Technical Report 2000–005
Department of Applied Mathematics and Physics
KYOTO UNIVERSITY

# A Comparison of Two Stochastic Subspace System Identification Methods

**Hidetoshi Kawauchi**†, **Alessandro Chiuso**‡
**Tohru Katayama**†* and **Giorgio Picci**‡

† *Department of Applied Mathematics and Physics, Graduate School of Informatics*

*Kyoto University, Kyoto, 606-8501, Japan*

E-mail: kawauchi@kuamp.kyoto-u.ac.jp, katayama@kuamp.kyoto-u.ac.jp

‡ *Dipartimento di Elettronica e Informatica, University of Padova, 35131 Padova, Italy*

E-mail: chiuso@dei.unipd.it, picci@dei.unipd.it

**Abstract**

In this paper, we analyze two subspace identification methods. The one is an N4SID-like algorithm which performs poorly in certain conditions where the past signal and future input spaces are nearly parallel. The other method, based on a preliminary orthogonal decomposition of output data space, is more robust and reliable than the first method in critical cases. Numerical results demonstrate a substantial improvement of performance in such a parallel case.

**Keywords**: canonical correlation analysis, orthogonal decomposition, parallel case, LQ factorization.

*Corresponding author

# 1 Introduction

We consider the identification problem for combined deterministic-stochastic systems with exogenous inputs as a continuation of the work presented at IFAC 99 by Chiuso and Picci[1]. The purpose of this paper is not to derive a new identification algorithm, but to make a comparison of two subspace identification methods: the one is an N4SID-like algorithm proposed by Katayama and Picci[4, 5], and the other is the orthogonal decomposition method due to Picci and Katayama[7, 8]. The former algorithm based on the classical canonical correlation analysis (CCA) contains oblique projections and needs the assumption that future inputs space and past signal space are not (nearly) parallel in the sense that the minimal principal angle [2] between the two spaces is not (nearly) zero. This situation may correspond, for example, to the case of inputs of relatively low power in the frequency band where the stochastic subsystem has resonances and the stochastic power spectrum of the output is relatively large with respect to the power spectrum of the deterministic input.

We consider the case where the assumption does not hold, and investigate the effects of this ill-conditioning on the estimates of system matrices. We show that the orthogonal decomposition method is more reliable and can successfully handle the case where the first algorithm fails. The results of simulations reported below show ill-conditioning effects on estimating deterministic and stochastic transfer functions. We can see from simulation results that under the ill-conditioning, the CCA algorithm has difficulties in separating the stochastic and deterministic subsystems and the estimates of the deterministic transfer functions become rather unrealistic; however, by using the orthogonal decomposition algorithm the estimates may become more realistic.

The orthogonal decomposition algorithm gives a state space model with the block-decoupled canonical form, see Section 3, or see [8] for more details. The choice of basis in the state space, the crucial step in subspace identification, can in this framework be made independently for the stochastic and the deterministic state spaces, thereby allowing the choice of the most convenient canonical (or pseudo-canonical) form for the two subsystems. In particular we can choose a state vector made of suitable canonical variates for the stochastic subsystem so as to get a stochastically balanced canonical form, while for the deterministic subsystem we can choose a suitably weighted deterministic balanced canonical form. The use of these canonical forms is instrumental in the order reduction step of the algorithm. It should be noted that the model reduction step in virtually all real-data subspace identification algorithms, commonly consisting in discarding the dynamics associated with "small" singular values, is well-behaved (positivity and stability-preserving etc.) only if the appropriate balanced truncation is used[3].

The organization of this paper is as follows. In Section 2, we review the realization of stochastic systems with exogenous inputs and discuss the effect of ill-conditioning of subspace identification by the CCA algorithm. In Section 3, an orthogonal decomposition identification algorithm is discussed. Section 4 presents numerical results for two cases: one is a white noise input case where the past signals and future input spaces are not parallel, and the other is a nearly parallel case. We can see that the CCA algorithm behaves well in the non-parallel case, but the estimates of transfer functions may become unrealistic in the parallel case. On the other hand, the results by orthogonal decomposition show that better estimates can be obtained even in such a parallel case. Section 5 contains the conclusions.

# 2 The CCA method

The realization for stochastic systems with exogenous inputs [4] is based on the classical CCA technique. We start with computing oblique projections of the future outputs on the space spanned by future inputs along the space spanned by past inputs and outputs, and of the future outputs on the past signal space along future inputs space. After computing these oblique projections, we can estimate the (extended) observability and controllability matrices by using conditional covariance matrices obtained from these projections. Using the controllability matrix, we obtain the estimates of states, and consequently determine system matrices by the least-squares (LS) method. The basic assumption is that there is no *feedback* from output $y$ to input $u$. See [8] for a discussion of this condition.

The resulting state space representation is given by the innovation model appearing in Kalman filtering theory, namely

$$\begin{aligned}\hat{x}(t+1) &= A\hat{x}(t) + Bu(t) + Ke(t)\\ y(t) &= C\hat{x}(t) + Du(t) + e(t)\end{aligned}$$

where $K$ is the steady state Kalman gain obtained by solving an algebraic Riccati equation (ARE).

The stochastic subspace identification algorithm is derived by adapting the realization theory to finite input-output data. Assume we have two finite time-series

$$\begin{gathered}\{u_0, u_1, \cdots, u_t, \cdots, u_{2k+N-1}\}\\ \{y_0, y_1, \cdots, y_t, \cdots, y_{2k+N-1}\}\end{gathered} \tag{1}$$

with $u_t \in \mathcal{R}^m$, $y_t \in \mathcal{R}^p$, measured during the experiment. Here, we shall assume that the data are generated by a true underlying system of order $n$, and that the sample size $N$ is large enough with $N \gg k$. It is assumed that the order $n$ is known so that we have no need to discuss order estimation.

With a present time $t = k$ fixed, input block Hankel matrices $U_{0|k-1}$, $U_{k|2k-1}$ of dimension $mk \times (N+1)$ are defined as

$$U_{0|k-1} := \begin{bmatrix} u_0 & u_1 & \cdots & u_N \\ u_1 & u_2 & \cdots & u_{N+1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k-1} & u_k & \cdots & u_{k+N-1} \end{bmatrix}$$

$$U_{k|2k-1} := \begin{bmatrix} u_k & u_{k+1} & \cdots & u_{k+N} \\ u_{k+1} & u_{k+2} & \cdots & u_{k+N+1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{2k-1} & u_{2k} & \cdots & u_{2k+N-1} \end{bmatrix}.$$

where $U_{0|k-1}$ denotes the past input data matrix and $U_{k|2k-1}$ denotes the future input data matrix. The row spaces $\mathcal{U}_{0|k-1}$, $\mathcal{U}_{k|2k-1}$ are the past and future input spaces generated by the past and future inputs, respectively. The block Hankel matrices formed by the output $y_t$ are defined as $Y_{0|k-1}$ and $Y_{k|2k-1}$ in the same way. In addition, as the past signal data matrix we shall define

$$P_{0|k-1} = \begin{bmatrix} U_{0|k-1} \\ Y_{0|k-1} \end{bmatrix}$$

and $\mathcal{P}_{0|k-1}$ as the past signal space.

We now explain the CCA algorithm, which is an N4SID-like algorithm proposed by Katayama and Picci[4, 5]. The basic object which is computed first in the CCA algorithm is the optimal predictor matrix of the future output based on the past observations and future inputs, namely

$$\begin{aligned} \hat{Y}_{k|2k-1} &:= \hat{E}\{Y_{k|2k-1} \mid \mathcal{P}_{0|k-1} \vee \mathcal{U}_{k|2k-1}\} \\ &= \Pi P_{0|k-1} + \Phi U_{k|2k-1}. \end{aligned}$$

It is shown that $\Pi$ and $\Phi$ satisfy the discrete Wiener-Hopf type equations

$$\Pi \Sigma_{pp|u} = \Sigma_{fp|u}, \qquad \Phi \Sigma_{uu|p} = \Sigma_{fu|p} \tag{2}$$

where each $\Sigma$ is a conditional covariance matrix and $f$, $u$ and $p$ respectively denotes future output, future input and past signal. The conditional covariance matrices above are obtained by computing projections of measured input and output data

$$\Sigma_{ab|c} = E\{(a|c^{\perp})(b|c^{\perp})'\} = \Sigma_{ab} - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{cb}' \tag{3}$$

where $(a|c^{\perp}) := a - \hat{E}\{a|c\}$. Because the data has finite length, the conditional covariance matrices are computed approximately, and in fact by using LQ factorization of Van Overschee and De Moor [9]. Estimates of $\Pi$ and $\Phi$ are obtained by solving equations (2). Using these $\Pi$, $\Phi$ and conditional covariance matrices $\Sigma$, system parameters $A$, $B$, $C$, $D$ and $K$ are obtained by developing subspace method. The procedure is as follows. See [4, 5] for a detailed discussion of this algorithm.

*Step 1*: Assume that the conditional covariance matrices $\Sigma_{ff|u}$, $\Sigma_{pp|u}$ and $\Sigma_{fp|u}$ are given by applying LQ factorization to the data matrices. Compute the SVD of normalized conditional covariance matrix

$$\Sigma_{ff|u}^{-1/2}\Sigma_{fp|u}\left(\Sigma_{pp|u}^{-1/2}\right)' = USV' \approx \hat{U}\hat{S}\hat{V}'$$

to get

$$\Sigma_{fp|u} \approx \Sigma_{ff|u}^{1/2}\hat{U}\hat{S}\hat{V}'\left(\Sigma_{pp|u}^{1/2}\right)'$$

where $\hat{S}$ is obtained by neglecting singular values which are "nearly zero," so that the dimension of the state vector equals dim $\hat{S}$.

*Step 2*: Set the extended observability and controllability matrices as

$$\mathcal{O}_k = \Sigma_{ff|u}^{1/2} \hat{U} \hat{S}^{1/2}, \quad \mathcal{C}_k = \hat{S}^{1/2} \hat{V}' \left(\Sigma_{pp|u}^{1/2}\right)'.$$

*Step 3*: The estimate of the state vector is given by

$$\hat{X}_k = \mathcal{C}_k \Sigma_{pp|u}^{-1} P_{0|k-1} = \hat{S}^{1/2} \hat{V}' \Sigma_{pp|u}^{-1/2} P_{0|k-1}.$$

Note that this state vector is an $n \times (N+1)$ matrix

$$\hat{X}_k = \begin{bmatrix} \hat{x}_k & \hat{x}_{k+1} & \cdots & \hat{x}_{N+k} \end{bmatrix}.$$

*Step 4*: The matrices $(A, B, C, D)$ are obtained by applying the LS method to the overdetermined equations

$$\begin{bmatrix} \hat{X}_{k+1} \\ Y_{k|k} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{X}_k \\ U_{k|k} \end{bmatrix} + \begin{bmatrix} W_k \\ E_k \end{bmatrix} \tag{4}$$

where $\hat{X}_{k+1}$ is the CCA state vector computed at time $k+1$. In order to be the time update of the CCA state selected at time $k$, this vector should be transformed by a change of basis so as to keep the same observability matrix at time $k$ and $k+1$, see Van Overschee and De Moor [9]. If the integer $k$ is large enough, the state vector is approximately stationary. Then $\hat{X}_{k+1}$ may be approximately given by the one-step shifted version of the vector $\hat{X}_k$ obtained in Step 3.

The matrices $W_k$, $E_k$ are residuals which represent LS estimates of system and observation noise terms.

*Step 5*: Define the covariance matrices of residuals as

$$\begin{bmatrix} \Sigma_{ww} & \Sigma_{we} \\ \Sigma_{ew} & \Sigma_{ee} \end{bmatrix} = \frac{1}{N+1} \begin{bmatrix} W_k W_k' & W_k E_k' \\ E_k W_k' & E_k E_k' \end{bmatrix}$$

and solve the ARE

$$P = APA' - (APC' + \Sigma_{we})(CPC' + \Sigma_{ee})^{-1}(APC' + \Sigma_{we})' + \Sigma_{ww} \tag{5}$$

to get the stabilizing solution $P \geq 0$ and the corresponding Kalman gain

$$K = (APC' + \Sigma_{we})(CPC' + \Sigma_{ee})^{-1}$$

making $A - KC$ is stable.

This procedure is correct for infinitely long data where $N \to \infty$. In this case, it follow from Lemma 3 of Katayama and Picci[4] that the exact relations $\Sigma_{ee} = E\{e(t)e(t)'\}$, $\Sigma_{ww} = K\Sigma_{ee}K'$, $\Sigma_{we} = K\Sigma_{ee}$ should hold, so that the unique stabilizing solution of the ARE exists and is actually zero.

For the finite data case, however, these exact relations do not hold and the sample covariance matrices computed in Step 5 vary with $k$. Note that the ARE of (5) has a unique stabilizing solution $P \geq 0$ from which we can estimate $K$. This is so, since by construction of the extended observability matrix $\mathcal{O}_k$, the pair $(C, A)$ is observable and we have $W_k W_k' > 0$ generically.

In the rest of this section, we will discuss the effects of the near-parallelism of the data on the estimation of the system matrices. Note that the "near-parallelism" means that the minimum principal angle [2] between the two spaces is nearly zero. We shall henceforth assume that the infinite data are available, namely, $N \to \infty$. Therefore it should be noted that the matrices $\hat{X}_{k+1}$ and $\hat{X}_k$ are $n \times \infty$, $Y_{k|k}$ is $p \times \infty$ and $U_{k|k}$ is $m \times \infty$. In Step 4, the sensitivity of the solution to (4), i.e. the variance of the parameter estimates, may depend on the condition number of the matrix

$$\begin{bmatrix} \hat{X}_k \\ U_{k|k} \end{bmatrix} \in \mathcal{R}^{(n+m)\times\infty}. \tag{6}$$

This fact means that the minimum singular value of this matrix should not be nearly zero in order to get a good estimate. It may be noted that the computation of system matrices $A$, $B$, $C$ and $D$ by applying the LS method to (4) is equivalent to solving the following discrete Wiener-Hopf type equation

$$\begin{bmatrix} \Sigma_{x'x} & \Sigma_{x'u} \\ \Sigma_{yx} & \Sigma_{yu} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \Sigma_{xx} & \Sigma_{xu} \\ \Sigma_{ux} & \Sigma_{uu} \end{bmatrix} \tag{7}$$

involving various covariance matrices $\Sigma$ with $x$, $x'$, $u$ and $y$ corresponding to $\hat{X}_k$, $\hat{X}_{k+1}$, $U_{k|k}$ and $Y_{k|k}$ respectively. Since it is assumed that the infinite data are available, these covariance matrices are correct. The sensitivity of the solution of (7) depends on the condition number of the symmetric matrix

$$\begin{bmatrix} \Sigma_{xx} & \Sigma_{xu} \\ \Sigma_{ux} & \Sigma_{uu} \end{bmatrix} \in \mathcal{R}^{(n+m)\times(n+m)} \tag{8}$$

Though it is difficult to examine the condition number of the above matrix, we have the following relation

$$\det \begin{bmatrix} \Sigma_{xx} & \Sigma_{xu} \\ \Sigma_{ux} & \Sigma_{uu} \end{bmatrix} = \det(\Sigma_{xx}) \det(\Sigma_{uu|x}) \tag{9}$$

where

$$\Sigma_{uu|x} = \Sigma_{uu} - \Sigma_{ux}\Sigma_{xx}^{-1}\Sigma_{xu} \tag{10}$$

Thus the (near) rank deficiency of the block matrix (8) is due to the (near) rank deficiency of $\Sigma_{uu|x}$, since it is obvious that no singular values of the matrix $\Sigma_{xx}$ are very small due to the procedure of defining $\hat{X}_k$ in Step 3. This means that the first $n$ rows of the matrix (8) are linearly independent and consequently the matrices $A$ and $C$ are estimated correctly. It can be seen from (10) that if the future inputs are highly correlated with the present states, the determinant of $\Sigma_{uu|x}$ becomes very small. It therefore follows from (9) that there exists a possibility that the block matrix (8) has a (near) rank deficiency, leading to the near parallelism of the last $m$ rows of the block matrix (8). From this reason, the near-parallelism of the data may affect the estimation of the matrices $B$ and $D$. The situation is relatively safe if $u$ is a white noise process; in this case formula (10) reduces to $\Sigma_{uu|x} = \Sigma_{uu}$.

The orthogonal decomposition method, discussed in the next section, do not involve the oblique projection so that it will avoid such situation where the near-parallelism of the data affects the estimation.

# 3 The orthogonal decomposition method

In this section, we present the orthogonal projection method based on the stochastic realization theory due to Picci and Katayama [8]. Let $\mathcal{U}$ be the space spanned by $\{u(t), t = 0, \pm 1, \cdots\}$, and $\mathcal{U}_t^-$ the space spanned by $\{u(s), s < t\}$. It is shown under the feedback-free condition [8] that the stationary process $y$ has an orthogonal decomposition

$$y(t) = y_s(t) + y_u(t), \quad y_s(t) \perp y_u(t)$$

where $y_u(t)$ and $y_s(t)$ are respectively called the "deterministic" and "stochastic" components of $y(t)$, which are defined by

$$y_u(t) = \hat{E}\{y(t) \mid \mathcal{U}\}, \quad y_s(t) = y(t) - y_u(t)$$

Thus a state space model of the output process $y$ can be obtained by just combining the two state space models for $y_u$ and $y_s$, namely,

$$\begin{bmatrix} x_u(t+1) \\ x_s(t+1) \end{bmatrix} = \begin{bmatrix} A_u & 0 \\ 0 & A_s \end{bmatrix} \begin{bmatrix} x_u(t) \\ x_s(t) \end{bmatrix} + \begin{bmatrix} B_u \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ B_s \end{bmatrix} e(t)$$
$$y(t) = \begin{bmatrix} C_u & C_s \end{bmatrix} \begin{bmatrix} x_u(t) \\ x_s(t) \end{bmatrix} + D_u u(t) + e(t)$$

where $e(t)$ is the one-step prediction error of the process $y_s$ based on its own past.

Now assume that we have two finite time series of (1). Thus the first step of the orthogonal decomposition method is to separate the output into the "deterministic" and "stochastic" components. This step is performed by computing the orthogonal projection of output onto the input space, namely

$$\begin{aligned} \hat{Y}^u_{0|2k-1} &:= \hat{E}\{Y_{0|2k-1} \mid \mathcal{U}_{0|2k-1}\} \\ \hat{Y}^s_{0|2k-1} &:= \hat{E}\{Y_{0|2k-1} \mid \mathcal{U}^\perp_{0|2k-1}\} \\ &= Y_{0|2k-1} - \hat{Y}^u_{0|2k-1} \end{aligned}$$

For this calculation we may take advantage of the LQ factorization. Hence the identification problem reduces to the separate problems: identification of the deterministic subsystem (i.e. $A_u, B_u, C_u, D_u$) using the data $\hat{Y}^u_{0|2k-1}$ and $U_{0|2k-1}$, and identification of the stochastic subsystem (i.e. $A_s, B_s, C_s$) using only the stochastic data of output $\hat{Y}^s_{0|2k-1}$.

**Lemma 1 ([8])** *Let $y_u$ be described by the deterministic subsystem. Then,*

$$\hat{E}\{y(t) \mid \mathcal{U}_{0|2k-1}\} = \hat{E}\{y_u(t) \mid \mathcal{U}_{0|2k-1}\} := \hat{y}_u(t)$$

*where $\hat{y}_u(t)$ is described by the same state space model but started at a different initial state, namely,*

$$\begin{aligned} \hat{x}_u(t+1) &= A_u\hat{x}_u(t) + B_u u(t), \\ \hat{y}_u(t) &= C_u\hat{x}_u(t) + D_u u(t), \\ \hat{x}_u(0) &= \hat{E}\{x_u(0) \mid \mathcal{U}_{0|2k-1}\}. \end{aligned} \tag{11}$$

**Proof:** Follows by noting that $y(t) = y_u(t) + y_s(t)$ and $y_s(t) \perp \mathcal{U}_{0|2k-1}$. ∎

It follows from this lemma that the deterministic subsystem matrices $(A_u, B_u, C_u, D_u)$ can be identified using standard deterministic identification procedure, for example, discussed in [11], or improved algorithm of [1]. Once the system matrices are computed, the estimate of the initial state $\hat{x}_u(0)$ must also be reconstructed. Because we can see from (11) that the estimate $\hat{x}_u(0)$ is clearly different from the true $x_u(0)$.

**Proposition 1** *Let $\tilde{y}_u(t) := y_u(t) - \hat{y}_u(t)$ (the "smoothing error" of $y_u(t)$), then*

$$\hat{y}_s(t) = y_s(t) + \tilde{y}_u(t), \quad 0 \le t \le 2k-1. \tag{12}$$

**Proof:** A proof is straightforward from $y(t) = y_u(t) + y_s(t) = \hat{y}_u(t) + \hat{y}_s(t)$. ∎

Hence for finite data length, the projection $\hat{y}_s(t)$ of the output on the complementary subspace $\mathcal{U}^{\perp}_{0|2k-1}$ does not coincide with the stochastic component $y_s(t)$, as it would instead have happened for infinite data length. The ideal projection $y_s(t)$ is affected by an additional "smoothing error" term $\tilde{y}_u(t)$ which depends on the error on the estimate of the initial state of the deterministic component, $\tilde{x}_u(0) := x_u - \hat{x}_u(0)$. It has the form

$$\tilde{y}_u(t) = C_u A_u^t \tilde{x}_u(0), \quad 0 \le t \le 2k-1.$$

This additional term is a source of difficulty in identifying of the stochastic part since if it is neglected it tends to produce a stochastic model of $y_s$ of a much higher than the true dimension $n_s$. In fact, due to the residual projection errors and noise effects, the estimated model will include also the dynamics of the deterministic subsystem, especially if the modes of $A_u$ do not decay to zero rapidly. From (12), we can write

$$\begin{bmatrix} \hat{y}_s(0) \\ \hat{y}_s(1) \\ \vdots \\ \hat{y}_s(2k-1) \end{bmatrix} = \begin{bmatrix} C_u \\ C_u A_u \\ \vdots \\ C_u A_u^{2k-1} \end{bmatrix} \tilde{x}_u(0) + \begin{bmatrix} y_s(0) \\ y_s(1) \\ \vdots \\ y_s(2k-1) \end{bmatrix}$$

and rewrite this equation as

$$\hat{\boldsymbol{y}}^s = \Gamma_{2k}\tilde{x}_u(0) + \boldsymbol{y}^s,$$

where $\Gamma_{2k}$ is the extended observability matrix. Solving by the LS method we have

$$\tilde{x}_u(0) = (\Gamma'_{2k}\Gamma_{2k})^{-1}\Gamma'_{2k}\hat{\boldsymbol{y}}^s \tag{13}$$

This estimate will not be optimal unless $y_s(t)$ is white noise, however it can be recomputed as a weighted LS estimate when the covariance of $Y^s_{0|2k-1}$ is available from the identification of the stochastic subsystem (so the estimation can be refined iteratively).

Denote the LS estimate of $Y^s_{0|2k-1}$ corresponding to (13) by $Z$. This matrix can be interpreted as a projection of $\hat{Y}^s_{0|2k-1}$ onto the orthogonal component of the column span of $\Gamma_{2k}$, i.e.

$$Z_{0|2k-1} = [I - \Gamma_{2k}(\Gamma'_{2k}\Gamma_{2k})^{-1}\Gamma'_{2k}]\hat{Y}^s_{0|2k-1}.$$

From this we see that a distortion of the "true" time-series $y_s$ is introduced. With some formal manipulations the last block-row of the estimated matrix $Z_{0|2k-1}$ can be seen as the output of a time invariant FIR filter which processes sequentially the sample values $\hat{y}_s(t)$, i.e.

$$z(t) = P(z^{-1})\hat{y}_s(t), \quad t \geq 2k-1$$

where the coefficients of $P(z^{-1})$ can be expressed in terms of the estimated parameters $(\hat{A}_u, \hat{C}_u)$. See Picci and Katayama [7] for a detailed description of this filtering step. The identification of the stochastic subsystem is done by processing the "filtered" data matrix $Z$ obtained in the previous step. The identification can be done by the algorithm described in Van Overschee and De Moor[9] or the CCA algorithm in Section 2. As explained in Lindquist and Picci[3], identification of state space models in a stochastically balanced canonical form is to be recommended in order to guarantee positivity after truncation by retaining only the larger singular values. This stochastically balanced form is automatic with the SVD procedure of the "stochastic" algorithm described in [9].

# 4 Simulation results

In this section we present simulation results for two models. The first model is proposed in [1] and the second in [12, 4].

*Example 1*

First we present simulation results which compare the CCA algorithm with the orthogonal decomposition algorithm. In both cases the order is supposed to be known so that no order estimation is performed. The transfer functions of deterministic and stochastic subsystems are

$$G_u(z) = \frac{0.1667 - 0.0500z^{-1} + 0.0917z^{-2} - 0.0542z^{-3}}{1 - 2.2000z^{-1} + 1.9525z^{-2} - 0.7043z^{-3}}$$
$$G_s(z) = \frac{1 - 0.20z^{-1} - 0.48z^{-2}}{1 + 0.4z^{-1} + 0.4z^{-2}}$$

and Figure 1 depicts the corresponding block diagram, where $u$ is the input and $y$ is the output. In order to show that the CCA algorithm behaves well in non-parallel case, the input signal $u$ for the deterministic subsystem is chosen as a white noise process with variance $\sigma_u^2 = 9$, see Figure 2(a) and the input signal $e$ for the stochastic subsystem as a white noise process with variance $\sigma_e^2 = 1$. The identification results by the CCA algorithm are shown in Figures 3, 4 and 5, where $N = 1000$, $k = 15$. Figure 3 shows that the poles are estimated very well and Figures 4 and 5 show that both deterministic and stochastic subsystems are estimated quite nicely. This implies that, for a white noise input, the CCA method gives the good estimates of both zeros and poles of deterministic and stochastic transfer functions.

Next we consider the parallel case in order to investigate how the parallelism of the data effects the estimation by the CCA algorithm. As a parallel case we choose the exogenous input as an ARMA process with colored spectrum as depicted in Figure 2(b). It should be noted that the stochastic subsystem has resonances in the frequency range where the input power spectrum is relatively low. Here the input process is chosen so that its variance is almost as large as that of the white noise input. Figure 7, which depicts the estimates of the deterministic subsystem, shows that although the amplitude of input power spectrum is quite large in the frequency band of the deterministic subsystem (in fact the shapes of the spectra are the same), the estimates of the deterministic subsystem may be rather poor. This happens in spite of the fact that the poles of the deterministic part are estimated rather well by the CCA algorithm, see Figure 6. This is due to the ill-conditioning of the estimates of $B$ and $D$, as seen from Proposition 1 in Section 2. On the other hand, the orthogonal decomposition algorithm avoids this kind of problems and also gives better performance in estimating the deterministic subsystem as shown in Figures 9 and 10. Moreover Figure 11 shows that the stochastic subsystem is identified more correctly.

*Example 2*

Next, we consider a fifth-order model with the block diagram shown in Figure 12, where $u$ is the input and $w$ and $v$ are additive white noises with zero mean. The transfer function is given by

$$G(z) = \frac{0.0275z^{-4} + 0.0551z^{-5}}{1 - 2.3443z^{-1} + 3.081z^{-2} - 2.5274z^{-3} + 1.241z^{-4} - 0.3686z^{-5}}.$$

Thus $G(z)$ has a zero at $z = -2$ and poles at $z = 0.9,\ 0.8e^{\pm j},\ 0.8e^{\pm 1.2j}$. The performance of each algorithm is measured by the norm of the estimation error

$$I_N = \frac{1}{M}\sum_{l=1}^{M}\left(\sum_{j=1}^{11}\left[\theta_j - \hat{\theta}_j(l,N)\right]^2\right)$$

$$N = 200,\ 500,\ 1000,\ 2000$$

where $\theta_j$ denotes the true parameter and $\hat{\theta}_j(l,N)$ is the estimate of $\theta_j$ at $l$th run with the number of data $N$, and $M$ denotes the number of simulation runs. In order to compare the effect of input signals, we consider two input processes. In the first case the input is a zero-mean white Gaussian noise with variance $\sigma_u^2 = 1$ and in the second case, the input is chosen as

$$u(t) = U_0 \sum_{i=1}^{20} \sin(\omega_i t)$$

where the frequencies $\omega_i$'s are uniformly spaced in the interval $(0.1,\ 3)$(rad) and where $U_0$ is adjusted to yield $\sigma_u^2 = 1$. The variances of $w$ and $v$ are $\sigma_w^2 = \sigma_v^2 = (0.05)^2$.

Figures 13, 14, 15 and 16 depict the simulation results for the case where the input is a white noise process and $N = 1000$ and the trancation index is $k = 15$. We see from these figures that the white noise input give good performance and we can say that in this case there is little difference between the estimation result using the CCA method and that using the orthogonal decomposition method.

On the other hand, Figures 17, 18, 19 and 20 depict the simulation results for the case where the input is the sinusoidal process and $N = 1000$, $k = 15$. Figures 17 and 19 show that both methods give good estimates of poles, while from the results of Figures 18 and 20 we see that the estimated transfer function by the CCA method is poorer than that by the orthogonal decomposition method in the lower frequency range. This implies that the zeros of the transfer function are estimated more correctly by the orthogonal decomposition method than the CCA method.

Figure 21 depicts the estimation errors $I_N$ vs. the number of the data $N = 200$, 500, 1000, 2000, where $M = 100$, $k = 15$. The solid lines denote the estimation errors by the orthogonal decomposition method and the dashed lines denote the estimation errors by the CCA method. In addition $\circ$ and $\times$ denote the white noise input and sinusoidal input, respectively. From this figure, in the case of white noise input both methods give similar performance irrespective of the number of data $N$. In the case of sinusoidal input the orthogonal decomposition method gives much better performance than the CCA method. This may imply that the convergence of the orthogonal decomposition method is quite rapid compared with the CCA method, especially for non-white noise inputs.

# 5 Conclusions

In this paper, we have analyzed the CCA based subspace method by Katayama and Picci[4, 5] and showed that it may perform poorly in a certain condition where the past signal and future input spaces are (nearly) parallel. However, we have showed that the orthogonal decomposition algorithm produces better estimates even under such ill-conditioned cases. This will be due to the fact that the latter algorithm contains computing orthogonal decomposition in order to separate measured output data into the deterministic and stochastic components, and the identification problem is divided into the estimation subproblems of identifying the deterministic and the stochastic subsystems separately.

# References


[1] A. Chiuso and G. Picci, "Subspace Identification by Orthogonal Decomposition," *Proc. 14th IFAC World Congress*, vol. H, pp. 241-246, 1999.

[2] G. H. Golub and C. R. Van Loan, *Matrix Computation (2nd ed.)*, The Johns Hopkins Univ. Press, 1989.

[3] A. Lindquist and G. Picci, "Canonical Correlation Analysis, Approximate Covariance Extension and Identification of Stationary Time Series," *Automatica*, **32**, pp. 709-733, 1996.

[4] T. Katayama and G. Picci, "Realization of Stochastic Systems with Exogenous Inputs and Subspace Identification Methods," *Automatica*, **35**, pp. 1635-1652, 1999.

[5] T. Katayama, S. Omori and G. Picci, "A Comparison of Some Subspace Identification Methods," *Proc. 37th IEEE Conference on Decision and Control*, pp. 1850-1852, 1998.

[6] H. Kawauchi, A. Chiuso, T. Katayama and G. Picci, "Comparison of Two Subspace Identification Methods for Combined Deterministic-Stochastic Systems," *Proc. 31st ISCIE International Symposium on Stochastic Systems Theory and its Applications,* Yokohama, Japan, 1999.

[7] G. Picci and T. Katayama, "A Simple Subspace Identification Method with Exogenous Inputs," *Proc. 13th IFAC World Congress*, vol. I, pp. 175-180, 1996.

[8] G. Picci and T. Katamaya, "Stochastic Realization with Exogenous Inputs and 'Subspace-Methods' Identification," *Signal Proccesing*, vol. 52, pp. 145-160, 1996.

[9] P. Van Overschee and B. De Moor, "Subspace Algorithm for the Stochastic Identification Problem," *Automatica*, **29**, pp. 649-660, 1993.

[10] P. Van Overschee and B. De Moor, "N4SID: Subspace Algorithm for the Identification of Combined Deterministic-Stochastic Systems," *Automatica*, **30**, pp. 75-93, 1994.

[11] M. Verhaegen, "Identification of the Deterministic Part of MIMO State Space Models Given in Innovations Form from Input-Output Data," *Automatica*, **30**, pp. 61-74, 1994.

[12] M. Viberg, "Subspace-Based Methods for the Identification of Linear Time-invariant Systems," *Automatica*, **31**, pp.1835-1851, 1995.

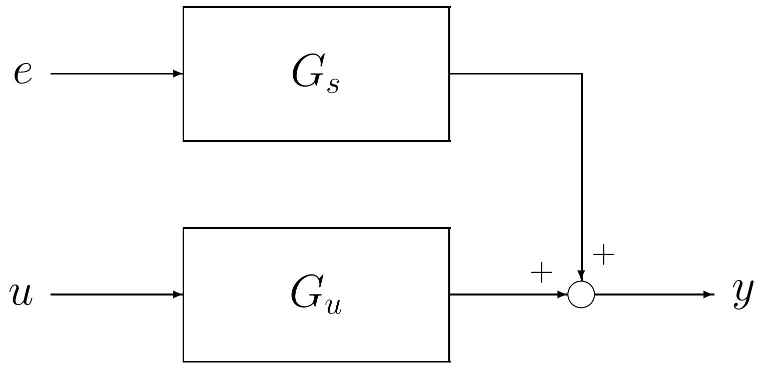


Figure 1: The plant model for Example 1.

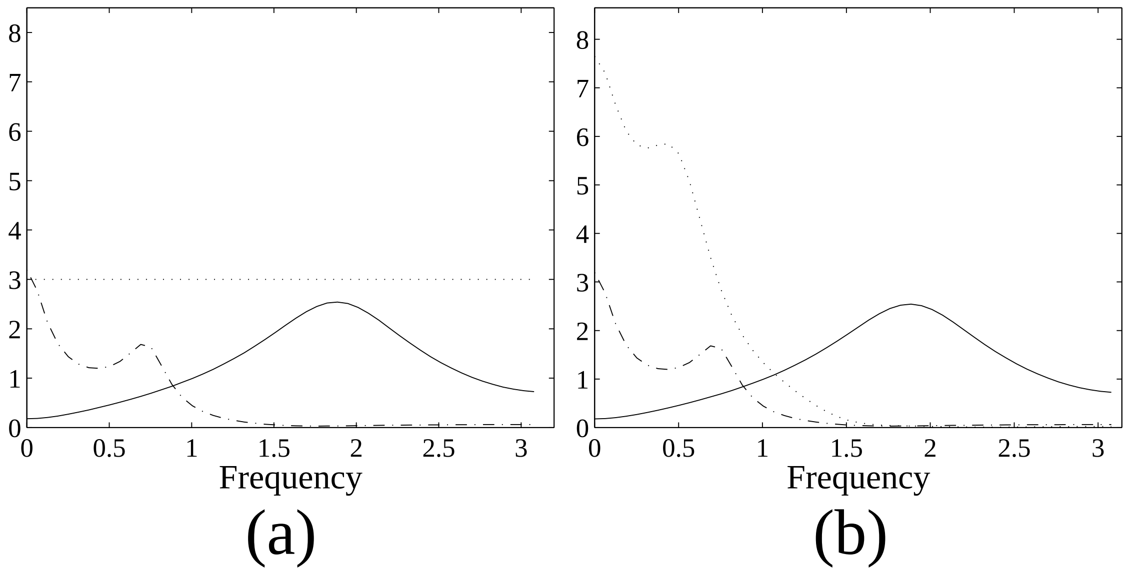


Figure 2: True spectra; (a) case 1; (b) case 2; solid: stochastic subsystem; dashed: deterministic subsystem; dotted: input.

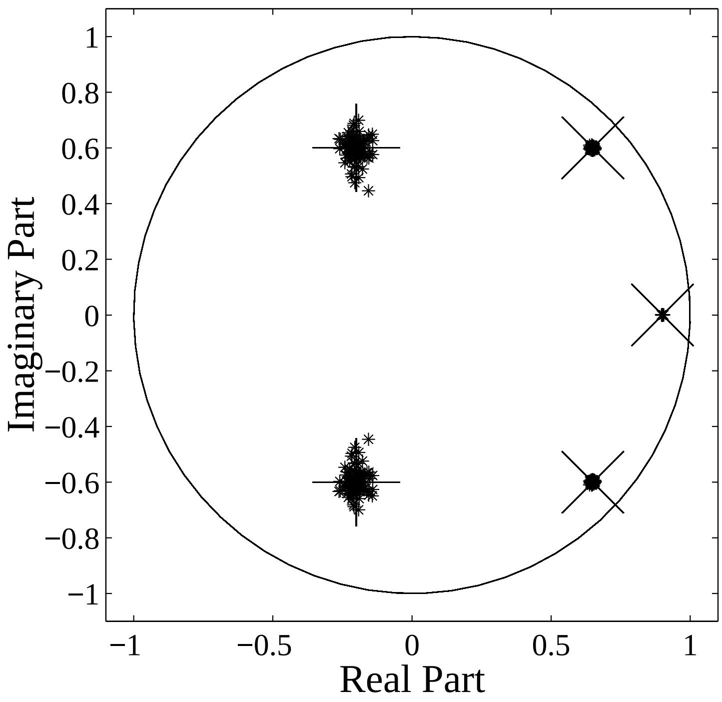


Figure 3: Non-parallel case: estimated poles in 100 trials by the CCA method, where $\times$ and $+$ denote poles of deterministic sybsystem and stoshastic subsystem, respectively.

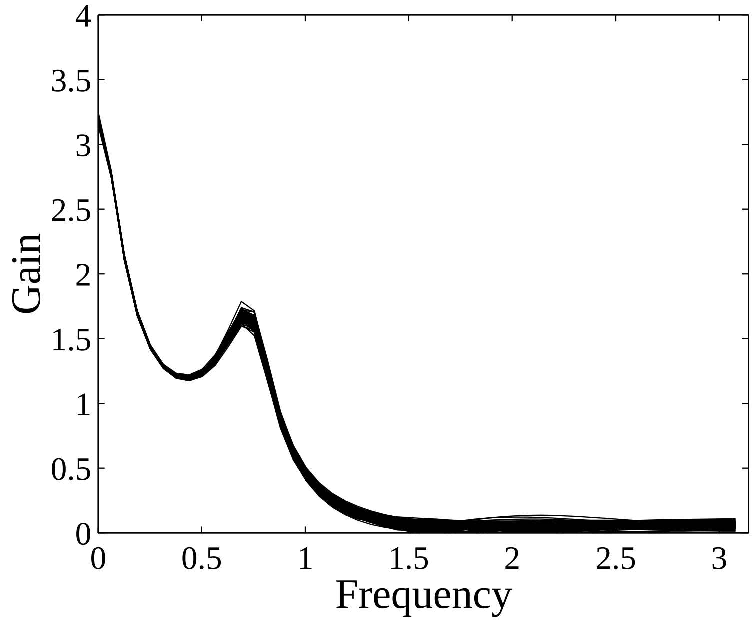


Figure 4: Non-parallel case: estimated deterministic subsystem in 100 trials by the CCA method.

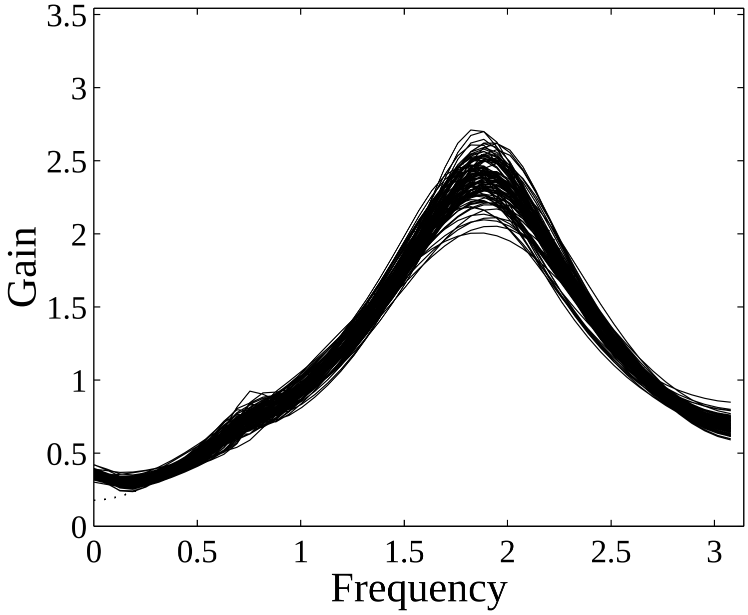


Figure 5: Non-parallel case: estimated stochastic subsystem in 100 trials by the CCA method.

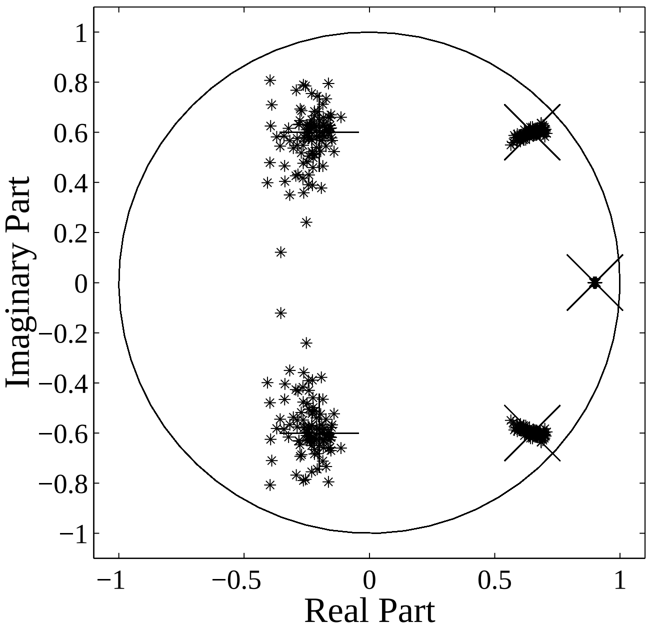


Figure 6: Parallel case: estimated poles in 100 trials by the CCA method, where $\times$ and $+$ denote poles of deterministic subsystem and stochastic subsystem, respectively.

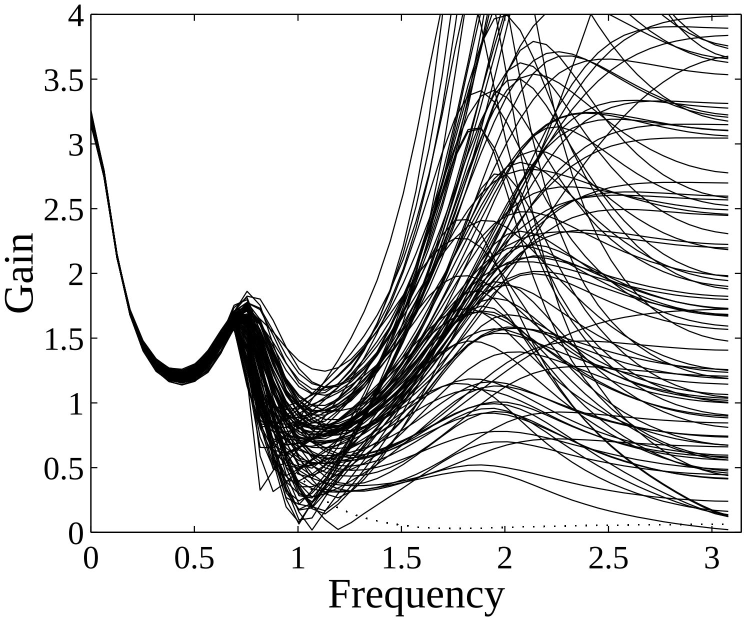


Figure 7: Parallel case: estimated deterministic subsystem in 100 trials by the CCA method.

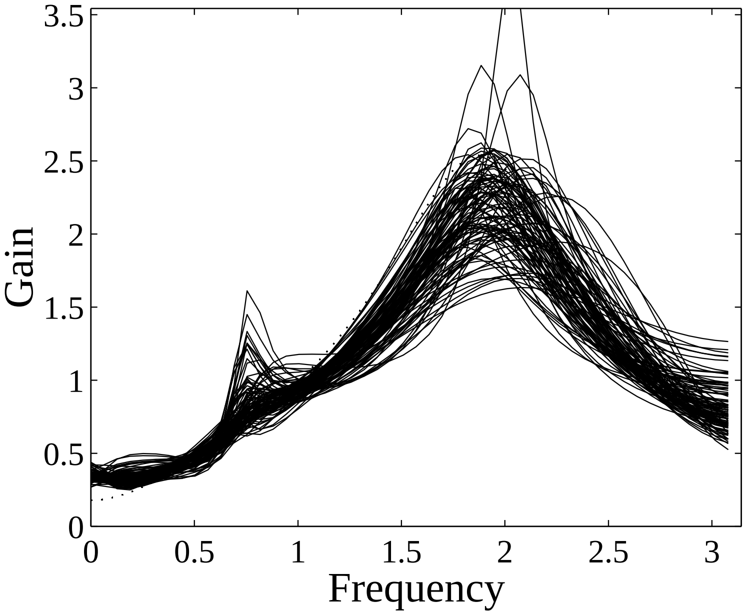


Figure 8: Parallel case: estimated stochastic subsystem in 100 trials by the CCA method.

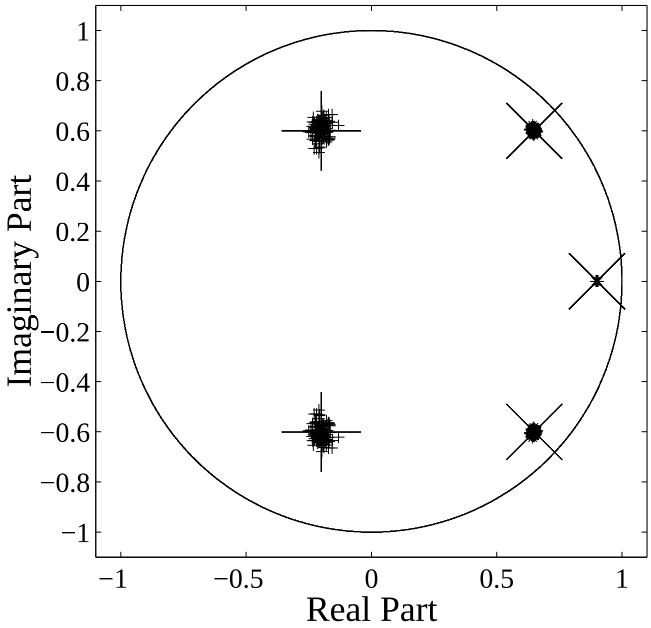


Figure 9: Parallel case: estimated poles in 100 trials by the orthogonal decomposition method, where $\times$ and $+$ denote poles of deterministic subsystem and stochstic subsytem, respectively.

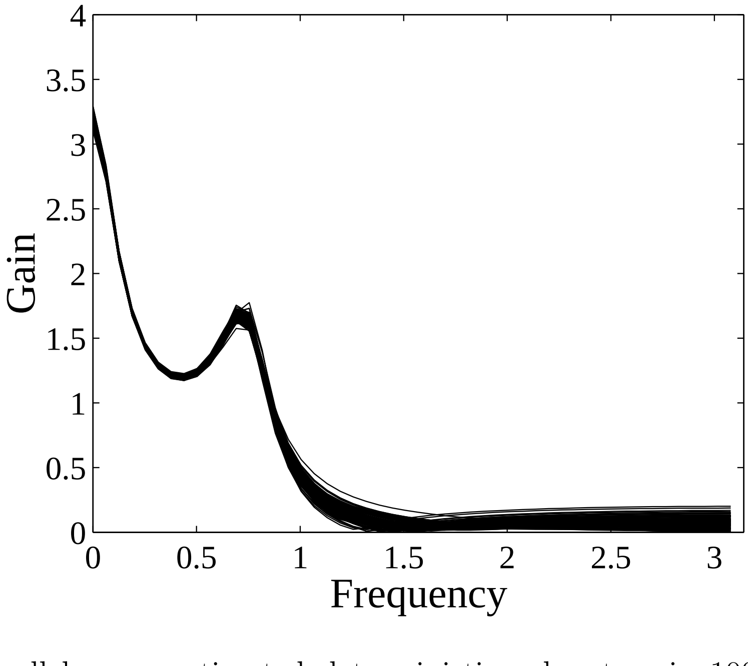


Figure 10: Parallel case: estimated deterministic subsystem in 100 trials by the orthogonal decomposition method.

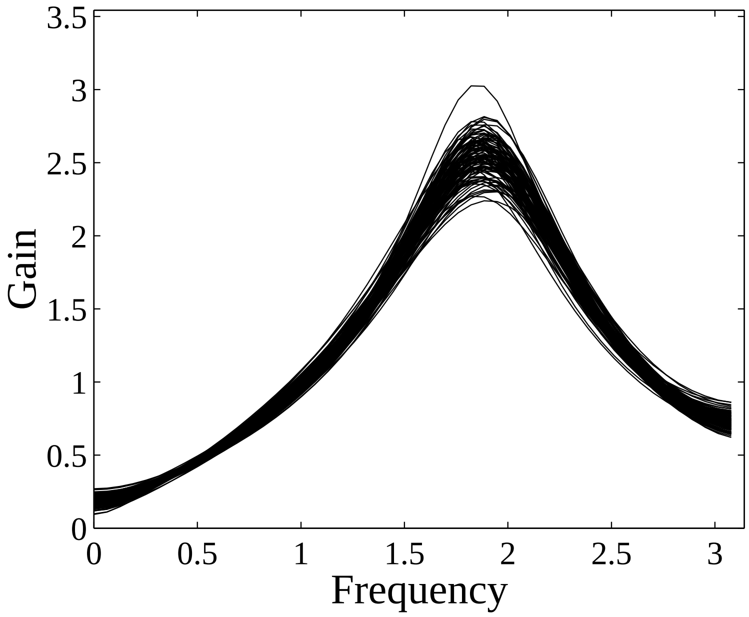


Figure 11: Parallel case: estimated stochastic subsystem in 100 trials by the orthogonal decomposition method.

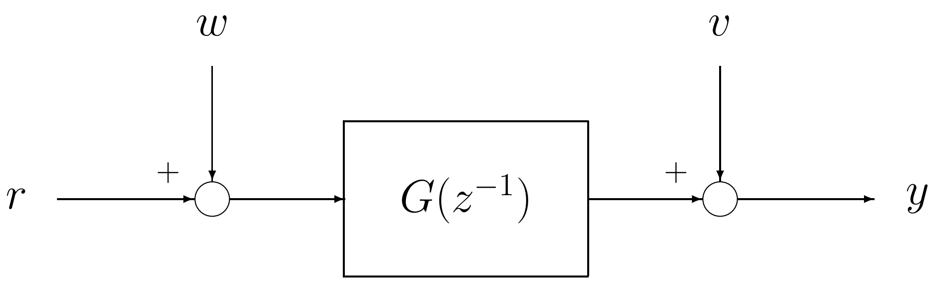


Figure 12: The plant model for Example 2

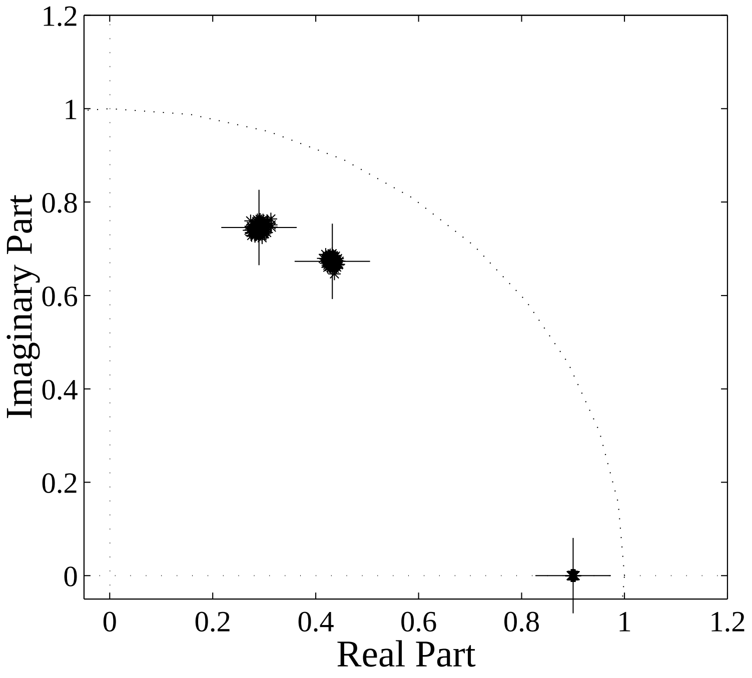


Figure 13: White noise input: estimated poles in 100 trials by the CCA method.

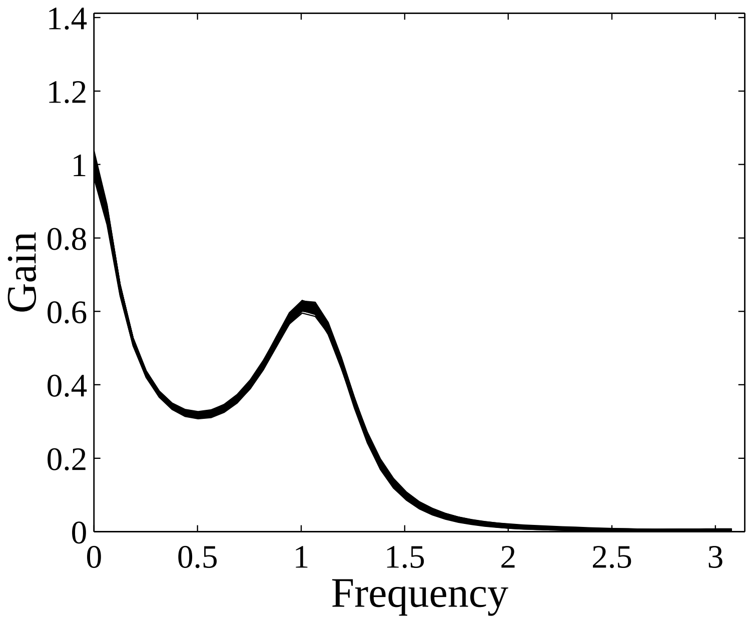


Figure 14: White noise input: estimated transfer function in 100 trials by the CCA method.

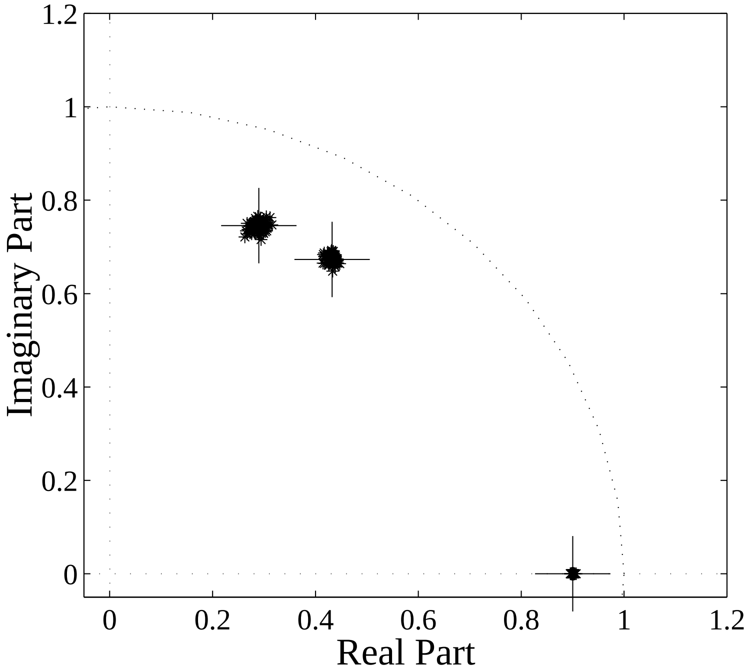


Figure 15: White noise input: estimated poles in 100 trials by the orthogonal decomposition method.

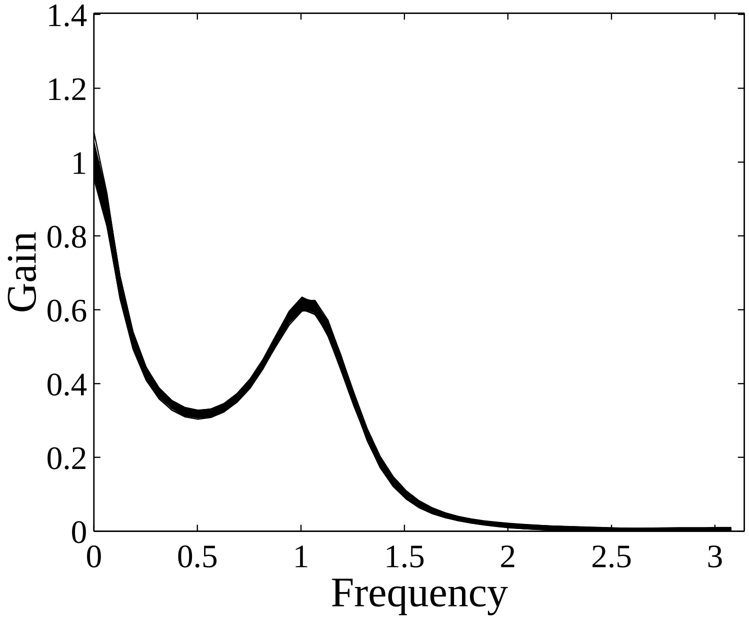


Figure 16: White noise input: estimated transfer function in 100 trials by the orthogonal decomposition method.

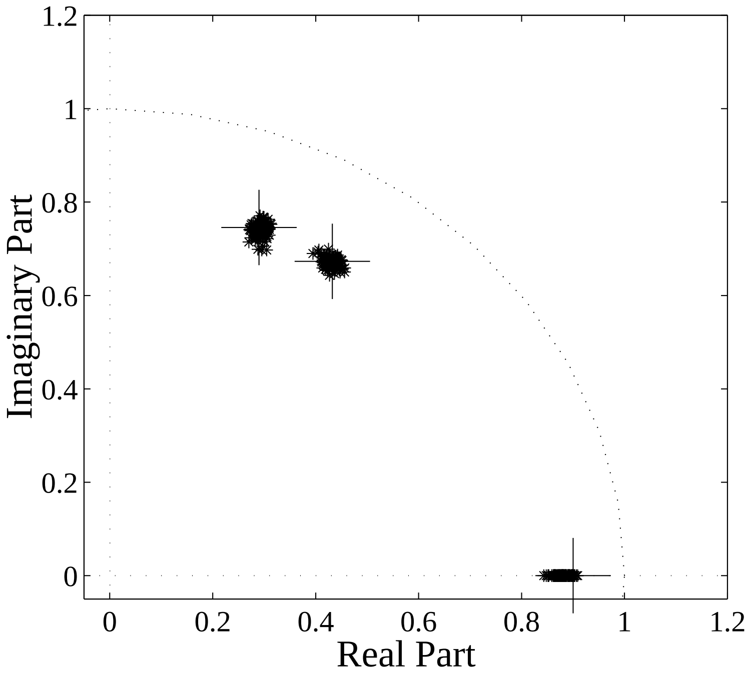


Figure 17: Sinusoidal input: estimated poles in 100 trials by the CCA method.

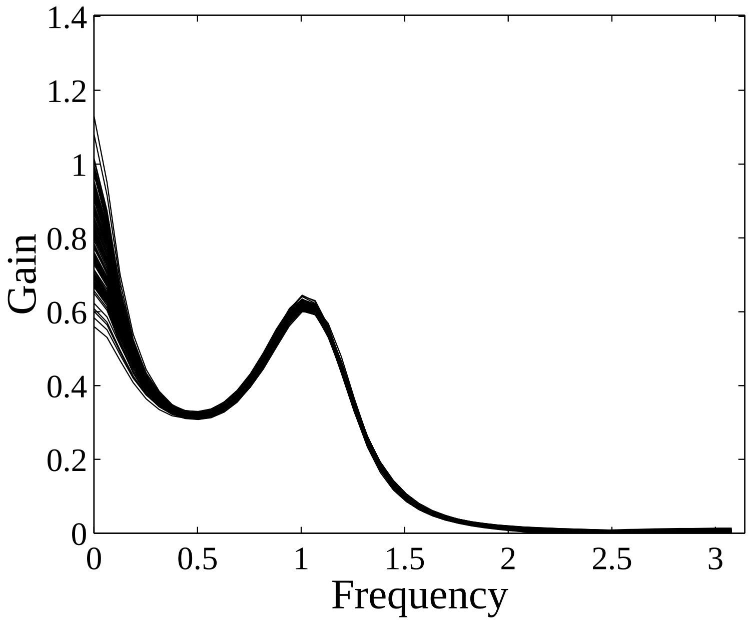


Figure 18: Sinusoidal input: estimated transfer function in 100 trials by the CCA method.

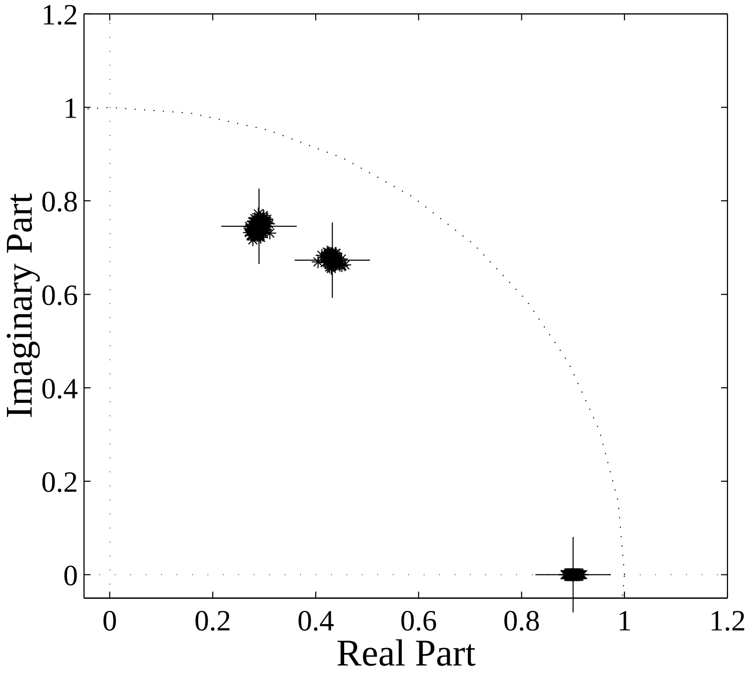


Figure 19: Sinusoidal input: estimated poles in 100 trials by the orthogonal decomposition method.

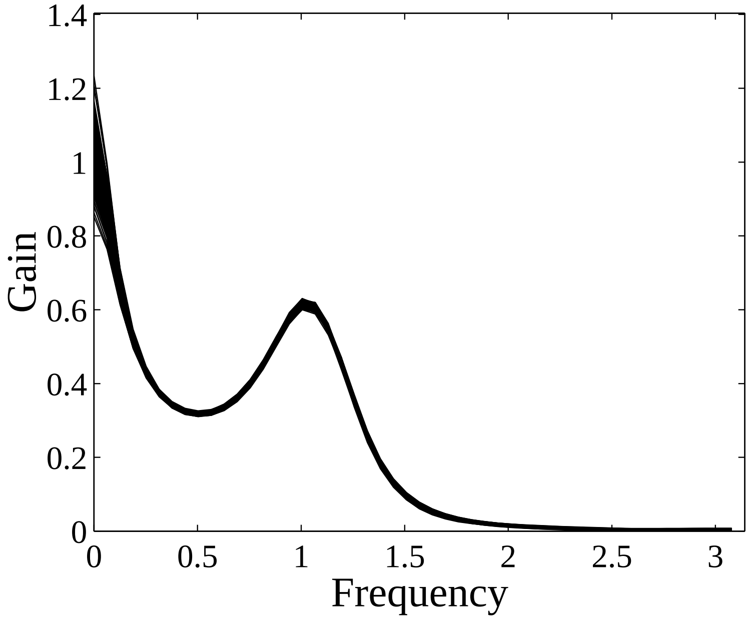


Figure 20: Sinusoidal input: estimated transfer function in 100 trials by the orthogonal decomposition method.

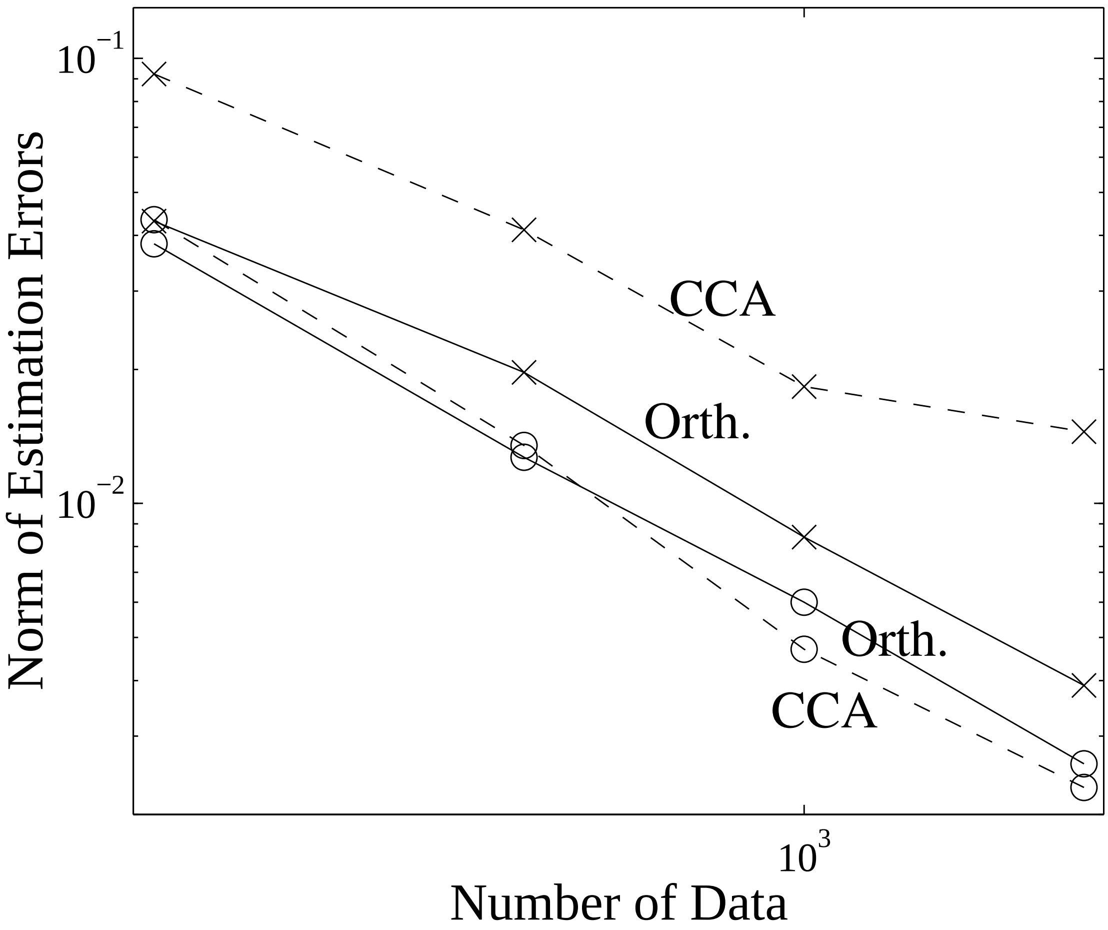


Figure 21: Performance of four cases; solid: the orthogonal decomposition method, dashed: the CCA method, ∘: white noise input case, ×: sinusoidal input case.