

ON ADAPTIVE ESTIMATION AND POLE ASSIGNMENT OF OVERPARAMETRIZED SYSTEMS

LIGA XIA, J. B. MOORE AND M. GEVERS

*Department of Systems Engineering, Research School of Physical Sciences, Australian National University,
G.P.O. Box 4, Canberra, A.C.T., Australia*

SUMMARY

In this paper a first step is taken to avoid ill-conditioning in adaptive estimation and pole assignment schemes for the case when there is a signal model overparametrization. Such a situation can occur in practice when an unknown model order is guessed too high so as to be on the 'safe' side. The methods proposed in the paper are relatively simple compared with on-line order determination, being based on introducing suitable excitation in the 'regression' vectors of the parameter estimation algorithms to ensure parameter convergence. For the case when the models are non-unique in that pole-zero cancellations can occur, the algorithms seek to estimate the unique model where the cancellations occur at the origin. Applying estimates of this (unique) model turns out to avoid ill-conditioning in central tendency adaptive pole assignment. For the case of one pole-zero cancellation the convergence theory of the algorithm is complete. For the more general case algorithms are readily devised which appear to work well but for which a complete theory is not available.

KEY WORDS Adaptive control Adaptive pole assignment Overparametrization

1. INTRODUCTION

In the practice of adaptive estimation and control there is a tendency to overparametrize signal models (plants) to be on the 'safe' side. However, for overparametrized models there is a danger of ill-conditioning of both the adaptive estimation and the adaptive control algorithms applied to such plants. Of course, there is the twin danger of underparametrization, particularly in the absence of appropriate preprocessing of signals. The effects of underparametrization could be catastrophic and, since this is widely known, overparametrization emerges as a common problem. This paper shows that (after appropriate preprocessing) certain ill-conditioning associated with overparametrization can be avoided, without the need to perform on-line order determination with its associated significant increase in computational complexity.

For overparametrized signal models there can be a lack of excitation in regression vectors employed in parameter estimation and consequent ill-conditioning in the algorithms. Also, insufficient excitation can lead to identification of non-uniquely parametrized models which include pole-zero cancellations in the complex z -plane. When recursive estimates of the parameters of such non-uniquely parametrized models are applied for adaptive control, ill-conditioning leading to excessive controls can easily rise, particularly in adaptive pole assignment schemes.

Adaptive pole assignment schemes are perhaps the simplest schemes for adaptively stabilizing

This paper was recommended for publication by editor E. Mosca

linear plants which are possibly non-minimum phase.¹ Also, they are the most natural form of adaptive scheme to use in some applications where it is required that the adaptive scheme behave as closely as possible to a nominal optimal design. However, a severe limitation for their application in practice has been their failure when the signal models are overparametrized. Adaptive pole assignment requires the solution of a linear algebraic equation which becomes ill-conditioned when estimates of the plant have near pole-zero cancellations. This is inevitable when the signal model is overparametrized. Some authors have proposed methods to cope with this difficulty using on-line estimation of plant order in some sense.² Such an approach increases the complexity of the adaptive scheme considerably.

The first contribution of this paper, in Section 2, is to introduce excitation signals into the regression vectors for recursive (least-squares-based) parameter estimation in such a manner as to avoid ill-conditioning even when the model is overparametrized. For the special case when there is a potential non-uniqueness in the signal model owing to a pole-zero cancellation on the real axis, it is shown how the excitation can be designed so that the parameter estimates converge to those of a unique signal model, if one exists, otherwise to a model with a pole-zero cancellation at the origin. The introduced excitation does not excite the plant as in the case of added persistence of excitation signals. The estimation result of Section 2 is useful when applied in conjunction with the second contribution of the paper, in Section 3, which shows that when the parameter estimates converge so that the identified plant has a pole-zero cancellation at the origin, the associated central tendency adaptive pole assignment controller converges without ill-conditioning.

Section 4 gives a novel property of Sylvester matrices required in the proof of the theory of Section 3, and Section 5 gives an illustrative simulation study. Conclusions are drawn in Section 6.

2. ALGORITHMS AND RESULTS — WHITE NOISE CASE

Signal model

Consider the following single-input, single-output (SISO) input-output (stochastic) signal model class (plant) in terms of the unit delay operator q^{-1} , input u_k , output y_k and white noise disturbances w_k :

$$\begin{aligned} A(q^{-1})y_k &= B(q^{-1})u_k + w_k \\ A(q^{-1}) &= 1 + a_1q^{-1} + \dots + a_nq^{-n} \quad B(q^{-1}) = b_1q^{-1} + \dots + b_mq^{-m} \end{aligned} \quad (1)$$

This can be rewritten as

$$\begin{aligned} y_k &= \boldsymbol{\theta}^T \bar{\mathbf{x}}_k + w_k & \boldsymbol{\theta}^T &= [a_1 \ a_2 \ \dots \ a_n \ b_1 \ b_2 \ \dots \ b_m] \ \dots \\ \bar{\mathbf{x}}^T &= [-y_{k-1} \ \dots \ -y_{k-n} \ u_{k-1} \ \dots \ u_{k-m}] \end{aligned} \quad (1')$$

The conditions on w_k are more precisely:

$$\left. \begin{aligned} &\text{The sequence } \{w_k\} \text{ is independent of } u_k \text{ with } E[w_k | F_{k-1}] = 0, \\ &E[w_k^2 | F_{k-1}] \leq \sigma_w^2 < \infty, \text{ where } F_k \text{ denotes the } \sigma\text{-algebra generated by } w_1, w_2, \dots, w_k. \end{aligned} \right\} (2)$$

Recursive least-squares (RLS) estimation

Consider that $\boldsymbol{\theta}$ is estimated recursively by minimizing a least-squares criterion

$$\bar{J}_k = \frac{1}{k} \left(\sum_{i=1}^k (y_i - \boldsymbol{\theta}^T \bar{\mathbf{x}}_i)^2 + (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_0)^T \bar{B}_0 (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_0) \right)$$

via an RLS scheme as

$$\begin{aligned}\bar{\theta}_k &= \bar{P}_k [\bar{B}_{k-1} \bar{\theta}_{k-1} + \bar{x}_k y_k] & \bar{B}_k &= \bar{B}_{k-1} + \bar{x}_k \bar{x}_k^T \\ \bar{P}_k &= \bar{P}_{k-1} - \bar{P}_{k-1} \bar{x}_k \bar{x}_k^T \bar{P}_{k-1} (1 + \bar{x}_k^T \bar{P}_{k-1} \bar{x}_k)^{-1} = \bar{B}_k^{-1}\end{aligned}\quad (3)$$

for some initial conditions $\bar{\theta}_0, \bar{B}_0 > 0$.

Convergence properties review

To achieve a simple analysis making connection with Kalman filter theory as in Reference 3, let us assume:

$$\begin{aligned}\{w_k\} &\text{ is normally distributed and the } a \text{ priori probability density associated with} \\ \theta &\text{ is } N[\bar{\theta}_0, \bar{P}_0] \text{ for some } \bar{P}_0 = \bar{B}_0^{-1} > 0.\end{aligned}\quad (4)$$

This assumption is not needed for a more general theory based on stochastic Lyapunov functions $\bar{\theta}_k^T \bar{B}_k \bar{\theta}_k$ in References 4 and 5, where $\bar{\theta}_k = \theta - \hat{\theta}_k$, but then the results are not quite as tidy.

For the model (1), (2) and the RLS scheme (3) under (4), Kalman filter theory tells us that, with $\bar{\theta}_k = \theta - \hat{\theta}_k$,

$$\sigma^2 \bar{P}_k = E[\bar{\theta}_k \bar{\theta}_k^T | F_{k-1}] \quad \bar{\theta}_k = E[\theta | F_{k-1}] \quad (5)$$

Moreover,³ there is almost sure convergence as

$$\lim_{k \rightarrow \infty} \bar{P}_k = \bar{P}^{LS} \quad \lim_{k \rightarrow \infty} \bar{\theta}_k = \bar{\theta}^{LS} \quad \text{a.s.} \quad (6)$$

for random variables $\bar{P}^{LS}, \bar{\theta}^{LS}$. With \bar{x}_k sufficiently exciting in that $\bar{P}^{LS} = 0$, then Reference 3 tells us that $\bar{\theta}^{LS} = \theta$. Also if $\bar{\theta}^{LS} = \theta$, then $\bar{\theta}_k \bar{\theta}_k^T \rightarrow 0$ as $k \rightarrow \infty$ and consequently, under (5), $\bar{P}_k \rightarrow 0$ as $k \rightarrow \infty$. Thus we have the following strong connection between sufficiency of excitation of \bar{x}_k and parameter convergence.

Lemma 1. For the RLS scheme (3) applied to the signal model (1), (2) under (4), then

$$\lim_{k \rightarrow \infty} \bar{P}_k = 0 \quad \Leftrightarrow \quad \lim_{k \rightarrow \infty} \bar{\theta}_k = \theta \quad (7)$$

Proof. The proof is as above based on results in Reference 3.

On sufficient excitation

In this subsection three specific excitation scenarios are studied using known results from References 4–6. These relate excitation of signal model inputs to outputs or states for reachable open-loop time-invariant plants, with or without (possibly time-varying) feedback. The first two cases are a review of known results for the case when there is no overparametrization, while the third case deals with the case of overparametrized models.

Case (i). The simplest case to study is the non-overparametrized case when

$$q^n A(q^{-1}), q^m B(q^{-1}) \text{ are coprime} \quad (8)$$

and u_k is suitably exciting in that

$$\lim_{k \rightarrow \infty} \left(\sum_1^k \bar{\mathbf{u}}_k \bar{\mathbf{u}}_k^T \right)^{-1} = 0 \quad \bar{\mathbf{u}}_k^T = [u_{k-1} \ u_{k-2} \ \dots \ u_{k-n-m}] \quad (9)$$

This latter condition is achieved when u_k includes at least $n + m/2$ distinct sinusoids decaying no faster than $1/k$, as when u_k is white or filtered white noise with a variance decaying no faster than $1/k$.

Under (8), the model (1) is uniquely parametrized, $\bar{\mathbf{x}}_k$ is reachable from u_k alone and excitation of the inputs as in (9) implies excitation of the states $\bar{\mathbf{x}}_k$.^{4,5} Thus (8), (9) ensure that $\bar{P}_k \rightarrow 0$ as $k \rightarrow \infty$ and in turn $\bar{\boldsymbol{\theta}}_k \rightarrow \boldsymbol{\theta}$ as $k \rightarrow \infty$. For this case then, w_k need not be sufficiently exciting in any sense.

Case (ii). Another simple case to study is when (8) does not necessarily hold, but w_k as well as u_k is suitably exciting in that (9) is satisfied and

$$\left(\sum_1^\infty E\{w_k^2 | F_{k-1}\} \right)^{-1} = 0 \quad (10)$$

Under (9), (10), the model (1) is uniquely parametrized, $\bar{\mathbf{x}}_k$ is reachable from u_k , w_k and is sufficiently exciting to guarantee that $\bar{P}_k \rightarrow 0$ as $k \rightarrow \infty$ ^{4,5} and in turn that $\bar{\boldsymbol{\theta}}_k \rightarrow \boldsymbol{\theta}$ as $k \rightarrow \infty$. For this case then, the convergence as such is independent of whether or not the coprimeness condition (8) is satisfied.

Case (iii). The possibility overparametrized signal model situation of particular interest in this paper is when (8) possibly fails and there is no *a priori* guarantee of sufficient excitation of w_k as in (10). In this case the model (1) may not be uniquely parametrized, having one or more pole-zero cancellations in the z -plane. Also, $\bar{\mathbf{x}}_k$ may not be sufficiently exciting to ensure that $\bar{P}_k \rightarrow 0$ as $k \rightarrow \infty$. Convergence can take place to a signal model with pole-zero cancellation anywhere in the complex z -plane. We seek to avoid such a situation and propose an RLS algorithm with additional excitation in the regression vector. It is derived using an alternative signal model formulation.

Alternative signal model formulation

Consider (1) reorganized as

$$y_k = \boldsymbol{\theta}^T \mathbf{x}_k + (w_k - \boldsymbol{\theta}^T \mathbf{v}_k) \quad \mathbf{x}_k = \bar{\mathbf{x}}_k + \mathbf{v}_k \quad (11)$$

where \mathbf{v}_k is an excitation term to ensure that \mathbf{x}_k is suitably exciting. Notice that \mathbf{v}_k has no influence on y_k , u_k . Its selection in the next subsection is in accordance with a parameter estimation error measure, so that when parameter estimates are converging to their true values, \mathbf{v}_k converges to zero.

RLS estimation with regression vector excitation

Consider that $\boldsymbol{\theta}$ is estimated recursively by minimizing a least-squares criterion

$$J_k = \frac{1}{k} \left(\sum_{i=1}^k (y_i - \boldsymbol{\theta}^T \mathbf{x}_i)^2 + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T B_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \right) \quad (12)$$

via an RLS scheme as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_k &= P_k [B_{k-1} \hat{\boldsymbol{\theta}}_{k-1} + \mathbf{x}_k y_k] & B_k &= B_{k-1} + \mathbf{x}_k \mathbf{x}_k^\top \\ P_k &= P_{k-1} - P_{k-1} \mathbf{x}_k \mathbf{x}_k^\top P_{k-1} (1 + \mathbf{x}_k^\top P_{k-1} \mathbf{x}_k)^{-1} = B_k^{-1}\end{aligned}\quad (13)$$

for some initial conditions $\boldsymbol{\theta}_0, B_0 > 0$.

Regression vector excitation selection

Consider the signal model (1) formulated as (11). Let us assume that either there is no overparametrization in that (8) holds, or that there is the possibility of overparametrization which includes one pole-zero cancellation, leading to a non-uniquely parametrized model. In the latter case we consider a unique parametrization with the properties (meaningful only when $n > 1, m > 1$)

$$a_n = b_m = 0 \quad q^{n-1}A(q^{-1}), q^{m-1}B(q^{-1}) \text{ are coprime} \quad (14)$$

For such a situation we propose a \mathbf{v}_k selection as follows.

$$\left. \begin{array}{l} \text{The sequence } \{\mathbf{v}_k\} \text{ is selected as an independent (Gaussian) zero mean white noise} \\ \text{excitation term such that its covariance } \Sigma_{\mathbf{v}}^2 = Dk^{-1} \text{tr}(P_{k-1}) \text{ for} \\ D = \text{diag}[0 \ 0 \ \dots \ 0 \ d_n^2 \ 0 \ \dots \ 0 \ d_{n+m}^2] \text{ where } d_n^2 < 0, d_{n+m}^2 > 0. \text{ Denote the} \\ \text{non-zero elements as } \sigma_{n,k}^2, \sigma_{n+m,k}^2 \end{array} \right\} \quad (15)$$

The Gaussian assumption on \mathbf{v}_k is to keep the analysis simple and is not a necessary condition. In practice a more efficient excitation would be where elements had values in a bounded domain. Also, results are readily derived for the case when \mathbf{v}_k is deterministic but containing a sufficient number of frequency components.

The following results are now a consequence of a straightforward application of results from References 4 and 5.

Lemma 2. Consider the linear signal model (1), (2) formulated as (11) with the \mathbf{v}_k selection of (15).

(i) Then, under (8), \mathbf{x}_k is the output of a linear time-invariant system reachable from u_k . Moreover, with u_k selected so that for some $\alpha > 0$

$$\lim_{k \rightarrow \infty} (\ln k)^{1+\alpha} \left(\sum_1^k \bar{\mathbf{u}}_i \bar{\mathbf{u}}_i^\top \right)^{-1} = 0 \quad (16)$$

then

$$\lim_{k \rightarrow \infty} (\ln k)^{1+\alpha} P_k = 0 \quad \lim_{k \rightarrow \infty} (\ln k)^{1+\alpha} \bar{P}_k = 0 \quad \text{a.s.} \quad (17a,b)$$

(ii) Also, under (14), \mathbf{x}_k is the output of a linear time-invariant system reachable from u_k and the elements $v_{n,k}, v_{n+m,k}$ of \mathbf{v}_k . Then, under (14), the excitation conditions on u_k of (4) and \mathbf{v}_k , namely,

$$\lim_{k \rightarrow \infty} \left(\sum_1^k \sigma_{n,k}^2 \right)^{-1} = \lim_{k \rightarrow \infty} \left(\sum_1^k \sigma_{n+m,k}^2 \right)^{-1} = 0 \quad (18)$$

translate to excitation of \mathbf{x}_k as

$$\lim_{k \rightarrow \infty} P_k = 0 \quad \text{a.s.} \quad (19)$$

Proof

(i) Under the coprimeness condition (8), it is immediate⁵ that \mathbf{x}_k is the output of a linear time-invariant system driven by u_k , \mathbf{v}_k , w_k and is reachable from u_k alone. Now should u_k be acting alone, then Lemma (3.2) of Reference 4 applies to give that, for all k and some $K > 0$,

$$\sum_1^{k+m+n} \mathbf{x}_i \mathbf{x}_i^T \geq K \sum_1^k \bar{\mathbf{u}}_i \bar{\mathbf{u}}_i^T$$

from which (for some $\alpha > 0$)

$$\begin{aligned} \lim_{k \rightarrow \infty} \left(\sum_1^k \bar{\mathbf{u}}_i \bar{\mathbf{u}}_i^T \right)^{-1} = 0 &\Rightarrow \lim_{k \rightarrow \infty} \left(\sum_1^k \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} = 0 \\ \lim_{k \rightarrow \infty} (\ln k)^{1+\alpha} \left(\sum_1^k \bar{\mathbf{u}}_i \bar{\mathbf{u}}_i^T \right)^{-1} = 0 &\Rightarrow \lim_{k \rightarrow \infty} (\ln k)^{1+\alpha} P_k = 0 \end{aligned}$$

Applying the results of Lemma (3.3) and its Remark 1 of Reference 4 now tells us that when bounded variance white noise inputs \mathbf{v}_k , w_k (independent of u_k) are also applied, the same implications hold, so that (16) implies (17a) as claimed. The result (17b) holds likewise.

(ii) Under (14), the elements $y_{k-1} \dots y_{k-n-1}$, $u_{k-1} \dots u_{k-m-1}$ of \mathbf{x}_k are reachable from u_k alone⁵ and the remaining elements $(y_{k-n} + v_{n,k})$, $(u_{k-m} + v_{n+m,k})$ of \mathbf{x}_k are reachable from the non-zero elements of \mathbf{v}_k , namely $v_{n,k}$ and $v_{n+m,k}$. Thus \mathbf{x}_k is reachable from u_k , \mathbf{v}_k under (14). Applying again the Lemma (3.3) and its Remark 1 of Reference 4 gives directly that (4), (18) together imply (19).

Main results of section

Theorem 1. Consider the signal model (1), (2) which is possibly overparametrized in that either (8) or (14) holds. Consider an RLS scheme (13) based on the alternative model formulation (11) with \mathbf{v}_k selected as in (15) and (4) holding. Consider also that u_k is sufficiently exciting in that (16) holds. Then there is parameter convergence as

$$\lim_{k \rightarrow \infty} \hat{\boldsymbol{\theta}}_k = \boldsymbol{\theta} \quad \text{a.s.} \quad (20)$$

where $\boldsymbol{\theta}$ is the unique parameter associated with (1) under (8) or (14). Moreover (19) also holds.

Proof

Part (i). In the case that (14) is satisfied, so that $a_n = b_m = 0$, then $\boldsymbol{\theta}^T \mathbf{v}_k = 0$. Now Lemma (1) applies with \mathbf{x}_k replacing $\bar{\mathbf{x}}_k$, so that (20) holds if and only if (19) holds. Assume that (19) does not hold: then since $P_k \leq P_{k-1}$ for all k , $\text{tr}(P_k)$ converges to a non-zero element and from (15) the variances of \mathbf{v}_k decay at a rate k^{-1} . Thus (18) holds and in turn (19) holds under Lemma 2. This contradicts the assumption, so that (19) and (20) hold.

Part (ii). In the case that (8) is satisfied, Lemma 2 tells us that (16) implies (17). As a consequence from (15) then

$$\lim_{k \rightarrow \infty} \sum_1^k \sigma_{n,k}^2 < \infty \quad \lim_{k \rightarrow \infty} \sum_1^k \sigma_{n+m,k}^2 < \infty$$

In turn we claim that

$$\lim_{k \rightarrow \infty} \sum_1^k (\boldsymbol{\theta}^T \mathbf{v}_k)^2 < \infty \quad \text{a.s.} \quad (21)$$

This follows since, as is readily established under (15),

$$\sum_1^k (v_{k,n}^2 - \sigma_{n,k}^2) \quad \sum_1^k (v_{n+m,k}^2 - \sigma_{n+m,k}^2)$$

are martingales bounded in L_2 and converge almost surely. Under (21) and (17b), we now claim

$$\lim_{k \rightarrow \infty} \bar{P}_k \sum_1^k w_i (\mathbf{v}_i^\top \boldsymbol{\theta}) = 0 \quad \text{a.s.} \quad (22)$$

To see this observe that

$$E[M_k | F_{k-1}] = M_{k-1} \quad M_k = \sum_1^k w_i (\mathbf{v}_i^\top \boldsymbol{\theta})$$

$$E[M_k^2] = E\left[\sum_{i=1}^k \sum_{j=1}^k \boldsymbol{\theta}^\top \mathbf{v}_i E[w_i w_j | F_{\min\{i-1, j-1\}}] \mathbf{v}_j^\top \boldsymbol{\theta}\right] \leq E\left[\sum_1^k (\boldsymbol{\theta}^\top \mathbf{v}_i)^2\right] \sigma_w^2$$

Thus M_k is a martingale on F_{k-1} , bounded in L_2 under (21), and so converges almost surely, so that (17b) implies (22).

Now, under (17b), Lemma (1) can be applied to yield

$$\lim_{k \rightarrow \infty} \bar{P}_k \sum_1^k \bar{x}_i w_i = 0 \quad \lim_{k \rightarrow \infty} \bar{P}_k \sum_1^k \bar{x}_i (\mathbf{v}_i^\top \boldsymbol{\theta}) = 0 \quad \text{a.s.} \quad (23)$$

The first result follows from (7) and the relationships

$$\bar{\boldsymbol{\theta}}_k = \bar{P}_k \sum_1^k \bar{x}_i y_i \quad \bar{\boldsymbol{\theta}}_k = -\bar{P}_k \left(\sum_1^k \bar{x}_i w_i - B_0 \right)$$

The second result follows from the first since $(\mathbf{v}_i^\top \boldsymbol{\theta})$ has the same essential properties as w_i in (2), (4).

The results (22), (23) lead in turn to the following convergence results:

$$\bar{P}_k B_k = \bar{P}_k \sum_1^k (\bar{x}_i \bar{x}_i^\top + \bar{x}_i \mathbf{v}_i^\top + \mathbf{v}_i \bar{x}_i^\top + \mathbf{v}_i \mathbf{v}_i^\top) \rightarrow \mathbf{I} \quad \text{a.s. as } k \rightarrow \infty$$

$$\hat{\boldsymbol{\theta}}_k = P_k \sum_1^k \mathbf{x}_i y_i = (\bar{P}_k B_k)^{-1} \bar{P}_k \sum_1^k (\bar{x}_i \bar{x}_i^\top \boldsymbol{\theta} + \boldsymbol{\theta}^\top \mathbf{v}_i \bar{x}_i^\top \boldsymbol{\theta} + \bar{x}_i w_i + \boldsymbol{\theta}^\top \mathbf{v}_i w_i) \rightarrow \boldsymbol{\theta} \quad \text{a.s. as } k \rightarrow \infty$$

so that (20) holds as claimed.

Remarks

1. The specific \mathbf{v}_k selection of the theorem is for the case when there is one possible pole-zero cancellation in the model. This is clearly one of the most important cases, since in selecting a model order there is a tendency when in doubt to merely increase a likely order by one for safety.
2. If a bank of estimators is employed conditioned on different model orders, then the results above tell us that only odd (or even) orders need to be covered. Such a saving is a factor of two.
3. Rather than work with banks of estimators as in Remark 2 above, an *ad hoc* approach is to relax (15) and have $v_{k,1}, v_{k,2}, \dots, v_{k,n}$ each independent and suitably exciting with the variance of $v_{k,i}$ increasing with i . This would force pole-zero cancellations to occur near the origin but would lead to biased estimates. We do not study this technique.

4. The algorithms and results of this section have been analysed for the simplest of stochastic models, namely, when $y_k - \theta^T x_k$ is white. For more general autoregressive moving-average exogenous input (ARMAX) models, extended least-squares (ELS) based algorithms can be employed. We claim that the technique of introducing v_k also can be made to extend the capability of ELS-based schemes. In particular, for those which are globally convergent under the coprimeness condition (9),⁷ with the modifications they are globally convergent also when there is a possible overparametrization, as when (14) holds. Essentially the same theoretical approach applies, but the technical details are more tedious so are not explored here.

3. ADAPTIVE POLE ASSIGNMENT

Pole assignment

Let us seek an adaptive pole assignment scheme associated with the signal model (1) so that there is asymptotic convergence to

$$H(q^{-1})y_k = KB(q^{-1})r_k \quad (24)$$

where $H(q^{-1}) = 1 + h_1q^{-1} + \dots + h_{n+m}q^{-n-m}$ is specified by the desired closed-loop poles, K is a constant and r_k is a reference input. This can be achieved by the following controller:²

$$H(q^{-1})E(q^{-1})u_k = -H(q^{-1})F(q^{-1})y_k + Kr_k \quad (25)$$

where $E(q^{-1}) = 1 + e_1q^{-1} + \dots + e_mq^{-m}$, $F(q^{-1}) = f_1q^{-1} + \dots + f_nq^{-n}$ are given from the solution of the Bezout equation

$$A(q^{-1})E(q^{-1}) + B(q^{-1})F(q^{-1}) = 1 \quad \text{or} \quad \mathbf{S}(n, m)\boldsymbol{\varphi} = \boldsymbol{\alpha} \quad (26a, b)$$

where

$$\boldsymbol{\varphi}^T = [e_1 \ e_2 \ \dots \ e_m \ f_1 \ f_2 \ \dots \ f_n] \quad \boldsymbol{\alpha}^T = [-a_1 \ -a_2 \ \dots \ a_n \ 0 \ \dots \ 0]$$

$$\mathbf{S}(n, m) = \begin{array}{c} \left[\begin{array}{cccccccc} 1 & 0 & \cdot & \cdot & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ a_1 & 1 & & & 0 & b_1 & & & & \cdot \\ \cdot & a_1 & \cdot & & \cdot & \cdot & b_1 & & & \cdot \\ \cdot & & & & \cdot & \cdot & \cdot & & & \cdot \\ a_n & & & & 1 & b_m & & \cdot & \cdot & \cdot \\ 0 & \cdot & & & \cdot & 0 & \cdot & & & b_1 \\ \cdot & & \cdot & & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & & & & \cdot & \cdot & \cdot & & & \cdot \\ 0 & \cdot & \cdot & 0 & a_n & 0 & \cdot & \cdot & & b_m \end{array} \right] \begin{array}{l} \uparrow \\ \\ \\ \\ \\ \downarrow \end{array} \\ \leftarrow \begin{array}{cc} m & n \end{array} \rightarrow \end{array} \quad (27)$$

It is known that solutions of (26) exist if and only if $\text{rank}[\mathbf{S} \ \boldsymbol{\alpha}] = \text{rank}[\mathbf{S}]$. Also,⁸

$$\mathbf{S}(n, m) \text{ is non-singular} \Leftrightarrow q^n A(q^{-1}), q^m B(q^{-1}) \text{ are coprime} \quad (28)$$

In certainty equivalence adaptive pole assignment, the estimates $\hat{\theta}_k$ are used in lieu of θ in (27) to compute on-line estimates $\hat{\boldsymbol{\varphi}}_k^{\text{CE}}$ of the controller parameters, so that in obvious notation

$$\hat{\mathbf{S}}_k^{\text{CE}}(n, m)\hat{\boldsymbol{\varphi}}_k^{\text{CE}} = \hat{\boldsymbol{\alpha}}_k^{\text{CE}} \quad (29)$$

When $q^n A(q^{-1}), q^m B(q^{-1})$ are coprime, $\hat{\mathbf{S}}_k(n, m)$ is non-singular and the solution of (29)

exists. Otherwise it may not. Ill-conditioning in $\hat{S}_k^{-1}(n, m)$ can cause excessive values for $\hat{\varphi}_k$. One modification to avoid large $\hat{\varphi}_k$ is to select $\hat{\varphi}_k = \hat{\varphi}_{k-1}$ during ill-conditioning, but for unknown plants it is not *a priori* clear how to quantify ill-conditioning to achieve a useful adaptive controller.

Central tendency adaptive control

In central tendency adaptive pole assignment,⁹ ill-conditioning in calculating the controller parameters is avoided without requiring prior information concerning the plant or controller. Suppose there is a Gaussian *a posteriori* probability density for the model parameters θ as $N[\hat{\theta}_k, \hat{\sigma}_k^2 P_k]$ where $\hat{\sigma}_k^2$ is an estimate of σ^2 ; then there is an associated non-Gaussian probability density for the pole assignment controller parameters φ . A central tendency selection $\hat{\varphi}_k^{\text{CT}}$ is one which maximizes this density or at least avoids the tails of this density. Practical implementations are given in Reference 9. Associated with $\hat{\varphi}_k^{\text{CT}}$ is some parameter estimate $\hat{\theta}_k^{\text{CT}}$ which is not in general $\hat{\theta}_k$. Thus in obvious notation

$$\hat{S}_k^{\text{CT}}(n, m)\hat{\varphi}_k^{\text{CT}} = \hat{\alpha}_k^{\text{CT}} \quad (30)$$

The estimate $\hat{\theta}_k^{\text{CT}}$ has the property⁹ that it is 'close' to $\hat{\theta}_k$ but 'far' from hypersurfaces for which $q^n \hat{A}_k(q^{-1}), q^m \hat{B}_k(q^{-1})$ are not coprime. As a consequence, the following property is claimed for central tendency adaptive pole assignment.

The selections $\hat{\theta}_k^{\text{CT}} \rightarrow \theta$ are such that $[\hat{S}_k^{\text{CT}}(n, m)]^{-1}$ exists for all k , and if θ belongs to the hypersurface (in θ -space) defined by $q^n A(q^{-1}), q^m B(q^{-1})$ not coprime, then, as $k \rightarrow \infty$, $\hat{\theta}_k^{\text{CT}}$ is contained in a cone centred at θ which excludes the tangent hyperplane at θ . } (31)

Remarks

1. To give a geometric interpretation of (31), consider Figures 1 and 2. The heavy arc is the pole-zero cancellation singular region for θ estimates in Θ -space. The light shaded area is a zone of all-conditioning control surrounding the singular arc. The heavy shaded area cones are the conic regions of possible central tendency control estimates $\hat{\theta}_k^{\text{CT}}$ of (31). Figure 1 depicts the situation when the solution (14) is at the conic intersection and the cones avoid the ill-conditioned regions. Figure 2 depicts a solution not satisfying (14) when the central tendency estimates $\hat{\theta}_k^{\text{CT}}$ become ill-conditioned when converging.
2. The property (31) can be viewed as a corollary of results rigorously proved in Reference

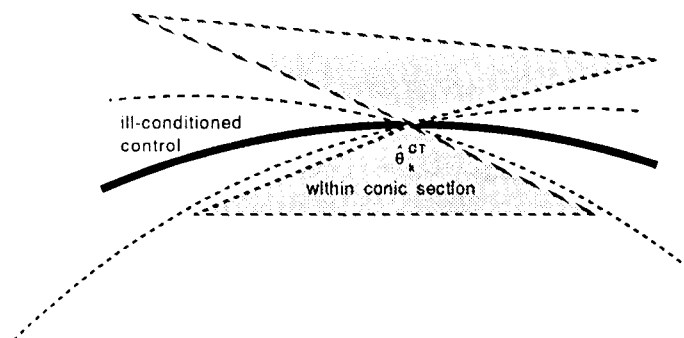


Figure 1. Avoidance of ill-conditioning

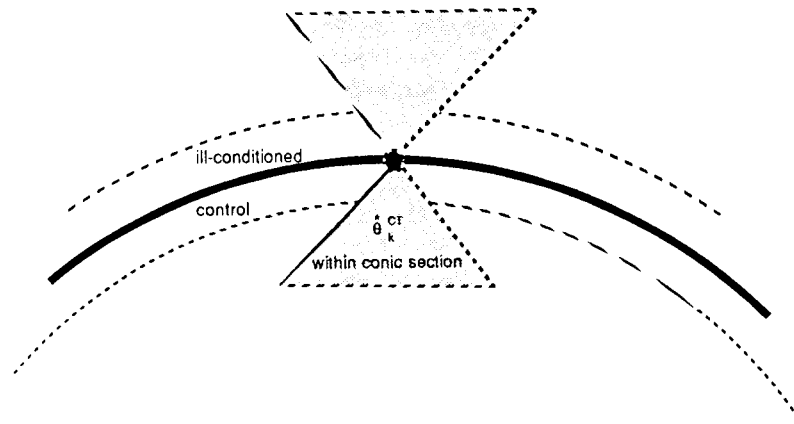


Figure 2. Ill-conditioned control

9, although specific reference has not been made to (31) there. The property is readily believable, but since it is not rigorously proved as such in Reference 9 and is beyond the scope of this paper, we here add the ‘qualification’ to central tendency control that (31) be satisfied.

Two cases are studied now.

Case (i). Known model order. We now claim the following:

Theorem 2. Consider the signal model (1) for the case when it is not overparametrized, so that $q^n A(q^{-1})$, $q^m B(q^{-1})$ are coprime. Consider also RLS parameter estimates $\hat{\theta}_k$ from (13) and associated certainty equivalence (or central tendency) pole assignment with controller parameters $\hat{\varphi}_k^{CE}$ (or $\hat{\varphi}_k^{CT}$) and Sylvester matrices $\hat{S}_k^{CE}(n, m)$ (or $\hat{S}_k^{CT}(n, m)$). Consider also that r_k is sufficiently exciting so that $\hat{\sigma}_k^2 P_k \rightarrow 0$ as $k \rightarrow \infty$ and there is parameter convergence with $\hat{\theta}_k \rightarrow \theta$ as $k \rightarrow \infty$. Then

$$\begin{aligned} \lim_{k \rightarrow \infty} [\hat{S}_k^{CE}(n, m)]^{-1} = S^{-1}(n, m) & \quad \text{or} & \quad \lim_{k \rightarrow \infty} [\hat{S}_k^{CT}(n, m)]^{-1} = S^{-1}(n, m) \\ \lim_{k \rightarrow \infty} \hat{\varphi}_k^{CE} = \varphi & \quad \text{or} & \quad \lim_{k \rightarrow \infty} \hat{\varphi}_k^{CT} = \varphi \end{aligned} \quad (32)$$

Proof. This is immediate from the coprimeness assumption and the property (28). Notice that there is no need of assumption (31).

Case (ii). Overparametrization. We now claim the following:

Theorem 3. Consider the signal model (1) with $n > 1$, $m > 1$ and (14) holding. Consider also RLS parameter estimates $\hat{\theta}_k$ from (13) and associated central tendency adaptive pole assignment controller parameters $\hat{\varphi}_k^{CT}$ given from (30) with (31) satisfied. Then, under RLS convergence of $\hat{\theta}_k^{CT}$ to θ as $k \rightarrow \infty$,

$$\lim_{k \rightarrow \infty} \det \hat{S}_k^{CT}(n, m) = 0 \quad \lim_{k \rightarrow \infty} \hat{\varphi}_k^{CT} = \varphi^* \quad (33)$$

where φ^* is a unique solution of $\mathbf{S}(n, m)\varphi^* = \alpha$ with zero elements as

$$\varphi^* = [e_1^* \ e_2^* \ \dots \ e_{m-1}^* \ 0 \ f_1^* \ f_2^* \ \dots \ f_{n-1}^* \ 0]^T \quad (34)$$

Proof. The proof is given in the Appendix, based on the Sylvester matrix property studied in Section 4.

Remarks

1. It might be thought that the results can be more directly proved from properties of Diophantine equations. Although certain progress can be made along these lines, and indeed the results can be stated in such terms, it does not appear straightforward to complete any proof without resort to Sylvester matrix properties as in Section 4.
2. This theorem result is dependent on the non-standard nature of the RLS algorithm with its internal perturbations \mathbf{v}_k . In the presence of overparametrization, standard RLS estimation (when $\mathbf{v}_k = \mathbf{0}$) will almost surely not converge to the unique limits solution (14). This means that there is inevitable ill-conditioning. This situation applies even when a central tendency adaptive control law is implemented, as depicted in Figure 2.
3. The above theorem result is also facilitated by the central tendency property (31) of the adaptive controller. In the presence of overparametrization, even when the modified RLS algorithm of the paper is implemented without (31) satisfied, there is a non-zero probability of ill-conditioning as suggested from Figure 1, at least during transients.
4. The theorem is developed in conjunction with the RLS estimation (4), which copes with possible overparametrization by the a_n, b_m . Should some RLS-based scheme cope with higher orders of overparametrization, the results of this theorem would still hold. The details of a more general proof are straightforward and are omitted here.
5. The result of the theorem also applies to the situation when (26) is replaced by

$$A(q^{-1})E(q^{-1}) + B(q^{-1})F(q^{-1}) = H(q^{-1}) \quad (35)$$

with $H(q^{-1})$ having degree no greater than $n + m - 2$. This is the usual situation when $A(q^{-1})$ is of degree $n - 1$ and $B(q^{-1})$ is of degree $m - 1$. *The proof details are a mild variation on that given here when (26) applies. The restriction on $H(q^{-1})$ implies that the associated α in the algebraic form of (35) corresponding to (26b) has its last two entries zero.*

6. The above result covers the cases when the plant has a delay of unity or greater. In these cases $B(q^{-1})$ specializes as having a factor q^{-N} where $N \geq 1$.

4. A PROPERTY OF SYLVESTER MATRICES

First recall (27) which associates with $\theta = [a_1 \ \dots \ a_n \ b_1 \ \dots \ b_m]^T$ a Sylvester matrix $\mathbf{S}(n, m)$. Let us denote the adjoint of $\mathbf{S}(n, m)$ as $\mathbf{M}(n, m)$ and determinant as $D(n, m)$. Now consider a (linear) trajectory in θ -space parametrized in terms of a scalar variable ξ as

$$\theta(\xi) = [a_1 \ \dots \ a_{n-1}(a\xi) \ b_1 \ \dots \ b_{m-1}(b\xi)]^T \quad |a| + |b| = 1 \quad (36)$$

Also denote the Sylvester matrix associated with $\theta(\xi)$ as $\mathbf{S}_\xi(a_n, b_m)$, its adjoint matrix as $\mathbf{M}_\xi(a_n, b_m)$ and its determinant as $D_\xi(a_n, b_m)$. Then we claim the following:

Lemma 3. Consider $\mathbf{S}_\xi^{-1}(a_n, b_m) = D_\xi^{-1}(a_n, b_m)\mathbf{M}_\xi(a_n, b_m)$ for the case that

$$D(n-1, m-1) \neq 0 \quad ab_{m-1} - a_{n-1}b \neq 0 \quad (37)$$

Then the following limits exist as $\xi \rightarrow 0$ for all $j < n + m - 1$:

$$\lim_{\xi \rightarrow 0} D_\xi^{-1}(a_n, b_m) [M_\xi(a_n, b_m)]_{i,j} = \begin{cases} [S^{-1}(n-1, m-1)]_{i,j} & \text{for } i < m \\ [S^{-1}(n-1, m-1)]_{i-1,j} & m < i < m+n \\ 0 & \text{for } i = m \text{ or } i = m+n \end{cases} \quad \begin{array}{l} (38a) \\ (38b) \\ (38c) \end{array}$$

Proof. We consider in turn expressions for $D_\xi(a_n, b_m)$ and $\mathbf{M}_\xi(a_n, b_m)$ in terms of order ξ and higher-order terms denoted $O(\xi^2)$. Simple manipulations give an expression for $D_\xi(a_n, b_m)$ in terms of the elements of the last row of $\mathbf{S}_\xi(a_n, b_m)$ and their minors, with the minors likewise expanded, as

$$D_\xi(a_n, b_m) = (-1)^n \xi (ab_{m-1} - a_{n-1}b) D(n-1, m-1) + O(\xi^2) \quad (39)$$

Noting that $[\mathbf{M}_\xi(a_n, b_m)]_{i,j}$ is a (signed) determinant of a submatrix of $\mathbf{S}_\xi(a_n, b_m)$, then again simple manipulations give an expression in terms of the elements of the last row of the submatrix of $\mathbf{S}_\xi(a_n, b_m)$ and minors of this submatrix as

$$[\mathbf{M}_\xi(a_n, b_m)]_{i,j} = (-1)^n (a\xi) \mathbf{M}_\xi(a_n, m-1)_{i,j} + (b\xi) [\mathbf{M}_\xi(a_n, b_m)]_{i,j} + O(\xi^2) \quad \text{for } i < m \quad j < n + m - 1 \quad (40)$$

Here $\mathbf{M}_\xi(a_n, m-1)$ denotes the adjoint matrix of the lower-dimensional Sylvester matrix $\mathbf{S}_\xi(a_n, m-1)$ which associates with the vector

$$[I_{n+m-1} \ 0] \boldsymbol{\theta}(\xi) = [a_1 \ \dots \ a_{n-1} \ (a\xi) \ b_1 \ \dots \ b_{m-1}]^T$$

Using derivations similar to those giving (40), then

$$\begin{aligned} [\mathbf{M}_\xi(a_n, m-1)]_{i,j} &= b_{m-1} [\mathbf{M}(n-1, m-1)]_{i,j} + O(\xi) \\ &= b_{m-1} D(n-1, m-1) [S^{-1}(n-1, m-1)]_{i,j} + O(\xi) \quad \text{for } i < m \quad j < n + m - 1 \end{aligned} \quad (41)$$

where $S^{-1}(n-1, m-1)$ exists under the assumption (37). The dual form of (41) is

$$[\mathbf{M}_\xi(n-1, b_m)]_{i,j} = (-1)^{n-1} a_{n-1} D(n-1, m-1) [S^{-1}(n-1, m-1)]_{i,j} + O(\xi) \quad \text{for } i < m \quad j < n + m - 1 \quad (42)$$

Substitution of (41), (42) into (40) yields

$$[\mathbf{M}_\xi(a_n, b_m)]_{i,j} = (-1)^n (ab_{m-1} - a_{n-1}b) \xi D(n-1, m-1) [S^{-1}(n-1, m-1)]_{i,j} + O(\xi^2) \quad \text{for } i < m \quad j < n + m - 1 \quad (43)$$

Dividing (43) by $D_\xi(a_n, b_m)$ from (39) and taking limits as $\xi \rightarrow 0$, under the assumption (37), leads to the result (38a). The result (38b) can be established along similar lines.

Similar arguments to the derivation of (40), (41) lead to

$$[\mathbf{M}_\xi(a_n, b_m)]_{m,j} = (b\xi) [\mathbf{M}_\xi(n-1, b_m)]_{m,j} + O(\xi^2) \quad \text{for } j < n + m - 1 \quad (44)$$

$$[\mathbf{M}_\xi(n-1, b_m)]_{m,j} = (b\xi) O(\boldsymbol{\theta}) \quad \text{for } j < n + m - 1 \quad (45)$$

where $O(\boldsymbol{\theta})$ denotes a quantity that is bounded in terms of $\boldsymbol{\theta}$. From (44), (45)

$$[\mathbf{M}_\xi(a_n, b_m)]_{m,j} = O(\xi^2) \quad \text{for } j < n + m - 1 \quad (46)$$

Similarly we have

$$[\mathbf{M}_\xi(a_n, b_m)]_{n+m,j} = O(\xi^2) \quad \text{for } j < n + m - 1 \quad (47)$$

Then dividing (46) and (47) by $D_\xi(a_n, b_m)$ from (39) and taking limits as $\xi \rightarrow 0$, under the assumption (37), we have (38c) as claimed.

Remarks

1. This result can be generalized to the case other a_i, b_i converge to zero using the same technique as above. Details are omitted.
2. The above result can be expressed in terms of an $(n+m) \times (n+m-2)$ matrix

$$[\mathbf{S}^+(\boldsymbol{\theta}, \xi)]_{i,j} = D_\xi^{-1}(a_n, b_m) [\mathbf{M}_\xi(a_n, b_m)]_{i,j} \quad (48)$$

Thus under (37)

$$\mathbf{S}^+(\boldsymbol{\theta}, 0) \triangleq \lim_{\xi \rightarrow 0} \mathbf{S}^+(\boldsymbol{\theta}, \xi) = \begin{bmatrix} I_{m-1} & 0 \\ 0 & 0 \\ 0 & I_{n-1} \\ 0 & 0 \end{bmatrix} \mathbf{S}^{-1}(n-1, m-1) \quad (49)$$

Moreover, simple manipulations yield

$$\mathbf{S}(n, m) \begin{bmatrix} I_{m-1} & 0 \\ 0 & 0 \\ 0 & I_{n-1} \\ 0 & 0 \end{bmatrix} = \left[\mathbf{S}(n-1, m-1) \right]_{n+m}^{\uparrow \downarrow}$$

so that

$$\mathbf{S}(n, m) \mathbf{S}^+(\boldsymbol{\theta}, 0) = \left[I_{n+m-2} \right]_{n+m}^{\uparrow \downarrow} \quad (50)$$

Corollary 1. Consider that $A(q^{-1}, \xi), B(q^{-1}, \xi)$ are the polynomials associated with the $\boldsymbol{\theta}(\xi)$ as in (36). Consider also that $E(q^{-1}, \xi), F(q^{-1}, \xi)$ are the solution to

$$A(q^{-1}, \xi)E(q^{-1}, \xi) + B(q^{-1}, \xi)F(q^{-1}, \xi) = 1 \quad (51)$$

with the degree of $E(q^{-1}, \xi)$ being m and the degree of $F(q^{-1}, \xi)$ being n . Then, under the same conditions as in Lemma 3, the last coefficients of $E(q^{-1}, \xi), F(q^{-1}, \xi)$ converge as

$$\lim_{\xi \rightarrow 0} e_m(\xi) = 0 \quad \lim_{\xi \rightarrow 0} f_n(\xi) = 0 \quad (52)$$

Proof. The proof is straightforward from the result of Lemma 3 and in particular from (49). However, to prove it by just using the properties of polynomial equations appears too formidable without resort to the result on the Sylvester matrix described in Lemma 3.

Lemma 4. Consider the Bezout equation (26) under (14) with $m > 1$. Then a unique solution $\boldsymbol{\varphi}^*$ of (42) exists as

$$\boldsymbol{\varphi}^* = \mathbf{S}^+(\boldsymbol{\theta}, 0) [I_{n+m-2} \ 0] \boldsymbol{\alpha} \quad (53)$$

with the property that $\boldsymbol{\varphi}^*$ has n th and $(n+m)$ th elements which are zero.

Proof. Under (14), of course, $\mathbf{S}^{-1}(n, m)$ does not exist, but (37) holds for a suitable selection of a, b . Also, the last two elements of $\boldsymbol{\alpha}$ are zero. Now Lemma 3 holds under (37) so that (49),

(50) apply. Thus with φ^* uniquely defined from (53)

$$\mathbf{S}(n, m)\varphi^* = \begin{bmatrix} I_{n+m-1} & 0c \\ 0 & 0 \end{bmatrix} \alpha = \alpha$$

and (26) is satisfied. Application of (49) under (51) gives guarantees that the n th and $(n + m)$ th elements of φ^* are zero.

5. SIMULATIONS

Consider now the application of the adaptive pole assignment scheme discussed in Sections 2 and 3 to the plant taken from Reference 10 with

$$y_k - 1.2y_{k-1} = u_{k-1} - 3.1u_{k-2} + 2.2u_{k-3} + w_k$$

$$K = 10 \quad H = 1$$

$$r(k) = \begin{cases} 1 & k = 1 \dots 10, 21 \dots 30, 41 \dots 50 \\ -1 & k = 11 \dots 20, 31 \dots 40, 51 \dots 60 \end{cases} \quad (54)$$

and variance of w_k decaying as k^{-2} .

Three cases are studied:

- (i) The plant is overparametrized by one, i.e. the plant is modelled as

$$y_k + a_1y_{k-1} + a_2y_{k-2} = b_1u_{k-1} + b_2u_{k-2} + b_3u_{k-3} + b_4u_{k-4} + w_k \quad (55)$$

and standard RLS estimation is used.

- (ii) The plant is overparametrized by one too, but the RLS estimation discussed in Section 2 is employed.
- (iii) The plant is not overparametrized, i.e. the plant is modelled as

$$y_k + a_1y_{k-1} = b_1u_{k-1} + b_2u_{k-2} + b_3u_{k-3} + w_k \quad (56)$$

Figures 3–5 show the poor performance in the output, the estimates of the plant and the estimates of the controller in case (i). Figures 6–8 give the comparison of the outputs, the estimates of the plant and the estimates of the controller in cases (ii) and (iii).

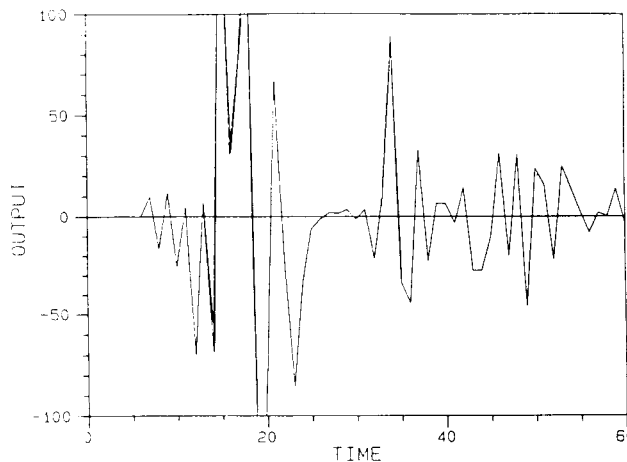


Figure 3. Output (y_k) in case (i)

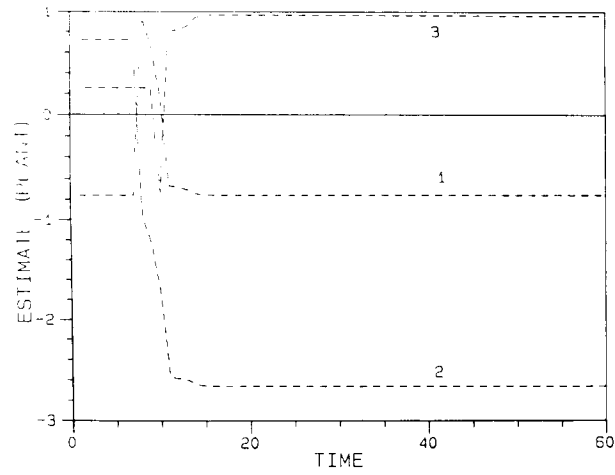


Figure 4. Estimates of the plant in case (i); curve 1, $\hat{a}_{1,k}$; 2, $\hat{b}_{2,k}$; 3, $\hat{b}_{4,k}$

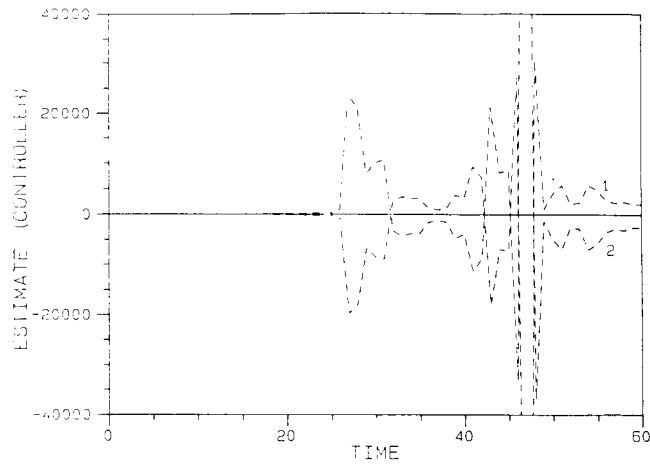


Figure 5. Estimates of the controller in case (i); curve 1, $\hat{f}_{1,k}$; 2, $\hat{f}_{2,k}$

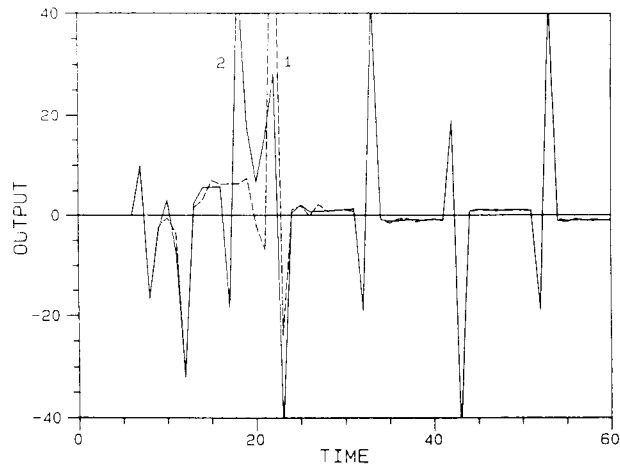


Figure 6. Comparison of the outputs in cases (ii) and (iii); curve 1, y_k in case (ii); 2, y_k in case (iii)

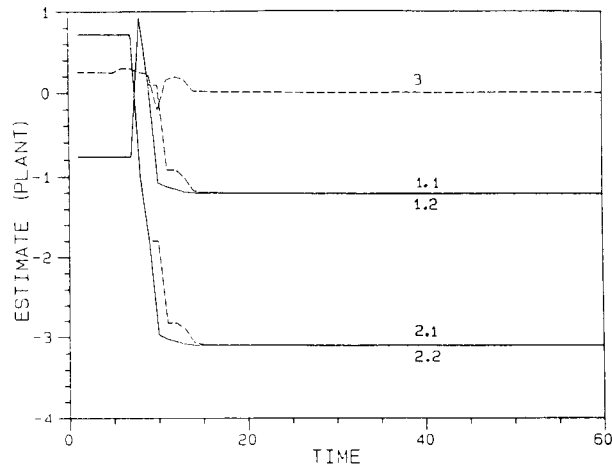


Figure 7. Comparison of the estimated plants in cases (ii) and (iii): curve 1.1, $\hat{a}_{1,k}$ in case (ii); 1.2, $\hat{a}_{1,k}$ in case (iii); 2.1, $\hat{b}_{2,k}$ in case (ii); 2.2, $\hat{b}_{2,k}$ in case (iii); 3, $\hat{b}_{4,k}$ in case (ii)

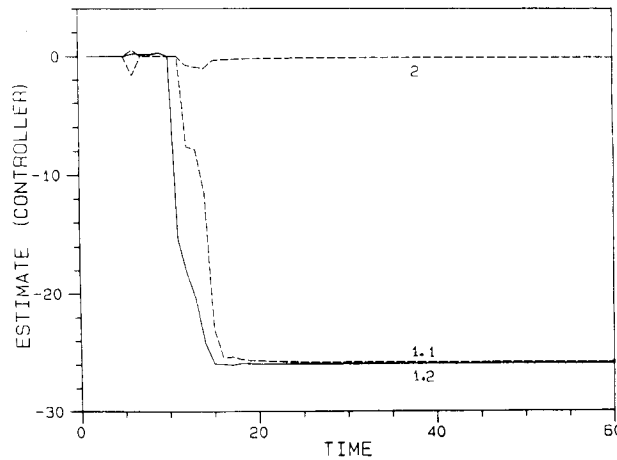


Figure 8. Estimated controllers in cases (ii) and (iii): curve 1.1, $\hat{f}_{1,k}$ in case (ii); 1.2, $\hat{f}_{1,k}$ in case (iii); 2, $\hat{f}_{2,k}$ in case (ii)

From the figures it is clearly shown that in the overparametrization presence the adaptive pole assignment scheme based on the standard RLS estimation of the plant could yield quite poor performance; however, the scheme proposed in this paper works very well. In fact, it converges to the optimal pole assignment controller as the standard adaptive pole assignment scheme does without overparametrization.

6. CONCLUSIONS

Non-standard adaptive estimation and control techniques have been proposed to avoid ill-conditioning which can arise when standard techniques are applied to signal models with overparametrization. The techniques avoid ill-conditioning and yield asymptotic optimality in the case when there is possibly one pole-zero cancellation in the assumed signal model. It seems reasonable to apply such techniques in conjunction with on-line model order estimation tech-

niques, since from finite data these possibly lead to overestimation of the order. The techniques have been studied for the case when exact pole-zero cancellation occurs but is known from simulations to avoid ill-conditioning when there are stable near pole-zero cancellations. The results of the paper are a starting point from which to cope with higher-order pole-zero cancellations.

APPENDIX

Consider first the following lemma.

Lemma 5. Under (13), (14), (31) (the conditions of Theorem 3),

$$\lim_{k \rightarrow \infty} \mathbf{S}^+(\hat{\boldsymbol{\theta}}_k^{\text{CT}}, \xi_k) = \mathbf{S}^+(\boldsymbol{\theta}, 0) \quad (57)$$

Proof. With (31) satisfied and (13)

$$\lim_{k \rightarrow \infty} \hat{\boldsymbol{\theta}}_k^{\text{CT}} = \boldsymbol{\theta}$$

so that when (14) holds

$$\lim_{k \rightarrow \infty} \hat{a}_{n,k}^{\text{CT}} = \lim_{k \rightarrow \infty} a_k \xi_k = 0 \quad \lim_{k \rightarrow \infty} \hat{b}_{m,k}^{\text{CT}} = \lim_{k \rightarrow \infty} b_k \xi_k = 0$$

Since $|a_k| + |b_k| = 1$, recalling (36), then

$$\lim_{k \rightarrow \infty} \xi_k = 0 \quad (58)$$

Now, under (31), (13), (14), we have

$$\lim_{\xi_k \rightarrow 0} \mathbf{S}^+(\hat{\boldsymbol{\theta}}_k^{\text{CT}}, \xi_k) = \mathbf{S}^+(\hat{\boldsymbol{\theta}}_k^{\text{CT}}, 0) = \begin{bmatrix} I_{m-1} & 0 \\ 0 & 0 \\ 0 & I_{n-1} \\ 0 & 0 \end{bmatrix} [\hat{\mathbf{S}}_k^{\text{CT}}(n-1, m-1)]^{-1} \quad (59)$$

Also under (31), (13), (14),

$$\lim_{k \rightarrow \infty} \mathbf{S}^+(\hat{\boldsymbol{\theta}}_k^{\text{CT}}, 0) = \mathbf{S}^+(\boldsymbol{\theta}, 0) \quad (60)$$

From continuity and (58), (57) is established

Remark. Without (31) satisfied, it cannot be guaranteed that $\mathbf{S}^+(\hat{\boldsymbol{\theta}}_k^{\text{CT}}, \xi_k)$ exists for all k . Also, with $\mathbf{S}^+(\hat{\boldsymbol{\theta}}_k, \xi_k)$ large, then $\hat{\boldsymbol{\varphi}}_k$ is large and ill-conditioning is said to occur.

Proof of Theorem 3

Consider the central tendency adaptive control with (31) satisfied and define for all k

$$[\hat{\mathbf{S}}_k^{\text{CT}}(n, m)]^{-1} = [\mathbf{S}^+(\hat{\boldsymbol{\theta}}_k^{\text{CT}}, \xi) *] \quad (61)$$

where * denotes terms not of interest. Thus for the signal model with $m > 1$, or in other

words the last two entries of $\hat{\alpha}_k^{\text{CT}}$ zero,

$$\begin{aligned}\hat{\alpha}_k^{\text{CT}} &= [\hat{\mathbf{S}}_k^{\text{CT}}(n, m)]^{-1} \hat{\alpha}_k^{\text{CT}} = [\hat{\mathbf{S}}_k^{\text{CT}}(n, m)]^{-1} [I_{n+m-2} \ 0]^T [I_{n+m-2} \ 0] \hat{\alpha}_k^{\text{CT}} \\ &= \mathbf{S}^+ (\hat{\boldsymbol{\theta}}_k^{\text{CT}}, \xi) [I_{n+m-2} \ 0] \hat{\alpha}_k^{\text{CT}}\end{aligned}\quad (62)$$

Now applying (57) when taking limits as $k \rightarrow \infty$, we get

$$\lim_{k \rightarrow \infty} \hat{\boldsymbol{\phi}}_k^{\text{CT}} = \lim_{k \rightarrow \infty} \mathbf{S}^+ (\hat{\boldsymbol{\theta}}_k^{\text{CT}}, \xi_k) [I_{n+m-2} \ 0] \hat{\alpha}_k^{\text{CT}} = \mathbf{S}^+ (\boldsymbol{\theta}, 0) [I_{n+m-2} \ 0] \boldsymbol{\alpha}$$

The desired results (33), (34) follow from application of Lemma 4.

REFERENCES

1. Goodwin, G. C., and K. S. Sin, *Adaptive Filtering Prediction and Control*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.
2. Shinnaka, S. and T. Suzuki, 'Adaptive pole placement applicable to nonminimum-phase systems of unknown order and unknown dead-time', *Denshi Tsushin Gakkai Ronbunshi*, **66A**, (12), 1228–1235 (1983).
3. Sternby, J., 'On consistency for the method of least squares using martingale theory', *IEEE Trans. Automatic Control* **AC-22**, 346 (1977).
4. J. B. Moore, Persistence of excitation in extended least squares, *IEEE Trans. Automatic Control*, **AC-28**, 60–68 (1983).
5. Green, M. and J. B. Moore, 'Persistence of excitation in linear systems', *Proc. American Control Conf.*, Boston, 1985, pp. 412–417.
6. Moore, J. B., 'A universability advantage of stochastic excitation signals for adaptive control', *Syst. Control Lett.*, **9**, 55–58 (1987).
7. Moore, J. B., 'Stochastic adaptive control via consistent parameter estimation' *Proc. IFAC Identification Conf.*, York, July 1985, pp. 611–616.
8. Kailath, T., *Linear Systems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1980.
9. Moore, J. B., T. Ryall and L. Xia, 'Central tendency pole assignment', *Proc. 25th CDC*, Athens, 1986, pp. 100–105.
10. Goodwin, G. C. and K. S. Sin, 'Adaptive control of non minimum phase plants', *IEEE Trans. Automatic Control*, **AC-26**, 478–483 (1981).