Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

An Estimation Method of Innovations Model in Closed-Loop Environment with Lower Horizons

Kenji Ikeda * **Hideyuki Tanaka** **

* *Tokushima University, Tokushima, 770-8506 Japan (e-mail: ikeda@is.tokushima-u.ac.jp).*
** *Hiroshima University, Higashi Hiroshima, 739-8524 Japan (e-mail: tanakalpha@hiroshima-u.ac.jp)*

Abstract: This paper proposes an estimation method of the innovations model in closed loop environment by using the estimate of the innovations process. The estimate of the innovations process from the finite interval of data has a bias, so are the estimate of the proposed method. However, it is analyzed that the bias can be reduced. The Kalman gain and the covariance of the innovations process are estimated by using a semi-definite programming problem previously proposed by the authors. Numerical simulation illustrates the proposed method gives better performance than Closed-Loop MOESP and PBSID when the data length is large and the past horizon is selected low.

Keywords: System identification, Subspace methods, Kalman filters, Semi-definite programming.

1. INTRODUCTION

Closed loop identification is important or required for production, economic, or safety reasons (Forssell and Ljung (1999)). In the literature on closed-loop subspace identification, utilization of the estimate of the innovations process has become major and common (Chiuso and Picci (2005); van der Veen et al. (2013); Mercère et al. (2016)). Because the estimate of the innovations process from a finite interval of data instead of an infinite interval of data has a bias, the method based on the estimate of the innovations process inevitably results in asymptotically consistent estimate (Knudsen (2001)), *i.e.*, not only the horizontal size (data length minus past and future horizons) but also the vertical size (determined by the past and future horizons) of the data matrix must go to infinity in order for the estimate to converge in probability to the true value. Increase of the horizons requires more computational cost. Thus the lower horizons is preferable from a viewpoint of computational cost.

An interesting method called Closed-Loop MOESP was proposed by van der Veen et al. (2010). It is based on the estimate of the innovations process, so the estimate has an asymptotic bias. However, it can be seen from our numerical simulations that the estimate of (A, C) has a small bias compared to the one in PBSID (Chiuso and Picci (2005)) when the data length is large and lower horizons are adopted. Unfortunately, the asymptotic bias of the estimate of (B, D, K) in Closed-Loop MOESP is not so small as the one of the estimate of (A, C). We will see this in Sec. 5.

The authors proposed a consistent estimate of the innovations model in open-loop environment which is based on a semi-definite programming problem on the squared residuals (Ikeda and Tanaka (2017, 2019)). It has been shown that the Kalman gain K and the covariance of the innovations process Ω_e can be estimated consistently when the consistent estimate of (A, B, C, D) is obtained by, *e.g.*, PO-MOESP method (Verhaegen (1994); Verhaegen and Verdult (2007)). As described above, an accurate estimate of (A, C) can be obtained in the closed-loop environment by using Closed-Loop MOESP. This suggests that a more accurate estimate of the innovations model can be obtained in the closed-loop environment, when a more accurate estimate of (B, D) can be obtained.

In this paper, an estimation method of the innovations model in closed-loop environment with low horizons is developed. First, we analyze the estimation error of the estimate of (A, C) in Closed-Loop MOESP and consider why it gives accurate estimate. In order to take advantage of the mechanism of this accuracy, an estimation method of (B, D) is proposed in which an idea is borrowed from ordinary MOESP (Verhaegen and Dewilde (1992)). For the estimation of K and Ω_e, the method proposed in Ikeda and Tanaka (2017, 2019) is extended and applied to the closed-loop identification. Tanaka and Ikeda (2018) applied the estimation method of K and Ω_e above to the joint-input-output approach in the closed-loop identification. However, the method has not yet been applied to the direct approach. The performance of the proposed method is illustrated by using numerical simulations, in which the bias of the proposed method is considereably reduced even when the past horizon is low and the data length is large.

This paper is organized as follows: Section 2 provides an innovations model as the system to be identified and some assumptions are made. Section 3 summarizes some preliminaries for the subspace identification. Section 4 introduces a proposed method with some analysis on the estimation error in Closed-Loop MOESP. Section 5 shows some numerical simulations in order to illustrate the performance of the proposed method compared to Closed-Loop MOESP and PBSID. Finally, Section 6 concludes the paper.

Notations: Let $X^\dagger$ be a pseudo inverse (Moore-Penrose generalized inverse) of X(Golub and van Loan (1989)).

Let $\lambda_i(A)$ be an eigenvalue of A.

Let $E\{\cdot\}$ denote a mathematical expectation.

Let $\boldsymbol{\mathcal{O}}_f(A,C)$ denote an extended observability matrix composed of the system matrices (A,C) for a given index $f > n$ where n is an order of the system. Namely,

$$\boldsymbol{\mathcal{O}}_f(A,C) := [C^\top, (CA)^\top, \ldots, (CA^{f-1})^\top]^\top. \quad (1)$$

Let $\boldsymbol{\mathcal{C}}_f(A,B)$ denote an extended reachability matrix in the reversed order as

$$\boldsymbol{\mathcal{C}}_f(A,B) := [A^{f-1}B, \ldots, AB, B]. \quad (2)$$

Let $\boldsymbol{\mathcal{T}}_f(A,B,C,D)$ be a block Toeplitz matrix composed of the Markov parameters of the system (A,B,C,D) as

$$\boldsymbol{\mathcal{T}}_f(A,B,C,D) := \begin{bmatrix} D & & & 0 \\ CB & D & & \\ \vdots & & \ddots & \\ CA^{f-2}B & CA^{f-3}B & \cdots & D \end{bmatrix}. \quad (3)$$

Block Hankel matrix composed of a time-series data $\{u_k\}$ is denoted by

$$\boldsymbol{\mathcal{U}}_{i|j} := \begin{bmatrix} u_i & u_{i+1} & \cdots & u_{i+N-1} \\ u_{i+1} & u_{i+2} & \cdots & u_{i+N} \\ \vdots & \vdots & & \vdots \\ u_j & u_{j+1} & \cdots & u_{j+N-1} \end{bmatrix}. \quad (4)$$

We define $\boldsymbol{\mathcal{Y}}_{i|j}$, $\boldsymbol{\mathcal{E}}_{i|j}$, $\boldsymbol{\mathcal{G}}_{i|j}$, and $\boldsymbol{\mathcal{Z}}_{i|j}$, in the same way with $\boldsymbol{\mathcal{U}}_{i|j}$, respectively by using $\{y_k\}$, $\{e_k\}$, $\{g_k = (e_k^\top, f_k^\top)^\top\}$, and $\{z_k = (u_k^\top, y_k^\top)^\top\}$.

2. INNOVATIONS MODEL

Consider the following innovations model (Anderson and Moore (2005)):

$$x_{k+1} = Ax_k + Bu_k + Ke_k, \quad (5)$$

$$y_k = Cx_k + Du_k + e_k, \quad (6)$$

where $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^\ell$, $e_k \in \mathbb{R}^\ell$, and $x_k \in \mathbb{R}^n$ are the input, the output, the noise, and the state, respectively, and $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{\ell\times n}$, $D \in \mathbb{R}^{\ell\times m}$, and $K \in \mathbb{R}^{n\times \ell}$ are the system matrices to be estimated.

The control input u_k is defined as an output of a linear feedback controller as follows:

$$x_{c,k+1} = A_c x_{c,k} + B_c y_k + K_c f_k, \quad (7)$$

$$u_k = C_c x_{c,k} + f_k, \quad (8)$$

where $x_c \in \mathbb{R}^{n_c}$ and $f_k \in \mathbb{R}^m$ are the state of the controller and the innovations process, respectively, and $A_c \in \mathbb{R}^{n_c\times n_c}$, $B_c \in \mathbb{R}^{n_c\times \ell}$, $C_c \in \mathbb{R}^{m\times n_c}$, and $K_c \in \mathbb{R}^{n_c\times m}$ are the system matrices of the controller. The following assumptions are made for this system:

(A1) Feedback system is stable, $|\lambda_i(A-KC)| < 1$, $i = 1,\ldots,n$, and $|\lambda_i(A_c - K_cC_c)| < 1$, $i = 1,\ldots,n_c$.
(A2) The innovations processes $\{e_k\}$ and $\{f_k\}$ are white Gaussian processes with means $E\{e_k\} = 0$ and $E\{f_k\} = 0$ and covariance matrices $E\{e_k e_l^\top\} = \Omega_e\delta_{kl}$ and $E\{f_k f_l^\top\} = \Omega_f\delta_{kl}$, respectively.
(A3) The processes $\{f_k\}$ and $\{e_k\}$ are mutually independent.
(A4) The processes $\{f_k\}$ and $\{e_k\}$ are ergodic and stationary (Anderson and Moore (2005)).

Note that the pair (K, Ω_e) is not arbitrary but defined by using a solution of some Riccati equation. From Assumptions above, the system is identifiable (Anderson and Gevers (1982)).

3. PRELIMINARIES

3.1 I/O data Equation

I/O data equation derived from the innovations model (5) and (6) plays an important roles in the analysis and the implementation of subspace identification methods:

$$\mathcal{Y}_f = \mathcal{O}_f\mathcal{X}_0 + \mathcal{T}_f\mathcal{U}_f + \mathcal{H}_f\mathcal{E}_f, \quad (9)$$

where $\mathcal{O}_f$, $\mathcal{T}_f$, $\mathcal{H}_f$, $\mathcal{X}_i$, $\mathcal{U}_f$, $\mathcal{Y}_f$, E_f, and $\mathcal{F}_f$ are given by

$$\mathcal{O}_f := \boldsymbol{\mathcal{O}}_f(A,C), \quad (10)$$

$$\mathcal{T}_f := \boldsymbol{\mathcal{T}}_f(A,B,C,D), \quad (11)$$

$$\mathcal{H}_f := \boldsymbol{\mathcal{T}}_f(A,K,C,I), \quad (12)$$

$$\mathcal{X}_i := [x_i \; x_{i+1} \; \cdots \; x_{i+N-1}], \quad (13)$$

$$\mathcal{U}_f := \boldsymbol{\mathcal{U}}_{0|f-1}. \quad (14)$$

We define $\mathcal{Y}_f$, $\mathcal{E}_f$, and $\mathcal{F}_f$ as in the same way with $\mathcal{U}_f$.

Most of the subspace methods based on the innovations model (5) and (6), *e.g.* PO-MOESP (Verhaegen (1994)) or N4SID (Van Overschee and De Moor (1994)), adopt the following instrumental variable matrix $\mathcal{Z}_p^-$ in order to reduce the asymptotic bias of the estimate where $\mathcal{Z}_p^- = \boldsymbol{\mathcal{Z}}_{-p|-1}$ is a block Hankel matrix composed of past $z_k = [u_k^\top, y_k^\top]^\top$. The state matrix is represented by using the instrumental variable matrix as:

$$\mathcal{X}_0 = \bar{A}^p\mathcal{X}_{-p} + \mathcal{K}_p\mathcal{Z}_p^- = \bar{A}^p\mathcal{X}_{-p} + \mathcal{X}_0^{(p)}, \quad (15)$$

where $\mathcal{K}_p := \boldsymbol{\mathcal{C}}_p(\bar{A}, [\bar{B}, K])$, $\bar{A} = A - KC$, $\bar{B} = B - KD$, and $\mathcal{X}_0^{(p)} = \mathcal{K}_p\mathcal{Z}_p^-$.

The I/O data equation (9) is rewritten as

$$\mathcal{Y}_f = \mathcal{O}_f\mathcal{X}_0^{(p)} + \mathcal{T}_f\mathcal{U}_f + \mathcal{H}_f\mathcal{E}_f + \mathcal{O}_f\bar{A}^p\mathcal{X}_{-p}, \quad (16)$$

$$= \mathcal{O}_f\mathcal{K}_p\mathcal{Z}_p^- + \mathcal{T}_f\mathcal{U}_f + \mathcal{H}_f\mathcal{E}_f + \mathcal{O}_f\bar{A}^p\mathcal{X}_{-p}. \quad (17)$$

3.2 Estimation of the Innovations Process

From (5) and (6), the following is derived.

$$x_{k+1} = \bar{A}x_k + \bar{B}u_k + Ky_k. \quad (18)$$

From this and (6), another I/O date equation is obtained.

$$\boldsymbol{y}_i = C\bar{A}^{i-1}\mathcal{K}_p\mathcal{Z}_p^- + C\mathcal{K}_i\mathcal{Z}_i + D\boldsymbol{u}_i + \boldsymbol{e}_i + C\bar{A}^{p+i}\mathcal{X}_{-p}, \quad (19)$$

where $\boldsymbol{y}_i$, $\boldsymbol{e}_i$, $\mathcal{Z}_i$ and $\mathcal{K}_i$ are given by

$$\boldsymbol{y}_i = \boldsymbol{\mathcal{Y}}_{i|i}, \quad \boldsymbol{e}_i = \boldsymbol{\mathcal{E}}_{i|i}, \tag{20}$$
$$\mathcal{Z}_i = \boldsymbol{\mathcal{Z}}_{0|i-1}, \tag{21}$$
$$\mathcal{K}_i := \boldsymbol{\mathcal{C}}_i(\bar{A}, [\bar{B}, K]). \tag{22}$$

Define Π_i and $\Pi_i^\perp$ as

$$\Pi_i = \Pi_{\mathcal{Z}_{(i)}} = \mathcal{Z}_{(i)}^\top (\mathcal{Z}_{(i)} \mathcal{Z}_{(i)}^\top)^{-1} \mathcal{Z}_{(i)}, \tag{23}$$
$$\Pi_i^\perp = \Pi_{\mathcal{Z}_{(i)}}^\perp = I - \Pi_i, \tag{24}$$

where $\mathcal{Z}_{(i)} = [(\mathcal{Z}_p^-)^\top, \mathcal{Z}_i^\top, \boldsymbol{u}_i^\top]^\top$ and $\boldsymbol{u}_i = \boldsymbol{\mathcal{U}}_{i|i}$.

From the above, the estimate of the innovations process is defined as

$$\hat{\boldsymbol{e}}_i = \boldsymbol{y}_i \Pi_i^\perp \tag{25}$$
$$= \boldsymbol{e}_i - \boldsymbol{e}_i \Pi_i + C \bar{A}^{i+p} \mathcal{X}_{-p} \Pi_i^\perp. \tag{26}$$

The second term in the r.h.s. of (26) is of order $1/\sqrt{N}$ because innovations process does not correlate with the past input and output. The third term in the r.h.s. of (26) might cause an asymptotic bias. An estimate of $\mathcal{E}_f$ is defined as

$$\hat{\mathcal{E}}_f = [\hat{\boldsymbol{e}}_0^\top, \ldots, \hat{\boldsymbol{e}}_{f-1}^\top]^\top \in \mathbb{R}^{f\ell \times N}. \tag{27}$$

Decompose $\hat{\mathcal{E}}_f$ as $\hat{\mathcal{E}}_f = \mathcal{E}_f + \tilde{\mathcal{E}}_f$, then, $\tilde{\mathcal{E}}_f$ is given by

$$\tilde{\mathcal{E}}_f = \underbrace{\begin{bmatrix} \boldsymbol{e}_0 \Pi_0 \\ \vdots \\ \boldsymbol{e}_{f-1} \Pi_{f-1} \end{bmatrix}}_{\tilde{\mathcal{E}}_{f1}} + \underbrace{\begin{bmatrix} C \bar{A}^p \mathcal{X}_{-p} \Pi_0^\perp \\ \vdots \\ C \bar{A}^{f-1+p} \mathcal{X}_{-p} \Pi_{f-1}^\perp \end{bmatrix}}_{\tilde{\mathcal{E}}_{f2}}. \tag{28}$$

Remark 1. The estimate $\hat{\mathcal{E}}_f$ above can be obtained numerically as follows. Compute the LQ decomposition as

$$\begin{bmatrix} \mathcal{Z}_p^- \\ \mathcal{Z}_f \end{bmatrix} = L_a Q_a^\top,$$

where $\mathcal{Z}_f = \boldsymbol{\mathcal{Z}}_{0|f-1}$, L_a is a lower triangle matrix, and $Q_a^\top Q_a = I_{(p+f)(m+\ell)}$. Let $\mathcal{G}_a = \boldsymbol{\mathcal{G}}_{-p|f-1}$ be a block Hankel matrix composed of $g_k = [f_k^\top, e_k^\top]^\top$. Then, its estimate is given as

$$\hat{\mathcal{G}}_a = M_a Q_a^\top,$$

where $M_a = \text{block-diag}(L_{a,0}, \ldots, L_{a,p+f})$, and $L_{a,i} = L_a(i(m+\ell)+1 : (i+1)(m+\ell), i(m+\ell)+1 : (i+1)(m+\ell))$ using MATLAB notation. Thus, $\hat{\mathcal{E}}_f$ is given as

$$\hat{\mathcal{E}}_f = [0_{f(m+\ell)\times p(m+\ell)}, I_f \otimes [0_{\ell\times m}, I_\ell]] M_a Q_a^\top. \tag{29}$$

4. PROPOSED METHOD

In this section, an estimation method of the innovations model is proposed. The estimates of A and C matrices are the same as Closed-Loop MOESP (van der Veen et al. (2010, 2013)). Borrowing an idea from ordinary MOESP (Verhaegen and Dewilde (1992)), an estimation method of B and D is proposed. It is also analyzed why the estimates above reduces the asymptotic bias. Finally, the Kalman gain K and the covariance of the innovations process Ω_e are estimated by extending and applying the previously proposed method by Ikeda and Tanaka (2017, 2019) to the closed loop identification.

4.1 Estimation of A and C

Numerical simulation shows that Closed-Loop MOESP (van der Veen et al. (2010, 2013)) gives accurate estimate of (A, C) compared to PBSID (Chiuso and Picci (2005)) when the data length is large and the future horizon f and past horizon p are not so large; we will show numerical simulation results in Sec. 5. In Closed-Loop MOESP, the extended observability matrix $\mathcal{O}_f$ is first estimated by projecting $\mathcal{Y}_f$ onto the complement of the space spanned by $[\mathcal{U}_f^\top, \hat{\mathcal{E}}_f^\top]^\top$ as

$$\mathcal{Y}_f \Pi^\perp_{\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}} = \mathcal{O}_f \mathcal{X}_0 \Pi^\perp_{\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}} - \mathcal{H}_f \tilde{\mathcal{E}}_f \Pi^\perp_{\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}}, \tag{30}$$

where $\Pi^\perp_{\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}}$ is a projection defined by

$$\Pi^\perp_{\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}} = I - \begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}^\top \left(\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix} \begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}^\top \right)^{-1} \begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}. \tag{31}$$

Because $\boldsymbol{e}_0, \ldots \boldsymbol{e}_{f-1}$ are uncorrelated to $\mathcal{X}_0$, $\tilde{\mathcal{E}}_{f1}$ in (28) will not cause an asymptotic bias. On the other hand, $\hat{\mathcal{E}}_{f2}$ causes an asymptotic bias. However, estimating $\mathcal{O}_f$ from the equation above, the bias term will be reduced as follows. Let

$$\bar{\mathcal{X}}_{-p}^\perp = \frac{1}{f} \sum_{i=0}^{f-1} \mathcal{X}_{-p} \Pi_i^\perp.$$

Adding and subtracting $\bar{\mathcal{X}}_{-p}^\perp$ to and from $\mathcal{X}_{-p}\Pi_i^\perp$, we obtain

$$\tilde{\mathcal{E}}_{f2} = \bar{\mathcal{O}}_f \bar{A}^p \bar{\mathcal{X}}_{-p}^\perp + \underbrace{\begin{bmatrix} C \bar{A}^p (\mathcal{X}_{-p} \Pi_0^\perp - \bar{\mathcal{X}}_{-p}^\perp) \\ \vdots \\ C \bar{A}^{f-1+p} (\mathcal{X}_{-p} \Pi_{f-1}^\perp - \bar{\mathcal{X}}_{-p}^\perp) \end{bmatrix}}_{\tilde{\mathcal{E}}_{f3}}. \tag{32}$$

From the fact that

$$\mathcal{O}_f = \mathcal{H}_f \bar{\mathcal{O}}_f, \tag{33}$$

the first term in the r.h.s. of (32) pre-multiplied by $\mathcal{H}_f$ has the same range of the first term in the r.h.s. of (30). Thus, it does not cause an asymptotic bias. The second term in the r.h.s. of (32), $\tilde{\mathcal{E}}_{f3}$, causes an asymptotic bias. However, since

$$\sum_{i=1}^{f-1} \left\| \mathcal{X}_{-p} \Pi_i^\perp - \bar{\mathcal{X}}_{-p}^\perp \right\|_F^2 \le \sum_{i=1}^{f-1} \left\| \mathcal{X}_{-p} \Pi_i^\perp \right\|_F^2, \tag{34}$$

it is expected that the bias caused by $\tilde{\mathcal{E}}_{f3}$ is small compared to the one caused by $\tilde{\mathcal{E}}_{f2}$. This effect will be seen in the numerical example in Sec. 5

Remark 2. The projection $\mathcal{Y}_f \Pi^\perp_{\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}}$ is numerically calculated as follows. Compute the LQ decomposition as

$$\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \\ \mathcal{Y}_f \end{bmatrix} = \begin{bmatrix} L_{11} & & \\ L_{21} & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} Q_1^\top \\ Q_2^\top \\ Q_3^\top \end{bmatrix}, \tag{35}$$

then, the projection is given by $\mathcal{Y}_f \Pi^\perp_{\begin{bmatrix} \hat{\mathcal{E}}_f \\ \mathcal{U}_f \end{bmatrix}} = L_{33} Q_3^\top$.

The extended observability matrix $\mathcal{O}_f$ is estimated by using the singular value decomposition (SVD) of L_{33} as

$$L_{33} = [\hat{U}_n, \hat{U}_n^{\perp}] \begin{bmatrix} \hat{S}_n & \\ & \hat{S}_n^{\perp} \end{bmatrix} \begin{bmatrix} \hat{V}_n^{\top} \\ (\hat{V}_n^{\perp})^{\top} \end{bmatrix} \sim \hat{U}_n \hat{S}_n \hat{V}_n^{\top}. \quad (36)$$

The estimate of $\mathcal{O}_f$ is defined as $\hat{\mathcal{O}}_f = \hat{U}_n$. The system matrices A and C are estimated from the shift invariance property as

$$\hat{C} = \hat{\mathcal{O}}_f(1:\ell,:), \quad (37)$$

$$\hat{A} = \hat{\mathcal{O}}_f(1:(f-1)\ell,:)^{\dagger}\hat{\mathcal{O}}_f(\ell+1:f\ell,:). \quad (38)$$

using MATLAB notation.

4.2 Estimation of B and D

In Closed-Loop MOESP, B, D, K and the initial state x_0 are estimated by solving a least squares problem. However, $\tilde{\mathcal{E}}_{f2}$ instead of the second term of the r.h.s. of (32) will cause an asymptotic bias. To reduce this asymptotic bias, $(\hat{\mathcal{O}}_f^{\perp})^{\top} = (\hat{U}_n^{\perp})^{\top}$ and $\Pi_{\hat{\mathcal{E}}_f}^{\perp} = I - \hat{\mathcal{E}}_f^{\top}(\hat{\mathcal{E}}_f\hat{\mathcal{E}}_f^{\top})^{-1}\hat{\mathcal{E}}_f$ are pre- and post-multiplied to the I/O data equation (9) and regressed to $\mathcal{U}_f\Pi_{\hat{\mathcal{E}}_f}^{\perp}$ as

$$(\hat{\mathcal{O}}_f^{\perp})^{\top}\mathcal{Y}_f\Pi_{\hat{\mathcal{E}}_f}^{\perp}(\mathcal{U}_f\Pi_{\hat{\mathcal{E}}_f}^{\perp})^{\dagger} = (\hat{\mathcal{O}}_f^{\perp})^{\top}\mathcal{T}_f$$
$$+(\hat{\mathcal{O}}_f^{\perp})^{\top}(\mathcal{O}_f\mathcal{X}_0 + \mathcal{H}_f\mathcal{E}_f)\Pi_{\hat{\mathcal{E}}_f}^{\perp}(\mathcal{U}_f\Pi_{\hat{\mathcal{E}}_f}^{\perp})^{\dagger} \quad (39)$$

The second term in the r.h.s. of (39) is an error term containing some bias. However, it is expected that the magnitude of the bias is reduced as in the bias of $\hat{\mathcal{O}}_f$.

In practice, B and D are estimated by solving the following least squares problem:

$$\underset{(\hat{B},\hat{D})}{\text{minimize}} \left\| (\hat{\mathcal{O}}_f^{\perp})^{\top}L_{32}L_{22}^{-1} - (\hat{\mathcal{O}}_f^{\perp})^{\top}\hat{\mathcal{S}}_f(I_f \otimes \begin{bmatrix}\hat{B}\\ \hat{D}\end{bmatrix}) \right\|_F, \quad (40)$$

where $\hat{\mathcal{S}}_f = \boldsymbol{\mathcal{T}}_f(\hat{A}, [I_n, 0_{n\times\ell}], \hat{C}, [0_{\ell\times n}, I_\ell])$.

4.3 Estimation of K and Ω_e

Estimate of $\mathcal{Y}_f$ is defined as $\hat{\mathcal{Y}}_f = \hat{\mathcal{T}}_f\mathcal{U}_f$ where $\hat{\mathcal{T}}_f = \boldsymbol{\mathcal{T}}_f(\hat{A},\hat{B},\hat{C},\hat{D})$. Squared sum of the residuals is defined

$$\hat{\Gamma} = \frac{1}{N}(\mathcal{Y}_f - \hat{\mathcal{Y}}_f)(\mathcal{Y}_f - \hat{\mathcal{Y}}_f)^{\top} \quad (41)$$
$$= \frac{1}{N}\mathcal{H}_f\mathcal{E}_f\mathcal{E}_f^{\top}\mathcal{H}_f^{\top} + \frac{1}{N}\mathcal{O}_f\mathcal{X}_0\mathcal{X}_0^{\top}\mathcal{O}_f^{\top} + \delta, \quad (42)$$

where δ is given by

$$\delta = \frac{1}{N}\Big\{\mathcal{O}_f\mathcal{X}_0\mathcal{E}_f^{\top}\mathcal{H}_f^{\top} + \mathcal{H}_f\mathcal{E}_f\mathcal{X}_0^{\top}\mathcal{O}_f^{\top} + \tilde{\mathcal{T}}_f\mathcal{U}_f\mathcal{U}_f^{\top}\tilde{\mathcal{T}}_f^{\top}$$
$$+\tilde{\mathcal{T}}_f\mathcal{U}_f\mathcal{E}_f^{\top}\mathcal{H}_f^{\top} + \mathcal{H}_f\mathcal{E}_f\mathcal{U}_f^{\top}\tilde{\mathcal{T}}_f^{\top}$$
$$+\tilde{\mathcal{T}}_f\mathcal{U}_f\mathcal{X}_0^{\top}\mathcal{O}_f^{\top} + \mathcal{O}_f\mathcal{X}_0\mathcal{U}_f^{\top}\tilde{\mathcal{T}}_f^{\top}\Big\}, \quad (43)$$

$$\tilde{\mathcal{T}}_f = \mathcal{T}_f - \hat{\mathcal{T}}_f. \quad (44)$$

The first term of the r.h.s. of (42) converges in probability to the following Γ as $N \to \infty$:

$$\Gamma = \mathcal{S}_f(I_f \otimes \Xi)\mathcal{S}_f^{\top}, \quad (45)$$

where $\mathcal{S}_f$ and Ξ are given by

$$\mathcal{S}_f = \boldsymbol{\mathcal{T}}_f(A, [I_n, 0_{n\times\ell}], C, [0_{\ell\times n}, I_\ell]), \quad (46)$$

$$\Xi = \begin{bmatrix} K\Omega_e K^{\top} & K\Omega_e \\ \Omega_e K^{\top} & \Omega_e \end{bmatrix}. \quad (47)$$

Pre- and post-multiplying $(\hat{\mathcal{O}}_f^{\perp})^{\top}$ and $\hat{\mathcal{O}}_f^{\perp}$, respectively, to $\hat{\Gamma}$, terms multiplied by $\mathcal{O}_f$ in (42) will be reduced. For simplicity, we assume that $[\hat{\mathcal{O}}_f, \hat{\mathcal{O}}_f^{\perp}]$ is an orthogonal matrix $[\hat{U}_n, \hat{U}_n^{\perp}]$. Thus, the estimate of Ξ is obtained by solving the following semi-definite programming problem.

$$\underset{\hat{\Xi}=\hat{\Xi}^{\top}\geq 0}{\text{minimize}} \left\| (\hat{\mathcal{O}}_f^{\perp})^{\top}(\hat{\mathcal{S}}_f(I_f\otimes\hat{\Xi})\hat{\mathcal{S}}_f^{\top} - \hat{\Gamma})\hat{\mathcal{O}}_f^{\perp} \right\|_F, \quad (48)$$

$$\text{subject to } \hat{\Theta} = \hat{\mathcal{O}}_f^{\top}(\hat{\Gamma} - \hat{\mathcal{S}}_f(I_f\otimes\hat{\Xi})\hat{\mathcal{S}}_f^{\top})\hat{\mathcal{O}}_f \geq 0. \quad (49)$$

Note that $\hat{\Theta}$ is an estimate of the covariance matrix of the state which must be positive semi-definite.

The solution $\hat{\Xi}$ is not an estimate of Ξ but an estimate of $\begin{bmatrix} Q & S \\ S^{\top} & R \end{bmatrix}$, where $Q \in \mathcal{R}^{n\times n}$ is a covariance of the process noise, $R \in \mathcal{R}^{\ell\times\ell}$ is a covariance of the measurement noise, and $S \in \mathcal{R}^{n\times\ell}$ is a cross covariance of the process noise and the measurement noise. To obtain the estimates of the Kalman gain K and the covariance of the innovations process Ω_e, we have to solve a Kalman filter problem. Divide $\hat{\Xi} = \begin{bmatrix} \hat{Q} & \hat{S} \\ \hat{S}^{\top} & \hat{R} \end{bmatrix}$, where $\hat{Q}$, $\hat{S}$, and $\hat{R}$ have the same dimension as Q, S, and R, respectively. Calculate the stabilizing solution $\hat{X}$ of the following Riccati equation:

$$\hat{X} = \hat{A}\hat{X}\hat{A}^{\top} + \hat{Q} - (\hat{A}\hat{X}\hat{C}^{\top} + \hat{S})$$
$$\cdot(\hat{C}\hat{X}\hat{C}^{\top} + \hat{R})^{-1}(\hat{A}\hat{X}\hat{C}^{\top} + \hat{S})^{\top}. \quad (50)$$

Then, the estimates of K and Ω_e are defined as

$$\hat{K} = (\hat{A}\hat{X}\hat{C}^{\top} + \hat{S})(\hat{C}\hat{X}\hat{C}^{\top} + \hat{\mathcal{R}})^{-1}, \quad (51)$$

$$\hat{\Omega}_e = \hat{C}\hat{X}\hat{C}^{\top} + \hat{R}. \quad (52)$$

Remark 3. Ikeda and Tanaka (2019) analyzed a condition for the problem above to be solved. If the future horizon f is selected such that $f \geq 2\lceil n/\ell\rceil + 1$, $\hat{K}$ and $\hat{\Omega}_e$ are uniquely determined for almost all $[C^{\top}, A^{\top}] \in \mathcal{R}^{(\ell+n)\times n}$. It is also shown that in the open loop environment, $\hat{K}$ and $\hat{\Omega}_e$ are consistently estimated by solving the SDP problem and the Kalman filter problem above when the system matrices (A, B, C, D) are consistently estimated by using PO-MOESP method.

5. NUMERICAL EXAMPLE

Consider the following 4-th order system:

$$x_{k+1} = \begin{bmatrix} 1.0135 & 0.0033 & 0.6026 & 0.1179 \\ 0.4154 & 1.1336 & 0.3614 & 0.3942 \\ -0.4166 & -0.4398 & 0.3449 & 0.2093 \\ 0.0237 & -0.1478 & -0.2212 & 0.8222 \end{bmatrix} x_k$$
$$+ \begin{bmatrix} -0.5345 \\ -1.1675 \\ 0.6770 \\ 1.7008 \end{bmatrix} u_k + \begin{bmatrix} -0.4442 & -0.3588 \\ 0.3814 & -0.5855 \\ -0.2308 & 0.1455 \\ 0.0633 & -0.0680 \end{bmatrix} e_k,$$

$$y_k = \begin{bmatrix} -0.8018 & 0.5371 & -0.1816 & 0.1890 \\ -0.2673 & -0.5838 & -0.7430 & -0.1890 \end{bmatrix} x_k + e_k,$$

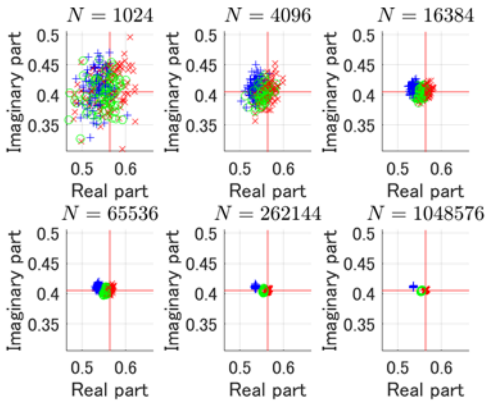

Fig. 1. Eigenvalue of $\hat{A} - \hat{K}\hat{C}$ for $f = p = 8$. $\times$: Proposed method, $+$: Closed-Loop MOESP, $\circ$: PBSID

where the innovations process e_k is given by a Gaussian random process as $e_k \sim N(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}8.7745 & 0.2255\\0.2255 & 6.6895\end{bmatrix})$. The system above is controlled by the following controller:

$$x_{c,k+1} = \begin{bmatrix} 0.5721 & -0.0059 & 0.2571 & 0.1908 \\ 0.5878 & 0.5033 & -0.0004 & 0.3356 \\ -0.5762 & -0.1825 & 0.4088 & 0.2084 \\ 0.0226 & -0.0996 & -0.2660 & 0.6166 \end{bmatrix} x_{c,k} - \begin{bmatrix} -0.4442 & -0.3588 \\ 0.3814 & -0.5855 \\ -0.2308 & 0.1455 \\ 0.0633 & -0.0680 \end{bmatrix} y_k + \begin{bmatrix} 0.0001121 \\ -0.0000917 \\ 0.0002494 \\ 0.0002080 \end{bmatrix} f_k,$$

$$u_k = [0.0198\ -0.0716\ 0.0034\ 0.1063]\, x_{c,k} + f_k,$$

where $f_k \sim N(0, 4.0003)$ is a Gaussian random process. The proposed method is compared with Closed-Loop MOESP (van der Veen et al. (2010, 2013)) and PBSID (Chiuso and Picci (2005)). Hundred times trials of estimation are performed for the data length $N = 2^{10} \sim 2^{20}$ and the future and past horizons $f = p = 8$. Fig. 1 shows the estimated $\lambda(\hat{A} - \hat{K}\hat{C})$. For larger N, the proposed method gives estimates closer to the true value than other two methods.

Fig. 2 shows MSE (mean squared error) and variance of $\lambda(\hat{A})$. The estimate $\hat{A}$ of the proposed method, which is the same as the one of Closed-Loop MOESP, gives better performance than the one of PBSID when $N = 2^{18}$ and $N = 2^{20}$.

MSE and variance of Markov parameters $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ is plotted in Fig. 3. MSE and variance of the proposed estimate are almost the same while those of Closed-Loop MOESP and PBSID are apart from each other. This means that the proposed estimate goes to the true value as N goes to large, while the estimates of Closed-Loop MOESP and PBSID converge to some value apart from the true value.

Figs. 4 and 5 show MSE's and variances of $\lambda(\hat{A} - \hat{K}\hat{C})$ and $\hat{\Omega}$, respectively. Similar tendency as in Fig. 3 can be seen.

49

Figs. 6 and 7 show the MSE's and variances of Markov parameters of the system $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ and of $\lambda(\hat{A} - \hat{K}\hat{C})$ when $f = p = 12$, respectively. This result means that

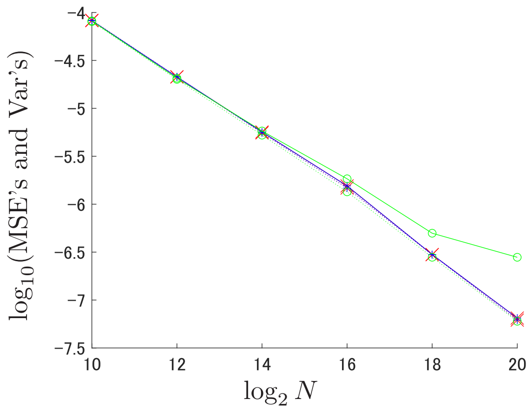

Fig. 2. MSE's (—) and Var's $(\cdots)$ of $\lambda(\hat{A})$ for $f = p = 8$. $\times$: Proposed method, $+$: Closed-Loop MOESP, $\circ$: PBSID

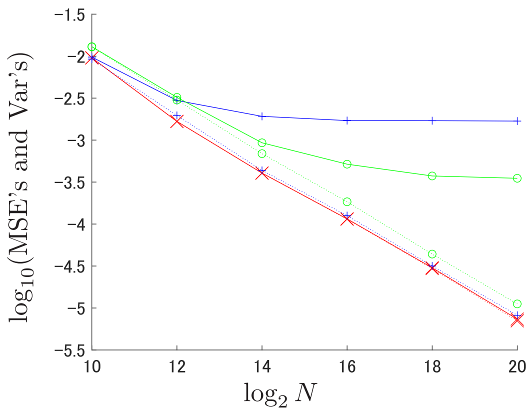

Fig. 3. MSE's (—) and Var's $(\cdots)$ of Markov parameters of the system $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ for $f = p = 8$. $\times$: Proposed method, $+$: Closed-Loop MOESP, $\circ$: PBSID

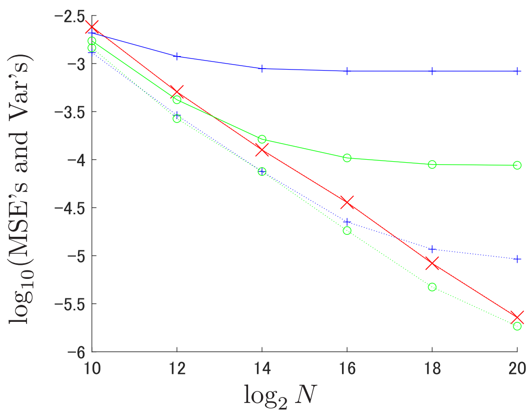

Fig. 4. MSE's (—) and Var's $(\cdots)$ of $\lambda(\hat{A} - \hat{K}\hat{C})$ for $f = p = 8$. $\times$: Proposed method, $+$: Closed-Loop MOESP, $\circ$: PBSID

taking large f and p, Closed-Loop MOESP and PBSID give better performance. However, the proposed method achieves better performance with lower f and p.

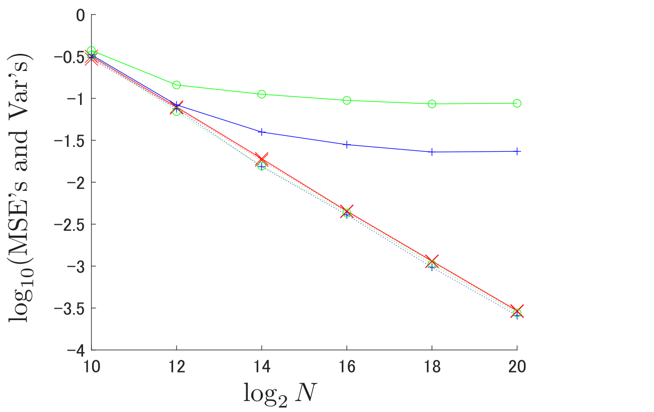

Fig. 5. MSE's (—) and Var's ($\cdots$) of $\hat{\Omega}_e$ for $f = p = 8$. ×: Proposed method, +: Closed-Loop MOESP, ○: PBSID

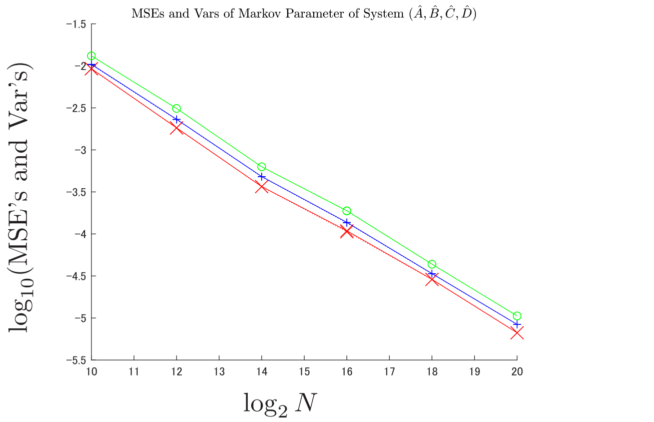

Fig. 6. MSE's (—) and Var's ($\cdots$) of Markov parameters of the system $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ for $f = p = 12$. ×: Proposed method, +: Closed-Loop MOESP, ○: PBSID

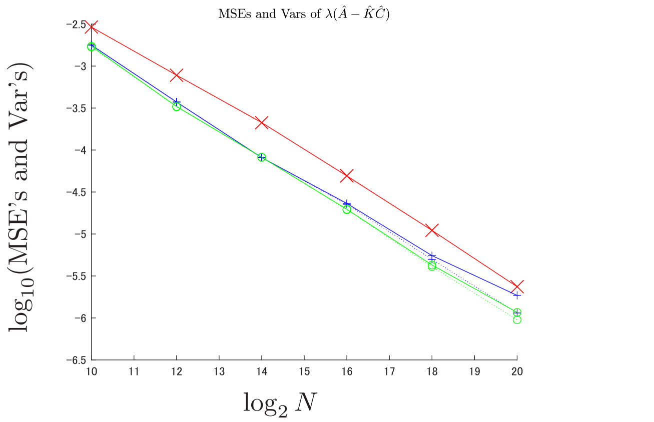

Fig. 7. MSE's (—) and Var's ($\cdots$) of $\lambda(\hat{A} - \hat{K}\hat{C})$ for $f = p = 12$. ×: Proposed method, +: Closed-Loop MOESP, ○: PBSID

6. CONCLUSION

Modifying the Closed-Loop MOESP (van der Veen et al. (2010, 2013)), a new estimation method of the innovations model is proposed for the closed loop environment. It is based on the innovations estimate $\hat{\mathcal{E}}_f$ which is estimated from the finite interval of data. instead of the infinite interval of data. Thus, $\hat{\mathcal{E}}_f$ has an asymptotic bias. In spite of this asymptotic bias, an asymptotic bias of $(\hat{A}, \hat{C})$ in Closed-Loop MOESP is shown to be small from the error analysis. In order to take advantage of the accuracy of $(\hat{A}, \hat{C})$ in Closed-Loop MOESP, B and D matrices are estimated by applying ordinary MOESP (Verhaegen and Dewilde (1992)) to $\mathcal{Y}_f \Pi^{\perp}_{\hat{\mathcal{E}}_f}$. For the estimation of K and Ω_e, an SDP problem formulation is introduced which is originally proposed by Ikeda and Tanaka (2017, 2019) for open loop environment. Numerical simulations illustrate that the proposed method gives better performance when N is large and f and p are selected low.

It is not presented in this paper that MSE's and variances of the estimates in the proposed method become large when the plant is unstable, though they become small as N goes to large as in the case of stable plant. This problem is left to the future research.

REFERENCES

B. D. O. Anderson and M. Gevers. Identifiability of Linear Stochastic Systems Operating Under Linear Feedback. *Automatica*, 18(2): 195–213, 1982.

B. D. O. Anderson and J .B. Moore. *Optimal Filtering*. originally published by Prentice-Hall in 1979, Dover, 2005.

A. Chiuso and G. Picci. Consistency analysis of some closed-loop subspace identification methods. *Automatica*, 41:377–391, 2005.

U. Forssell and L. Ljung. Closed-loop identification rivisited. *Automatica*, 35(7):1215–1241, 1999.

G. H. Golub and C. F. van Loan. *Matrix Computations (2nd Ed.)*. The Johns Hopkins University Press, 1989.

K. Ikeda and H. Tanaka. An SDP formulation for consistent estimate of innovations model. In *11th ASCC*, pages 1772–1777, Gold Coast, December 2017.

K. Ikeda and H. Tanaka. On the uniqueness of the estimate of innovations model. In *12th ASCC*, pages 1313–1318, Kitakyushu, 5 2019.

T. Knudsen. Consistency analysis of subspace identification methods based on a linear regression approach. *Automatica*, 37(1):81–89, 2001.

G. Mercère, I. Markovsky, and J. A. Ramos. Innovation-based subspace identification in open- and closed-loop. In *55th IEEE CDC*, pages 2951–2956, Las Vegas, 12 2016.

H. Tanaka and K. Ikeda. Closed-loop subspace identification taking initial state into account. In *SYSID 2018*, pages 604–609, Stockholm, 2018.

G. van der Veen, J.W. van Wingerden, and M. Verhaegen. Closed-loop MOESP subspace model identification with parametrisable disturbances. In *47th IEEE CDC*, pages 2813–2818, 2010.

G. van der Veen, J.W. van Wingerden, M. Bergamasco, M. Lovera, and M. Verhaegen. Closed-loop subspace Subspace identification methods: an overview. *IET Control Theory and Applications*, 7: 1339–1358, 2013.

P. Van Overschee and B. De Moor. N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems. *Automatica*, 30(1):75–93, 1994.

M. Verhaegen. Identification of the Deterministic Part of MIMO State Models given in Innovations Form from Input-Output Data. *Automatica*, 30:61–74, 1994.

M. Verhaegen and P. Dewilde. Subspace model identification Part 1. The output-error state-space model identification class of algorithms. *International Journal of Control*, 56:1187–1210, 1992.

M. Verhaegen and V. Verdult. *Filtering and System Identification*. Cambridge University Press, 2007.