21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Error Analysis of MOESP Type Algorithm

Kenji Ikeda[1]

***Abstract*— Explicit formulae of dominant parts of the estimation errors of A and C matrices in PO-MOESP are derived based on a proposing lemma on perturbations to the singular subspaces. The formulae are derived directly from a slightly modified PO-MOESP algorithm. Based on the formulae, the effect of the weighting matrices in the singular value decomposition is analyzed. It is also shown that the original PO-MOESP method is optimal in a sense that singular values of some matrix in the Frobenius norm of the estimation error are minimized.**

I. INTRODUCTION

Subspace identification method is remarkably developed and widely applied to real systems. Some famous approaches are CCA([13]),N4SID([17]), MOESP([20], [21], [18], [19]), etc., and the asymptotic variances of the estimates in these methods are analyzed and are now emerging. ([3], [4], [6], [5], [7], [2], [14], [9], [12]).

In the analysis of the variance, [12] derives a formula of the variance of the estimates in the MOESP type algorithm, while [9] derives more comprehensive formulae of the variance by using a kind of state approach. It is also shown in [8] on some equivalence relation between the robust N4SID and MOESP type methods. But to the best of the author's knowledge, there are no direct derivation of a comprehensive formulae of the variance in MOESP type methods. The difficulty comes from the complexity of the perturbation formula to the singular vectors. For this problem, it reveals that a more simple formula can be given for the perturbations to the singular subspace instead of the singular vectors([11]). The advantage comes from considering a gap between two singular subspaces and the fact that the perturbation on the estimates of the system matrices can be analyzed by using the gap. In this paper, an explicit formula of the estimation errors of A and C matrices in PO-MOESP method is to be derived by using a lemma on the perturbations to the singular subspaces.

In the estimation of A and C matrices, weighting matrices $\hat{W}_f^+$ and $\hat{W}_p^-$ are introduced as in [7]. The effect of $\hat{W}_p^-$ is anlyzed by using the obtained formulae and the estimation error will be compared to that in the original PO-MOESP method.

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 gives formulae of the estimation errors of A and C matrices in PO-MOESP methods. Section 5 analyzes the effect of the weighting matrix based on the obtained formulae. Finally Section 6 concludes the paper.

Notations: Let $X^\dagger$ denote a pseudo inverse (Moore-Penrose generalized inverse) of X ([10]).
Let a spectral radius of a matrix A is denoted by $\rho(A) = \max_i(|\lambda_i(A)|)$ where $\lambda_i(A)$ is an eigenvalue of A.
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 $\mathcal{O}_f(A,C)$ 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,

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

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

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

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

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

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

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

II. ARMAX MODEL

Consider the following innovations (ARMAX) model:

$$x_{k+1} = Ax_k + Bu_k + Ke_k, \quad (5)$$
$$y_k = Cx_k + Du_k + e_k, \quad (6)$$

where $u_k \in \mathrm{R}^m$, $y_k \in \mathrm{R}^l$, $e_k \in \mathrm{R}^l$, 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) $|\lambda_i(A)| < 1$, $|\lambda_i(A - KC)| < 1$, $i = 1, \ldots, n$.
(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.

[1]Kenji Ikeda is with Faculty of Engineering, The University of Tokushima, 2-1 Minamijosanjima, Tokushima Japan ikeda@is.tokushima-u.ac.jp

ISBN: 978-90-367-6321-9

(A4) The processes $\{x_k\}$, $\{u_k\}$, and $\{e_k\}$ are ergodic and stationary ([1]).

(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 ([22]) are automatically satisfied.

III. Preliminaries

A. 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:

$$\mathcal{Y}_f = \mathcal{O}_f \mathcal{X}_0 + \mathcal{T}_f \mathcal{U}_f + \mathcal{H}_f \mathcal{E}_f, \tag{7}$$

where

$$\mathcal{O}_f := \boldsymbol{\mathcal{O}}_f(A, B), \tag{8}$$
$$\mathcal{T}_f := \boldsymbol{\mathcal{T}}_f(A, B, C, D), \tag{9}$$
$$\mathcal{H}_f := \boldsymbol{\mathcal{T}}_f(A, K, C, I), \tag{10}$$
$$\mathcal{X}_i := [\, x_i \quad x_{i+1} \quad \cdots \quad x_{i+N-1} \,], \tag{11}$$
$$\mathcal{U}_f := \boldsymbol{\mathcal{U}}_{0|f-1}. \tag{12}$$

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

B. Instrumental Variable Matrix

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

$$\mathcal{X}_0 = \mathcal{X}_0^{(p)} + \bar{A}^p \mathcal{X}_{-p}, \tag{13}$$
$$\mathcal{X}_0^{(p)} = \mathcal{K}_p \mathcal{Z}_p^- \tag{14}$$

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

$$\mathcal{Y}_f = \mathcal{O}_f \mathcal{K}_p \mathcal{Z}_p^- + \mathcal{T}_f \mathcal{U}_f + \mathcal{H}_f \mathcal{E}_f + \mathcal{O}_f \bar{A}^p \mathcal{X}_{-p} \tag{15}$$

C. LQ Decomposition

Subspace identification methods are based on the following LQ decomposition:

$$\begin{bmatrix} \mathcal{U}_f \\ \mathcal{Z}_p^- \\ \mathcal{Y}_f \end{bmatrix} =: \begin{bmatrix} L_{11} & & \\ L_{21} & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} Q_1^\top \\ Q_2^\top \\ Q_3^\top \end{bmatrix}, \tag{16}$$

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,

$$[\, \hat{\beta}_u \quad \hat{\beta}_z \,] = \mathcal{Y}_f \begin{bmatrix} \mathcal{U}_f \\ \mathcal{Z}_p^- \end{bmatrix}^\dagger. \tag{17}$$

Then,

$$\hat{\beta}_u = L_{31} L_{11}^{-1} - L_{32} L_{22}^{-1} \cdot L_{21} L_{11}^{-1}, \tag{18}$$
$$\hat{\beta}_z = L_{32} L_{22}^{-1}. \tag{19}$$

From Eq. (15), $\hat{\beta}_z$ 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

$$\tilde{\beta}_z = \underbrace{\mathcal{H}_f \mathcal{E}_f (\mathcal{Z}_p^- \Pi_{\mathcal{U}_f}^\perp)^\dagger}_{\tilde{\beta}_{z1}} + \underbrace{\mathcal{O}_f \bar{A}^p \mathcal{X}_{-p} (\mathcal{Z}_p^- \Pi_{\mathcal{U}_f}^\perp)^\dagger}_{\tilde{\beta}_{z2}}, \tag{20}$$
$$\tilde{\beta}_u = \mathcal{H}_f \mathcal{E}_f (\mathcal{U}_f \Pi_{\mathcal{Z}_p^-}^\perp)^\dagger + O(\bar{A}^p/\sqrt{N}), \tag{21}$$

where

$$\Pi_{\mathcal{U}_f}^\perp = I - \mathcal{U}_f^\top (\mathcal{U}_f \mathcal{U}_f^\top)^{-1} \mathcal{U}_f = I - Q_1 Q_1^\top, \tag{22}$$
$$\Pi_{\mathcal{Z}_p^-}^\perp = I - (\mathcal{Z}_p^-)^\top (\mathcal{Z}_p^- (\mathcal{Z}_p^-)^\top)^{-1} \mathcal{Z}_p^-. \tag{23}$$

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

$$\begin{aligned} \lim_{N\to\infty} \tilde{\beta}_{z2} = {} & \mathcal{O}_f \bar{A}^p \Omega_{xx} \mathcal{O}_p^\top \\ & \times \left[\mathcal{O}_p \Omega_{xx} \mathcal{O}_p^\top + \mathcal{H}_p (I \otimes \Omega_{ee}) \mathcal{H}_p^\top \right]^{-1} \\ & \times [\, \mathcal{T}_p \quad I \,]. \end{aligned} \tag{24}$$

This term will cause an asymptotic bias in the estimation of $\mathcal{K}_p$ while it will be neglected in the estimation of $\mathcal{O}_f$.

D. SVD

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

$$\hat{W}_f^+ \hat{\beta}_z \hat{W}_p^- = \hat{U}_n \hat{Z}_n \hat{V}_n^\top + \hat{R}. \tag{25}$$

Weighting matrices:

$$\hat{W}_f^+ = \left(\hat{\Gamma}_f^{+\Pi}\right)^{-\frac{1}{2}} := \left(\frac{1}{N} \mathcal{Y}_f \Pi_{\mathcal{U}_f}^\perp \mathcal{Y}_f^\top\right)^{-\frac{1}{2}}, \tag{26}$$
$$\hat{W}_p^- = \left(\hat{\Gamma}_p^{-\Pi}\right)^{\frac{1}{2}} := \left(\frac{1}{N} \mathcal{Z}_p^- \Pi_{\mathcal{U}_f}^\perp (\mathcal{Z}_p^-)^\top\right)^{\frac{1}{2}}, \tag{27}$$

or identity matrices are often used.

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

Lemma 1: [11] Let X and $\hat{X} = X + \tilde{X}$ have the singular value decompositions as:

$$X = [\, U_n \quad U_n^\perp \,] \begin{bmatrix} \Sigma_n & \\ & 0 \end{bmatrix} [\, V_n \quad V_n^\perp \,]^\top, \tag{28}$$
$$\hat{X} = [\, \hat{U}_n \quad \hat{U}_n^\perp \,] \begin{bmatrix} \hat{\Sigma}_n & \\ & \hat{\Sigma}_n^\perp \end{bmatrix} [\, \hat{V}_n \quad \hat{V}_n^\perp \,]^\top. \tag{29}$$

Then, for any SVD's of X and $\hat{X}$, the following equations hold

$$(U_n^\perp)^\top \tilde{U}_n = (U_n^\perp)^\top \tilde{X} \hat{V}_n \hat{\Sigma}_n^{-1} \tag{30}$$
$$\tilde{V}_n^\top V_n^\perp = \hat{\Sigma}_n^{-1} \hat{U}_n^\top \tilde{X} V_n^\perp \tag{31}$$

where $\tilde{U}_n = \hat{U}_n - U_n$, $\tilde{V}_n = \hat{V}_n - V_n$. Furthermore, when $\|\tilde{X}\|$ is smaller enough than the smallest singular value of Σ_n, then there exist SVD's (U_n, Σ_n, V_n) of X and $([\hat{U}_n, \hat{U}_n^\perp], \mathrm{diag}(\hat{\Sigma}_n, \hat{\Sigma}_n^\perp), [\hat{V}_n, \hat{V}_n^\perp])$ of $\hat{X}$ such that

$$(U_n^\perp)^\top \tilde{U}_n = (U_n^\perp)^\top \tilde{X} V_n \Sigma_n^{-1} + O(\|\tilde{X}\|^2), \tag{32}$$
$$\tilde{V}_n^\top V_n^\perp = \Sigma_n^{-1} U_n^\top \tilde{X} V_n^\perp + O(\|\tilde{X}\|^2). \tag{33}$$

Proof: See the appendix.

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

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

$U_n^\top \tilde{U}_n$ does not affect the left singular subspace.

In the shift invariance approach, $\hat{\mathcal{O}}_f$ is estimated from the left singular subspace by using an appropriate nonsingular matrix T_M:

$$\widehat{\mathcal{O}_f \mathcal{K}_p} = \underbrace{(\hat{W}_f^+)^{-1} \hat{U}_n T_M}_{\hat{\mathcal{O}}_f} \cdot T_M^{-1} \hat{\Sigma}_n \hat{V}_n^\top (\hat{W}_p^-)^{-1}. \quad (34)$$

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

$$\hat{\mathcal{O}}_f = \underbrace{(\hat{W}_f^+)^{-1} U_n (I + U_n^\top \tilde{U}_n) T_M}_{\mathcal{O}'_f} + \underbrace{(\hat{W}_f^+)^{-1} U_n^\perp (U_n^\perp)^\top \tilde{U}_n T_M}_{\tilde{\mathcal{O}}_f}. \quad (35)$$

The difference between $\mathcal{O}'_f$ and $\mathcal{O}_f$ comes from the difference of the coorinate systems: $(A', B', C', D') = (T^{-1}AT, T^{-1}B, CT, D)$ for $T = T_M^{-1}(I + U_n^\top \tilde{U}_n) T_M$. However, the effect of this difference is small enough because

$$\tilde{\mathcal{O}}'_f = \tilde{\mathcal{O}}_f T = \tilde{\mathcal{O}}_f + O(\tilde{\beta}_z^2). \quad (36)$$

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

Lemma 2: Under the assumptions (A1)$\sim$ (A5), the estimation error $\tilde{\mathcal{O}}_f$ defined in Eq. (35) is given by

$$\begin{aligned}\tilde{\mathcal{O}}_f =& (\hat{W}_f^+)^{-1} \Pi^\perp_{\mathcal{O}_f^\top \hat{W}_f^+} \hat{W}_f^+ \mathcal{H}_f \mathcal{E}_f (\mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f})^\dagger \\ &\times \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top (\mathcal{K}_p \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top)^{-1} \\ &+ O(\tilde{\beta}_{z1}^2), \quad (37)\\ \Pi^\perp_{\mathcal{O}_f^\top \hat{W}_f^+} =& I - \hat{W}_f^+ \mathcal{O}_f (\mathcal{O}_f^\top (\hat{W}_f^+)^\top \hat{W}_f^+ \mathcal{O}_f)^{-1} (\hat{W}_f^+ \mathcal{O}_f)^\top. \quad (38)\end{aligned}$$

Proof: Because $\mathcal{T}_M^{-1} \Sigma_n V_n^\top (\hat{W}_p^-)^{-1} = \mathcal{K}_p$ and $\mathcal{K}_p$ has a full row rank,

$$V_n \Sigma_n^{-1} T_M = (K_p \hat{W}_p^-)^\dagger. \quad (39)$$

From this and Lemma 1, $\tilde{\mathcal{O}}_f$ is given by

$$\tilde{\mathcal{O}}_f = (\hat{W}_f^+)^{-1} U_n^\perp (U_n^\perp)^\top \hat{W}_f^+ \tilde{\beta}_z \hat{W}_p^- (\mathcal{K}_p \hat{W}_p^-)^\dagger. \quad (40)$$

By using $U_n = \hat{W}_f^+ \mathcal{O}_f T_M^{-1}$,

$$\begin{aligned}U_n^\perp (U_n^\perp)^\top =& I - U_n U_n^\top = I - U_n (U_n^\top U_n)^{-1} U_n^\top \\ =& \Pi^\perp_{\mathcal{O}_f^\top \hat{W}_f^+}. \quad (41)\end{aligned}$$

Substituting this into Eq. (40), Eq. (37) is obtained. Note that the term $\tilde{\beta}_{z2}$ disappears because $\Pi^\perp_{\mathcal{O}_f^\top \hat{W}_f^+} \hat{W}_f^+ \mathcal{O}_f = 0$. This proves the lemma.

When $\hat{W}_f^+ = I$ and Eq. (27) for $\hat{W}_p^-$ are adopted,

$$\begin{aligned}\tilde{\mathcal{O}}_f =& \Pi^\perp_{\mathcal{O}_f^\top} \mathcal{H}_f \mathcal{E}_f (\mathcal{K}_p \mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f})^\dagger + O(\tilde{\beta}_{z1}^2), \quad (42)\\ \Pi^\perp_{\mathcal{O}_f^\top} =& I - \mathcal{O}_f (\mathcal{O}_f^\top \mathcal{O}_f)^{-1} \mathcal{O}_f^\top. \quad (43)\end{aligned}$$

IV. Estimation of System Matrices

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

$$\begin{aligned}\hat{A} =& \underline{\hat{\mathcal{O}}_f}^\dagger \overline{\hat{\mathcal{O}}_f} \quad (44)\\ =& \underbrace{\underline{\mathcal{O}_f}^\dagger \overline{\mathcal{O}_f}}_{A} + \underbrace{\underline{\mathcal{O}_f}^\dagger (\overline{\tilde{\mathcal{O}}_f} - \underline{\tilde{\mathcal{O}}_f} A) + O(1/N)}_{\tilde{A}}, \quad (45)\\ \hat{C} =& [I_l, O] \hat{\mathcal{O}}_f \quad (46)\\ =& \underbrace{[I_l, O] \mathcal{O}_f}_{C} + \underbrace{[I_l, 0] \tilde{\mathcal{O}}_f}_{\tilde{C}} \quad (47)\end{aligned}$$

where $\underline{\hat{\mathcal{O}}_f} = [I_{l(f-1)}, 0] \hat{\mathcal{O}}_f$ and $\overline{\hat{\mathcal{O}}_f} = [0, I_{l(f-1)}] \hat{\mathcal{O}}_f$. It is also used that $\tilde{\mathcal{O}}_f = O(1/\sqrt{N})$. The following lemma gives a formula of $\tilde{A}$.

Lemma 3: Under the assumptions (A1) $\sim$ (A5), together with $\hat{\mathcal{O}}_f$ in (34), the estimation errors of $\hat{A}$ and $\hat{C}$ in the shift invariance approach (44) and (46) are given by

$$\begin{aligned}\tilde{A} =& \left\{ [0, \mathcal{O}_{f-1}^\dagger] - A (\hat{W}_f^+ \mathcal{O}_f)^\dagger \hat{W}_f^+ \right\} \mathcal{H}_f \mathcal{E}_f (\mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f})^\dagger \\ &\times \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top (\mathcal{K}_p \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top)^{-1} \\ &+ O(\lambda^f / \sqrt{N}) + O(1/N), \quad (48)\\ \tilde{C} =& \left\{ [I_l, 0] - C (\hat{W}_f^+ \mathcal{O}_f)^\dagger \hat{W}_f^+ \right\} \mathcal{H}_f \mathcal{E}_f (\mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f})^\dagger \\ &\times \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top (\mathcal{K}_p \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top)^{-1} \\ &+ O(1/N), \quad (49)\\ \lambda =& \max\{\rho(A), \rho(\bar{A})\}. \quad (50)\end{aligned}$$

Proof: The estimation error of C matrix can be obtained straightforwardly. The estimation error of A matrix can be obtained as follows. Because $\underline{\mathcal{O}_f}^\dagger \underline{\Pi^\perp_{\mathcal{O}_f^\top}} = O(A^f)$,

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

From Lemma 2, the 1st term in the R.H.S. of the equation above is calculated as

$$\begin{aligned}\underline{\mathcal{O}_f}^\dagger \overline{\tilde{\mathcal{O}}_f} =& \mathcal{O}_{f-1}^\dagger ([0_{(f-1)l \times l}, I] - \mathcal{O}_{f-1} A (\hat{W}_f^+ \mathcal{O}_f)^\dagger \hat{W}_f^+) \\ &\times \mathcal{H}_f \mathcal{E}_f (\mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f})^\dagger \\ &\times \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top (\mathcal{K}_p \hat{W}_p^- (\hat{W}_p^-)^\top \mathcal{K}_p^\top)^{-1} \\ &+ O(1/N). \quad (51)\end{aligned}$$

This proves the lemma.

Remark 2: In (48) and (49), the noise $\{e_f\}$ appears as

$$\mathcal{E}_f (\mathcal{Z}_p^- \Pi^\perp_{\mathcal{U}_f})^\dagger = \frac{1}{N} \mathcal{E}_f \Pi^\perp_{\mathcal{U}_f} (\mathcal{Z}_p^-)^\top \left(\hat{\Gamma}_p^{-\Pi} \right)^{-1} \quad (52)$$

$\hat{\Gamma}_p^{-\Pi}$ converges to a constant matrix as $N \to \infty$ while $\frac{1}{\sqrt{N}} \mathcal{E}_f \Pi^\perp_{\mathcal{U}_f} (\mathcal{Z}_p^-)^\top$ converges in distribution to a gaussian random matrix of mean 0 and a finite covariance from the central limit theorem. Thus, each of $\tilde{A}$ and $\tilde{C}$ converges to a gaussian random matrix of mean 0 and covariance of order $1/N$.

Remark 3: When the weighting function $\hat{W}_f^+ = I$ and $\hat{W}_p^-$ in Eq. (27) is adopted,

$$\tilde{A}=\left\{[0, \mathcal{O}_{f-1}^\dagger] - A\mathcal{O}_f^\dagger\right\}\mathcal{H}_f\mathcal{E}_f(\mathcal{X}_0^{(p)}\Pi_{\mathcal{U}_f}^\perp)^\dagger + O(\lambda^f/\sqrt{N}) + O(1/N), \tag{53}$$

$$\tilde{C}=\left\{[I_l, 0] - C\mathcal{O}_f^\dagger\right\}\mathcal{H}_f\mathcal{E}_f(\mathcal{X}_0^{(p)}\Pi_{\mathcal{U}_f}^\perp)^\dagger + O(1/N), \tag{54}$$

where $\mathcal{X}_0^{(p)}$ is in Eq. (14). In the original PO-MOESP method, L_{32} instead of $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$ is decomposed to the singular subspaces. Let a SVD of L_{22} be $L_{22} = USV^\top$. Then $\hat{W}_p^-$ in Eq. (27) can be represented as $\hat{W}_p^- = USU^\top/\sqrt{N}$. On the other hand, in order for $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$ to be identical to L_{32}, the weighting functions must be chosen

$$\hat{W}_f^+ = I \quad \text{and} \quad \hat{W}_p^- = USU^\top \cdot UV^\top.$$

In this case, $\hat{W}_p^-(\hat{W}_p^-)^\top = US^2U^\top = N\hat{\Gamma}_p^{-\Pi}$. Thus, the estimation errors of $\hat{A}$ and $\hat{C}$ matrices in the original PO-MOESP method are identical to Eqs. (53) and (54), respectively.

V. On the weighting matrix $\hat{W}_p^-$

In [7], an optimal weighting matrix $\hat{W}_f^+$ in the state approach is analyzed. Similar discussion can be done for $\hat{W}_p^-$ in the shift invariance approach. The estimation errors $\tilde{A}$ and $\tilde{C}$ in Lemma 3 can be vectorized by using raw span as follows:

$$\tilde{\theta}=\text{vec}\left(\begin{bmatrix}\tilde{A}\\ \tilde{C}\end{bmatrix}^\top\right) \tag{55}$$

$$=\text{vec}(\mathbf{H}_p^\top\mathcal{E}_f^\top\mathbf{F}_f^\top) \tag{56}$$

$$=(\mathbf{F}_f \otimes \mathbf{H}_p^\top)\,\text{vec}(\mathcal{E}_f^\top) \tag{57}$$

where

$$\mathbf{F}_f=\left\{\begin{bmatrix}0 & \mathcal{O}_{f-1}^\dagger\\ I_l & O\end{bmatrix} - \begin{bmatrix}A\\ C\end{bmatrix}(\hat{W}_f^+\mathcal{O}_f)^\dagger\hat{W}_f^+\right\}\mathcal{H}_f, \tag{58}$$

$$\mathbf{H}_p=(\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp)^\dagger\hat{W}_p^-(\hat{W}_p^-)^\top\mathcal{K}_p^\top(\mathcal{K}_p\hat{W}_p^-(\mathcal{K}_p\hat{W}_p^-)^\top)^{-1}. \tag{59}$$

Then,

$$\|[\tilde{A}^\top\tilde{C}^\top]^\top\|_F^2 = \tilde{\theta}^\top\tilde{\theta} \tag{60}$$

$$=\text{vec}(\mathcal{E}_f^\top)^\top(\mathbf{F}_f^\top \otimes \mathbf{H}_p)(\mathbf{F}_f \otimes \mathbf{H}_p^\top)\,\text{vec}(\mathcal{E}_f^\top) \tag{61}$$

$$=\text{vec}(\mathcal{E}_f^\top)^\top((\mathbf{F}_f^\top\mathbf{F}_f)\otimes(\mathbf{H}_p\mathbf{H}_p^\top))\,\text{vec}(\mathcal{E}_f^\top). \tag{62}$$

On the other hand, $\mathbf{H}_p^\top\mathbf{H}_p$ will be minimized when $\hat{W}_p^-(\hat{W}_p^-)^\top$ is selected to be $\hat{\Gamma}_p^{-\Pi}$ as in Eq. (27). Let $\hat{W}_p^-(\hat{W}_p^-)^\top = \hat{\Gamma}_p^{-\Pi} + \Delta$ and $\hat{\Omega}_{xx}^{(p)} = \mathcal{K}_p\hat{\Gamma}_p^{-\Pi}\mathcal{K}_p^\top$. Then,

$$\begin{aligned}&\mathbf{H}_p^\top\mathbf{H}_p\\ &=\frac{1}{N}(\hat{\Omega}_{xx}^{(p)} + \mathcal{K}_p\Delta\mathcal{K}_p^\top)^{-1}\mathcal{K}_p(\hat{\Gamma}_p^{-\Pi}+\Delta)(\hat{\Gamma}_p^{-\Pi})^{-1}\\ &\quad\times(\hat{\Gamma}_p^{-\Pi}+\Delta)\mathcal{K}_p^\top(\hat{\Omega}_{xx}^{(p)} + \mathcal{K}_p\Delta\mathcal{K}_p^\top)^{-1}\end{aligned} \tag{63}$$

$$\begin{aligned}&=\frac{1}{N}(\hat{\Omega}_{xx}^{(p)})^{-1} + (\hat{\Omega}_{xx}^{(p)})^{-1}\mathcal{K}_p\Delta(\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp(\mathcal{Z}_p^-)^\top)^{-1}\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp\\ &\quad\times\underbrace{\left(I - \Pi_{\mathcal{U}_f}^\perp(\mathcal{X}_0^{(p)})^\top(\mathcal{X}_0^{(p)}\Pi_{\mathcal{U}_f}^\perp(\mathcal{X}_0^{(p)})^\top)^{-1}\mathcal{X}_0^{(p)}\Pi_{\mathcal{U}_f}^\perp\right)}_{\geq 0}\\ &\quad\times\Pi_{\mathcal{U}_f}^\perp(\mathcal{Z}_p^-)^\top(\mathcal{Z}_p^-\Pi_{\mathcal{U}_f}^\perp(\mathcal{Z}_p^-)^\top)^{-1}\Delta^\top\mathcal{K}_p^\top(\hat{\Omega}_{xx}^{(p)})^{-1} + O(\Delta^4)\end{aligned} \tag{64}$$

Thus, $\Delta = 0$ minimizes $\mathbf{H}_p^\top\mathbf{H}_p$. Since the non-zero singular values of $\mathbf{H}_p^\top\mathbf{H}_p$ are identical to the non-zero sngular values of $\mathbf{H}_p\mathbf{H}_p^\top$, $\hat{W}_p^-(\hat{W}_p^-)^\top = \hat{\Gamma}_p^{-\Pi}$ also minimize $\mathbf{H}_p\mathbf{H}_p^\top$. Thus, from Remark 3, the original PO-MOESP method gives an optimal estimate in the sense that the singular values of $\mathbf{H}_p\mathbf{H}_p^\top$ are minimized.

VI. Conclusion

An explicit formula of the dominant part of the estimation error on each of A and C matrix in PO-MOESP is derived. An optimal weighting matrix $\hat{W}_p^-$ is analyzed and it is shown that Eq. (27) gives an optimal weight in the sense that the Frobenius norm of the estimation errors are minimized. It is also shown that the original PO-MOESP method gives identical estimates with this optimal weighting matrix $\hat{W}_p^-$.

In the analysis of the estimation error terms, mathematical expectation or limitting operations of N, f or p are avoided in order to clarify the effect of the paraemters.

The formulae (53) and (54) have similar expression to the one in [9]. So, the variance of $\hat{A}$ and $\hat{C}$ can be calculated as in [9].

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. originally published by Prentice-Hall in 1979, Dover, 2005.
[2] D. Bauer, "Asymptotic properties of subspace estimators," *Automatica*, vol. 41, pp. 395–376, 2005.
[3] D. Bauer, M. Deistler, and W. Scherrer, "The analysis of the asymptotic variance of subspace algorithms," in *Proceedings of SYSID'97*, 1997, pp. 1087–1091.
[4] ——, "Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs," *Automatica*, vol. 35, pp. 1243–1254, 1999.
[5] ——, "On the impact of weighting matrices in subspace algorithms," in *Proceedings of the IFAC conference SYSID'00*, Santa Barbara, 2000, pp. 97–102.
[6] D. Bauer and M. Jansson, "Analysis of the asymptotic properties of the MOESP type of subspace algorithms," *Automatica*, vol. 36, pp. 497–509, 2000.
[7] D. Bauer and L. Ljung, "Some facts about the choise of the weighting matrices in Larimore type of subspace algorithms," *Automatica*, vol. 38, pp. 763–773, 2002.
[8] A. Chiuso and G. Picci, "On the ill-conditioning of subspace identification with inputs," *Automatica*, vol. 40, no. 4, pp. 575–589, 2004.
[9] ——, "The asymptotic variance of subspace estimates," *Journal of Econometrics*, vol. 118(1-2), pp. 257–291, 2004.
[10] G. H. Golub and C. F. van Loan, *Matrix Computations (2nd Ed.)*. The Johns Hopkins University Press, 1989.

[11] K. Ikeda, "Asymptotic Properties of MOESP-type Methods," in *Preprints of the 11th IFAC International Workshop on Adaptation and Learning in Contgrol and Signal Processing*, 2013, pp. 92–97.
[12] M. Jansson, "Asymptotic Variance Analysis of Subspace Identification Methods," in *Proceedings of the IFAC conference SYSID'00*, Santa Barbara, 2000, pp. 91–96.
[13] W. E. Larimore, "Canonical Variate Analysis in Identification, Filtering, and Adaptive Control," in *Proceedings of the 29th Conference on Decision and Control*, Honolulu, Hawaii, 1990, pp. 596–604.
[14] W. Scherrer, "Local Optimality of Minimum Phase Balanced Truncation," in *Proceedings of the 15th IFAC World Congress*, 2002, pp. 440–454.
[15] G. W. Stewart, "Error and perturbation bound for subspaces associated with certain eigenvalue problems," *SIAM Review*, vol. 15, no. 4, pp. 727–764, 1973.
[16] G. W. Stewart and J. guang Sun, *Matrix Perturbation Theory*. Academic Press, 1990.
[17] P. Van Overschee and B. De Moor, "N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.
[18] M. Verhaegen, "Subspace model identification Part 3. Analysis of the ordinary output-error state-space model identification algorithm," *International Journal of Control*, vol. 58, pp. 555–586, 1993.
[19] ——, "Identification of the Deterministic Part of MIMO State Models given in Innovations Form from Input-Output Data," *Automatica*, vol. 30, pp. 61–74, 1994.
[20] M. Verhaegen and P. Dewilde, "Subspace model identification Part 1. The output-error state-space model identification class of algorithms," *International Journal of Control*, vol. 56, pp. 1187–1210, 1992.
[21] ——, "Subspace model identification Part 2. Analysis of the elementary output-error state-space model identification algorithms," *International Journal of Control*, vol. 56, pp. 1211–1241, 1992.
[22] M. Verhaegen and V. Verdult, *Filtering and System Identification*. Cambridge University Press, 2007.

PROOF OF LEMMA1

Let $i \le n$ and X and $\hat{X} = X + \tilde{X}$ have SVD's as

$$X{=}[U_i, U_i^{\perp}]\,\mathrm{diag}(\Sigma_i, \Sigma_i^{\perp})[V_i, V_i^{\perp}]^{\top}, \tag{65}$$
$$\hat{X}{=}[\hat{U}_i, \hat{U}_i^{\perp}]\,\mathrm{diag}(\hat{\Sigma}_i, \hat{\Sigma}_i^{\perp})[\hat{V}_i, \hat{V}_i^{\perp}]^{\top}. \tag{66}$$

where $\Sigma_i \in R^{i\times i}$ and $\Sigma_i^{\perp}$ with appropriate dimensions are composed of the singular values of X and $[U_i, U_i^{\perp}]$, $[V_i, V_i^{\perp}]$ are the unitary matrices. $\hat{\Sigma}_i \in R^{i\times i}$ and $\hat{\Sigma}_i^{\perp}$ are composed of the singular values of $\hat{X}$ and $[\hat{U}_i, \hat{U}_i^{\perp}]$ and $[\hat{V}_i, \hat{V}_i^{\perp}]$ are unitary matrices. Let σ_i denote an i-th singular value of X and $\hat{\sigma}_i$ denote an i-th singular value of $\hat{X}$.

An upper bound of the perturbations on the singular values is given by

$$\left\| \begin{bmatrix} \Sigma_n & \\ & 0 \end{bmatrix} - \begin{bmatrix} \hat{\Sigma}_n & \\ & \hat{\Sigma}_n^{\perp} \end{bmatrix} \right\|_2 \le \|\tilde{X}\|_2. \tag{67}$$

(Mirsky's theorem. See [16] for example.) Especially, $\|\hat{\Sigma}_n^{\perp}\|_2 \le \|\tilde{X}\|_2$.

As for the perturbations on the singular subspaces, the gap[15] between the invariant subspaces $\mathrm{Range}\,(U_i)$ and $\mathrm{Range}\,(\hat{U}_i)$ plays an important role. The gap is given by

$$\begin{aligned} &\gamma(\mathrm{Range}\,(U_i), \mathrm{Range}\,(\hat{U}_i)) \\ =&\|U_iU_i^{\top} - \hat{U}_i\hat{U}_i^{\top}\|_2 && (68)\\ =&\|[U_i, U_n^{\perp}]^{\top}(U_iU_i^{\top} - \hat{U}_i\hat{U}_i^{\top})[\hat{U}_i, \hat{U}_i^{\perp}]\|_2 && (69)\\ =&\left\| \begin{bmatrix} O & U_i^{\top}\tilde{U}_i^{\perp} \\ -(U_i^{\perp})^{\top}\tilde{U}_i & O \end{bmatrix} \right\|_2 && (70)\\ =&\max\left\{ \|(U_i^{\perp})^{\top}\tilde{U}_i\|_2, \|U_i^{\top}\tilde{U}_i^{\perp}\|_2 \right\} && (71) \end{aligned}$$

The fact that the pre- and post-multiplications of the unitary matrices $[U_i, U_i^{\perp}]^{\top}$ and $[\hat{U}_i, \hat{U}_i^{\perp}]$ do not affect the 2-norm is used in the calculation above.

The first argument of max function in Eq. (71) is given as follows. Calculate $(U_i^{\perp})^{\top}\hat{X}\hat{V}_i$ by using the SVD of $\hat{X}$ (66) as

$$(U_i^{\perp})^{\top}\hat{X}\hat{V}_i{=}(U_i^{\perp})^{\top}\left\{\hat{U}_i\hat{\Sigma}_i\hat{V}_i + \hat{U}_i^{\perp}\hat{\Sigma}_i^{\perp}(\hat{V}_i^{\perp})^{\top}\right\}\hat{V}_i \tag{72}$$
$$={(U_i^{\perp})^{\top}\hat{U}_i\hat{\Sigma}_i} \tag{73}$$
$$={(U_i^{\perp})^{\top}\tilde{U}_i\hat{\Sigma}_i}. \tag{74}$$

On the other hand, calculate $(U_i^{\perp})^{\top}\hat{X}\hat{V}_i$ by using $\hat{X} = X + \tilde{X}$ and the SVD of X (65) as

$$(U_i^{\perp})^{\top}\hat{X}\hat{V}_i{=}(U_i^{\perp})^{\top}\left\{U_i\Sigma_iV_i^{\top} + U_i^{\perp}\Sigma_i^{\perp}(V_i^{\perp})^{\top} + \tilde{X}\right\}\hat{V}_i \tag{75}$$
$$={\Sigma_i^{\perp}(V_i^{\perp})^{\top}\tilde{V}_i + (U_i^{\perp})^{\top}\tilde{X}\hat{V}_i}. \tag{76}$$

From Eqs. (74) and (76),

$$(U_i^{\perp})^{\top}\tilde{U}_i\hat{\Sigma}_i = \Sigma_i^{\perp}(V_i^{\perp})^{\top}\tilde{V}_i + (U_i^{\perp})^{\top}\tilde{X}\hat{V}_i. \tag{77}$$

Calculating $(V_i^{\perp})^{\top}\hat{X}^{\top}\hat{U}_i$ in two ways similar to the above gives

$$(V_i^{\perp})^{\top}\tilde{V}_i\hat{\Sigma}_i = \Sigma_i^{\perp}(U_i^{\perp})^{\top}\tilde{U}_i + (V_i^{\perp})^{\top}\tilde{X}^{\top}\hat{U}_i. \tag{78}$$

From (77) and (78), the following is obtained:

$$\begin{aligned} &(U_i^{\perp})^{\top}\tilde{U}_i(\hat{\Sigma}_i)^2 - (\Sigma_i^{\perp})^2(U_i^{\perp})^{\top}\tilde{U}_i \\ =&(U_i^{\perp})^{\top}\tilde{X}\hat{V}_i\hat{\Sigma}_i + \Sigma_i^{\perp}(V_i^{\perp})^{\top}\tilde{X}^{\top}\hat{U}_i. \end{aligned} \tag{79}$$

If $\hat{\sigma}_i > \sigma_{i+1}$, the Lyapunov equation above has a unique solution for $(U_i^{\perp})^{\top}\tilde{U}_i$ and it is of order $\|\tilde{X}\|_2$. When $i = n$,

$$(U_n^{\perp})^{\top}\tilde{U}_n\hat{\Sigma}_n = (U_n^{\perp})^{\top}\tilde{X}\hat{V}_n \tag{80}$$

is obtained because $\Sigma_n^{\perp} = 0$. Note that the condition $\sigma_n > \hat{\sigma}_{n+1}$ is not required for $i = n$. This proves Eq. (30).

The second argument of max function in Eq. (71) is given as follows: Calculating $U_i^{\top}X\hat{V}_i^{\perp}$ and $Vi^{\top}X^{\top}\hat{U}_i^{\perp}$ by using the SVD of X (65) and by using $X = \hat{X} - \tilde{X}$ and the SVD of $\hat{X}$ (66), a Lyapunov equation similar to Eq. (79) is obtained:

$$\begin{aligned} &U_i^{\top}\tilde{U}_i^{\perp}(\hat{\Sigma}_i^{\perp})^2 - (\Sigma_i)^2U_i^{\top}\tilde{U}_i^{\perp} \\ =&U_i^{\top}\tilde{X}\hat{V}_i^{\perp}\hat{\Sigma}_i^{\perp} + \Sigma_iV_i^{\top}\tilde{X}^{\top}\hat{U}_i^{\perp}. \end{aligned} \tag{81}$$

If $\sigma_i > \hat{\sigma}_{i+1}$, the Lyapunov equation above has a unique solution for $U_i^{\top}\tilde{U}_i^{\perp}$ and it is of order $\|\tilde{X}\|_2$.

From Eqs. (79) and (81) and the Mirsky's thoerem (67), if $\|\tilde{X}\|_2 < (\sigma_i - \sigma_{i+1})/2$, then

$$\gamma(\mathrm{Range}\,(U_i), \mathrm{Range}\,(\hat{U}_i)) = O(\|\tilde{X}\|_2). \tag{82}$$

Eq. (31) and the gap between $\mathrm{Range}\,(V_i)$ and $\mathrm{Range}\,(\hat{V}_i)$ can be seen similarly.

When σ_i is a distinct singular value, *i.e.*, $\sigma_{i-1} > \sigma_i > \sigma_{i+1}$, the choise of the corresponding singular vectors u_i and v_i in real vectors is just a sign. Because the gap between $\mathrm{Range}\,(u_i)$ and $\mathrm{Range}\,(\hat{u}_i)$ is of order $\|\tilde{X}\|_2$ and so is the gap between $\mathrm{Range}\,(v_i)$ and $\mathrm{Range}\,(\hat{v}_i)$, for a given pair of $\hat{u}_i$ and $\hat{v}_i$, there is a pair of u_i and v_i such that $\|u_i - \hat{u}_i\|_2 = O(\|\tilde{X}\|_2)$ and $\|v_i - \hat{v}_i\|_2 = O(\|\tilde{X}\|_2)$.

When σ_i is a multiple singular value, say $\sigma_{i-1} > \sigma_i = \sigma_{i+1} = \cdots = \sigma_{i+m} > \sigma_{i+m+1}$, there is a degree of freedom of the choise of the sign and the Givens rotation of the set of corresponding singular vectors $[u_i, \ldots, u_{i+m}]$ and $[v_i, \ldots, v_{i+m}]$. As in the distinct singular value case, Because the gap between $\mathrm{Range}\,([u_i, \ldots, u_{i+m}])$ and $\mathrm{Range}\,([\hat{u}_i, \ldots, \hat{u}_{i+m}])$ is of order $\|\tilde{X}\|_2$ and so is the gap between $\mathrm{Range}\,([v_i, \ldots, v_{i+m}])$ and $\mathrm{Range}\,([\hat{v}_i, \ldots, \hat{v}_{i+m}])$, for a given pair of $[\hat{u}_i, \ldots, \hat{u}_{i+m}]$ and $[\hat{v}_i, \ldots, \hat{v}_{i+m}]$, there is a pair of $[u_i, \ldots, u_{i+m}]$ and $[v_i, \ldots, v_{i+m}]$ such that $\|[u_i, \ldots, u_{i+m}] - [\hat{u}_i, \ldots, \hat{u}_{i+m}]\|_2 = O(\|\tilde{X}\|_2)$ and $\|[v_i, \ldots, v_{i+m}] - [\hat{v}_i, \ldots, \hat{v}_{i+m}]\|_2 = O(\|\tilde{X}\|_2)$

From the discussions above and Mirsky's theorem, there exist (U_n, Σ_n, V_n) and $(\hat{U}_n, \hat{\Sigma}_n, \hat{V}_n)$ such that $\hat{U}_n \hat{\Sigma}_n^{-1} = U_n \Sigma_n^{-1} + O(\|\tilde{X}\|)$ and $\hat{V}_n \hat{\Sigma}_n^{-1} = V_n \Sigma_n^{-1} + O(\|\tilde{X}\|)$. This proves the second part of the lemma.