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

On the precision of the plant estimates in some subspace identification methods

Kenji Ikeda *

* *The University of Tokushima, Tokushima, 770-8506, Japan (Tel: +81-88-656-7504; e-mail: ikeda@is.tokushima-u.ac.jp).*

Abstract: Explicit formulae of dominant parts of the estimation errors of plant A matrix in PI-MOESP, PO-MOESP, and one of N4SID methods are derived based on a proposing lemma on perturbations to the singular subspaces. By considereing a gap between the true and a perturbed singular subspaces instead of the perturbation on the subspace, a simple formulae of the estimation error is obtained. The order of the subordinate terms are also analysed.

Keywords: Subspace methods; Identification for control; Estimation and filtering.

1. INTRODUCTION

Subspace identification method is remarkably developed and widely applied to real systems. Some famous approaches are CCA(Larimore (1990)),N4SID(Van Overschee and De Moor (1994)), MOESP(Verhaegen and Dewilde (1992a,b); Verhaegen (1993, 1994)), etc., and the asymptotic variances of the estimates in these methods are analyzed and are now emerging. (Bauer et al. (1997, 1999); Bauer and Jansson (2000); Bauer et al. (2000); Bauer and Ljung (2002); Bauer (2005); Scherrer (2002); Chiuso and Picci (2004b); Jansson (2000)).

In the analysis of the variance, Chiuso and Picci (2004a) shows an equivalency between Robust N4SID and PO-MOESP methods but to the best of the author's knowledge, there are no comparison study by using explicit formulae of the estimation errors. The difficulty comes from the complexity of the perturbation formula to the basis of the singular subspaces. For this problem, it reveals that a more simple formula can be given for the perturbations to the singular subspace itself instead of its basis(Ikeda (2013)). In this paper, an explicit formula of the estimation error of A matrix in each of PI-MOESP, PO-MOESP, and one of N4SID methods is to be derived by using a lemma on the perturbations to the singular subspaces.

It is often the case that the purpose of the identification is to obtain a plant model. For such cases, the estimation error on the A matrix of the plant model is also to be analyzed under the assumption that the plant model and the noise model does not have a common pole.

This paper is organized as follows. Section 2 formulates the problem and the assumptions. Section 3 describes some preliminaries on the subspace identification methods while Section 4 analyzes the estimation errors of A matrix in PI-MOESP, PO-MOESP, and one of N4SID methods. Finally Section 5 concludes the paper.

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

Big O notation is adopted to describe the error term in an approximation, *i.e.* the least-significant terms are summarized in a single big O term.

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 a degree of the system. Namely,

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

Let $\boldsymbol{\mathcal{C}}_f(A,B)$ denote an extended controllability matrix as

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

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

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

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

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

2. ARMAX AND OE MODELS

Consider the following innovations (ARMAX) model:

$$x_{k+1} = Ax_k + Bu_k + Ke_k, \tag{5}$$

$$y_k = Cx_k + Du_k + e_k, \tag{6}$$

where $u_k \in \mathrm{R}^m$, $y_k \in \mathrm{R}^l$, $e_k \in \mathrm{R}^l$, and $x_k \in \mathrm{R}^n$ are the input, the output, the noise, and the state, respectively and $A \in \mathrm{R}^{n\times n}$, $B \in \mathrm{R}^{n\times m}$, $C \in \mathrm{R}^{l\times n}$, $D \in \mathrm{R}^{l\times m}$, and $K \in \mathrm{R}^{n\times l}$ are the system matrices to be estimated. The following assumptions are made for this system:

(A1) $|\lambda_i(A)| < 1$, $|\lambda_i(A-KC)| < 1$, $i = 1,\dots,n$.
(A2) The innovations process $\{e_k\}$ is a white Gaussian process with mean $E\{e_k\} = 0$ and covariance $E\{e_k e_l^\top\} = \Omega_{ee}\delta_{kl}$.
(A3) The processes $\{u_k\}$ and $\{e_k\}$ are mutually independent.
(A4) The processes $\{x_k\}$, $\{u_k\}$, and $\{e_k\}$ are ergodic and stationary (Anderson and Moore (2005)).

For the sake of simplicity, the input process $\{u_k\}$ is assumed to be a white Gaussian process as:

(A5) $\{u_k\}$ is a white Gaussian process with mean $E[u_k] = 0$ and covariance matrix $E[u_k u_l^\top] = \Omega_{uu}\delta_{kl}$ where $\Omega_{uu} = \sigma_u^2 I_m$.

From the assumption (A5), the required PE (persistence of excitation) conditions required (Verhaegen and Verdult (2007)) are automatically satisfied.

In this paper, the estimates of A matrices of the plant model are analyzed under the assumption that the plant and noise models do not have a common pole.

(A6) $A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}$, $B = \begin{pmatrix} B_1 \\ 0 \end{pmatrix}$, $K = \begin{pmatrix} 0 \\ K_2 \end{pmatrix}$, and $C = (\, C_1 \quad C_2 \,)$.

According to the decomposition of the matrices above, the state is decomposed as $x_k = (x_{u,k}^\top, x_{e,k}^\top)^\top$.

On the other hand, the OE model under the assumptions above is given by

$$x_{u,k+1} = A_1 x_{u,k} + B_1 u_k, \tag{7}$$
$$y_k = C_1 x_{u,k} + v_k, \tag{8}$$

where

$$v_k = \left[C_2(zI - A_2)^{-1}K_2 + I\right] e_k. \tag{9}$$

3. PRELIMINARIES

3.1 I/O Data Equation

I/O data equation derived from the innovations model (5) and (6) plays important roles in analyzing and implementing subspace identification methods:

$$\mathcal{Y}_f = \mathcal{O}_f \mathcal{X}_0 + \mathcal{T}_f \mathcal{U}_f + \mathcal{H}_f \mathcal{E}_f, \tag{10}$$

where

$$\mathcal{O}_f := \boldsymbol{\mathcal{O}}_f(A, B), \tag{11}$$
$$\mathcal{T}_f := \boldsymbol{\mathcal{T}}_f(A, B, C, D), \tag{12}$$
$$\mathcal{H}_f := \boldsymbol{\mathcal{T}}_f(A, K, C, I), \tag{13}$$
$$\mathcal{X}_i := [\, x_i \quad x_{i+1} \quad \cdots \quad x_{i+N-1} \,], \tag{14}$$
$$\mathcal{U}_f := \boldsymbol{\mathcal{U}}_{0|f-1}. \tag{15}$$

$\mathcal{Y}_f$ and $\mathcal{E}_f$ are defined similarly to $\mathcal{U}_f$.

From the assumption (A6), $\mathcal{O}_f$ is divided as

$$\mathcal{O}_f = [\mathcal{O}_{f1}, \mathcal{O}_{f2}] := [\boldsymbol{\mathcal{O}}_f(A_1, C_1), \boldsymbol{\mathcal{O}}_f(A_2, C_2)], \tag{16}$$

while $\mathcal{T}_f$ and $\mathcal{H}_f$ are given by using smaller matrices:

$$\mathcal{T}_{f1} := \boldsymbol{\mathcal{T}}_f(A_1, B_1, C_1, D) = \mathcal{T}_f, \tag{17}$$
$$\mathcal{H}_{f2} := \boldsymbol{\mathcal{T}}_f(A_2, K_2, C_2, I) = \mathcal{H}_f. \tag{18}$$

3.2 Instrumental Variable Matrix

From the innovations model (5) and (6), the state matrix is given by:

$$\mathcal{X}_0 = \mathcal{X}_0^{(p)} + \bar{A}^p \mathcal{X}_{-p}, \tag{19}$$
$$\mathcal{X}_0^{(p)} = \mathcal{K}_p \mathcal{Z}_p^- \tag{20}$$

where $\mathcal{K}_p := [\boldsymbol{\mathcal{C}}_p(\bar{A}, \bar{B}), \boldsymbol{\mathcal{C}}_p(\bar{A}, K)]$, $\bar{A} := A - KC$, $\bar{B} := B - KD$, $\mathcal{Z}_p^- := [(\mathcal{U}_p^-)^\top, (\mathcal{Y}_p^-)^\top]^\top$, $\mathcal{U}_p^- := \boldsymbol{\mathcal{U}}_{-p|-1} \in \mathrm{R}^{mp\times N}$, and $\mathcal{Y}_p^- := \boldsymbol{\mathcal{Y}}_{-p|-1} \in \mathrm{R}^{lp\times N}$. Thus, the I/O data equation becomes

$$\mathcal{Y}_f = \mathcal{O}_f \mathcal{K}_p \mathcal{Z}_p^- + \mathcal{T}_f \mathcal{U}_f + \mathcal{H}_f \mathcal{E}_f + \mathcal{O}_f \bar{A}^p \mathcal{X}_{-p}. \tag{21}$$

From the assumption (A6),

$$\bar{A}^p = \begin{bmatrix} A_1^p & 0 \\ -\mathcal{C}_{p2}\mathcal{O}_{p1} & \bar{A}_2^p \end{bmatrix}, \tag{22}$$

$$\mathcal{K}_p = \begin{bmatrix} \mathcal{C}_{p1} & \\ -\mathcal{C}_{p2}\mathcal{T}_{p1} & \mathcal{C}_{p2} \end{bmatrix}, \tag{23}$$

where $\bar{A}_2 := A_2 - K_2 C_2$, and $\mathcal{C}_{p1} := \boldsymbol{\mathcal{C}}_p(A_1, B_1)$, $\mathcal{C}_{p2} := \boldsymbol{\mathcal{C}}_p(A_2, K_2)$.

3.3 LQ Decomposition

Subspace identification methods are based on the following LQ decomposition:

$$\begin{bmatrix} \mathcal{U}_f \\ \mathcal{Z}_p^- \\ \mathcal{Y}_f \end{bmatrix} =: \begin{bmatrix} L_{11} & & \\ L_{21} & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} Q_1^\top \\ Q_2^\top \\ Q_3^\top \end{bmatrix}, \tag{24}$$

where Q_i, $i = 1, 2, 3$ is an orthonormal matrix which satisfies $Q_i^\top Q_i = I$ and $Q_i^\top Q_j = 0$ for $i \neq j$. L_{ij} is a matrix with appropriate dimensions.

Let $[\hat{\beta}_u, \hat{\beta}_z]$ be a projection of $\mathcal{Y}_f$ onto $[\mathcal{U}_f^\top, (\mathcal{Z}_p^-)^\top]^\top$, namely,

$$[\, \hat{\beta}_u \quad \hat{\beta}_z \,] = \mathcal{Y}_f \begin{bmatrix} \mathcal{U}_f \\ \mathcal{Z}_p^- \end{bmatrix}^\dagger. \tag{25}$$

Then,

$$\hat{\beta}_u = L_{31}L_{11}^{-1} - L_{32}L_{22}^{-1} \cdot L_{21}L_{11}^{-1}, \tag{26}$$
$$\hat{\beta}_z = L_{32}L_{22}^{-1}. \tag{27}$$

From Eq. (21), $\hat{\beta}_z$ and $\hat{\beta}_u$ become estimates of $\beta_z = \mathcal{O}_f \mathcal{K}_p$ and $\beta_u = \mathcal{T}_f$, respectively and their estimation errors are given by

$$\tilde{\beta}_z = \underbrace{\mathcal{H}_f\mathcal{E}_f(\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp)^\dagger}_{\tilde{\beta}_{z1}} + \underbrace{\mathcal{O}_f\bar{A}^p\mathcal{X}_{-p}(\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp)^\dagger}_{\tilde{\beta}_{z2}}, \quad (28)$$

$$\tilde{\beta}_u = \mathcal{H}_f\mathcal{E}_f(\mathcal{U}_f\Pi_{\mathcal{Z}_p^-}^\perp)^\dagger + O(\bar{A}^p/\sqrt{N}), \quad (29)$$

where

$$\Pi_{\mathcal{U}_f}^\perp = I - \mathcal{U}_f^\top(\mathcal{U}_f\mathcal{U}_f^\top)^{-1}\mathcal{U}_f = I - Q_1Q_1^\top, \quad (30)$$

$$\Pi_{\mathcal{Z}_p^-}^\perp = I - (\mathcal{Z}_p^-)^\top(\mathcal{Z}_p^-(\mathcal{Z}_p^-)^\top)^{-1}\mathcal{Z}_p^-. \quad (31)$$

The error term $\tilde{\beta}_{z1} = O(1/\sqrt{N})$, while $\tilde{\beta}_{z2} = O(\bar{A}^p)$. Furthermore,

$$\lim_{N\to\infty}\tilde{\beta}_{z2} = \mathcal{O}_f\bar{A}^p\Omega_{xx}\mathcal{O}_p^\top\left[\mathcal{O}_p\Omega_{xx}\mathcal{O}_p^\top + \mathcal{H}_p(I\otimes\Omega_{ee})\mathcal{H}_p^\top\right]^{-1} \times [\,\mathcal{T}_p \quad I\,]. \quad (32)$$

This means, $\hat{\beta}_z$ has an asymptotic bias.

3.4 SVD

In order to estimate $\mathcal{O}_f$ or $\mathcal{K}_p$, pre- and post-multiplying appropriate positive definite matrices $\hat{W}_f^+$ and $\hat{W}_p^-$ to $\hat{\beta}_z$ and decompose $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$ into singular spaces as:

$$\hat{W}_f^+\hat{\beta}_z\hat{W}_p^- = \hat{U}_n\hat{\Sigma}_n\hat{V}_n^\top + \hat{R}. \quad (33)$$

Weighting matrices:

$$\hat{W}_f^+ = \left(\hat{\Gamma}_f^{+\Pi}\right)^{-\frac{1}{2}} = \left(\frac{1}{N}\mathcal{Y}_f\Pi_{\mathcal{U}_f}^\perp\mathcal{Y}_f^\top\right)^{-\frac{1}{2}}, \quad (34)$$

$$\hat{W}_p^- = \left(\hat{\Gamma}_p^{-\Pi}\right)^{\frac{1}{2}} = \left(\frac{1}{N}\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp(\mathcal{Z}_p^-)^\top\right)^{\frac{1}{2}}, \quad (35)$$

or identity matrices are often used.

The following lemma on the perturbations to the singular subspaces is useful.

Lemma 1. Ikeda (2013) Let the singular value decomposition of X and $\hat{X} = X + \tilde{X}$ be:

$$X = [\,U_n \quad U_n^\perp\,]\begin{bmatrix}\Sigma_n & \\ & 0\end{bmatrix}[\,V_n \quad V_n^\perp\,]^\top, \quad (36)$$

$$\hat{X} = [\,\hat{U}_n \quad \hat{U}_n^\perp\,]\begin{bmatrix}\hat{\Sigma}_n & \\ & \hat{\Sigma}_n^\perp\end{bmatrix}[\,\hat{V}_n \quad \hat{V}_n^\perp\,]^\top. \quad (37)$$

When $\|\tilde{X}\|$ is smaller enough than the smallest singular value of Σ_n, then

$$(U_n^\perp)^\top\tilde{U}_n = (U_n^\perp)^\top\tilde{X}\hat{V}_n\hat{\Sigma}_n^{-1} \quad (38)$$

$$= (U_n^\perp)^\top\tilde{X}V_n\Sigma_n^{-1} + O(\|\tilde{X}\|^2) \quad (39)$$

$$\tilde{V}_n^\top V_n^\perp = \hat{\Sigma}_n^{-1}\hat{U}_n^\top\tilde{X}V_n^\perp \quad (40)$$

$$= \Sigma_n^{-1}U_n^\top\tilde{X}V_n^\perp + O(\|\tilde{X}\|^2) \quad (41)$$

where $\tilde{U}_n = \hat{U}_n - U_n$, $\tilde{V}_n = \hat{V}_n - V_n$.

Proof: These equations are obtained straightforwardly from the calculations of $(U_n^\perp)^\top\tilde{U}_n$ and $\tilde{V}_n^\top V_n^\perp$.

Remark 2. It is often the case that the gap between the two subspaces spanned by the left (or right) singular subspaces is more important than the perturbation of the subspace $\tilde{U}_n$ (or $\tilde{V}_n$) itself. In such cases, Lemma 1 is useful because taking into account, for example,

$$\tilde{U}_n = U_nU_n^\top\tilde{U}_n + U_n^\perp(U_n^\perp)^\top\tilde{U}_n,$$

$U_n^\top\tilde{U}_n$ just affects the coordinate transformation as is seen in Remark 3 below.

In some of the state approaches, $\hat{\mathcal{K}}_p$ is defined from the right singular subspace by using an appropriate nonsingular matrix T_N as

$$\widehat{\mathcal{O}_f\mathcal{K}}_p = (\hat{W}_f^+)^{-1}\hat{U}_n\hat{\Sigma}_n\hat{V}_n^\top(\hat{W}_p^-)^{-1} \quad (42)$$

$$= (\hat{W}_f^+)^{-1}\hat{U}_n\hat{\Sigma}_nT_N\cdot\underbrace{T_N^{-1}\hat{V}_n^\top(\hat{W}_p^-)^{-1}}_{\hat{\mathcal{K}}_p}. \quad (43)$$

Some typical choises of T_N are $T_N = \hat{\Sigma}_n^{-\frac{1}{2}}$, or $T_N = I$. The choise of T_N determines the coordinate system of the state space representation. Thus, the choise of T_N affects the magnitude of the estimation error, but it does not affect the representation of the error term as is seen in the rest of this paper. The estimate $\hat{\mathcal{K}}_p$ can be decomposed into the signal/noise components as

$$\hat{\mathcal{K}}_p = \underbrace{T_N^{-1}(I + \tilde{V}_n^\top V_n)V_n^\top(\hat{W}_p^-)^{-1}}_{\mathcal{K}_p'} + \underbrace{T_N^{-1}\tilde{V}_n^\top V_n^\perp(V_n^\perp)^\top(\hat{W}_p^-)^{-1}}_{\tilde{\mathcal{K}}_p}, \quad (44)$$

where U_n, Σ_n, and V_n are a SVD of $\hat{W}_f^+\beta_z\hat{W}_p^- = U_n\Sigma_nV_n^\top$.

Remark 3. The difference between $\mathcal{K}_p'$ and $\mathcal{K}_p$ comes from the difference of the coordinate systems: $(A', B', C', D') = (T^{-1}AT = A', T^{-1}B, CT, D)$, for $T = T_N^{-1}(I+\tilde{V}_n^\top V_n)^{-1}T_N$. However, the effect of this difference is small enough because

$$\tilde{\mathcal{K}}_p' = T\tilde{\mathcal{K}}_p = \tilde{\mathcal{K}}_p + O(\tilde{\beta}_z^2). \quad (45)$$

The following lemma gives a formula for $\tilde{\mathcal{K}}_p$ when $\hat{W}_p^-$ is defined as in Eq. (35).

Lemma 4. Under the assumptions (A1) $\sim$ (A5), together with the weighting function $\hat{W}_p^-$ in Eq. (35), the estimation error $\tilde{\mathcal{K}}_p$ defined in Eq. (44) is given by

$$\tilde{\mathcal{K}}_p = (\hat{W}_f^+\mathcal{O}_f)^\dagger\hat{W}_f^+\mathcal{H}_f\mathcal{E}_f\Pi_{\mathcal{U}_f}^\perp(\mathcal{Z}_p^-)^\top\Gamma_p + \bar{A}^p\mathcal{X}_{-p}\Pi_{\mathcal{U}_f}^\perp(\mathcal{Z}_p^-)^\top\Gamma_p + O(\tilde{\beta}_z^2) \quad (46)$$

$$\Gamma_p = (\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp(\mathcal{Z}_p^-)^\top)^{-1} - \mathcal{K}_p^\top(\mathcal{X}_0^{(p)}\Pi_{\mathcal{U}_f}^\perp(\mathcal{X}_0^{(p)})^\top)^{-1}\mathcal{K}_p. \quad (47)$$

Proof: Applying Lemma 1 to $\tilde{\mathcal{K}}_p$ in Eq. (44),

$$\tilde{\mathcal{K}}_p = T_N^{-1}\Sigma_n^{-1}U_n^\top\hat{W}_f^+\tilde{\beta}_z\hat{W}_p^-V_n^\perp(V_n^\perp)^\top(\hat{W}_p^-)^{-1} + O(\tilde{\beta}_z^2). \quad (48)$$

Because $\hat{W}_f^+\mathcal{O}_f = U_n\Sigma_nT_N$,

$$T_N^{-1}\Sigma_n^{-1}U_n^\top = (\hat{W}_f^+\mathcal{O}_f)^\dagger. \quad (49)$$

Substituting $\hat{W}_p^-$ by Eq. (35),

$$
\begin{aligned}
&\hat{W}_p^- V_n^\perp (V_n^\perp)^\top (\hat{W}_p^-)^{-1} \\
&= \hat{W}_p^- (I - V_n V_n^\top)(\hat{W}_p^-)^{-1} \\
&= I - (\hat{W}_p^-)^2 \mathcal{K}_p^\top (\mathcal{K}_p (\hat{W}_p^-)^2 \mathcal{K}_p^\top)^{-1} \mathcal{K}_p \\
&= (\mathcal{Z}_p^- \Pi_{\mathcal{U}_f}^\perp (\mathcal{Z}_p^-)^\top)\Gamma_p. \qquad (50)
\end{aligned}
$$

Substituting (49), (28) and (50) into (48), the lemma is obtained.

Similar to the estimation of $\hat{\mathcal{K}}_p$, $\hat{\mathcal{O}}_f$ is estimated from the left singular subspace by using an appropriate nonsingular matrix T_M:

$$
\widehat{\mathcal{O}_f \mathcal{K}_p} = \underbrace{(\hat{W}_f^+)^{-1} \hat{U}_n T_M}_{\hat{\mathcal{O}}_f} \cdot T_M^{-1} \hat{\Sigma}_n \hat{V}_n^\top (\hat{W}_p^-)^{-1}. \qquad (51)
$$

$\hat{\mathcal{O}}_f$ is decomposed into the signal/noise components as

$$
\begin{aligned}
\hat{\mathcal{O}}_f &= \underbrace{(\hat{W}_f^+)^{-1} U_n (I + U_n^\top \tilde{U}_n) T_M}_{\mathcal{O}_f'} \\
&\quad + \underbrace{(\hat{W}_f^+)^{-1} U_n^\perp (U_n^\perp)^\top \tilde{U}_n T_M}_{\tilde{\mathcal{O}}_f}. \qquad (52)
\end{aligned}
$$

The estimation error of $\hat{\mathcal{O}}_f$ is given by the following lemma.

Lemma 5. Under the same assumptions in Lemma 4, the estimation error $\tilde{\mathcal{O}}_f$ defined in Eq. (52) is given by

$$
\begin{aligned}
\tilde{\mathcal{O}}_f &= (\hat{W}_f^+)^{-1} \Pi_{\mathcal{O}_f^\top \hat{W}_f^+}^\perp \hat{W}_f^+ \mathcal{H}_f \mathcal{E}_f (\mathcal{K}_p \mathcal{Z}_p^- \Pi_{\mathcal{U}_f}^\perp)^\dagger \\
&\quad + O(\tilde{\beta}_{z1}^2), \qquad (53) \\
\Pi_{\mathcal{O}_f^\top \hat{W}_f^+}^\perp &= I - \hat{W}_f^+ \mathcal{O}_f (\mathcal{O}_f^\top (\hat{W}_f^+)^2 \mathcal{O}_f)^{-1} \mathcal{O}_f^\top \hat{W}_f^+. \qquad (54)
\end{aligned}
$$

Proof: Using $V_n \Sigma_n^{-1} T_M = (\mathcal{K}_p \hat{W}_p^-)^\dagger$ and Lemma 1, $\tilde{\mathcal{O}}_f$ is given by

$$
\begin{aligned}
\tilde{\mathcal{O}}_f &= (\hat{W}_f^+)^{-1} \Pi_{\mathcal{O}_f^\top \hat{W}_f^+}^\perp \hat{W}_f^+ \mathcal{H}_f \mathcal{E}_f (\mathcal{Z}_p^- \Pi_{\mathcal{U}_f}^\perp)^\dagger \\
&\quad \times (\hat{W}_p^-)^2 \mathcal{K}_p^\top (\mathcal{K}_p (\hat{W}_p^-)^2 \mathcal{K}_p^\top)^{-1} + O(\tilde{\beta}_{z1}^2). \qquad (55)
\end{aligned}
$$

Note that the term $\tilde{\beta}_{z2}$ disappears because $\Pi_{\mathcal{O}_f^\top \hat{W}_f^+}^\perp \hat{W}_f^+ \mathcal{O}_f = 0$. Adopting (35) for $\hat{W}_p^-$, the lemma is obtained.

When $\hat{W}_f^+ = I$,

$$
\begin{aligned}
\tilde{\mathcal{O}}_f &= \Pi_{\mathcal{O}_f^\top}^\perp \mathcal{H}_f \mathcal{E}_f (\mathcal{K}_p \mathcal{Z}_p^- \Pi_{\mathcal{U}_f}^\perp)^\dagger + O(\tilde{\beta}_{z1}^2), \qquad (56) \\
\Pi_{\mathcal{O}_f^\top}^\perp &= I - \mathcal{O}_f (\mathcal{O}_f^\top \mathcal{O}_f)^{-1} \mathcal{O}_f^\top. \qquad (57)
\end{aligned}
$$

In the following, $\hat{W}_f^+ = I$ will be adopted.

4. ESTIMATION OF A MATRIX

4.1 State Approach

In this paper, the following regression is adopted:

$$
\begin{bmatrix} \hat{A} & [B, \widehat{0, \ldots, 0}] \\ \hat{C} & [D, \widehat{0, \ldots, 0}] \end{bmatrix} = \begin{bmatrix} \hat{\mathcal{X}}_1 \\ \mathcal{Y}_{0|0} \end{bmatrix} \begin{bmatrix} \hat{\mathcal{X}}_0 \\ \mathcal{U}_f \end{bmatrix}^\dagger, \qquad (58)
$$

where $\hat{\mathcal{X}}_0 = \hat{\mathcal{K}}_p \mathcal{Z}_p^-$, $\hat{\mathcal{X}}_1 = \hat{\mathcal{K}}_p \mathcal{Z}_p^{\leftarrow}$, $\mathcal{Z}_p^{\leftarrow} = \mathcal{Z}_p^- J_N$, $J_N = \begin{bmatrix} 0 & 0_{1\times 1} \\ I_{N-1} & 0 \end{bmatrix}$. A regression onto $\begin{bmatrix} \hat{\mathcal{X}}_0 \\ \mathcal{U}_{0|0} \end{bmatrix}$ is adopted in the ordinary N4SID. However, the difference is $O(1/N)$ under the assumption (A5) and does not affect the analysis of the dominant parts of the estimation errors.

The following theorem gives the estimation error $\tilde{A} = \hat{A} - A' = \hat{A} - T^{-1}AT$ in the state approach where T is a transformation matrix in Remark 3.

Theorem 6. Under the assumptions (A1) $\sim$ (A5), together with $\hat{\mathcal{K}}_p$ in (43), the weighting functions $\hat{W}_p^-$ in Eq. (35) and $\hat{W}_f^+ = I$, the estimation error of $\hat{A}$ in the state approach (58) is given by

$$
\begin{aligned}
\tilde{A} &= \mathcal{O}_f^\dagger \mathcal{H}_f \mathcal{E}_f \Pi_{\mathcal{U}_f}^\perp (\mathcal{Z}_p^-)^\top \Big[-\mathcal{K}_p^\top (\mathcal{X}_0^{(p)} \Pi_{\mathcal{U}_f}^\perp (\mathcal{X}_0^{(p)})^\top)^{-1} A \\
&\quad + \begin{bmatrix} J_p^\top \otimes I_m & \\ & J_p^\top \otimes I_l \end{bmatrix} \mathcal{K}_p^\top (\mathcal{X}_0^{(p)} \Pi_{\mathcal{U}_f}^\perp (\mathcal{X}_0^{(p)})^\top)^{-1} \Big] \\
&\quad + K \mathcal{E}_1 (\mathcal{X}_0^{(p)} \Pi_{\mathcal{U}_f}^\perp)^\dagger + O(\lambda^{2p}) + O(\frac{\lambda^p}{\sqrt{N}}) + O(\frac{1}{N}). \qquad (59)
\end{aligned}
$$

where $\lambda = \max\{\rho(A), \rho(\bar{A})\}$ and $\rho(A)$ is a spectral radius of A.

Proof: See Appendix A.

Remark 7. In order for $O(\lambda^{2p})$ to be $O(1/N)$, the past horizon p must be taken as $p = -\frac{1}{2\log\lambda}\log N$. When the sampling becomes faster and λ approaches to 1, very large p will be required. If $N = 10000$ and $\lambda = 0.98$, the required p is $p = 228$. If p is less than the half of its required value, $O(\lambda^{2p})$ terms might be greater than $O(1/\sqrt{N})$ terms.

Next, the estimation error of the plant model part will be analyzed. If the nonsingular matrix T_N is known, $\tilde{A}_1 = [I, O]\tilde{A}[I, O]^\top$ can be calculated directly. Thus, the following theorem is obtained.

Theorem 8. In addition to the assumptions in Theorem 6, assume (A6). Then, the estimation error of the plant model $\tilde{A}_1 = [I, 0]\tilde{A}[I, 0]^\top$ in the state approach is given by

$$
\begin{aligned}
\tilde{A}_1 &= \left[\mathcal{O}_f^\dagger\right]_1 \mathcal{H}_f \mathcal{E}_f \Pi_{\mathcal{U}_f}^\perp (\mathcal{U}_p^-)^\top \\
&\quad \times \Big[-\mathcal{C}_{p1}^\top (\mathcal{X}_{u0}^{(p)} \Pi_{\mathcal{U}_f}^\perp (\mathcal{X}_{u0}^{(p)})^\top)^{-1} A_1 \\
&\quad + (J^\top \otimes I) \mathcal{C}_{p1}^\top (\mathcal{X}_{u0}^{(p)} \Pi_{\mathcal{U}_f}^\perp (\mathcal{X}_{u0}^{(p)})^\top)^{-1} \Big], \\
&\quad + O(\lambda^{2p}) + O(\frac{\lambda^p}{\sqrt{N}}) + O(\frac{1}{N}), \qquad (60)
\end{aligned}
$$

where

$$
\left[\mathcal{O}_f^\dagger\right]_1 = \mathcal{O}_{f1}^\dagger - \mathcal{O}_{f1}^\dagger \mathcal{O}_{f2} (\mathcal{O}_{f2}^\top \Pi_{\mathcal{O}_{f1}^\top}^\perp \mathcal{O}_{f2})^{-1} \mathcal{O}_{f2}^\top \Pi_{\mathcal{O}_{f1}^\top}^\perp, \qquad (61)
$$

$$
\Pi_{\mathcal{O}_{f1}^\top}^\perp = I - \mathcal{O}_{f1} (\mathcal{O}_{f1}^\top \mathcal{O}_{f1})^{-1} \mathcal{O}_{f1}^\top, \qquad (62)
$$

$$
\Pi_{\mathcal{O}_{f2}^\top}^\perp = I - \mathcal{O}_{f2} (\mathcal{O}_{f2}^\top \mathcal{O}_{f2})^{-1} \mathcal{O}_{f2}^\top, \qquad (63)
$$

$$
\mathcal{X}_{u0}^{(p)} = \mathcal{C}_{p1} \mathcal{U}_p^-. \qquad (64)
$$

Proof: This theorem is a straight consequence of Theorem 6 and the following equation:

$$(\mathcal{O}_f^\top \mathcal{O}_f)^{-1} = \begin{bmatrix} (\mathcal{O}_{f1}^\top \Pi^\perp_{\mathcal{O}_{f2}^\top} \mathcal{O}_{f1})^{-1} & -\mathcal{O}_{f1}^\dagger \mathcal{O}_{f2}(\mathcal{O}_{f2}^\top \Pi^\perp_{\mathcal{O}_{f1}^\top} \mathcal{O}_{f2})^{-1} \\ * & (\mathcal{O}_{f2}^\top \Pi^\perp_{\mathcal{O}_{f1}^\top} \mathcal{O}_{f2})^{-1} \end{bmatrix}. \tag{65}$$

4.2 Shift Invariance Approach

In the shift invariance approach, $\hat{A}$ is defined by

$$\hat{A} = \underline{\hat{\mathcal{O}}_f}^\dagger \overline{\hat{\mathcal{O}}_f} \tag{66}$$

$$= \underbrace{\underline{\mathcal{O}_f}^\dagger \overline{\mathcal{O}_f}}_{A} + \underbrace{\underline{\mathcal{O}_f}^\dagger (\overline{\tilde{\mathcal{O}}_f} - \underline{\tilde{\mathcal{O}}_f} A) + O(1/N)}_{\tilde{A}}, \tag{67}$$

where $\underline{\hat{\mathcal{O}}_f} = [I, 0]\hat{\mathcal{O}}_f$ and $\overline{\hat{\mathcal{O}}_f} = [0, I]\hat{\mathcal{O}}_f$. It is also used that $\tilde{\mathcal{O}}_f = O(1/\sqrt{N})$. The following theorem gives a formula of $\tilde{A}$ under the same assumptions in Theorem 6.

Theorem 9. Under the assumptions (A1) $\sim$ (A5), together with $\hat{\mathcal{O}}_f$ in (51), the weighting functions $\hat{W}_p^-$ in Eq. (35) and $\hat{W}_f^+ = I$, the estimation error of $\hat{A}$ in the shift invariance approach (66) is given by

$$\tilde{A} = \left\{[0, \mathcal{O}_{f-1}^\dagger] - A\mathcal{O}_f^\dagger\right\} \mathcal{H}_f \mathcal{E}_f (\mathcal{X}_0^{(p)} \Pi^\perp_{\mathcal{U}_f})^\dagger + O(\lambda^f/\sqrt{N}) + O(1/N). \tag{68}$$

Proof: From Lemma (5) and $\underline{\mathcal{O}_f}^\dagger \Pi^\perp_{\underline{\mathcal{O}_f^\top}} = O(A^f)$,

$$\tilde{A} = \underline{\mathcal{O}_f}^\dagger \overline{\tilde{\mathcal{O}}_f} + O(\lambda^f/\sqrt{N}) + O(1/N).$$

The 1st term in the R.H.S. of the equation above is calculated as

$$\underline{\mathcal{O}_f}^\dagger \overline{\tilde{\mathcal{O}}_f} = \mathcal{O}_{f-1}^\dagger ([0_{(f-1)l \times l}, I] - \mathcal{O}_{f-1} A (\mathcal{O}_f^\top \mathcal{O}_f)^{-1} \mathcal{O}_f^\top) \times \mathcal{H}_f \mathcal{E}_f (\mathcal{X}_0^{(p)} \Pi^\perp_{\mathcal{U}_f})^\dagger + O(\frac{1}{N}). \tag{69}$$

This proves the lemma.

As in the state approach, the estimation error of the plant model $\tilde{A}_1 = [I, O]\tilde{A}[I, O]^\top$ is given by the following theorem.

Theorem 10. In addition to the assumptions in Theorem 9, assume (A6). Then, the estimation error of the plant model $\tilde{A}_1 = [I, 0]\tilde{A}[I, 0]^\top$ in the shift invariance approach is given by

$$\tilde{A}_1 = \left([0, [\mathcal{O}_{f-1}^\dagger]_1] - A_1 [\mathcal{O}_f^\dagger]_1\right) \mathcal{H}_f \mathcal{E}_f (\mathcal{X}_{u0}^{(p)} \Pi^\perp_{\mathcal{U}_f})^\dagger + O(\frac{\lambda^f}{\sqrt{N}}) + O(\frac{1}{N}). \tag{70}$$

4.3 PI-MOESP Method

In the PI-MOESP method, $\hat{\mathcal{O}}_{f1}$ is defined by using an appropriate nonsingular matrix T_I as

$$\widehat{\mathcal{O}_{f1}\mathcal{C}_{p1}} = \mathcal{Y}_f \Pi^\perp_{\mathcal{U}_f} (\mathcal{U}_p^-)^\top (\mathcal{U}_p^- \Pi^\perp_{\mathcal{U}_f} (\mathcal{U}_p^-)^\top)^{-1} \tag{71}$$

$$= \underbrace{(\hat{W}_f^+)^{-1} \hat{U}_n T_I}_{\hat{\mathcal{O}}_{f1}} \cdot T_I^{-1} \hat{\Sigma}_n \hat{V}_n^\top (\hat{W}_p^-)^{-1} + \hat{R} \tag{72}$$

The estimated extended observability matrix is decomposed into the following signal/noise components

$$\hat{\mathcal{O}}_{f1} = \underbrace{(\hat{W}_f^+)^{-1} U_n (I + U_n^\top \tilde{U}_n) T_I}_{\mathcal{O}'_{f1}} + \underbrace{(\hat{W}_f^+)^{-1} U_n^\perp (U_n^\perp)^\top \tilde{U}_n T_I}_{\tilde{\mathcal{O}}_{f1}}. \tag{73}$$

On the other hand, $\widehat{\mathcal{O}_{f1}\mathcal{C}_{p1}}$ is given by

$$\widehat{\mathcal{O}_{f1}\mathcal{C}_{p1}} = \mathcal{O}_{f1}\mathcal{C}_{p1} + \tilde{\beta}, \tag{74}$$

where

$$\tilde{\beta} = \mathcal{H}_f \mathcal{E}_f \Pi^\perp_{\mathcal{U}_f} (\mathcal{U}_p^-)^\top (\mathcal{U}_p^- \Pi^\perp_{\mathcal{U}_f} (\mathcal{U}_p^-)^\top)^{-1} + \mathcal{O}_{f2}\mathcal{C}_{p2}\mathcal{H}_p \mathcal{E}_p^- \Pi^\perp_{\mathcal{U}_f} (\mathcal{U}_p^-)^\top (\mathcal{U}_p^- \Pi^\perp_{\mathcal{U}_f} (\mathcal{U}_p^-)^\top)^{-1}. \tag{75}$$

Thus, the following lemma is obtained.

Lemma 11. Under the assumptions (A1) $\sim$ (A6) together with the weighting matrices $\hat{W}_f^+ = I$ and $\hat{W}_p^- = \left(\frac{1}{N}\mathcal{U}_p^- \Pi^\perp_{\mathcal{U}_f} (\mathcal{U}_p^-)^\top\right)^{\frac{1}{2}}$, the estimation error of the extended observability matrix in PI-MOESP method in (73) is given by

$$\tilde{\mathcal{O}}_{f1} = \Pi^\perp_{\mathcal{O}_{f1}^\top} \mathcal{H}_f \mathcal{E}_f (\mathcal{X}_{u0}^{(p)} \Pi^\perp_{\mathcal{U}_f})^\dagger + \Pi^\perp_{\mathcal{O}_{f1}^\top} \mathcal{O}_{f2}\mathcal{C}_{p2}\mathcal{H}_p \mathcal{E}_p^- (\mathcal{X}_{u0}^{(p)} \Pi^\perp_{\mathcal{U}_f})^\dagger + O(\frac{1}{N}) \tag{76}$$

Proof: Because $T_I^{-1}\Sigma_n V_n^\top = \mathcal{C}_{p1}\hat{W}_p^-$, $(\mathcal{C}_{p1}\hat{W}_p^-)^\dagger = V_n \Sigma_n^{-1} T_I$. From this, $U_n^\perp (U_n^\perp)^\top = \Pi^\perp_{\mathcal{O}_f^\top \hat{W}_f^+}$, and Lemma 1,

$$\tilde{\mathcal{O}}_{f1} = (\hat{W}_f^+)^{-1} \Pi^\perp_{\mathcal{O}_f^\top \hat{W}_f^+} \hat{W}_f^+ \tilde{\beta} \hat{W}_p^- (\mathcal{C}_{p1}\hat{W}_p^-)^\dagger + O(\tilde{\beta}^2). \tag{77}$$

Substituting $\hat{W}_f^+$ and $\hat{W}_p^-$ proves the lemma.

The estimation error of $\hat{A}_1$ in PI-MOESP method is given by the following theorem.

Theorem 12. Under the same assumptions in Lemma 11, the estimation error of $\hat{A}_1$ in PI-MOESP method is given by

$$\tilde{A}_1 = \left([0, \mathcal{O}_{f-1,1}^\dagger] - A_1 \mathcal{O}_{f1}^\dagger\right) \left[\mathcal{H}_f \mathcal{E}_f (\mathcal{X}_{u0}^{(p)} \Pi^\perp_{\mathcal{U}_f})^\dagger + \mathcal{O}_{f2}\mathcal{C}_{p2}\mathcal{H}_p \mathcal{E}_p^- (\mathcal{X}_{u0}^{(p)} \Pi^\perp_{\mathcal{U}_f})^\dagger\right] + O(\frac{\lambda^f}{\sqrt{N}}) + O(\frac{1}{N}) \tag{78}$$

Proof: Similar to the proof of Theorem 9, so omitted here.

5. CONCLUSION

An explicit formula of the dominant part of the estimation error on A matrix in each of PI-MOESP, PO-MOESP, and one of N4SID methods is derived. As reported in Ikeda (2013), the difference between the estimation errors in $\hat{A}_1$'s in PO-MOESP and PI-MOESP methods are that not only $\mathcal{E}_f \Pi^\perp_{\mathcal{U}_f}(\mathcal{U}_p^-)^\top$ but also $\mathcal{E}_p \Pi^\perp_{\mathcal{U}_f}(\mathcal{U}_p^-)^\top$ are counted as noise terms in PI-MOESP method and that $\mathcal{O}^\dagger_{f1}$ and $\mathcal{O}^\dagger_{f-1,1}$ in PI-MOESP method are replaced by the 1-1 blocks of $\mathcal{O}^\dagger_f$ and $\mathcal{O}^\dagger_{f-1}$ in PO-MOESP method. The difference between the dominant noise terms in N4SID and PO-MOESP methods lies in the difference of the terms pre- and post-multiplied to $\mathcal{H}_f \mathcal{E}_f \Pi^\perp_{\mathcal{U}_f}(\mathcal{U}_p^-)^\top$, whose orders are the same when $f = p$. However, in N4SID, there is an asymptotic bias in $\hat{\mathcal{K}}_p$ of order $O(\lambda^{2p})$.

REFERENCES

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

Bauer, D. (2005). Asymptotic properties of subspace estimators. *Automatica*, 41, 395–376.

Bauer, D., Deistler, M., and Scherrer, W. (1997). The analysis of the asymptotic variance of subspace algorithms. In *Proceedings of SYSID'97*, 1087–1091.

Bauer, D., Deistler, M., and Scherrer, W. (1999). Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. *Automatica*, 35, 1243–1254.

Bauer, D., Deistler, M., and Scherrer, W. (2000). On the impact of weighting matrices in subspace algorithms. In *Proceedings of the IFAC conference SYSID'00*, 97–102. Santa Barbara.

Bauer, D. and Jansson, M. (2000). Analysis of the asymptotic properties of the MOESP type of subspace algorithms. *Automatica*, 36, 497–509.

Bauer, D. and Ljung, L. (2002). Some facts about the choise of the weighting matrices in Larimore type of subspace algorithms. *Automatica*, 38, 763–773.

Chiuso, A. and Picci, G. (2004a). On the ill-conditioning of subspace identification with inputs. *Automatica*, 40(4), 575–589.

Chiuso, A. and Picci, G. (2004b). The asymptotic variance of subspace estimates. *Journal of Econometrics*, 118(1-2), 257–291.

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

Ikeda, K. (2013). Asymptotic Properties of MOESP-type Methods. In *Preprints of the 11th IFAC International Workshop on Adaptation and Learning in Contgrol and Signal Processing*, 92–97.

Jansson, M. (2000). Asymptotic Variance Analysis of Subspace Identification Methods. In *Proceedings of the IFAC conference SYSID'00*, 91–96. Santa Barbara.

Larimore, W.E. (1990). Canonical Variate Analysis in Identification, Filtering, and Adaptive Control. In *Proceedings of the 29th Conference on Decision and Control*, 596–604. Honolulu, Hawaii.

Scherrer, W. (2002). Local Optimality of Minimum Phase Balanced Truncation. In *Proceedings of the 15th IFAC World Congress*, 440–454.

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

Verhaegen, M. (1993). Subspace model identification Part 3. Analysis of the ordinary output-error state-space model identification algorithm. *International Journal of Control*, 58, 555–586.

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

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

Verhaegen, M. and Dewilde, P. (1992b). Subspace model identification Part 2. Analysis of the elementary output-error state-space model identification algorithms. *International Journal of Control*, 56, 1211–1241.

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

Appendix A. PROOF OF THEOREM 6

From Eqs. (5) and (19),

$$\hat{\mathcal{X}}_0 = \mathcal{X}_0 - \bar{A}^p \mathcal{X}_{-p} + \tilde{\mathcal{K}}_p \mathcal{Z}_p^-, \tag{A.1}$$

$$\hat{\mathcal{X}}_1 = \mathcal{X}_1 - [0,\ldots,0,x_N] - \bar{A}^p \mathcal{X}_{-p+1} + \tilde{\mathcal{K}}_p \mathcal{Z}_p^{\leftarrow} + \bar{A}^p[0,\ldots,0,x_{N-p+1}]. \tag{A.2}$$

From the equation above, the following equation is obtained:

$$\hat{\mathcal{X}}_1 = A\hat{\mathcal{X}}_0 + B\mathcal{U}_{0|0} + \mathcal{W}_1, \tag{A.3}$$

where

$$\mathcal{W}_1 = \tilde{\mathcal{K}}_p \mathcal{Z}_p^{\leftarrow} + K\mathcal{E}_{0|0} - A\tilde{\mathcal{K}}_p \mathcal{Z}_p^- + A\bar{A}^p \mathcal{X}_{-p} - \bar{A}^p \mathcal{X}_{-p+1} - [0,\ldots,0,x_N] + \bar{A}^p[0,\ldots,0,x_{N-p+1}]. \tag{A.4}$$

The estimation error of $\hat{A}$ in the state approach is given by

$$\tilde{A} = \mathcal{W}_1(\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f})^\dagger. \tag{A.5}$$

It will be claimed that the major part of $\tilde{A}$ is composed of $(\tilde{\mathcal{K}}_p \mathcal{Z}_p^{\leftarrow} + K\mathcal{E}_{0|0})(\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f})^\dagger$ while the other terms in (A.5) will be involved in $O(\tilde{\beta}_z^2) = O(\lambda^{2p}) + O(\lambda^p/\sqrt{N}) + O(1/N)$ terms.

Because $\Gamma_p \mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f}(\mathcal{Z}_p^-)^\top \mathcal{K}_p = 0$, the 3rd term of $\mathcal{W}_1(\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f})^\dagger$ is given by

$$A\tilde{\mathcal{K}}_p \mathcal{Z}_p^-(\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f})^\dagger = O(\tilde{\beta}_z^2). \tag{A.6}$$

As for the 4th term of $\mathcal{W}_1(\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f})^\dagger$,

$$\begin{aligned} & A\bar{A}^p \mathcal{X}_{-p} \Pi^\perp_{\mathcal{U}_f} \hat{\mathcal{X}}_0 (\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f} \hat{\mathcal{X}}_0^\top)^{-1} \\ &= A\bar{A}^p \mathcal{X}_{-p} \Pi^\perp_{\mathcal{U}_f} (\mathcal{K}_p \mathcal{Z}_p^- + \tilde{\mathcal{K}}_p \mathcal{Z}_p^-)^\top (\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f} \hat{\mathcal{X}}_0^\top)^{-1} \\ &= O(\lambda^p/\sqrt{N}) + O(\lambda^{2p}) + O(\lambda^p \tilde{\beta}_z) \end{aligned} \tag{A.7}$$

The 5th term of $\mathcal{W}_1(\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f})^\dagger$ can be analyzed in a similar way to the 4th term. It is easy to see the 6th and 7th terms of $\mathcal{W}_1(\hat{\mathcal{X}}_0 \Pi^\perp_{\mathcal{U}_f})^\dagger$ is $O(1/N)$.

Finally, by using the following equations:

$$\begin{aligned} & \lim_{N\to\infty} (\mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f} (\mathcal{Z}_p^-)^\top)^{-1} \mathcal{Z}_p^{\leftarrow} \Pi^\perp_{\mathcal{U}_f} (\mathcal{Z}_p^-)^\top \\ &= \begin{bmatrix} J_p^\top \otimes I_m & \\ & J_p^\top \otimes I_l \end{bmatrix} + \begin{bmatrix} -\mathcal{T}_p^\top (\hat{\Gamma}_p^{+\Pi})^{-1} \mathcal{O}_p \\ (\hat{\Gamma}_p^{+\Pi})^{-1} \mathcal{O}_p \end{bmatrix} \\ & \times [B\Omega_{uu},\ 0_{n\times pm},\ A\Omega_{xx}C^\top + K\Omega ee, 0_{n\times pl}], \end{aligned}$$

and

$$\left[B\Omega_{uu},\ 0_{n\times pm},\ A\Omega_{xx}C^\top + K\Omega ee, 0_{n\times pl}\right]\mathcal{K}_p = O(\lambda^p),$$

the lemma is obtained.