Subspace Identification of Combined Deterministic-Stochastic Systems by LQ Decomposition

Tohru Katayama

Abstract—In this paper, we revisit the realization-based subspace identification problem for combined deterministic-stochastic discrete-time LTI systems. Under the assumption that the past horizon is sufficiently large, we reveal a block lower triangular structure of an L-factor related to the stochastic component in the LQ decomposition of data matrix based on an asymptotic analysis and covariance factorization. Adapting the theoretical result to finite input-output data, we then develop a method of computing the steady state Kalman gain and the covariance of innovation process, where K is obtained by a method similar to computing (B, D) parameters in the MOESP method. Thus, under the assumption that the input is persistently exciting (PE) of sufficiently high order, we can compute all the systems parameters from L-factors of a single LQ decomposition of the data matrix.

I. INTRODUCTION

The LQ decomposition (transpose of the QR decomposition), together with the singular value decomposition (SVD), has effectively been used as a numerical tool in subspace system identification methods [1], [2], [3]. We have employed the LQ decomposition for a preliminary orthogonal decomposition of the output process into deterministic and stochastic components in order to develop a stochastic realization theory in the presence of exogenous inputs [4]. Moreover, clarified was the role of LQ decomposition in the subspace identification of deterministic systems [5].

It is well known that in the PO-MOESP method [6], the system parameters are obtained by using L-factors related to the deterministic component in the LQ decomposition of a data matrix. For combined deterministic-stochastic systems [7], the steady state Kalman gain is derived, after the identification of deterministic component, by solving an algebraic Riccati equation that arises in stochastic realization theory [8], [9]. This shows that the L-factor related to the stochastic component in the LQ decomposition has not been fully utilized in subspace identification algorithms for combined deterministic-stochastic systems.

In [10], a method of computing the steady state Kalman gain directly from the LQ decomposition has been derived by utilizing L-factors related to the stochastic component in a different setting than standard subspace identification methods; however, the result is obscure since no clear explanations of the algorithm were provided.

In this paper, under the assumption that the past horizon is sufficiently large and the number of input-output data goes to infinity, we show that an L-factor in the LQ decomposition related to the stochastic component has a block lower triangular structure by using a technique developed in [6]. Adapting this theoretical result to finite input-output data, we obtain a new approximate procedure for identifying the innovation model, thereby showing that all the systems parameters including stochastic components are computed from L-factors only in a single LQ decomposition of the data matrix. This result is not unexpected, in view of the fact that the LQ decomposition transforms a given data matrix into a product of a lower triangular matrix and an orthogonal matrix, where the former carries the information useful for system identification (least-squares), while the latter provides orthogonal bases of the row space of the data matrix.

The rest of the paper is organized as follows. The problem is stated in Section II. Section III reviews matrix state-input-output equations employed in subspace methods. In Section IV, we introduce the LQ decomposition of the data matrix composed of the past and future of input-output data, and derive a subspace identification method for the deterministic component. In Section V, under the assumptions that the past horizon is sufficiently large and the number of data goes to infinity, we analyze a block lower triangular structure of a certain L-factor related to the stochastic component. Based on this analysis, we then derive a method of estimating the steady state Kalman gain and the covariance of innovation process. In Section VI, we briefly discuss the case where there are no exogenous inputs. Some numerical results are included in Section VII to show the applicability of the LQ decomposition-based method. We conclude this paper in Section VIII.

II. PROBLEM STATEMENT

Consider a stochastic system of the innovation form

\begin{align}
    x(t + 1) &= Ax(t) + Bu(t) + Ke(t) \\
    y(t) &= Cx(t) + Du(t) + e(t)
\end{align}

where \( x \in \mathbb{R}^n \) is the state vector, \( u \in \mathbb{R}^m \) the exogenous input, \( y \in \mathbb{R}^p \) the output vector, \( e \in \mathbb{R}^p \) the innovation vector, and \( A \in \mathbb{R}^{n \times n} \), \( B \in \mathbb{R}^{n \times m} \), \( K \in \mathbb{R}^{n \times p} \), \( C \in \mathbb{R}^{p \times n} \), \( D \in \mathbb{R}^{p \times m} \) are constant matrices. In the following, we assume that \((A, B)\) and \((A, K)\) are reachable and \((C, A)\) is observable. The innovation process \( e \) is a white noise vector with mean zero and covariance matrix

\[ E[e(t)e^T(s)] = \Lambda \delta_{ts}, \quad \Lambda > 0, \quad \Lambda \in \mathbb{R}^{p \times p} \]

and is uncorrelated with the past state, i.e.

\[ E[e(t)x^T(s)] = 0, \quad t \geq s \]
The problem is to identify the deterministic component \((A, B, C, D)\) and the stochastic component \((K, \Lambda)\) based on finite input-output data \((y(t), u(t), t = 0, 1, \cdots, T)\) by using the realization-based subspace method [11]. It is assumed that there is no feedback from the output \(y\) to the input \(u\) [4], so that we consider an open-loop subspace identification problem. In Sections V and VI, we further assume that \(y, e\) and \(u\) are stationary processes in order to analyze the structure of L-factors related to the stochastic component.

III. MATRIX STATE-INPUT-OUTPUT EQUATION

We derive matrix state-input-output equations, on which most subspace identification methods are based.

Let \(k > n\). We define the stacked vectors as

\[
y_k(t) := \begin{bmatrix} y(t) \\ y(t + 1) \\ \vdots \\ y(t + k - 1) \end{bmatrix} \in \mathbb{R}^{kp}
\]

\[
u_k(t) := \begin{bmatrix} u(t) \\ u(t + 1) \\ \vdots \\ u(t + k - 1) \end{bmatrix} \in \mathbb{R}^{km}
\]

\[
e_k(t) := \begin{bmatrix} e(t) \\ e(t + 1) \\ \vdots \\ e(t + k - 1) \end{bmatrix} \in \mathbb{R}^{kp}
\]

Then, from (1) and (2), the stacked vectors satisfy the following augmented equation

\[
y_k(t) = O_k x(t) + T_k u_k(t) + K_k e_k(t) \tag{4}
\]

where \(O_k\) is the extended observability matrix given by

\[
O_k = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix} \in \mathbb{R}^{kp \times n}
\]

and where \(T_k\) is the block lower triangular Toeplitz matrix defined by impulse responses

\[
T_k = \begin{bmatrix} D & 0 \\ CB & D \\ \vdots & \ddots & \ddots \\ CA^{k-2} B & \cdots & CB & D \end{bmatrix} \in \mathbb{R}^{kp \times km} \tag{5}
\]

and \(K_k\) is the block Toeplitz matrix formed by the impulse responses from \(e\) to \(y\), i.e.

\[
K_k = \begin{bmatrix} I_p & 0 \\ CK & I_p \\ \vdots & \ddots & \ddots \\ CA^{k-2} K & \cdots & CK & I_p \end{bmatrix} \in \mathbb{R}^{kp \times kp} \tag{6}
\]

We also define block Hankel matrices

\[
Y_{0k-1} = \begin{bmatrix} y(0) & y(1) & \cdots & y(N - 1) \\ y(1) & y(2) & \cdots & y(N) \\ \vdots & \vdots & \ddots & \vdots \\ y(k - 1) & y(k) & \cdots & y(k + N - 2) \end{bmatrix} \in \mathbb{R}^{kp \times N}
\]

\[
U_{0k-1} = \begin{bmatrix} u(0) & u(1) & \cdots & u(N - 1) \\ u(1) & u(2) & \cdots & u(N) \\ \vdots & \vdots & \ddots & \vdots \\ u(k - 1) & u(k) & \cdots & u(k + N - 2) \end{bmatrix} \in \mathbb{R}^{km \times N}
\]

and

\[
E_{0k-1} = \begin{bmatrix} e(0) & e(1) & \cdots & e(N - 1) \\ e(1) & e(2) & \cdots & e(N) \\ \vdots & \vdots & \ddots & \vdots \\ e(k - 1) & e(k) & \cdots & e(k + N - 2) \end{bmatrix} \in \mathbb{R}^{kp \times N}
\]

Similarly, we define \(Y_{2k-1}, Y_{4k-1}, E_{2k-1}\).

It follows from (4) that matrix state-input-output equations are given by

\[
Y_{0k-1} = O_kX_0 + T_kU_{0k-1} + K_kE_{0k-1} \tag{7}
\]

\[
Y_{2k-1} = O_kX_k + T_kU_{2k-1} + K_kE_{2k-1} \tag{8}
\]

where

\[
X_0 = [x(0) \ x(1) \ \cdots \ x(N - 1)] \in \mathbb{R}^{n \times N}
\]

\[
X_k = [x(k) \ x(k + 1) \ \cdots \ x(k + N - 1)] \in \mathbb{R}^{n \times N}
\]

are the initial state matrices for (7) and (8), respectively.

The next step is to derive a subspace identification procedure for the deterministic component \((A, B, C, D)\) by applying the LQ decomposition.

IV. IDENTIFICATION OF DETERMINISTIC COMPONENT

In this section, we assume that the input \(u\) is PE with order \(2k\), i.e. \(U_{02k-1}\) has full row rank. Also, that \(X_0\) has full row rank is assured by the reachability of \((A, [B\ K])\).

Define the past input-output data

\[
W_p = \begin{bmatrix} U_{0k-1} \\ Y_{0k-1} \end{bmatrix} \in \mathbb{R}^{(k+p) \times N}
\]

which was used as an instrument to remove noise effects from the matrix state-input-output equation [11]. The actual computation is performed by the LQ decomposition [6]

\[
\begin{bmatrix} U_{2k-1} \\ W_p \\ Y_{2k-1} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 \\ R_{21} & R_{22} & 0 \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{bmatrix} \tag{9}
\]

where \(R_{11} \in \mathbb{R}^{km \times km}, R_{22} \in \mathbb{R}^{(k+p) \times (k+p)}, R_{33} \in \mathbb{R}^{kp \times kp}\) are block lower triangular matrices, and \(Q_1 \in \mathbb{R}^{N \times km}, Q_2 \in \mathbb{R}^{N \times (k+p)}, Q_3 \in \mathbb{R}^{N \times kp}\) are orthogonal matrices.

Post-multiplying (8) by \(Q_1\) and \(Q_2\) respectively yields

\[
Y_{2k-1}Q_1 = O_kX_0Q_1 + T_kU_{2k-1}Q_1 + K_kE_{2k-1}Q_1 \tag{10}
\]

\[
Y_{2k-1}Q_2 = O_kX_kQ_2 + T_kU_{2k-1}Q_2 + K_kE_{2k-1}Q_2 \tag{11}
\]
Thus, the innovation process $e$ is a white noise and that there is no feedback from the output $y$ to the input $u$. Thus, the innovation process $E_{4(k-1)}$ is uncorrelated with the past inputs and outputs $W_{0(k-1)}$ and future inputs $U_{k|2k-1}$, i.e.

$$\lim_{N \to \infty} \frac{1}{N} E_{4|2k-1} [U_{k|2k-1}^T W_p^T] = 0$$

Moreover, it follows from (9) that

$$\begin{bmatrix} U_{k|2k-1}^T \\ W_p^T \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}$$

where $R_{11}$ and $R_{22}$ are nonsingular. Thus we have

$$\lim_{N \to \infty} \frac{1}{N} E_{4|2k-1} [Q_1 Q_2] = 0$$

This implies that the orthogonal matrices $Q_1^T$ and $Q_2^T$ are uncorrelated with the future noise $E_{4|2k-1}$ [6].

Letting $N \to \infty$ in (14) and (15), we get asymptotically

$$R_{31} = O_k X_k Q_1 + T_k R_{11} \tag{17}$$

$$R_{32} = O_k X_k Q_2 \tag{18}$$

where the common factor $1/\sqrt{N}$ in the above equations is suppressed. A subspace method of obtaining $(A, B, C, D)$ from the above equation is well known in the MOESP method [3], so that we briefly summarize the procedure.

**Algorithm A: Identification Method of $(A, B, C, D)$**

1. Compute the SVD of $R_{32}/\sqrt{N}$, i.e.

$$\frac{1}{\sqrt{N}} R_{32} = [U_1 \\ U_2] \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} V^T \in \mathbb{R}^{k \times (m+p)}$$

where we assume that $S_1 \in \mathbb{R}^{\nu \times \nu}$ is positive definite, and the singular values of $S_2$ are negligibly small.

2. Under the assumption that $X_k Q_2$ has full rank, we see from (18) that $\text{Im}(O_k) = \text{Im}(R_{32})$, so that we define $O_k = U_1 S_1^{-1/2}$. Then, we have

$$A = O_k^\dagger \tilde{O}_k, \quad C = O_k(1: p, 1 : n)$$

where $\tilde{O}_k = O_k(p + 1 : k p, :)$.

3. Pre-multiplying (17) by $U_2^T$ yields

$$U_2^T R_{31} = U_2^T T_k (B, D) R_{11} \tag{19}$$

Since, from (5), $T_k (B, D)$ is linear with respect to $(B, D)$, we can compute their estimates by applying the least-squares method to (19).

**V. IDENTIFICATION OF STOCHASTIC COMPONENT**

Recall from (9) that

$$Y_{4|2k-1} = R_{31} Q_1^T + R_{32} Q_2^T + R_{33} Q_3^T \tag{20}$$

As shown in (17) and (18), the matrices $R_{31}$ and $R_{32}$ contain information about the deterministic component $(A, B, C, D)$. Hence, it is clear that $R_{33}$ carries information about the stochastic component; but it seems that this fact was overlooked in the literature.

In [6], as shown in (16), the correlation between the noise $E_{4|2k-1}$ and orthogonal factors $Q_1^T, Q_2^T$ has been evaluated to derive the basic equations like (17), (18) satisfied by the deterministic component. However, the correlation between $E_{4|2k-1}$ and $Q_3^T$ has not been considered in [6], which is the main topic in this section.

For the asymptotic analysis of the structure of $R_{33}$, we assume that $y, u$ and $e$ are 2nd-order stationary processes and that the past horizon is sufficiently large. Thus we define

$$W_p = \begin{bmatrix} U_{k-pk-1} \\ Y_{k-pk-1} \end{bmatrix}$$

where the length of past horizon $p$ is sufficiently large; however, the length of future horizon is $k$ as before.

We see from (8) and (20) that

$$O_k X_k + T_k U_{k|2k-1} - R_{31} Q_1^T - R_{32} Q_2^T = R_{33} Q_3^T - K_k E_{4|2k-1} \tag{21}$$

The first term $O_k X_k$ in the above equation is the oblique projection of $Y_{4|2k-1}$ onto the entire past $W_{\infty}$ along $U_{k|2k-1}$ [1], [12]. Hence, the row vectors of $O_k X_k$ nearly lie in the row space of $W_p$ for sufficiently large $p$, so that $O_k X_k$ is expressed in terms of $Q_1^T$ and $Q_2^T$. Also, it follows from (9) that $U_{k|2k-1}$ is expressed in terms of $Q_1^T$. Thus, the left-hand side of (21) is expressed as a linear combination of $Q_1^T$ and $Q_2^T$, if $p$ is sufficiently large. On the other hand, we see from (16) that $E_{4|2k-1}$ is orthogonal to $Q_1^T$ and $Q_2^T$ asymptotically. Thus, since $E_{4|2k-1} \in \text{rowspace}(Q_1^T \oplus Q_2^T \oplus Q_3^T)$, we see that $E_{4|2k-1}$ is expressed in terms of $Q_3^T$, i.e. $E_{4|2k-1} = H_k Q_3^T, H_k \in \mathbb{R}^{k \times (m+p)}$, implying that $E_{4|2k-1} R_{32}^T = H_k^T$ holds asymptotically.

Post-multiplying (21) by $Q_3$ yields

$$R_{33} = K_k E_{4|2k-1} Q_3 = K_k H_k \tag{22}$$

Thus, under the ergodicity assumption [13], we see from (3) and (22) that

$$\lim_{N \to \infty} \frac{1}{N} R_{33} R_{33}^T = K_k \left[ \lim_{N \to \infty} \frac{E_{4|2k-1} E_{4|2k-1}^T}{N} \right] K_k^T$$

$$= \begin{bmatrix} \Lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Lambda_k \end{bmatrix}$$

where $\Lambda_1 = \cdots = \Lambda_k$ holds for $N \to \infty$. It should be however noted that $\Lambda_i$ take different values for finite $N$, since

$$\Lambda_i(N) := \frac{1}{N} \sum_{r=k-N+1}^{k-N+i} e(t) e^T(t), \quad i = 1, \cdots, k \tag{24}$$

are averages over sliding windows.
Since $\Lambda_i > 0$, we see from (23) that for $N \to \infty$, a Cholesky factor is given by
\[
\frac{1}{\sqrt{N}} R_{33} = \mathcal{K}_k \begin{bmatrix} F_1 & 0 & \cdots & 0 \\ CKF_1 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{k-2} KF_1 & CA^{k-3} KF_2 & \cdots & F_k \end{bmatrix} \in \mathbb{R}^{kp\times p}
\] (25)

By definition, both $R_{33}$ and $\mathcal{K}_k$ are block lower triangular, so that $F_i, i = 1, \cdots, k$ are also block lower triangular and satisfy
\[
\lim_{N \to \infty} F_i F_i^T = \Lambda_i, \quad i = 1, \cdots, k
\] (26)
Moreover, since $R_{33}$ in the LQ decomposition is unique up to a signature matrix, so are $F_i, i = 1, \cdots, k$. A main result of the paper is summarized as follows.

**Theorem 1** Assume that the length of past horizon $p$ is sufficiently large. Then, for $N \to \infty$, we see that $R_{33}$ has a block lower triangular structure of the form
\[
\frac{1}{\sqrt{N}} R_{33} = \begin{bmatrix} F_1 & 0 & \cdots & 0 \\ CKF_1 & F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{k-2} KF_1 & CA^{k-3} KF_2 & \cdots & F_k \end{bmatrix}
\] (27)

**Proof:** A proof is immediate from (6) and (25). $\blacksquare$

We can easily derive a procedure of identifying $K$ and $\Lambda$ from Theorem 1 by using the finite horizon LQ decomposition of (9).

For simplicity, we define $\bar{R}_{33} = R_{33}/\sqrt{N}$. It then follows from (27) that
\[
F_i = \bar{R}_{33}(i-1)p + 1 : ip, (i-1)p + 1 : ip), \quad i = 1, \cdots, k
\]
Also, from (27),
\[
\begin{align*}
O_{k-1} KF_1 &= \bar{R}_{33}(p + 1 : kp, 1 : p) \\
O_{k-2} KF_2 &= \bar{R}_{33}(2p + 1 : kp, p + 1 : 2p) \\
& \vdots \\
O_1 KF_{k-1} &= \bar{R}_{33}((k-1)p + 1 : kp, (k-2)p + 1 : (k-1)p)
\end{align*}
\] (28)

**Algorithm B: Identification Method of ($K, \Lambda$)**

Step 1: Define the right-hand members of (28) as
\[
\begin{align*}
R_i &= \bar{R}_{33}(i-1)p + 1 : ip, (i-1)p + 1 : ip) F_i^{-1} \\
\end{align*}
\]
where $F_i, i = 1, \cdots, k - 1$ are nonsingular. Then (28) is reduced to
\[
\begin{bmatrix}
O_{k-1} \\
O_{k-2} \\
\vdots \\
O_1
\end{bmatrix} K = \begin{bmatrix} R_1 \\
R_2 \\
\vdots \\
R_{k-1}
\end{bmatrix}
\] (29)

Step 2: Apply the least-squares method to (29) to get an estimate of the steady-state Kalman gain. It should be noted that the stability of $A - KC$ is not assured$^1$.

$^1$A way of getting a stabilizing $K$ is to solve a Riccati equation using noise covariances computed from residuals [12]; see also [7], [14].

Step 3: It follows from (26) that an estimate of the covariance matrix $\Lambda$ is given by
\[
\hat{\Lambda} = \frac{1}{k} \sum_{i=1}^{k} F_i F_i^T
\] (30)

It should be noted that the identification method of $K$ in Algorithm B is quite similar to that of $(B, D)$ in the MOESP algorithm [3] for solving (19), which utilizes the block structure of the Toeplitz matrix $T_k$ of (5).

We have established a realization-based subspace identification method (LQD-based method) for combined deterministic-stochastic systems by fully utilizing the property of LQ decomposition of the data matrix, which is summarized as Algorithms A and B.

**VI. STOCHASTIC SYSTEM**

We briefly discuss the stochastic system ($u = 0$)
\[
\begin{align*}
x(t + 1) &= Ax(t) + Ke(t) \\
y(t) &= Cx(t) + e(t)
\end{align*}
\] (31) (32)
where we assume that $y$ and $e$ are 2nd-order stationary processes. Also, it is assumed that the past horizon is sufficiently large, so that the past is given by $W_p = Y_{k-p+1}$ with $p$ sufficiently large.

It then follows from (8) that the matrix state-input-output equations are reduced to
\[
Y_{k+1} = O_k X_k + \mathcal{K}_k E_{k+1}
\] (33)

The LQ decomposition is then given by$^2$
\[
\begin{align*}
Y_{k-p+1} &= L_{11} Q_1^T \\
Y_{k+1} &= L_{21} Q_1^T + L_{22} Q_2^T
\end{align*}
\] (34) (35)

From (33) and (35), we have
\[
L_{21} Q_1^T - O_k X_k = \mathcal{K}_k E_{k+1} - L_{22} Q_2^T
\] (36)

It is well known [8], [16] that the optimal predictor of $Y_{k+1}$ based on the past $Y_{k-p+1}$ is given by the orthogonal projection
\[
\hat{Y}_{k+1} = \hat{E}[Y_{k+1} | Y_{k-p+1}]
\]

Also, $\hat{E}[Y_{k+1} | Y_{k-p+1}] \to O_k X_k$ as $p \to \infty$, and
\[
\hat{E}[Y_{k+1} | Y_{k-p+1}] = Y_{k+1} (Q_1 Q_1^T) = L_{21} Q_1^T
\]
This implies that for $p \to \infty$,
\[
L_{21} Q_1^T = O_k X_k
\] (37)
so that from (36) we have
\[
L_{22} Q_2^T = \mathcal{K}_k E_{k+1}
\] (38)

It follows from (37) that for $p \to \infty$,
\[
\lim_{N \to \infty} \frac{1}{N} L_{21} L_{21}^T = O_k \left( \lim_{N \to \infty} \frac{X_k X_k^T}{N} \right) O_k^T = O_k \Sigma_k O_k^T
\]

$^2$See also [15], in which a stochastic realization method is developed based on the LQ decomposition in Hilbert space.
where $\Sigma$ is the covariance matrix of the state vector $x$ of the stochastic system of (31) and (32). Similarly, from (38),

$$
\lim_{N \to \infty} \frac{1}{N} L_{22}^T = \mathcal{K}_k \left( \lim_{N \to \infty} \frac{E_{4l_{2k-1}} E_{4l_{2k-1}}^T}{N} \right) \mathcal{K}_k^T
$$

so that for $N \to \infty$, a Cholesky factor is given by

$$
\frac{1}{\sqrt{N}} L_{22} = \mathcal{K}_k \begin{bmatrix}
    \bar{F}_1 \\
    \vdots \\
    \bar{F}_k
\end{bmatrix}
$$

where

$$
\lim_{N \to \infty} \frac{1}{\sqrt{N}} \bar{F}_i \bar{F}_i^T = \Lambda_i, \quad i = 1, \ldots, k
$$

Thus, we have the following result.

**Theorem 2** Assume that the length of past horizon $p$ is sufficiently large. Then, for $N \to \infty$, we have

$$
\frac{1}{\sqrt{N}} L_{21} = O_{kT} T^{-1} \Sigma^{1/2} Z^T
$$

where $T \in \mathbb{R}^{p \times p}$ is nonsingular and $Z \in \mathbb{R}^{p \times n}$ is orthogonal, and

$$
\frac{1}{\sqrt{N}} L_{22} = \begin{bmatrix}
    \bar{F}_1 & 0 \\
    C \bar{K}_1 \bar{F}_1 & \bar{F}_2 \\
    \vdots & \vdots \\
    C A^{k-2} K \bar{F}_1 & C A^{k-3} K \bar{F}_2 & \cdots & 0
\end{bmatrix}
$$

**Proof:** A proof is immediate from above. \hfill \blacksquare

Thus, for a stochastic system, we have an approximate identification method by using the finite horizon LQ decomposition.

**Algorithm C: Identification Method of (A, C, K, 0)**

Step 1: Compute the SVD

$$
\frac{1}{\sqrt{N}} L_{21} = [U_1 \quad U_2] \begin{bmatrix}
    S_1 \\
    S_2
\end{bmatrix} V^T
$$

where $S_1 \in \mathbb{R}^{n \times n}$ is positive definite, while the singular values of $S_2$ are negligibly small.

Step 2: From (39), we see that $\text{Im}(O_k) = \text{Im}(L_{21})$, and hence $\text{Im}(O_k) = \text{Im}(U_1)$. A simplest way of choosing the extended observability matrix is to put

$$
O_k = U_1, \quad \Sigma^{1/2} = S_1
$$

Then, we have

$$
A = O_k^T \bar{O}_k, \quad C = O_k(1 : p, 1 : n)
$$

Step 3: Since (40) has exactly the same form as (27), we use Algorithm B for estimating $(K, \Lambda)$.

We can apply the above algorithm to the identification of stationary time series, and to the identification of stochastic component in the ORT method [4].

**VII. NUMERICAL EXAMPLE**

Consider a 2-input, 2-output system ([10])

$$
A = \begin{bmatrix}
    1.5 & 1 & 0.1 \\
    -0.7 & 0 & 0.1 \\
    0 & 0 & 0.85
\end{bmatrix}, \quad B = \begin{bmatrix}
    0 & 0 \\
    0 & 1 \\
    1 & 0
\end{bmatrix}
$$

$$
K = \begin{bmatrix}
    0 & 1 \\
    0.1 & 0.1 \\
    0 & 0.2
\end{bmatrix}, \quad C = \begin{bmatrix}
    3 & 0 & -0.6 \\
    0 & 1 & 1 \\
\end{bmatrix}, \quad D = \begin{bmatrix}
    0 & 0 \\
    0 & 0
\end{bmatrix}
$$

Let the input $u$ and the noise $e$ be white noise vectors with means 0 and covariances $\Sigma_u = \text{diag}(1, 1)$ and $\Lambda = \text{diag}(1/4, 1/4)$, respectively. For $N = 4000$, we compare the results of identification by the direct CCA (regression approach [12]) and the LQD-based method for the row indices $k = 10, 15, 20, 25$.

Though not shown here, identification results of the deterministic component $P(z) = D + C(zI - \Lambda)^{-1} B$ by both methods are quite good and similar. Hence, in the following, we show the identification results of the stochastic component $H(z) = I + C(zI - A)^{-1} K$ and $\Lambda$, where

$$
H_{11}(z) = (1 - 2.35 z^{-1} + 2.275 z^{-2} - 0.85 z^{-3})/d(z)
$$

$$
H_{12}(z) = (0.18 z^{-1} - 0.015 z^{-2} - 0.024 z^{-3})/d(z)
$$

$$
H_{21}(z) = (0.1 z^{-1} - 0.235 z^{-2} + 0.1275 z^{-3})/d(z)
$$

$$
H_{22}(z) = (1 - 2.15 z^{-1} + 1.625 z^{-2} - 0.4395 z^{-3})/d(z)
$$

$$
d(z) = 1 - 2.35 z^{-1} + 1.975 z^{-2} - 0.595 z^{-3}
$$

and where the poles are $0.85, 0.75 \pm 0.3708j$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>Direct CCA $\Sigma$</th>
<th>LQD-based method $\Sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.2819 0.0113 0.2568 0.0030</td>
<td>0.2819 0.0113 0.2568 0.0030</td>
</tr>
<tr>
<td>15</td>
<td>0.2653 0.0057 0.2473 0.0006</td>
<td>0.2653 0.0057 0.2473 0.0006</td>
</tr>
<tr>
<td>20</td>
<td>0.2570 0.0027 0.2424 0.0004</td>
<td>0.2570 0.0027 0.2424 0.0004</td>
</tr>
<tr>
<td>25</td>
<td>0.2534 0.0012 0.2395 0.0001</td>
<td>0.2534 0.0012 0.2395 0.0001</td>
</tr>
</tbody>
</table>

Table I shows the covariance $\Lambda$ estimated by the two methods, where the estimates are the mean of 30 simulation runs. We see that the estimates by the direct CCA get better as $k$ increases, and that, though not bad, the values of estimates by the LQD-based method get smaller as $k$ increases. The latter is observed as the ratio $k/N$ increases; see (24). But, this could be avoided by taking a future horizon shorter than the past horizon.

Table II shows the estimates of the steady state gain, where we cannot see much difference in numerical values. To make the difference visible, Bode plots of the identified models are also shown in Figs. 1 3. We see that Bode plots of $H_{11}$ and $H_{21}$ identified by the LQD-based method is less biased than those by the direct CCA method. In particular, a notch of $H_{11}$ around 0.7 rad/sec is better estimated by the LQD-based...
method. This implies that the block lower triangular structure of $R_{33}$ in Theorem 1 is effectively used in the new subspace identification Algorithm B. If $k$ gets larger, however, it seems that the difference of the two method disappears as shown in Fig. 3. This fact may be further clarified analytically by using nonstationary (or finite-interval) Kalman filter approach [7].

VIII. CONCLUSIONS

We have examined a special role of LQ decomposition in subspace identification of stochastic systems with exogenous inputs. In particular, under the assumptions that the past horizon is sufficiently large, we have shown that $R_{33}$ has a block lower triangular structure as the number of data goes to infinity. Adapting the theoretical result to finite input-output data, we have then developed a method of computing the steady state Kalman gain and the covariance of innovation process, where $K$ is computed by a method similar to computing $(B, D)$ parameters in the MOESP method. We have thus completed a realization-based subspace identification method for combined deterministic-stochastic systems by fully utilizing the property of LQ decomposition of the data matrix. Moreover, we have briefly treated the case where there exist no exogenous inputs. Also, numerical results show the applicability of the LQD-based method.

We need to derive finite-interval results to study the robustness of the algorithms to the length of past horizon.

IX. ACKNOWLEDGMENTS

The author is grateful to H. Tanaka for his valuable comments and to M. Takeno for assistance in programming.

REFERENCES


![Fig. 1. Bode plots: k = 10. (a) Direct CCA. (b) LQD-based method.](image1)

![Fig. 2. Bode plots: k = 15. (a) Direct CCA. (b) LQD-based method.](image2)

![Fig. 3. Bode plots: k = 20. (a) Direct CCA. (b) LQD-based method.](image3)