Subspace Identification of Linear Systems with Observation Outliers*

Hideyuki Tanaka†, Jaafar AlMutawa† and Tohru Katayama†

In this paper, we consider a subspace identification method for linear stochastic systems subject to observation outliers, where the observation noise contains large values with a low probability. We derive a subspace identification method by combining the orthogonal decomposition-based subspace identification method (ORT-method) and a weighted LQ decomposition. We apply the ORT-method to the input-output data, coupled with the standard LQ decomposition to obtain residuals of the output sequence. By using the median of residuals, outliers are detected by a simple scheme in robust statistics. Based on detected outliers, a weighting matrix is generated automatically, and is incorporated in the weighted LQ decomposition to get an improved estimate of the system matrices. A numerical example is included to show the effectiveness of the proposed method.

1. Introduction

Subspace identification methods have extensively been studied since early 1990s (e.g. MOESP by Verhaegen and Dewilde[19], N4SID by Van Overschee and De Moor[18]). The subspace methods compute signal subspace of data matrices via LQ decomposition and singular value decomposition (SVD), and identify state space models by using shift-invariance property of the block Hankel matrix[17]. A great advantage is that the difficult model selection problem inherent in the classical parameter optimization-based approaches to MIMO identification[9] is completely avoided, except for the model order determination.

A subspace identification method based on orthogonal decomposition (ORT-method) has been developed in Refs. [12,13], where the realization of stochastic systems with exogenous inputs and its adaptation to the subspace methods have been considered by decomposing the output vector into deterministic and stochastic components. It may be noted that numerical results in Refs. [1,6] show that in the case where the exogenous input is colored, the ORT-method is more robust than the N4SID-like algorithms[18,20]. It is well known that in real world, there are cases where large errors are contained in observed data with a low probability, so that a standard Gaussian assumption for observation noises may fail. The least-squares estimate is quite sensitive to this type of non-Gaussian disturbances called outliers. In this sense, the performance of the ORT-method degrades since it decomposes the output sequence in the least-squares sense into the deterministic and stochastic components. Thus, it will be an important topic to develop a simple method to cope with outliers when we use the ORT-method. Though some system identification problems subject to missing data have been studied in Refs. [3,4] for ARX model setting, and in Ref. [14] for a frequency domain identification, no subspace methods of state space identification in the presence of outliers or missing data have been studied in the literature to our knowledge.

In this paper, we consider a subspace identification method for linear stochastic systems subject to observation outliers, where the observation noise takes on large values with a low probability. We derive a subspace identification method by combining the ORT-method and a weighted LQ decomposition. We develop a method of detecting outliers in the observed output sequence by using the ORT-method and a sim-
ple scheme for robust estimation methods in statistics[11]. An earlier version of this paper, based on the MOESP method, is presented at the IFAC workshop[16].

This paper is organized as follows. In section 2, we state the problem of subspace identification in the presence of outliers. In section 3, some results of the ORT-method are reviewed. Section 4 considers a weighted orthogonal decomposition and gives an algorithm for computing the weighted LQ decomposition. In section 5, we present a method of detecting outliers and construct a weighting matrix. Section 6 gives a subspace identification method based on the ORT-method coupled with the weighted LQ decomposition. Section 7 includes a numerical simulation result. We conclude the paper in section 8. The basic subspace identification algorithm and proofs of theorems are deferred in Appendix.

2. Problem Statement and Orthogonal Decomposition

2.1 System Model and Problem Statement

Consider the discrete-time stochastic system described by

\[
\begin{bmatrix}
    x_{t+1} \\
    y_t
\end{bmatrix} = \begin{bmatrix}
    A & B \\
    C & D
\end{bmatrix} \begin{bmatrix}
    x_t \\
    u_t
\end{bmatrix} + \begin{bmatrix}
    w_t \\
    v_t
\end{bmatrix}
\]

(1)

where \(x_t \in \mathbb{R}^n\) is the state vector and \(u_t \in \mathbb{R}^m\) is the exogenous input vector, \(y_t \in \mathbb{R}^p\) is the output observation vector, and \(w_t \in \mathbb{R}^n\) is a white Gaussian noise. It is assumed that \((A, B, C, D)\) are constant matrices with appropriate dimensions. Also \(v_t \in \mathbb{R}^p\) is a white noise that contains outliers with a low probability:

\[
v_t = (1 - \alpha_t)v_t^0 + \alpha_t v_t^0,
\]

(2)

where \(\alpha_t = \{0, 1\}\) and \(\text{Prob}\{\alpha_t = 1\}\) is small, and where \(v_t^0\) and \(v_t^0\) are Gaussian white noises with

\[
v_t^0 \in \mathcal{N}(0, \sigma_0^2), \quad v_t^0 \in \mathcal{N}(0, \sigma_0^2).
\]

(3)

It is assumed that \(\sigma_n^2\) is much larger than \(\sigma_0^2\), or \(\sigma_n^2 \ll \sigma_0^2\). We also assume that \(w_t, v_t^0\) and \(v_t^0\) are zero mean white Gaussian noise vectors with finite variances.

Under these conditions, the output equation is written as (see Fig. 1)

\[
y_t = (1 - \alpha_t)y_t^0 + \alpha_t y_t^0,
\]

where \(y_t^0\) and \(y_t^0\) are defined by

\[
y_t^0 := Cx_t + Du_t + v_t^0
\]

\[
y_t^0 := Cx_t + Du_t + v_t^0.
\]

In the following, \(y_t^0\) is called the normal output, whereas \(y_t^0\) the outliers output. It should be, however, noted that \(y_t^0\) and \(y_t^0\) are fictitious outputs never observed.

Let \(\tau\) be a positive integer, and \(\nu\) be sufficiently large. Then, the problem to be considered in this paper is stated as follows: Given the observed data sequence

\[
\{u_t, y_t : t = 1, \ldots, 2\tau + \nu - 1\}
\]

where the output \(y_t\) contains observation outliers, we identify the system matrices \((A, B, C, D)\) within the freedom of equivalent transforms by using the ORT-method.

2.2 Assumptions

Let \(\mathcal{U}, \mathcal{Y}^m\) and \(\mathcal{Y}^o\) denote vector spaces obtained by taking all finite linear combinations of the input \(\{\sum \alpha_j u_{t_j} \mid \alpha_j \in \mathbb{R}^m, j, t_j \in \mathbb{Z}\}\), the normal output \(\{\sum \alpha_j y_{nt_j} \mid \alpha_j \in \mathbb{R}^p, j, t_j \in \mathbb{Z}\}\) and the outliers output \(\{\sum \alpha_j y_{ot_j} \mid \alpha_j \in \mathbb{R}^p, j, t_j \in \mathbb{Z}\}\), respectively[12,5]. And we define \(\mathcal{U} \cup \mathcal{Y}^m \cup \mathcal{Y}^o\) by all finite linear combinations of the input and the normal and outliers outputs:

\[
\mathcal{U} \cup \mathcal{Y}^m \cup \mathcal{Y}^o = \{\alpha_j y_{nt_j} \mid \alpha_j \in \mathbb{R}^{m+2p}, j, t_j \in \mathbb{Z}\}.
\]

Then, we obtain a Hilbert space, by closing the vector space \(\mathcal{U} \cup \mathcal{Y}^m \cup \mathcal{Y}^o\) with respect to the norm induced by the scalar product \(\langle \xi, \zeta \rangle = E\{\xi \zeta\}\) for \(\xi, \zeta \in \mathcal{U} \cup \mathcal{Y}^m \cup \mathcal{Y}^o\) where \(E\{\cdot\}\) denotes mathematical expectation. The orthogonal projection onto the subspace \(\mathcal{A}\) is denoted by the symbol \(\mathbb{E}\{\cdot \mid \mathcal{A}\}\).

The spaces \(\mathcal{U}, \mathcal{Y}^m\) and \(\mathcal{Y}^o\) are commonly denoted by

\[
\mathcal{U} = \text{span}\{u_x\mid -\infty < \sigma < \infty\},
\]

\[
\mathcal{Y}^m = \text{span}\{y_x^m\mid -\infty < \sigma < \infty\},
\]

\[
\mathcal{Y}^o = \text{span}\{y_x^o\mid -\infty < \sigma < \infty\}.
\]

We define the linear space of second order random variables spanned by the infinite past and future of the input vector at the present time \(t\) as:

\[
\mathcal{U}_t := \text{span}\{u_x \mid \sigma < t\},
\]

\[
\mathcal{U}_t^+ := \text{span}\{u_x \mid \sigma \geq t\}.
\]

Also \(\mathcal{Y}_t^m, \mathcal{Y}_t^m +, \mathcal{Y}_t^o, \mathcal{Y}_t^o +\) are similarly defined by \(y_x^m\) and \(y_x^o\), respectively.

The following assumptions are made throughout the paper.

(A) There is no feedback from \(y_x^m\) and \(y_x^o\) to the input \(u_t\), which is equivalent to the fact that the future of \(u\) is conditionally uncorrelated with the past of \(y_x^m\) and \(y_x^o\) given the past of \(u\); this can

![Fig. 1 Setting for identification](image-url)
be written as Ref. [12]
\[ U_i^+ \perp \gamma_i^{(n)} \parallel U_i, \]
\[ U_i^+ \perp \gamma_i^{(o)} \parallel U_i, \]
where the notation \( A \perp B | X \) means that \( A \) and \( B \) are conditionally orthogonal given a third subspace \( X \), i.e. \( \langle \xi - \hat{E}\{\xi|X\}, \zeta - \hat{E}\{\zeta|X\} \rangle = 0 \) for \( \xi \in A, \zeta \in B \).

Consequently, there is no feedback from \( y \) to \( u \).

(B) The input \( u \) satisfies the sufficiently high PE condition. This condition is satisfied if the spectral density matrix \( \Phi_{uu}(z) \) of \( u \) is positive definite on the unit circle \( |z| = 1 \).[15, 8]

3. Orthogonal Projection

In this section we review a realization based on the orthogonal decomposition method (ORT-method) under the assumption that there are no outliers[12], i.e. \( \text{Prob}\{\alpha_t = 1\} = 0 \) or \( y_t = y_t^p \) in (2).

3.1 Orthogonal Decomposition by Infinite Data

When \( z \) is vector valued, we use the notation \( E\{z|A\} \):
\[ E\{z|A\} = \begin{bmatrix} E\{z_1|A\} \\ E\{z_2|A\} \\ \vdots \\ E\{z_p|A\} \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{bmatrix} \]
for \( z_i \in U \cup \gamma^m \cup \gamma^o \).

Under the assumption (B), it can be shown[12] that
\[ E\{y_t^n|U\} = E\{y_t^o|U_{t+1}\}. \]

This implies that if there is no feedback from \( y^n \) to \( u \), then the estimate \( E\{y_t^o|U\} \) is causal, so that it can be expressed as a linear combination of the present and past inputs.

Now, define the projection as
\[ \hat{y}_t^n = E\{y_t^n|U_{t+1}\} = E\{y_t^n|U\}. \tag{4} \]
We see that \( \hat{y}_t^n \) is the part of \( y_t^n \) that is linearly related to the past inputs; thus \( \hat{y}_t^n \) is expressed as a moving average of \( u_t, u_{t+1}, \ldots \). Hence, \( \hat{y}_t^n \) is called the deterministic component of \( y_t^n \). The orthogonal projection of \( y_t^n \) onto the complementary space \( U^\perp \) is given by
\[ \hat{y}_t^n = y_t^n - \hat{E}\{y_t^n|U_{t+1}\} = y_t^n - \hat{E}\{y_t^n|U\} \tag{5} \]
where the residual \( \hat{y}_t^n \) is called the stochastic component.

[Proposition 1] [12] Under the assumption (A), the normal output process \( y^n \) has the orthogonal decomposition
\[ y_t^n = \hat{y}_t^n + \tilde{y}_t^n \tag{6} \]
where \( \hat{y}_t^n \) and \( \tilde{y}_t^n \) are mutually uncorrelated, i.e.
\[ E\{\hat{y}_i^n(\tilde{y}_j^n)^\top\} = 0, \forall i, j \in \mathbb{Z}. \]

It has been shown[12] that the system matrices \( (A, B, C, D) \) are obtained from the deterministic component \( \hat{y}_t^n \).

3.2 Orthogonal Projection by Finite Data

Let the orthogonal projection of \( X \in \mathbb{R}^{n \times p} \) onto the row space of a full row rank matrix \( U \in \mathbb{R}^{n \times p} \) be given by
\[ \hat{E}\{X|U\} := XU^\top(UU^\top)^{-1}U, \]
where the projection is calculated by the standard LQ decomposition[18].

Define input and output matrices as
\[ u_t = \begin{bmatrix} u_t \\ u_{t+1} \\ \vdots \\ u_{t+p-1} \end{bmatrix} \in \mathbb{R}^{n \times p}, \tag{7} \]
\[ y_t^n = \begin{bmatrix} y_t^n \\ y_{t+1}^n \\ \vdots \\ y_{t+p-1}^n \end{bmatrix} \in \mathbb{R}^{p \times p}. \tag{8} \]

Also, define stacked matrices as
\[ Y_{1|2r}^n = \begin{bmatrix} y_1^n \\ \vdots \\ y_{2r}^n \end{bmatrix}, \quad U_{1|2r} := \begin{bmatrix} u_1 \\ \vdots \\ u_{2r} \end{bmatrix}. \tag{9} \]

According to (4), we calculate projections
\[ \hat{y}_t^n := \hat{E}\{y_t^n|U_{1|2r}\}, \quad t = 1, \ldots, 2r. \]

The stochastic component of the normal output is estimated by
\[ \tilde{y}_t^n := y_t^n - \hat{E}\{y_t^n|U_{1|2r}\}, \tag{10} \]
so that \( \hat{y}_t^n(\tilde{y}_t^n)^\top = 0 \) holds. Since \( \hat{y}_t^n \) is the estimation of the noise which is generated by \( w_t \) and \( v_t^n \), we see that the elements of \( \hat{y}_t^n \) are of Gaussian distribution.

4. Weighted Orthogonal Projection

In order to consider outliers in the framework of subspace identification method, we define an inner product and define a weighted least-squares (WLS) problem.

For a nonnegative definite \( \Theta \in \mathbb{R}^{r \times r} \), define a product of two matrices \( X \in \mathbb{R}^{p \times n} \) and \( Y \in \mathbb{R}^{q \times n} \) as
\[ \langle X, Y \rangle_\Theta = X \Theta Y^\top. \tag{11} \]

Then, we define a quadratic form
\[ \Phi(\hat{Y}) := \text{Tr}(Y - \hat{Y}, Y - \hat{Y})_\Theta \]
where \( Y, \hat{Y} \in \mathbb{R}^{q \times r} \).

[WLS problem] Find \( \hat{Y} \), satisfying
\[ \hat{Y} = \arg\min_{\hat{Y} \in \mathcal{U}} \Phi(\hat{Y}) \tag{12} \]
where \( \mathcal{U} := \text{row space}(U) = \text{span}\{U(i,:)|i = 1, \ldots, q\} \), and where \( \langle U, U \rangle_\Theta > 0 \) holds with \( U \in \mathbb{R}^{q \times r} \).

Suppose that matrices \( Y \in \mathbb{R}^{p \times n} \) and \( U \in \mathbb{R}^{q \times n} \) are given, and that \( \mathcal{U} \) is the space spanned by the row
vectors of $U$. If there exits a $P \in \mathbb{R}^{p \times q}$ such that
\[ Y = \hat{Y} + \hat{Y}, \quad \hat{Y} = PU, \quad (\hat{Y}, U)\Theta = 0, \quad (13) \]
then $\hat{Y}$ is the weighted orthogonal projection of $Y$ onto $U$, which is also denoted by $\hat{E}_\Theta(Y|U)$.

**Lemma 1.** Suppose that matrices $\Theta \geq 0, U \in \mathbb{R}^{q \times n}$ and $Y \in \mathbb{R}^{p \times n}$ are given. Then, there always exists a $\hat{Y}$ that minimizes $\Phi(\hat{Y})$. Furthermore, if $(U, U)\Theta$ is non-singular, there exists a unique minimizer $\hat{Y}$ to $\Phi(\hat{Y})$, which is given by $\hat{Y} = \hat{E}_\Theta(Y|U)$.

*(Proof)* See Appendix 1.

**Theorem 1.** Let $(U, U)\Theta \in \mathbb{R}^{q \times q}$ be non-singular. Suppose $[U^T \quad Y^T]^T$ is decomposed as
\[ [U^T \quad Y^T]^T = [Q^T_u \quad Q^T_y] \Theta [I \quad 0 \quad J], \quad (14) \]
where $Q_u$ and $Q_y$ satisfy
\[ \begin{bmatrix} Q^T_u \\ Q^T_y \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix}, \quad \Theta = \begin{bmatrix} I & 0 \\ 0 & J \end{bmatrix}. \quad (15) \]
where $J = \text{block-diag}(I, 0)$ and $L_{uu} \in \mathbb{R}^{q \times q}$. Then, it follows that $L_{uu}$ is non-singular, and
\[ \hat{Y} = \hat{E}_\Theta(Y|U) = (Y, U)\Theta(U, U)\Theta^{-1}U = L_{yy}Q_u^T = L_{yy}L_{uu}^{-1}U. \quad (16) \]

*(Proof)* See Appendix 2.

We present an algorithm for “WLQ” (weighted LQ) decomposition in (14) and (15), where $\Theta$ is a singular matrix and $(U, U)\Theta > 0$ holds.

Compute the eigenvalue decomposition of $\Theta$:
\[ \Theta = V \begin{bmatrix} \Theta_1 & 0 \\ 0 & 0 \end{bmatrix} V^T, \quad V \in \mathbb{R}^{p \times p} \quad (17) \]
where $\Theta_1 = \Theta_1^{1/2} \Theta_1^{1/2} > 0, V^T V = I, \Theta_1 \in \mathbb{R}^{n_1 \times n_1}$ and $n = n_1 + n_2$. Define matrices $G_u$ and $G_y$ as
\[ [G_u^T \quad G_y^T]^T = [U^T \quad Y^T] V \begin{bmatrix} \Theta_1^{1/2} \\ 0 \end{bmatrix}_{n_2 \times n_1}, \quad (18) \]
Compute the LQ decomposition of $[G_u^T \quad G_y^T]^T$:
\[ [G_u^T \quad G_y^T]^T = \begin{bmatrix} A_{uu} & A_{uy} \\ A_{yu} & A_{yy} \end{bmatrix} \begin{bmatrix} Q_u^T \\ Q_y^T \end{bmatrix}. \quad (19) \]

**Theorem 2.** From (17)-(19), we have
\[ \hat{E}_\Theta(Y|U) = A_{yu}A_{uu}^{-1}U. \quad (20) \]

*(Proof)* See Appendix 3.

### 5. Detection of Outliers and an Associated Weighting Matrix

In this section, we give a method of detecting outliers and of disregarding the data subject to outliers. To this end, we employ an idea from robust estimation(11).

Define the output data Hankel matrices
\[ Y_{1|2r} = \begin{bmatrix} y_1 \\ \vdots \\ y_{r+1} \\ y_r \end{bmatrix}, \quad Y_{r+1|2r} = \begin{bmatrix} y_{r+2} \\ \vdots \\ y_{2r} \end{bmatrix}, \quad (21) \]
and $Y_{1|2r} = [Y_{1|2r}, Y_{r+1|2r}]^T$.

#### Detection of outliers

**Step 1** Define the following matrix from the output data
\[ y_1 = [y_1, y_2, \cdots, y_{2r}] \in \mathbb{R}^{p \times n}. \]

Projecting $y_{2r}$ onto the row space of block Hankel matrix of the input data yields
\[ \hat{y}_{2r} = \hat{E}_\Theta(y_{2r}|U_{1|2r}). \]

**Step 2** Calculate the difference of the actual and the projected output data matrix as
\[ \hat{y}_{2r} = y_{2r} - \hat{y}_{2r}. \]

The elements of $\hat{y}_{2r}$ are expressed as
\[ \hat{y}_{2r} = [\hat{y}_{2r}, \hat{y}_{2r+1}, \cdots, \hat{y}_{2r+n-1}], \quad (21) \]
and are called residuals which are the estimates of stochastic components; see (10). It should be here noted that they are only rough approximation due to possible existence of outliers.

**Step 3** Obtain the median of the residuals
\[ s = \text{median}(\|\hat{y}_t\| : t = 2r, \cdots, 2r + n - 1). \quad (22) \]

According to a robust estimation method, if $\|\hat{y}_t\| > cs$, we say that $\|\hat{y}_t\|$ contains an outlier with high probability, where $c = 5 \sim 9^2$.

In order to give an idea to construct a weighting matrix, we assume for a while that an outlier exists in $y_t$ only. In this case, due to its block Hankel structure, the matrix $Y_{1|2r}$ is written as
\[ Y_{1|2r} = \begin{bmatrix} y_1 & \cdots & y_{r+1} \\ \vdots & \ddots & \vdots \\ y_{2r} & \cdots & y_{2r+n-1} \end{bmatrix}. \]

\[ \hat{y}_{2r} = \hat{E}_\Theta(y_{2r}|U_{1|2r}). \]

\[ w(\hat{y}_t) = \begin{cases} (1 - (\hat{y}_t/cs)^2)^2 & \text{for } \|\hat{y}_t\| < cs \\ 0 & \text{for } \|\hat{y}_t\| \geq cs \end{cases} \]
where $c = 5 \sim 9$ is recommended(11). Here we employed a simple cut method with the threshold $cs$, where the actual value of $c$ is borrowed from the biweight method.
Thus, we must delete all the data from \((i-2T+1)\)-th to \(\min\{i,\nu\}\)-th columns from \(Y_{1|2T}\). Hence, if \(\Theta\) is defined as
\[
\Theta = \text{diag}(1, \ldots, 1, 0, 1, \ldots, 1),
\]
then, we see that \(Y_{1|2T} = Y_{1|2T}^n\Theta\) holds where \(Y_{1|2T}^n\) is defined in (9), implying that \(Y_{1|2T}\Theta\) is free from outliers. Thus, we can estimate the deterministic component based on the smoothed data \(Y_{1|2T}\Theta\).

Construction of a weighting matrix

We now construct the following diagonal matrix
\[
\Theta = \text{diag}(\theta_1, \theta_2, \ldots, \theta_v) \in \mathbb{R}^{v \times v}
\] (23)

where \(\theta_i \in \mathbb{R}\).

Step 1 Let \(D\) be a set of indices \(i\) of detected outliers based on the median of (22) for \(2T \leq i \leq 2T + \nu - 1\), or \(D = \{i|y_i \text{ has outliers} \}\). Since \(Y_{1|2T}\) is a Hankel matrix, each \(y_i\) appears \(2T\) times in it for \(2T \leq i \leq \nu\). We therefore define another index set as
\[
I = \{j|y_i \in D, \quad i \leq j \leq \min\{i+2T-1, \nu+2T-1\}\},
\]
where \(I\) contains all the information on the deleted columns in the output data matrix \(Y_{1|2T}\), and \(D \subseteq I\) holds.

Step 2 According to the detection scheme based on the median of (22), we define the weighting matrix \(\Theta\) by setting \(\theta_j\) \((1 \leq j \leq \nu)\) as
\[
\theta_j = \begin{cases} 
0 & \text{for } j+2T-1 \in I, \\
1 & \text{for } j+2T-1 \notin I.
\end{cases}
\] (25)

6. Subspace Identification Method with Weighting Matrix

In this section, we present a subspace method of identifying the system matrices in the presence of outliers. Define the stacked matrix as
\[
\begin{bmatrix} 
Y_{r+1|2T} \\
Y_{1|2T} \\
Y_{r+1|2T}
\end{bmatrix} := \tilde{E}_\Theta(Y_{r+1|2T} | U_{1|2T})
\] (26)

where \(\tilde{Y}_{r+1|2T}\) is the deterministic component obtained by attenuating noise by the weighted orthogonal projection with a given \(\Theta\).

Taking the weighted orthogonal projection of \(Y_{r+1|2T}\) onto the data space \(U_{1|2T}\), we have approximately the following basic equation[1,5]
\[
\hat{Y}_{r+1|2T} = \mathcal{O}_r \hat{X}_r + \Psi_r U_{r+1|2T}
\] (27)

where \(\hat{X}_r\) is the weighted orthogonal projection of the state vector \(X_r := [x_{r+1} \ x_{r+2} \ \cdots \ x_{r+r}]\) onto the input data space \(\text{span}(U_{1|2T})\), and \(\mathcal{O}_r\) is the extended observability matrix defined by
\[
\mathcal{O}_r := \begin{bmatrix} 
C \\
CA \\
\vdots \\
CA^{r-1}
\end{bmatrix} \in \mathbb{R}^{p \times n},
\]
and where \(\Psi_r\) is the lower triangular Toeplitz matrix defined by
\[
\Psi_r := \begin{bmatrix} 
D & CB & D \\
CAB & CB & D \\
\vdots & \ddots & \ddots \\
CA^{r-2}B & \cdots & CB & D
\end{bmatrix} \in \mathbb{R}^{p \times rm}.
\]

For the basic equation of (27), we assume that the well-known rank conditions are satisfied[10]:
(C-1) \(\text{rank}(X_r) = n\),
(C-2) \(\text{span}\{X_r\} \cap \text{span}\{U_{r+1|2T}\} = \emptyset\),
(C-3) \(\text{rank}(\{U_{1|2T}, U_{1|2T}\} | \Theta) = 2Tm\).

Condition (C-1) implies that the state vector is sufficiently excited, and (C-2) shows that there is no linear feedback from \(x\) to the input \(u\). Also, (C-3) is the PE condition.

A subspace identification algorithm with a scheme of attenuating the effect of outliers is given by the following.

Subspace algorithm with a scheme that attenuates outliers

Step 1 Detect outliers based on the median of (22).

Step 2 Define \(\Theta\) of (23) as (25).

Step 3 Compute the weighted LQ decomposition of the input output data matrix
\[
\begin{bmatrix} 
U_{r+1|2T} \\
Y_{1|2T} \\
Y_{r+1|2T}
\end{bmatrix} = \begin{bmatrix} 
L_{11} & 0 & 0 \\
L_{21} & L_{22} & 0 \\
L_{31} & L_{32} & L_{33} & 0 \\
L_{41} & L_{42} & L_{43} & L_{44}
\end{bmatrix}
\begin{bmatrix} 
Q_1^T \\
Q_2^T \\
Q_3^T \\
Q_4^T
\end{bmatrix}
\] (28)

where \(L_{11}, L_{22} \in \mathbb{R}^{m \times m}, L_{33}, L_{44} \in \mathbb{R}^{p \times p}\) are lower triangular, and \(Q_1, Q_2 \in \mathbb{R}^{p \times m}, Q_3, Q_4 \in \mathbb{R}^{p \times p}\) satisfy
\[
\begin{bmatrix} 
Q_1^T \\
Q_2^T \\
Q_3^T \\
Q_4^T
\end{bmatrix} = \begin{bmatrix} 
I_{2Tm} & 0 \\
0 & J
\end{bmatrix}
\]
where \(J = \text{block-diag}(I, 0) \in \mathbb{R}^{2T \times 2T}\), depending on the number of outliers.

Step 4 Compute a state space realization \((A, B, C, D)\) as in Appendix 4.

7. Numerical Simulation

In this section, we present a simulation result to show the effectiveness of the subspace identification method coupled with a scheme of attenuating outliers. In our simulations, two subspace identification
algorithms, ORT-method with and without the attenuation scheme of outliers, are compared.

Let $G(z)$ be a fifth-order system described by ([21])

$$G(z) = \frac{G_n(z)}{G_d(z)}$$

where $G_n(z)$ and $G_d(z)$ are polynomials such that

$$G_n(z) = 0.0275 z^{-4} + 0.0551 z^{-5},$$

$$G_d(z) = 1 - 2.3443 z^{-1} + 3.081 z^{-2} - 2.5274 z^{-3} + 1.2415 z^{-4} - 0.3686 z^{-5}.$$  

Simulated data are generated by

$$y_t = G(z) u_t + v_t$$

where $u_t$ is the known exogenous input and $y_t$ is the measured output corrupted by noise $v_t$. We assume that $\text{Prob}(\alpha = 1) = 0.05$ in (2), and the variances are given by $\sigma^2_{2n} = 0.01$ and $\sigma^2_{20} = 1$ in (3), and that the outliers are contained only in the measured output. A sample realization of the measurement noise $v_t$ containing outliers is shown in Fig. 2.

Fig. 3 shows the Bode plots of the estimated system for a particular run using the ORT-method with the standard LQ and the WLQ decompositions, respectively, where $\nu = 1981$, $k = 20$, and $c = 8$. This figure shows the Bode plots are improved by using the WLQ decomposition technique not only in high frequency but also in low frequency ranges.

In Figs. 4 and 5, we plot the estimated poles for 30 simulation runs carried out with different noise realizations. We see that the pole estimates by the ORT-method coupled with the WLQ decomposition are clustered around the true poles, whereas those by the standard LQ decomposition are widely scattered.

We see from the results shown above that the ORT-method with WLQ decomposition outperforms the standard ORT-method with LQ decomposition. Also, a numerical result by using a MOESP-based robust identification scheme is found in Ref. [16].

8. Conclusions

In this paper, we have considered a subspace identification problem with observation outliers, and derived a subspace identification method by combining the ORT-method and a WLQ decomposition. More precisely, we apply the ORT-method to the input-output data to obtain residuals of the output sequence. By using the median of the residuals, outliers are detected by a simple scheme in robust statistics. Based on the detected outliers, a weighting matrix is generated automatically, and is incorporated in the WLQ decomposition to get an estimate of the system matrices. A numerical example has demonstrated the effectiveness of the proposed method.

References


Appendix

Appendix 1. Proof of Lemma 1

The following Propositions 2 and 3 are well known.

[Proposition 2] [2] There exists a $\Theta$-orthogonal projection from $Y$ onto $U$ if and only if there exists a matrix $P$ such that $\langle Y, U \rangle_\Theta = P \langle U, U \rangle_\Theta$ holds. If it exists, the $\Theta$-orthogonal projection is given by

$$\tilde{Y} = P U$$

(A1)

where $P$ is an arbitrary solution satisfying $\langle Y, U \rangle_\Theta = P \langle U, U \rangle_\Theta$.

[Proposition 3] [7] If $\Theta \geq 0$, there exists a $P$, such that $\langle Y, U \rangle_\Theta = P \langle U, U \rangle_\Theta$.

Now we prove Lemma 1. From Propositions 2 and 3, we get

$$\Phi(PU) = \text{Tr}(Y - PU, Y - PU)$$

$$= \text{Tr} \left[ -P^T I \right] Y \Theta U^T Y \Theta U^T \left[ -P^T I \right]$$

$$= \text{Tr}(P(U - P) \langle U, U \rangle_\Theta (P^T U - P^T) + \text{Tr}((Y, Y) \langle U, U \rangle - P \langle U, U \rangle_\Theta P^*)).$$

Since $\Theta \geq 0$, $\Phi(PU)$ is minimized when $PU = PU$.

When $\langle U, U \rangle_\Theta > 0$ holds, $P$ is a unique minimizer given by $\hat{E}_\Theta(Y|U)$, since $PU$ satisfies (A1).

Appendix 2. Proof of Theorem 1

If $\langle U, U \rangle_\Theta > 0$, then $Y_0 = PU$ is a unique minimizer and $P$ is given by $\langle Y, U \rangle_\Theta (U, U)_\Theta^{-1}$. Since, from $\langle U, U \rangle_\Theta = L_u L_u^T > 0$, $L_u$ is non-singular, (16) is obtained by noting that the rows of $Q_u^*$ form a $\Theta$-orthonormal basis for the row space of $U$.

Appendix 3. Proof of Theorem 2

Define matrices $A_{uz}$ and $A_{yz}$ as

$$\begin{bmatrix} A_{uz} \\ A_{yz} \end{bmatrix} := \begin{bmatrix} U \\ Y \end{bmatrix} V \begin{bmatrix} 0_{n_1 \times p_2} \\ I_{p_2} \end{bmatrix}. \quad (A2)$$

Also define matrices $Q_u$, $Q_y$ and $Q_z$ as

$$\begin{bmatrix} Q_u^* \\ Q_y^* \\ Q_z^* \end{bmatrix} := \begin{bmatrix} Q_u^* \Theta_1^{-1/2} A^{\top}_w A_w \\ Q_y^* \Theta_1^{-1/2} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0_{n_1 \times p_2} \\ I_{p_2} \end{bmatrix} V^T. \quad (A3)$$

From (17), (18), (19) and (A2), we have

$$\begin{bmatrix} U \\ Y \end{bmatrix} V \begin{bmatrix} \Theta_1^{1/2} & 0 \\ 0 & I_{p_2} \end{bmatrix} = A_{uw} 0 \\ A_{uy} A_{yz} - A_{uy} A_{uw} A_{uz} \end{bmatrix} \begin{bmatrix} Q_u^* \Theta_1^{-1} A_w \\ 0 \\ 0 \end{bmatrix}.$$

Using (A3), we have the following equations:
\[
\begin{bmatrix}
Y
\end{bmatrix} =
\begin{bmatrix}
A_{uu} & 0 \\
A_{yu} & A_{yy} - A_{yu}A_{uu}^{-1}A_{ux}
\end{bmatrix}
\begin{bmatrix}
Q_u^T \\
Q_y^T \\
Q_x^T
\end{bmatrix},
\]

and
\[
\begin{bmatrix}
Q_u^T \\
Q_y^T \\
Q_x^T
\end{bmatrix} = \begin{bmatrix}
I & 0 & 0 \\
0 & I & 0 \\
0 & 0 & 0
\end{bmatrix}.
\]

The result is immediate from these equations.

**Appendix 4. Subspace Identification Algorithm**

It follows from (27) and (28) that
\[
\dot{Y}_{t+1|2t} = L_{41}Q_1^T + L_{42}Q_2^T = \Psi_r L_{11}Q_1^T + \mathcal{O}_r \hat{X}_r.
\]

Post-multiplying (A4) by \(Q_2^T\) yields \(L_{42} = \mathcal{O}_r \hat{X}_r Q_2^T\). Since, from the assumption (C-1), \(\hat{X}_r Q_2^T\) has row full rank, we have
\[
\text{column space}(L_{42}) = \text{column space}(\mathcal{O}_r).
\]

Similarly, pre-multiplying (A4) by \((\mathcal{O}_r^T)^T\) and post-multiplying by \(Q_1^T\) yield
\[
(\mathcal{O}_r^T)^T L_{41} = (\mathcal{O}_r^T)^T \Psi_r L_{11}.
\]

We obtain \(A\) and \(C\) from \(\mathcal{O}_r\) in (A5), and \(B\) and \(D\) from (A6) by using the identification algorithm given below.

**Step 1** Compute the SVD
\[
L_{42} = U S V^T = \begin{bmatrix} \hat{U} & \hat{U} \end{bmatrix} \begin{bmatrix} \hat{S} & 0 \\ 0 & \hat{S} \end{bmatrix} \begin{bmatrix} \hat{V}^T \\ \hat{V}^T \end{bmatrix}
\]

where \(\hat{O}_r = \mathcal{O}_r(p+1:p+1:m)\), which is obtained by deleting the first \(p\) rows from \(\mathcal{O}_r\).

Given the extended observability matrix, the Toeplitz matrix \(\Psi_r\) becomes linear with respect to \(B\) and \(D\). Thus, \(B\) and \(D\) are derived by solving (A6), where we can use \(U^T\) of (A7) for \((\mathcal{O}_r^T)^T\).

**Step 2** Compute \(A\) and \(C\) as
\[
A = \hat{O}_{r-1} \hat{O}_r^T;
C = \text{the first } p \text{ rows of } \mathcal{O}_r.