Theoretical and Experimental Study of Order Estimation

by J. Pan and W.S. Levine

TECHNICAL RESEARCH REPORT

SYSTEMS RESEARCH CENTER

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

TR 92-31

Theoretical and Experimental Study of Order Estimation[1]

Jiqin Pan and William S. Levine
Systems Research Center
and
Electrical Engineering Department
University of Maryland
College Park, MD 20742

Abstract

We consider estimation of system order and parameters for ARX systems with martingale difference noise. The order estimation criterion used is the accumulated prediction error (or predictive least-squares error) criterion which generates strongly consistent order estimates. High computational load and over-complex (when real data are processed) models are two glaring problems occurring in order estimation. We develop a fast and parallel algorithm for strongly consistent estimation of system order and parameters. This shows that order estimation by minimizing the APE cost function could be performed on-line. We also modify the APE criterion by adding a device for tradeoffs between model complexity and model accuracy. This modification helps to solve the over-complexity problem. Furthermore, we present a simulation study on order estimation to complement the theoretical analysis.

1 Introduction

Estimating the order of a dynamic system from knowledge of its inputs and outputs is an important problem in adaptive control and adaptive IIR (infinite impulse response) filtering, where order estimation amounts to the determination of the number of poles and perhaps the number of zeros of a linear model based on sequences of measurements. Neither having a more complex model than the system itself nor having an overly simple model is desirable in practice. A model should cover the dominant part of system dynamics. Too many poles and zeros in the model demand a stronger persistent excitation condition, lead to a larger variance of parameter estimates, and significantly increase the condition number involved in parameter estimation. Moreover, the extra poles and

[1]This research was supported by the Systems Research Center under National Science Foundation Grant CDR 8803012.

zeros often cluster together, which may lead to severe problems in many approaches for designing adaptive controllers and filters. It is equally bad when a model is overly simple and cannot represent the dominant system dynamics. This may significantly reduce the performance of IIR filters [1] and cause instability problems in adaptive control[2,3,4].

There have been many attempts to study the order estimation problem and some consistent estimation criteria have been developed for linear systems, provided that the system inputs and outputs are stationary and ergodic random signals [5]. Order estimation for linear systems with feedback inputs is, however, more difficult because the conventional stationarity and ergodicity assumption is no longer valid in this case. This difficulty was not overcome until very recently, Chen et al. and Hemmerly and Davis, respectively, develop two *strongly consistent* order estimation approaches for *adaptive control systems* [6,7,8,9].

The consistency of order estimation is a common theoretical goal. It brings, however, some potential problems to practical applications. Note that the consistency of an order estimation algorithm requires the algorithm to chase the "true" model. On the other hand, in practical situations, the "true" model is very likely to be a very high order linear model or even infinite-dimensional because physical systems often have some slight nonlinearity, parasitic high frequency dynamics, and/or complicated noise structure. Therefore, the "true" model is often technically too complicated to handle or is inefficient and uneconomical in designing and implementing controllers and filters. Thus, consistent order estimation often results in a model which is much more complicated than what is wanted. We will call this problem the over-complexity problem.

High computational complexity is another glaring problem occurring in the order estimation approaches for linear systems with feedback inputs. This is due to the fact that LS parameter estimates for a finite group of interesting models, say, ARX models with different numbers of poles and zeros, have to be determined before order estimation is performed.

In this paper, we show some modifications to the strongly consistent order estimation approach by Hemmerly and Davis [8,9] to (i) produce dominant models with model accuracy as great as desired and (ii) reduce the computational complexity without significantly sacrificing the performance of order estimation. The basic class of models discussed is ARX models. We also include results of numerical simulation upon short and medium length data sequences to provide further understanding of order estimation. Our main results are as follows:

(1) A fast and parallel order estimation approach, exploiting a structural property of least squares, ensures that the parameter and order estimation is strongly consistent, when the approach

is applied to the ARX systems considered in [9].

(2) A method of balancing model complexity and model accuracy, attached to the above theoretically consistent order estimation approaches, generates dominant models with arbitrarily small model inaccuracy.

(3) An investigation of the previous and proposed approaches, through numerical simulation upon computer-generated and real data.

The rest of this paper is organized as follows. Section 2 reviews and analyses the APE criterion. A fast and parallel algorithm for strongly consistent estimation of system order and parameters is developed in Section 3. To overcome the over-complexity problem occurring in consistent order estimation, the concept of dominant models is introduced by means of accumulated prediction error in Section 4. This concept is then applied to the APE criterion leading to the quasi-APE criterion. Section 5 conducts a simulation study of order estimation, including APE and quasi-APE criteria, to complement the theoretical convergence analysis. In order to illustrate the feasibility of order estimation in practical situations, the APE and quasi-APE criteria are applied to a laboratory-scale process in Section 6. Finally, Section 7 provides some conclusions.

2 Background on Accumulated Prediction Error Criterion

In this section, we will review some previous results on the APE criterion in terms of consistency of parameter and order estimation and the assumptions involved. These facts will be used in subsequent sections to prove consistency of our fast algorithm and correctness of our method for identifying dominant models. The generic form of the APE criterion for a linear regression model is expressed as follows

$$\min_k \sum_{t=1}^{n} \|y_t - \hat{y}_t(\hat{\theta}_{t-1}(k))\|^2$$

where $\hat{\theta}_{t-1}(k)$ is a least squares parameter estimate determined by data measurements up to time $t-1$, k denotes the number of parameters, and n is the time at which the latest datum arrives. Thus $\hat{y}_t(\hat{\theta}_{t-1}(k)) \triangleq \hat{\theta}_{t-1}(k)\phi_{t-1}^T(k)$ is the "*honest*" least squares prediction of y_n, where $\phi_{t-1}(k)$ is the regression vector associated with the linear regression model.

The APE criterion has been applied to order estimation for an m-output and l-input ARX system[9]:

$$\mathbf{y}_n + A_1\mathbf{y}_{n-1} + \cdots + A_{p_t}\mathbf{y}_{n-p_t} = C_0\mathbf{u}_n + C_1\mathbf{u}_{n-1} + \cdots + C_{q_t}\mathbf{u}_{n-q_t} + \omega_n \tag{1}$$

where the order (p_t, q_t) and parameter matrix

$$\theta^T(p_t, q_t) \triangleq [-A_1 \ \cdots \ -A_{p_t} \ \ C_0 \ \cdots \ C_{q_t}]$$

are unknown. The ARX system is assumed to satisfy:

A-I The noise $\{\omega_n\}$ is a martingale difference process satisfying

$$E\|\omega_n\|^2|\mathcal{F}_{n-1}) = \sigma^2 \ \ a.s. \tag{2}$$

$$\sup_n E\|\omega_n\|^\alpha|\mathcal{F}_{n-1}) < \infty \ a.s., \ \text{ for some } \alpha > 2. \tag{3}$$

A-II The true order (p_t, q_t) belongs to a known finite set, which we call the model complexity set,

$$\mathcal{O}_{p_0,q_0}^{p^*,q^*} \triangleq \{(p,q); p_0 \leq p \leq p^*, q_0 \leq q \leq q^*, p \text{ and } q \text{ are positive integers }\} \tag{4}$$

A-III The matrices A_{p_t} and C_{q_t} are of full row rank.

Remark 2.1 *(Martingale assumptions)* The martingale assumption on the noise is less restrictive in modeling problems, especially those corresponding to feedback systems, than the stationary or ergodic assumption on outputs and inputs of the systems in question. The stationary or ergodic assumption on output is usually not satisfied for feedback control systems. However, the assumption of martingale stochastic noise is true for many closed-loop systems. For example, consider a stochastic control system described in (1) with initial condition: $\{\mathbf{y}_{-1} \ldots \mathbf{y}_{-p_t} \ \mathbf{u}_0 \ \mathbf{u}_{-1} \ldots \mathbf{u}_{-q_t}\}$. The system input $\mathbf{u}_n$ at time $n \geq 1$ is a measurable function of past system outputs and the current reference signals, i.e., $\mathbf{u}_n = \mathbf{c}(\{\mathbf{y}_t\}_{t=-p_t}^{n-1}; \mathbf{r}_n)$, where $\mathbf{r}_n, n \geq 1$, are reference signals. Suppose that (i). the initial condition is independent of noise $\{\omega_n\}, n \in Z_+$. (ii). the reference signals $\mathbf{r}_n, n \in Z_+$, are independent of noise $\{\omega_n\}, n \in Z_+$. (iii). the model noise $\{\omega_n\}$ is a sequence of *independent* random variables with zero mean. Then, the model noise $\{\omega_n\}$ is a martingale difference process w.r.t. an increasing family of σ-fields $\{\mathcal{F}_n\}$ generated by available output/input measurements at time n and noise $\omega_i, \ i \leq n$. ∎

Remark 2.2 *(Finite-dimensional ARX systems)* Assumption **A-III** is made to insure the uniqueness of the true order (p_t, q_t). Assumption **A-II** implies that (1) finite-dimensional ARX systems are considered, and (2) the true system is in the model set to be considered. This assumption is needed for studying the consistency of order estimation. However, this assumption could be violated in practical situations. This issue will be addressed in Section 4.

Order estimation in the Hemmerly and Davis approach is performed in two steps:

1. Determine LS parameter estimates,

$$\hat{\theta}_t^T(p,q) \triangleq [-\hat{A}_{t,1},\ldots,-\hat{A}_{t,p},\hat{C}_{t,0},\ldots,\hat{C}_{t,q}], \quad t=0,\cdots,n-1,$$

   for each ARX model in (1) with order $(p,q) \in \mathcal{O}_{p_0,p_0}^{p^*,q^*}$.

2. Determine order estimates,

$$(\hat{p}_n,\hat{q}_n) = \arg \min_{(p,q)\in\mathcal{O}_{p_0,q_0}^{p^*,q^*}} \text{LS–APE}(p,q,n),$$

   where

$$\text{LS–APE}(p,q,n) \triangleq \sum_{t=1}^{n} \|\mathbf{y}_t - \hat{\theta}_{t-1}^T(p,q)\phi_{t-1}(p,q)\|^2$$

   and

$$\phi_{t-1}(p,q) \triangleq (\mathbf{y}_{t-1}^T \cdots \mathbf{y}_{t-p}^T \ \mathbf{u}_t^T \cdots \mathbf{u}_{t-q}^T)^T.$$

A significant advantage of using the APE criterion to estimate ARX model order is that strongly consistent order estimates can be generated even when the noise is not stationary and ergodic. This is shown by Hemmerly and Davis and is presented as follows.

Theorem 2.1 [Hemmerly and Davis, [9]] Suppose that ARX system (1) satisfies assumptions **A-I**— **A-III**. Let $V_n(p,q) \triangleq \sum_{t=1}^{n} \phi_{t-1}(p,q)\phi_{t-1}^T(p,q)$ and $\lambda_{min}(p,q,n)$ and $\lambda_{max}(p,q,n)$ denote the minimum and maximum eigenvalues of $V_n(p,q)$. If for each model order $(p,q) \in \mathcal{O}_{p_0,q_0}^{p^*,q^*}$

$$\phi_n^T(p,q)V_n^{-1}(p,q)\phi_n(p,q) \to 0 \ \ a.s. \ \text{ as } \ n\to\infty \tag{5}$$

$$\lambda_{min}(p,q,n) \to \infty \ \ a.s. \ \text{ as } \ n \to \infty \tag{6}$$

and

$$\lambda_{max}(p,q,n) = \mathbf{O}(\lambda_{min}(p,q,n)(\log \lambda_{min}(p,q,n))^{\gamma}) \ \ a.s., \gamma < 1 - \frac{2}{\alpha}, \tag{7}$$

where α is the constant appearing in (3), then for n big enough,

$$\text{LS–APE}(p,q,n) > \text{LS–APE}(p_t,q_t,n) \ \ a.s. \ \ \forall (p,q) \neq (p_t,q_t) \text{ and } (p,q) \in \mathcal{O}_{p_0,q_0}^{p^*,q^*} \tag{8}$$

As a result of (8),

$$(\hat{p}_n,\hat{q}_n) \to (p_t,q_t) \ \ a.s. \ \text{ as } \ n \to \infty, \tag{9}$$

and, also due to Theorem 3.1 in [5, chap. 6],

$$\hat{\theta}_n(\hat{p}_n,\hat{q}_n) \to \theta(p_t,q_t) \ \ a.s. \ \text{ as } \ n\to\infty, \tag{10}$$

with a convergence rate of $\mathbf{O}((\frac{\log(\lambda_{max}(p,q,n))}{\lambda_{min}(p,q,n)})^{1/2})$. ∎

Remark 2.3 *(Persistent excitation)* It follows from Lemma 3.1(i) in [5, chap. 6] that assumption (5) implies that $\det(V_n(p,q))$ cannot increase too fast as time n gets large. Note that $\det(V_t(p,q))$ is bounded below by $\lambda_{min}(p,q,n)$ to some power and above by $\lambda_{max}(p,q,n)$ to some power. Thus, $\lambda_{max}(p,q,n)$ cannot increase too fast. In addition, when the input and output processes $\mathbf{u}$ and $\mathbf{y}$ are almost surely and uniformly bounded, assumption (5) is implied by assumption (6). Assumption (6) is a sort of persistent excitation condition and is also a necessary condition for consistent parameter estimation. Assumption (7) means that the condition number of the normal matrix $V_n(p,q)$ could approach infinity, but not faster than some power of $\log(\lambda_{min}(p,q,n))$. From the point of view of computational accuracy, the condition number must be finite; otherwise the computational error involved in determining LS parameter estimates could become unbounded. ∎

Remark 2.4 *(Constraints on eigenvalues of $V_n(p,q)$)* As seen from (10), the larger $\lambda_{min}(n) \equiv \lambda_{min}(p,q,n)$ is, the faster the convergence rate of parameter estimation could be. However, $\lambda_{min}(n)$ can not be arbitrarily large. For example, the almost sure boundedness of input and output processes implies that $\lambda_{min}(n) = \mathbf{O}(n)$ *a.s.*. In addition, a big growth rate of $\lambda_{min}(n)$ is not always permitted because $\lambda_{min}(n)$ is bounded above by the "total energy" of input signals:

$$\lambda_{min}(n) \equiv \lambda_{min}(p,q,n) \leq \sum_{t=1}^{n} \|\mathbf{u}_t\|^2, \qquad \forall p > 0,\ q > 0, \tag{11}$$

for any data sequences $\{\mathbf{y}_k\}_{k=0}^{n}$ and $\{\mathbf{u}_k\}_{k=0}^{n}$. Remember that for any $p \geq 0$ and $q \geq 0$, $\lambda_{min}(p,q,n)$ is the minimum eigenvalue of normal matrix $V_n(p,q) = \sum_{t=1}^{n} \phi_{t-1}(p,q)\phi_{t-1}^T(p,q)$ and $\phi_{t-1}(p,q) = (\mathbf{y}_{t-1}^T \cdots \mathbf{y}_{t-p}^T\ \mathbf{u}_t^T \cdots \mathbf{u}_{t-q}^T)^T$. Specifically, $V_n(0,0) = \sum_{t=1}^{n} \mathbf{u}_t \mathbf{u}_t^T$. To show (11), recall that for a symmetric matrix R the following relationships hold:

$$\lambda_{min} = \min_{\|\mathbf{x}\|=1} \mathbf{x}^T R \mathbf{x} \qquad \text{and} \qquad \lambda_{max} = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T R \mathbf{x}. \tag{12}$$

So, if $\{R_m\}$ is a sequence of nested symmetric matrices such that $R_{m+1} = \begin{pmatrix} R_m & * \\ * & * \end{pmatrix}$, where the asterisks represent the elements of R_{m+1} which are not interesting to us, it follows from (12) that

$$\lambda_{min}(R_{m+1}) \leq \lambda_{min}(R_m) \qquad \text{and} \qquad \lambda_{max}(R_{m+1}) \geq \lambda_{max}(R_m) \tag{13}$$

and the condition number satisfies:

$$\kappa(R_{m+1}) \geq \kappa(R_m). \tag{14}$$

Obviously, $V_n(p,q)$ is nested matrices when either p or q is decreased. It follows from (13) that

$$\sum_{t=1}^{n} \|\mathbf{u}_t\|^2 = \operatorname{trace}(V_n(0,0)) \geq \lambda_{min}(0,0,n) \geq \lambda_{min}(p,q,n). \tag{15}$$

■

Remark 2.5 *(Effect of extra parameter estimates)* Obviously, any parameter estimator is not consistent if the corresponding system is strictly undermodeled , i.e., $p < p_t$ or $q < q_t$. Now consider the overmodeled case, i.e., parameter estimators of order (p,q), $p \geq p_t$, $q \geq q_t$. Recall from (1) that an ARX system with order (p_t, q_t) can be described, through introducing some zero coefficients, by a linear regression model of order (p,q) satisfying $p \geq p_t$ and $q \geq q_t$. And the parameter matrix of the model is equal to

$$\theta^T(p,q) = (-A_1 \ \cdots - A_{p_t} \ \underbrace{0 \cdots 0}_{p-p_t} \ \ C_0 \ \cdots C_{q_t} \ \underbrace{0 \cdots 0}_{q-q_t}).$$

For such a linear regression model, the least-squares parameter estimate, $\hat{\theta}_n(p,q)$, is equal to the solutions to the following normal matrices:

$$\underbrace{(\sum_{t=1}^{n} \phi_{t-1}(p,q)\phi_{t-1}^T(p,q))}_{V_n(p,q)}\hat{\theta}_n(p,q) = \sum_{t=1}^{n} \phi_{t-1}(p,q)\mathbf{y}_t^T.$$

Thus, the convergence rate can be described as

$$||\hat{\theta}_n(p,q) - \theta(p,q)|| = \mathbf{O}((\frac{\log(\lambda_{max}(p,q,n))}{\lambda_{min}(p,q,n)})^{1/2}) \quad a.s. \tag{16}$$

provided that the assumptions of Theorem 2.1 hold, where $\lambda_{min}(p,q,n)$ and $\lambda_{max}(p,q,n)$ denote the minimum and maximum eigenvalues of $V_n(p,q)$. Note from the structure of the matrix $V_n(p,q)$ that in the strictly overmodeled case: $p \geq p_t$, $q \geq q_t$, $p + q > p_t + q_t$, $V_n(p_t,q_t)$ is always a nested symmetric matrix of $V_n(p,q)$. Applying (13) to (16), we have the expected result that estimation of extra zero coefficients would slow the convergence rate of parameter estimation for all coefficients.

In addition, it follows from (14) that $\kappa(V_n(p_t,q_t)) \leq \kappa(V_n(p,q))$. So, assumption (7) could become very restrictive to consistent order estimation when p^* (or q^*) is much bigger than p_t (or q_t). ■

Remark 2.6 *(Computational complexity of order estimation)* When order estimation is approached by minimizing the APE criterion, a finite group of LS ARX models of different order should be determined. The computation of the APE's associated with all the models usually does not form a major computational load for each time iteration. However, as discussed in the following section, the computational complexity involved in the determination of all the LS ARX models is as high as $\mathbf{O}(\sum_{p=p_0}^{p^*}\sum_{q=q_0}^{q^*}(mp+lq)^2)$. Thus, order estimation could be prohibitive for on-line computation if there are many candidate models. This is often the case when the prior knowledge of the model is poor. ■

3 A Fast and Parallel Algorithm for Order Estimation

The LS parameter estimation for a group of models with both poles and zeros is computationally expensive. This is due to the fact that most currently available fast algorithms, for instance, the recursive LS lattice algorithm [10], the QR-based fast algorithms [11], and the algebraic methods for Toeplitz-like systems [12,13], cannot be applied in this case because the data matrix involved is not a Toeplitz matrix and the Yule-Walker matrix involved has no explicit displacement representation. As a result, the computational complexity is as high as $\mathbf{O}(p^{*^3}q^* + p^*q^{*^3})$ when the standard time recursive LS algorithm is used, where p^* and q^* represents, respectively, the maximum number of poles and zeros. Such a computational complexity is not practical in adaptive control and adaptive IIR filtering. Researchers have proposed several methods to reduce the computational complexity. An over-parameterization method is suggested in [14], where only models with the same numbers of poles and zeros are used for purposes of parameter estimation. Chen et al. suggested a scheme of truncating parameter estimates of the model having the maximum numbers of poles and zeros and then using the truncated estimates to replace the desired. However, as discussed in Remark 2.5 and shown in [15], the performance of these methods is not good because the over-parametrization causes the parameter estimates to have large variance and this could delay the arrival of the right order estimates which are consistent with the true model order. Karaboyas et al. developed an efficient ARX identification algorithm but it is inordinately complex at present [16].

Recently, a fast and parallel algorithm has been developed for approximating LS parameter estimates of a group of ARX models[15]. The algorithm is called the time and order recursive algorithm (TORA). The TORA can generate desired parameter estimates, which we will call the TORA estimates throughout this paper, in $\mathbf{O}((p^* + q^*)^2)$ flops. Equally importantly, the TORA estimates converge to the corresponding LS estimates under some mild conditions. Therefore, it is natural to consider replacing the LS estimates in the LS–APE by the TORA estimates. This idea results in the following *fast* method for simultaneous estimation of system order and parameters.

1. Obtain TORA parameter estimates

$$\tilde{\theta}_t^T(p,q) = [-\tilde{A}_{t,1} \ \cdots \ -\tilde{A}_{t,p} \ \ \tilde{C}_{t,0} \ \ \tilde{C}_{t,1} \ \cdots \ \tilde{C}_{t,q}] \quad t = 0, \cdots, n-1 \tag{17}$$

for each ARX system in (1) with order $(p,q) \in \mathcal{O}_{p_0,q_0}^{p^*,q^*}$. The tilde emphasizes that the parameter estimates are determined by using the fast *time and order recursive algorithm.*

2. Obtain order estimates

$$(\tilde{p}_n, \tilde{q}_n) = \arg \min_{(p,q)\in\mathcal{O}^{p^*,q^*}_{p_0,q_0}} \sum_{t=1}^{n} \|y_t - \tilde{\theta}^T_{t-1}(p,q)\phi_{t-1}(p,q)\|^2. \tag{18}$$

Note that the only difference between the proposed method and Hemmerly and Davis approach presented in Section 2 is the parameter estimates used. In their approach, LS parameter estimates $\hat{\theta}_{t-1}(p,q)$ are used. Instead, in the proposed method, TORA parameter estimates $\tilde{\theta}_{t-1}(p,q)$ are used to reduce the computational burden. Besides the computational efficiency, the proposed method preserves the asymptotic property of the Hemmerly and Davis approach when the input and output processes are uniformly bounded for each realization. This is illustrated in the following statement.

Theorem 3.1 Consider ARX systems in (1). Suppose that the input and output processes are uniformly bounded for each realization. If system order and parameters are estimated following the procedure described in (17) and (18), then , under assumptions **A-I** — **A-III** and (5) — (7),

$$(\tilde{p}_n, \tilde{q}_n) \to (p_t, q_t) \quad a.s. \quad \text{as} \quad n \to \infty, \tag{19}$$

and

$$\tilde{\theta}_n(\tilde{p}_n, \tilde{q}_n) \to \theta(p_t, q_t) \quad a.s. \quad \text{as} \quad n \to \infty, \tag{20}$$

with a convergence rate of $\mathbf{O}((\frac{\log(\lambda_{max}(p,q,n))}{\lambda_{min}(p,q,n)})^{1/2})$. ∎

This theorem can be proved by analyzing some previous results.

Observation 3.1: Suppose that the driving noise ω_n satisfying **A–I** and the "honest" prediction $\hat{\mathbf{y}}_n(p,q)$ is measurable with respect to $\mathcal{F}_{n-1}$. Let $C_n \triangleq \sum_{t=1}^{n} \|\mathbf{y}_t - \hat{\mathbf{y}}_t(p,q) - \omega_t\|^2$ denote the accumulated pure prediction error (APPE). Then applying Lemma 2(iii) in [17] to the APE immediately yields:

$$\sum_{t=1}^{n} \|\mathbf{y}_t - \hat{\mathbf{y}}_t(p,q)\|^2 = \sum_{t=1}^{n} \|\omega_t\|^2 + C_n(1+\mathbf{o}(1)) \quad a.s. \tag{21}$$

on the set $\{C_n \to \infty\}$ and

$$\sum_{t=1}^{n} \|\mathbf{y}_t - \hat{\mathbf{y}}_t(p,q)\|^2 = \sum_{t=1}^{n} \|\omega_t\|^2 + C_n(1+\mathbf{O}(1)) \quad a.s. \tag{22}$$

on the set $\{\lim_{n\to\infty} C_n < \infty\}$. $C_n\mathbf{o}(1)$ and $C_n\mathbf{O}(1)$ denote $\mathbf{o}(C_n)$ and $\mathbf{O}(C_n)$. (21) and (22) reflect that the limit behavior of APE (p,q,n) is decided by the APPE $C_n(p,q)$ and the accumulated energy of noise.

Specifically, we define the LS–APPE as

$$C_n(p,q) \triangleq \sum_{t=1}^{n} \|\mathbf{y}_t - \hat{\theta}_{t-1}^T(p,q)\phi_{t-1}(p,q) - \omega_t\|^2. \tag{23}$$

Then,

$$\sum_{t=1}^{n} \|\mathbf{y}_t - \hat{\theta}_{t-1}^T(p,q)\phi_{t-1}(p,q)\|^2 - \sum_{t=1}^{n} \|\mathbf{y}_t - \hat{\theta}_{t-1}^T(p_t,q_t)\phi_{t-1}(p_t,q_t)\|^2$$
$$= (C_n(p,q) - C_n(p_t,q_t))(1+\mathbf{o}(1)) \tag{24}$$

on the set $\{C_n(p,q) \to \infty, C_n(p_t,q_t) \to \infty\}$. *Thus, the APPE becomes a key factor for consistent order estimation when the LS–APE is used as an order estimation criterion.* ∎

Observation 3.2: It has been shown in Lemma 2.2 [9] that under the assumptions of Theorem 2.1, LS–APPE $C_n(p,q)$ defined in (23) has the following limit behavior:

$$C_n(p,q) = (1+\mathbf{o}(1))\sigma^2 \log\det(V_n(p,q)) \;\; a.s. \tag{25}$$

when $p \geq p_t$ and $q \geq q_t$. It has also been proved [(67), [9]] that in the undermodeled case, for instance, $p < p_t$,

$$\text{LS–APE}(p,q,n) \geq \sum_{t=1}^{n} \|w_t\|^2 + (1+\mathbf{o}(1))\|A_{p_t}\|^2\lambda_{min}(p,q,n) + \mathbf{O}(1). \;\; a.s.$$

Thus, (6), (21), and (22) imply that $C_n(p,q) \to \infty$ *a.s.* as $n \to \infty$ and

$$C_n(p,q) \geq (1+\mathbf{o}(1))\|A_{p_t}\|^2\lambda_{min}(p,q,n) \;\; a.s. \tag{26}$$

in the undermodeled case. ∎

Lemma 3.1 Suppose that $\mathbf{y}$ and $\mathbf{u}$ are almost surely and uniformly bounded. Then, under the assumptions of Theorem 2.1,

$$\sum_{t=1}^{n} \|\mathbf{y}_t - \tilde{\theta}_{t-1}^T(p,q)\phi_{t-1}(p,q) - \omega_t\|^2 \sim C_n(p,q) \;\; a.s. \tag{27}$$

∎

To prove Lemma 3.1, we need the following lemma.

Lemma 3.2 Let $\{a_t, t \in Z_+\}$ be a sequence of positive numbers. If the partial sums $\{b_t = \sum_{i=0}^{t} a_i, t \in Z_+\}$ are divergent, then

$$\sum_{t=0}^{n} a_t/b_t \leq 1 + \log b_n - \log b_0. \tag{28}$$

In addition, if $\{x_t, t \in Z_+\}$ is a sequence of positive numbers which converges to zero, then

$$\sum_{t=0}^{n} x_t a_t = \mathbf{o}(\sum_{t=0}^{n} a_t). \tag{29}$$

∎

Proof:

(i) $\sum\limits_{t=0}^{n} a_t/b_t = 1 + \sum\limits_{t=1}^{n}(b_t - b_{t-1})/b_t \leq 1 + \sum\limits_{t=1}^{n} \log(b_t/b_{t-1}) = 1 + \log b_n - \log b_0.$

(ii) For any given $\epsilon > 0$, $\exists \epsilon', N > 0$ such that $x_n < \epsilon' < \epsilon$, $n \geq N$. Hence, for any $n > N$,

$$0 \leq \frac{\sum_{t=0}^{n} x_t a_t}{\sum_{t=0}^{n} a_t} \leq \frac{\sum_{t=0}^{N} x_t a_t}{\sum_{t=0}^{n} a_t} + \frac{\epsilon' \sum_{t=N+1}^{n} a_t}{\sum_{t=0}^{n} a_t} \leq \frac{\sum_{t=0}^{N} x_t a_t}{\sum_{t=0}^{n} a_t} + \epsilon'.$$

Letting n be big enough, we have $0 \leq \frac{\sum_{t=0}^{N} x_t a_t}{\sum_{t=0}^{n} a_t} < \epsilon$ since $\sum_{t=0}^{n} a_t \to \infty$ as $n \to \infty$. ∎

Lemma 3.3 [Pan and Levine, [15]] Suppose that $\mathbf{y}$ and $\mathbf{u}$ are almost surely and uniformly bounded. If (i) $\lambda_{min}(p, q, n) \to \infty$ a.s. as $n \to \infty$, (ii) the LS parameter estimates $\hat{\theta}_n(p, q), n \leq 1$, are almost surely and uniformly bounded; i.e., $\exists K_1$ such that $||\hat{\theta}_n(p, q)|| < K_1$ *a.s.*, $\forall n \geq 1$, then

$$||\hat{\theta}_n(p, q) - \tilde{\theta}_n(p, q)|| = \mathbf{O}(\lambda_{min}^{-1}(p, q, n)) \;\; a.s. \tag{30}$$

where $\tilde{\theta}_n(p, q)$, $n \leq 1$, are the corresponding TORA parameter estimates. ∎

Returning now to Lemma 3.1, from the definition of $C_n(p, q)$ in (23) we have that for any (p, q),

$$\begin{aligned} C_n(p, q) &+ \sum_{t=1}^{n} ||(\hat{\theta}_{t-1}^T(p, q) - \tilde{\theta}_{t-1}^T(p, q))\phi_{t-1}(p, q)||^2 \\ &\geq \sum_{t=1}^{n} ||\mathbf{y}_t - \tilde{\theta}_{t-1}^T(p, q)\phi_{t-1}(p, q) - \omega_t||^2 \\ &\geq C_n(p, q) - \sum_{t=1}^{n} ||(\hat{\theta}_{t-1}^T(p, q) - \tilde{\theta}_{t-1}^T(p, q))\phi_{t-1}(p, q)||^2. \end{aligned}$$

This implies that

$$\begin{aligned} &|C_n(p, q) - \sum_{t=1}^{n} ||\mathbf{y}_t - \tilde{\theta}_{t-1}^T(p, q)\phi_{t-1}(p, q) - \omega_t||^2| \\ &\leq \sum_{t=1}^{n} ||\hat{\theta}_{t-1}^T(p, q) - \tilde{\theta}_{t-1}^T(p, q)||^2 ||\phi_{t-1}(p, q)||^2. \end{aligned}$$

At this point we care about the limit behavior of the sum

$$\sum_{t=1}^{n} ||\hat{\theta}_{t-1}^T(p, q) - \tilde{\theta}_{t-1}^T(p, q)||^2 ||\phi_{t-1}(p, q)||^2.$$

It follows from Lemma 3.3 and (7) that for almost every realization of $(\mathbf{y}, \mathbf{u})$, there exists an finite number N such that

1. $||\hat{\theta}_n(p,q) - \tilde{\theta}_n(p,q)|| \leq K_1\lambda_{min}^{-1}(p,q,n)$ for some finite positive number K_1,

2. $\lambda_{max}(p,q,n) \leq K_2\lambda_{min}^{3/2}(p,q,n)$ for some finite positive number K_2.

Hence, for any $n > N$,

$$\sum_{t=1}^{n} ||\hat{\theta}_{t-1}(p,q) - \tilde{\theta}_{t-1}(p,q)||^2||\phi_{t-1}(p,q)||^2$$

$$\leq \sum_{t=1}^{N} ||\hat{\theta}_{t-1}(p,q) - \tilde{\theta}_{t-1}(p,q)||^2||\phi_{t-1}(p,q)||^2 + \sum_{t=N+1}^{n} K_1^2\lambda_{min}^{-2}(p,q,t)||\phi_{t-1}(p,q)||^2. \tag{31}$$

Note that the first term on the right side of (31) is finite. The second term satisfies

$$K_1^2 \sum_{t=N+1}^{n} \lambda_{min}^{-2}(p,q,t)||\phi_{t-1}(p,q)||^2$$

$$\leq K_1^2K_2\sum_{t=1}^{n} \lambda_{min}^{-1/2}(p,q,t)\lambda_{max}^{-1}(p,q,t)||\phi_{t-1}(p,q)||^2. \tag{32}$$

Using assumption (6) and (29) in Lemma 3.2, we have that

$$\sum_{t=1}^{n} ||\hat{\theta}_{t-1}(p,q) - \tilde{\theta}_{t-1}(p,q)||^2||\phi_{t-1}(p,q)||^2$$

$$= \mathbf{o}(\sum_{t=1}^{n} \lambda_{max}^{-1}(p,q,t)||\phi_{t-1}(p,q)||^2) + \mathbf{O}(1)$$

$$\leq \mathbf{o}(\sum_{t=1}^{n} K_3\frac{||\phi_{t-1}(p,q)||^2}{\mathrm{trace}(\sum_{l=1}^{t}\phi_{l-1}(p,q)\phi_{l-1}^T(p,q))}), \tag{33}$$

where K_3 is the dimension of matrix $\sum_{t=1}^{n}\phi_{t-1}(p,q)\phi_{t-1}^T(p,q)$. Applying (28) in Lemma 3.2 to (33) and then using (7) yield

$$\sum_{t=1}^{n} ||\hat{\theta}_{t-1}(p,q) - \tilde{\theta}_{t-1}(p,q)||^2||\phi_{t-1}(p,q)||^2 \leq \mathbf{o}(\log\mathrm{trace}(V_n(p,q))) + \mathbf{O}(1)$$

$$= \mathbf{o}(\log\lambda_{max}(p,q,n)) + \mathbf{O}(1) \leq \mathbf{o}((1+\mathbf{o}(1))\log\lambda_{min}(p,q,n)) + \mathbf{O}(1). \tag{34}$$

Comparing (34) with either (25) or (26) yields

$$\sum_{t=1}^{n} ||\hat{\theta}_{t-1}(p,q) - \tilde{\theta}_{t-1}(p,q)||^2||\phi_{t-1}(p,q)||^2 = \mathbf{o}(C_n(p,q)) \quad a.s.$$

for $\forall(p,q) \in \mathcal{O}_{p_0,q_0}^{p^*,q^*}$ ∎

Proof of Theorem 3.1: Note that the assumptions of Theorem 2.1 are implied by the assumptions of Theorem 3.1. Lemma 3.1 and Observation 3.1 illustrate that the LS–APE(p,q,n) used in the Hemmerly and Davis approach has the same limit behavior as the following quantity

$$\sum_{t=1}^{n} ||y_t - \tilde{\theta}_{t-1}^T(p,q)\phi_{t-1}(p,q)||^2,$$

which we will call the TORA–APE. Thus, Theorem 3.1 holds. ∎

Remark 3.1 *(Applications to adaptive control)* The LS–APE, combined with the adaptive control strategy devised in [18], has been used in self-tuning control of ARX systems with martingale difference noise[9]. It is shown there that, under the assumptions of Theorem 2.1, the parameters and order are estimated in a strongly consistent way while the optimal cost of the whole adaptive control system is achieved asymptotically. This implies that the TORA-APE can also be applied to self-tuning control of ARX systems. ∎

Remark 3.2 *(Generalization to other fast algorithms)* It is worth pointing out that Theorem 3.1 is an extension to Theorem 2.1 and there are only two extra assumptions used in proving Theorem 3.1. They are the boundedness of outputs and inputs and the upper bound on the convergence rate of TORA parameter estimates to the corresponding LS parameter estimates. Thus, Theorem 3.1 holds if the TORA estimates in the procedure (described in (17) and (18)) are replaced by some other parameter estimates which, under assumptions of Theorem 3.1, converge to corresponding LS parameter estimates at the same or faster rate than that in (30). Hence, we call this kind of parameter estimates the approximate LS parameter estimates. ∎

4 Dominant Models

Besides the over-complexity problem pointed out in Section 1, the order estimation algorithm via the APE criterion has a special problem in that the model order is often overestimated for a quite long sequence of data. This phenomenon can be observed both experimentally, as shown later in Section 5, and theoretically. Recall that order estimates are virtually decided by the accumulated *pure* prediction error (APPE) defined below (22). It follows from Observation 3.2 that the APPE has the following limit behavior: for any $(\tilde{p}, \tilde{q}) \in \mathcal{O}^{p^*,q^*}_{p_0,q_0}$ with $\tilde{p} \geq p_t$ and $\tilde{q} \geq q_t$, and for any $(p,q) \in \mathcal{O}^{p^*,q^*}_{p_0,q_0}$ with $p < p_t$ or $q < q_t$,

$$\begin{aligned} C_n(\tilde{p},\tilde{q}) &= \mathbf{O}(\log \lambda_{max}(\tilde{p},\tilde{q},n)) \ \ a.s. && \text{(by (25))} \\ &\leq \mathbf{O}(\log \lambda_{max}(\tilde{p} \vee p, \tilde{q} \vee q, n)) \ \ a.s. && \text{(by (12))} \\ &= \mathbf{O}(\log \lambda_{min}(\tilde{p} \vee p, \tilde{q} \vee q, n) + \gamma \log\log \lambda_{min}(\tilde{p} \vee p, \tilde{q} \vee q, n)) \ \ a.s. && \text{(by (7))} \end{aligned}$$

and

$$\begin{aligned} C_n(p,q) &\geq (1+\mathbf{o}(1))K\lambda_{min}(p,q,n) \ \ a.s. && \text{(by (26))} \\ &\geq (1+\mathbf{o}(1))K\lambda_{min}(\tilde{p} \vee p, \tilde{q} \vee q, n) \ \ a.s. && \text{(by (12))}, \end{aligned}$$

where γ is the constant appearing in (7) and K is a positive number. This implies that the APPE increases in the overmodeled case *at most* at the rate of $\log \lambda_{min}(\tilde{p} \vee p, \tilde{q} \vee q, n)$, but it gets big in the undermodeled case *at least* at the rate of $\lambda_{min}(\tilde{p} \vee p, \tilde{q} \vee q, n)$. Therefore, the undermodeled models can be easily rejected. However, in the overmodeled case, the APPE $C_n(\tilde{p}, \tilde{q})$ is not that sensitive to the model order. As a result, overestimation of system order remains true for a long time until (p_t, q_t) becomes the argument minimizing the APE criterion over the model complexity set.

As a result of the above arguments, some improvement should be made to have the APE criterion fit engineering applications. Before coming up to some modification, we explore some interpretations of consistency of order estimation.

Remark 4.1 *(Model accuracy and model complexity)* Consider an ARX system of order (p_t, q_t) (expressed in (1)) subject to **A-I** – **A-III**. Suppose that the output and input processes satisfy the persistent excitation condition defined in (5) – (7) for any model order $(p, q) \in \mathcal{O}_{0,-1}^{p_u, q_u}$, where p_u and q_u are two finite integers. Theorem 2.1, Lemma 3.1, and Observation 3.1 state that the APE criterion has the following limit behavior:

$$APE(p, q, n) > APE(p_t, q_t, n) \quad a.s. \tag{35}$$

for any $(p, q) \in \mathcal{O}_{0,-1}^{p_u, q_u}$ if $(p_t, q_t) \in \mathcal{O}_{p_0, q_0}^{p^*, q^*} \subset \mathcal{O}_{0,-1}^{p_u, q_u}$, where the APE can be either LS-APE or TORA-APE. Note that the order estimate (p_n, q_n) is the argument minimizing the order estimation criterion over the model complexity set. This implies that the order estimate (p_n, q_n) and the associated parameter estimate $\theta_n(p_n, q_n)$ represent the *most accurate model of interest.* The model accuracy is then measured in terms of the APE (accumulated prediction error). When the "true" model order $(p_t, q_t) \in \mathcal{O}_{0,-1}^{p_u, q_u}$ is out of the model complexity set, which is the primary case we will discuss[2], the estimates (p_n, q_n) and $\theta_n(p_n, q_n)$ stand for the *most accurate approximation of the "true" model* or the *most accurate approximate model.* This is because the inequality (35) holds and then asymptotically,

$$(p_n, q_n) = \arg \min_{(p,q) \in \mathcal{O}_{p_0, q_0}^{p^*, q^*}} APE(p, q, n) = \arg \min_{(p,q) \in \mathcal{O}_{p_0, q_0}^{p^*, q^*}} |APE(p, q, n) - APE(p_t, q_t, n)| \quad a.s..$$

∎

As discussed before, the most accurate model could be overly complex. In other words, a slight sacrifice of accuracy could gain a large reduction of model complexity. Here, we use the number of

[2]Due to considerations of computational burden and, in particular, the usefulness of a model in the intended application, the model complexity set cannot be too large. In other words, the upper bound (p^*, q^*) cannot be too big. Thus, in practice, the "true model" could be out of the model set.

poles to measure the model complexity. For this reason, we define a set of acceptable approximate models:

$$\mathcal{M}_n(\rho) \triangleq \left\{(p,q) \in \mathcal{O}_{p_0,q_0}^{p^*,q^*} \mid \frac{APE_n(p,q)-\alpha_n}{\epsilon+\alpha_n} \leq \rho\right\}, \tag{36}$$

where the quantity α_n, representing the accuracy of the most accurate approximate model, is defined as

$$\alpha_n \triangleq \min_{(p,q)\in\mathcal{O}_{p_0,q_0}^{p^*,q^*}} APE(p,q,n). \tag{37}$$

The small positive number ϵ is introduced to prevent the divisor in (36) from being zero when n is very small. The positive number ρ will be explained below.

Remark 4.2 *(Properties of the set of acceptable approximate models $\mathcal{M}_n(\rho)$)* The constant ρ represents the tolerated amount of model accuracy degeneration in terms of the relative accuracy of a model with respect to the most accurate approximate model. The set $\mathcal{M}_n(\rho)$ shrinks as the threshold ρ is reduced. When $\rho = 0$, the set $\mathcal{M}_n(\rho)$ contains one element that is the order of the most accurate approximate model. This implies that the set $\mathcal{M}_n(\rho)$ is a non-empty set which contains at least the order of the most accurate approximate model. As will be shown, the threshold affects the tradeoff between model accuracy and model complexity. ∎

To solve the over-complexity problem arising in consistent order estimation via the APE criterion, we suggest the use of a less complex but reasonably accurate model, which we call the dominant model, to replace the most accurate approximate model. The order of the dominant model is determined below:

$$\begin{aligned} p_n &= \arg\min\{p \mid (p,q) \in \mathcal{M}_n(\rho)\} \\ q_n &= \begin{cases} p_n - 1, & \text{if } (p_n,q_n) \in \mathcal{M}_n(\rho) \\ \arg\min\{q \mid (p_n,q) \in \mathcal{M}_n(\rho)\}, & \text{otherwise.} \end{cases} \end{aligned} \tag{38}$$

The parameters of the dominant model are equal to $\theta_n(p_n,q_n)$ which has been computed during the process of order estimation. (38) states that the dominant model is an acceptable approximate model with lowest model complexity. The determination of the number of zeros comes from the observation that when sampling a continuous time system, the zero-order-hold equivalent of the continuous time system generically has same number of numerator coefficients and denominator coefficients.[3] As shown in the definition of the set of acceptable approximate models, the order

[3]The determination of the number of zeros may be refined, which is not pursued in this report. An interested reader may refer to [19, pp. 88-90].

estimation criterion described in (36)–(38) is based on the APE criterion. So, we call the criterion the quasi-APE criterion.

Remark 4.3 *(Relation between the dominant model and the threshold ρ)* The order estimation procedure in (38) manifests that the order of a dominant model is heavily dependent on the threshold ρ. The smaller ρ is, the more accurate the dominant model is with respect to the most accurate approximate model. When ρ is reduced to zero, the resulting dominant model is simply the most accurate approximate model. It should be emphasized that the intended application of a dominant model ought be taken into account and some "physical entities" need be resorted to in the determination of the value of ρ. Our experimental study suggests a value of about 1 for ρ when the LS-APE is used and a value of about 0.5 for ρ when the TORA-APE is used. ∎

5 Experimental Study of System Identification Using the APE Criterion

A large number of simulation studies has been done on identifying most accurate approximate models and dominant models for SISO ARX systems. The parameter estimates are determined by using either the time-recursive LS algorithm or the TORA. The order estimation is performed by means of the APE or quasi-APE criterion[4].

The simulation study is carried out for three purposes. The first is to observe the transient performance of parameter and order estimation for short sequences of data. The second is to observe the effect of various system characteristics on parameter and order estimation. These system characteristics include stability, controllability and observability, fast dynamics, and frequency content of exciting signals. The third is to further investigate the meaning of dominant models in terms of some other "physical entities" such as Bode plots, step responses, and the grammian of a balanced realization.

5.1 Stability

Stability is the most important system characteristic in the design of control systems and IIR filters. To explore its effect on order estimation, the following two systems have been considered.

[4]The threshold of 1.0 is used for the quasi-LS–APE criterion and 0.5 for the quasi-TORA–APE criterion.

System 1: The first system is a stable system. The system input-output relation is described as

$$y_n + 0.68y_{n-1} + 0.7y_{n-2} - 0.2692y_{n-3} = u_{n-1} - 0.09u_{n-3} + e_n. \tag{39}$$

The poles of the system reside at $-0.4792 \pm 0.8586i$ and 0.2784 and they have magnitude of $0.9833, 0.9833$, and 0.2784, respectively.

System 2: The second system is an unstable system. Its ARX description is

$$y_n + 0.09y_{n-1} + 0.7y_{n-2} - 0.289y_{n-3} = u_{n-1} - 0.09u_{n-3} + e_n \tag{40}$$

with two unstable poles $0.5902 \pm 0.8262i$ and one stable pole 0.2803.

The inputs of the systems are white binary signals with variance of 1.0 and the system noises are white noises with variance of 0.5. The statistics of order estimation are presented in Table 1, which are obtained through 20 replications of numerical simulation. There are 10 blocks of data in the table and each of them represents the distribution of order estimates for one of the two systems using one particular order estimation criterion. For instance, the upper-left block represents the distribution of order estimates for System 1 through minimizing the LS-APE criterion.

A. Order Estimation via APE Criterion

The consistency of order estimation via the LS–APE criterion can be observed from Table 1. When only 500 data points are available, the LS–APE criterion generates, in a 55% of the tests, order estimates consistent with the true system order for the stable system. Another interesting observation is that the correct number of poles is found in 80% of the tests. This implies that it is easier to identify the number of poles than the number of zeros. As indicated by Table 1, the TORA-APE criterion is no better in performance than the LS-APE criterion when 500 data points are used. This matches the observation that the TORA sacrifices somewhat the accuracy of LS parameter estimation for computational efficiency and the sacrifice becomes smaller and smaller as more and more data arrive [20]. An encouraging message from Table 1 is that if only the number of model poles is interesting, the performance of the TORA-APE criterion is close to that of the LS-APE criterion even when only 500 data points are available.

Stability has a great impact on order estimation. First, the TORA-APE criterion is subject to the boundedness of data points. As a result, the TORA–APE criterion cannot be applied to order estimation for unstable systems. Secondly, the data measurements of larger and larger magnitude make the condition number of the Yule-Walker matrices involved increase extremely fast. However,

q	1	2	3	1	2	3	1	2	3	1	2	3
p	LS-APE			TORA-APE			quasi-LS-APE			quasi-TORA-APE		
System (39): a stable system; 500 data points												
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1.00	0	0	1.00	0	0
3	0.20	*0.55*	0.05	0.55	*0.05*	0	0	0	0	0	0	0
4	0.20	0	0	0.40	0	0	0	0	0	0	0	0
System (40): an unstable system; 200 data points												
1	0	0	0	–	–	–	0	0	0	–	–	–
2	0.30	0	0	–	–	–	1.00	0	0	–	–	–
3	0.25	*0.15*	0	–	–	–	0	0	0	–	–	–
4	0.25	0.05	0	–	–	–	0	0	0	–	–	–

Table 1: The distribution of order estimates for the stable system (39) and unstable system (40). The order estimation is conducted by minimizing the LS/TORA–APE criterion or by using the quasi-LS/TORA–APE criterion.

as the computer simulation results shows, the LS–APE criterion works well when the measurements are in the numerical range allowed by the computer.

B. Order Estimation via Quasi-APE Criterion

The order of the dominant models for the stable system (35) is $(2, 1)$ almost surely no matter which quasi-APE criterion is used. The same story happens to the unstable system (36) when the quasi-LS–APE criterion is adopted. This seems quite reasonable for the unstable system because the two unstable poles are definitely dominant modes. To justify the dominant models of the stable system, we randomly pick one replication of the simulation and draw, in Figure 1, Bode plots of the dominant model determined by using the quasi-LS-APE criterion, based on 25 or 1000 data points. Figure 1 reflects that comparing with the system itself, the dominant models are rather accurate in terms of Bode plot, in particular when many data points are available. The curves of order estimates versus the number of data points used are plotted in Figure 2. By comparing Figure 2 with Figure 1, an interesting fact shows up that the order of a dominant model starts to be constant much earlier than parameters of the model converge. This would be helpful in applications because changing structure of controllers or signal processors is much more difficult than changing their parameters.

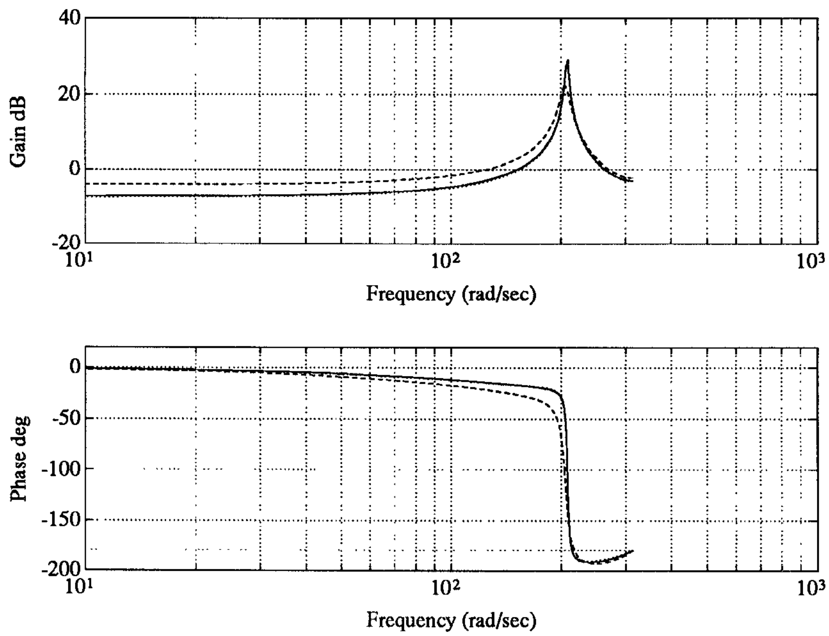

Figure 1: Comparison between Bode plots of different models. The solid line represents the Bode plot of the stable system (39). For the dominant model of order (2,1), identified based on 1000 data points, the Bode plot is indistinguishable from the solid line. The Bode plot of the dominant model of order (2,1), identified based on the first 25 data points, is drawn in the dashed line.

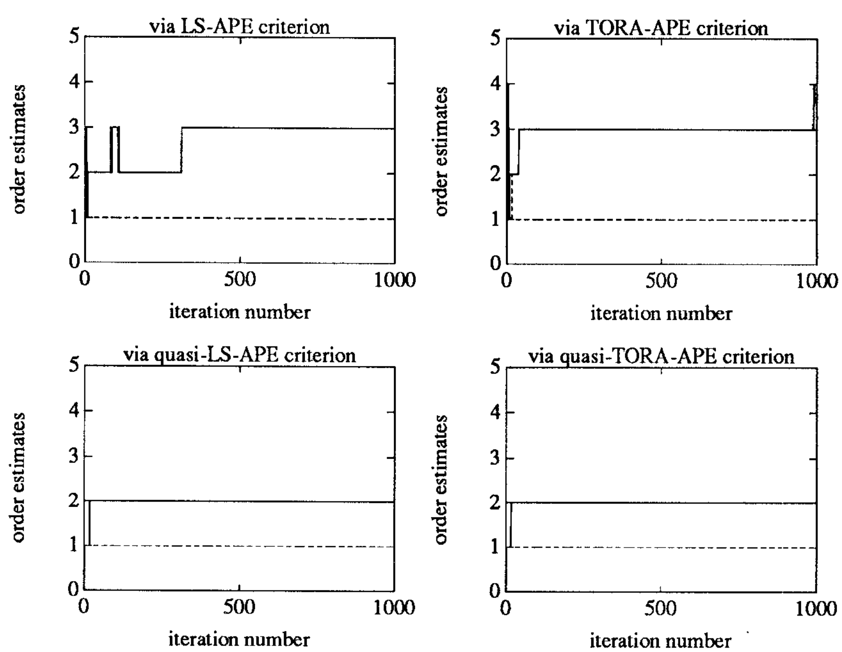

Figure 2: Order estimates for system (39). The solid line represents estimates of the number of poles; the dashed line that of zeros.

q	1	2	3	1	2	3	1	2	3	1	2	3
p	via LS-APE			via TORA-APE			via quasi-LS-APE			via quasi-TORA-APE		
1	0	0	0	0	0	0	0.95	0	0	0	0	0
2	0.25	0	0	0.20	0	0	0.05	0	0	1.00	0	0
3	0.65	0.10	0	0.80	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0

Table 2: The distribution of order estimates for the ARX system (41) via LS/TORA and quasi-LS/TORA-APE criteria. The distribution is computed throught 20 duplicate tests and 1000 data points are used for each test.

5.2 Controllability and Observability

In this example, our attention is devoted to an uncontrollable ARX system:

$$y_n - 1.6386y_{n-1} + 0.8646y_{n-2} - 0.1225y_{n-3} = 0.0367u_{n-1} + 0.0218u_{n-2} - 0.0067u_{n-3} + e_n, \quad (41)$$

where the noise is a white noise with a variance of 0.2 and the white binary input has a unit variance. The poles of the system are located at 0.2231 and $0.7077 \pm 0.2189i$ and the zeros reside at -0.8182 and 0.2231. The distribution of order estimates is presented in Table 2.

At first glance, Table 2 looks surprising because the LS–APE and TORA–APE criteria have 75% chance or more to generate pole estimates which are consistent with the true system. This indicates that even for uncontrollable SISO ARX systems, the APE criterion is still able to asymptotically determine the right order of the system. Note that the system output is the summation of two parts. The first part is stimulated by the system input and is not affected by the cancelable pole and zero. The second part is equal to the output of a pure AR model excited by the system noise. The denominator of the pure AR model is exactly the same as the denominator of the original ARX system and it contains the cancelable pole. Consequently, the true order of the uncontrollable system could be identified. However, uncontrollability indeed causes big trouble to order estimation in aspect of the number of the data points needed for consistent order estimation. As seen from Table 2, even when long sequences of 1000 data points have been available, there is still very small chance of 0.1 that the order estimates are equal to the true order.

The order of the dominant models suggested by the quasi-LS–APE and TORA–APE criteria is mainly equal to $(1,1)$ and $(2,1)$. respectively. The order $(2,1)$ appears to be a right choice in terms of model input-output relation because the cancelable pole and zero have no effect on the *system response to system input* and compared with system noises, system inputs usually contribute to

q	1	2	3	1	2	3
p	LS-APE			TORA-APE		
1	71.6777	71.8812	72.4218	71.5529	71.5691	72.1031
2	39.8992	40.2544	40.5511	40.5053	41.0865	41.6239
3	*39.3537*	39.5944	39.9123	*39.9523*	40.5676	41.1423
4	39.6032	39.9410	40.3547	40.1792	40.7686	41.3350

Table 3: The LS–APE and TORA–APE of one test when 1000 data points are used. The ARX system considered in order estimation is described in (41).

system outputs dominantly. The order $(1,1)$ comes from the improper choice of the threshold for the quasi-LS–APE criterion. As you see from Table 3, when the threshold is reduced from 1.0 to 0.7, the order of the dominant model, estimated via the quasi-LS–APE criterion, becomes $(2,1)$ instead of $(1,0)$.

5.3 Fast Dynamics

Order estimation has also been done for two discrete time ARX systems with fast dynamics. Their transfer functions are the zero-order-hold (zoh) equivalents of the following two continuous time systems:

$$\frac{s^2 + 0.2s + 4.01}{(s^2 + 0.2s + 1.01)(s^2 + 2 \cdot 0.4472 \cdot 11.1803s + 11.1803^2)} \tag{42}$$

and

$$\frac{s^2 + 0.2s + 4.01}{(s^2 + 0.2s + 1.01)(s^2 + 2 \cdot 0.0100 \cdot 10.0005s + 10.0005^2)}. \tag{43}$$

The sampling interval is, respectively, 0.1 sec for system (42) and 0.05 sec for system (43). These ARX systems are excited by white binary inputs with unit variance. The noise of these systems is white noise having variance of 0.3 and independent on the model inputs. The distribution of order estimates based on 500 data points is presented in Table 4. The LS–APE criterion always gives the right estimate of the number of poles. In contrast, the performance of the TORA-APE criterion is not good for the systems (42) and (43). To have better performance, we can use the modified TORA in the APE criterion [20]. Of course, as predicted by the previous theoretical analysis, their performance should be the same asymptotically when data sequences are uniformly bounded.

According to both the quasi-LS–APE criterion and the quasi-TORA–APE criterion, the number of poles of the dominant models should be 2 for the first case and 4 for the second case. Intuitively, this is quite believable because the fast dynamics in the first case has a damping ratio of 0.4472, but 0.01 in the second case. This very small damping ratio causes a big jump in the magnitude

q	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5
p	via LS-APE				via TORA-APE				via quasi-LS-APE				via quasi-TORA-APE				
	Case 1: fast dymanics with a mediate damping ratio																
2	0	0	0	0	0.	0	0	0	1.00	0	0	0	1.00	0	0	0	0
3	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0
4	0.75	0.20	0.05	0	0.60	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0.30	0	0	0	0	0	0	0	0	0	0	0	0
	Case 2: fast dymanics with a small damping ratio																
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0.10	0	0	0
4	0.70	0.20	0.10	0	0	0	0	0	0	0	1.00	0	0	0	0.50	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	0
6	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0.20

Table 4: The distribution of order estimates via LS/TORA and quasi-LS/TORA APE criteria. It is obtained throught 20 replications of tests.

of the transfer function of the system at the frequency of 10.0005 rad/sec. Also, we can use a balanced realization to evaluate the order of the dominant models. The balanced realization of the z-o-h equivalents of systems (42) and (43) has been generated by using MATLAB. The realization has four states. By comparing the diagonal elements of the grammian of the balanced realization, reduced-order models having one state, two states and three states are generated and then the step responses of the equivalents and their reduced-order models are drawn in Figure 3. There is no noticeable difference in step response between the z-o-h equivalent of system (42) and its reduced-order models except for the one-state model. For system (43), no model reduction can be done without obviously changing the model step response. This exactly matches the order of the dominant models estimated through the quasi-LS/TORA–APE criteria.

5.4 Rohrs' Example

Adaptive control algorithms which are theoretically stable could become unstable in the presence of *mild* unmodeled dynamics. This fact was first reported in the example of Rohrs et al. [3]. The adaptive control scheme of Rohrs' example is designed based on a first-order model with transfer function $g(s) = \frac{2}{s+1}$. The actual plant is a third-order system with a transfer function

$$P(s) = \frac{2}{s+1} \cdot \frac{229}{s^2 + 30s + 229}. \tag{44}$$

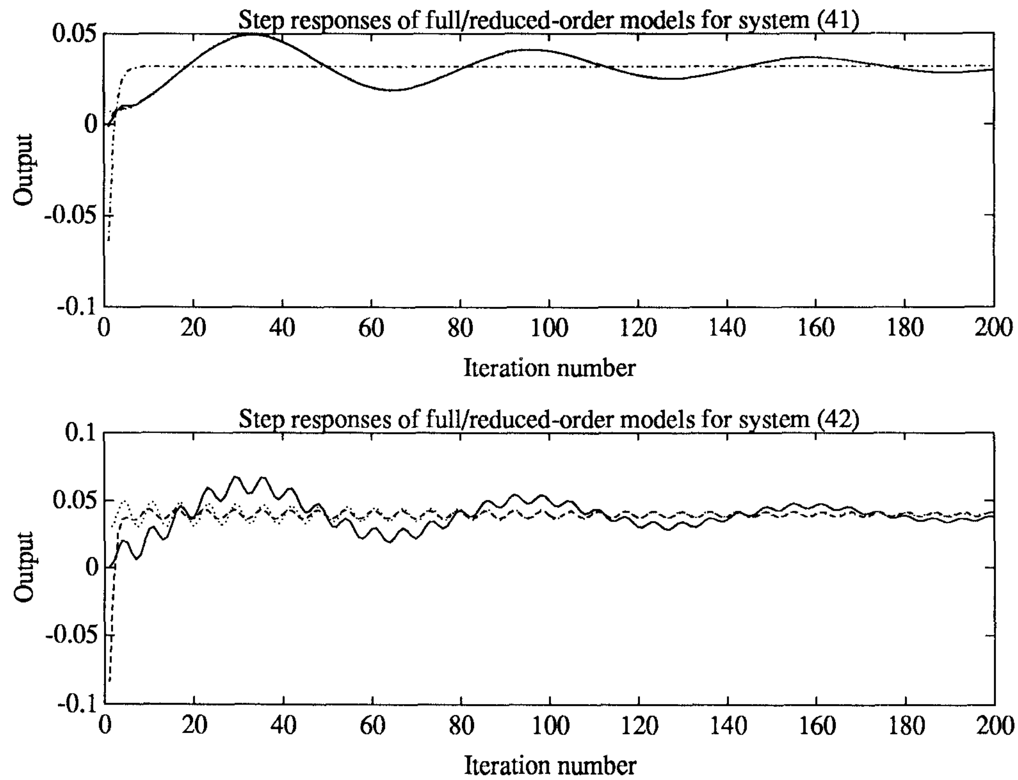

Figure 3: Step responses of the z-o-h equivalents of Systems (42) and (43) as well as their reduced-order models. The z-o-h equivalents have four states and their step responses are drawn in solid lines. The step responses of reduced-order models of three states, two states, and one state (only for System (42)) are plotted in dashed, dotted, and dashdot lines, respectively.

When the sampling interval is 0.05 sec, its z-o-h equivalent is

$$P(z^{-1}) = \frac{0.0976}{1-0.9512z^{-1}} \cdot \frac{0.0676(1+2.5703z^{-1})(1+0.1792z^{-1})}{1-0.9400z^{-1}+0.2231z^{-2}}. \tag{45}$$

Much research on robust adaptive control has been stimulated by Rohrs' example [21,22]. A major difficulty in robust adaptive control is that the amount of unmodeled dynamics tolerated by a robust adaptive control algorithm is very problem-dependent. In this section, we investigate, through numerical simulation, what model order should be estimated for the system (45).

An ARX system with transfer function $P(z^{-1})$ in (45) is considered in the simulation:

$$y_t = P(z^{-1})u_t + P_e(z^{-1})e_t \tag{46}$$

where z^{-1} represents the one-step delay operator and

$$P_e(z^{-1}) = \frac{1}{(1-0.9512z^{-1})(1-0.9400z^{-1}+0.2231z^{-2})}. \tag{47}$$

We have modified Rohrs' example by the addition of a driving noise term. Such a term is likely to be present in real applications. Of course, in a real system, there would also be some measurement noise. We have not included measurement noise yet. The Bode plots of $P(z^{-1})$ and $P_e(z^{-1})$ are drawn in Figure 4. The system input and noise are sinusoidal signals:

$$u_t = 2 + \sum_{i=1}^{4} 0.1\sin(\omega_i T_s t) \quad \text{and} \quad e_t = a_e \sin(\omega_e T_s t). \tag{48}$$

Specifically, four combinations of system input and noise with different frequency content are used[5]:

(1) no noise and high frequency input: $\omega_i = 1, 5, 10, 15$; $a_e = 0$,

(2) low frequency noise and low frequency input: $\omega_i = 1, 1.3, 1.6, 2$; $a_e = 0.2, \omega_e = 5$,

(3) low frequency noise and high frequency input: $\omega_i = 1, 5, 10, 15$; $a_e = 0.2, \omega_e = 5$,

(4) high frequency noise and low frequency input: $\omega_i = 1, 1.3, 1.6, 2$; $a_e = 0.2, \omega_e = 16.1$.

The simulation results are summarized in Table 5 and Table 6. After 25 iterations, order (1,1) or (1,0) is not suggested as the order of any dominant model except for one case in which the threshold ρ is larger than 200%. Surprisingly, this is quite different from the model used in most

[5]The signals considered here are out of the range covered by the convergence theorems at present. However, these signals, including input signals and noise signals, are quite typical in engineering. Most importantly, we want to explore the effect of frequency content of input and noise signals on order estimation by means of these sinusoidal noise and dither.

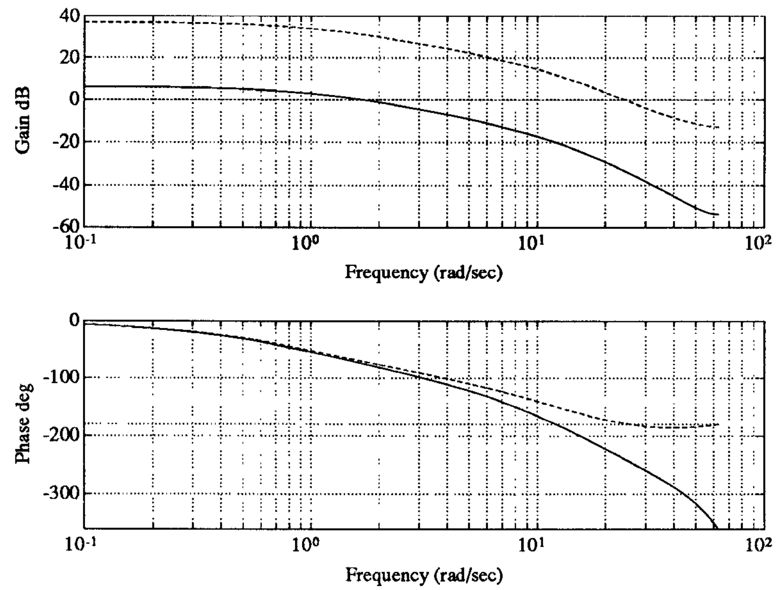

Figure 4: Bode plots of $P(z^{-1})$ (in solid lines) and $P_e(z^{-1})$ (in dashed lines)

case	1		2		3		4	
n	[25, 31]	[32, 500]	[25, 31]	[32, 500]	[25, 31]	[32, 500]	[25, 31]	[32, 500]
MAM	(2,1)	(2,1)	(2,1)	(3,1)	(2,1)	(3,1)	(4,3)	(4,3)
DM	(2,1)	(2,1)	(2,1)	(2,1)	(2,1)	(2,1)	(3,2)	(3,2)

Table 5: Order estimates of System (46) by using LS–APE and quasi-LS–APE criteria. The MAM stands for the order of most accurate models and the DM represents the order of dominant models, where the threshold ρ is set as 1.0.

case	q	1	2	3	4	ρ	order
	p	accumulated prediction error					estimate
1	1	0.0729	0.0585	0.0712	0.0946	0	(2,1)
	2	0.0232	0.0406	0.0654	0.0926	0.3	(2,1)
	3	0.0235	0.0411	0.0660	0.0927	1.0	(2,1)
	4	0.0238	0.0414	0.0662	0.0930	2.0	(2,1)
	5	0.0238	0.0414	0.0663	0.0931	5.0	(1,1)
2	1	104.6901	104.9540	105.2291	105.4041	0	(3,1)
	2	0.6335	0.9139	1.2201	1.4655	0.3	(3,1)
	3	0.3828	0.6671	0.9705	1.1990	1.0	(2,1)
	4	0.3843	0.6704	0.9738	1.1955	2.0	(2,1)
	5	0.3864	0.6718	0.9755	1.1975	5.0	(2,1)
3	1	104.3822	104.6759	104.6472	103.9399	0	(3,1)
	2	0.5273	0.8131	1.1408	1.4258	0.3	(3,1)
	3	0.3807	0.6761	1.0083	1.2794	1.0	(2,1)
	4	0.3825	0.6792	1.0108	1.2751	2.0	(2,1)
	5	0.3847	0.6803	1.0123	1.2770	5.0	(2,1)
4	1	31.5488	31.6364	31.7723	31.7963	0	(4,3)
	2	14.0029	13.9382	13.8531	13.4541	0.3	(3,2)
	3	0.6285	0.5959	0.5983	0.6357	1.0	(3,2)
	4	0.6356	0.6005	0.5898	0.6382	2.0	(3,2)
	5	0.6352	0.6055	0.5911	0.6443	5.0	(3,2)

Table 6: Order of dominant models and corresponding model accuracy when 500 data points are used

considerations of Rohrs' example [3,4,21,22]. In addition, the order of dominant models varies as exciting signals change. This desired feature would help to reduce model uncertainty to a level that a robust adaptive control algorithm can tolerate.

Before explaining the simulation results, we analyze the time responses of the ARX system (46) to system inputs and noise described in (48). It is easy to show from (46) and (48) that

$$\begin{aligned} y_t = |P(1)| \quad & + \sum_{i=1}^{4} |P(e^{-j\omega_i T_s})| 0.1 \sin(\omega_i T_s t + \angle P(e^{-j\omega_i T_s})) \\ & + |P_e(e^{-j\omega_e T_s})| a_e \sin(\omega_e T_s t + \angle P_e(e^{-j\omega_e T_s})), \end{aligned} \tag{49}$$

where T_s is the sampling interval.

A. Case 1: no noise and high frequency input

All the APE's are very small in this case because there is no system noise. Note that the frequency range of input signals is from 0 to 15 rad/sec and as shown in Figure 5, the magnitude of $P(j\omega)$ decays about 30 dB within the range. So, all the three modes of the system (45) are excited and only the slow mode, e^{-T_s}, is energetically stimulated. This results in that all the APE's are in same order.

To justify the dominant model of order $(2,1)$, we compare, in terms of Bode plots, the dominant model with the original system (44) and the model $g(s) = \frac{2}{s+1}$, which is used for controller design in Rohrs' example. Figure 5 shows that the dominant model of order (2,1) is more accurate than the model $g(s)$.

B. Case 2 and Case 3: low/high frequency input and low frequency noise

It seems puzzling that APE's are nearly same in Case 2 and Case 3 even though all the three modes, e^{-T_s} and $e^{T_s(-15\pm j2)}$, are excited in Case 3 instead of only the slow mode e^{-T_s} in Case 2. However, the excitation of the slow mode and the pair of fast modes is much different in the degree of strongness. Note that the mode e^{-T_s} is excited by both input signals and noise signals. As shown in Fig. 4, $|P_e(e^{-5T_s})|$ is about 15dB larger than $|P(e^{-T_s})|$ and $|P(e^{-T_s})|$ is about 25 dB larger than $|P(e^{-15T_s})|$. Consequently, it follows from (49) that the dominant frequency components of y_t are composed of the system response to low frequency input signals and noise in these cases. In other words, only the slow mode e^{-T_s} is strongly excited rather than the fast modes. That is the reason why we have nearly same APE's in both cases.

People may ask why the dominant models listed in Table 6 always have more than one pole since only the slow mode is strongly excited. Remember that dominant models are some approximation of a most accurate model. As illustrated in Figure 6, the most accurate model has a sharp resonance

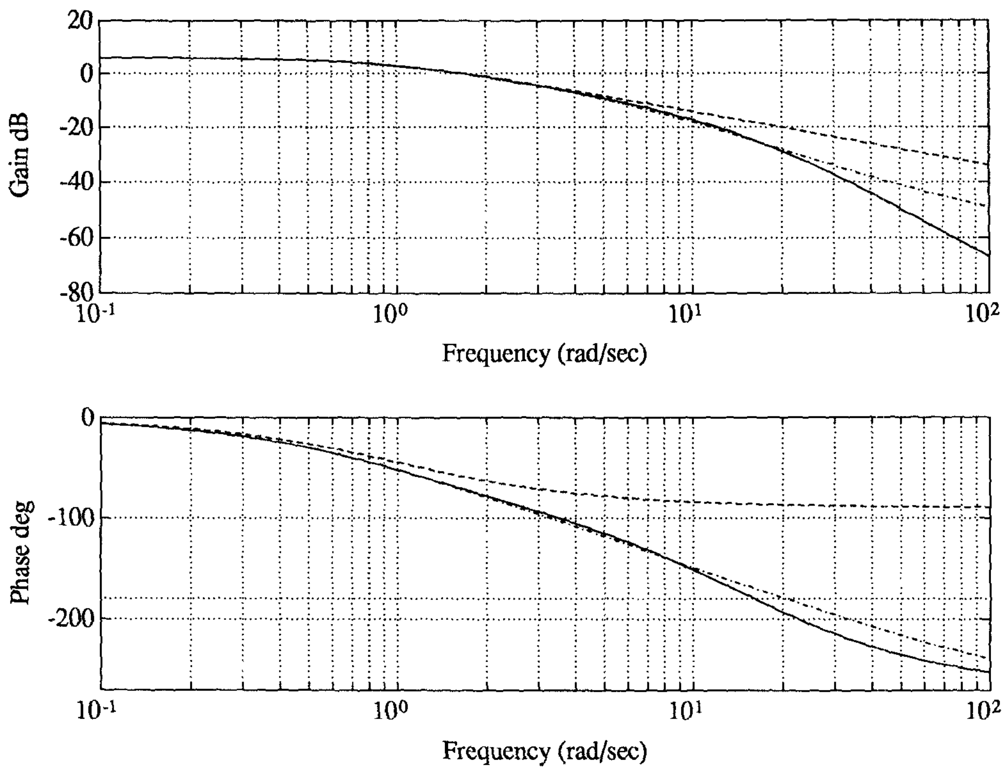

Figure 5: Bode plots of the System (44), the first-order model $g(s) = \frac{2}{s+1}$, and the continuous-time equivalent of the dominant model of order $(2,1)$. They are, respectively, drawn in solid, dashed, and dashdot lines. The output sequence is generated as the noise-free measurements of the output of system (44) which is excited by the high frequency input described in the case 1.

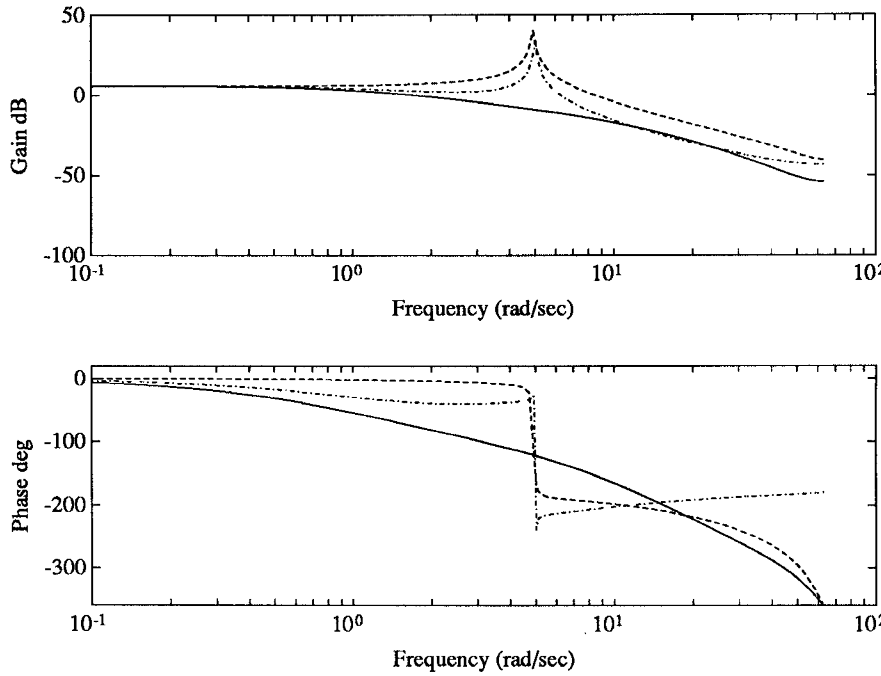

Figure 6: Bode plots of the System (45), the most accurate model and second-order dominant model in Case 2. The plots are drawn in solid, dashdot, and dashed lines, respectively.

mode[6]. Therefore, any good approximation of the model should have two or more poles.

C. Case 4: low frequency input and high frequency noise

In this case, the slow mode is excited by the low frequency input signal and the pair of fast modes is stimulated by the high frequency noise. Note that $|P(e^{-2T_s})|$ is about the same as $|P_e(e^{-16.1T_s})|$. This results in that all the three modes are excited nearly at a same level. Therefore, the dominate model should have three poles; otherwise the APE's would be increased significantly, as shown in Table 6.

[6]The most accurate model depicted in Figure 6 is not accurate in terms of Bode plot. This is due to a small signal-to-noise ratio which is equal to

$$\frac{\sum_{t=1}^{500}[|P(1)| + \sum_{i=1}^{4}|P(e^{-j\omega_i T_s})|0.1\sin(\omega_i T_s t + \angle P(e^{-j\omega_i T_s}))]^2}{\sum_{t=1}^{500}[|P_e(e^{-j\omega_e T_s})|a_e\sin(\omega_e T_s t + \angle P_e(e^{-j\omega_e T_s}))]^2}.$$

In Case 2, the SNR has a value of 4.3149.

5.5 Summary

As illustrated in the above simulation results and heuristic analyses, order estimates usually quickly enter a steady state after tens iterations. The number of poles of dominant models is dependent on the threshold ρ and in particular, the number of (unknown) system poles which are energetically excited by exogenous signals and noise. The number of poles of a dominant model is often equal to, or is one more than, the number of the energetically excited poles when the threshold ρ is not larger than 2. The performance of a dominant model measured in terms of Bode plots is subject to that of the most accurate model.

6 Application to a Laboratory-Scale Process

To illustrate the feasibility of order estimation in practical situations, we performed order estimation on real data. The data were collected from output-input measurements of a laboratory temperature-control process and were stored in the System Identification Toolbox by Ljung [23]. Ljung has performed system identification on the data using several identification techniques in [23,24]. By means of the standard performance measure of parameter estimation, Akaike loss function, and some physical entities like time response and Bode plots, a conclusion has been drawn that the proper model for the process is an ARX model with *two* poles, *one* zero, and *three* steps of input delay. Here, we use the data again to perform system identification using the LS/TORA–APE and quasi-LS/TORA–APE criteria. Specifically, the data will partitioned into two parts. The first part, which will be directly used for identification, consists of 1000 pairs of output-input measurements. The input actually is a realization of a binary random signal with a shifting probability of 0.2. The second part will be used to evaluate the identified model and is composed of input-output measurements of the process excited by a binary random signal with a shifting probability of 0.5. The simulation results are summarized in Table 7 and Figure 7.

Order estimation is conducted through an ARX model with a presumably assigned value of the step of model input delay. When three steps of input delay are assumed, the LS–APE criterion suggests an over-estimated model order $(5, 2)$ and the TORA–APE criterion generates a less over-estimated model order $(3, 1)$. This matches our concern on over-estimation of model order discussed in Section 4. It is worth pointing out that, as stressed in [24], this temperature-control process is "well behaved: it has reasonably simple dynamics with quite small distances." So, for industrial-scale processes, the over-estimation problem of the APE criterion may become more severe.

q	1	2	3	1	2	3	order
p	**LS–APE**			**TORA–APE**			estimate
	no input delay						
1	37.1025	31.0525	15.2163	38.6826	33.2420	17.9491	LS: (3,3)
2	11.9800	9.9968	4.1492	13.9334	12.7757	*7.4572*	QL: (2,3)
3	10.3306	8.9175	*4.0111*	12.5401	11.9558	7.6617	TO: (2,3)
4	10.3671	8.9504	4.0245	12.7541	12.1687	7.8810	QT: (2,1)
	one step input delay						
1	30.6897	14.5275	5.3256	32.4458	17.1604	7.9843	LS: (4,3)
2	9.6751	3.5141	2.6993	11.8406	6.6487	*5.7242*	QL: (2,2)
3	8.5958	3.3816	2.5991	10.9813	6.6912	5.9484	TO: (2,3)
4	8.6293	3.3905	*2.5751*	11.1511	6.8591	6.1715	QT: (3,1)
	two step input delay						
1	14.3304	4.7496	2.5228	16.3053	7.1459	5.2024	LS: (3,3)
2	3.3529	2.1344	2.0359	5.5914	*4.7864*	4.9953	QL: (2,1)
3	3.1921	2.0239	*1.9132*	5.5939	4.8360	5.1664	TO: (2,2)
4	3.1914	1.9984	1.9170	5.7237	5.0093	5.4212	QT: (2,1)
	three step input delay						
1	4.6565	2.4755	2.0608	6.3974	4.4363	4.2813	LS: (5,2)
2	2.0473	2.0014	1.9581	3.8839	4.1392	4.4098	QL: (2,1)
3	1.9772	1.8876	1.9140	*3.8755*	4.1378	4.5797	TO: (3,1)
4	1.9594	1.8809	1.9111	3.9839	4.3035	4.7586	QT: (2,1)
5	1.9627	*1.8680*	1.9095	4.1188	4.4780	4.9766	
6	1.9762	1.8852	1.9281	4.2654	4.6748	5.2240	

LS— LS–APE; QL — quasi-LS–APE; TO — TORA–APE; QT — quasi-TORA–APE

Table 7: APE's and order estimates via LS/TORA–APE and quasi-LS/TORA APE criteria

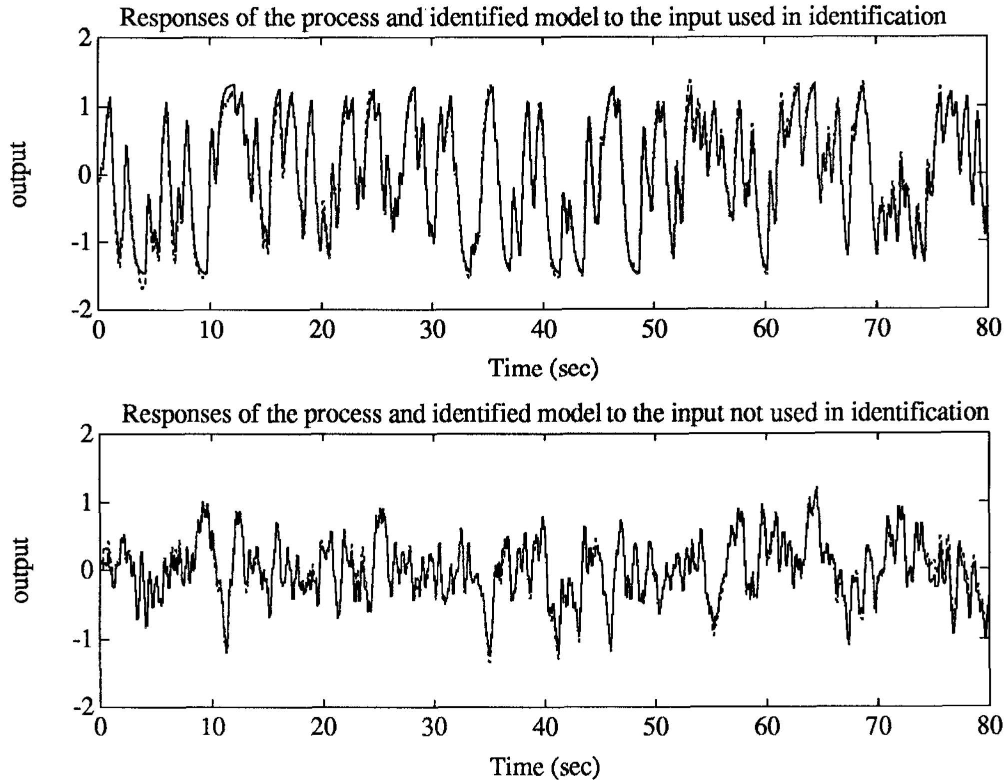

Figure 7: Comparison of time responses of the process and dominant model

In contrast, the quasi-APE criteria yield the order estimate of $(2, 1)$ no matter which quasi-APE criterion is used. This matches the results obtained in [23,24]. To further illustrate the accuracy of the dominant models, time responses of the original process and the dominant model determined by using the quasi-LS-APE criterion have been plotted in Figure 7. The dominant model matches the dynamics of the process in time response very well.

Another interesting fact is that, as the model input delay assumed in order estimation is successively increased up to the true one, the values of the corresponding minimum APE's become smaller and smaller. For instance, as shown in Table 7, the minimum LS–APE falls down from 4.0111 to 1.8680 when the assumed input delay approaches the true input delay. This indicates that the APE criterion may also be used to estimate input delay.

7 Conclusions

This paper presents three-fold results. We have developed a fast and parallel algorithm for on-line estimation of ARX system order and parameters. Compared with the algorithm in [9], the proposed algorithm reduces the computational complexity of order estimation significantly while preserving the strong consistency of order estimation. This algorithm is, however, limited to uniformly bounded data sequences.

We have invented a device attached to consistent order estimation criteria for solving the over-complexity problem. The device makes tradeoffs between model complexity (measured in terms of the number of model poles) and model accuracy (measured by means of order estimation criteria). As a result, a dominant model with a satisfactory accuracy, instead of a most accurate but over-complex model, is estimated. Simulation results show that the order of dominant models determined through order and parameter estimation rather precisely matches the order of conventional dominant models which are justified by means of Bode plots, time responses, and balanced realization, etc..

We have conducted a comprehensive simulation study to help understanding of order estimation. Many factors which are not studied in theoretical analysis have been carefully considered in simulation. These factors include stability, controllability/observability, and fast dynamics of unknown systems, as well as the feasibility of order estimation in practical situations. The simulation results show that the quasi-APE criterion — a modified version of the APE criterion has satisfactory performance in various situations.

References

[1] G. Long, D. Shwed and D. O. Falconer, "Study of a Pole-zero Adaptive Echo Canceller," *IEEE Trans. Circuits Systems*, CAS-34, no. 7, pp. 765–769, 1987.

[2] B. Egardt, "Stability of Adaptive Controllers," *Lecture Notes in Control and Information Science*, 20, 1979.

[3] C. E. Rohrs, L. Valavani, M. Athans and G. Stein, "Robustness of Adaptive Control Algorithms in the Presence of Unmodeled Dynamics," in *Proc. of the 21st IEEE CDC*. Florida: pp. 3–11, 1982.

[4] ———, "Robustness of Continus-time Adaptive Control Algorithms in the Presence of Unmodeled Dynamics," *IEEE Trans. AC*, AC-30, no. 9, pp. 881–889, September 1985.

[5] P. E. Caines, *Linear Stochastic Systems*. John Wiley & Sons, Inc., 1988.

[6] H. F. Chen and L. Guo, "Consistent Estimation of the Order of Stochastic Control Systems," *IEEE Trans. AC*, 32, no. 6, pp. 531–535, 1987.

[7] H. F. Chen and J. F. Zhang, "Identification and Adaptive Control for Systems with Unknown Orders, Delay, and Coefficients," *IEEE AC'*, 35, no. 8, pp. 866 – 877, 1990.

[8] E. M. Hemmerly and M. H. A. Davis, "Strong Consistency of the PLS Criterion for Order Determination of Autoregressive Processes," *Ann. Statist.*, 17, pp. 941–946, 1989.

[9] ———, "Recursive Order Estimation of Stochastic Control Systems," *Math. Systems Theory*, 22, pp. 323–346, 1989.

[10] A. A. Giordano and F. M. Hsu, *Least Square Estimation with Applications to Digital Signal Processing.* Prentice-Hall, Inc., 1985.

[11] M. G. Bellanger, "The FLS-QR Algorithm for Adaptive Filtering," *Signal Processing*, 17, pp. 291–304, 1989.

[12] B. Friedlander, Morf, T. Kailath and L. Ljung, "New Inversion Formulas for Matrices Classified in Terms of Their Distance from Toeplitz Matrices," *Linear Algebra and Its Applications*, 27, pp. 31–60, 1979.

[13] T. Kailath, S. Y. Kung and M. Morf, "Displacement Ranks of Matrices and Linear Equations," *J. Math Anay. App.*, 68, pp. 395–407, 1979.

[14] E. J. Hannan and J. Rissanen, "Recursive Estimation of Mixed Autoregressive-moving Average Order," *Biometrika*, 69, no. 1, pp. 81–94, 1982.

[15] J. Pan and W. S. Levine, "A Fast Time and Order Recursive Parameter Estimation Algorithm for a Set of ARX Models," in *Proc. of ACC.* Boston, MA: pp. 2832–2838, 1991.

[16] S. Karaboyas, N. Kalouptsidis and C. Caroubalos, "Efficient ARX Identification Algorithms with Full Parallelism," *IEEE Trans on ASSP*, 38, no. 11, pp. 1902–1913, Nov. 1990.

[17] T. L. Lai and C. Z. Wei, "Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems," *Ann. Statist.*, 10, pp. 154–166, 1982.

[18] H. F. Chen and L. Guo, "Convergence Rate of Least-Squares Identification And Adaptive Control for Stochastic Systems," *Int. J. Control*, 44, no. 5, pp. 1459–1476, 1986.

[19] M. Lundh, "Robust Adaptive Control," Department of Automatic Control, Lund Institute of Technology, Ph.D. Thesis, December, 1991.

[20] J. Pan, "Fast Recursive Estimation of System Order and Parameters for Adaptive Control and IIR Filtering," Electrical Engineering Department, University of Maryland at College Park, Ph.D. Thesis, August, 1991.

[21] B. D. O. Anderson, R. R. Bitmead, C. R. J. Jr., P. V. Kokotovic, R. L. Kosut, I. M. Mareels, L. Praly and B. D. Riedle, *Stability of Adaptive Systems.* Cambridge, Massachusetts, The MIT Press, 1986.

[22] S. Sastry and M. Bodson, *Adaptive Control — Stability, Convergence, and Robustness.* Prentice-Hall, Inc., 1989.

[23] L. Ljung, *System Identification Toolbox — for use with MATLAB.* The MathWorks, Inc., April 1988.

[24] L. Ljung, *System Identification: Theory for the User.* Prentice Hall, 1987.