Proceedings of the 37th IEEE
Conference on Decision & Control
Tampa, Florida USA • December 1998

A Comparison of Some Subspace Identification Methods

Tohru Katayama†, Shogo Omori† and Giorgio Picci‡

† Department of Applied Mathematics and Physics, Kyoto University
Kyoto 606-01, Japan; e-mail: katayama@kuamp.kyoto-u.ac.jp
‡ Department of Electronics and Informatics, University of Padova
35131 Padova, Italy; e-mail: picci@dei.unipd.it

Abstract

Recently, we have derived stochastic realization methods for a system with exogenous inputs [4, 1] and the relevance of stochastic realization to subspace identification of state-space systems has shown in [4]. In this paper, we briefly review the basis of stochastic subspace identification algorithm [1, 2] and present some simulations results to compare the performance and computational loads of the realization based algorithms, the CLS algorithm [3], and the basic 4SID [7].

Stochastic Realization with Exogenous Inputs

In this section, we summarize the stochastic realization method based on [1, 2]. Consider a discrete-time stochastic linear system with the $m \times 1$ input vector $u(t)$ and the $p \times 1$ output vector $y(t)$. It is assumed that $\{u(t),\ y(t),\ t = 0, \pm 1, \cdots\}$ are jointly wide sense stationary processes with zero mean and finite covariance matrices.

Let t be the present time and k a positive integer. We then define the stacked vectors of past and future inputs as

$$u_-(t) := \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \\ \end{bmatrix}, \quad u_+(t) := \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+k-1) \end{bmatrix}$$

and $y_-(t)$ and $y_+(t)$, the stacked vectors of past and future of outputs, are defined similarly. For notational simplicity, we also define the *past* and *future* as

$$p(t) := \begin{bmatrix} u_-(t) \\ y_-(t) \end{bmatrix}, \quad f(t) := y_+(t)$$

Theorem 1 Suppose that $p(t) \cap u_+(t) = 0$. Then the optimal LS predictor $\hat{f}(t)$, of the future output vector $f(t)$, based on the past input-output data $p(t)$ and future inputs $u_+(t)$, is given by the orthogonal projection

$$\hat{\mathbf{f}}(t) = f(t)|p(t) \vee u_+(t) = \Pi p(t) + \Phi u_+(t) \qquad (1)$$

where Π, Φ are given by

$$[\Pi \ \ \Phi] = [\Sigma_{fp} \ \ \Sigma_{fu}] \begin{bmatrix} \Sigma_{pp} & \Sigma_{pu} \\ \Sigma_{up} & \Sigma_{uu} \end{bmatrix}^{-1} \qquad (2)$$

It can be also shown that the operators Π and Φ satisfy the discrete Wiener-Hopf type equations

$$\Pi \Sigma_{pp|yu} = \Sigma_{fp|u}, \quad \Phi \Sigma_{uu|p} = \Sigma_{fu|p} \qquad (3)$$

where $\Sigma_{pp|u}$, $\Sigma_{uu|p}$ are the conditional covariance operators of the past vector $p(t)$ given $u_+(t)$ and of the future input $u_+(t)$ given the past $p(t)$, and are defined by

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

where $a|c^\perp := a - (a|c)$. □

Let $\{y(t), u(t)\}$ be the jointly stationary regular full rank process. Suppose that there is no feedback from y to u. Then it can be shown that Φ is block lower-triangular, so that it is a causal operator.

Let $\text{rank}\Sigma_{fp|u} = n$. Consider the Cholesky factorizations $\Sigma_{pp|u} = L_p L_p^T$ and $\Sigma_{ff|u} = L_f L_f^T$. Define $\varepsilon_+(t) := L_f^{-1}(f|u_+{}^\perp)(t)$, $\varepsilon_-(t) := L_p^{-1}(p|u_+{}^\perp)(t)$. It then follows that $E\{\varepsilon_+(t)\varepsilon_-^T(t)\} = L_f^{-1}\Sigma_{fp|u}L_p^{-T}$. Suppose that the SVD of the normalized block Hankel matrix $L_f^{-1}\Sigma_{fp|u}L_p^{-T}$ be given by

$$L_f^{-1}\Sigma_{fp|u}L_p^{-T} = U\tilde{\Sigma}V^T \qquad (4)$$

where $U^T U = I_n$, $V^T V = I_n$ and $\tilde{\Sigma} = \text{diag}(\tilde{\sigma}_1, \cdots, \tilde{\sigma}_n)$ is a diagonal matrix with nonzero singular values ($1 \geq \tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_n > 0$). We see that $\sigma_i's$ are the canonical correlation coefficients between the *conditional* random vectors $(f|u_+{}^\perp)(t)$ and $(p|u_+{}^\perp)(t)$.

For the SVD of (4), we define the extended observability and controllability matrices as

$$\mathcal{O} := L_f U \tilde{\Sigma}^{1/2}, \quad \mathcal{C} := \tilde{\Sigma}^{1/2} V^T L_p^T \qquad (5)$$

where $\text{rank}\mathcal{O} = \text{rank}\mathcal{C} = n$. Then the block Hankel matrix $\Sigma_{fp|u}$ has a decomposition $\Sigma_{fp|u} = \mathcal{O}\mathcal{C}$. Since $\Pi = \Sigma_{fp|u}\Sigma_{pp|u}^{-1}$, the oblique projection is expressed as

$$\Pi p(t) = \mathcal{O}x(t) \qquad (6)$$

where the state vector is now defined to be the $n \times 1$ vector

$$x(t) = C\Sigma_{pp|u}^{-1}p(t) = \tilde{\Sigma}^{1/2}V^T L_p^{-1}p(t) \quad (7)$$

Theorem 2 Suppose that there is no feedback from the output $y(t)$ to the input $u(t)$. We assume that $\text{rank}\Sigma_{fp|u} = n$. Then in terms of a state vector $x(t)$ of (7), we have a stochastic realization of the form

$$x(t+1) = Ax(t) + Bu(t) + Ke(t) \quad (8)$$

$$y(t) = Cx(t) + Du(t) + e(t) \quad (9)$$

Simulation Results

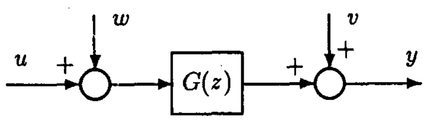

Fig. 1 The plant model

Some results of computer simulations are presented to show the performance of five subspace identification algorithms.

1) Basic 4SID is due to Verhaegen[7], where the Cholesky factorization is used to get L factor.
2) COV-a is the algorithm based on the stochastic realization, where the system matrices are estimated by using the estimate of *state vector* (see (7)).
3) COV-b is the algorithm based on the stochastic realization, where the system matrices are estimated by using $\mathcal{O}$ and Φ (see (3) and (5)).
4) CLS-a is the algorithm based on the constraint least-squares algorithm due to Peternell *et al.*[3] and using the estimation of *state vector* (see (7)).
5) CLS-b is the constraint least-squares algorithm and using $\mathcal{O}$ and Φ (see (3) and (5)).

We consider a 5th-order SISO system shown in Fig. 1 [9], where $u(t)$ is the input, and $w(t)$ and $v(t)$ are white noises with mean zeros. The transfer function is given by $G(z) = B(z)/A(z)$, where

$$B(z) = 0.0275z^{-4} + 0.551z^{-5}$$
$$A(z) = 1 - 2.3443z^{-1} + 3.081z^{-2} - 2.5274z^{-3} + 1.2415z^{-4} - 0.3686z^{-5}$$

The $G(z)$ has a zero at $z = -2$ and poles at $z = 0.9$, $0.8e^{\pm j}$, $0.8e^{\pm 1.2j}$.

Table 1: The number of flops for 50 simulation runs

	COV-a	CLS-a	Basic 4SID	LQ
Flops	1.58×10^8	2.23×10^9	5.26×10^7	6.42×10^9

In the present simulation studies, the input is chosen as $u(t) = U_0 \sum_{i=1}^{10} \sin(\omega_i t)$, where the frequencies ω_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 noise variances are chosen as $\sigma_w^2 = \sigma_v^2 = (0.05)^2$. It follows from the PE condition for $U_{0|2k-1}$ that $k \leq 10$. The performance is evaluated by the mean square error

$$I_N = \frac{1}{M}\sum_{l=1}^{M}\left(\sum_{j=1}^{10}[\theta_j - \hat{\theta}_j(l,N)]^2\right)$$

where $N = 200, 400, 1000, 2000$, and θ_j denotes the true parameter and $\hat{\theta}_j(l,N)$ is the estimate of θ_j at l-th run with the number of data N, and where M denotes the number of simulation runs.

Fig. 2 depicts the performance of five algorithms, where $k = 8$, $M = 100$. In this case, COV-a and COV-b show similar performance, but the performance of CLS-a and CLS-b is rather different. In order to analyze this fact, we have simulated CLS-a and CLS-b methods for several different k's, where the input is a sum of 15 sinusoids and $M = 50$, $N = 1000$. We see from Fig. 3 that both methods give similar performance for k greater than 10, but for the smaller k, CLS-a shows better performance. In Figs. 4 and 5, the pole estimates by COV-a and CLS-a methods are depicted for $k = 8$, $N = 1000$. We see that COV-a gives a rather scattered pole estimates, but CLS-a yields better pole estimates with a smaller variability.

Table 1 shows the number of flops of four algorithms, where LQ denotes the algorithm based on the LQ factorization of the Hankel matrix [5, 6, 7]. The number of flops includes all the computations for the whole simulations by each algorithm for $k = 8$, $M = 50$, $N = 1000$. It therefore follows that by using the Cholesky factorization [2], we get a great computational saving over the method based on LQ factorization. Also, it is rather surprising to find that CLS-a is ten times more expensive than COV-a.

References

[1] T. Katayama and G. Picci, "An Approach to Realization of Stochastic Systems with Exogenous Input", *Preprints of 11th IFAC Symposium on System Identification*, Kitakyushu, Japan, July 1997, pp. 1107-1112.

[2] T. Katayama and G. Picci, "Realization of Stochastic Systems with Exogenous Inputs and Subspace Identification Methods," 1998 (submitted).

[3] K. Peternell, W. Scherrer and M. Deistler, "Statistical Analysis of Novel Subspace Identification Methods," *Signal Processing*, vol. 52, no. 2, July 1996, pp. 161-177.

[4] G. Picci and T. Katayama, "Stochastic Realization with Exogenous Inputs and "Subspace Methods" Identification," *Signal Processing*, vol. 52, no. 2, July 1996, pp. 145-160.

[5] P. Van Overschee, P. and B. De Moor, "N4SID - Subspace Algorithms for the Identification of Combined Deterministic - Stochastic Systems," *Automatica*, vol. 30, no, 1, 1994, pp. 75-93.

[6] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems*, Kluwer Academic Publications, 1996.

[7] M. Verhaegen and P. Dewilde, "Subspace Model Identification (Parts 1 and 2)," *Int. J. Control*, vol. 56, 1992, pp. 1187-1210 & pp. 1211-1241.

[8] 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, 1994, pp. 61-74.

[9] M. Viberg, "Subspace-based Methods for the Identification of Linear Time-invariant Systems," *Automatica*, vol. 31, no. 12, 1995, pp. 1835-1851.

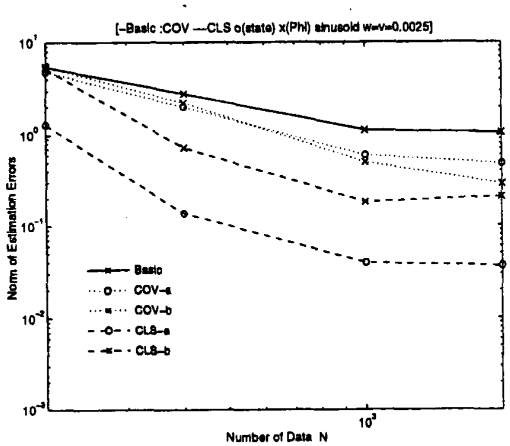

Fig. 2 The performance of five algorithms

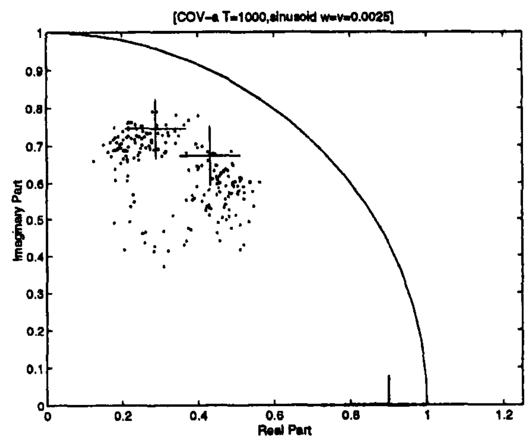

Fig. 4 The pole estimates by COV-a

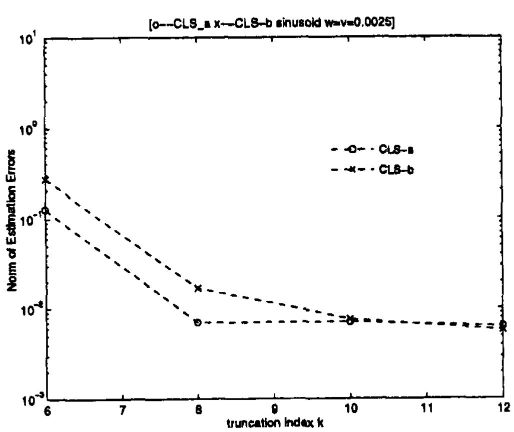

Fig. 3 The performance vs. number of rows

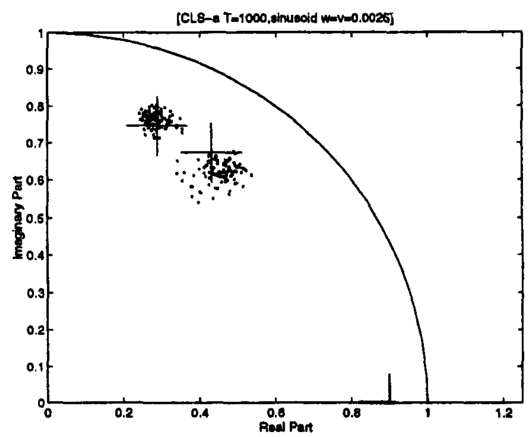

Fig. 5 The pole estimates by CLS-a