ON ADAPTIVE ESTIMATION AND POLE ASSIGNMENT
OF OVERPARAMETERIZED SYSTEMS

J.B. Moore, Lige Xia, and M. Gevers

Department of Systems Engineering
Research School of Physical Sciences, Australian National University
G.P.O. Box 4, Canberra, A.C.T., 2600, Australia.

ABSTRACT

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 overparameterization. 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 to on-line order determination, being based on introducing suitable excitation in the "regression vectors of the parameters estimation unhindered by a ensure parameter convergence. For the case when the models are nonunique 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 (unconditioned when estimates of the plant have near pole-zero cancellations a common problem. This paper shows that (after appropriate preprocessing), certain ill-conditioning associated with overparameterization can be avoided, without the need to perform on-line order determination with its associated significant increase in computational complexity.

For overparameterized 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 nonuniquely parameterized models which include pole-zero cancellations in the complex z-plane. When recursive estimates of the parameters of such nonuniquely parameterized models are applied for adaptive control, then ill-conditioning leading to excessive controls can easily arise, particularly in adaptive pole assignment schemes.

Adaptive pole assignment schemes are perhaps the simplest schemes for adaptively stabilizing linear plants which are possibly nonminimum 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 overparameterized. Adaptive pole assignment requires the solution of a linear algebraic equation which becomes ill-conditioned even when the plant has near pole-zero cancellation. Such is inevitable when the signal model is overparameterized. 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 signal model is overparameterized. For the special case when there is a potential nonuniqueness in the signal model due 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. Proofs are presented 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 wk, output yk and white noise disturbances wk.

\[
A(q^{-1})y_k = B(q^{-1})u_k + w_k \\
A(q^{-1}) = 1 + \alpha_q q^{-1} + \ldots + \alpha_0 q^{-n}, \\
B(q^{-1}) = b_0 q^{-1} + \ldots + b_{m-1} q^{-m}
\]

This can be rewritten as

\[
y_k = \theta^T x_k + w_k, \quad \theta^T = [a_0, a_1, \ldots, a_{n-1}, b_1, b_2, \ldots, b_m] \]

The conditions on wk are more precisely

The sequence \{wk\} is independent of wk with \E[w_k F_k^{-1}] = 0, \E[w_k^2] < \infty, \|w_k\| < \infty where F_k denotes the \sigma^-algebra generated by wk, w_k, \ldots, wk.

Recessive Least Squares (RLS) Estimation
Consider that \theta is estimated recursively minimizing a least squares criterion:

\[
\hat{\theta}_k = \hat{\theta}_{k-1} - \frac{1}{\lambda_k} \sum_{l=1}^{k} (y_l - \theta^T_k x_l) x_l^T \hat{\theta}_{k-1} = \hat{\theta}_{k-1} + \lambda_k \hat{\theta}_{k-1}
\]

via an RLS scheme as

\[
\hat{\theta}_k = \hat{\theta}_{k-1} - \lambda_k \hat{\theta}_{k-1} + \lambda_k \hat{\theta}_{k-1} x_k^T F_k^{-1} + \lambda_k \hat{\theta}_{k-1} x_k^T F_k^{-1} \lambda_k \hat{\theta}_{k-1} = \hat{\theta}_{k-1}
\]

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

Convergence Properties Review To achieve a simple analysis making connection with Kalman filter theory as in [3], let us assume

When wk is normally distributed, and the a priori probability density distribution with \theta is

\[
N(\hat{\theta}_0, \Sigma_0) \text{ for some } P_0 = B_0^{-1} > 0.
\]

The assumption is not needed for a more general theory based on stochastic Lyapunov functions \phi_k^T P_k \phi_k in [4,5].

Moreover, from [3] there is almost sure convergence as

\[
\lim_{k \to \infty} \hat{\theta}_k = \hat{\theta}_k \text{ a.s.}
\]
for random variables $\tilde{P}_k, \tilde{Q}_k$. With $\tilde{x}_k$ sufficiently existing in that $P \tilde{S} \leq 0$, then $\tilde{S} \leq \theta$. Also if $P \tilde{S} = \theta$, then $\theta_k \tilde{R} \geq 0$ as $k \to \infty$. Thus the following strong connection between sufficiency of excitation of $x_k$ and parameter convergence.

**Lemma 2.1**

For the RLS scheme (2.3) applied to the signal model (2.1) (2.2) under (2.4), then

$$1 \implies \begin{cases} P_k \to 0 & \Longleftrightarrow \begin{cases} \theta_k \to \theta \\ k \to +\infty \end{cases} \end{cases}$$

**Proof:** As above based on results in [3].

**On Sufficient Excitation**

In this subsection three specific excitation scenarios are studied using known results from [4,5,6]. 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 overparameterization, while the third case deals with the case of overparameterized models.

**Case (i):** The simplest case to study is the nonoverparametrized case when

$$q^2A(q^{-1}), q^mB(q^{-1})$$

are coprime.

(2.8)

and $u_k$ is suitably exciting in that

$$k \rightarrow \begin{cases} \begin{cases} u_k \in \mathbb{C} \cup \mathbb{C}^T \\ \sum_{-1}^{u_k} + u_k \in \mathbb{C}^{m+1} \end{cases} & \text{as } k \to \infty \end{cases}$$

(2.9)

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

Under (2.8), the model (2.1) is uniquely parametrized, $x_k$ is reachable from $u_k$, and excitation of the inputs as in (2.9) implies excitation of the states $\tilde{x}_k$ [4,5]. Thus (2.8),(2.9) ensures that $P_k \to 0$ as $k \to \infty$, and in turn $\theta_k \to \theta$ as $k \to \infty$. For this case then, $x_k$ need not be sufficiently exciting in any sense.

**Case (ii):** Another simple case to study is when (2.8) does not necessarily hold, but $u_k$ and $s_k$ are suitably exciting in that (2.9) is satisfied and

$$\sum_{k=0}^{\infty} \mathbb{E}[u_k^2] = \infty$$

(2.10)

Under (2.9),(2.10), the model (2.1) is uniquely parametrized, $x_k$ is reachable from $u_k$, and $x_k$ is sufficiently exciting to guarantee that $P_k \to 0$ as $k \to \infty$, and in turn that $\theta_k \to \theta$ as $k \to \infty$. For this case then, the convergence as such is independent of whether or not the coprime condition (2.8) is satisfied.

**Case (iii):** The possibly overparametrized signal model situation of particular interest in this paper, is when (2.8) possibly fails and there is no a priori guarantee of sufficient excitation of $u_k$ as in (2.10). In this case the model (2.1) may not be uniquely parametrized, having one or more pole-zero cancellations in the $z$-plane. Also, $x_k$ may not be sufficiently exciting to ensure that $P_k \to 0$ as $k \to \infty$. 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 (2.1) re-organized as

$$y_k = \sum_{i=1}^{k} \sum_{i=1}^{y_k} + u_k = x_k + v_k$$

(2.11)

where $v_k$ is an excitation term to ensure that $x_k$ is suitably exciting. Notice that $v_k$ has no influence on $u_k$, $x_k$. 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, $v_k$ converges to zero.

**RLS Estimation with Regression Vector Excitation**

Consider that $\theta$ is estimated recursively minimizing a least squares criterion:

$$J_k = \frac{1}{2}\left[ \sum_{i=1}^{k} \sum_{i=1}^{y_k} + (\theta - \theta_0)^T B_0 (\theta - \theta_0) \right]$$

(2.12)

via an RLS scheme as

$$\begin{cases} \theta_k = P_k (\theta_k - \theta_k - 1 + x_k)^T \\ B_k = B_k (\theta_k + x_k x_k)^T \\ P_k = P_k - P_k (I_k - x_k x_k^T) P_k (I_k - x_k x_k^T)^{-1} \\ = B_k B_k \end{cases}$$

(2.13)

for some initial conditions $\theta_0, B_0 > 0$.

**Regression Vector Excitation Selection**

Consider the signal model (2.1) formulated as (2.11). Let us assume that either there is no overparameterization in (2.8) holds, or that there is the possibility of overparameterization which includes one pole-zero cancellation, leading to a nonuniquely parametrized model. In the latter case, we consider a unique parametrization with the properties (meaningful only when $n=m$):

$$\theta_k = \theta_k = \begin{cases} \mathbb{E}[u_k] & \text{as } \mathbb{E}[u_k] = \theta_k \end{cases}$$

(2.14)

For such a situation, we propose a $v_k$ selection as follows.

The sequence $\{v_k\}$ is selected as an independent (Gaussian) zero mean white noise excitation term subject to the covariance $E_k = D_k (v_k - n)$ for $D_k = \text{diag}(0 0 \ldots 0 d_k^0 \ldots 0 d_k^m)$ where $d_k^0 > 0$, $d_k^m > 0$. Denote the nonzero elements at $\hat{v}_n(k), n = n, n = m(k)$

(2.15)

The Gaussian assumption on $v_k$ 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 $v_k$ is deterministic but containing a sufficient number of frequency components.

The following results are now a consequence of a straightforward application of results from [4,5].

**Lemma 2.2**

Consider the linear signal model (2.1) (2.2) formulated as (2.11) with the $v_k$ selection of (2.15). (i) Then under (2.8), $x_k$ is the output of a linear time-invariant system reachable from $u_k$. Moreover, with $u_k$ selected so that for some $\alpha > 0$

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \\ k \to \infty \end{cases}$$

(2.16)

then

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \\ k \to \infty \end{cases}$$

(2.17)

(ii) Also under (2.14), $x_k$ is the output of a linear time-invariant system reachable from $u_k$ defined by the excitation conditions on $u_k$ of (2.4) and $v_k$ namely,

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \\ k \to \infty \end{cases}$$

(2.18)

translate to excitation of $x_k$ as

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \\ k \to \infty \end{cases}$$

(2.19)

**Proof:**

(i) Under the coprime condition (2.8), then from [5] it is immediate that $x_k$ is the output of a linear time-invariant system driven by $u_k$, $x_k$, and is reachable from $u_k$ alone. Now should $u_k$ be acting alone, then Lemma (3.2) of [4] applies to give that for all $k$ and some $k > 0$

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \end{cases}$$

(2.20)

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

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \end{cases}$$

(2.21)

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \end{cases}$$

(2.22)

$$\begin{cases} \mathbb{E}[u_k] = \theta_k \end{cases}$$

(2.23)

Applying the results of Lemma (3.3) and its Remark 1 of [4], now tells us that when bounded variance white noise inputs $u_k, w_k$ (independent of $u_k$ and $w_k$ are also applied, the same implications hold, so that (2.16) implies (2.17a) as claimed. The result (2.17b) holds likewise.

(ii) Under (2.14), the elements $v_k = \ldots v_k + n - 1, v_k = \ldots v_k + n + 1$ of $x_k$ are reachable from $u_k$ alone [5], and the remaining elements $(v_k - n + v_n, (v_k - m + v_n + m))$ of $x_k$ are reachable from $u_k$ alone [5], and the remaining elements $(v_k - n + v_n, (v_k - m + v_n + m))$ of $x_k$,
are reachable from the nonzero elements of $v_k$, namely $v_{n,k}$ and $v_{n+m,k}$. Thus $v_k$ is reachable from $v_{k}$, $v_{k}$ under (2.14). Applying again the Lemma (3.3) and its Remark 1 of [4] gives directly that (2.4), (2.18) together imply (2.19).

## Main Results of Section

**Theorem 2.1**

Consider the signal model (2.1), (2.2) which is possibly overparameterized in that either (2.8) or (2.14) holds. Consider an RLS scheme (2.13) based on the alternative model formulation (2.11) with $v_k$ selected as in (2.15) and (2.4) holding. Consider also that $v_k$ is sufficiently exciting in that (2.16) holds. Then there is parameter convergence as

$$
lim_{k \to \infty} \theta_k = \theta \quad \text{a.s.}$$

(2.20)

where $\theta$ is the unique parameter associated with (2.1) under (2.8) or (2.14). Moreover (2.19) also holds.

**Proof:**

Part (i) In the case that (2.14) is satisfied, so that $\delta_0 = b_{2n} = 0$, then $\delta^0 V_k = 0$. Now Lemma (2.1) applies with $V_k$ replacing $V_k$, so that (2.20) holds if and only if $v_0$ holds. Assume that (2.19) does not hold, then since $P_k < P_{k-1}$ for all $k$, $tr(P_k) \to 0$ and from (2.15) the variance of $v_k$ decay as a rate $k^{-1}$. Thus (2.18) holds, and in turn (2.19) holds under Lemma 2.2. This contradicts the assumption, so that (2.19) and (2.20) hold.

Part (ii) In the case that (2.8) is satisfied, Lemma 2.2 tells us that (2.16) implies (2.17). As a consequence from (2.15) then

$$
lim_{k \to \infty} \sigma^2_{n,k} < \infty, \quad \lim_{k \to \infty} \sigma^2_{n+m,k} < \infty$$

(2.21)

In turn we claim that

$$
limit_{k \to \infty} \sum_{i=1}^{k} (e_i^T v_k) \leq 0 \quad \text{a.s.}$$

(2.22)

This follows, since as is readily established under (2.15)

$$
\sum_{i=1}^{k} (v_{i}^T v_k - \sigma^2_{n,k}) \leq \sum_{i=1}^{k} (v_{i}^T v_k - \sigma^2_{n+m,k}) \leq \sum_{i=1}^{k} \sigma^2_{n,m} \leq \sum_{i=1}^{k} \sigma^2_{n+m,k}
$$

(2.23)

are martingales bounded in $L_2$ and converge almost surely. Under (2.21) and (2.17b) we now claim

$$
limit_{k \to \infty} \sum_{i=1}^{k} e_i^T v_k \leq 0 \quad \text{a.s.}$$

(2.24)

To see this observe that

$$E[M_k] = \sum_{i=1}^{k} \sum_{j=1}^{i} e_i^T v_k \leq \sum_{i=1}^{k} \sum_{j=1}^{i} (v_{i}^T v_k - \sigma^2_{n,k}) = E[M_k] - \sum_{i=1}^{k} \sum_{j=1}^{i} \sigma^2_{n,k}
$$

Thus $M_k$ is a martingale on $P_{k-1}$, bounded in $L_2$, under (2.21) and so converges almost surely, so that (2.17b) implies (2.22).

Now under (2.17b) Lemma (2.1) can be applied to yield the following

$$
limit_{k \to \infty} \sum_{i=1}^{k} e_i^T v_k \leq 0 \quad \text{a.s.}$$

(2.25)

The former results follow from (2.7) and the relationships

$$
\theta_k = \sum_{i=1}^{k} x_i y_i, \quad \theta_k = \sum_{i=1}^{k} x_i (w_i - b_p)
$$

(2.26)

The second result follows from the first since $(v_i^T \theta)$ has the same essential properties as $w_i$ in (2.2), (2.4). The results (2.22),(2.23) lead in turn to the following convergence results

$$
P_k (P_k + \Sigma (x_i x_i^T) + v_i v_i^T + v_i v_i^T + v_i v_i^T) + 1 \quad \text{as} \quad k \to \infty
$$

(2.27)

$$
\theta_k = \sum_{i=1}^{k} \sum_{j=1}^{i} v_i \quad \text{as} \quad k \to \infty
$$

(2.28)

so that (2.23) holds as claimed.

**Remarks:**

1. The specific $v_k$ 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. Rather than work with banks of estimators as in Remark 2 above, an ad hoc approach is to relax (2.15) and have $v_k$, $v_{k-1}$, ..., $v_k$, each independent and suitably exciting with the variance of $v_k$ increasing with $k$. Such would force pole-zero cancellations to occur near the origin, but would lead to biased estimates. We do not study this technique.

3. Algorithms and results of this section have been analysed for the simplest of stochastic signal models, namely, when $(y_k - \delta^0 v_k)$ 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 $v_k$ can also be used to extend the capability of ELS based schemes. In particular, for those which are globally convergent under the coprimeness condition (2.9), [7] with the modifications they are globally convergent also when there is a pole-zero overparameterization as when (2.14) 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 (2.1) so that there is asymptotic convergence to

$$H(q^{-1})y_k = K(q^{-1})r_k$$

(3.1)

where $H(q^{-1}) = 1 + b_1 q^{-1} + \cdots + b_{n+m} q^{-1} - m$ 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].

$$H(q^{-1})y_k = -H(q^{-1})F(q^{-1})y_k + Kr_k$$

(3.2)

where $E[q^{-1}] = 1 + e_1 q^{-1} + \cdots + e_{n+m} q^{-1}$, $F(q^{-1}) = f_1 q^{-1} + \cdots + f_n q^{-1}$ are given from the solution of the Bezout equation.

$$A(q^{-1})E(q^{-1}) = B(q^{-1})F(q^{-1}) = 1 \quad \text{and} \quad S(n,m) = c(3.3)$$

where $v_i = [v_i, e_1, e_2, \cdots, e_m, f_1, f_2, \cdots, f_n]$ and

$$a^T = [-a_1, -a_2, \cdots, -a_m, 0, \cdots, 0]$$

$$S(n,m) = \begin{bmatrix} a_1 & \cdots & b_1 & \cdots & a_n & \cdots & b_m & \cdots & a_{n+m} \end{bmatrix}$$

(3.4)

It is known that solutions of (3.3) exist if and only if $r = \text{rank}[S] = \text{rank}[S]$. Also from [8]

$$S(n,m) = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

(3.5)

$S(n,m)$ is nonsingular if and only if $q^T A(q^{-1})$, $q^T B(q^{-1})$ are coprime.
In certainty equivalence adaptive pole assignment, the estimates $\hat{\theta}_k$ are used in lieu of $\theta$ in (3.4) to compute on-line estimates $\hat{S}_k^{CE}(n,m)\hat{P}_k^{CE}$ of the controller parameters, so that in obvious notation

$$\hat{S}_k^{CE}(n,m)\hat{P}_k^{CE} = \tilde{\theta}_k^{CE}$$  \hspace{1cm} (3.6)

When $q^\alpha A(q^{-1})$, $q^\alpha B(q^{-1})$ are coprime then $\tilde{\theta}_k(n,m)$ is nonsingular and the solution of (3.6) exists. Otherwise it may not. III-conditioning in $\tilde{\theta}_k(n,m)$ can cause excessive values for $k^{\alpha}$. One modification to avoid large $k^{\alpha}$ is to select $k^{\alpha} = 1$ during III-conditioning, but for unknown plants, it is not a priori clear how to quantify III-conditioning to achieve a useful adaptive controller.

Central Tendency Adaptive Control: In central tendency adaptive pole assignment [9], III-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 $\theta = \{\theta_k^{CE}, \theta_k^{CT}\}$ where $\theta_k^{CT}$ is an estimate of $\sigma^2$, then there is an associated noncentral probability density for the pole assignment controller parameters $\theta_k^{CT}$. A central tendency selection $\theta_k^{CT}$ is one which maximizes this density, or at least avoids the tails of this density. Practical implementations are given in [9]. Associated with $\theta_k^{CT}$ are the pole estimates $k^{\alpha}$ which is not in general $\hat{\theta}_k$. Then, in obvious notation

$$\hat{S}_k^{CT}(n,m)\hat{P}_k^{CT} = \tilde{\theta}_k^{CT}$$  \hspace{1cm} (3.7)

The estimate $\tilde{\theta}_k^{CT}$ has the property that it is _"close"_ to $\theta_k^{CE}$ but is far from hypersurfaces for which $q^\alpha A(q^{-1})$, $q^\alpha B(q^{-1})$ are not coprime. As a consequence, the following property is claimed for central tendency adaptive pole assignment.

The selections $\theta_k^{CT} \to \theta$ are such that $[\hat{S}_k^{CT}(n,m)]^{-1}$ exists for all $k$, and if $\theta$ belongs to the hypersurface (in $\theta$-space) defined by $q^\alpha A(q^{-1})$, $q^\alpha B(q^{-1})$ not coprime, then as $k \to \infty$, $\theta_k^{CT}$ is contained in a cone centered at $\theta$ which excludes the tangent hyperplane at $\theta$. \hspace{1cm} (3.8)

Remarks:

1. To give a geometric interpretation of (3.8), consider Figures 3.1, 3.2. The heavy arc is the pole zero cancellation singular region for $\theta$ estimates in $\Theta$-space. The light shaded area is a zone of III-conditioning control surrounding the singular arc. The double shaded area contains the cone region of possible central tendency control estimates $\theta_k^{CT}$ of (3.8). Figure 3.1 depicts the situation when the solution (2.14) is at the conic intersection and the cones avoid the III-conditioned regions. Figure 3.2 depicts a solution not satisfying (2.14) when the central tendency estimates $\theta_k^{CT}$ are not conditioned when converging.

2. The property (3.8) can be viewed as a corollary of results rigorously proved in [9], although specific reference has not been made to (3.8) in [9]. The property is readily believable, but since it is not rigorously proved as such in [9] and is beyond the scope of this paper, we here add the "qualification" to central tendency control that (3.8) be satisfied.

Then two cases are studied now.

Case (i) Known Model Order: We now claim the following

Theorem 3.1

Consider the signal model (2.1) for the case when it is not overparameterized, so that $q^\alpha A(q^{-1})$, $q^\alpha B(q^{-1})$ are coprime. Consider also RLS parameter estimates $\hat{\theta}_k$ from (2.13), and associated certainty equivalence [or central tendency] pole assignment with controller parameters $\hat{S}_k^{CE}(n,m)$ or $\hat{S}_k^{CT}(n,m)$ and Sylvester matrices $S_k^{CE}(n,m)$ or $S_k^{CT}(n,m)$.

Consider also that $\hat{\theta}_k$ is sufficiently exciting so that $\lim_{k \to \infty} \hat{\theta}_k \to \theta$ as $k \to \infty$. Then

$$\lim_{k \to \infty} [\hat{S}_k^{CE}(n,m)]^{-1} = S^{-1}(n,m), \quad \text{or}$$

$$\lim_{k \to \infty} [\hat{S}_k^{CT}(n,m)]^{-1} = S^{-1}(n,m)$$

(3.9)

Proof: This immediate from the coprime ness assumption and the property (3.5). Notice that there is no need of assumption (3.8).

Case (ii) Overparameterization: We now claim the following

Theorem 3.2

Consider the signal model (2.1) with $n > 1$, $m > 1$, (2.14) holding. Consider also RLS parameter estimates $\hat{\theta}_k$ of (2.13) and associated central tendency adaptive pole assignment controller parameters $\hat{S}_k^{CT}(n,m)$ given from (3.7) with (3.8) satisfied. Then under RLS convergence of $\theta_k^{CT}$ to $\theta$ as $k \to \infty$,

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

(3.10)

where $\varphi^*$ is a unique solution of $S(n,m)\varphi = 0$ with zero elements as

$$\varphi^* = [\varphi_1^{*1}, \ldots, \varphi_m^{*1}]^T$$

(3.11)

Proof: The proof is given in the Appendix, based on 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 nonstandard nature of the RLS algorithm with its internal perturbations $V_k$. In the presence of overparameterization, standard RLS estimation (when $V_k = 0$) will almost surely not converge to the unique limits solution (2.14). This means that there is inevitable III-conditioning. This situation applies even when a central tendency adaptive control law is implemented, as depicted in Figure 3.2.

3. The above theorem result is also facilitated by the central tendency property (3.8) of the adaptive controller. In the presence of overparameterization, even when the modified RLS algorithm of the paper is implemented without (3.8) satisfied, there is a nonzero probability of III-conditioning as suggested from Figure 3.1, at least during transients.

4. The theorem is developed in conjunction with the RLS estimation (2.4) which cope with possible overparameterization, by the $a_k b_m$. Should some RLS based scheme cope with higher orders of overparameterization, 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 apply to the situation when (3.3) is replaced by

$$A(q^{-1})B(q^{-1}) + B(q^{-1})F(q^{-1}) = H(q^{-1})$$

(3.12)

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 (3.3) applies. The restriction on $F(q^{-1})$ implies that the associated $z$ in the algebraic form of (3.12) corresponding to (3.14) 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 $N > 1$.

4. A PROPERTY OF SYLVESTER MATRICES

First recall (3.4) which associates with $\theta = [a_1, \ldots, a_n, b_1, \ldots, b_m]^T$ a Sylvester matrix $S(n,m)$. Let us denote the adjoint of $S(n,m)$ as $M_S(n,m)$ and determinant as $D(n,m)$. Now consider a (linear) trajectory in $\theta$-space parametrized in terms of a scalar variable $\xi$, as

$$\xi(t) = [a_1, \ldots, a_n, (at) b_1, \ldots, b_m, (bt)]^T$$

Then

(4.1)

Also denote the Sylvester matrix associated with $\xi(t)$ as $S_\xi(n,m)$ its adjoint matrix as $M_S(\xi(n,m))$ and determinant as $D(\xi, n,m)$. Then we claim the following:

Lemma 4.1

Consider

$$S_{\xi(n,m)} = D(t)S(\xi(n,m))M(\xi(n,m))$$

for the case that $D(t-1, m-1) = 0, ab_{m-1} - a_{n-1}b_t = 0$.

Then the following limits exist as $\xi \to 0$ for all $j < n + m - 1$
\[
\lim_{\xi \to 0} D_F^{-1}(a_n, b_m)[M(a_n, b_m)]_{i,j}, \quad \xi \to 0 \tag{4.3a}
\]
\[
[S^{-1}(n-1,m-1)]_{i,j} \quad \text{for } i < m 
\]
\[
[S^{-1}(n-1,m-1)]_{i-1,j} \quad m < i < n+m \tag{4.3b}
\]
\[
0 \quad \text{for } i = m \text{ or } i = m+n \tag{4.3c}
\]

Proof: We consider in turn expressions for \(D_T(a_n, b_m)\) and \(M_T(a_n, b_m)\) in terms of order \(\xi\) and higher order terms denoted by \(O(\xi^2)\). Simple manipulations give an expression for \(D_T(a_n, b_m)\) in terms of the elements of the last row of \(S_T(a_n, b_m)\) and their minors, with the minors likewise expanded, as
\[
D_T(a_n, b_m) = (-1)^{m-n}b_m a_{n-m} D(n-1,m-1) + O(\xi^2) \tag{4.4}
\]
Noting that \(M_T(a_n, b_m)\) is a (signed) determinant of a submatrix of \(S_T(a_n, b_m)\), then again simple manipulations give an expression in terms of the elements of the last row of the submatrix of \(S_T(a_n, b_m)\) and minors of this submatrix as
\[
[M_T(a_n, b_m)]_{i,j} = (-1)^{m-n+i-j}M_T(a_n, b_m-m-1)]_{i,j} + O(\xi^2), \quad \text{for } i < m, j < n+m-1 \tag{4.5}
\]
Here \(M_T(a_n, b_m-m-1)\) denotes the adjoint matrix of the lower dimensional Sylvester matrix \(S_T(a_n, b_m-m-1)\) which associates with the vector \(\{a_{n-m}, a_{n-m+1}, a_{n-m+2}, \ldots, a_n\}\). Using derivations similar to those giving (4.5), then
\[
[M_T(a_n, b_m-m-1)]_{i,j} = b_{n-m-1} D(n-1,m-1) \times n+m-1 \tag{4.6}
\]

where \(S^{-1}(n-1,m-1)\) exists under the assumption (4.2). The dual form of (4.6) is
\[
[M_T(a_n, b_m-m-1)]_{i,j} = (-1)^{m-n+i-j}M_T(a_n, b_m-m-1)]_{i,j} + O(\xi^2), \quad \text{for } i < m, j < n+m-1 \tag{4.7}
\]
Substitution of (4.6) into (4.7) yields
\[
[M_T(a_n, b_m)]_{i,j} = (-1)^{m-n+i-j}M_T(a_n, b_m-m-1)]_{i,j} + O(\xi^2), \quad \text{for } i < m, j < n+m-1 \tag{4.8}
\]
Dividing (4.8) by \(D_T(a_n, b_m)\) from (4.4) and taking limits as \(\xi \to 0\) under the assumption (4.2) leads to the result (4.3a). The result (4.3b) can be established along similar lines. Similar arguments to the derivation of (4.5) leads to
\[
[M_T(a_n, b_m)]_{i,j} = b_{i} M_T(a_n, b_m-m-1)]_{i,j} + O(\xi^2), \quad \text{for } j < n+m-1 \tag{4.9}
\]
Similarly, we have
\[
[M_T(a_n, b_m)]_{i,j} = O(\xi^2), \quad \text{for } j < n+m-1 \tag{4.10}
\]
where \(O(\xi^2)\) denotes a quantity that is bounded in term of \(\xi^2\).
\[
[M_T(a_n, b_m)]_{i,j} = O(\xi^2), \quad \text{for } j < n+m-1 \tag{4.11}
\]
Similarly, we have
\[
[M_T(a_n, b_m)]_{i,j} = O(\xi^2), \quad \text{for } j < n+m-1 \tag{4.12}
\]
Then dividing (4.11) and (4.12) by \(D_T(a_n, b_m)\) from (4.4) and taking limits as \(\xi \to 0\) under the assumption (4.2), we have (4.3c) as claimed.

Remarks:
1. This result can be generalized to the case other \(a_i, b_j\) converging to zero using the same technique as above. Details are omitted.
2. The above result can be expressed in terms of an \((n+m)\) by \((n+m-2)\) matrix.
\[
[S^{-1}(n, m)]_{i,j} = D_F^{-1}(a_n, b_m)[M_T(a_n, b_m)]_{i,j} \tag{4.13}
\]
Thus under (4.2)
\[
S^*(a, 0) \triangleq \lim_{\xi \to 0} S^*(a, \xi) = \begin{bmatrix} I_{m-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{4.14}
\]
Moreover, simple manipulations yield
\[
S(n, m) = \begin{bmatrix} I_{m-1} & 0 \\ 0 & I_{n-1} \end{bmatrix} \tag{4.15}
\]
so that
\[
S(n, m)S^*(a, 0) = \begin{bmatrix} I_{m-1} & 0 \\ 0 & I_{n-1} \end{bmatrix} \tag{4.16}
\]

Corollary 4.2
Consider that \(A(q^{-1}, \xi), B(q^{-1}, \xi)\) are the polynomials associated with the \(s(\xi)\) as in (4.1). Consider also that \(E(q^{-1}, \xi), F(q^{-1}, \xi)\) are the solution to \(A(q^{-1}, \xi)E(q^{-1}, \xi) + B(q^{-1}, \xi)F(q^{-1}, \xi) = 0\) with the degree of \(E(q^{-1}, \xi)\) being \(m\) and the degree of \(E(q^{-1}, \xi)\) being \(n\). Under the same conditions as in Lemma 4.1, then the last coefficients of \(E(q^{-1}, \xi), F(q^{-1}, \xi)\) converge as
\[
\lim_{\xi \to 0} e_n(\xi) = 0, \quad \lim_{\xi \to 0} f_n(\xi) = 0 \tag{4.17}
\]
Proof: It is straightforward from the result of Lemma 4.1, and in particular from (4.14). 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 4.1.

Lemma 4.3
Consider the Bezout equation (3.3) under (2.14) with \(m > 1\). Then a unique solution \(\phi^*\) of (3.7) exists as
\[
\phi^* = S^*(a, 0)I_{n+m-2} \theta_0 \tag{4.18}
\]
with the property that \(\phi^*\) has \(n\)th and \((n+m)\)th elements which are zero.

Proof: Under (2.14), of course \(S^*(n, m)\) does not exist, but (4.2) holds for a suitable selection of \(a, b\). Also, the last two elements of \(\theta\) are zero. Now Lemma 4.1 holds under (4.2) so that (4.14),(4.15) apply. Thus with \(\phi^*\) uniquely defined from (4.18)
\[
S(n, m)\phi^* = \begin{bmatrix} I_{n+m-2} & 0 \\ 0 & 0 \end{bmatrix} \alpha = \alpha \tag{4.19}
\]
and (3.3) is satisfied. Application of (4.14) under (4.16) gives guarantees that the \(n\)th and \((n+m)\)th elements of \(\phi^*\) are zero.

5. SIMULATIONS
Presented at Conference.

6. CONCLUSIONS
Nonstandard adaptive estimation and control techniques have been proposed to avoid ill-conditioning which can arise when standard techniques are applied to signal models with overparameterization. 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 techniques 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.

REFERENCES

APPENDIX

Consider first the following lemma.

**Lemma A.1**

Under (2.13), (2.14), (3.8) [the conditions of Theorem 3.2]

\[
\lim_{k \to \infty} S^*(\hat{\theta}_k^{CT}, \xi_k) = S^*(\theta, 0)
\]  

(A.1)

**Proof:**

With (3.8) satisfied and (2.13)

\[
\hat{\theta}_k^{CT} = \theta
\]

so that when (2.14) holds

\[
\lim_{k \to \infty} \hat{\theta}_n^{CT} = \lim_{k \to \infty} \theta_n \xi_k = 0
\]

\[
\lim_{k \to \infty} \hat{b}_m^{CT} = \lim_{k \to \infty} \theta_b \xi_k = 0
\]

Since \( \theta_n \xi_k + \theta_b \xi_k = 1 \), recalling (4.1), then

\[
\lim_{k \to \infty} \xi_k = 0
\]

(A.2)

Now, under (3.8), (2.13), (2.14), we have

\[
\lim_{k \to \infty} S^*(\hat{\theta}_k^{CT}, \xi_k) = S^*(\theta, 0)
\]

\[
\hat{\theta}_k^{CT}, \xi_k \to 0
\]

\[
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\]

\[
\hat{\theta}_k^{CT}(n-1, m-1)
\]

(A.3)

Also under (3.8), (2.13), (2.14)

\[
\lim_{k \to \infty} S^*(\hat{\theta}_k^{CT}, 0) = S^*(\theta, 0)
\]

(A.4)

From continuity and (A.2), (A.1) is established.

**Remark**

Without (3.8) satisfied, it cannot be guaranteed that \( S^*(\hat{\theta}_k^{CT}, \xi_k) \) exists for all \( k \). Also, with \( S^*(\hat{\theta}_k, \xi_k) \) large then \( \xi_k \) is large, and ill-conditioning is said to occur.

**Proof of Theorem 3.2**

Consider the central tendency adaptive control with (3.8) satisfied and define for all \( k \)

\[
[S_k^{CT}(n, m)]^{-1} = [S^*(\hat{\theta}_k^{CT}, \xi_k)]^{-1}
\]

(A.5)

where \( * \) denotes terms not of interest. Thus for the signal model with \( m > 1 \), or in other words the last two entries of \( \hat{\theta}_k^{CT} \) zero

\[
\hat{\theta}_k^{CT} = [S_k^{CT}(n, m)]^{-1} [\xi_k^{CT}]
\]

\[
[S_k^{CT}(n, m)]^{-1} [\xi_k^{CT}] = [S^*(\hat{\theta}_k^{CT}, \xi_k)]^{-1} [\xi_k^{CT}]
\]

(A.6)

Now applying (A.1) when taking limits as \( k \to \infty \)

\[
\lim_{k \to \infty} \hat{\theta}_k^{CT} = \lim_{k \to \infty} S^*(\hat{\theta}_k^{CT}, \xi_k) [I_{n+m-2} 0] \hat{\theta}_k^{CT}
\]

\[
\theta\to_0\lim_{k \to \infty} = S^*(\theta, 0) [I_{n+m-2} 0] \hat{\theta}_k^{CT}
\]

The desired results (3.10) (3.11) follow from application of Lemma 4.3.

---

![Fig. 3.1 Avoidance of ill-conditioning](https://via.placeholder.com/150)

**Fig. 3.1** Avoidance of ill-conditioning

![Fig 3.2 Ill-conditioned control](https://via.placeholder.com/150)

**Fig 3.2** Ill-conditioned control