Linear Multivariable System Identification: Multi-experiments Case

Wael SULEIMAN André MONIN
LAAS - CNRS
7 avenue du Colonel Roche
31077 Toulouse Cedex 4
France
{*suleiman, monin*}*@laas.fr*

***Abstract*— In this paper, we propose a method for identifying the linear model of a system in the case of multi-experiments. The method is based on the minimization of output error cost function of all the input/output data sets simultaneously. The implementation of the proposed method is based on a local parameterization of the linear state space model in order to minimize the number of gradient search iterations. The optimization process is initialized by an extension of a classical subspace method. We show that, the estimated linear models provided by the proposed method are more accurate than that obtained by the actual methods of linear systems identification. Moreover, we figure out that, the method can handle the case of short multi-experiments by increasing the model's vector of parameters to also include the initial conditions of the internal states.**

***Index Terms*— Identification; linear systems; subspace methods; optimization.**

I. Introduction

Consider the linear state space model, which has the following structure

$$\begin{aligned} x_t &= Ax_{t-1} + Bu_t \\ y_t &= Cx_t + v_t \end{aligned} \tag{1}$$

where x_t, u_t, y_t and v_t are the internal state of the system, the input signal, the output signal and the measurement noise. The identification of the above model is extremely studied in the literature [3, 4, 7, 2].

On the other hand, the case of short multi-experiments is infrequently addressed. The raison why this is important is that the identification of real systems is done by measuring the output response to various input sequences. The objective of each sequence is exciting one or more modes of the system. In real experiments, sometimes we obtain short sets of input/output data that due to a large time sampling period or short tracked phenomena. In these cases, we should exploit all the data sets to obtain the model of the system. Moreover, in the case of short experiments, the effect of initial conditions can not be neglected, so we should also estimate them.

The paper is organized as follows, Section 2 gives an overview of PO-MOESP method, which is a classical subspace method. In Section 3 an extension of the PO-MOESP method is described in order to deal with multi-experiments case. The problem of output-error identification is defined in Section 4. Some illustrative examples are given in Section 5.

II. Overview of PO-MOESP method

The PO-MOESP method is a class of subspace model identification [6]. Given a linear state space model (1), the intent of the method is to calculate an estimation of the triple $[A, B, C]$. First, the input and output are stocked in Hankel matrices form

$$U_{1,\alpha,N} \triangleq \begin{pmatrix} u_1 & u_2 & \cdots & u_{N-\alpha+1} \\ u_2 & u_3 & \cdots & u_{N-\alpha+2} \\ \vdots & \vdots & \ddots & \vdots \\ u_\alpha & u_{\alpha+1} & \cdots & u_N \end{pmatrix} \tag{2}$$

where $(1, \alpha)$ are the first subscript refers to the index of the first data sample and the number of rows in the matrix respectively. N refers to the index of the last data sample in the matrix. The number of rows should be chosen to be greater than the system order n [7]. The output Hankel matrix $Y_{1,\alpha,N}$ and the noise Hankel matrix $V_{1,\alpha,N}$ are defined analogously. The following input-output equation is then easily derived from the system description (1)

$$Y_{1,\alpha,N} = \Gamma_\alpha X_{0,N-\alpha} + \Phi_\alpha U_{1,\alpha,N} + V_{1,\alpha,N} \tag{3}$$

Where Γ_α is the extended observability matrix of the system, Φ_α is a block-triangular matrix

$$\Gamma_\alpha = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{(\alpha-1)} \end{bmatrix}$$

$$\Phi_\alpha = \begin{bmatrix} CB & 0 & 0 & \cdots & 0 \\ CAB & CB & 0 & \cdots & 0 \\ CAB & CAB & CB & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ CA^{(\alpha-1)}B & \cdots & \cdots & CAB & CB \end{bmatrix}$$

and $X_{0,N-\alpha} = \begin{bmatrix} x_0 & x_1 & \cdots & x_{N-\alpha} \end{bmatrix}$.

PO-MOESP method uses the past input and output data as an instrumental variable to remove the effects of the noise term. Consider the following QR factorization

$$\begin{bmatrix} U_{1+\alpha,\alpha,N} \\ U_{1,\alpha,N-\alpha} \\ Y_{1,\alpha,N-\alpha} \\ Y_{1+\alpha,\alpha,N} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 & 0 \\ R_{21} & R_{22} & 0 & 0 \\ R_{31} & R_{32} & R_{33} & 0 \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} Q^T \quad (4)$$

where the R_{ii} are lower-triangular matrices and the column-unitary matrix Q ($Q^TQ = I$) is partitioned according to the dimension of lower-triangular matrices R_{ii} as

$$Q = \begin{bmatrix} Q_1 \ Q_2 \ Q_3 \ Q_4 \end{bmatrix}$$

If we consider the quantity

$$\begin{aligned} \begin{bmatrix} R_{42} \ R_{43} \end{bmatrix} &= Y_{1+\alpha,\alpha,N} \begin{bmatrix} Q_2 \ Q_3 \end{bmatrix} \\ &= (\Gamma_\alpha X_{\alpha,N-\alpha+1} + V_{1+\alpha,\alpha,N}) \begin{bmatrix} Q_2 \ Q_3 \end{bmatrix} \end{aligned}$$

it is not difficult to see that

$$\lim_{N\to\infty} \frac{1}{N} V^N_{1+\alpha,\alpha} \begin{bmatrix} Q_2 \ Q_3 \end{bmatrix} = 0 \quad w.p.1$$

where $w.p.1$ denotes with probability of one. Consequently, the effects of noise will vanish, thus Γ_α is extracted from the matrix $\begin{bmatrix} R_{42} \ R_{43} \end{bmatrix}$ using a singular value decomposition (SVD) as follows

$$\begin{bmatrix} R_{42} \ R_{43} \end{bmatrix} = U_n S_n V_n^T + U_n^\perp S_2 {V_n^\perp}^T \quad (5)$$

where the matrix S_n contains the principals singular values (further than threshold), the dimension of this matrix yields n (the dimension of the internal state x_t). Then the matrix Γ_α can be estimated as follow

$$\Gamma_\alpha = U_n S_n^{1/2} \quad (6)$$

and so we estimate the matrices C and A directly from Γ_α

$$\begin{aligned} C =& \Gamma_\alpha(1:p,:) \\ \Gamma_\alpha(1:(\alpha-1)p,:)A =& \Gamma_\alpha(p+1:p\alpha,:) \end{aligned} \quad (7)$$

where $\Gamma_\alpha(i:j,:)$ stands for the submatrix of Γ_α which contains the columns from i^{th} to j^{th} columns.
In order to find B, we consider the least-squares solution to the overdetermined system of equations

$$\left(U_n{}^\perp\right)^T \begin{bmatrix} R_{31} \ R_{32} \ R_{41} \end{bmatrix} = \left(U_n{}^\perp\right)^T \Phi_\alpha \begin{bmatrix} R_{21} \ R_{22} \ R_{11} \end{bmatrix} \quad (8)$$

which provides a consistent estimate of Φ_α, from which B is easily calculated.

III. Dealing with multiple data sets

It is clear that the model representation (3) holds for an arbitrary non-zero initial conditions. For that, non-zero initial conditions have no effect at all on the calculations of the triple $[A, B, C]$. As a result, dealing with multiple data sets does not introduce an additional problem. Therefore we adapt PO-MOESP to deal with multiple data sets as follows. Consider the following data sets

$$\left\{u_t^i, y_t^i\right\}_{t=1}^{N_i} \text{ and } i = 1, 2, \cdots, K \quad (9)$$

where each data set corresponds to one experiment. For each input/output data set, we obtain the following data equation

$$Y^i_{1,\alpha,N_i} = \Gamma_\alpha X^i_{1,N_i-\alpha+1} + \Phi_\alpha U^i_{1,\alpha,N_i} + V^i_{1,\alpha,N_i} \quad (10)$$

Then the data equations can easily be combined as follows

$$\begin{aligned} [Y_{1,\alpha,N_1} \cdots Y_{1,\alpha,N_K}] = \Gamma_\alpha \left[X^1_{1,N_1-\alpha+1} \cdots X^K_{1,N_K-\alpha+1}\right] + \\ \Phi_\alpha \left[U^1_{1,\alpha,N_1} \cdots U^K_{1,\alpha,N_K}\right] + \left[V^1_{1,\alpha,N_1} \cdots V^K_{1,\alpha,N_K}\right] \end{aligned} \quad (11)$$

The structure of this equation is similar to that of the original data equation (3), for that we still can use the main body of PO-MOSEP algorithm by computing QR factorization of the following matrix

$$\begin{aligned} \begin{bmatrix} U^1_{1+\alpha,\alpha,N_1} | U^2_{1+\alpha,\alpha,N_2} | \cdots | U^K_{1+\alpha,\alpha,N_K} \\ U^1_{1,\alpha,N_1-\alpha} | U^2_{1,\alpha,N_2-\alpha} | \cdots | U^K_{1,\alpha,N_K-\alpha} \\ Y^1_{1,\alpha,N_1-\alpha} | Y^2_{1,\alpha,N_2-\alpha} | \cdots | Y^K_{1,\alpha,N_K-\alpha} \\ Y^1_{1+\alpha,\alpha,N_1} | Y^2_{1+\alpha,\alpha,N_2} | \cdots | Y^K_{1+\alpha,\alpha,N_K} \end{bmatrix} = \\ \begin{bmatrix} \mathcal{R}_{11} & 0 & 0 & 0 \\ \mathcal{R}_{21} & \mathcal{R}_{22} & 0 & 0 \\ \mathcal{R}_{31} & \mathcal{R}_{32} & \mathcal{R}_{33} & 0 \\ \mathcal{R}_{41} & \mathcal{R}_{42} & \mathcal{R}_{43} & \mathcal{R}_{44} \end{bmatrix} \mathcal{Q}^{\mathcal{T}} \end{aligned} \quad (12)$$

Using a logic analogous to the case of single experiment, we consider the following SVD

$$\begin{bmatrix} \mathcal{R}_{42} \ \mathcal{R}_{43} \end{bmatrix} = \mathcal{U}_n \mathcal{S}_n \mathcal{V}_n^T + \mathcal{U}_n^\perp \mathcal{S}_2 {\mathcal{V}_n^\perp}^T \quad (13)$$

where the matrix $\mathcal{S}_n$ contains the principals singular values (further than threshold). Then the matrix Γ_α can be estimated as follows

$$\Gamma_\alpha = \mathcal{U}_n \mathcal{S}_n^{1/2} \quad (14)$$

The matrices C, A and B can be estimated by using Equations (7) and (8).

The initial conditions (x_0^i) where $i = 1, 2, \cdots, K$ can be estimated for a data set $\left\{u_t^i, y_t^i\right\}$ from the following equation

$$\begin{bmatrix} y_1^i \\ y_2^i \\ \vdots \\ y_\alpha^i \end{bmatrix} = \Gamma_\alpha x_0^i + \Phi_\alpha \begin{bmatrix} u_1^i \\ u_2^i \\ \vdots \\ u_\alpha^i \end{bmatrix} \quad (15)$$

As Γ_α is a full rank matrix, the above equation provides an estimation of x_0^i by using the pseudo-inverse of Γ_α.

IV. output-error identification

As it is well known that the model obtained by the above approach is not optimal, because the input signals are short and as well the effects of noise on the observed output signals. Therefore, the optimization of the obtained model is a necessary step to obtain a reliable model. The model's parameters obtained with the PO-MOESP technique will just be used as an initial guess of the parameters to be optimized.

If we consider that, we have K experiments, then the vector of parameters of the linear model is the following

$$\theta = \begin{bmatrix} vec(A) \\ vec(B) \\ vec(C) \\ x_0^1 \\ x_0^2 \\ \vdots \\ x_0^K \end{bmatrix} \tag{16}$$

The optimization problem which consider all the data sets simultaneously can be formulated as follow

$$J_K(\theta) = \frac{1}{K}\sum_{j=1}^{K}\frac{1}{Nj}\sum_{k=1}^{N_j}\left\|y_k^j - \hat{y}_k^j(\theta)\right\|_2^2 = \frac{1}{K}E_K(\theta)^T E_K(\theta) \tag{17}$$

where

$$E_K(\theta) = \left[E_{N_1}^1(\theta)^T \; E_{N_2}^2(\theta)^T \cdots E_{N_K}^K(\theta)^T\right]^T \tag{18}$$

and

$$E_{N_i}^i(\theta) = \frac{1}{\sqrt{N_i}}\left[e^i(1)^T \; e^i(2)^T \cdots e^i(N_i)^T\right]^T \tag{19}$$

is the error vector in which $e^i(k) = y_k^i - \hat{y}_k^i(\theta)$.
Assume that we have the sets $\{u_t^i, y_t^i : t = 1, 2, \cdots, N_i \; and \; i = 1, 2, \cdots, K\}$. The estimated output $\hat{y}_t^j(\hat{\theta})$ of the data set number j is given by the following model

$$\begin{aligned} \hat{x}_t^j &= A(\hat{\theta})\hat{x}_{t-1}^j + B(\hat{\theta})u_t^j \\ \hat{y}_t^j(\hat{\theta}) &= C(\hat{\theta})\hat{x}_t^j \end{aligned} \tag{20}$$

The minimization of (17) can be calculated by using the recursive gradient search method as follows

$$\theta^{l+1} = \theta^l - (\psi_K^T(\theta^l)\psi_K(\theta^l) + \lambda^l I)^{-1}\psi_K^T(\theta^l)E_K(\theta^l) \tag{21}$$

where

$$\psi_K(\theta) = \begin{bmatrix} \psi_{N_1}^1(\theta) \\ \psi_{N_2}^2(\theta) \\ \vdots \\ \psi_{N_K}^K(\theta) \end{bmatrix} \tag{22}$$

and

$$\psi_{N_i}^i(\theta) \triangleq \frac{\partial E_{N_i}^i(\theta)}{\partial \theta^T} \tag{23}$$

is the jacobian of the error vector $E_{N_i}^i$.

A. Local parameterization

In point of fact, it is well known that, the state-space representation of linear systems is not unique. As a consequence the minimization of $J_K(\theta)$ does not have a unique solution. Indeed, the optimal solution can be made unique by choosing a suitable parameterization. Ribarits *et al* [5] have proposed for the linear systems, in the case of single experiment and without considering the initial condition, a local parameterization of the state space model, in which the directions that do not change the cost function of output error are identified and projected out at each iteration, for that only the active parameters are updated. Analogously to previous method, we will develop a local parameterization of the linear systems in the case of multi-experiments case.

Let $T \in \mathbb{R}^{n\times n}$ be a nonsingular matrix. Consider that the state x_t^i of the data set number i is subjected to the following transformation

$$z_t^i = (T)^{-1}\, x_t^i \;,\; i = 1, 2, \cdots, K \tag{24}$$

Then the novel system's vector of parameters is $\bar{\theta}$ is given as follows

$$\begin{aligned} \bar{A} &= \left(T^i\right)^{-1} AT^i \\ \bar{B} &= \left(T^i\right)^{-1} B \\ \bar{C} &= CT^i \\ z_0^i &= \left(T^i\right)^{-1} x_0^i \end{aligned} \tag{25}$$

From Eq. 25, we obtain that the subset of equivalent models is parameterized by the transformation matrix T. This subset defines a manifold. In order to identify the tangent plane of the manifold, we linearize the relation (25) around the identity matrix $T = I_n$. Considering a small perturbation $T = I_n + \Delta T$, by using the approximation $(I_n + \Delta T)^{-1} \simeq I_n - \Delta T$ and neglecting all second-order terms, we obtain

$$\begin{aligned} \bar{A} &= A - \Delta TA + A\Delta T \\ \bar{B} &= B - \Delta TB \\ \bar{C} &= C + C\Delta T \\ z_0^i &= x_0^i - \Delta T x_0^i \end{aligned} \tag{26}$$

If we consider the following model's vectors of parameters

$$\theta = \begin{bmatrix} vec(A) \\ vec(B) \\ vec(C) \\ x_0^1 \\ x_0^2 \\ \vdots \\ x_0^K \end{bmatrix} \quad and \quad \bar{\theta} = \begin{bmatrix} vec(\bar{A}) \\ vec(\bar{B}) \\ vec(\bar{C}) \\ z_0^1 \\ z_0^2 \\ \vdots \\ z_0^K \end{bmatrix} \tag{27}$$

Then we obtain

$$\bar{\theta} = \theta + \begin{bmatrix} vec\left(-\Delta TA + A\Delta T\right) \\ vec\left(-\Delta TB\right) \\ vec\left(C\Delta T\right) \\ vec\left(-\Delta T x_0^1\right) \\ vec\left(-\Delta T x_0^2\right) \\ \vdots \\ vec\left(-\Delta T x_0^K\right) \end{bmatrix} \tag{28}$$

Using the property $vec(ABC) = \left(C^T \otimes A\right) vec(B)$, we obtain the relation between θ and $\bar{\theta}$

$$\bar{\theta} = \theta + M_\theta vec(\Delta T) \tag{29}$$

where

$$M_\theta = \begin{bmatrix} -A^T \otimes I_n + I_n \otimes A \\ -B^T \otimes I_n \\ I_n \otimes C \\ -{x_0^1}^T \otimes I_n \\ \vdots \\ -{x_0^K}^T \otimes I_n \end{bmatrix} \tag{30}$$

The equation (29) shows that the tangent space of the manifold of all systems similar to $(A, B, C, x_0^1, \cdots, x_0^K)$ is equal to the column space of the matrix M_θ (30). Since the left null space of the matrix M_θ is orthogonal complement to the column space, the directions in which the value of the cost function of output error changes are those related to the left null space of M_θ.
Note that, the left null space of M_θ can be efficiently obtained by singular Value Decomposition (SVD). Let the SVD of M_θ be given by

$$M_\theta = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ \left(V_1^T\right)^\perp \end{bmatrix} \tag{31}$$

then the null space of $M(\theta)$ is U_2. In order to exploit this property, the update rule should be modified such that we project out the directions in which the cost function does not change. The new update rule becomes

$$\theta^i = \theta^{i-1} - U_2 \left(U_2^T \psi_K^T \psi_K U_2 + \lambda^i I\right)^{-1} U_2^T \psi_K^T E_K \tag{32}$$

where U_2 and ψ_K depend on θ^{i-1}. Note that, since U_2 depends on the past parameter θ^{i-1} the SVD (31) must be computed at each iteration.

B. Computing the iterative parameter update

In order to compute the update rule (32), the following quantities $E_K(\theta)$ and $\psi_K(\theta)$ must be computed. For that, we simulate the systems (20) that corresponds to θ^{k-1}. Note that this simulation brings out the state $\hat{x}_t^i$ and $\hat{y}_t^i$ for $i = 1, 2, \cdots, K$. In order to simulate $\psi_K(\theta^{i-1})$, we should compute the derivative of $\left\{\hat{y}_t^i : i = 1, 2, \cdots, K\right\}$ with respect to θ^{k-1}. Let us define

$$\zeta_t^{i,j} = \frac{\partial \hat{x}_t^i}{\partial \theta_j} \tag{33}$$

where θ_j is the j^{th} element of the vector θ. The computation of $\frac{\partial \hat{y}_t^i}{\partial \theta} = \begin{bmatrix} \frac{\partial \hat{y}_t^i}{\partial \theta_1} & \cdots & \frac{\partial \hat{y}_t^i}{\partial \theta_q} \end{bmatrix}$, where q is the number of parameters in θ, can be made using the following model

$$\begin{aligned} \zeta_t^{i,j} &= A\zeta_{t-1}^{i,j} + \frac{\partial A}{\partial \theta_j}\hat{x}_{t-1}^i + \frac{\partial B}{\partial \theta_j}u_t \\ \frac{\partial \hat{y}_t^i}{\partial \theta_j} &= C\zeta_t^{i,j} + \frac{\partial C}{\partial \theta_j}\hat{x}_t^i \end{aligned} \tag{34}$$

Note that, as the initial condition is included into the vector of parameters then $\zeta_0^{i,j} = \frac{\partial \hat{x}_0^i}{\partial \theta_j} \neq \mathbf{0}$.

V. ILLUSTRATIVE EXAMPLES

In this section, we compare the results of the proposed method with the actual methods for identifying linear systems. First, we compare the results obtained by a classic subspace method (N4SID) [4]. Second, we study the effects of the number of data sets on the obtained results. For the two cases, we consider the linear system (1), which corresponds to the following matrices

$$\begin{aligned} A &= \begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.8 & -0.4 & 0 \\ 0 & 0 & 0 & 0.8 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.7 \end{bmatrix} \\ B &= \begin{bmatrix} -0.4 & 1.2 \\ -1.7 & -0.1 \\ 0.2 & 0.3 \\ 0.3 & 0.2 \\ -1.1 & -0.2 \\ 1.2 & 0.7 \end{bmatrix} \\ C &= \begin{bmatrix} -0.6 & -0.1 & 1.1 & -0.1 & 0.3 & 0.7 \\ 2.2 & 0.1 & 0.1 & -0.8 & -1.3 & 1.6 \end{bmatrix} \end{aligned} \tag{35}$$

The input signals are two dimensional uniform white noise with zero mean. The measurement noise is a Gaussian white noise.

A. Comparison with N4SID method

We have simulated the system 20 times with random initial conditions x_0^i, and the length of each of them is equal to 40 samples. In order to compare the performance of the proposed method and the actual methods, as it is well known that the eigenvalues of the matrix A are invariant, we calculate the eigenvalues of the obtained matrix $\hat{A}$, and we compare them with those of A. The results obtained by N4SID from MATLAB system identification toolbox [1] for each data set are given in Figure 1, the results obtained by considering all the data sets without optimization are given in Figure 2, and the optimized results in Figure 3.
From Figures 1, 2 and 3, we conclude that

1) The N4SID method is not able to identify the linear model of system by treating the data sets separately.
2) The extension of PO-MOESP method gives a good initial estimation of the linear model, but it is not accurate.
3) Using the proposed optimization algorithm improves the accuracy of the estimated linear model.

B. The effects of the number of data sets

We consider a data sets of small length, which is equals to 30 samples. Note that, the method N4SID does not work in this case and gives an error message. This because the data length is small. assume that, we have simulated the linear system 3, 5, 10, and 15 times with random initial conditions, and the measurement noise is scaled such that $SNR = 15\ dB$. The eigenvalues of the obtained matrix $\hat{A}$ by considering all the data sets and minimizing the output error function are given in Figure 4, 5, 6, and 7.
It is clear that, the accuracy of the estimated model increases with increasing the number of data sets.

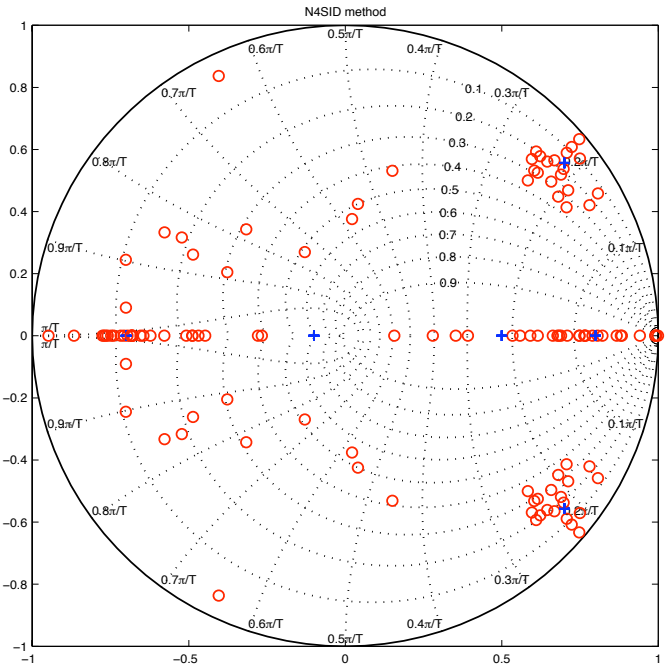

Fig. 1. The eigenvalues of matrix A (plus mark) and those of $\hat{A}$ obtained by N4SID method for each data set (circle mark) are superimposed.

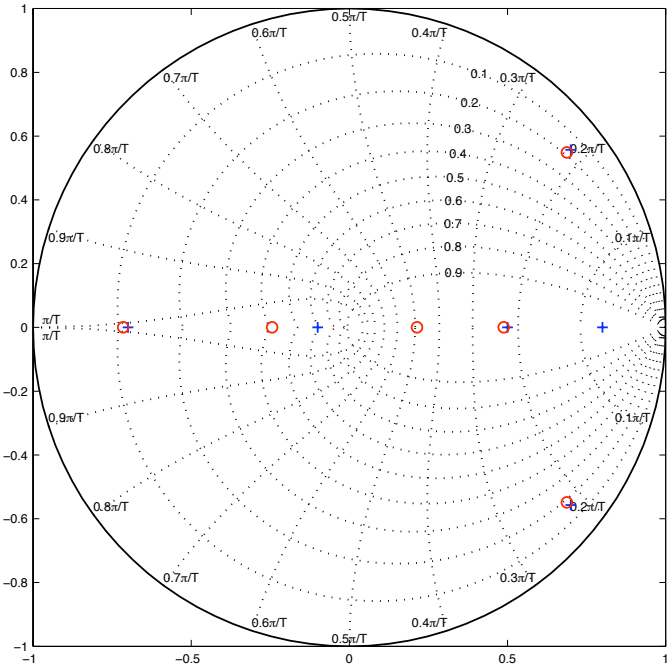

Fig. 2. The eigenvalues of matrix A (plus mark) and those of $\hat{A}$ obtained by the extension of PO-MOESP method to deal with multi-experiments (circle mark) are superimposed.

VI. Conclusions

In this paper, we have proposed a method to deal with the identification of linear multivariable systems in the case of multi-experiments. The results have pointed out that when the classical methods of subspace methods fail, the proposed method found the model of the linear system accurately. In the literature of system identification theory, the problem of dealing with multi-experiments is infrequently treated. However, we believe that this problem should be given more attention because it is the case in practice.

References

[1] L. Ljung, *system Identification Toolbox User's Guide*, MathWorks, August 1995.

[2] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall Informations and Systems Sciences, Jan 1999.

[3] M. Moonen, B. Moor, L. Vandenberghe, and J. Vandewalle, "On- and off-line identification of linear state-space models," *Int. J. Control*, vol. 49, no. 1, pp. 219–232, July 1989.

[4] P. V. Overschee and B. D. Moor, "N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems," *Automatica*, vol. 30, no. 01, pp. 75–93, Jan 1994.

[5] T. Ribarits, M. Deistler, and T. McKelvey, "An analysis of the parametrization by data driven local coordinates for multivariable linear systems," *Automatica*, vol. 40, pp. 789–803, 2004.

[6] M. Verhaegen, "Identification of the deterministic part of mimo state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.

[7] M. Viberg, "Subspace-based methods for the identification of linear time-invariant systems," *Automatica*, vol. 31, no. 12, pp. 1835–1851, 1995.

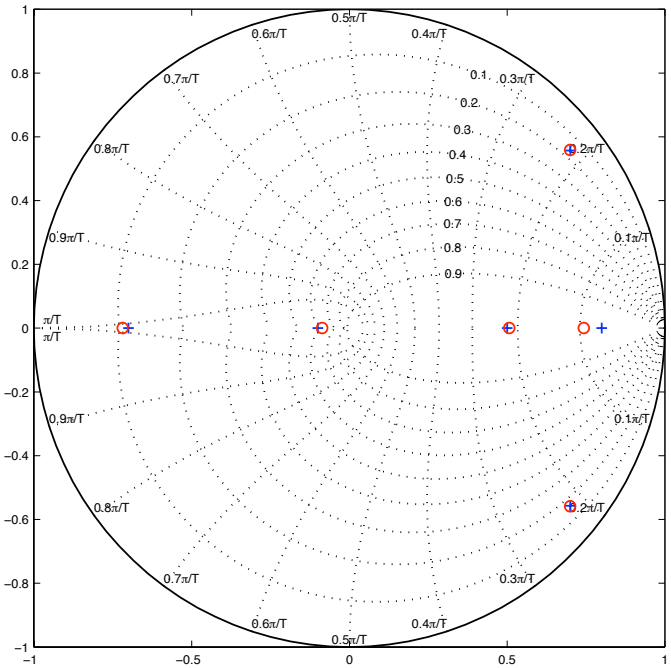

Fig. 3. The eigenvalues of matrix A (plus mark) and those of $\hat{A}$ obtained by the minimization of output error function (circle mark) are superimposed.

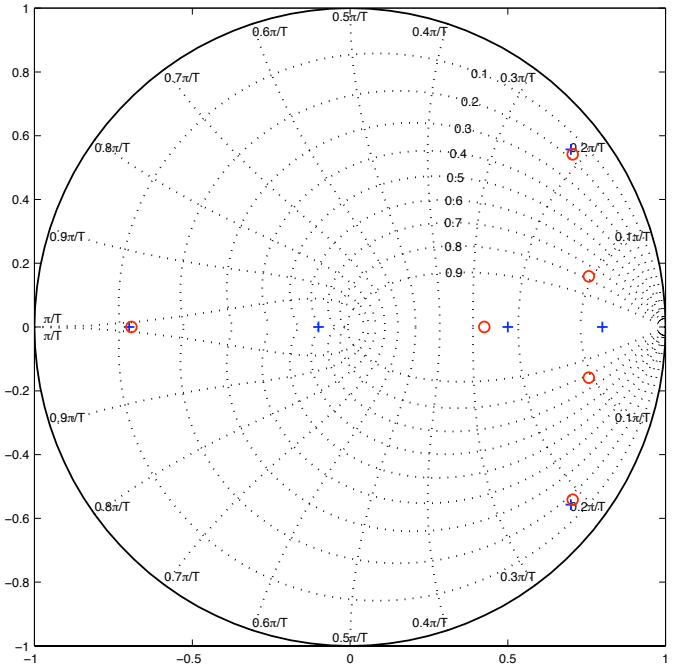

Fig. 4. The eigenvalues of matrix A (plus mark) and those of $\hat{A}$ obtained by the minimization of output error function (circle mark), the number of data sets is equal to 3.

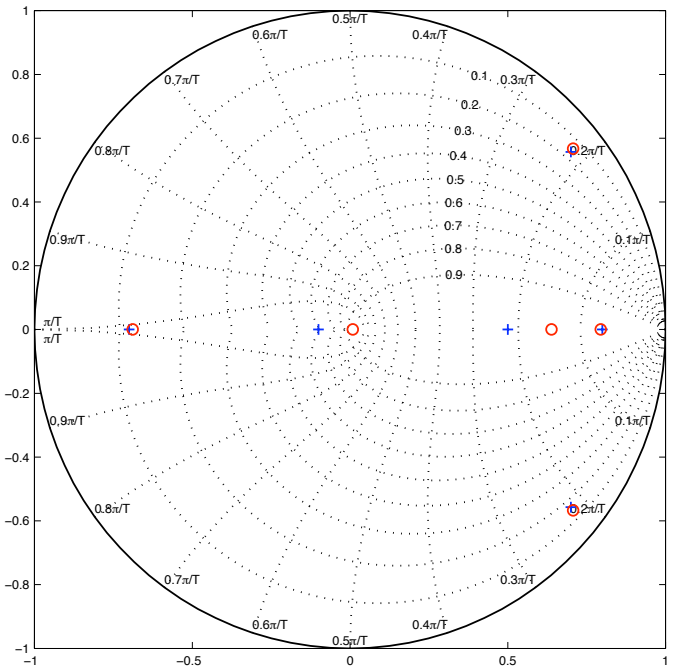

Fig. 5. The eigenvalues of matrix A (plus mark) and those of $\hat{A}$ obtained by the minimization of output error function (circle mark), the number of data sets is equal to 5.

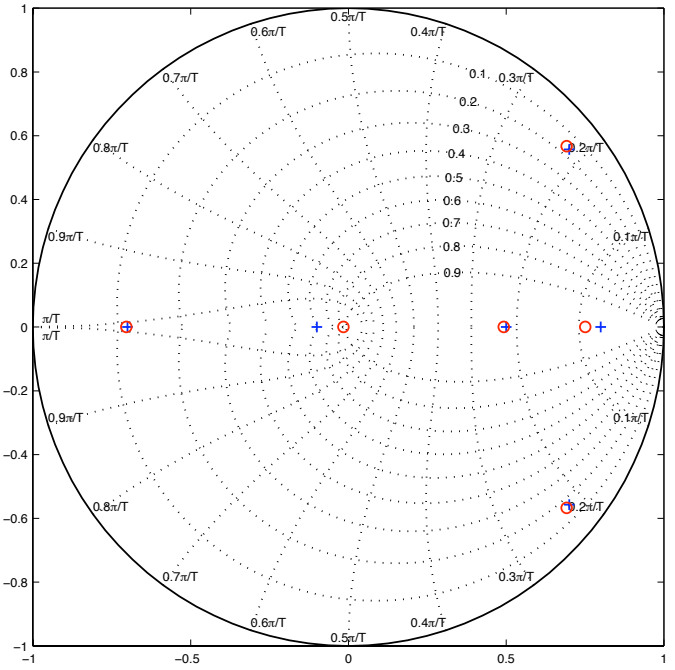

Fig. 6. The eigenvalues of matrix A (plus mark) and those of $\hat{A}$ obtained by the minimization of output error function (circle mark), the number of data sets is equal to 10.

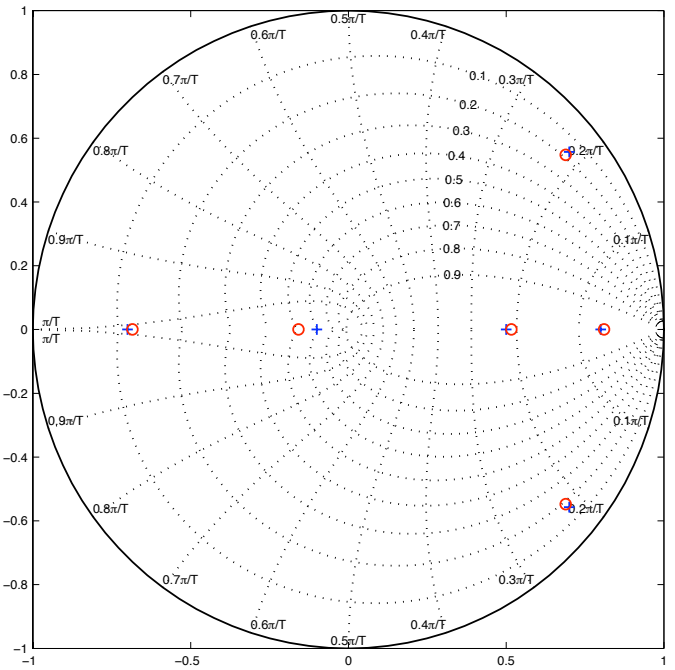

Fig. 7. The eigenvalues of matrix A (plus mark) and those of $\hat{A}$ obtained by the minimization of output error function (circle mark), the number of data sets is equal to 15.