2010 American Control Conference
Marriott Waterfront, Baltimore, MD, USA
June 30-July 02, 2010

A Comparison of Least Squares Algorithms for Estimating Markov Parameters

Matthew S. Fledderjohn[1], Matthew S. Holzel[2], Harish J. Palanthandalam-Madapusi[3], Robert J. Fuentes[4], and Dennis S. Bernstein[5]

***Abstract*— The purpose of this work is to compare model structures and identification algorithms for estimating Markov parameters in the presence of uncorrelated and correlated input, process, and output noise. We consider several least-squares variants with ARX and μ-Markov model structures, which are compared with white noise identification signals.**

I. INTRODUCTION

Linear system identification techniques have been extensively developed, and numerous techniques exist for identifying systems in state space and ARMAX form [1–3]. The present paper is motivated by control design techniques that require estimates of the Markov parameters [4–8]. In addition, Markov parameters are related to fundamental properties of linear systems, such as invariant zeros [9]. These parameters can be estimated by a wide variety of methods, such as ARMAX fits, subspace methods, and, by using the inverse FFT, frequency domain techniques.

The purpose of the present paper is to provide a detailed comparison of the accuracy, consistency, and biasedness of least-squares techniques for Markov parameters. Our goal is to present numerical evidence to compare the accuracy of these methods when used with various model structures and noise configurations. It is our hope that these comparisons will be useful to practitioners in the field, while providing guidance for future theoretical developments.

We consider include several variants of least-squares identification applied to ARMAX models as well as a subspace method. For the least-squares algorithms we consider an ARX and μ-Markov structure. The latter structure, which is derived from the ARX structure, explicitly displays μ Markov parameters [7, 10]. Although the identification techniques are applicable to MIMO systems, we confine our attention to SISO plants. Finally, since our goal is to estimate Markov parameters and not all system parameters, we do not require that the true system be contained in the model class used for identification.

In these numerical studies, we explore the effects of the noise configuration, signal-to-noise ratio (SNR), data length, and assumed plant order. We do not consider model prediction accuracy since we are concerned only with the problem of estimating Markov parameters.

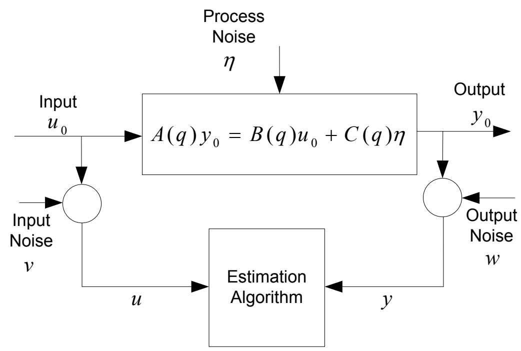


Fig. 1. Plant model and identification architecture, where noise can enter in three separate locations.

To assess the accuracy of these algorithms, we consider a setup where the input, measurement of the input, and output are corrupted by white noise signals. This setting allows us to examine whether the estimates of Markov parameters are consistent, that is, convergent to the true parameter values in the limit of infinite data. The algorithms we consider include batch least squares, total least squares, structured total least squares, multistage least squares, quadratically constrained least squares, and N4SID.

II. IDENTIFICATION CONSIDERATIONS

Among the factors that can affect the accuracy of system identification, we limit our attention to noise properties, model structure, model order, and the amount of data. We consider zero-mean Gaussian white noise input signals with correlated and uncorrelated zero-mean Gaussian white noise processes superimposed on the input, measurement of the input, and output with specified SNRs. The SNR is taken to be the ratio of the RMS value of the true signal to the RMS value of the noise superimposed on that signal. We also consider ARX, μ-Markov, and state space models. Since the algorithms require an estimate of the order of the plant generating the data, we under- and over-estimate the order of the plant. Lastly, we vary the amount of data to determine whether the parameter estimates appear consistent.

III. PROBLEM SETUP

We consider three noise signals affecting the true input u_0 and true output y_0, namely, input noise v, process noise η, and output noise w, as seen in Figure 1. The problem can be modeled with the ARMAX structure

$$A(\mathbf{q})y_0(k) = B(\mathbf{q})u_0(k) + C(\mathbf{q})\eta(k), \tag{1}$$

This work was partially sponsored by NASA under IRAC grant NNX08AB92A.

[1]M.S. Student, University of Michigan, Ann Arbor, MI.
[2]Ph.D. Student, University of Michigan, Ann Arbor, MI.
[3]Assistant Professor, Deptartment of Mechanical and Aerospace Engineering, Syracuse University, Syracuse, NY.
[4]Flight Controls Department, Raytheon Missle Systems, Tucson, AZ.
[5]Professor, Deptartment of Aerospace Engineering, University of Michigan, Ann Arbor, MI, dsbaero@umich.edu.

978-1-4244-7425-7/10/$26.00 ©2010 AACC

where $\mathbf{q}$ is the forward shift operator, $u(k) = u_0(k) + v(k)$, and $y(k) = y_0(k)+w(k)$. Except where noted, $B(\mathbf{q}) = C(\mathbf{q})$.

A. Uncorrelated v, η, and w Noise Signals

When noise signals are uncorrelated, they are generated independently and then scaled to a specified SNR. The realizations of v and η are scaled relative to u_0, while w is scaled relative to y_0.

B. Correlated v, η, and w Noise Signals

When noise signals are correlated, the same random sequence is used for both signals, which are then scaled to the specified SNR. We consider correlated v and w as well as correlated η and w. We do not consider correlated v and η because in this case $v = \eta$, and thus $u_0 + v = u_0 + \eta$. If $B(\mathbf{q}) = C(\mathbf{q})$, then this scenario is equivalent to having no input or process noise.

IV. Model Structures for Identification

The ARX model is given by

$$y(k) = \sum_{j=0}^{n_{\rm mod}} b_j u(k-j) - \sum_{j=1}^{n_{\rm mod}} a_j y(k-j), \qquad (2)$$

where the Markov parameters are obtained by impulsing (2).

The μ-Markov model is given by

$$y(k) = \sum_{j=0}^{\mu-1} H_j u(k-j) + \sum_{j=\mu}^{n_{\rm mod}+\mu-1} b'_j u(k-j) - \sum_{j=\mu}^{n_{\rm mod}+\mu-1} a'_j y(k-j), \qquad (3)$$

where $\mu \geq 1$ and $H_0, \ldots, H_{\mu-1}$ are the first μ Markov parameters.

Both models can be written as an over-determined equation set of the form

$$Ax = b + e, \qquad (4)$$

where e is a vector of residuals, x is a vector of model parameters, A is a regression matrix of measured inputs and outputs, and b is the predicted variable $y(k)$.

V. Identification Methods

All of the proceeding algorithms will be used in conjunction with both ARX and μ-Markov model structures, except for N4SID, which is inherently a state-space technique.

A. Batch Least Squares (BLS)

Batch least squares determines the model coefficients x that minimize the cost $\|Ax - b\|_2$. It can be shown that BLS with an ARX model structure is identical to the OKID algorithm for estimating Markov parameters [1].

B. Total Least Squares (TLS)

From (4), if $e = 0$, then there exists at least one solution x of $Ax = b$, that is, $\begin{bmatrix} A & b \end{bmatrix}$ has a nontrivial null space containing $\begin{bmatrix} x & -1 \end{bmatrix}^T$. However, when $e \neq 0$, then there does not exist x satisfying (4). Therefore, total least squares determines $\hat{A}$ and $\hat{b}$ which minimize $\left\| \begin{bmatrix} \hat{A} & \hat{b} \end{bmatrix} - \begin{bmatrix} A & b \end{bmatrix} \right\|_F$ such that there exists $\hat{x}$ satisfying $\hat{A}\hat{x} = \hat{b}$. From the Eckart-Young-Mirsky theorem [11], the minimizing $\begin{bmatrix} \hat{A} & \hat{b} \end{bmatrix}$ is obtained by zeroing the smallest singular value of $\begin{bmatrix} A & b \end{bmatrix}$. Note that since TLS perturbs the entries of A and b, it considers errors in both the observed input and output.

C. Structured Total Least Squares (STLS)

Similar to TLS, structured total least squares accounts for errors in both A and b, while constraining $\begin{bmatrix} \hat{A} & \hat{b} \end{bmatrix}$ to the same Toeplitz pattern as $\begin{bmatrix} A & b \end{bmatrix}$.

D. Multistage Least Squares (MLS)

The first step of multistage least squares consists of a high-order BLS to obtain the estimate $\hat{x}$. Next, the residual $r = b - A\hat{x}$ is used as a known input in a second least-squares computation. Specifically, we fit the low-order TIMO model

$$A(\mathbf{q})y(k) = \begin{bmatrix} B(\mathbf{q}) & C(\mathbf{q}) \end{bmatrix} \begin{bmatrix} u(k) \\ r(k) \end{bmatrix}. \qquad (5)$$

E. Quadratically Constrained Least Squares (QCLS)

Quadratically constrained least squares [12, 13] assumes knowledge of the noise statistics, and estimates the model parameters by solving a generalized-eigenvalue problem. QCLS is an extension of the Koopmans-Levin algorithm.

F. N4SID

The Matlab N4SID function is a subspace algorithm which estimates a state space model of the system [14].

VI. Performance Metric

We consider a metric based on the truncated Markov-parameter Toeplitz matrix $\mathscr{T}_\mu$ [15, p. 202] given by

$$\mathscr{T}_\mu \triangleq \begin{bmatrix} H_0 & \cdots & 0 \\ \vdots & \ddots & 0 \\ H_{\mu-1} & \cdots & H_0 \end{bmatrix}, \qquad (6)$$

where $H_0, \ldots, H_{\mu-1}$ are the first μ Markov parameters. The performance metric is the largest singular value of the difference between the estimated and actual truncated Markov block-Toeplitz matrices, that is, $\varepsilon_{\mathscr{T}_\mu} \triangleq \sigma_{\max}\big(\mathscr{T}_\mu - \hat{\mathscr{T}}_\mu\big)$.

VII. Numerical Experiments

We consider the discrete-time transfer function model

$$G(z) = \frac{(z-0.75)(z-0.85)(z^2-1.6z+0.6425)}{(z-0.8)(z^2+0.01)(z^2+0.04)(z^2+0.9025)}, \qquad (7)$$

which has high-frequency modes, lightly damped modes, and approximate pole/zero cancellation. With z replaced by the forward shift operator $\mathbf{q}$, the numerator and denominator of $G(z)$ correspond to $B(\mathbf{q})$ and $A(\mathbf{q})$, respectively. All simulations have zero initial conditions.

A. Effect of μ on the Metric $\varepsilon_{\mathscr{T}_\mu}$

Increasing the number of identified Markov parameters μ creates a larger truncated Markov block-Toeplitz matrix. To investigate the effect of μ on the metric $\varepsilon_{\mathscr{T}_\mu}$, we run each algorithm 100 times at each μ with different uncorrelated v and w realizations. The input realization u_0 remains the same for all simulations. For each μ, we compute the average the error $\bar{\varepsilon}_{\mathscr{T}_\mu} = 1/100 \sum_{j=1}^{100} \varepsilon_{\mathscr{T}_{\mu_j}}$, where j represents the simulations index. Figure 2 shows that the mean error $\bar{\varepsilon}_{\mathscr{T}_\mu}$ increases as μ increases for BLS, TLS, STLS, and N4SID. We thus consider $\mu = 10$ for subsequent comparison.

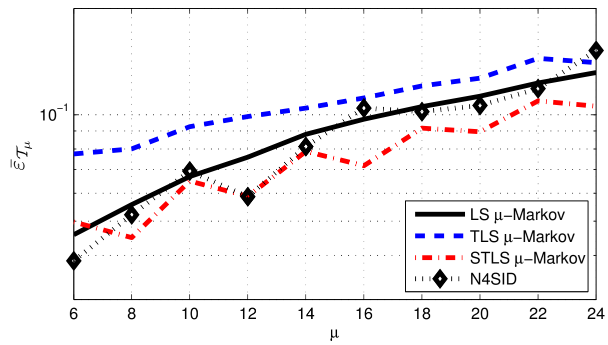
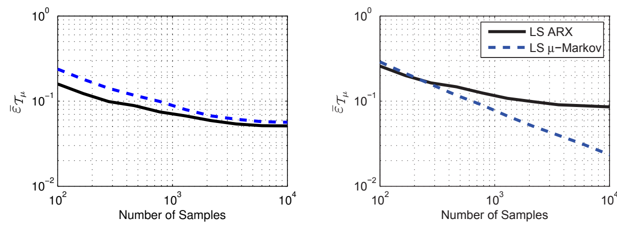


Fig. 2. Average error $\bar{\varepsilon}_{\mathcal{T}_\mu}$ for a range of μ for BLS, TLS, STLS, and N4SID with uncorrelated v and w, SNR = 10, $n_{\text{mod}} = 7$, and 5000 samples.



(a) v only (b) w only

Fig. 3. Comparison of ARX and μ-Markov on BLS with SNR = 10, $\mu = 10$, and $n_{\text{mod}} = 7$.

B. Comparison of ARX and μ-Markov Structures

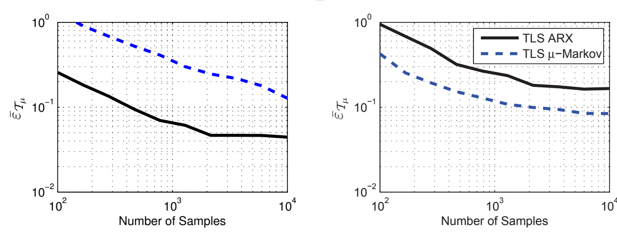
Figure 3 indicates that with v only, BLS with ARX and μ-Markov models have comparable errors, while with w only, the μ-Markov structure provides a smaller error.

Figure 4a indicates that with v only, TLS with a μ-Markov model is not as accurate as with an ARX model. Figure 4b indicates that with w only, TLS with a μ-Markov model is more accurate than with an ARX model.

Figure 5 indicates that STLS with μ-Markov and ARX models are similar up to 1000 samples, but, as the number of samples increases, STLS with a μ-Markov model is more accurate than with an ARX model.

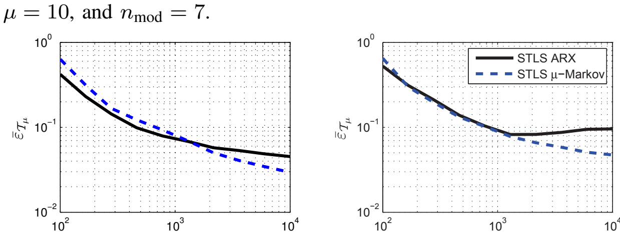
C. Comparison of BLS and MLS

Here we consider colored process noise with



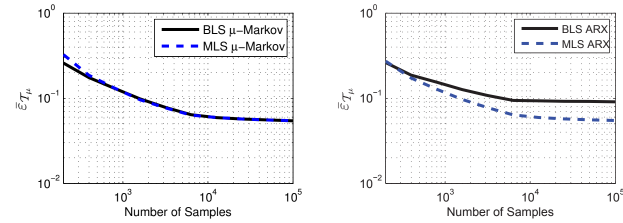
(a) v only (b) w only

Fig. 4. Comparison of ARX and μ-Markov on TLS with SNR = 10, $\mu = 10$, and $n_{\text{mod}} = 7$.



(a) v only (b) w only

Fig. 5. Comparison of ARX and μ-Markov on STLS with SNR = 10, $\mu = 10$, and $n_{\text{mod}} = 7$.



(a) μ Markov (b) ARX

Fig. 6. BLS and MLS with uncorrelated v and w, SNR = 10, $\mu = 10$, and $n_{\text{mod}} = 7$. In (a), μ-Markov models. In (b), ARX models.

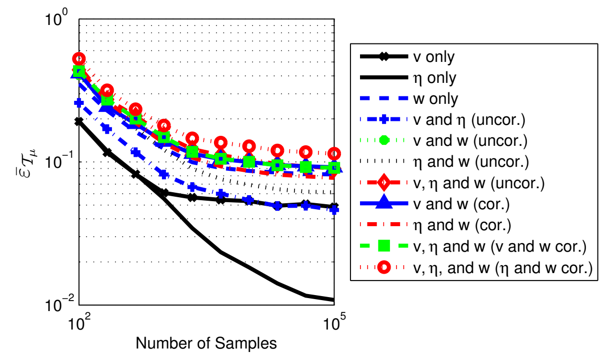


Fig. 7. Consistency of BLS ARX with SNR = 10, $\mu = 10$, $n_{\text{mod}} = 7$.

$$C(\mathbf{q}) = \mathbf{q}^5 - 0.82\mathbf{q}^4 + 0.31\mathbf{q}^3 - 0.40\mathbf{q}^2 + 0.40\mathbf{q} - 0.07. \quad (8)$$

Figure 6a suggests that MLS provides no improvement over BLS when used with the μ-Markov structure. However, Figure 6b shows that MLS can improve the accuracy of BLS when used with the ARX structure.

D. Consistency of BLS μ-Markov, BLS ARX, TLS, STLS, QCLS, and N4SID

We investigate the consistency properties of all algorithms by plotting the error $\varepsilon_{\mathcal{T}_\mu}$ as the number of samples increases. A constant downward slope of the error on a log-log plot suggests that the error approaches zero in the limit as the number of samples approaches infinity, and hence the Markov parameter estimates are consistent. If the error approaches a constant value on the graph, then we infer (based on this finite data) that the estimator is not consistent.

Figure 7 indicates that for all noise configurations, the BLS Markov parameter estimates from a 7^{th} order ARX model are not consistent.

Figure 8 indicates that with w only, BLS with an ARX model can provide consistent Markov parameter estimates when n_{mod} is large. This and other examples lead to the conjecture that with w only, η only, or w and η, BLS ARX provides consistent estimates if $\mu \leq n_{\text{mod}} + 1$.

Figure 9a suggests that for all combinations of η and w, BLS with a μ-Markov model yields consistent Markov parameter estimates. However, Figure 9b indicates that when v is also present, the estimates are not consistent.

Next, for QCLS we require knowledge of the autocorrelation of the noise $R \triangleq \mathbb{E}\left[\psi^T(k)\psi(k)\right]$ affecting u_0 and y_0 to within a positive scalar multiple. Here $\mathbb{E}$ is the expectation,

$$\psi(k) = \left[\ \zeta(k) \ \cdots \ \zeta(k-n) \ \ -v(k) \ \cdots \ -v(k-n)\ \right]^T,$$

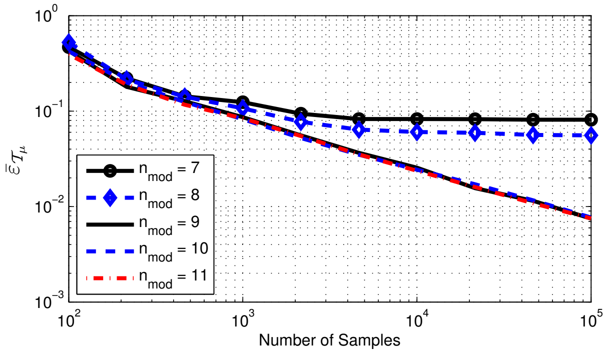
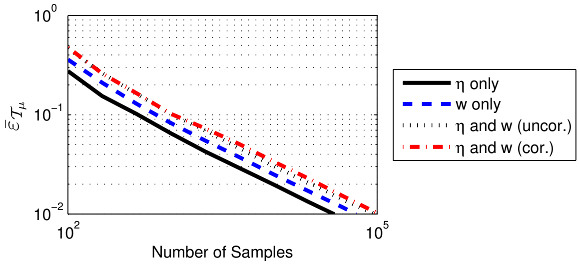
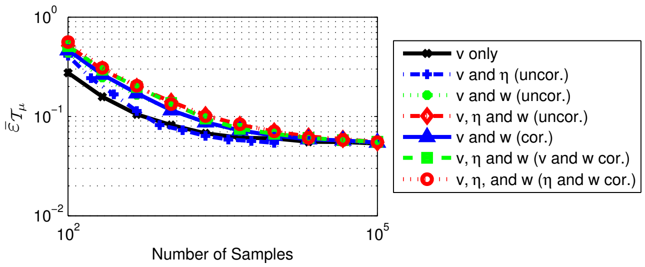


Fig. 8. Consistency of BLS ARX with w only, SNR = 10, $\mu = 10$.



(a) η and/or w only



(b) v present

Fig. 9. Consistency of BLS μ-Markov with SNR = 10, $\mu = 10$, $n_{\text{mod}} = 7$.

for the ARX model, $\zeta(k) = w(k) + A(\mathbf{q})^{-1}C(\mathbf{q})\eta(k)$,

$$\psi(k) = [\ \zeta(k) \quad \zeta(k-\mu) \quad \cdots \quad \zeta(k-\mu-n+1)$$
$$-v(k) \quad \cdots \quad -v(k-\mu-n+1)\ ]^T,$$

for the μ-Markov model, ζ affect y_0, and $-v$ affects u_0. Furthermore, since ζ and v are stationary, R is independent of k. We partition R as

$$R = \begin{bmatrix} R_{\zeta\zeta} & -R_{\zeta v} \\ -R_{\zeta v} & R_{vv} \end{bmatrix}. \tag{9}$$

The construction of R for all noise configurations is shown in Table I. The table also summarizes the consistency of BLS with μ-Markov and ARX models. Table I also shows that QCLS produces consistent estimates with $\hat{R} = \alpha R$. The alternate choice $\hat{R} = N_{LS}$ does not use noise properties, and produces the BLS result. When BLS is consistent, QCLS is consistent with $\hat{R} = \alpha R$ or $\hat{R} = N_{LS}$.

Figure 10 suggests that for all noise configurations, QCLS yields consistent Markov parameter estimates when R is known to within a constant positive multiple.

Figure 11 suggests that for all noise configurations, neither the STLS or TLS Markov parameter estimates are consistent.

Figure 12 suggests that N4SID is consistent with η and/or w but not consistent with v present.

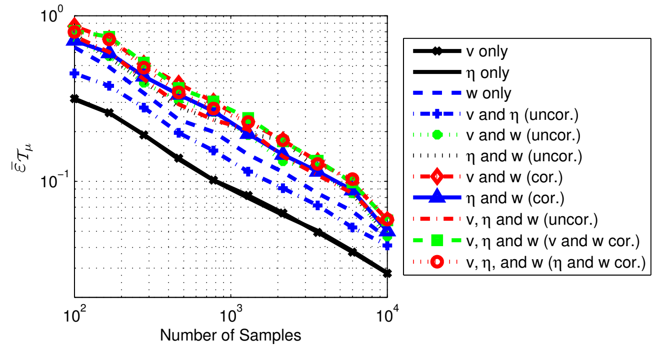
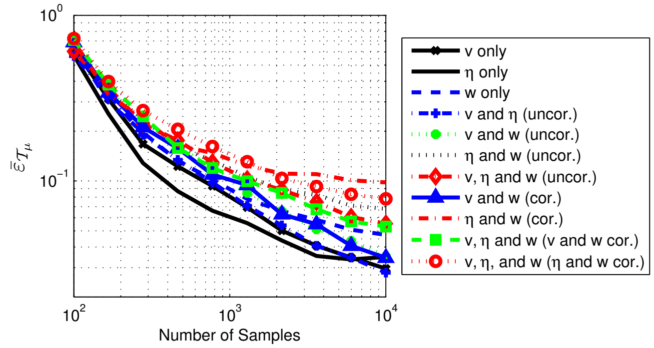
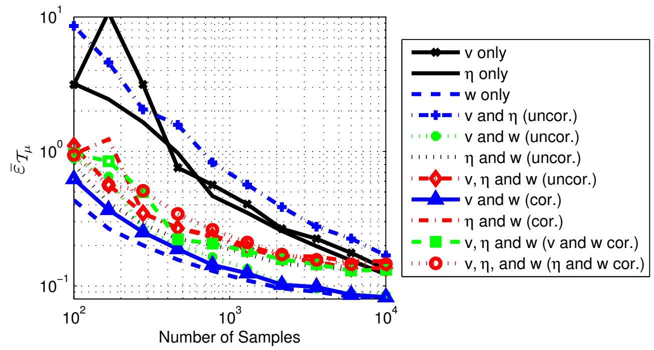


Fig. 10. QCLS consistency with $\hat{R} = \alpha R$, SNR = 10, $\mu = 10$, $n_{\text{mod}} = 7$.

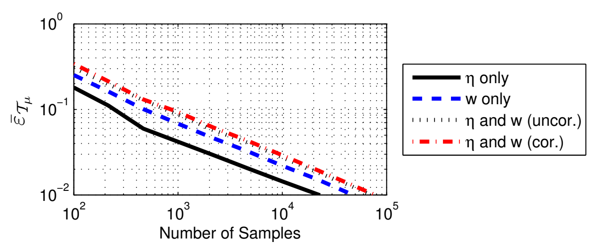


(a) STLS μ-Markov

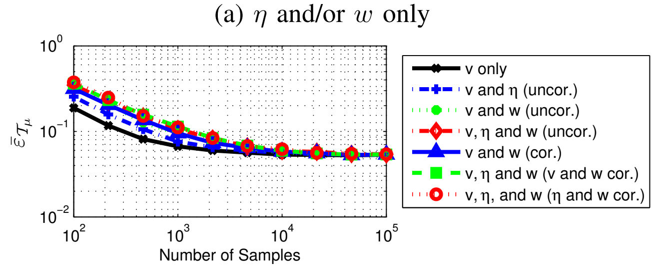


(b) TLS μ-Markov

Fig. 11. Consistency of STLS and TLS with a μ-Markov model with SNR = 10, $\mu = 10$, $n_{\text{mod}} = 7$.



(a) η and/or w only



(b) v present

Fig. 12. Consistency of N4SID with SNR = 10, $\mu = 10$, $n_{\text{mod}} = 7$.

Noise	R	BLS ARX Consistent	BLS μ-Markov Consistent	QCLS ARX Consistent	QCLS μ-Markov Consistent
v	$\begin{bmatrix} 0 & 0 \\ 0 & R_{vv} \end{bmatrix}$	Never	Never	$\hat{R} = \alpha R$	$\hat{R} = \alpha R$
η	$\begin{bmatrix} R_{G\eta G\eta} & 0 \\ 0 & 0 \end{bmatrix}$	$\mu \leq n_{\text{mod}} + 1$	Always	$\hat{R} = \alpha R$	$\hat{R} = \alpha R$ or $\hat{R} = N_{LS}$
w	$\begin{bmatrix} R_{ww} & 0 \\ 0 & 0 \end{bmatrix}$	$\mu \leq n_{\text{mod}} + 1$	Always	$\hat{R} = \alpha R$	$\hat{R} = \alpha R$ or $\hat{R} = N_{LS}$
v,η	$\begin{bmatrix} R_{G\eta G\eta} & -R_{vG\eta} \\ -R_{vG\eta} & R_{vv} \end{bmatrix}$	Never	Never	$\hat{R} = \alpha R$	$\hat{R} = \alpha R$
v,w	$\begin{bmatrix} R_{ww} & -R_{wv} \\ -R_{wv} & R_{vv} \end{bmatrix}$	Never	Never	$\hat{R} = \alpha R$	$\hat{R} = \alpha R$
η,w	$\begin{bmatrix} R_{\zeta\zeta} & 0 \\ 0 & 0 \end{bmatrix}$	$\mu \leq n_{\text{mod}} + 1$	Always	$\hat{R} = \alpha R$	$\hat{R} = \alpha R$ or $\hat{R} = N_{LS}$
v,η,w	$\begin{bmatrix} R_{\zeta\zeta} & -R_{\zeta v} \\ -R_{\zeta v} & R_{vv} \end{bmatrix}$	Never	Never	$\hat{R} = \alpha R$	$\hat{R} = \alpha R$

TABLE I

WITH v ABSENT, BLS ARX PROVIDES CONSISTENT ESTIMATES OF THE MARKOV PARAMETERS IF $\mu \leq n_{\text{MOD}} + 1$ AND KNOWLEDGE OF R IS NOT REQUIRED. WHEN R IS KNOWN UP TO A POSITIVE SCALAR MULTIPLE, QCLS PROVIDES CONSISTENT ESTIMATES OF THE MARKOV PARAMETERS.

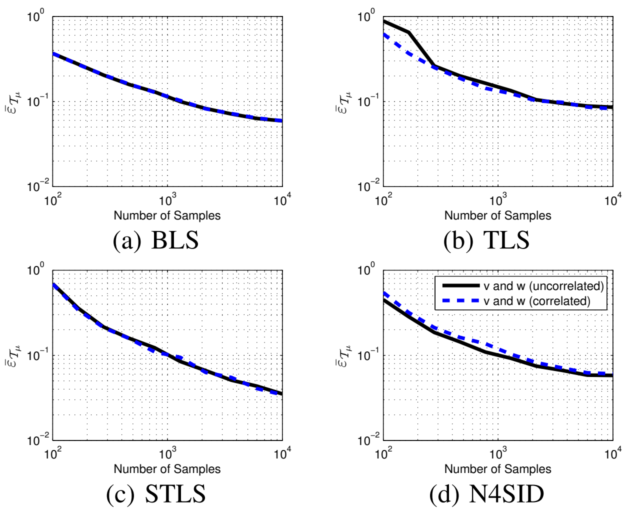


Fig. 13. Effect of correlation between v and w on BLS, TLS, STLS, and N4SID with a μ-Markov model. We use SNR = 10, $\mu = 10$, and $n_{\text{mod}} = 7$.

E. Accuracy versus Correlation

Figure 13 indicates that correlation between v and w has little effect on BLS, TLS, STLS, and N4SID.

Figure 14 shows that correlation between η and w decreases the accuracy of TLS, STLS, and N4SID, while BLS remains unaffected.

F. Analysis of Bias

In addition to consistency of these algorithms, we analyze bias for finite data sets. To test bias, each method is run with 50 different noise realizations. The closeness of the mean of the 50 estimates to the actual Markov parameter is used as a measure of bias, that is, $\varepsilon_{H_i} \triangleq H_i - \sum_{j=1}^{50} \hat{H}_{i_j}/50$, where $j = 1, \ldots, 50$ is each test with a different noise realization and $i = 0, \ldots, \mu - 1$ is the Markov parameter index.

With w only, Figure 15 suggests that BLS with a μ-Markov model is unbiased, along with QCLS. However, both TLS and STLS appear to be biased.

With v only, Figure 16, suggests that BLS, TLS and STLS are biased, although QCLS is not.

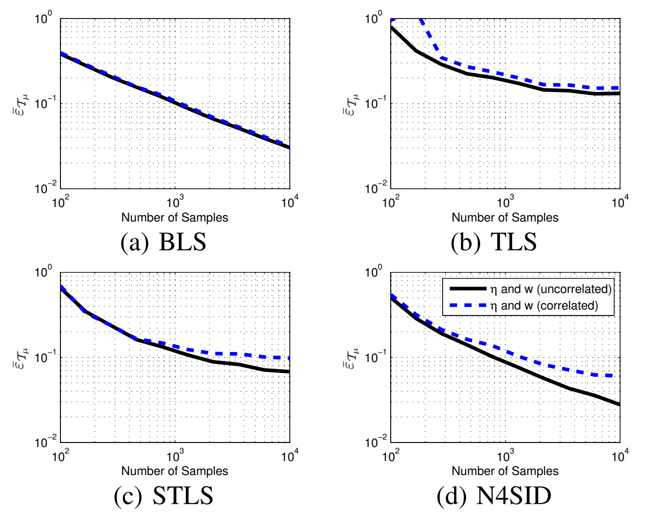


Fig. 14. The effects of correlation between η and w on BLS, TLS, STLS, and N4SID. We use SNR = 10, $\mu = 10$, and $n_{\text{mod}} = 7$.

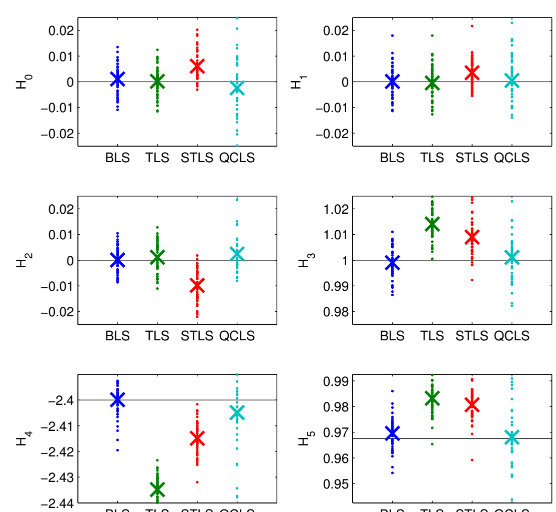


Fig. 15. Bias analysis of BLS, TLS, STLS, and QCLS with a μ-Markov model, w only, SNR = 10, $\mu = 10$, $n_{\text{mod}} = 7$, and 10,000 samples. The horizontal line is the actual Markov parameter, and each dot is the result of simulation with a different noise realization. For each algorithm, 50 simulations are run. 'x' denotes the average value of the simulations.

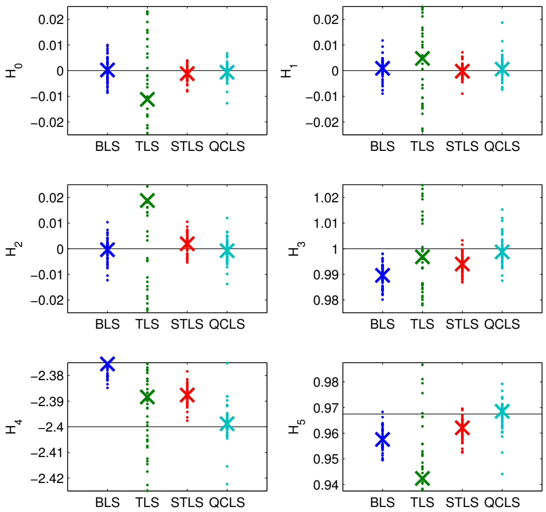
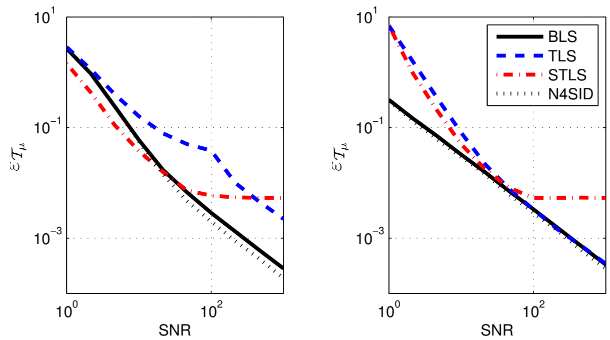


Fig. 16. Bias analysis of BLS, TLS, STLS, and QCLS with a μ-Markov model, v only, SNR = 10, $\mu = 10$, $n_{\rm mod} = 7$, and 10,000 samples. The horizontal line is the actual Markov parameter, and each dot is the result of simulation with a different noise realization. For each algorithm, 50 simulations are run. 'x' denotes the average value of the simulations.



(a) v only (b) w only

Fig. 17. Effect of SNR on BLS, TLS, STLS, and N4SID with various noise configurations, $\mu = 10$, and $n_{\rm mod} = 7$, and 5000 samples.

G. Accuracy versus SNR

The effect of SNR is examined in Figure 17. BLS and N4SID appear to produce the best estimates across all SNRs. The STLS estimates appear to be biased, even for high SNRs.

H. Accuracy versus Estimated Order

The effect of under- and over-estimating the order of a 7^{th} order plant (7) is shown in Figure 18. BLS is the least affected by underestimating the plant order. STLS is the most accurate when overestimating the plant order. BLS and N4SID appear to be unaffected by overestimating the order.

VIII. Conclusion

The purpose of this study is to investigate the properties of several least-squares identification algorithms with different model structures to assess the accuracy of Markov parameter estimates. The numerical results indicate that μ-Markov is a better model structure than ARX.

If the autocorrelation of the noise is known to within a scalar multiple, QCLS provides the best estimates for all noise cases and consistency under all noise configurations.

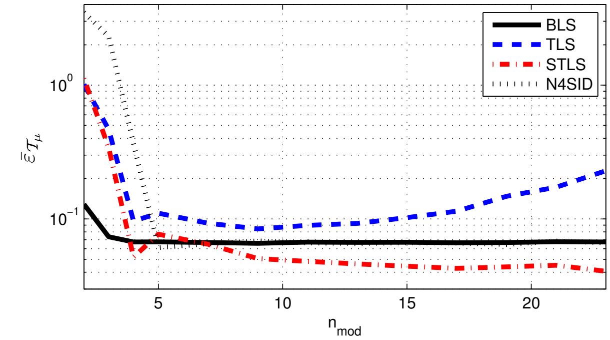


Fig. 18. Effect of order of BLS, TLS, STLS, and N4SID with uncorrelated v and w, SNR = 10, $\mu = 10$, and 5000 samples.

With only η and/or w, BLS with a μ-Markov structure is consistent and unbiased. With varying SNRs, noise configurations, and order estimates, BLS with a μ-Markov model is the most accurate or close to the most accurate algorithm. BLS with a μ-Markov structure is not affected by correlated noise signals and is not greatly affected by order estimate.

STLS with a μ-Markov model can produce good estimates of the Markov parameters, but it is significantly effected by SNR, noise configuration, and the amount of data.

TLS with a μ-Markov model generally yields estimates which are biased, not consistent, and negatively affected by overestimating the plant order and correlated noise signals.

The subspace identification method N4SID is comparable to BLS. The estimates are close amidst varying conditions.

MLS with an ARX structure provides a slight advantage over BLS with an ARX structure, but MLS with a μ-Markov structure is very similar to BLS with a μ-Markov structure.

References

[1] J.-N. Juang, *Applied System Identification*. Prentice Hall, 1993.

[2] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer, 1996.

[3] L. Ljung, *System Identification: Theory for the User*.

[4] R. E. Skelton and G. Shi, "The Data-Based LQG Control Problem," in *Proc. IEEE Conf. Dec. Contr.*, December 1994, pp. 1447–1452.

[5] K. Furuta and M. Wongsaisuwan, "Discrete-Time LQG Dynamic Controller Design Using Plant Markov Parameters," *Automatica*, vol. 31, pp. 1317–1324, 1995.

[6] M. A. Santillo and D. S. Bernstein, "Adaptive Control Based on Retrospective Cost Optimization," *AIAA J. Guid. Contr. Dyn.*, vol. 33, pp. 289–304, 2010.

[7] M. Kamrunnahar, B. Huang, and D. B. Fisher, "Estimation of markov parameters and time-delay/interactor matrix," *Chemical Engineering Science*, vol. 55, pp. 3353–3363, 1999.

[8] J. Huang and N. E. Wood, "Validation of a small signal probing concept for prognosis on a nonlinear model for longitudinal motion of a Boeing-747," in *Proc. Amer. Contr. Conf.*, June 10-12, 2009.

[9] J. Tokarzewski, *Finite Zeros in Discrete Time Control Systems*.

[10] T. H. Van Pelt and D. S. Bernstein, "Least Squares Identification Using μ-Markov Parameterizations," in *Conf. Dec. Contr.*

[11] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Anaylsis*. SIAM, 1987.

[12] T. H. Van Pelt and D. S. Bernstein, "Quadratically Constrained Least Squares Identification," in *Amer. Contr. Conf.*, 2001, pp. 3684–3689.

[13] H. J. Palanthandalam-Madapusi, T. H. Van Pelt, and D. S. Bernstein, "Parameter Consistency and Quadratically Constrained Errors-in-Variables Least-Squares Identification," *submitted for publication*.

[14] P. V. Overschee and B. De Moor, "N4SID: Subspace algorithms for the identification of combined deterministic stochastic systems," *Automatica*, vol. 30, pp. 75–93, 1994.

[15] K. Zhou, *Robust and Optimal Control*. Pretice Hall, 1995.