrix $\mathbf{A}_0$, call it $\mathbf{A}$, and hence its row dimension d are given by the basis and dimension of the null space of $\mathbf{X}$, respectively. These can be obtained either through an SVD of $\mathbf{X}$ or via an eigenvalue analysis of the sample covariance matrix $\mathbf{S_x} \triangleq \frac{1}{N}\mathbf{X}^T\mathbf{X}$ as follows[2]:

$$\mathbf{S_x V} = \mathbf{V}\Lambda, \quad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_{M-d} & \\ & & & \mathbf{0} \end{bmatrix}, \lambda_1 > \cdots > \lambda_{M-d} \tag{A.1}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_M \end{bmatrix}. \quad \mathbf{A} = \begin{bmatrix} \mathbf{v}_{M-d+1} & \cdots & \mathbf{v}_M \end{bmatrix}^T \tag{A.2}$$

The projections of columns of $\mathbf{X}$ onto the basis vectors $\mathbf{V}$ collected in $\mathbf{P} = \mathbf{XV}$ are termed as the *principal components* of $\mathbf{X}$. They satisfy $\mathbf{p}_i^T\mathbf{p}_j = 0$, $\forall i \neq j = 1, \cdots, M$.

Note that $\mathbf{A}$ and $\mathbf{A}_0$ differ by a similarity transformation $\mathbf{T}$ s.t. $\mathbf{T}^T\mathbf{T} = \mathbf{I}$. In general, this is not an issue and is completely alleviated in regression since the variable set can be partitioned into d dependent and $(M - d)$ independent variables and the regression coefficients can be uniquely recovered as follows:

$$\mathbf{x}[k] = \begin{bmatrix} \mathbf{x}_I[k] \\ \mathbf{x}_D[k] \end{bmatrix}^T, \qquad \mathbf{A} = \begin{bmatrix} \mathbf{A}_D & \mathbf{A}_I \end{bmatrix} \tag{A.3}$$

$$\implies \mathbf{x}_D[k] = \underbrace{-\mathbf{A}_D^{-1}\mathbf{A}_I}_{\mathbf{B}}\,\mathbf{x}_I[k], \qquad \mathbf{A}_D \in \mathbb{R}^{d\times d} \tag{A.4}$$

The coefficient matrix $\mathbf{B}$ is invariant to a similarity transformation of $\mathbf{A}$, i.e., both $\mathbf{A}$ and $\mathbf{TA}$ yield the same $\mathbf{B}$.

[2] It is a common practice to normalize columns of $\mathbf{X}$ to account for disparities in the units of variables.