Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Stochastic subspace identification of linear systems with observation outliers

Hideyuki Tanaka, Jaafar ALMutawa and Tohru Katayama

Department of Applied Mathematics and Physics
Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan.
email: {htanaka, jaafar, katayama}@amp.i.kyoto-u.ac.jp

Abstract— **This paper considers a problem of identifying stochastic linear systems subject to observation outliers, where the observation noise contains large values with a low probability. A stochastic subspace identification method for the problem is developed based on a block LQ decomposition, introducing a weighting matrix to delete outputs which are suspected as outliers. The weighting matrix is generated automatically, and is incorporated in the block LQ decomposition to get improved estimates of the forward innovation representation. A numerical simulation result is included to show effectiveness of the proposed method.**

I. Introduction

Stochastic subspace identification algorithms [1], [2] compute stochastic state space systems from finite strings of time-series data, where the numerical operations include not only the singular value decomposition (SVD) and QR decomposition (or LQ decomposition), but also computation of the stabilizing solution of an associated Riccati equation. It is well-known that the stochastic subspace identification algorithms [1], [2] are derived on the basis of the stochastic realization theories [3], [4], [5].

Recently, we have re-visited stochastic realization theory [6]. Moreover, adapting stochastic realization algorithms to a finite string of time-series data, we have presented a stochastic subspace identification algorithm [7], and given an algorithm to guarantee stability and minimum phase property of the resulting forward innovation representation [8].

In real applications, there are cases where large errors are contained in observed data with a low probability, which is however significantly higher than that given by a single Gaussian distribution, so that a standard Gaussian assumption for observation noises may fail [9]. The least-squares estimate is quite sensitive to this type of non-Gaussian disturbances called outliers, and thus numerical performance of LQ decomposition may degrade. Therefore subspace identification methods which are based on the standard LQ decomposition may not give good models under the observation outliers. It will be thus an important topic to develop a simple method to cope with outliers for stochastic subspace identification methods. For the purely non-deterministic system, however, subspace identification methods have not been studied well under observation outliers to the author's knowledge.

In this paper, we consider a stochastic 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 stochastic subspace identification method for outliers by using a weighting matrix for the block LQ decomposition, and develop a simple method of deleting outliers with the help of a scheme for robust estimation methods in statistics [10]. We give a numerical simulation result to show effectiveness of the proposed method.

II. Problem Statement

In this section, we consider a problem of identifying stochastic linear systems subject to observation outliers.

Consider the unknown stochastic system

$$\begin{bmatrix} x_{t+1} \\ y_t^{\mathrm{n}} \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} x_t + \begin{bmatrix} w_t \\ v_t^{\mathrm{n}} \end{bmatrix}, \tag{1}$$

where $\begin{bmatrix} w_t^T & (v_t^{\mathrm{n}})^T \end{bmatrix}^T$ is a zero-mean stationary white Gaussian noise sequence with a covariance matrix

$$\mathrm{E}\left\{ \begin{bmatrix} w_s \\ v_s^{\mathrm{n}} \end{bmatrix} \begin{bmatrix} w_t \\ v_t^{\mathrm{n}} \end{bmatrix}^T \right\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{st}. \tag{2}$$

The output $\{y_t^{\mathrm{n}},\ t = 0,\ \pm 1,\ \pm 2,\ \ldots\}$ is a p-dimensional non-deterministic process with mean zero and covariance matrices

$$\Lambda_k = \mathrm{E}\left\{ (y_{t+k}^{\mathrm{n}})(y_t^{\mathrm{n}})^T \right\}, \quad k = 0, \pm 1, \pm 2, \ldots,$$

where Λ_k $(k = 0,\ \pm 1,\ \ldots)$ is a positive real sequence: $\sum_{i,j} u_i^T \Lambda_{i-j} u_j > 0,\ u_i \not\equiv 0$. Defining $G = \mathrm{E}\{x_{t+1}(y_t^{\mathrm{n}})^T\}$, we have $\Lambda_k = CA^{k-1}G$. We moreover assume that $(A,\ G,\ C)$ is a minimal realization with order n.

Suppose that we observe the output of the stochastic system (1) which contains outliers:

$$\begin{bmatrix} x_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} x_t + \begin{bmatrix} w_t \\ v_t \end{bmatrix}, \tag{3}$$

where v_t is a white noise which contains outliers

$$v_t = (1 - \alpha_t) v_t^{\mathrm{n}} + \alpha_t v_t^{\mathrm{o}} \tag{4}$$

where $\alpha_t = \{0, 1\}$ and $\Pr\{\alpha_t = 1\}$ is small and v_t^{o} is a Gaussian white noise with a covariance matrix

$$\mathrm{E}\left\{ (v_t^{\mathrm{o}})(v_t^{\mathrm{o}})^T \right\} = R^{\mathrm{o}}$$

0-7803-9568-9/05/$20.00 ©2005 IEEE

which satisfies

$$R^{\mathrm{o}} \gg R. \tag{5}$$

Under these conditions, the output equation is written as

$$y_t = (1-\alpha_t)y_t^{\mathrm{n}} + \alpha_t y_t^{\mathrm{o}},$$

where y_t^{o} is defined as $y_t^{\mathrm{o}} := Cx_t + v_t^{\mathrm{o}}$. In the following, y_t^{n} is called the normal output, whereas y_t^{o} the outliers output. It should be, however, noted that y_t^{n} and y_t^{o} are fictitious outputs never observed.

The problem in this paper is stated as follows: Let τ be a positive integer with $\tau > n$, and ν be sufficiently large. Suppose that a finite string of time-series data

$$\{y_t \,:\, t = 0,\, 1,\, \ldots,\, 2\tau+\nu-2\} \tag{6}$$

is given. Estimate then a realization $(A,\ \hat{K},\ C,\ \hat{R})$ of a forward innovation representation of the system (1),

$$\begin{aligned} &\begin{bmatrix} \hat{x}_{t+1} \\ y_t^{\mathrm{n}} \end{bmatrix} = \begin{bmatrix} A & \hat{K} \\ C & I \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ \hat{v}_t \end{bmatrix}, \\ &\mathrm{E}\{\hat{v}_s \hat{v}_t^T\} = \hat{R}\delta_{st}, \end{aligned} \tag{7}$$

where $\hat{v}_t$ is an innovation process of the normal output y_t^{n}.

It should noted that we are interested in also detecting the outliers output y_t^{o}, though we only develop an algorithm to estimate (7) in this paper. We check if the output contains outliers or not in (5), by using the forward innovation representation, since there exist Markov representations of y_t^{n} generated by the stochastic system (1) [1].

III. Identification free from outliers

We review a stochastic subspace identification method based on a block LQ decomposition, under the assumption that there are no outliers i.e. $\Pr(\alpha_t = 1) = 0$ or $y_t = y_t^{\mathrm{n}}$.

A. Stochastic realization

In this subsection, we review stochastic realization theory for an observed data [11], [12], assuming that we have data $y_t = y_t^{\mathrm{n}}$ in (6) with $\nu \to \infty$ and $\tau \to \infty$.

[1]Consider a Lyapunov equation

$$P - APA^T = Q.$$

We then have $G - APC^T = S$ and $\Lambda_0 - CPC^T = R$. Every Markov representation is given by solving LMI [11]:

$$M(P) := \begin{bmatrix} P - APA^T & G - APC^T \\ (G - APC^T)^T & \Lambda_0 - CPC^T \end{bmatrix} \geq 0 \tag{8}$$

Factorizing $M(P)$ as

$$M(P) = \begin{bmatrix} L \\ D \end{bmatrix} \begin{bmatrix} L \\ D \end{bmatrix}^T, \quad \begin{bmatrix} L \\ D \end{bmatrix} : \text{full column rank},$$

a stochastic realization is given by

$$\begin{bmatrix} x_{t+1} \\ y_t^{\mathrm{n}} \end{bmatrix} = \begin{bmatrix} A & L \\ C & D \end{bmatrix} \begin{bmatrix} x_t \\ e_t \end{bmatrix}, \tag{9}$$

where e_t is a zero-mean white noise with a unit variance. Among stochastic realizations (9), the forward innovation representation gives the largest $R = DD^T$, since the stabilizing solution P_* of the associated Riccati equation gives the smallest solution to the LMI (8): $P_* \leq P$. See also Section III.

According to [11], [12], we define the tail matrix from observed data as

$$\boldsymbol{y}_t := \begin{bmatrix} y_t & y_{t+1} & y_{t+2} & \cdots \end{bmatrix} \in \mathbb{R}^{p\times\infty}.$$

We also define a vector space spanned by all finite linear combinations of row vectors of $\boldsymbol{y}_t$ as

$$\mathcal{Y}^\infty := \{ \sum a_k^T \boldsymbol{y}_k \mid a_k \in \mathbb{R}^p,\ k = 0,\, \pm1,\, \pm2,\, \ldots \}.$$

For $a^T\boldsymbol{y}_i$ and $b^T\boldsymbol{y}_j \in \mathcal{Y}^\infty$, we define an inner product as

$$\langle a^T\boldsymbol{y}_i, b^T\boldsymbol{y}_j \rangle_{\frac{I}{\infty}} = a^T \Lambda_{i-j} b.$$

By completing the vector space $\mathcal{Y}^\infty$ with respect to convergence in the norm induced by the inner product, we get a Hilbert space, which is also written as $\mathcal{Y}^\infty$.

Let $\mathcal{U}$ be a Hilbert subspace of $\mathcal{Y}^\infty$ spanned by row vectors of a matrix $U \in \mathbb{R}^{\bullet\times\infty}$, and the orthogonal projection of $\boldsymbol{\xi} \in \mathcal{Y}^\infty$ onto the space $\mathcal{U}$ be denoted by $\hat{\mathrm{E}}_{\frac{I}{\infty}}(\boldsymbol{\xi} \,|\, U)$.

Define the past and the future matrices as

$$Y_t^- := \begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \boldsymbol{y}_{t-3} \\ \vdots \end{bmatrix}, \quad Y_t^+ := \begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{y}_{t+1} \\ \boldsymbol{y}_{t+2} \\ \vdots \end{bmatrix}.$$

The block Hankel matrix of Λ_i is then given by

$$\mathcal{H} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \Lambda_3 & \cdots \\ \Lambda_2 & \Lambda_3 & \Lambda_4 & \cdots \\ \Lambda_3 & \Lambda_4 & \Lambda_5 & \cdots \\ \Lambda_4 & \Lambda_5 & \Lambda_6 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \langle Y_t^+, Y_t^- \rangle_{\frac{I}{\infty}},$$

and $\mathcal{H} = \mathcal{O}\mathcal{C}$ holds, where $\mathcal{O}$ and $\mathcal{C}$ are extended observability and reachability matrices described as

$$\begin{aligned} \mathcal{O} &= \begin{bmatrix} C^T & (CA)^T & (CA^2)^T & \cdots \end{bmatrix}^T, \\ \mathcal{C} &= \begin{bmatrix} G & AG & A^2G & A^3G & \cdots \end{bmatrix}. \end{aligned}$$

Define moreover variables as

$$\begin{aligned} \hat{\boldsymbol{x}}_t &:= \mathcal{C}\langle Y_t^-, Y_t^- \rangle_{\frac{I}{\infty}}^{-1} Y_t^-, \\ \hat{\boldsymbol{v}}_t &:= \boldsymbol{y}_t - \hat{\mathrm{E}}_{\frac{I}{\infty}}(\boldsymbol{y}_t | Y_t^-). \end{aligned}$$

The projection of the future onto the past is then given as [3], [4]

$$\hat{\mathrm{E}}_{\frac{I}{\infty}}(Y_t^+ | Y_t^-) = \mathcal{O}\hat{\boldsymbol{x}}_t,$$

and $\boldsymbol{y}_t$ is expressed as a forward innovation representation in the Hilbert space $\mathcal{Y}^\infty$ [2], [11],

$$\begin{bmatrix} \hat{\boldsymbol{x}}_{t+1} \\ \boldsymbol{y}_t \end{bmatrix} = \begin{bmatrix} A & \hat{K} \\ C & I \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}}_t \\ \hat{\boldsymbol{v}}_t \end{bmatrix}, \tag{10}$$

where $\hat{\boldsymbol{v}}_t$ satisfies $\hat{R}\delta_{st} = \langle \hat{\boldsymbol{v}}_s, \hat{\boldsymbol{v}}_t \rangle_{\frac{I}{\infty}}$ and $\hat{K}$ and $\hat{R}$ are defined as

$$\hat{K} := (G - AP_*C^T)(\Lambda_0 - CP_*C^T)^{-1}, \tag{11}$$

$$\hat{R} := \Lambda_0 - CP_*C^T, \tag{12}$$

where P_* is the stabilizing solution [2] of a Riccati equation

$$P = APA^T + (G - APC^T) \times (\Lambda_0 - CPC^T)^{-1}(G - APC^T)^T.$$

From the system (10), we have

$$\begin{bmatrix} \hat{\boldsymbol{x}}_{t+1} \\ \hat{\boldsymbol{v}}_t \end{bmatrix} = \begin{bmatrix} A - \hat{K}C & \hat{K} \\ -C & I \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}}_t \\ \boldsymbol{y}_t \end{bmatrix}. \tag{13}$$

Since (13) is a purely non-deterministic system, $\hat{\boldsymbol{x}}_t$ is expressed as [2]

$$\hat{\boldsymbol{x}}_t = \begin{bmatrix} \hat{x}_t & \hat{x}_{t+1} & \hat{x}_{t+2} & \cdots \end{bmatrix}, \tag{14}$$

where $\hat{x}_t$ is determined from the steady state Kalman filter,

$$\hat{x}_{t+1} = A\hat{x}_t + \hat{K}(y_t - C\hat{x}_t),$$

and $\hat{K}$ is the Kalman gain defined as (11). Moreover, from (14), $\hat{\boldsymbol{v}}_t$ is given as

$$\hat{\boldsymbol{v}}_t = \begin{bmatrix} \hat{v}_t & \hat{v}_{t+1} & \hat{v}_{t+2} & \cdots \end{bmatrix}, \tag{15}$$

where $\hat{v}_t = y_t - C\hat{x}_t$.

Now, define matrices as

$$\hat{L}_0 = I; \quad \hat{L}_k = CA^{k-1}\hat{K}, \quad k = 1, 2, \ldots. \tag{16}$$

In terms of $\hat{L}_k$ in (16), we define a block Hankel matrix as

$$\hat{\mathcal{S}} = \begin{bmatrix} \hat{L}_1 & \hat{L}_2 & \hat{L}_3 & \cdots \\ \hat{L}_2 & \hat{L}_3 & \hat{L}_4 & \cdots \\ \hat{L}_3 & \hat{L}_4 & \hat{L}_5 & \cdots \\ \hat{L}_4 & \hat{L}_5 & \hat{L}_6 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

and block upper and lower triangular matrices as

$$\hat{\mathcal{L}}^- := \begin{bmatrix} \hat{L}_0 & \hat{L}_1 & \hat{L}_2 & \hat{L}_3 & \cdots \\ & \hat{L}_0 & \hat{L}_1 & \hat{L}_2 & \cdots \\ & & \hat{L}_0 & \hat{L}_1 & \cdots \\ & & & \hat{L}_0 & \cdots \\ 0 & & & & \ddots \end{bmatrix},$$

$$\hat{\mathcal{L}}^+ := \begin{bmatrix} \hat{L}_0 & & & & 0 \\ \hat{L}_1 & \hat{L}_0 & & & \\ \hat{L}_2 & \hat{L}_1 & \hat{L}_0 & & \\ \hat{L}_3 & \hat{L}_2 & \hat{L}_1 & \hat{L}_0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Define moreover variables as

$$\hat{V}_t^- = \begin{bmatrix} \hat{\boldsymbol{v}}_{t-1} \\ \hat{\boldsymbol{v}}_{t-2} \\ \hat{\boldsymbol{v}}_{t-3} \\ \vdots \end{bmatrix}, \quad \hat{V}_t^+ = \begin{bmatrix} \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{v}}_{t+1} \\ \hat{\boldsymbol{v}}_{t+2} \\ \vdots \end{bmatrix},$$

and matrices as

$$\hat{\mathcal{R}} := \text{block-diag}(\hat{R}, \hat{R}, \ldots),$$
$$\hat{\mathcal{F}} := \begin{bmatrix} \hat{K} & A\hat{K} & A^2\hat{K} & \cdots \end{bmatrix}.$$

[2]The stabilizing solution P_* is the smallest solution in the LMI (8): $P_* \leq P$ [3], thus the forward innovation representation (10) gives the largest $R = DD^T$ among Markov representations (9).

Proposition 1: [6] In terms of the innovation processes $\hat{V}_t^-$ and $\hat{V}_t^+$, the past Y_t^- and the future Y_t^+ are decomposed as

$$\begin{bmatrix} Y_t^- \\ Y_t^+ \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{L}}^- & 0 \\ \hat{\mathcal{S}} & \hat{\mathcal{L}}^+ \end{bmatrix} \begin{bmatrix} \hat{V}_t^- \\ \hat{V}_t^+ \end{bmatrix}$$

where $\hat{V}_t^-$ and $\hat{V}_t^+$ satisfy

$$\left\langle \begin{bmatrix} \hat{V}_t^- \\ \hat{V}_t^+ \end{bmatrix}, \begin{bmatrix} \hat{V}_t^- \\ \hat{V}_t^+ \end{bmatrix} \right\rangle_{\frac{I}{\infty}} = \begin{bmatrix} \hat{\mathcal{R}} & 0 \\ 0 & \hat{\mathcal{R}} \end{bmatrix}.$$

Moreover, $\hat{\mathcal{S}}$ has a decomposition

$$\hat{\mathcal{S}} = \mathcal{O}\hat{\mathcal{F}}. \tag{17}$$

The data matrix defined by stacking $\boldsymbol{y}_t$ has equivalently a block LQ decomposition

$$\begin{bmatrix} \vdots \\ \boldsymbol{y}_{t-2} \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_t \\ \boldsymbol{y}_{t+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & & & & & \\ \cdots & \hat{L}_0 & & & & \\ \cdots & \hat{L}_1 & \hat{L}_0 & & & \\ \cdots & \hat{L}_2 & \hat{L}_1 & \hat{L}_0 & & \\ \cdots & \hat{L}_3 & \hat{L}_2 & \hat{L}_1 & \hat{L}_0 & \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \hat{\boldsymbol{v}}_{t-2} \\ \hat{\boldsymbol{v}}_{t-1} \\ \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{v}}_{t+1} \\ \vdots \end{bmatrix}, \tag{18}$$

where $\hat{\boldsymbol{v}}_t$ satisfies

$$\left\langle \begin{bmatrix} \vdots \\ \hat{\boldsymbol{v}}_{t-1} \\ \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{v}}_{t+1} \\ \vdots \end{bmatrix}, \begin{bmatrix} \vdots \\ \hat{\boldsymbol{v}}_{t-1} \\ \hat{\boldsymbol{v}}_t \\ \hat{\boldsymbol{v}}_{t+1} \\ \vdots \end{bmatrix} \right\rangle_{\frac{I}{\infty}} = \begin{bmatrix} \ddots & & & \\ & \hat{R} & & \\ & & \hat{R} & \\ & & & \ddots \end{bmatrix}. \tag{19}$$

We can thus compute $\hat{L}_t$ and $\hat{R}$ by means of the block LQ decomposition (18), and obtain $(A,\ \hat{K},\ C)$ from (17).

B. Stochastic subspace identification

Adapting stochastic realization theory to a finite string of time-series data, we have a stochastic subspace identification method [7], [8].

Define a matrix as

$$\tilde{\boldsymbol{y}}_t = \begin{bmatrix} y_t & y_{t+1} & \cdots & y_{t+\nu-1} \end{bmatrix} \in \mathbb{R}^{p\times\nu}. \tag{20}$$

We also define a bilinear product of two matrices $X \in \mathbb{R}^{p\times\nu}$ and $Y \in \mathbb{R}^{q\times\nu}$ as

$$\langle X, Y\rangle_{\frac{I}{\nu}} = \frac{1}{\nu}XY^T.$$

We then have an approximation,

$$\Lambda_{i-j} \approx \langle \boldsymbol{y}_i, \boldsymbol{y}_j\rangle_{\frac{1}{\nu}}.$$

For $\tau > n$, also define a matrix as

$$\tilde{Y}_0^+ = \begin{bmatrix} \tilde{\boldsymbol{y}}_0 \\ \tilde{\boldsymbol{y}}_1 \\ \vdots \\ \tilde{\boldsymbol{y}}_{2\tau-1} \end{bmatrix}. \tag{21}$$

Assume $\langle \tilde{Y}_0^+, \tilde{Y}_0^+\rangle_{\frac{I}{\nu}} > 0$. An identification algorithm is given then as follows [8]. We compute the standard LQ decomposition of $\tilde{Y}_0^+$ which defined in (21):

$$\frac{1}{\sqrt{\nu}}\tilde{Y}_0^+ = LQ^T, \tag{22}$$

or equivalently we compute a decomposition

$$\tilde{Y}_0^+ = L\tilde{Q}^T, \quad \langle \tilde{Q}^T, \tilde{Q}^T \rangle_{\frac{I}{\nu}} = I, \tag{23}$$

where $\tilde{Q} := \sqrt{\nu}Q$. Partition L as

$$L = \begin{bmatrix} L_{0,0} & & 0 \\ \vdots & \ddots & \\ L_{2\tau-1,0} & \cdots & L_{2\tau-1,2\tau-1} \end{bmatrix} = \begin{bmatrix} L_{pp} & 0 \\ L_{fp} & L_{ff} \end{bmatrix},$$

where $L_{i,j} \in \mathbb{R}^{p\times p}$ and L_{pp}, L_{fp}, $L_{ff} \in \mathbb{R}^{p\tau\times p\tau}$. Define a matrix as

$$D_L := \text{block-diag}(L_{0,0}, \ldots, L_{2\tau-1,2\tau-1}).$$

The matrix D_L is then non-singular from the assumption.

Define matrices as

$$\tilde{\mathcal{L}}_0^+ := LD_L^{-1}, \tag{24}$$
$$\tilde{V}_0^+ := D_L\tilde{Q}^T, \tag{25}$$
$$\tilde{\mathcal{R}}_0^+ := D_LD_L^T. \tag{26}$$

We then have a block LQ decomposition of the matrix $\tilde{Y}_0^+$,

$$\tilde{Y}_0^+ = \tilde{\mathcal{L}}_0^+\tilde{V}_0^+, \tag{27}$$
$$\langle \tilde{V}_0^+, \tilde{V}_0^+ \rangle_{\frac{I}{\nu}} = \tilde{\mathcal{R}}_0^+, \tag{28}$$

where $\tilde{\mathcal{L}}_0^+$ and $\tilde{\mathcal{R}}_0^+$ are described as

$$\tilde{\mathcal{L}}_0^+ = \begin{bmatrix} \tilde{L}_{0,0} & & 0 \\ \vdots & \ddots & \\ \tilde{L}_{2\tau-1,0} & \cdots & \tilde{L}_{2\tau-1,2\tau-1} \end{bmatrix}, \tag{29}$$

$$\tilde{\mathcal{R}}_0^+ = \text{block-diag}(\tilde{R}_0, \ldots, \tilde{R}_\tau, \ldots, \tilde{R}_{2\tau-1}), \tag{30}$$

where $\tilde{R}_t \in \mathbb{R}^{p\times p}$, $\tilde{L}_{i,j} = L_{i,j}L_{j,j}^{-1}$ and $\tilde{L}_{i,i} = I_p$.

A stochastic subspace identification method is summarized as follows [7], [8].

A stochastic subspace identification algorithm

Step 1: Compute the standard LQ decomposition (22) and obtain $\tilde{\mathcal{L}}_0^+$, $\tilde{\mathcal{R}}_0^+$ as (24) and (26), respectively. Determine moreover a matrix as

$$\tilde{\Psi}_\tau := L_{fp}L_{fp}^T + L_{ff}L_{ff}^T. \tag{31}$$

Step 2: Calculate the SVD of a matrix $(\tilde{\Psi}_\tau)^{-\frac{1}{2}}L_{fp}$ as

$$(\tilde{\Psi}_\tau)^{-\frac{1}{2}}L_{fp} = \tilde{U}_\tau\tilde{\Sigma}_\tau\tilde{V}_J^T, \quad \tilde{\Sigma}_\tau \in \mathbb{R}^{\tilde{n}\times\tilde{n}}, \tag{32}$$

where $\tilde{U}_\tau^T\tilde{U}_\tau = I$, $\tilde{V}_J^T\tilde{V}_J = I$ and $\operatorname{rank}\tilde{\Sigma}_\tau = \tilde{n}$. Based on the SVD (32), define $\tilde{\mathcal{O}}_\tau$ as $\tilde{\mathcal{O}}_\tau = \left(\tilde{\Psi}_\tau\right)^{\frac{1}{2}}\tilde{U}_\tau\tilde{\Sigma}_\tau^{\frac{1}{2}}$.

Step 3: Compute $\tilde{C}$ and $\tilde{A}$ as

$$\tilde{C} = \tilde{\mathcal{O}}_\tau(1:p,:), \quad \tilde{A} = \tilde{\mathcal{O}}_{\tau-1}^\dagger\tilde{\mathcal{O}}_\tau(p+1:p\tau,:),$$

where $\tilde{\mathcal{O}}_{\tau-1} := \tilde{\mathcal{O}}_\tau(1:p(\tau-1),:)$ and $(\cdot)^\dagger$ expresses Moore-Penrose pseudo-inverse.

Step 4: Define $\tilde{R}_\tau$ from (30), and compute $\tilde{K}_\tau$ from

$$\tilde{K}_\tau = \tilde{\mathcal{O}}_{\tau-1}^\dagger \begin{bmatrix} \tilde{L}_{\tau+1,\tau} \\ \tilde{L}_{\tau+2,\tau} \\ \vdots \\ \tilde{L}_{2\tau-1,\tau} \end{bmatrix},$$

where $\tilde{L}_{\tau+1,\tau}$, $\tilde{L}_{\tau+2,\tau}$, …, $\tilde{L}_{2\tau-1,\tau}$ are found in the matrix $\tilde{\mathcal{L}}_0^+$ in (29).

Step 5: Define transfer functions

$$\tilde{\Upsilon}_\tau(z) := \tilde{W}_\tau(z)\tilde{W}_\tau^T(z^{-1}),$$
$$\tilde{W}_\tau(z) := (\tilde{C}(zI-\tilde{A})^{-1}\tilde{K}_\tau + I)\tilde{R}_\tau^{\frac{1}{2}}.$$

The transfer function $\tilde{\Upsilon}_\tau(z)$ is positive real, and can be a good approximation to the true spectral density of y_t for large ν and τ [7]. Find a stable and minimum phase spectral factor from a canonical spectral factorization of $\tilde{\Upsilon}_\tau(z)$, solving two Riccati equations, where the procedure is given in [8].

Partition the matrix $\tilde{V}_0^+ \in \mathbb{R}^{2\tau p\times\nu}$ as

$$\tilde{V}_0^+ = \begin{bmatrix} \tilde{\boldsymbol{v}}_0 \\ \tilde{\boldsymbol{v}}_1 \\ \vdots \\ \tilde{\boldsymbol{v}}_{2\tau-1} \end{bmatrix}, \quad \tilde{\boldsymbol{v}}_i \in \mathbb{R}^{p\times\nu}. \tag{33}$$

From (15), estimates of innovations $\hat{v}_{2\tau-1}$, $\hat{v}_{2\tau}$, …, $\hat{v}_{2\tau+\nu-2}$ are given by

$$\tilde{\boldsymbol{v}}_{2\tau-1} = \begin{bmatrix} \tilde{v}_{2\tau-1} & \tilde{v}_{2\tau} & \cdots & \tilde{v}_{2\tau+\nu-2} \end{bmatrix}.$$

The identification method shown above estimates forward innovations $\hat{v}_t$ by means of the block LQ decomposition (27), which is computed in the first step of the algorithm summarized above.

IV. Identification under outliers

We give a method of identifying systems subject to observation outliers. To this end, we present an identification method by using a weighting matrix.

A. A weighting matrix

We define a bilinear product of two matrices $X \in \mathbb{R}^{p\times\nu}$ and $Y \in \mathbb{R}^{q\times\nu}$ as

$$\langle X, Y \rangle_\Theta = X\Theta Y^T$$

for a nonnegative definite $\Theta \in \mathbb{R}^{\nu\times\nu}$, where

$$\Theta = r \times \operatorname{diag}(\theta_0, \theta_1, \cdots, \theta_{\nu-1}) \in \mathbb{R}^{\nu\times\nu} \tag{34}$$

for r, $\theta_i \in \mathbb{R}$.

In order to give an idea to construct a weighting matrix, we assume for a while that an outlier exists in "y_i" only and that it is detected exactly. For simplicity, we restrict p to $p = 1$. Define $\boldsymbol{y}_t$ as (20) and $\tilde{Y}_0^+$ as (21). The matrix $\tilde{Y}_0^+$ is then written as

$$\tilde{Y}_0^+ = \left[\begin{array}{ccc|ccc|ccc} y_0 & \cdots & & y_{i-2\tau+1} & \cdots & y_i & \cdots & y_{\nu-1} \\ \vdots & & & & \iddots & & & \vdots \\ y_{2\tau-1} & \cdots & & y_i & \cdots & y_{i+2\tau-1} & \cdots & y_{2\tau+\nu-2} \end{array}\right]. \tag{35}$$

Since $\tilde{Y}_0^+$ has a Hankel structure, y_i appears 2τ times in it for $2\tau-1 \le i \le \nu-1$, and we must delete all the data from $(i-2\tau+1)$-th to $\min\{i, \nu-1\}$-th columns from $\tilde{Y}_0^+$. Thus, if $\Theta \in \mathbb{R}^{\nu\times\nu}$ is defined as

$$\Theta = \frac{1}{\nu-2\tau} \times \operatorname{diag}(\underbrace{1, \cdots, 1}_{i-2\tau+1}, \underbrace{0, \cdots, 0}_{2\tau}, \underbrace{1, \cdots, 1}_{\nu-i-1}),$$

then, we can delete outliers y_i in (35) in order to compute covariance matrices approximately,

$$\Lambda_{i-j} \approx \langle \tilde{\boldsymbol{y}}_i, \tilde{\boldsymbol{y}}_j \rangle_\Theta,$$

and we hence have a good approximation

$$\begin{bmatrix} \Lambda_0 & \Lambda_1^T & \cdots & \Lambda_{2\tau-1}^T \\ \Lambda_1 & \Lambda_0 & \cdots & \Lambda_{2\tau-2}^T \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{2\tau-1} & \Lambda_{2\tau-2} & \cdots & \Lambda_0 \end{bmatrix} \approx \langle \tilde{Y}_0^+, \tilde{Y}_0^+ \rangle_\Theta.$$

B. Constructing weighting matrix

We show how to construct a weighting matrix Θ, by deleting all the outputs suspected as outliers instead of detecting outliers exactly. To this end, we use the stochastic subspace identification algorithm [7], [8] which computes the innovation $\tilde{V}_0^+$ in the first step.

Construction of a weighting matrix

Step 1: Define $\Theta = \frac{1}{\nu} I$ and $\mathcal{D} = \emptyset$.

Step 2: Compute a weighted block LQ decomposition,

$$\tilde{Y}_0^+ = \tilde{\mathcal{L}}_0^+ \tilde{V}_0^+,$$
$$\langle \tilde{V}_0^+, \tilde{V}_0^+ \rangle_\Theta = \tilde{\mathcal{R}}_0^+,$$

where $\tilde{\mathcal{L}}_0^+$ and $\tilde{\mathcal{R}}_0^+$ have block structures as (29) and (30), and $\tilde{L}_{ii} = I$.

Step 3: Partition the matrix $\tilde{V}_0^+$ as (33) and find estimates $\tilde{v}_t$.

Step 4: Obtain the median of $\|\tilde{v}_t\|$

$$s = \text{median}\{\|\tilde{v}_t\| \,:\, t = 2\tau - 1, \cdots, 2\tau + \nu - 2\}.$$

If $\|\tilde{v}_t\| > cs$, then y_t possibly contains outliers [3], where $c = 5 \sim 9$ is used [4], and hence add t to the set $\mathcal{D}$.

Step 5: Define an index set as

$$\mathcal{I} = \{\, j \,|\, i \in \mathcal{D}, \quad i \le j \le m \,\},$$

where $m = \min\{i + 2\tau - 1, \nu + 2\tau - 2\}$. It may be noted that $\mathcal{D} \subseteq \mathcal{I}$ holds.

Step 6: Set θ_i $(0 \le i \le \nu - 1)$ as

$$\theta_i = \begin{cases} 0 & \text{for} \quad i + 2\tau - 1 \in \mathcal{I}, \\ 1 & \text{for} \quad i + 2\tau - 1 \notin \mathcal{I}. \end{cases}$$

Compute $r = 1/(\sum_{i=0}^{\nu-1} \theta_i)$, and define Θ as (34). Go to Step 2, and compute iteratively until the possible outliers are not newly detected.

[3]The process $\tilde{v}_t$ is computed as $\tilde{v}_t = y_t - C\tilde{x}_t$, where $\tilde{x}_t$ is a linear combination of $\{y_{t-2\tau-1}, y_{t-2\tau}, \ldots, y_{t-1}\}$. It thus follows from (3), (4) and (5) that $\{y_t\}$ or $\{y_{t-1}, \ldots, y_{t-2\tau-1}\}$ may contain outliers, if $\|\tilde{v}_t\| > cs$. We delete however only the output $\{y_t\}$ in each iteration in this algorithm.

[4]For the biweight robust estimation, the weighting coefficients are chosen as

$$w(z) = \begin{cases} (1 - (z/cs)^2)^2 & \text{for } \|z\| < cs \\ 0 & \text{for } \|z\| \ge cs \end{cases}$$

where $c = 5 \sim 9$ is recommended [10]. Here we employed a simple cut method with the threshold cs, where the actual value of c is borrowed from the biweight method.

The set $\mathcal{D}$ contains indices of y_t which are suspected as outliers in each iteration, while the set $\mathcal{I}$ contains columns of $\tilde{Y}_0^+$ to be deleted.

The number of indices contained in $\mathcal{D}$ increases monotonically in Step 4 in each iteration. When the possible outliers are not newly detected, the matrix Θ is determined so that all the columns of $\tilde{Y}_0^+$ related with outliers are expected to be deleted. It should be noted that probability of outliers in the given data must be not very high in order that this algorithm successfully finishes.

The present algorithm computes iteratively a weighting matrix, since $\tilde{Y}_0^+$ is contaminated by outliers. The set $\mathcal{D}$ thus contains possibly some normal outputs, which are suspected as outliers in the proposed algorithm. Computing iteratively, however, the algorithm captures outliers, and most of the outliers outputs are expected to be contained in $\mathcal{D}$.

C. Identification based on the weighting matrix

Assume $\langle \tilde{Y}_0^+, \tilde{Y}_0^+ \rangle_\Theta > 0$. We present a stochastic subspace identification method by using the weighting matrix Θ constructed in Subsection IV-B.

Identification algorithm under outliers

Step 1: Compute the weighted LQ decomposition

$$\tilde{Y}_0^+ = L\tilde{Q}^T, \quad \langle \tilde{Q}^T, \tilde{Q}^T \rangle_\Theta = I,$$

instead of (23), and obtain $\tilde{\mathcal{L}}_0^+$, $\tilde{\mathcal{R}}_0^+$ and $\tilde{\Psi}_\tau$ from (24), (26) and (31), respectively.

Steps 2-5: Compute Steps 2-5 as in the **stochastic subspace identification algorithm** shown in Section III-B.

Step 1 is a major difference between the proposed algorithm and the one in Section III-B. The weighting matrix Θ is thus incorporated in the stochastic subspace identification method based on the block LQ decomposition [7], [8] in a simple way.

V. Numerical simulation

In this section, we present a simulation result to show effectiveness of the stochastic subspace identification coupled with a scheme of attenuating outliers. Two stochastic subspace identification algorithms, the algorithm shown in Subsection III-B and the proposed one, are compared.

Finite strings of time-series data y_t are given by different noise realizations, where y_t contains outliers output y_t^{o} with the probability $\Pr(\alpha_t = 1) = 0.01$. The outliers output y_t^{o} is Gaussian noise with $y_t^{\text{o}} \sim \text{N}(0, 6.5)$, and the normal output y_t^{n} is generated by $y_t^{\text{n}} = W(z)e_t$, where e_t is also Gaussian white noise with $e_t \sim \text{N}(0, 1)$. The transfer function $W(z)$ is a second-order system given by

$$W(z) = \frac{1 - 0.20z^{-1} - 0.48z^{-2}}{1 + 0.75z^{-1} + 0.81z^{-2}}.$$

A covariance matrix $\Lambda_0 = \text{E}\{(y_t^{\text{n}})(y_t^{\text{n}})^T\}$ is obtained by solving a Lyapunov equation, and is equal to 6.4449. A sample realization of the measurement output y_t ($t = 6070$, $\ldots$, 6120) is shown in Fig. 1, and the outliers occur at

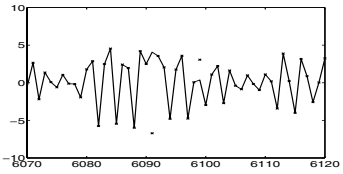


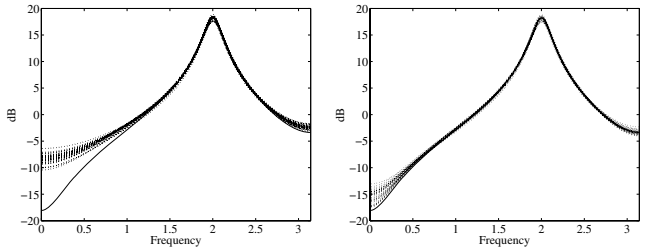
Fig. 1. A sample noise of y_t ($t = 6070, \ldots, 6120$): the solid line expresses y_t^{n}, while "$\times$" expresses y_t.

$t = 6091, 6099$. It seems to be difficult to detect outliers from only distribution of y_t, since $\Lambda_0 = 6.449$ and $y_t^{\mathrm{o}} \sim \mathrm{N}(0, 6.5)$.

We estimated the system for 30 simulation runs carried out with different noise realizations where $\tau = 6$, $\nu = 7,000$ and $\tilde{n} = 2$. In order to construct weighting matrices, we have set c as $c = 2.1\pi = 6.5973$. In average, the number of deleted columns is 601, which is smaller than $2\tau\nu \times \Pr(\alpha_t = 1) = 2 \times 6 \times 7000 \times 0.01 = 840$.

In every simulation, the algorithm successfully finished after 10 iterations to construct a weighting matrix. It should be however noted that the proposed algorithm failed for the data subject to outliers with a higher probability $\Pr(\alpha_t = 1) = 0.05$.

Fig. 2 shows the Bode plots of the estimated system for 30 simulations. These figures show that the Bode plots are improved in the low frequency range.



(a) Algorithm [8] (b) Proposed method

Fig. 2. Bode plots of the systems estimated by the algorithm [8] and the proposed method, where the dotted line expresses the plots of estimates for 30 simulation and the dashed one gives the plot of $W(z)$.

VI. Conclusion

We have considered a stochastic subspace identification method for linear stochastic systems subject to observation outliers, where the observation noise takes on large values with a low probability. We have developed a stochastic subspace identification method based on a weighted block LQ decomposition, where a weighting matrix is automatically constructed by means of an iterative calculation using robust estimation. A numerical simulation result has shown effectiveness of the proposed method.

VII. Acknowledgments

This work was partially supported by Research Project Grant-in-Aid for Scientific Research, JSPS.KAKENHI (15760316) and (17560389), Japan Society for the Promotion of Science.

References

[1] P. Van Overschee and B. De Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29(3):649–660, 1993.

[2] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems*. Kluwer Academic Pub., 1996.

[3] P. L. Faurre. Stochastic realization algorithms. In R. Mehra and D. G. Lainiotis, editors, *System Identification: Advances and Case Studies*, pages 1–25. Academic Press, 1976.

[4] H. Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM J. Control*, 13(1):162–173, 1975.

[5] U. B. Desai, D. Pal, and R. D. Kirkpatrick. A realization approach to stochastic model reduction. *Int. J. Control*, 42(4):821–838, 1985.

[6] H. Tanaka and T. Katayama. A stochastic realization in a Hilbert space based on "LQ decomposition" with application to subspace identification (submitted to Automatica). *Proc. of the 13th IFAC Symposium on System Identification (SYSID 2003)*, pages 899–904, 2003.

[7] H. Tanaka and T. Katayama. Stochastic subspace identification via "LQ decomposition". *Proc. of the 42nd IEEE Conference on Decision and Control (CDC 2003)*, pages 3467–3472, 2003.

[8] H. Tanaka and T. Katayama. Stochastic subspace identification algorithm guaranteeing stability and minimum phase. *Proc. of the 16th IFAC world congress in Prague (IFAC2005)*, 2005.

[9] E.-W. Bai. A random least-trimmed-squares identification algorithm. *Automatica*, 39(9):1651–1659, 2003.

[10] T. Nakagawa and Y. Koyanagi. *Data analysis by the least-squares method (in Japanese)*. The University of Tokyo Press, 1982.

[11] A. Lindquist and G. Picci. Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica*, 32(5):709–733, 1996.

[12] A. Lindquist and G. Picci. Geometric methods for state space identification. In S. Bittanti and G. Picci, editors, *Identification, Adaptation, Learning*, pages 1–69. Springer-Verlag, 1996.