INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING, VOL. 4, 389–414 (1990)

FACTORIZATIONS THAT RELAX THE POSITIVE REAL CONDITION IN CONTINUOUS-TIME AND FAST-SAMPLED ELS SCHEMES

VIKRAM KRISHNAMURTHY
Department of Systems Engineering, Australian National University, Canberra, ACT 2601, Australia

BO WAHLBERG
Department of Electrical Engineering, Linköping University, Linköping, Sweden

AND

JOHN B. MOORE
Department of Systems Engineering, Australian National University, Canberra, ACT 2601, Australia

SUMMARY

This paper proposes extended least-squares (ELS) for ARMAX model identification of continuous-time and certain discrete-time systems. The schemes have a relaxed strictly positive real (SPR) condition for global convergence. The relaxed SPR scheme is achieved by introducing overparametrization and prefiltering but without introducing ill-conditioning. The schemes presented are the first such proposed for continuous-time systems.

The concepts developed in continuous time carry through to fast-sampled continuous-time systems and associated discrete-time ELS algorithms. For such situations, in comparison with previously proposed discrete-time schemes, the degree of overparametrization required in the proposed scheme of this paper is significantly lower. The reduction is achieved by using more suitable prefiltering and overparametrization techniques than previously proposed.

We also establish the persistence of excitation (PE) of the regression vectors in the proposed ELS schemes to assure strong consistency, obtain convergence rates and provide robustness to unmodelled dynamics. To prove the PE of continuous-time regression vectors, we develop output reachability characterization for MIMO linear continuous-time systems.

KEY WORDS Extended least squares Identification Strictly positive real Persistent excitation Overparametrization Coloured noise

1. INTRODUCTION

There are two widely used classes of recursive identification schemes for linear stochastic systems. These are the recursive prediction error (RPE) and the extended least-squares (ELS) methods including their stochastic approximation versions. The RPE schemes require projection into a stability domain for convergence, and although attractive in open-loop stable system identification, cannot be used confidently in adaptive control. The ELS schemes require a strictly positive real (SPR) condition on a filtered coloured noise model for their convergence. The construction of the filter to achieve the SPR condition is in general more difficult than

This paper was recommended for publication by editor B. Wittenmark

0890–6327/90/050389–26$13.00

Received September 1989
Revised May 1990

projection into a stability domain for open-loop identification, and so perhaps renders the ELS approach less attractive to use then RPE methods in this situation. However, for adaptive control the ELS approach is the only approach known with guaranteed convergence results. (See Reference 1 for discrete-time and Reference 2 and 3 for continuous-time results.)

An obstacle towards guaranteeing convergence of an ELS scheme is selecting a filter to satisfy the SPR condition. In the usual discrete-time ARMAX model notation where the noise model is characterized in terms of a polynomial C (of degree n), a filter with transfer function W^{-1} must be chosen such that

$$\frac{W}{C}-\tfrac{1}{2} \text{ is strictly positive real (SPR)}$$

In discrete time, it is often the case the the prefilter $z^{-n}W(z)=1$ is chosen, and consequently the SPR condition is satisfied only when the noise is 'near' white. When there are deterministic disturbances such as constant biases, ramps and sinusoids, then inevitably $z^nC^{-1}(z)-\frac{1}{2}$ cannot be SPR, even where noise disturbances are white.

In continuous time the SPR condition is even more restrictive than in the corresponding discrete-time case. Taking $s^{-n}W(s)=1$, to correspond to the discrete-time example above, means that the SPR condition cannot be satisfied with any $s^{-n}C(s)$ other than a constant greater than $\frac{1}{2}$. Of course, in this case $W(s)$ is not asymptotically stable and is *a priori* not a reasonable prefilter to use. It turns out that selecting W is only straightforward if C is known, and otherwise is a formidable task.

Several modifications of the ELS algorithm have been proposed to relax the SPR condition (see References 4 and 5 and references therein for discrete-time results). In Reference 5 the SPR condition is side-stepped by transforming the discrete-time ARMAX model into an equivalent and unique overparametrized form. Then $\bar{C}^{-1}(z^{-1})\triangleq z^nC^{-1}(z)$ is expanded in the form $\mathscr{F}(z^{-1})+z^{-D}\mathscr{G}(z^{-1})\bar{C}^{-1}(z^{-1})$ for a suitably large delay D so that the relevant condition that $\mathscr{F}(z^{-1})\bar{C}^{-1}(z^{-1})-\frac{1}{2}$ be SPR is satisfied. A problem with this approach is that if the zeros of $C(z)$ lie in a region offset from the centre of the unit circle in the z-plane, then D is a large number and an unrealistically large number of parameters have to be estimated. The value of D is inevitably large, for example, in fast-sampled continuous-time systems[6] where the zeros of $C(z)$ lie close to $z=1$ in the unit circle; likewise in systems with a deterministic' disturbance such as a constant or ramp bias or sinusoidal disturbance.

This paper proposes ELS schemes with a relaxed SPR condition for ARMAX model identification of both continuous-time and discrete-time systems. The relaxed SPR scheme is achieved by introducing overparametrization and prefiltering but without introducing ill-conditioning. (Stochastic approximation versions can be treated similarly; however, details are omitted.) For discrete-time systems where the zeros of $C(z)$ lie in a region offset from the centre of the unit z-circle, we show that instead of expanding C in the delay operator, which is the existing approach,[5] using more suitable operators yields significantly lower-order regression vectors with fewer parameters to be estimated. One motivation to develop results for continuous-time schemes is their relevance for discrete-time schemes derived from fast-sampled continuous-time schemes. It is important to establish that no insurmountable problems arise should the sampling rate increase.

This paper is organized as follows. In Section 2 we first present the class of signal models of interest and then propose novel expansions and factorizations for C^{-1} for the continuous-time and discrete-time cases. In Section 3 we describe the continuous-time transformed ELS scheme and study its relaxed SPR property. In Section 4 we analyse the analogous discrete-time algorithm. We establish the persistence of excitation of the associated regression vectors to assure strong consistency of the identification schemes and therefore the robustness of the schemes to unmodelled dynamics. In Section 5 the convergence properties of the algorithm and a scheme for recovering the parameters are discussed. Some conclusions are drawn in Section 6.

2. SIGNAL MODELS AND FACTORIZATION OF $C^{-1}(s)$, $C^{-1}(q)$

Signal models

We first work with a continuous-time version of the ARMAX model:

$$\begin{gathered} A(p)y(t) = B(p)u(t) + C(p)e(t) \\ A(p) = p^n + a_1 p^{n-1} + \cdots + a_n, \qquad B(p) = b_1 p^{n-1} + \cdots + b_n \\ C(p) = p^n + c_1 p^{n-1} + \cdots + c_n \end{gathered} \tag{1}$$

Here p denotes the differentiation operator, $u(t)$ and $y(t)$ are the input and output signals respectively and $e(t)$ is a disturbance modelled here formally as 'white' noise. A more rigorous signal model than one driven by 'white' noise can be formulated using Ito equations:

$$\begin{aligned} d\phi(t) &= \mathscr{A}\phi(t)\, dt + \mathscr{B}u(t)\, dt + \mathscr{K}\, dv(t) \\ d\bar{y}(t) &= \mathscr{C}\phi(t)\, dt + dv(t) \end{aligned}$$

where $v(t)$ is a Wiener process. The measurable physical output of (1) is

$$\bar{y}(t) = \frac{1}{\Delta}\int_{t-\Delta}^{t} y(\tau)\, d\tau = \frac{1}{\Delta}\int_{t-\Delta}^{t} d\bar{y}(\tau), \quad \Delta > 0 \tag{2}$$

Notice that using $\bar{y}(t)$ avoids problems associated with the fact that since the degree of $C(p)$ is equal to that of $A(p)$, $y(t)$ in (1) includes a 'white' noise component. The actual structures of $\mathscr{A}$, $\mathscr{B}$, $\mathscr{K}$ and $\mathscr{C}$ are not important at this stage.

We will also develop results for the following discrete-time ARMAX model:

$$\begin{gathered} A(q)y(k) = B(q)u(k) + C(q)e(k) \\ A(q) = q^l + a_1 q^{l-1} + \cdots + a_l, \qquad B(q) = b_1 q^{l-1} + b_2 q^{l-2} + \cdots + b_m q^{l-m}, \quad m \leqslant l \\ C(q) = q^n + c_1 q^{n-1} + \cdots + c_n \end{gathered} \tag{3}$$

where q is the forward shift operator and $e(k)$ is a discrete white noise process.

An important case where existing relaxed ELS schemes are inefficient is when (3) is obtained by fast integrated sampling of the continuous-time process (1). It is proved in Reference 7 that for small sampling intervals, all the zeros of resulting $C(q)$, which without loss of generality lie inside the unit circle, tend to $z = 1$ exponentially fast. An example of this is a deterministic output error system $A(q)z(k) = B(q)u(k)$ with measurements $y(k)$ of $z(k)$ contaminated by added noise $e(k)$, i.e. $y(k) = z(k) + e(k)$. Then (3) holds with

$$A(q) = C(q), \qquad z(k) \triangleq y(k) - e(k), \qquad A(q)z(k) = B(q)u(k) \tag{4}$$

Proposed factorization for continuous-time models

Here we first develop factorizations for continuous-time models. Then corresponding discrete-time models are derived. The following key lemma proposes an expansion of $C^{-1}(p)$ which will be exploited subsequently.

Lemma 1

With $C(p)$ defined in (1), assume $C^{-1}(p)$ to be asymptotically stable. Then in Laplace transform notation, for some real converging sequence r_k and any $a > 0$,

$$\frac{1}{C(s)} = \frac{1}{(s+a)^n} \sum_{k=0}^{\infty} r_k \left(\frac{s-a}{s+a}\right)^k, \quad \mathrm{Re}(s) \geqslant 0 \tag{5}$$

Proof. Consider the bilinear transformation

$$z = \frac{s+a}{s-a} \Leftrightarrow s = a\,\frac{z+1}{z-1} \tag{6}$$

Setting

$$P(z) = C^{-1}\left(a\,\frac{z+1}{z-1}\right)$$

then $P(z)$ is analytic outside the unit circle since (6) maps the left half-plane into the unit disc. Furthermore, $P(z)$ has precisely n zeros at $z = 1$. This is because the n zeros of $C^{-1}(s)$ at $s = \infty$ map precisely to the n zeros of $P(z)$ at $z = 1$. Hence for some real converging sequence p_k,

$$P(z) = (1 - z^{-1})^n \sum_{k=0}^{\infty} p_k z^{-k}, \quad |z| \geqslant 1 \tag{7}$$

Substituting back $z = (s+a)/(s-a)$ now proves the lemma. □

Remark. The expansion (5) is closely related to Laguerre function representations; see Reference 6 for details.

Corollary 1

The transfer function $C^{-1}(s)$ in Lemma 1 can be uniquely factorized as

$$\frac{1}{C(s)} = \frac{F(s)}{L(s)} + \frac{G(s)H(s)}{L(s)C(s)} \Leftrightarrow L(s) = F(s)C(s) + G(s)H(s) \tag{8a}$$

where $H(s) = (s-a)^{N-n+1}$, $L(s) = (s+a)^N$, $F(s)$ is of degree $N-n$ and $G(s)$ is of degree $n-1$, i.e.

$$F(s) = f_{N-n} + \cdots + f_0 s^{N-n}, \qquad G(s) = g_{n-1} + \cdots + g_0 s^{n-1} \tag{8b}$$

Furthermore, given by $\varepsilon > 0$, there exists $N \geqslant n$ such that

$$\left\| \frac{G(s)H(s)}{L(s)C(s)} \right\|_\infty \leqslant \varepsilon \tag{9}$$

where for any function f, $\| f(s) \|_\infty \triangleq \sup_w | f(s) |_{s=jw}$.

Proof. From Lemma 1

$$\frac{1}{C(s)} = \frac{1}{(s+a)^n} \sum_{k=0}^{N-n} r_k \left(\frac{s-a}{s+a}\right)^k + \frac{1}{(s+a)^n} \sum_{k=N-n+1}^{\infty} r_k \left(\frac{s-a}{s+a}\right)^k, \quad \mathrm{Re}(s) \geqslant 0$$

The first term on the right-hand side equals $F(s)L^{-1}(s)$. The second term can be made arbitrarily small by choosing N sufficiently large. It equals $H(s)L^{-1}(s)$ times a strictly proper transfer function; this strictly proper transfer function has its poles at the zeros of $C(s)$. □

Remark 1. The choice of a has a significant effect on the size of N that satisfies (8) and (9). Let z_i, $i=1,\dots,n$, denote the zeros of $C(s)$. Then p_k in (7) is of order $O(\max_i |(z_i+a)/(z_i-a)|^k)$. Hence to obtain a fast convergence rate in the series expansion, the value of $-a$ should be chosen close to the zeros of $C(s)$. Also note that choosing a too large or small results in a slow convergence rate. We give a comprehensive design rule for selecting N and a in Section 3.

Remark 2. From Remark 1 it follows that a large value of N must be chosen if the zeros of $C(s)$ are scattered. This can be circumvented if $C^{-1}(s)$ is expanded around several as, i.e. use

$$L(s)=\prod_{i=1}^{m}(s+a_i)^{N_i},\qquad \sum_{i=1}^{m}N_i=N$$

$$H(s)=\prod_{i=1}^{m}(s-a_i)^{M_i},\qquad \sum_{i=1}^{m}M_i=N-n+1$$

in (8).

Remark 3. Since $F(s)$ is in general non monic, we shall in the sequel work with the monic polynomial

$$\bar{F}(s)\triangleq F(s)/f_0 \tag{10}$$

Note that by equating the coefficients of the s^n terms in (8a), we have $f_0+g_0=1$.

Proposed factorization for discrete-time models

We now give the discrete-time versions of Lemma 1 and Corollary 1.

Lemma 2

With $C(q)$ defined in (3), assume $C^{-1}(q)$ to be asymptotically stable. Then for some real converging sequence r_k,

$$\frac{1}{C(q)}=\frac{1}{(q-a)^n}\sum_{k=0}^{\infty}r_k\left(\frac{1-aq}{q-a}\right)^k,\quad |q|\geqslant 1 \tag{11}$$

for $|a|<1$.

Proof. We use the transformation

$$z=\frac{q-a}{1-aq}\Leftrightarrow q=\frac{z+a}{1+az} \tag{12}$$

Setting

$$P(z)=C^{-1}\left(\frac{z+a}{1+az}\right)$$

then $P(z)$ is analytic outside the unit circle since (12) maps the unit disc to itself. Furthermore, $P(z)$ has precisely n zeros at $z^{-1} = -a$. Then for some real converging sequence p_k,

$$P(z) = (z^{-1} + a)^n \sum_{k=0}^{\infty} p_k z^{-k} \tag{13}$$

Substituting back $z = (q - a)/(1 - aq)$ now proves the lemma. □

Corollary 2

The transfer function $C^{-1}(q)$ in Lemma 2 can be uniquely factorized as

$$\frac{1}{C(q)} = \frac{F(q)}{L(q)} + \frac{G(q)H(q)}{L(q)C(q)} \Leftrightarrow L(q) = F(q)C(q) + G(q)H(q) \tag{14a}$$

where $H(q) = (1 - aq)^{N-n+1}$, $L(q) = (q - a)^N$, $F(q)$ is of degree $N - n$ and $G(q)$ is of degree $n - 1$, i.e.

$$F(q) = f_{N-n} + \cdots + f_0 q^{N-n}, \qquad G(q) = g_{n-1} + \cdots + g_0 q^{n-1} \tag{14b}$$

Furthermore, given any $\varepsilon > 0$, there exists $N \geqslant n$ such that

$$\left\| \frac{G(z)H(z)}{L(z)C(z)} \right\|_{\infty} \leqslant \varepsilon \tag{15}$$

where for any function $f(z)$, $\| f(z) \|_{\infty} \triangleq \sup_w | f(z) |_{z=e^{jw}}$. □

Proof. The proof is similar to that of Corollary 1.

Remark. The discrete-time versions of Remarks 1, 2 and 3 (following Corollary 1) hold. If z_i, $i = 1, \dots, n$, denote the zeros of $C(q)$, then p_k in (13) is of order $O(\max_i | (z_i - a)/(1 - az_i) |^k)$. Hence to obtain a fast convergence rate the value of a should be chosen close to the zeros of $C(q)$. Also, if the zeros of $C(q)$ are scattered within the unit disc, instead of choosing N large, $C^{-1}(q)$ can be expanded around several as. We give a comprehensive design rule for selecting N and a in Section 4.

In the sequel we shall find it convenient to work with the monic polynomial

$$\bar{F}(q) = F(q)/f_0 \tag{16}$$

Note that from (14a) we have $f_0 + (-a)^{N-n+1} g_0 = 1$.

In the rest of this paper we shall implicitly assume $| a | < 1$ in discrete-time results and $a > 0$ in continuous-time results.

Rapprochement with the factorization in Reference 5

Here we seek relationships for our proposed factorizations (8) and (14) with the unique factorization in Reference 5:

$$\frac{1}{\bar{C}(z^{-1})} = \mathscr{F}(z^{-1}) + z^{-(N-n+1)} \frac{\mathscr{G}(z^{-1})}{\bar{C}(z^{-1})} \tag{17a}$$

where $\bar{C}(z^{-1})$ and $\mathscr{F}(z^{-1})$ are monic polynomials of degree n and $N - n$ respectively and

$\mathscr{G}(z^{-1})$ is a polynomial of degree $(n-1)$, i.e.

$$\mathscr{F}(z^{-1}) = 1 + \sum_{i=1}^{N-n} \delta_i z^{-i}, \qquad \mathscr{G}(z^{-1}) = \sum_{i=0}^{n-1} \gamma_i z^{-i} \tag{17b}$$

We shall exploit these relationships in Sections 3 and 4 to propose rules for selecting a suitable N and a in our ELS schemes.

1. Continuous time. The following lemma shows the equivalence of our unique continuous-time factorization (8) and the unique factorization (17a).

Lemma 3

The factorizations (17) and (8) are equivalent under the bilinear transformation (6) and the following definitions of $\bar{C}$, $\mathscr{F}$ and $\mathscr{G}$ in terms of C, F and G or vice versa:

$$\frac{C(s)}{C(a)} = \frac{(s+a)^n}{(2a)^n}\,\bar{C}\left(\frac{s-a}{s+a}\right), \qquad G(s) = (s+a)^{n-1}\mathscr{G}\left(\frac{s-a}{s+a}\right)$$
$$F(s) = \frac{(2a)^n}{C(a)}\,(s+a)^{N-n}\mathscr{F}\left(\frac{s-a}{s+a}\right) \tag{18a}$$

or equivalently,

$$\bar{C}(z^{-1}) = \frac{(1-z^{-1})^n}{C(a)}\,C\left(a\,\frac{z+1}{z-1}\right), \qquad \mathscr{G}(z^{-1}) = \frac{(1-z^{-1})^{n-1}}{(2a)^{n-1}}\,G\left(a\,\frac{z+1}{z-1}\right),$$
$$\mathscr{F}(z^{-1}) = \frac{(1-z^{-1})^{N-n}}{(2a)^N}\,C(a)F\left(a\,\frac{z+1}{z-1}\right) \tag{18b}$$

Proof. We first show that (17) transforms to (8) under (6) and (18). Multiplying (17) by $(1-z^{-1})^n$ we have

$$\frac{(1-z^{-1})^n}{\bar{C}(z^{-1})} = (1-z^{-1})^n\mathscr{F}(z^{-1}) + (1-z^{-1})^n z^{-(N-n+1)}\,\frac{\mathscr{G}(z^{-1})}{\bar{C}(z^{-1})}$$

Substituting (18b) and (6) leads to

$$\frac{C(a)}{C(s)} = C(a)\,\frac{F(s)}{(s+a)^N} + \left(\frac{s-a}{s+a}\right)^{N-n+1}\frac{G(s)}{(s+a)^{n-1}}\,\frac{C(a)}{C(s)}$$

which yields (8). □

The converse holds likewise.

Remark. By its definition in (18), $\bar{C}(z^{-1})$ is a monic polynomial because $C(s)$ is monic. Also from (18a), with γ_i and δ_i defined in (17b), simple manipulations yield

$$f_0 = \frac{(2a)^n}{C(a)}\,(1+\delta_1+\cdots+\delta_{N-n}), \qquad g_0 = \gamma_0+\gamma_1+\cdots+\gamma_{n-1} \tag{19}$$

2. Discrete time. Here we seek a relationship between the unique factorizations (14a) and (17). Note that our factorization (14a) specializes to (17) when $a=0$ and our proposed ELS scheme then specializes to that in Reference 5.

Lemma 4

The factorizations (14a) and (17) are equivalent under the transformation (12) and the following definitions of $\bar{C}$, $\mathcal{F}$ and $\mathcal{G}$ in terms of C, F and G or vice versa:

$$\bar{C}(z^{-1}) = \frac{(1-a^2)^n}{\sum_{i=0}^{n} c_i a^i} \frac{C(q)}{(q-a)^n}, \qquad \mathcal{G}(z^{-1}) = \frac{G(q)}{(q-a)^{n-1}}$$
$$\mathcal{F}(z^{-1}) = \frac{\sum_{i=0}^{n} c_i a^i}{(1-a^2)^n} \frac{F(q)}{(q-a)^{N-n}}, \qquad c_0 = 1 \tag{20a}$$

or

$$C\left(\frac{z+a}{1+az}\right) = \frac{(\sum_{i=1}^{n} c_i a^i)\bar{C}(z^{-1})}{(z^{-1}+a)^n}, \qquad G\left(\frac{z+a}{1+az}\right) = (1-a^2)^{n-1} \frac{\mathcal{G}(z^{-1})}{(z^{-1}+a)^{n-1}}$$
$$F\left(\frac{z+a}{1+az}\right) = \frac{(1-a^2)^N}{\sum_{i=0}^{n} c_i a^i} \frac{\mathcal{F}(z^{-1})}{(z^{-1}+a)^{N-n}} \tag{20b}$$

Proof. We first show that (17) transforms to (14a) under (20) and (12). Substituting (20a) into (17) are dividing the resulting equation by $(q-a)^n$ directly yields (14a). The converse holds likewise. □

Remark. By its definition in (20), $\bar{C}(z^{-1})$ is a monic polynomial because $C(q)$ is monic. Also from (20a), with γ_i and δ_i defined in (17b), simple manipulations yield

$$f_0 = \frac{(1-a^2)^n}{\sum_{i=0}^{n} c_i a^i}(1 - a\delta_i + \cdots + (-a)^{N-n}\delta_{N-n}), \qquad g_0 = \gamma_0 - a\gamma_1 + \cdots + (-a)^{n-1}\gamma_{n-1} \tag{21}$$

3. CONTINUOUS-TIME ELS SCHEME WITH RELAXED SPR CONDITION

In this section we first derive a transformed ELS scheme and then interpret its associated SPR condition. We then develop conditions for the PE of the regression vectors and determine the convergence rates of the scheme. Finally we show that a companion least-squares scheme can be used to recover the original parameters.

Time domain equations

Let us consider a filtering operation on (1) in terms of the exponentially stable filter

$$\frac{1}{W(s)} = \frac{F(s)}{L(s)} \tag{22}$$

According to Corollary 1, (22) is a good approximation of $C^{-1}(s)$ provided N is large enough. However, because $F(s)$ is unknown, we do not use (22) in the actual implementation of the estimation scheme. Applying the filter (22) with the normalized $\bar{F}(s)$ defined in (10) replacing $F(s)$, (1) becomes

$$A(p)\frac{\bar{F}(p)}{L(p)}y(t) = B(p)\frac{\bar{F}(p)}{L(p)}u(t) + C(p)\frac{\bar{F}(p)}{L(p)}e(t)$$

or equivalently, using (8),

$$p^{-N}\bar{A}(p)y_{LN}(t) = p^{-N}\bar{B}(p)u_{LN}(t) + \bar{g}_0 e(t) - p^{-(n-1)}\bar{G}(p)e_H(t) + e(t) \tag{23}$$

under the following definitions:

$$\begin{gathered} y_{LN}(t) = p^N L^{-1}(p)y(t), \qquad u_{LN}(t) = p^N L^{-1}(p)u(t) \\ e_H(t) = p^{n-1}H(p)L^{-1}(p)e(t), \qquad \bar{A}(p) = A(p)\bar{F}(p), \qquad \bar{B}(p) = B(p)\bar{F}(p) \\ \bar{G}(p) = G(p)/f_0 \quad \text{where } \bar{G}(p) = \bar{g}_{n-1} + \cdots + \bar{g}_0 s^{n-1} \end{gathered} \tag{24}$$

We now estimate the coefficients of $\bar{A}(p)$, $\bar{B}(p)$ and $\bar{G}(p)$ as $\hat{\bar{A}}(p)$, $\hat{\bar{B}}(p)$ and $\hat{\bar{G}}(p)$ using ELS. To formulate the ELS scheme, let us consider the more precisely defined stochastic Ito form state equations.[2] With $\varepsilon(t) \triangleq e_H(t) - e(t)$, consider (23) reformulated as

$$\begin{aligned} \mathrm{d}x(t) &= Ax(t)\,\mathrm{d}t - e_1\,\mathrm{d}y_{LN}^{(1)}(t) + e_{N+1}u_{LN}(t)\,\mathrm{d}t - e_{2N+1}\,\mathrm{d}\varepsilon(t) - e_{2N+2}\,\mathrm{d}v(t) \\ \mathrm{d}\bar{y}_{LN}(t) &= \theta_N^{\mathrm{T}}x(t)\,\mathrm{d}t + \mathrm{d}v(t), \qquad x(0) = 0 \end{aligned} \tag{25a}$$

where $e_i^{\mathrm{T}} = (0 \ldots 0\ 1\ 0 \ldots 0)$ with the 1 in the ith position, $\bar{y}_{LN}(t)$ is defined similar to (2) and

$$\begin{aligned} x(t) &= (-y_{LN}^{(1)}(t) \ldots -y_{LN}^{(N)}(t)\ u_{LN}^{(1)}(t) \ldots u_{LN}^{(N)}(t)\ \varepsilon(t)\ -e_H^{(1)}(t) \ldots -e_H^{(n-1)}(t))^{\mathrm{T}} \\ \theta_N &= (\bar{a}_1 \ldots \bar{a}_N\ \bar{b}_1 \ldots \bar{b}_N\ \bar{g}_0 \ldots \bar{g}_{n-1})^{\mathrm{T}} \end{aligned} \tag{25b}$$

where $y_{LN}^{(i)}(t) = \int_0^t y_{LN}^{(i-1)}(\tau)\,\mathrm{d}\tau$, $y_{LN}^{(1)}(t) \triangleq \int_0^t y_{LN}(\tau)\,\mathrm{d}\tau$ and $u_{LN}^{(i)}(t)$ and $e_H^{(i)}(t)$ are defined likewise. Also $v(t)$ is defined in the Ito formulation of (1) and

$$A = \text{block diag}(E_N, E_N, E_n), \qquad E_i = \begin{bmatrix} 0 & 0 \\ I_{i-1} & 0 \end{bmatrix} \tag{25c}$$

Note that the components of $x(t)$ are measurable.

Transformed ELS scheme

Consider the ELS estimation of $\theta_N(t)$:

$$\begin{gathered} \mathrm{d}\hat{x}(t) = A\hat{x}(t)\,\mathrm{d}t - e_1\,\mathrm{d}y_{LN}^{(1)}(t) + e_{N+1}u_{LN}(t)\,\mathrm{d}t - e_{2N+1}\,\mathrm{d}\hat{\varepsilon}(t) - e_{2N+2}\,\mathrm{d}\hat{v}(t) \\ \mathrm{d}\hat{v}(t) \triangleq \mathrm{d}\bar{y}_{LN}(t) - \hat{\theta}_N^{\mathrm{T}}(t)\hat{x}(t)\,\mathrm{d}t, \qquad \mathrm{d}\hat{\theta}_N(t) = \hat{P}_t\hat{x}(t)\,\mathrm{d}\hat{v}(t) \\ \mathrm{d}P_t^{-1} = \hat{x}(t)\hat{x}^{\mathrm{T}}(t)\,\mathrm{d}t, \qquad \mathrm{d}\hat{P}_t = -\hat{P}_t\hat{x}(t)\hat{x}(t)^{\mathrm{T}}\hat{P}_t\,\mathrm{d}t, \qquad P_0 > 0 \end{gathered} \tag{26}$$

suitably initialized with $\hat{x}(0)$, $\hat{\theta}_N(0)$ and some $P_0 > 0$. The state estimate $\hat{x}(t)$ above is defined by (25b) with $e^{(\cdot)}(t)$, $e_H^{(i)}(t)$ replaced by $\hat{e}^{(\cdot)}(t)$, $\hat{e}_H^{(i)}(t)$. Also $\hat{e}_H^{(i)} \triangleq p^{n-1}H(p)L^{-1}(p)\hat{e}(t)$, $\hat{\varepsilon}(t) \triangleq \hat{e}_H(t) - \hat{e}(t)$ and $\hat{P}_t^{-1} = \int_0^t \hat{x}^{\mathrm{T}}(\tau)\,\mathrm{d}\tau + \hat{P}_0^{-1}$.

It can be shown[2] that a sufficient condition for the ELS scheme to converge is that

$$\frac{L(s)}{\bar{F}(s)C(s)} - \tfrac{1}{2} \text{ is SPR}$$

or equivalently (easily shown by using (8) above and equation (4.160) in Reference 8,

$$\left\| \frac{\bar{G}(s)H(s)}{L(s)} - \bar{g}_0 \right\|_\infty < 1 \quad \text{(strictly bound real (SBR) condition)} \tag{27}$$

Overparametrization selection to satisfy SPR condition

From Corollary 1 we know that

$$\left\| \frac{G(s)H(s)}{L(s)C(s)} \right\|_\infty \tag{28}$$

can be made arbitrarily small by choosing N large enough. By restricting the zeros of $C(s)$ to lie inside a given *a priori* compact set, it is possible to specify N and a such that (27) is satisfied for all $C(s)$ whose zeros lie inside the set. In this subsection we specify N, a and the compact set. We seek to do so by achieving a continuous-time version of the discrete-time result in Reference 5.

Lemma 5

Consider any polynomial (in z^{-1})

$$\bar{C}(z^{-1}) = \sum_{i=0}^{n} \bar{c}_i z^{-i} = \prod_{i=1}^{n} (1 - z_i z^{-i}) \tag{29}$$

with $\bar{c}_0 = 1$ such that $|z_i| \leqslant R < 1$ for all i. Consider also for any N, a polynomial pair $(\mathscr{F}(z^{-1}), \mathscr{G}(z^{-1}))$ with degrees $N-n$ and $n-1$ respectively, defined uniquely by the factorization (17). Then there exists $N_0(R)$ such that for all $N \geqslant N_0(R)$, $\mathscr{G}(z^{-1})$ is SBR.

Proof. See Reference 5. □

Notice that under (6) and (18), $z^{-(N-n+1)}\mathscr{G}(z^{-1})$ transforms to $G(s)H(s)L^{-1}(s)$. So

$$\left| \frac{G(s)H(s)}{L(s)} \right|_{s=jw} = |\mathscr{G}(z^{-1})|_{|z|=1} \tag{30a}$$

Moreover, since stability is preserved under the bilinear transformation,

$$\frac{g_0}{f_0} - \frac{1}{f_0}\frac{G(s)H(s)}{L(s)} \text{ SBR} \Leftrightarrow \frac{g_0 - \mathscr{G}(z^{-1})}{f_0} \text{ SBR} \tag{30b}$$

□

We now present the continuous-time version of lemma 5.

Lemma 6

Consider a polynomial $C(s) = \Pi_{i=1}^{n}(s - s_i)$, $s_i < 0$, such that its zeros lie in a circle with centre $x_0 = -a(1+R^2)/(1-R^2)$ on the real axis and radius

$$r = \frac{2aR}{1-R^2} \tag{31a}$$

or equivalently such that

$$|s_i - x_0| \leqslant r, \qquad R < 1 \qquad \forall i \tag{31b}$$

Consider also for any N, the polynomials $F(s)$, $G(s)$ and $H(s)$ are uniquely defined as in (8). Then there exists $N_0(r)$ such that for all $N \geqslant N_0(r)$,

$$\left| \frac{g_0}{f_0} \right| + \left| \frac{1}{f_0}\frac{G(s)H(s)}{L(s)} \right|_{s=jw} < 1 \tag{32}$$

and the SBR condition (27) is satisfied.

Proof. Using (6) with r defined in (31) and z_i defined in Lemma 5, straightforward manipulations yield

$$|z_i| \leqslant R < 1 \Leftrightarrow \left| s_i + a\,\frac{1+R^2}{1-R^2} \right| \leqslant r$$

Since (27) is implied by (32), we shall look for upper bounds on $|g_0/f_0|$ and

$$\left| \frac{G(s)H(s)}{f_0 L(s)} \right|_{s=jw} = \left| \frac{\mathscr{G}(z^{-1})}{f_0} \right|_{|z|=1}$$

It is proved in Reference 5 that

$$|g_0| \leqslant \sum_{i=0}^{n-1} |\gamma_i| < R^{N-n+1}\,\frac{(N+n)!}{N!}\,(1+R)^n \triangleq f(N)$$

and thus $|\mathscr{G}(z^{-1})|_{|z|=1} < f(N)$. Also, it is easily shown that $f(N)$ is monotonic decreasing if

$$N \geqslant \frac{R(1+n)-1)}{1-R} \tag{33a}$$

In addition, a minimum bound for $|f_0|$ can be obtained as follows. If N is chosen sufficiently large so that $f(N) < 1$, then since $f_0 + g_0 = 1$, $|f_0| \geqslant 1 - f(N)$. So

$$\left| \frac{g_0}{f_0} \right| + \left| \frac{\mathscr{G}(z^{-1})}{f_0} \right|_{|z|=1} < \frac{2f(N)}{1-f(N)}$$

Hence for any N, if $2f(N)/(1-f(N)) < 1$, i.e.

$$3R^{N-n+1}\,\frac{(N+n)!}{N!}\,(1+R)^n < 1 \tag{33b}$$

and (33a) holds, then the SBR condition (27) is satisfied. Note that $N_0(r)$ can be defined as the smallest value of N for which (33) holds.

On persistence of excitation using output controllability characterization

In Reference 9 the regression vectors of discrete-time ELS schemes are shown to be persistently exciting, leading to consistent parameter estimation using output reachability characterizations for MIMO systems. These characterizations translate excitation properties of system inputs to excitation properties of regression vectors. Here we develop continuous-time versions of the results of Reference 9 by first generalizing certain continuous-time results of Reference 10 for single-input systems to multi-input systems. The notation used in this subsection is independent of the rest of the paper.

Excitation and output controllability. Consider the MIMO continuous-time linear system

$$\begin{gathered} \dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0 \\ y(t) = Cx(t) \\ A \in \mathbb{C}^{\alpha\times\alpha}, \qquad B \in \mathbb{C}^{\alpha\times\beta}, \qquad C \in \mathbb{C}^{\gamma\times\alpha}, \qquad u \in L_e^1(\mathbb{C}^\beta) \end{gathered} \tag{34a}$$

and associated proper transfer function

$$T(s) = C(sI - A)^{-1}B \tag{34b}$$

Note. $L_e^1(\mathbb{C}^n)$ is the space of functions $f: \Re^+ \to \mathbb{C}^n$ which are Lebesgue-integrable over any finite interval. $L_e^2(\mathbb{C}^n)$ is the space of functions as above which are also square-Lebesgue-integrable over any finite interval.

Definition. The system (34a) will be called output-controllable[11] (equivalently y is controllable from u) at time t_0 if for any y_1 there exists a finite $t_1 > t_0$ and an input $u[t_0, t_1]$ that transfers the output from $y(t_0) = 0$ to $y(t_1) = y_1$.

Let $\mathrm{OC}(\alpha, \beta, \gamma)$ denote the class of output-controllable systems specified by (34). Let v be the McMillan degree of the system (34a) and let $M_1, \ldots, M_v$ be its Markov parameters. Define the matrix M by

$$M = [M_0 \; M_1 \ldots M_v] = [CB \; CAB \ldots CA^{v-1}B]$$

Recall[11] that M has full rank iff the system is output-controllable.

Lemma 7

The following statements are equivalent.

(a) The system (34) is output-controllable, i.e. $y(t)$ is controllable from $u(t)$.
(b) M has full rank.
(c) $T(z)$ has full rank over $\Re$.

Proof. The proof follows from straightforward continuous-time generalizations of results in Reference 9. □

We now find conditions on the system input under which the output is sufficiently rich for adaptive identification purposes. We closely follow the notation and results in Reference 10 and generalize them for the MIMO case.

Definition 1. The function $y \in L_e^2(\mathbb{C}^\gamma)$ is said to be *sufficiently rich* iff there exist positive constants ε_1, T such that for all $\tau \geqslant 0$,

$$\int_\tau^{T+\tau} y(t) y^*(t) \, \mathrm{d}t > \varepsilon_1 I_\gamma$$

T is termed the excitation period of y. (y^* is the Hermitian transpose of y.)

Definition 2. The input $u \in L_e^1(\mathbb{C}^\beta)$ is said to be persistently exciting for the class $\mathrm{OC}(\alpha, \beta, \gamma)$ iff for any system in $\mathrm{OC}(\alpha, \beta, \gamma)$ is produces a sufficiently rich output y uniformly in y_0 (i.e. ε_1 and T in Definition 1 are independent of the initial condition y_0).

We now proceed to define a *richness* property of the input that will characterize the class of $\mathrm{OC}(\alpha, \beta, \gamma)$ inputs. Let $f_\tau(t) = f(t + \tau)$ denote the translation of a function f along the real axis. Using the notation

$$(I^n u_\tau)(t) = \int_0^t \mathrm{d}\sigma_1 \int_0^{\sigma_1} \mathrm{d}\sigma_2 \ldots \int_0^{\sigma_{n-1}} \mathrm{d}\sigma_n \, u(\sigma_n + \tau), \qquad (Iu_\tau)(t) = \int_0^t f(\sigma_1 + \tau) \, \mathrm{d}\sigma_1$$

we define

$$V_t(t) \triangleq [Iu_\tau^T, \ldots, I^\alpha u_\tau^T]^T(t), \qquad W_\tau(\mathcal{M}, t) \triangleq V_\tau(t) + \mathcal{M}\theta(t)$$

where $\mathcal{M} \in \mathbb{C}^{\alpha\beta\times\alpha}$ is some constant matrix and $\theta(t) \triangleq [1 \ t \ t^2 \dots t^{\alpha-1}]^T$. Also define

$$J_\tau(\mathcal{M}, T) = \int_0^T W_\tau(\mathcal{M}, t) W_\tau^*(\mathcal{M}, t) \, \mathrm{d}t$$

Definition 3. The input $u \in L_e^1(\mathbb{C}^\beta)$ is said to be rich of order α iff there exist positive constants ε_3, T such that for all $\tau \geqslant 0$ and $\forall C \in \mathbb{C}^{\gamma\times\alpha}$

$$CJ_\tau(\mathcal{M}_\tau, T)C^* \geqslant \varepsilon_3 \lambda_{\min}(CC^*)I \quad \forall \tau \geqslant 0$$

where $\mathcal{M}_\tau \triangleq -(\int_0^T [V_\tau(t)\theta^T(t)\,\mathrm{d}t)N_0^{-1}$, $N_0 \triangleq \int_0^T \theta(t)\theta^T(t)\,\mathrm{d}t$. (See Reference 10 for the motivation of this definition.)

Theorem 1

The input $u \in L_e^1(\mathbb{C}^\beta)$ is persistently exciting for the class $\mathrm{OC}(\alpha, \beta, \gamma)$ iff it is rich of order α.

Proof. The proof is the same as that in Reference 10 except that the corresponding MIMO terms defined above must be used. □

Excitation of regression vectors. Theorem 1 reduces the question of the transfer of excitation from inputs to outputs to one of controllability of outputs from the inputs. In ELS estimation we require the regression vectors to be exciting. Thus we develop conditions for these regression vectors to be reachable. With Theorem 1 established, the results are straightforward generalizations of the discrete-time results in Reference 9.

Let $T(s)$ be a $\gamma \times \beta$ proper rational matrix with Markov expansion $T(s) = \Sigma_{i=0}^{\infty} M_i s^{-i}$. Define $T^i(s)$ as the ith row of $T(s)$ and $y_i(t)$ as the ith component of $y(t)$. For arbitrary integers $l_i \geqslant 1$ we define the general regression vector

$$\varphi_{l_1 \dots l_\gamma}(t) = [y_1^{(1)}(t) \dots y_1^{(l_1-1)}(t) \ \ y_2^{(1)}(t) \dots y_2^{(l_2-1)}(t) \dots y_\gamma^{(1)}(t) \dots y_\gamma^{(l_\gamma-1)}(t)]$$

where $y^{(i)}(t) = \int_0^t y^{(i)}(\tau)\,\mathrm{d}\tau$ and $y^{(1)}(t) = \int_0^t y(\tau)\,\mathrm{d}\tau$. Define

$$T_{l_1 \dots l_\gamma}(s) = [T^1(s)^{\mathrm{T}} \ s^{-1}T^l(s)^{\mathrm{T}} \dots s^{-l_1+1}T^1(s)^{\mathrm{T}} \dots T^\gamma(s)^{\mathrm{T}} \dots s^{-l_\gamma+1}T^\gamma(s)^{\mathrm{T}}]^{\mathrm{T}}$$
$$P(l_1 \dots l_\gamma) = \{\mathcal{F}(s) \in P^\gamma, \ \mathcal{F}(s) = [\mathcal{F}_1(s) \dots \mathcal{F}_\gamma(s)]^{\mathrm{T}}, \deg \mathcal{F}_i(s) \leqslant l_i - 1\}$$
$$\mathcal{H}(s) = \begin{bmatrix} s^{l-l_1}T^1(s) \\ \vdots \\ s^{l-l_\gamma}T^\gamma(s) \end{bmatrix}, \qquad l = \min\{l_i\} \tag{35}$$

Theorem 2

With the definitions (35), the following are equivalent.

(a) $\varphi_{l_1 \dots l_\gamma}(t)$ is reachable from $u(t)$.
(b) $T_{l_1 \dots l_\gamma}(s)$ is full row rank over $\mathbb{C}$.
(c) $P(l_1 \dots l_\gamma) \cap N(\mathcal{H}) = 0$, the zero polynomial, and $N(\mathcal{H})$ denotes the left null space of $\mathcal{H}(s)$.

Proof. The proof is a straightforward generalization of the discrete-time proof in Reference 9. □

Persistence of excitation of $x(t)$

We require the regression vector $x(t)$ and its estimate $\hat{x}(t)$ in (25) and (26) to be suitably exciting to assure strong consistency of the identification scheme and therefore its robustness to unmodelled dynamics.[12]

Lemma 8

A necessary and sufficient condition for $x(t)$ in (25b) associated with signal model (25a) to be controllable from inputs $u(t)$, $v(t)$ is that $A(s)K(s)$, $B(s)K(s)$ and $C(s)$ are coprime where $K(s) \triangleq H(s) - (s+a)^{N-n+1}$. ($A(s), B(s), C(s)$ are defined in (1) and $H(s)$ in (8).)

Proof. Because of the similar nature of the continuous-time PE conditions in Theorem 2 and the discrete-time PE conditions,[9] we provide the discrete-time proof in Section 4, Lemma 10. With $F(s)$, $G(s)$, $H(s)$ and $K(s)$ defined as above, replacing z by s, l and m by n in Lemma 10 yields the continuous-time proof. □

Convergence of modified continuous-time ELS scheme

Standard techniques[13] apply to achieve convergence properties of the transformed ELS scheme. We summarize as the following theorem.

Theorem 3

Consider the overparametrized signal (23), (24) with Ito form (25) and ELS estimation scheme (26). If N is chosen sufficiently large so that the SBR condition (27) is satisfied, then as $t \to \infty$

$$\| \theta_N - \hat{\theta}_N(t) \|^2 = O\left(\frac{\log \lambda_{\max} \hat{P}_t^{-1}}{\lambda_{\min} \hat{P}_t^{-1}}\right) \quad \text{a.s.} \tag{36a}$$

$$\| v(t) \|^2 = O(\log \lambda_{\max} \hat{P}_t^{-1}) \quad \text{a.s.} \tag{36b}$$

Moreover, with $u(t)$, $v(t)$ suitably exciting, i.e.

$$\lim_{t \to \infty} \frac{\log \lambda_{\max} P_t^{-1}}{\lambda_{\min} P_t^{-1}} = 0 \quad \text{a.s.} \tag{37}$$

then as $t \to \infty$

$$\| \theta_N - \hat{\theta}_N(t) \| = O\left(\frac{\log \lambda_{\max} P_t^{-1}}{\lambda_{\min} P_t^{-1}}\right) \quad \text{a.s.} \tag{36c}$$

$$\| v(t) \|^2 = O(\log \lambda_{\max} P_t^{-1}) \quad \text{a.s.} \tag{36d}$$

Furthermore let AK, BK, C be coprime (i.e. $x(t)$ is reachable from $u(t)$, $v(t)$ (Lemma 8)) and $C(s)$ be strictly minimum phase. For stable signal models (i.e. no finite escape time for $x(t)$ and $\lambda_{\max} P_t^{-1} \leqslant O(t)$) and with $u(t)$, $v(t)$ suitably exciting as in (37), then as $t \to \infty$

$$\| \theta_N - \theta_N(t) \|^2 = O(t^{-1} \log t) \quad \text{a.s.}$$

Proof. The results (36a)–(36d) are implicitly established in Reference 13. Although the signal model in Reference 13 has different interpretations for θ_N, $x(t)$ than here (namely a specialization of the model used here when $N=2$, $a=0$), the proofs are invariant of such interpretations as long as the subsystem with input $\tilde{\theta}_N^{\mathrm{T}}(t)\hat{x}(t)$ and output $\theta_N^{\mathrm{T}}\tilde{x}(t)+\frac{1}{2}\tilde{\theta}_N^{\mathrm{T}}(t)\hat{x}(t)$ is strictly passive. ($\tilde{x}(t)\triangleq x(t)-\hat{x}(t)$ and $\tilde{\theta}_N(t)\triangleq\theta_N-\hat{\theta}_N(t)$.) □

Least-squares parameter recovery

The state space Ito representation of (1) prefiltered by the exponentially stable filter $L^{-1}(s)$ is

$$\begin{aligned} \mathrm{d}\varphi(t) &= \Gamma\varphi(t)\,\mathrm{d}t - e_1\,\mathrm{d}y_{Ln}^{(1)}(t) + e_{n+1}u_{Ln}(t)\,\mathrm{d}t + e_{2n+1}\,\mathrm{d}e_L^{(1)}(t) \\ \mathrm{d}\bar{y}_{Ln}(t) &= \varphi^{\mathrm{T}}(t)\theta\,\mathrm{d}t + \mathrm{d}e_L^{(1)}(t), \qquad \varphi(0)=0 \end{aligned} \tag{38a}$$

where e_i^T is as defined in (25). Also

$$\begin{gathered} \varphi(t) = (-y_{Ln}^{(1)}(t)\dots -y_{Ln}^{(n)}(t)\;\; u_{Ln}^{(1)}(t)\dots u_{Ln}^{(n)}(t)\;\; e_L^{(1)}(t)\dots e_L^{(n)}(t))^{\mathrm{T}} \\ \theta = (a_1\dots a_n\; b_1\dots b_n\; c_1\dots c_n)^{\mathrm{T}}, \qquad e_L^{(i)}(t) = p^n\,\frac{e^{(i)}(t)}{L(p)} \end{gathered} \tag{38b}$$

where $y_{Ln}(t)=p^nL^{-1}(p)y(t)$, etc. The superscripts in (38a,b) are defined in (25b) and

$$\Gamma = \text{block diag}(E_n, E_n, E_n), \qquad E_n = \begin{bmatrix} 0 & 0 \\ I_{n-1} & 0 \end{bmatrix} \tag{38c}$$

So far we have proved that consistent estimates of parameters of the transformed signal model (25) can be obtained under the relaxed SBR condition (27) and PE condition (37). Identification of the original signal model parameters θ can also be accomplished under the same conditions if the following least-squares (LS) algorithm operating in parallel with the ELS algorithm is utilized:

$$\begin{aligned} \mathrm{d}\hat{\theta}(t) &= \bar{P}_t\bar{\varphi}(t)(\mathrm{d}\bar{y}_{Ln}(t) - \hat{\theta}^{\mathrm{T}}(t)\bar{\varphi}(t)\,\mathrm{d}t) \\ \bar{P}_t &= \left(\int_0^t \bar{\varphi}(\tau)\bar{\varphi}(\tau)^T\,\mathrm{d}\tau + \bar{P}_0^{-1}\right)^{-1}, \qquad \bar{P}_0>0 \end{aligned} \tag{39a}$$

where

$$\bar{\varphi}(t) = (-y_{Ln}^{(1)}(t)\dots -y_{Ln}^{(n)}(t)\;\; u_{Ln}^{(1)}(t)\dots u_{Ln}^{(n)}(t)\;\; \hat{e}_L^{(1)}(t)\dots \hat{e}_L^{(n)}(t))^{\mathrm{T}} \tag{39b}$$

and $\hat{\theta}(t)$ are the parameter estimates of θ.

Remark. It may be thought that $\mathrm{d}\bar{y}_{Ln}(t)-\varphi^{\mathrm{T}}(t)\theta\,\mathrm{d}t$ in (38a) should be the differential of a Wiener process for the LS scheme to converge. However, as shown in Reference 2, this is not the case as long as $L^{-1}(p)$ is exponentially stable and $\mathrm{d}e^{(1)}(t)(=\mathrm{d}v(t)$; see (1)) is the differential of a Wiener process.

Note that the noise terms $e_L^{(i)}(t)$ are obtained from the ELS scheme. Thus the scheme (39), despite its similarity to (26), has an almost standard least-squares form. The only non-standard feature of the proposed scheme is that the regression vector (39b) differs from the *true* one where $e_L^{(\cdot)}(t)$ would be present instead of $\hat{e}_L^{(\cdot)}(t)$. We no prove that as long as the ELS scheme converges, this discrepancy is asymptotically negligible, i.e. it does not affect either the consistency or the asymptotic rate of convergence of the LS scheme.

Theorem 4

Consider the least-squares scheme (39) applied to the signal model (1) with Ito form (38) under the relaxed SBR condition (27), where $e_L^{(i)}(t)$ is obtained from the ELS scheme (26). Then as $t \to \infty$

$$\| \theta - \hat{\theta}(t) \|^2 = O\left(\frac{\log \lambda_{\max} \bar{P}_t^{-1}}{\lambda_{\min} \bar{P}_t^{-1}}\right) \quad \text{a.s.} \tag{40}$$

Moreover, under (37) with P_t defined as in (26), as $t \to \infty$

$$\| \theta - \hat{\theta}(t) \|^2 = O\left(\frac{\log \lambda_{\max} P_t^{-1}}{\lambda_{\min} P_t^{-1}}\right) \quad \text{a.s.} \tag{41}$$

Furthermore, for suitably rich bounded variance inputs $u_L(t)$, $e_L^{(1)}(t)$, then as $t \to \infty$, lim inf $\lambda_{\min} P_t^{-1}/t$ and for stable models with bounded inputs

$$\| \theta - \hat{\theta}(t) \|^2 = O(t^{-1} \log t) \quad \text{a.s.} \tag{42}$$

Proof. See Appendix. □

4. ELS ALGORITHM WITH RELAXED SPR CONDITION FOR DISCRETE-TIME SYSTEMS

We present a transformed discrete-time ELS algorithm for discrete-time systems, interpret its associated SPR condition and finally discuss PE and convergence.

Transformed ELS algorithm

Consider the filtering operation on (3) in terms of the filter

$$\frac{1}{W(q)} = \frac{F(q)}{L(q)} \tag{43}$$

According to Corollary 2, (43) is a good approximation of $C^{-1}(q)$ provided N is large enough. Because $F(q)$ is unknown, we do not use (43) in the actual implementation of the ELS algorithm. Applying the filter (43) with the normalized $\bar{F}(q)$ defined in (16) replacing $F(q)$, (3) becomes

$$A(q)\frac{\bar{F}(q)}{L(q)}\, y(k) = B(q)\, \frac{\bar{F}(q)}{L(q)}\, u(k) + C(q)\, \frac{\bar{F}(q)}{L(q)}\, e(k)$$

or equivalently, using (14a),

$$\bar{A}(q) y_L(k) = \bar{B}(q) u_L(k) + \bar{g}_0(-a)^{N-n+1} e(k) - \bar{G}(q) e_H(k) + e(k) \tag{44}$$

The subscript L in (44) denotes filtering with $L^{-1}(q)$, i.e. $y_L(k) = L^{-1}(q) y(k)$, $u_L(k) = L^{-1}(q) u(k)$. Also $e_H(k) = H(q) L^{-1}(q) e(k)$, $\bar{A}(q) = A(q)\bar{F}(q)$, $\bar{B}(q) = B(q)\bar{F}(q)$ and

$$\bar{G}(q) = \bar{g}_0 q^{n-1} + \cdots + \bar{g}_{n-1}$$

We then estimate the coefficients of $\bar{A}(q)$, $\bar{B}(q)$ and $\bar{G}(q)$ as $\hat{\bar{A}}(q)$, $\hat{\bar{B}}(q)$ and $\hat{\bar{G}}(q)$ using ELS. In ELS we use the *a posteriori* estimate of $e(k)$ defined as

$$\hat{e}(k) = \hat{\bar{A}}(q) y_L(k) - \hat{\bar{B}}(q) u_L(k) - \hat{\bar{g}}_0(-a)^{N-n+1} \hat{e}(k) + \hat{\bar{G}}(q) \hat{e}_H(k) \tag{45}$$

where $\hat{e}_H(k) = H(q) L^{-1}(q) \hat{e}(k)$. Note that on the right-hand side of (45), $\hat{e}(k)$ cancels out in the expression $-\hat{g}_0(-a)^{N-n+1}\hat{e}(k)+\hat{\bar{G}}(q)\hat{e}_H(k)$. Thus (45) is causal. Now define

$$\begin{aligned}\phi(k)=(&-y_L(k-1)\dots -y_L(k+n-l-N)\ u_L(k-1)\dots u_L(k+n-m-N)\\ &[(-a)^{N-n+1}-q^{n-1}H(q)/L(q)]\,e(k+n-l-N)\ \ -e_H(k+2n-l-N-2)\ \dots\\ &-e_H(k+n-l-N))^{\mathrm{T}}\\ \theta=(&\bar{a}_1\dots\bar{a}_{l+N-n}\ \bar{b}_1\dots\bar{b}_{m+N-n}\ \bar{g}_0\dots\bar{g}_{n-1})^{\mathrm{T}}\end{aligned}\tag{46}$$

Define $\hat{\phi}(k)$ similar to $\phi(k)$ with $\hat{e}(.)$, $\hat{e}(.)$, $\hat{e}_H(.)$ replacing $e(.)$, $e_H(.)$. The ELS estimate $\theta(k)$ is defined from

$$\begin{aligned}&\hat{e}(k-l-N+n)=y_L(k)-\hat{\phi}^{\mathrm{T}}(k)\theta(k-1)\\ &\theta(k)=\theta(k-1)+\hat{P}_k\hat{\phi}(k)e(k-l-N+n)\\ &\hat{P}_k=\hat{P}_{k-1}-\frac{\hat{P}_{k-1}\hat{\phi}(k)\hat{\phi}^T(k)\hat{P}_{k-1}}{1+\hat{\phi}^T(k)\hat{P}_{k-1}\hat{\phi}(k)},\qquad \hat{P}_0>0\end{aligned}\tag{47}$$

suitably initialized with $\hat{\phi}(0)$ and $\theta(0)$. It can be shown[5] that a sufficient condition for the ELS scheme to converge is that

$$\frac{L(q)}{\bar{F}(q)C(q)}-\tfrac{1}{2}\ \text{is SPR}$$

or equivalently,

$$\left\|\frac{\bar{G}(q)H(q)}{L(q)}-g_0(-a)^{N-n+1}\right\|_\infty<1\quad\text{(strictly bound real (SBR) condition)}\tag{48}$$

Remark. For output error (OE) systems defined as in (4), setting $A=C$ in (44), the OE equivalent of (45) is

$$\hat{e}(k)=y(k)+\hat{g}_0(-a)^{N-n+1}(y(k)-e(k))-\hat{\bar{G}}(q)(y_H(k)-\hat{e}_H(k))-\hat{B}(q)u_L(k)$$

where $y_H(k)=H(q)L^{-1}(q)y(k)$. So the OE regression vector (cf. (46)) is

$$\begin{aligned}\phi_{\mathrm{OE}}(k)=[&\{-(-a)^{N-n+1}(y(k-n)-w(k-n))+y_H(k-1)-e_H(k-1)\}\\ &(y_H(k-2)-e_H(k-2))\dots(y_H(k-n)-e_H(k-n))\ u_L(k+N-n-1)\dots u_L(k-m)]^{\mathrm{T}}\end{aligned}$$

Also

$$\theta_{\mathrm{OE}}=(\bar{g}_0\dots\bar{g}_{n-1}\ \bar{b}_1\dots\bar{b}_{m+N-n})^{\mathrm{T}}$$

and $\hat{\phi}_{\mathrm{OE}}(k)$ is defined similar to $\phi_{\mathrm{OE}}(k)$ with $\hat{e}_H(k)$, $\hat{e}(k)$ replacing $e_H(k)$, $e(k)$. The ELS estimate $\theta_{\mathrm{OE}}(k)$ is defined from

$$\begin{aligned}&\theta_{\mathrm{OE}}(k)=\theta_{\mathrm{OE}}(k-1)+\hat{P}_k\hat{\phi}_{\mathrm{OE}}(k)(y(k-n)-\hat{\phi}_{\mathrm{OE}}^{\mathrm{T}}(k)\theta_{\mathrm{OE}}(k-1)\\ &\hat{P}_k=\hat{P}_{k-1}-\frac{\hat{P}_{k-1}\hat{\phi}_{\mathrm{OE}}(k)\hat{\phi}^{\mathrm{T}}{}_{\mathrm{OE}}(k)\hat{P}_{k-1}}{1+\hat{\phi}_{\mathrm{OE}}^{\mathrm{T}}(k)\hat{P}_{k-1}\hat{\phi}_{\mathrm{OE}}(k)},\qquad \hat{P}_0>0\end{aligned}$$

The SPR condition (48) also holds for OE systems.[5]

Overparametrization selection to satisfy SPR condition

From Corollary 2 we know that

$$\left\|\frac{G(q)H(q)}{L(q)C(q)}\right\|_\infty\tag{49}$$

can be made arbitrarily small by choosing N large enough. By restricting the zeros of $C(q)$ to lie inside a given *a priori* compact set, it is possible to specify N and a such that (27) is satisfied for all $C(q)$ whose zeros lie inside the set. In this subsection we specify N, a and the compact set.

Notice that under (12) and (20), $z^{-(N-n+1)}\mathscr{G}(z^{-1})$ transforms to $G(q)H(q)L^{-1}(q)$ and vice versa. So

$$\left|\frac{G(q)H(q)}{L(q)}\right|_{|q|=1} = |\mathscr{G}(z^{-1})|_{|z|=1} \tag{50a}$$

Moreover, since stability is preserved under the transformation (12),

$$\frac{(-a)^{N-n+1}}{f_0} - \frac{1}{f_0}\frac{G(q)H(q)}{L(q)} \text{ SBR} \Leftrightarrow \frac{(-a)^{N-n+1}g_0 - \mathscr{G}(z^{-1})}{f_0} \text{ SBR} \tag{50b}$$

We now show how to select N and a to satisfy the SBR condition (48).

Lemma 9

Consider a polynomial $C(z) = \Pi_{i=1}^{n} (z - q_i)$ such that its zeros lie in a circle with centre

$$x_0 = a\,\frac{1 - R^2}{1 - R^2a^2}$$

and radius

$$r = R\,\frac{1 - a^2}{1 - R^2a^2} \tag{51a}$$

or equivalently,

$$|q_i - x_0| \leqslant r, \qquad R < 1 \qquad \forall i \tag{51b}$$

Consider also for any N that the polynomials $F(q)$, $G(q)$ and $H(q)$ are uniquely defined as in (14). Then there exists $N_0(r)$ such that for all $N \geqslant N_0(r)$

$$\left|\frac{g_0(-a)^{N-n+1}}{f_0}\right| + \left|\frac{1}{f_0}\frac{G(q)H(q)}{L(q)}\right| < 1 \tag{51c}$$

and the SBR condition (48) is satisfied.

Proof. Using (12) with r defined as in (51a) and z_i defined as in Lemma 5, straightforward manipulations yield

$$|z_i| \leqslant R < 1 \Leftrightarrow \left|q_i + a\,\frac{1 - R^2}{1 - R^2a^2}\right| \leqslant r$$

While a proof similar to that of Lemma 6, it follows that for any N if

$$N \geqslant \frac{R(1+n) - 1}{1 - R} \quad \text{and} \quad 3R^{N-n+1}\,\frac{(N+n)!}{N!}\,(1+R)^n < 1 \tag{52}$$

then the SBR condition (48) is satisfied. Note that $N_0(r)$ can be defined as the smallest value of N for which (52) holds. □

Discussion. Unless N is chosen large (and so an unrealistically large number of parameters are estimated), existing relaxed SPR algorithms such as in Reference 5 are not efficient when any of the zeros of $C(z)$ lie in a region offset from the centre of the unit disc. Lemma 9 shows the usefulness of our novel factorization (14a), when, for example, the zeros of C lie close to $z = 1$ in the z-plane as a result of fast sampling of the continuous-time process (1). By choosing $0 \leqslant a < 1$ to satisfy (51a), we can shift the SPR bounds (the circles with centre x_0 and radius r in (51)) and the actual SPR regions (i.e. the location of the zeros of $C(q)$ when (48) is satisfied) along the real axis of the z-plane to the region where the fast-sampled zeros lie *without* increasing N. Figure 1 shows the SPR regions for second-order C-polynomials ($n = 2$) for $N = 2, 3$.

In Figure 1, with $a = 0$, the resulting SPR regions are those obtained in the relaxed SPR algorithm of Reference 5 and are unsuitable for fast-sampled systems. Notice that selecting $a = 0{\cdot}5$, $N = 2$ in our scheme adequately covers the region close to $z = 1$. The algorithm proposed in Reference 2 theoretically requires infinitely large N to be SPR at $z = 1$.

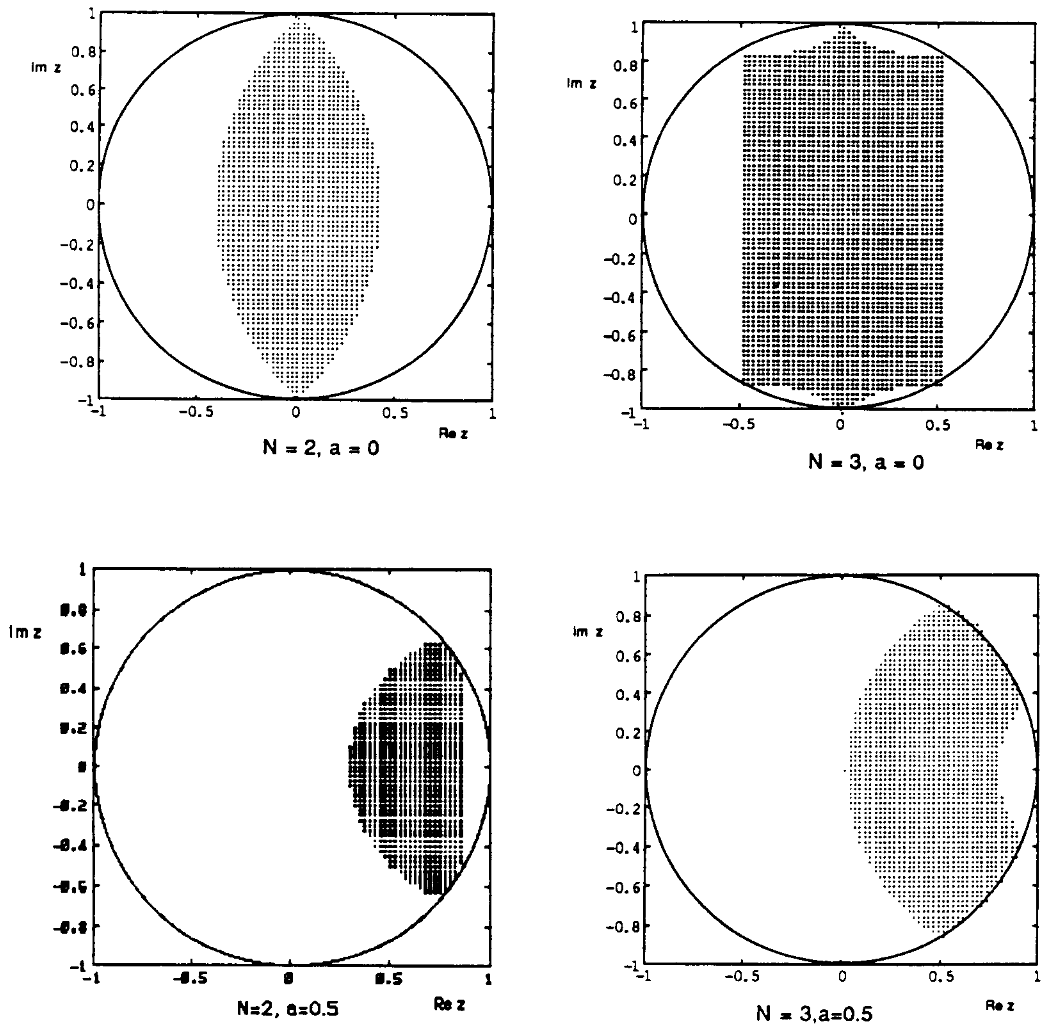


Figure 1. SPR regions in z-plane for second order C-polynomials ($n = 2$)

Note that the SPR regions and bounds can also be shifted along the negative real z-axis of the unit disc by selecting $-1 < a < 0$. The SPR regions and bounds are obtained by reflecting the SPR regions and bounds for corresponding positive values of a through the imaginary z-axis.

Persistence of excitation

We require the regression vector $\phi(k)$ and its estimate $\hat{\phi}(k)$ in (46) to be suitably exciting for two reasons.[9] First, if $\hat{\phi}(k)$ is not suitably exciting then the estimation algorithm and in particular the calculation of P_k will suffer from numerical ill-conditioning. Secondly, the PE of $\phi(k)$ is needed to assure strong consistency of the identification scheme and therefore its robustness to unmodelled dynamics.[12]

It is known[14] that there is almost sure parameter convergence in the ELS algorithm if in addition to SPR condition (48) the following PE condition holds for $\phi(k)$:

$$\lim_{r \to \infty} \frac{\log \lambda_{\max} P_r^{-1}}{\lambda_{\min} P_r^{-1}} = 0 \quad \text{a.s.} \tag{53}$$

where $P_r^{-1} = \sum_{k=1}^{r} \phi(k)\phi(k)^{\mathrm{T}} + P_0^{-1}$ for some $P_0 > 1$. In order to translate this condition on $\phi(k)$ to conditions on the external signals $u(k)$, $w(k)$, it is necessary and sufficient[9] that $\phi(k)$ be reachable from $u(k)$, $w(k)$.

Lemma 10

A necessary and sufficient condition for $\phi(k)$ in (46) associated with signal model (44) to be reachable from inputs $u(k)$, $w(k)$ is that $A(z)K(z)$, $B(z)K(z)$ and $C(z)$ are coprime where $K(z) \triangleq H(z) - (z-a)^{N-n+1}(-a)^{N-n+1}$.

Proof. See Appendix. □

Remark. It is easily shown for output error systems from Lemma 10 that since $A = C$, a necessary and sufficient condition for $\phi_{\mathrm{OE}}(k)$ to be reachable from input $u(k)$ is that A and BK are coprime.

Convergence of the transformed ELS algorithm

We summarize in the following theorem the convergence properties of our transformed ELS algorithm.

Theorem 5

Consider the transformed ELS algorithm (45)–(47) associated with signal model (44). If N is chosen sufficiently large such that the SBR condition (48) is satisfied, then as $k \to \infty$

$$\begin{aligned} \| \theta - \theta(k) \|^2 &= O\left(\frac{\log \lambda_{\max} \hat{P}_k^{-1}}{\lambda_{\min} \hat{P}_k^{-1}}\right) \quad \text{a.s.} \\ \sum_{i=1}^{k} \| e(i) - \hat{e}(i) \|^2 &= O(\log \lambda_{\max} \hat{P}_k^{-1}) \quad \text{a.s.} \end{aligned} \tag{54}$$

Moreover, under (53), as $k \to \infty$

$$\|\theta - \theta(k)\|^2 = O\left(\frac{\log \lambda_{\max} P_k^{-1}}{\lambda_{\min} P_k^{-1}}\right) \to 0 \quad \text{a.s.}$$
$$\sum_{i=1}^{k} \|e(i) - \hat{e}(i)\|^2 = O(\log \lambda_{\max} P_k^{-1}) \quad \text{a.s.} \tag{55}$$

Furthermore, if AK, BK and C are coprime (K is defined in Lemma 10), $\phi(k)$ is reachable from inputs $u(k)$, $w(k)$ and for suitably rich bounded variance inputs $u(k)$, $w(k)$, then as $k \to \infty$, lim inf $\lambda P_k^{-1}/k > 0$ and for stable signal models with bounded inputs

$$\|\theta - \theta(k)\| = \mathrm{O}(k^{-1} \log k) \quad \text{a.s.} \tag{56}$$

Proof. The proof of the convergence of our transformed ELS algorithm is the same as that in Reference 5. □

Least-squares parameter recovery

So far we have proved that consistent estimates of parameters of the transformed signal model (44) can be obtained under the relaxed SBR condition (48) and PE condition (53). Identification of the original signal model parameters A, B and C can also be accomplished under the same conditions if the following least-squares (LS) algorithm operating in parallel with the ELS algorithm is utilized:[5]

$$\bar{\theta}(k) = \bar{\theta}(k-1) + \bar{P}_k \bar{\phi}(k)^{\mathrm{T}} \bar{\theta}(k-1)$$
$$\bar{P}_k = \bar{P}_{k-1} - \frac{\bar{P}_{k-1}\bar{\phi}(k)\bar{\phi}(k)^{\mathrm{T}}\bar{P}_{k-1}}{1 + \bar{\phi}(k)^{\mathrm{T}}\bar{P}_{k-1}\bar{\phi}(k)}, \qquad \bar{P}_0 > 0 \tag{57}$$

where the regression vector $\bar{\phi}(k)$ is given in terms of $y(.)$, $u(.)$ and the noise estimate $\hat{e}(.)$ from the transformed ELS scheme as

$$\bar{\phi}_k \triangleq (y(k-1) \ldots y(k-l) \;\; u(k-1) \ldots u(k-m) \;\; \hat{e}(k+n-l-1) \ldots \hat{e}(k-l))^{\mathrm{T}} \tag{58}$$

Also $\bar{\theta}$ contains the coefficients of A, B and C suitably arranged. Note that the noise terms $\hat{e}(k)$ in the regression vector (58) are regarded as measurable. Thus the algorithm (57), despite its similarity to (45)–(47), has an almost standard least-squares form. The only non-standard feature of the proposed scheme is that the regression vector (58) differs from the *true* one where $e(.)$ would be present instead of $\hat{e}(.)$. It is proved in Reference 5 that as long as the ELS algorithm converges, this discrepancy is asymptotically negligible, i.e. it does not affect either the consistency or the asymptotic rate of convergence of the LS scheme.

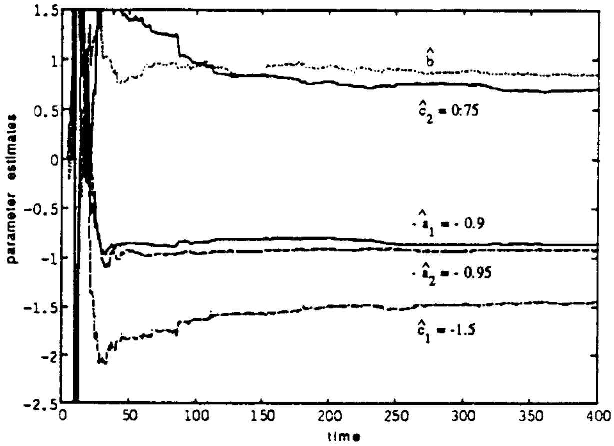
Remark. In OE systems

$$\bar{\phi}_{\mathrm{OE}}(k) \triangleq ((y(k-1) - \hat{e}(k-1)) \ldots (y(k-l) - \hat{e}(k-l)) \;\; u(k-1) \ldots u(k-m))^{\mathrm{T}}$$
$$\bar{\theta}_{\mathrm{OE}} = (a_1 \ldots a_l \;\; b_1 \;\; \ldots \;\; b_m)^{\mathrm{T}}$$

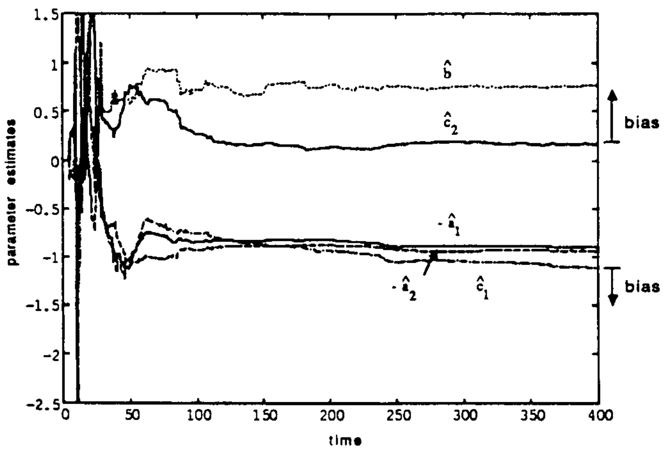
would replace $\bar{\phi}(k)$ and $\bar{\theta}(k)$.

5. SIMULATIONS

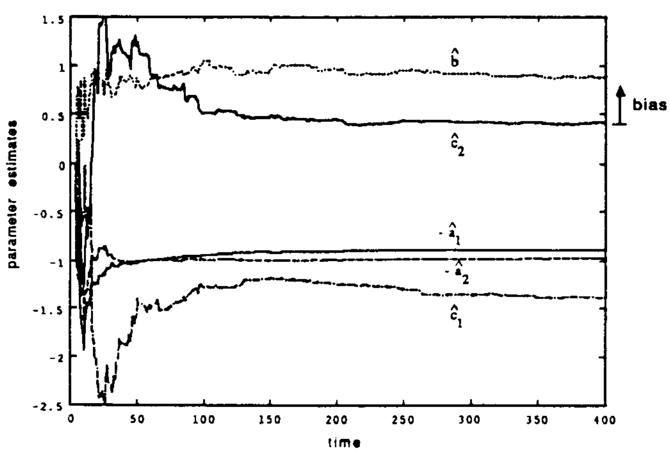
To give insights into the behaviour of the transformed ELS algorithm, a number of simulations have been made. Here we report a typical example.



(a) N = 2, a = 0.5



(b) N = 2, a = 0 (Standard ELS)



N=4, a = 0.

Figure 2. Comparison of estimates of system (59)

We consider a system with ARMAX representation

$$y(k)+0\cdot 9y(k-1)+0\cdot 95y(k-2)=u(k-1)+e(k)-1\cdot 5e(k-1)+0\cdot 75e(k-2) \quad (59)$$

The system fails to satisfy the SPR condition (48) when $a=0$, $N=2$. This corresponds to the standard ELS scheme and simulations (see Figure 2(b)) show a significant bias in the estimates. Assume we know *a priori* that $C(q)$ was obtained by fast sampling a continuous-time process (1), i.e. its zeros lie near $z=1$. We choose $a=0\cdot 5$, $N=2$ (see Figure 1). This selection is reasonable since the SPR region with $a=0\cdot 5$, $N=2$ covers a sizable amount of the right half of the unit disc. Figure 2(a) shows the estimates of the system parameters using our proposed scheme with $N=2$, $a=0\cdot 5$. It is clear that the parameter estimates converge to the 'true' values. Figure 2(c) shows that if the scheme in Reference 5 is used, even with $N=4$, there is still bias in the estimates. Simulations show that unless $N>6$ there is significant bias in the parameter estimates if the algorithm in Reference 5 is used.

6. CONCLUSIONS

The strength of ELS schemes and their stochastic approximation versions is that they hold out hope for global convergence in stochastic adaptive control, at least in the constant-parameter case with no unmodelled dynamics. Also, under persistence of excitation, they hold out hope of local stability in the presence of unmodelled dynamics. The Achilles heel of such schemes is the SPR convergence condition. This paper has addressed this issue for continuous-time schemes and certain fast-sampled continuous-time schemes, building on earlier work for discrete-time schemes. We have achieved a realistic trade-off between increasing algorithmic complexity and avoiding drift or bias and guaranteeing convergence. This has been illustrated by simulation studies in the most common applications to low-order models.

Some of the generalizations developed here of known PE results to continuous time are of independent interest. Also, the systematic application here of such results to the class of overparametrized systems is illustrative of the power of such results. It could be claimed that in adaptive control, parameter convergence and thus persistence of excitation is not strictly necessary to guarantee convergence to the optimal control. However, as is now well known, the spectre of the lack of robustness to unmodelled dynamics then looms large.

ACKNOWLEDGEMENTS

This work is partially supported by DSTO, Australia and Boeing Commercial Airplanes, U.S.A.

APPENDIX

Proof of Theorem 4

Considering (38) and (39), define

$$\bar{\theta}(t) \triangleq \theta - \hat{\theta}(t), \qquad \tilde{\varphi}(t) \triangleq \varphi(t) - \hat{\varphi}(t)$$

Then

$$\begin{aligned} \mathrm{d}\tilde{\theta}(t) &= -\mathrm{d}\hat{\theta}(t) = -\bar{P}_t\bar{\varphi}(t)(\mathrm{d}\bar{y}_{Ln}(t) - \hat{\theta}(t)^{\mathrm{T}}\bar{\varphi}(t)\,\mathrm{d}t) \\ &= [-\bar{P}_t\bar{\varphi}(t)(\bar{\varphi}(t)\theta\,\mathrm{d}t + \mathrm{d}e_L^{(1)}(t) - \hat{\theta}(t)^{\mathrm{T}}\bar{\varphi}(t)\,\mathrm{d}t)] - \bar{P}_t\bar{\varphi}(t)\tilde{\varphi}(t)^{\mathrm{T}}\theta\,\mathrm{d}t \end{aligned} \quad (60)$$

The term enclosed in square brackets in (60) is the least-squares parameter error associated with the model $d\bar{y}_{Ln}(t) = \bar{\varphi}(t)^{\mathrm{T}}\theta \, dt + de_L^{(1)}(t)$, so that by known results[13] (in fact an appropriate specialization of Theorem 3),

$$\| \bar{P}_t \bar{\varphi}(t)(\varphi^{\mathrm{T}}(t)\theta \, dt + dv_L(t) - \hat{\theta}^{\mathrm{T}}(t)\bar{\varphi}(t) \, dt) \|^2 = O\left(\frac{\log \lambda_{\max}\bar{P}_t^{-1}}{\lambda_{\min}\bar{P}_t^{-1}}\right) \tag{61}$$

The second term in (60) is bounded as follows. Using the Schwartz inequality

$$\left\| \int_0^T \bar{P}_t\bar{\varphi}(t)\tilde{\varphi}^{\mathrm{T}}(t)\theta \, dt \right\|^2 \leqslant \int_0^T | \bar{P}_t\bar{\varphi}(t) |^2 \, dt \int_0^T | \tilde{\varphi}^{\mathrm{T}}(t)\theta |^2 \, dt \tag{62}$$

It is established in Reference 13 (Lemma 3.1, Theorem 3.1) that for constants k_1 and k_2,

$$\int_0^T \| \tilde{\varphi}(t) \|^2 \, dt \leqslant k_1 \int_0^T \| \tilde{\theta}^{\mathrm{T}}(t)\varphi(t) \|^2 \, dt + k_2 = O(\log \operatorname{tr} \hat{P}_t^{-1})$$

where $\hat{P}_t^{-1}$ is defined as in (26). Thus

$$\int_0^T \| \theta^{\mathrm{T}}\tilde{\varphi}(t) \|^2 \, dt = O(\log \operatorname{tr} \hat{P}_t^{-1}) \tag{63}$$

Also

$$\int_0^T \| \bar{P}_t\bar{\varphi}(t) \|^2 \, dt = \int_0^T \operatorname{tr} \, [\bar{P}_t\bar{\varphi}(t)\bar{\varphi}(t)^{\mathrm{T}}\bar{P}_t^{\mathrm{T}}] \, dt = -\operatorname{tr} \, \bar{P}_t \tag{64}$$

where the last equality follows from differentiating $\bar{P}_t\bar{P}_t^{-1} = I$ with respect to time and using (39a). Therefore with (63) and (64) substitued in (62)

$$\left\| \int_0^T \bar{P}_t\bar{\varphi}(t)\tilde{\varphi}(t)^{\mathrm{T}}\theta \, dt \right\|^2 = O((\log \operatorname{tr} \hat{P}_t^{-1})(\operatorname{tr} \bar{P}_t)) = O\left(\frac{\log \lambda_{\max}\hat{P}_t^{-1}}{\lambda_{\min}\bar{P}_t^{-1}}\right) \tag{65}$$

But, recalling the definitions of $\hat{x}(t)$ in (25b) and $\bar{\varphi}(t)$ in (39),

$$\lambda_{\max}\hat{P}_t^{-1} = O\left(\operatorname{tr} \int_0^t \hat{x}(\tau)\hat{x}(\tau)^{\mathrm{T}} \, d\tau\right) = O\left(\operatorname{tr} \int_0^t \bar{\varphi}(\tau)\bar{\varphi}(\tau)^{\mathrm{T}} \, d\tau\right) = O(\lambda_{\max}\bar{P}_t^{-1})$$

because $L^{-1}(s)$ and $H(s)L^{-1}(s)$ in $x(t)$ are stable transfer functions. The desired result (40) then follows from (61) and (65). Also, under (37) we have[13]

$$\left(\frac{\log \lambda_{\max}\hat{P}_t^{-1}}{\lambda_{\min}\hat{P}_t^{-1}}\right) = O\left(\frac{\log \lambda_{\max}P_t^{-1}}{\lambda_{\min}P_t^{-1}}\right) \to 0 \quad \text{as } t \to \infty$$

thus establishing (41). The same arguments[13] used to prove Theorem 3 establish (42). □

Proof of Lemma 10

The proof is along the lines in References 5 and 9. Taking z-transforms we rewrite (3) as

$$y(z) = A^{-1}(z)B(z)u(z) + A^{-1}(z)C(z)e(z) = T_1(z)u(z) + T_2(z)e(z) = T(z)v(z)$$

where $v(z) = (u(z) \;\; e(z))^{\mathrm{T}}$ and $T(z) = (T_1(z) \;\; T_2(z))$. Observe that

$$\begin{pmatrix} y(z) \\ u(z) \\ e(z) \end{pmatrix} = \begin{pmatrix} T(z) \\ e_1 \\ e_2 \end{pmatrix} v(z)$$

where $e_1 = (1 \;\; 0)$ and $e_2 = (0 \;\; 1)$. Consider the regression vector $\phi(k)$ in (46) and note that $\theta(z) = T^{\mathrm{TEE}}(z)v(z)$ where

$$T^{\mathrm{TEE}}(z) = \frac{1}{L(z)} \, (-z^{l+N-n}T^{\mathrm{T}}(z) \cdots - T^{\mathrm{T}}(z) \;\; z^{l+N-n-1}e_1^{\mathrm{T}} \ldots z^{l-m}e_1^{\mathrm{T}}$$
$$(-H(z)z^{n-1} + L(z)(-a)^{N-n+1})e_2^{\mathrm{T}} \;\; -H(z)z^{n-2}e_2^{\mathrm{T}} \ldots -H(z)e_2^{\mathrm{T}})^{\mathrm{T}}$$

We know[9] that $\phi(.)$ is reachable from $v(.)$ iff $T^{\mathrm{TEE}}(z)$ has full rank over R for all z. Also, manipulations show that if for any $\alpha = (\alpha_1^{\mathrm{T}} \; \alpha_2^{\mathrm{T}} \; \alpha_3^{\mathrm{T}})^{\mathrm{T}}$, $\alpha_1 = (\alpha_{1,1} \ldots \alpha_{1,l+N-n})^{\mathrm{T}}$, $\alpha_2 = (\alpha_{2,1} \ldots \alpha_{2,m+N-n-1})^{\mathrm{T}}$,

$\alpha_3 = (\alpha_{3,1} \dots \alpha_{3,n-1})^{\mathrm{T}}$, the reachability condition is equivalent to

$$\alpha^{\mathrm{T}}(z)M(z) \triangleq (\alpha_1(z)\ \ \alpha_2(z)\ \ \alpha_3(z))\begin{pmatrix} T_1(z) & T_2(z) \\ 1 & 0 \\ 0 & K(z) \end{pmatrix} = 0 \tag{66}$$

where

$$K(z) \triangleq H(z) - (q-a)^{N-n+1}(-a)^{N-n+1}$$

$$\alpha_1(z) = \alpha_{1,1}z^{l+N-n-1} + \cdots + \alpha_{1,l+N-n},\ \alpha_2(z) = \alpha_{2,1}z^{l+N-n-1} + \cdots + \alpha_{2,m+N-n}z^{l-m},$$
$$\alpha_3(z) = \alpha_{3,1}z^{n-2} + \cdots + \alpha_{3,n-1}.$$

Note that $K(z)$ is of degree $N-n$.

Let $N(M)$ denote the left null-space of $M(z)$, i.e. the set of all polynomial vectors $\alpha(z)$ obeying (66). Define also

$$P(k_1, k_2, k_3) = \{\text{polynomials } (\alpha_1(z)\ \ \alpha_2(z)\ \ \alpha_3(z))^{\mathrm{T}}, \deg \alpha_i(z) < k_i\}$$

Then it follows from (66) that $T^{\mathrm{TEE}}(z)$ will be full row rank iff

$$P(l+N-n, l+N-n, n-1) \cap N(M) = 0 \tag{67}$$

We now show that the condition $A(z)K(z)$, $B(z)K(z)$ and $C(z)$ are coprime is necessary and sufficient for (67) to hold.

Necessity. Straightforward calculation show that

$$(-A(z)K(z)\ \ B(z)K(z)\ \ C(z))^{\mathrm{T}}M(z) = 0$$

and so $(A(z)K(z)\ \ B(z)K(z)\ \ -C(z))^{\mathrm{T}} \in N(M)$. Also

$$(-A(z)K(z)\ \ B(z)K(z)\ \ C(z))^{\mathrm{T}} \in P(l+N-n+1, l+N-n, n+1)$$

Thus $A(z)K(z)$, $B(z)K(z)$ and $C(z)$ coprime is a necessary condition for (67) to hold.

Sufficiency. Since for $A(z)K(z)$, $B(z)K(z)$ and $C(z)$ coprime, the vector $(-A(z)K(z)\ \ B(z)K(z)\ \ -C(z))^{\mathrm{T}}$ forms a basis for $N(M)$, any other element of $N(M)$ can be obtained as

$$(\mathscr{A}(z)\mathscr{B}(z)\mathscr{C}(z))^{\mathrm{T}} = \beta(z)(-A(z)K(z)\ \ B(z)K(z)\ \ -C(z))$$

where $\deg \beta(z) \geqslant 1$. However, since $\deg \mathscr{C}(z) > n+1$, coprimeness of AK, BK and C implies (67).

REFERENCES

1. Goodwin, G. C., and K. S. Sin, *Adaptive Filtering, Prediction and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
2. Moore, J. B., 'Convergence of continuous-time stochastic ELS parameter estimation', *Stoch. Process. Appl.*, **27**, 195–215 (1988).
3. Gevers, M., V. Wertz and G. C. Goodwin, 'A parameter estimation algorithm for continuous time stochastic adaptive control,' *Proc. 27th IEEE Conf. on Decision and Control*, December 1988, pp. 1922–1923.
4. Shah, S., and G. F. Franklin, 'On satisfying strict positive real condition for convergence by overparametrisation', *IEEE Trans. Automatic Control*, **AC-27**, 715–16 (1982).
5. Moore, J. B., M. Niedzwiecki and L. Xia, 'Identification/prediction algorithms for ARMAX models with relaxed positive real conditions', *Int. j. adapt. control signal process.*, **4**, 49–68 (1990).
6. Wahlberg, B., 'System identification using high-order models, revisited', *Technical Report EE/CICS8903*, University of Newcastle, 1989.
7. Wahlberg, B., 'Limit results for sampled systems', *Int. J. Control*, **48**, 1267–1283 (1988).
8. Ljung, L., and T. Soderstrom, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.
9. Green, M., and J. B. Moore, 'Persistence of excitation in linear systems', *Syst. Control Lett.*, **7**, 351–360 (1986).
10. Shimkin, N., and A. Feuer, 'Persistency of excitation in continuous time systems', *Syst. Control Lett.*, **9**, 225–233 (1987).

11. Chen, C. T., *Introduction to Linear Systems Theory*, Holt Rinehart Winston, New York, 1970.
12. Chen, H., and L. Guo, 'A robust stochastic adaptive controller', *IEEE Trans. Automatic Control*, **33**, 1035–1043 (1988).
13. Chen, H., and J. B. Moore, 'Convergence rates for continuous-time stochastic ELS parameter estimation', *IEEE Trans. Automatic Control*, **AC-32**, 267–269 (1987).
14. Chen, H., and L. Guo, 'Convergence rate of least squares identification and adaptive control for stochastic systems', *Int. J. Control*, **44**, 1459–1476 (1986).