Factorizations that relax the Positive Real Condition in Continuous-time ELS Schemes

Bo Wahlberg
Dept. Electrical Engineering, Linkoping Univ., Linkoping, Sweden.
Vikram Krishnamurthy, John B. Moore
Dept. Systems Engg., Australian National Univ., Canberra, ACT 2601, Australia.

Abstract: This paper proposes Extended Least Squares (ELS) schemes for ARMAX model identification of continuous-time systems. The schemes have a relaxed Strictly Positive Real (SPR) condition for global convergence. The relaxed SPR scheme is achieved by introducing overparametrisation and prefiltering but without introducing ill-conditioning. The schemes presented are the first such proposed for continuous-time systems. The concepts developed here carry through to output-error, fast-sampled continuous-time systems and associated discrete-time ELS algorithms. We also state conditions for the persistence of excitation (P.E.) of the regression vectors in the proposed ELS schemes to assure strong consistency and obtain convergence rates.

Keywords: Extended Least Squares, identification, Strictly Positive Real, Persistent Excitation, overparametrisation.

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. 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 colored noise model for their convergence. The construction of the filter to achieve the SPR condition is in general more difficult than projection into a stability domain for open-loop identification, and so perhaps renders the ELS approach less attractive to use than RPE methods in this situation. However, for adaptive control, the ELS approach is the only approach known with guaranteed convergence results. (See [1] for discrete-time and [2], [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

$$\frac{W}{C} - \frac{1}{2} \text{ is Strictly Positive Real (SPR).}$$

In discrete time, it is often the case that 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 $z^n C^{-1}(z) - 1/2$ cannot be SPR, even where noise disturbances are white.

In continuous time, the SPR condition is even more restrictive than in the correponding 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 1/2. Of course, in this case $W(s)$ is not asymptotically stable and is apriori 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 [13], [4] and the references therein for discrete-time results). In [4], the SPR condition is side stepped by transforming the discrete-time ARMAX model into an equivalent and unique overparametrised form. Then $\overline{C}^{-1}(z^{-1}) \triangleq z^n C^{-1}(z)$ is expanded in the form $\mathcal{F}(z^{-1})+z^{-D}\mathcal{G}(z^{-1})\overline{C}^{-1}(z^{-1})$ for suitably large delay D so that the relevant condition that $\mathcal{F}(z^{-1})\overline{C}^{-1}(z^{-1})-1/2$ be SPR is satisfied. A problem with this approach is that if the zeros of $C(z)$ lie in a region offset from the center 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 [5] 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 continuous time systems. The relaxed SPR scheme is acheived by introducing overparametrization and prefiltering but without introducing ill conditioning. The techniques developed in this paper have been extended to discrete-time systems where the zeros of $C(z)$ lie in a region offset from the centre of the unit z-circle [12], for example, in fast sampled output error (OE) continuous-time systems [5] where the zeros of $C(z)$ are close to the unit circle. It is shown in [12] that instead of expanding C in the delay operator which is the existing approach [4], using more suitable operators yield 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 organised as follows: In Sec.2 we first present the class of signal models of interest and then propose novel expansions and factorizations for C^{-1}. In Sec.3 we describe the continuous-time transformed ELS scheme and study its relaxed SPR property. Also the convergence properties of the algorithm and a scheme for recovering the parameters are discussed. Some conclusions are drawn in Sec.4.

2 Signal Model and Factorization of $C^{-1}(s)$

Signal Model

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

$$A(p)\,y(t) = B(p)\,u(t) + C(p)\,e(t)$$
$$A(p) = p^n + a_1 p^{n-1} + \cdots + a_n, \quad B(p) = b_1 p^{n-1} + \cdots + b_n$$

$$C(p) = p^n + c_1 p^{n-1} + \cdots + c_n. \qquad (2.1)$$

Here p denotes the differentiation operator, $u(t)$ and $y(t)$ are the input and output signals respectively, $e(t)$ 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 is given below as

$$d\phi(t) = \mathcal{A}\,\phi(t)\,dt + \mathcal{B}\,u(t)\,dt + \mathcal{K}\,dv(t)$$

$$d\overline{y}(t) = \mathcal{C}\,\phi(t)\,dt + dv(t)$$

where $v(t)$ is a Wiener process. The measurable physical output of (2.1) is

$$\overline{y}(t) = \frac{1}{\Delta}\int_{t-\Delta}^{t} y(\tau)\,d\tau = \frac{1}{\Delta}\int_{t-\Delta}^{t} d\,\overline{y}(\tau), \quad \Delta > 0. \qquad (2.2)$$

Notice that using $\overline{y}(t)$, avoids problems associated with the fact that since the degree of $C(p)$ is equal to that of $A(p)$, $y(t)$ in (2.1) includes a 'white' noise component. The actual structure of $\mathcal{A}$, $\mathcal{B}$, $\mathcal{K}$ and $\mathcal{C}$ are not important at this stage.

Factorization for Continuous-time Models

Here we develop factorizations for continuous-time models. Corresponding discrete-time factorizations are derived in [12]. The following key lemma proposes an expansion of $C^{-1}(p)$ which will be exploited subsequently.

Lemma 2.1: With $C(p)$ defined in (2.1), assume $C^{-1}(p)$ to be asymptotically stable. Then in Laplace transform notation, for some real converging sequence r_k and any $a > 0$,

$$\frac{1}{C(s)} = \frac{1}{(s+a)^n}\sum_{k=0}^{\infty} r_k \left(\frac{s-a}{s+a}\right)^k, \; Re\, s \geq 0 \qquad (2.5)$$

Proof: Consider the bilinear transformation

$$z = \frac{s+a}{s-a} \iff s = a\frac{z+1}{z-1}. \qquad (2.6)$$

Setting $P(z) = C^{-1}\left(a\frac{z+1}{z-1}\right)$, then $P(z)$ is analytic outside the unit circle since (2.6) maps the left half plane into the unit disk. 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,

$$P(z) = (1 - z^{-1})^n \sum_{k=0}^{\infty} p_k z^{-k}; \; |z| \geq 1. \qquad (2.7)$$

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

Remark: The expansion (2.5) is closely related to Laguerre function representations; see [5] for details.

Corollary 2.1: The transfer function $C^{-1}(s)$ in Lemma 2.1 can be uniquely factorized as

$$\frac{1}{C(s)} = \frac{F(s)}{L(s)} + \frac{G(s)\,H(s)}{L(s)\,C(s)}$$

$$\iff L(s) = F(s)\,C(s) + G(s)\,H(s) \qquad (2.8a)$$

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.,

$$F(s) = f_{N-n} + \cdots + f_0\, s^{N-n}, \quad G(s) = g_{n-1} + \cdots + g_0\, s^{n-1} \qquad (2.8b)$$

Furthermore, given any $\epsilon > 0$, $\exists N \geq n$ such that

$$\left\|\frac{G(s)\,H(s)}{L(s)\,C(s)}\right\|_\infty \leq \epsilon \qquad (2.9)$$

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

Proof: From Lemma 2.1

$$\frac{1}{C(s)} = \frac{1}{(s+a)^n}\sum_{k=0}^{N-n} r_k\left(\frac{s-a}{s+a}\right)^k + \frac{1}{(s+a)^n}\sum_{k=N-n+1}^{\infty} r_k\left(\frac{s-a}{s+a}\right)^k; \; Re(s) \geq 0.$$

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 (2.8) and (2.9). Let $z_i, i = 1, \ldots, n$ denote the zeros of $C(s)$. Then p_k in (2.7) is of order $O(\max_i \left|\frac{z_i+a}{z_i-a}\right|^k)$. 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 Sec.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 a's, i.e., use

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

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

in (2.8).

Remark 3: Since $F(s)$ is in general not monic, we shall in the sequel work with the monic polynomial

$$\overline{F}(s) \triangleq \frac{F(s)}{f_0}. \qquad (2.10)$$

Note that by equating the coefficients of the s^n terms in (2.8a), we have $f_0 + g_0 = 1$.

In the rest of this paper we shall implicitly assume $a > 0$.

3 Continuous-time ELS scheme with Relaxed SPR condition

In this section we first achieve a transformed ELS scheme and then interpret its associated SPR condition. We state conditions for the P.E. 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 (2.1) in terms of the exponentially stable filter

$$\frac{1}{W(s)} = \frac{F(s)}{L(s)}. \qquad (3.1)$$

According to the Corollary 2.1, (3.1) is a good approximation of $C^{-1}(s)$ provided N is large enough. However, because $F(s)$ is unknown we do not use (3.1) in the actual implementation of the estimation scheme. Applying the filter (3.1) with the normalized $\overline{F}(s)$ defined in (2.10) replacing $F(s)$, (2.1) becomes

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

or equivalently, using (2.8),

$$p^{-N}\,\overline{A}(p)\,y_{LN}(t) = p^{-N}\,\overline{B}(p)\,u_{LN}(t) + \overline{g}_0\,e(t) - p^{-(n-1)}\,\overline{G}(p)\,e_H(t) + e(t). \qquad (3.2)$$

under the following definitions

$y_{LN}(t) = p^N L^{-1}(p)\,y(t)$, $u_{LN}(t) = p^N L^{-1}(p)\,u(t)$; $e_H(t) = p^{n-1} H(p)\,L^{-1}(p)\,e(t)$

$$\overline{A}(p) = A(p)\,\overline{F}(p), \; \overline{B}(p) = B(p)\,\overline{F}(p); \; \overline{G}(p) = G(p)/f_0$$

$$\text{where } \overline{G}(p) = \overline{g}_{n-1} + \cdots + \overline{g}_0\, s^{n-1}. \qquad (3.3)$$

We now estimate the coefficients of $\overline{A}(p)$, $\overline{B}(p)$ and $\overline{G}(p)$ as $\hat{\overline{A}}(p)$, $\hat{\overline{B}}(p)$ and $\hat{\overline{G}}(p)$ using ELS. To formulate the ELS scheme, let us consider the more precisely defined stochastic Ito form state equations [2]. With $\epsilon(t) \triangleq e_H(t) - e(t)$, consider (3.2) reformulated as

$$dx(t) = A\,x(t)\,dt - e_1\,dy_{LN}^{(1)}(t) + e_{N+1}\,u_{LN}(t)\,dt$$
$$-e_{2N+1}\,d\epsilon(t) - e_{2N+2}\,dv(t)$$
$$d\overline{y}_{LN}(t) = \theta_N^T\,x(t)\,dt + dv(t),\ x(0) = 0 \qquad (3.4a)$$

where $e_i^T = (0\ldots010\ldots0)$ with the 1 in the i^{th} position, $\overline{y}_{LN}(t)$ is defined similar to (2.2) and

$$x(t) = \left(-y_{LN}^{(1)}(t)\ldots - y_{LN}^{(N)}(t)\ u_{LN}^{(1)}(t)\ldots u_{LN}^{(N)}(t)\right.$$
$$\left.\epsilon(t)\ \ -e_H^{(1)}(t)\ldots - e_H^{(n-1)}(t)\right)^T$$
$$\theta_N = (\overline{a}_1\ldots\overline{a}_N\ \overline{b}_1\ldots\overline{b}_N\ \overline{g}_0\ldots\overline{g}_{n-1})^T \qquad (3.4b)$$

where $y_{LN}^{(i)}(t) = \int_0^t y_{LN}^{(i-1)}(\tau)\,d\tau$, $y_{LN}^{(1)}(t) \triangleq \int_0^t y_{LN}(\tau)\,d\tau$ and $u_{LN}^{(i)}(t)$ and $e_H^{(i)}(t)$ are defined likewise. Also, $v(t)$ is defined in the Ito formulation of (2.1) and

$$A = block\,diag\,(E_N, E_N, E_n),\ E_i = \begin{bmatrix} 0 & 0 \\ I_{i-1} & 0 \end{bmatrix}. \qquad (3.4c)$$

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

Transformed ELS scheme

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

$$d\hat{x}(t) = A\,\hat{x}(t)\,dt - e_1\,dy_{LN}^{(1)}(t) + e_{N+1}\,u_{LN}(t)\,dt$$
$$-e_{2N+1}\,d\hat{\epsilon}(t) - e_{2N+2}\,d\hat{v}(t)$$
$$d\hat{v}(t) \triangleq d\overline{y}_{LN}(t) - \hat{\theta}_N^T(t)\,\hat{x}(t)\,dt,\ \ d\hat{\theta}_N(t) = \hat{P}_t\,\hat{x}(t)\,d\hat{v}(t),$$
$$dP_t^{-1} = \hat{x}(t)\,\hat{x}^T(t)\,dt,\ d\hat{P}_t = -\hat{P}_t\,\hat{x}(t)\,\hat{x}(t)^T\,\hat{P}_t\,dt,\ P_0 > 0,\ (3.5)$$

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 (3.4b) with $e^{(\cdot)}(t)$, $e_H^{(\cdot)}(t)$ replaced by $\hat{e}^{(\cdot)}(t)$, $\hat{e}_H^{(\cdot)}(t)$. Also $\hat{e}_H^{(\cdot)}(t) \triangleq p^{n-1}\,H(p)\,L^{-1}(p)\,\hat{e}(t)$, $\hat{\epsilon}(t) \triangleq \hat{e}_H(t) - \hat{e}(t)$ and $\hat{P}_t^{-1} = \int_0^t \hat{x}(\tau)\,\hat{x}^T(\tau)\,d\tau + \hat{P}_0^{-1}$.

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

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

or equivalently

$$\left\|\frac{\overline{G}(s)\,H(s)}{L(s)} - \overline{g}_0\right\|_\infty < 1(\text{Strictly Bound Real (SBR) condition}). \qquad (3.6)$$

Overparametrization Selection to satisfy SPR condition

From Corollary 2.1 we know that

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

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 (3.6) is satisfied for all $C(s)$ whose zeros lie inside the set. In this subsection we specify N, a and the compact set.

We first seek a relationship between the unique factorization (2.8) and the unique factorization in [4]

$$\frac{1}{\overline{C}(z^{-1})} = \mathcal{F}(z^{-1}) + z^{-(N-n+1)}\,\frac{\mathcal{G}(z^{-1})}{\overline{C}(z^{-1})} \qquad (3.7a)$$

where $\overline{C}(z^{-1})$ and $\mathcal{F}(z^{-1})$ are monic polynomials of degree n and $(N-n)$ respectively and $\mathcal{G}(z^{-1})$ is a polynomial of degree $(n-1)$, i.e.,

$$\mathcal{F}(z^{-1}) = 1 + \sum_{i=1}^{N-n}\delta_i\,z^{-i};\ \ \mathcal{G}(z^{-1}) = \sum_{i=0}^{n-1}\gamma_i\,z^{-i} \qquad (3.7b).$$

The motivation is to achieve a continuous-time version of the discrete-time result in [4] namely:

Lemma 3.1: Consider any polynomial

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

with $\overline{c}_0 = 1$ such that $|z_i| \leq R < 1$ for all i. Consider also for any N, a polynomial pair $(\mathcal{F}(z^{-1}), \mathcal{G}(z^{-1}))$ with degrees $N-n$ and $n-1$ respectively, defined uniquely by the factorization (3.7). Then there exists $N_0(R)$ such that for all $N \geq N_0(R)$, $\mathcal{G}(z^{-1})$ is SBR.

Proof: See [4]. •

To lead into a continuous-time version of Lemma 3.1., we introduce the following lemma.

Lemma 3.2: The factorizations (3.7) and (2.8) are equivalent under the bilinear transformation (2.6) and the following definitions of $\overline{C}$, $\mathcal{F}$ and $\mathcal{G}$ in terms of C, F and G or vice versa:

$$\frac{C(s)}{C(a)} = \frac{(s+a)^n}{(2\,a)^n}\,\overline{C}\left(\frac{s-a}{s+a}\right),\ G(s) = (s+a)^{n-1}\,\mathcal{G}\left(\frac{s-a}{s+a}\right)$$
$$F(s) = \frac{(2\,a)^n}{C(a)}\,(s+a)^{N-n}\,\mathcal{F}\left(\frac{s-a}{s+a}\right), \qquad (3.8a)$$

or

$$\overline{C}(z^{-1}) = \frac{(1-z^{-1})^n}{C(a)}\,C\left(a\frac{z+1}{z-1}\right),$$
$$\mathcal{G}(z^{-1}) = \frac{(1-z^{-1})^{n-1}}{(2a)^{n-1}}\,G\left(a\frac{z+1}{z-1}\right),$$
$$\mathcal{F}(z^{-1}) = \frac{(1-z^{-1})^{N-n}}{(2a)^N}\,C(a)\,F\left(a\frac{z+1}{z-1}\right). \qquad (3.8b)$$

Proof: We first show that (3.7) transforms to (2.8) under (2.6) and (3.8). Multiplying (3.7) by $(1-z^{-1})^n$ we have

$$\frac{(1-z^{-1})^n}{\overline{C}(z^{-1})} = (1-z^{-1})^n\,\mathcal{F}(z^{-1}) + (1-z^{-1})^n\,z^{-(N-n+1)}\,\frac{\mathcal{G}(z^{-1})}{\overline{C}(z^{-1})}.$$

Substituting (3.8b) and (2.6) 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 (2.8).

The converse holds likewise. •

Remark 1: By its definition in (3.8), $\overline{C}(z^{-1})$ is a monic polynomial because $C(s)$ is monic. Also from (3.8a), with γ_i and δ_i defined in (3.7b), simple manipulations yield

$$f_0 = \frac{(2\,a)^n}{C(a)}(1+\delta_1+\cdots+\delta_{N-n});\ \ g_0 = \gamma_0+\gamma_1+\cdots+\gamma_{n-1}. \ (3.9a)$$

Remark 2: Note that under (2.6) and (3.8), $z^{-(N-n+1)}\,\mathcal{G}(z^{-1})$ transforms to $G(s)\,H(s)\,L^{-1}(s)$. So

$$\left|\frac{G(s)\,H(s)}{L(s)}\right|_{s=j\omega} = |\mathcal{G}(z^{-1})|_{|z|=1}. \qquad (3.9b)$$

Moreover, since stability is preserved under the bilinear transformation,

$$\frac{g_0}{f_0} - \frac{1}{f_0}\,\frac{G(s)\,H(s)}{L(s)}\ \text{SBR} \iff \frac{g_0 - \mathcal{G}(z^{-1})}{f_0}\ \text{SBR}. \qquad (3.9c)$$

o

We now present the continuous-time version of Lemma 3.1.

Lemma 3.3: Consider a polynomial $C(s) = \prod_{i=1}^{n}(s - s_i)$, $s_i < 0$ such that, its zeros lie in a circle with centre $x_0 = -a\frac{(1+R^2)}{(1-R^2)}$ and radius

$$r = \frac{2aR}{1-R^2} \tag{3.10a}$$

or equivalently such that,

$$|s_i - x_0| \le r,\ R < 1\ \forall i. \tag{3.10b}$$

Consider also that for any N, the polynomials $F(s)$, $G(s)$ and $H(s)$ are uniquely defined as in (2.8). Then there exists $N_0(r)$ such that for all $N \ge N_0(r)$,

$$\left|\frac{g_0}{f_0}\right| + \left|\frac{1}{f_0}\frac{G(s)\,H(s)}{L(s)}\right| < 1 \tag{3.11}$$

and the SBR condition (3.6) is satisfied.

Proof: Using (2.6) with r defined in (3.10) and z_i defined in Lemma 3.1, straight-forward manipulations yield

$$|z_i| \le R < 1 \iff \left|s_i + a\frac{(1+R^2)}{(1-R^2)}\right| \le r.$$

Since (3.6) is implied by (3.11), we shall look for upper bounds on $|g_0/f_0|$ and

$$|\frac{G(s)\,H(s)}{f_0\,L(s)}|_{s=j\omega} = |\frac{\mathcal{G}(z^{-1})}{f_0}|_{|z|=1}.$$

It is proved in [4] that

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

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

$$N \ge \frac{R(1+n)-1}{1-R} \tag{3.12a}$$

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

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

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

$$3R^{N-n+1}\frac{(N+n)!}{N!}(1+R)^n < 1 \tag{3.12b}$$

and (3.12a) holds then the SBR condition (3.6) is satisfied. Note that $N_0(r)$ can be defined as the smallest value of N for which (3.12) holds.

Persistence of Excitation of $x(t)$

We require the regression vector $x(t)$ and its estimate $\hat{x}(t)$ in (3.4) and (3.5) to be suitably exciting to assure strong consistency of the identification scheme and therefore its robustness to unmodelled dynamics [11].

Lemma 3.4: A necessary and sufficient condition for $x(t)$ in (3.4b) associated with signal model (3.4a) to be controllable from inputs $u(t)$, $v(t)$ is that

$$A(s)K(s),\ \ B(s)K(s) \text{ and } C(s) \text{ are coprime} \tag{3.13}$$

where

$$K(s) \triangleq H(s) - (s+a)^{N-n+1} \tag{3.14}$$

($A(s)$, $B(s)$, $C(s)$ are defined in (2.1) and $H(s)$ in (2.8).)

Proof: See [12].

Convergence of modified continuous-time ELS scheme

Standard techniques [8] apply to achieve convergence properties of the transformed ELS scheme. We summarize as the following lemma.

Theorem 3.1: For signal model (3.4) and ELS estimation scheme (3.5), if N is chosen sufficiently large so that the SBR condition (3.6) is satisfied then as $t \to \infty$

$$||\theta_N - \hat{\theta}_N(t)||^2 = O\left(\frac{\log \lambda_{max}\,\hat{P}_t^{-1}}{\lambda_{min}\,\hat{P}_t^{-1}}\right) \text{ a.s.} \tag{3.15a}$$

$$||v(t)||^2 = O(\log \lambda_{max}\,\hat{P}_t^{-1}) \text{ a.s.} \tag{3.15b}$$

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

$$\lim_{t\to\infty} \frac{\log \lambda_{max}\,P_t^{-1}}{\lambda_{min}\,P_t^{-1}} = 0 \text{ a.s.} \tag{3.16}$$

then as $t \to \infty$,

$$||\theta_N - \hat{\theta}_N(t)||^2 = O\left(\frac{\log \lambda_{max}\,P_t^{-1}}{\lambda_{min}\,P_t^{-1}}\right) \text{ a.s.} \tag{3.15c}$$

$$||v(t)||^2 = O(\log \lambda_{max}\,P_t^{-1}) \text{ a.s.} \tag{3.15d}$$

Furthermore let AK, BK, C are coprime (i.e., $x(t)$ is reachable from $u(t)$, $v(t)$ (Lemma 3.4)) and $C(s)$ be strictly minimum phase. Then as $t \to \infty$ for stable signal models (and therefore no finite escape time for $x(t)$ and $\lambda_{max}P_t^{-1} \le O(t)$) with $u(t)$, $v(t)$ suitably exciting (3.16),

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

Proof: The results (3.15a) to (3.15d) are implicitly established in [8]. Although the signal model in [8] 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 $\tilde{\theta}_N^T(t)\,\hat{x}(t)$ and output $\theta_N^T\,\tilde{x}(t) + \frac{1}{2}\tilde{\theta}_N^T(t)\,\tilde{x}(t)$ 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 (2.1) prefiltered by the exponentially stable filter $L^{-1}(s)$ is

$$d\varphi(t) = \Gamma\,\varphi(t)\,dt - e_1\,dy_{Ln}^{(1)}(t) + e_{n+1}\,u_{Ln}(t)\,dt + e_{2n+1}\,de_L^{(1)}(t)$$

$$d\overline{y}_{Ln}(t) = \varphi^T(t)\,\theta\,dt + de_L^{(1)}(t),\ \ \varphi(0) = 0 \tag{3.17a}$$

where e_i^T is defined in (3.4). Also

$$\varphi(t) = \Big(-y_{Ln}^{(1)}(t)\ldots - y_{Ln}^{(n)}(t)\ u_{Ln}^{(1)}(t)\ldots u_{Ln}^{(n)}(t)$$
$$e_L^{(1)}(t)\ldots e_L^{(n)}(t)\Big)^T$$

$$\theta = (a_1\ldots a_n\ b_1\ldots b_n\ c_1\ldots c_n)^T,\ \ e_L^{(i)}(t) = p^n\frac{e^{(i)}(t)}{L(p)} \tag{3.17b}$$

where $y_{Ln}(t) = p^n\,L^{-1}(p)\,y(t)$, etc. The superscripts in (3.17a,b) are defined in (3.4b), and

$$\Gamma = block\,diag\,(E_n, E_n, E_n),\ E_n = \begin{bmatrix} 0 & 0 \\ I_{n-1} & 0 \end{bmatrix}. \tag{3.17c}$$

So far we have proved that consistent estimates of parameters of the transformed signal model (3.4) can be obtained under the relaxed SBR condition (3.6) and P.E. condition (3.16). 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 to the ELS algorithm is utilized:

$$d\hat{\theta}(t) = \overline{P}_t\,\overline{\varphi}(t)\,(d\overline{y}_{Ln}(t) - \hat{\theta}^T(t)\,\overline{\varphi}(t)\,dt)$$

$$\overline{P}_t = \left[\int_0^t \overline{\varphi}(\tau)\,\overline{\varphi}(\tau)^T\,d\tau + \overline{P}_0^{-1}\right]^{-1},\ \ \overline{P}_0 > 0 \tag{3.18a}$$

where

$$\overline{\varphi}(t) = (-y_{Ln}^{(1)}(t)\ldots - y_{Ln}^{(n)}(t)\ u_{Ln}^{(1)}(t)\ldots u_{Ln}^{(n)}(t)\ \hat{e}_L^{(1)}(t)\ldots\hat{e}_L^{(n)}(t))^T \tag{3.18b}$$

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

Remark: It may be thought that $d\bar{y}_{Ln}(t) - \varphi^T(t)\,\theta\,dt$ in (3.17a) should be the differential of a Wiener process for the L.S. scheme to converge. However, as shown in [2], this is not the case as long as $L^{-1}(p)$ is exponentially stable and $de^{(1)}(t)$ $(= dv(t)$, see (2.1)) 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 (3.18), despite its similarity to (3.5), has an almost standard least squares form. The only nonstandard feature of the proposed scheme is that the regression vector (3.18b) differs from the *true* one where $e_L^{(\cdot)}(t)$ would be present instead of $\hat{e}_L^{(\cdot)}(t)$. We now 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 3.2: Consider the least square scheme (3.18) with signal model (2.1) and (3.17) under the relaxed SBR condition (3.6), where $v_L^{(i)}(t)$ is obtained from the ELS scheme (3.5). Then as $t \to \infty$,

$$\|\theta - \hat{\theta}(t)\|^2 = O\left(\frac{\log \lambda_{max} \overline{P}_t^{-1}}{\lambda_{min} \overline{P}_t^{-1}}\right) \quad \text{a.s.} \qquad (3.19)$$

Moreover, under (3.16) with P_t defined in (3.5), as $t \to \infty$

$$\|\theta - \hat{\theta}(t)\|^2 = O\left(\frac{\log \lambda_{max} P_t^{-1}}{\lambda_{min} P_t^{-1}}\right) \quad \text{a.s.} \qquad (3.20)$$

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

$$\|\theta - \hat{\theta}(t)\|^2 = O(t^{-1} \log t) \quad \text{a.s.} \qquad (3.21)$$

Proof: Considering (3.17)and (3.18) define

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

Then

$$\begin{aligned} d\tilde{\theta}(t) &= -d\hat{\theta}(t) = -\overline{P}_t \overline{\varphi}(t)\,(d\overline{y}_{Ln}(t) - \hat{\theta}(t)^T \overline{\varphi}(t)\,dt) \\ &= [-\overline{P}_t \overline{\varphi}(t)\,(\overline{\varphi}(t)\,\theta\,dt + de_L^{(1)}(t) - \hat{\theta}(t)^T \overline{\varphi}(t)\,dt)] \\ &\quad - \overline{P}_t \overline{\varphi}(t)\,\tilde{\varphi}(t)^T \theta\,dt \end{aligned} \qquad (3.22)$$

The term enclosed in square brackets in (3.22) is the least squares parameter error associated with the model $d\bar{y}_{Ln}(t) = \overline{\varphi}(t)^T \theta\,dt + de_L^{(1)}(t)$, so that by known results [8] (in fact an appropriate specialization of Theorem 3.3),

$$\begin{aligned} &\|\overline{P}_t \overline{\varphi}(t)(\overline{\varphi}^T(t)\,\theta\,dt + dv_L(t) - \hat{\theta}^T(t)\,\overline{\varphi}(t)\,dt)\|^2 \\ &= O\left(\frac{\log \lambda_{max} \overline{P}_t^{-1}}{\lambda_{min} \overline{P}_t^{-1}}\right) \end{aligned} \qquad (3.23)$$

The second term in (3.22) is bounded as follows: Using the Schwartz inequality

$$\|\int_0^T \overline{P}_t \overline{\varphi}(t)\,\tilde{\varphi}^T(t)\,\theta\,dt\|^2 \le \int_0^T |\overline{P}_t \overline{\varphi}(t)|^2\,dt \int_0^T |\tilde{\varphi}^T(t)\,\theta|^2\,dt \qquad (3.24)$$

It is established in [8, Lemma 3.1, Theorem 3.1] that for constants k_1 and k_2,

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

where $\hat{P}_t^{-1}$ is defined in (3.5). Thus

$$\int_0^T \|\theta^T \tilde{\varphi}(t)\|^2\,dt = O(\log tr \hat{P}_t^{-1}). \qquad (3.25)$$

Also,

$$\int_0^T \|\overline{P}_t \overline{\varphi}(t)\|^2\,dt = \int_0^T tr\,[\overline{P}_t \overline{\varphi}(t)\,\overline{\varphi}(t)^T \overline{P}_t^T]\,dt = -tr\,\overline{P}_t \qquad (3.26)$$

where the last equality follows from differentiating $\overline{P}_t \overline{P}_t^{-1} = I$ with respect to time and using (3.18a). Therefore with (3.25) and (3.26) substituted in (3.24),

$$\begin{aligned} \|\int_0^T \overline{P}_t \overline{\varphi}(t)\,\tilde{\varphi}(t)^T \theta\,dt\|^2 &= O((\log tr \hat{P}_t^{-1})(tr \overline{P}_t)) \\ &= O\left(\frac{\log \lambda_{max} \hat{P}_t^{-1}}{\lambda_{min} \overline{P}_t^{-1}}\right). \end{aligned} \qquad (3.27)$$

But, recalling the definitions of $\hat{x}(t)$ in (3.4b) and $\overline{\varphi}(t)$ in (3.18),

$$\begin{aligned} \lambda_{max} \hat{P}_t^{-1} &= O(tr \int_0^t \hat{x}(\tau)\,\hat{x}(\tau)^T\,d\tau) \\ &= O(tr \int_0^t \overline{\varphi}(\tau)\,\overline{\varphi}(\tau)^T\,d\tau) = O(\lambda_{max} \overline{P}_t^{-1}) \end{aligned}$$

because $L^{-1}(s)$ and $H(s)\,L^{-1}(s)$ in $x(t)$ are stable transfer functions. The desired result (3.19) then follows from (3.23) and (3.27). Also, under (3.16), we have [8]

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

thus establishing (3.20). The same arguments [8] used to prove Theorem 3.3, establish (3.21).

4 Conclusions

The strength of ELS schemes 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, 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. 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.

References

[1] G.C. Goodwin and K.S. Sin, *Adaptive Filtering, Prediction and Control*, Prentice Hall, 1984.

[2] J.B. Moore, *Convergence of Continuous-time Stochastic ELS Parameter Estimation*, Stochastic Processes and their Applications, Vol.27, pp 195-215, 1988.

[3] M. Gevers, V. Wertz, G.C. Goodwin, *A Parameter Estimation Algorithm for Continuous time Stochastic Adaptive Control*, Proc. 27th IEEE Conf. on Decision and Control, pp 1922-1923, Dec. 1988.

[4] J.B. Moore, M. Niedzwiecki, L. Xia, *Identification/Prediction Algorithms for ARMAX models with Relaxed Positive Real Conditions*, 1988.

[5] B. Wahlberg, *System Identification using high-order models, revisited*, Technical Report EE/CICS8903, Univ. of Newcastle, 1989.

[6] M. Green, J.B. Moore, *Persistence of Excitation in Linear Systems*, Systems and Control Letters, Vol.7, pp 351-360, 1986.

[7] H. Chen and L. Guo, *Convergence rate of Least Squares Identification and Adaptive control for Stochastic Systems*, Int. J. Contr., Vol.44, pp 1459-1476, 1986.

[8] H. Chen, J.B. Moore, *Convergence rates for Continuous-time Stochastic ELS Parameter Estimation*, IEEE Trans. Auto. Contr., Vol.AC-32, No.3, March 1987.

[9] N. Shimkin, A Feuer, *Persistency of Excitation in Continuous time Systems*, System and Control Letters, Vol.9, No.3, Sept. 1987.

[10] C.T. Chen, *Introduction to Linear Systems Theory*, Holt Rinehart Winston, 1970.

[11] H. Chen, L. Guo, *A Robust Stochastic Adaptive Controller*, IEEE Trans. Auto. Contr., Vol.33, No.11, pp 1035-1043, Nov. 1988.

[12] V. Krishnamurthy, B. Wahlberg, J.B. Moore, *Factorizations that relax the Positive Real Condition in Continuous-time and Fast-Sampled ELS Schemes*, July 1989.

[13] S. Shah, G.F. Franklin, *On satisfying strict positive real condition for convergence by overparametrisation*, IEEE Trans. Auto. Contr., Vol.AC-27, pp 715-16.