International Journal of Control

ISSN: 0020-7179 (Print) 1366-5820 (Online) Journal homepage: https://www.tandfonline.com/loi/tcon20

Robust recursive identification of multidimensional linear regression models

LEI GUO , LIGE XIA & JOHN B. MOORE

To cite this article: LEI GUO , LIGE XIA & JOHN B. MOORE (1988) Robust recursive identification of multidimensional linear regression models, International Journal of Control, 48:3, 961-979, DOI: 10.1080/00207178808906229

To link to this article: https://doi.org/10.1080/00207178808906229

Published online: 18 Jan 2007.

INT. J. CONTROL, 1988, VOL. 48, NO. 3, 961–979

Robust recursive identification of multidimensional linear regression models

LEI GUO†, LIGE XIA† and JOHN B. MOORE†

Stochastic adaptive estimation and control algorithms involving recursive prediction estimates have guaranteed convergence rates when the noise is not 'too' coloured, as when a positive-real condition on the noise model is satisfied. Moreover, the whiter the noise environment the more robust are the algorithms. This paper shows that for linear regression signal models, the suitable introduction of white noise into the estimation algorithm can make it more robust without compromising on convergence rates. Indeed, there are guaranteed attractive convergence rates independent of the process noise colour. No positive-real condition is imposed on the noise model.

1. Introduction

Precise convergence rates are known for a number of stochastic adaptive schemes under a certain noise model positive-real condition (Chen and Guo 1986, Lai and Wei 1986 a) first exposed as a convergence condition by Ledwich and Moore (1977) and Ljung (1977). Robustness results are also known (Chen and Guo 1987 a). The whiter the process noise, the more likely the positive-real condition is satisfied, and the more robust are the algorithms.

In an earlier paper (Moore 1982), a method was proposed to side-step the positive-real condition for scalar variable noise models in stochastic adaptive estimation and control. The method has as a starting point the addition of white noise into the processing. Such additions ensure a whiter noise environment, which in turn ensures convergence and lends a certain robustness. The added noise can be seen as dominating unmodelled dynamics or unmodelled coloured noise. The method is made more powerful by additional processing involving on-line spectral factorization and parallel processing involving pre-whitening filters. Simulations support the ideas of Moore (1982), although his theory is incomplete.

In this paper and companion papers (Guo and Moore 1987), a number of the ideas of Moore (1982) are re-packaged in the context of a precise convergence analysis with the view to quantifying the extent of robustness enhancement and convergence rates. The techniques build on Kalman filtering theory, spectral factorization theory and expand on those used for extended least-squares convergence by Chen and Guo (1986) and Lai and Wei (1986 a). The earlier work (Moore 1982) is non-trivially generalized to cope with multivariable signal models. Convergence rates are guaranteed without imposition of a positive-real condition on the coloured noise model.

Received 18 September 1987.

† Department of Systems Engineering, Research School of Physical Sciences, Australian National University, G.P.O. Box 4, Canberra, A.C.T., 2601, Australia.

2. Algorithm description and main results

2.1. *Stochastic model*

Consider the following m-dimensional linear regression model:

$$y(t) = \theta_0 x(t) + \varepsilon(t), \quad t \geq 0 \tag{2.1}$$

where $y(t)$, $x(t)$ and $\varepsilon(t)$ are the m-, p- and m-dimensional observation vector, regression vector and modelling error, respectively, and where θ_0 is the $m \times p$ unknown parameter matrix.

Assume that the system noise $\varepsilon(t)$ is a moving average (MA) process

$$\varepsilon(t) = w(t) + C_1 w(t-1) + \dots + C_r w(t-r), \quad t \geq 0 \tag{2.2}$$

with unknown matrix coefficients C_i, $1 \leq i \leq r$, where the driven noise $\{w(t)\}$ is assumed to be a gaussian white noise sequence with

$$Ew(t) = 0, \quad Ew(t)w^\tau(t) = R_w > 0, \quad t \geq 0 \tag{2.3}$$

Let us denote all the unknown parameters by

$$\theta = [\theta_0, C_1, \dots, C_r]^\tau \tag{2.4}$$

2.2. *Introduced noise*

To dominate unmodelled dynamics and/or noise that is highly coloured, consider the introduction of a gaussian white noise sequence $\{v(t)\}$ that is independent of $\{w(t)\}$ with properties

$$Ev(t) = 0, \quad Ev(t)v^\tau(t) = \sigma_v^2 I_m, \quad \sigma_v^2 > 0 \tag{2.5}$$

The 'pre-whitening' idea proposed by Moore (1982) is to formulate the predictor in the identification algorithm by using the following 'pre-whitened' process

$$z(t) = y(t) + v(t), \quad t \geq 0 \tag{2.6}$$

together with the output sequence $\{y(t)\}$.

2.3. *Prediction error algorithm*

Consider the prediction error algorithm processing (2.6)

$$\hat{\theta}(t+1) = \hat{\theta}(t) + P(t)\psi(t)[z^\tau(t+1) - \psi^\tau(t)\hat{\theta}(t)] \tag{2.7}$$

$$P(t) = P(t-1) - \frac{P(t-1)\psi(t)\psi^\tau(t)P(t-1)}{1 + \psi^\tau(t)P(t-1)\psi(t)}, \quad P(0) > 0 \tag{2.8}$$

$$\psi(t) = [x^\tau(t) \quad z^\tau(t) - \psi^\tau(t-1)\hat{\theta}(t) \quad \dots \quad z^\tau(t-r+1) - \psi^\tau(t-r)\hat{\theta}(t-r+1)]^\tau \tag{2.9}$$

Estimates of the covariance of prediction errors are given by the following residual statistics

$$\hat{R}_{\bar{w}}(t) = \frac{1}{t}\sum_{i=0}^{t-1} [z(i+1) - \hat{\theta}^\tau(t)\psi(i)][z(i+1) - \hat{\theta}^\tau(t)\psi(i)]^\tau \tag{2.10}$$

the terms of which have convenient recursive forms. Notice that when the introduced noise $v(t)$ in (2.6) is set to zero, so that $z(t) = y(t)$, then (2.9) reduces to the more 'standard' regression vector.

2.4. *Theorems*

Let us denote $\lambda_{\min}(X)[\lambda_{\max}(X)]$ as the minimum [maximum] eigenvalue of a matrix X and $\|X\| = [\lambda_{\max}(XX^*)]^{1/2}$ its norm, where X^* is the transpose complex conjugate of X. Let us also denote $\zeta(t) = \varepsilon(t) + v(t)$ and set

$$G_t^0 = \sigma\{\zeta(i), i \leqslant t\}, \quad t \geqslant 0$$

Assume that the regression vector sequence $\{x(t), F_{t-1}\}$ is any adapted random sequence where

$$F_t = \sigma\{G_t^0 \cup G_t^1\}, \quad t \geqslant 0$$

with $\{G_t^1\}$ being any family of non-decreasing σ-algebras such that G_t^1 is independent of G_{t+1}^0 for any $t \geqslant 0$.

Theorem 2.1

For the system and algorithm described by (2.1)–(2.10), if in the pre-whitening of (2.5), (2.6), σ_v^2 is chosen to satisfy

$$\sigma_v^2 > r\|R_w\| \cdot \|[C_1, \ldots, C_r]\|^2 - \lambda_{\min}(R_w) \tag{2.11}$$

then the following convergence rates hold

$$\text{(i)} \quad \|\hat{\theta}(t+1) - \bar{\theta}\| = 0\left(\frac{\log \lambda_{\max}(t)}{\lambda_{\min}(t)}\right)^{1/2}, \quad \text{a.s.} \quad t \to \infty \tag{2.12}$$

$$\text{(ii)} \quad \|\hat{R}_{\bar{w}}(t) - R_{\bar{w}}\| = 0\left(\frac{\log \log t}{t}\right)^{1/2} + 0\left(\frac{\log \lambda_{\max}(t)}{t}\right), \quad \text{a.s.} \quad t \to \infty \tag{2.13}$$

Here

$$\bar{\theta} = [\theta_0 \quad D_1 \quad \ldots \quad D_r]^\tau \tag{2.14}$$

and $[D_i, 1 \leqslant i \leqslant r, R_{\bar{w}}\}$ satisfies

$$D(z)R_{\bar{w}}D^\tau(z^{-1}) = C(z)R_w C^\tau(z^{-1}) + \sigma_v^2 I \tag{2.15}$$

with

$$C(z) \triangleq I + C_1 z + \ldots + C_r z^r \tag{2.16}$$

$$D(z) \triangleq I + D_1 z + \ldots + D_r z^r \tag{2.17}$$

Here also, $\lambda_{\max}(t)[\lambda_{\min}(t)]$ denotes the maximum [minimum] eigenvalues of

$$\sum_{i=0}^{t} \psi(i)\psi^\tau(i) + \varepsilon I, \quad \varepsilon > 0$$

Theorem 2.2

Consider that the conditions of Theorem 1.1 apply and

$$\log \lambda_{\max}^0(t) = o[\lambda_{\min}^0(t)], \quad \text{a.s.} \quad t \to \infty \tag{2.18}$$

Then

$$\|\hat{\theta}(t+1) - \bar{\theta}\| = 0\left(\frac{\log \lambda_{\max}^0(t)}{\lambda_{\min}^0(t)}\right)^{1/2}, \quad \text{a.s.} \quad t \to \infty \tag{2.19}$$

and

$$\|\hat{R}_{\bar{w}}(t) - R_{\bar{w}}\| = 0\left(\frac{\log\log t}{t}\right)^{1/2} + 0\left(\frac{\log \lambda^0_{\max}(t)}{t}\right), \quad \text{a.s.} \quad t \to \infty \tag{2.20}$$

Here, $\lambda^0_{\max}(t)\,[\lambda^0_{\min}(t)]$ denotes the maximum [minimum] eigenvalue of

$$\sum_{i=0}^{t} \psi^0(i)\psi^0(t)^\tau + \varepsilon I$$

with

$$\psi^0(t) \triangleq [x^\tau(t) \quad \bar{w}^\tau(t) \quad \dots \quad \bar{w}^\tau(t-r+1)]^\tau \tag{2.21}$$

and $[\bar{w}(t), F_t\}$ is a gaussian martingale difference sequence with

$$E\bar{w}(t)\bar{w}^\tau(t) \xrightarrow[t\to\infty]{} R_{\bar{w}}$$

and satisfies, under (2.16)

$$C(z)w(t) + v(t) = D(z)\bar{w}(t) + 0\,(\exp(-\alpha t)), \quad \text{for some } \alpha > 0$$

Remark 1

The classical linear regression model considered in mathematical statistics is a specialization of (2.1) with the so-called 'design vector' $x(t)$ deterministic and with the noise $\varepsilon(t)$ white. Obviously the stochastic model (2.1), (2.2) considered in this paper is a more general one, namely, we allow the regression vector $x(t)$ to be a class of random vectors and the modelling error $\varepsilon(t)$ to be correlated. However, the restriction that $x(t) \in F_{t-1}$ is essential to the convergence analysis excludes the specialization of (2.1), (2.2) to general ARMAX models—a crucial point not observed by Moore (1982).

Remark 2

For the case when the added noise $v(t)$ in (2.6) is zero, then the condition (2.11) is usually replaced by a positive-real condition on the noise model (even for the case where one is only interested in identifying θ_0). In particular it is required that

$$[C^{-1}(z) - \tfrac{1}{2}I] \quad \text{is strictly positive-real} \tag{2.23}$$

(This condition is equivalent to $[C(z) - I]$ is strictly bounded-real. To see the equivalence, recall that $X(z)$ is bounded-real if and only if $Z(z) = [I - X(z)]$ $[I + X(z)]^{-1}$ is positive-real.) It is the addition of sufficient noise into the algorithm that obviates the need for such a condition in Theorem 2.2. In identifying (2.1) with $C(z)$ unknown, (2.23) cannot be checked *a priori*. In contrast, the condition (2.11) can be satisfied *a priori* with only a limited knowledge of the 'unknown' process, namely some upper bound on $\|R_w\|$ and $\|[C_1, \dots, C_r]\|$. In the scalar variable case, an upper bound on the term $\|[C_1, \dots, C_r]\|$ is numerically readily obtained since, without loss of generality, $C(z)$ can be minimum phase. In this case, it is readily shown (see the Appendix) that

$$\|[C_1, \dots, C_r]\| < [(2r)!(r!)^{-2} - 1]^{1/2} \tag{2.24}$$

Remark 3

Estimates $\hat{C}(t)$ and $\hat{R}_w(t)$ converging at the rates above to $C(z)$ and R_w can be determined from estimates $\hat{D}(t)$, $\hat{R}_{\bar{w}}(t)$ by an on-line spectral factorization corresponding to the off-line version (2.15). Details are given by Guo and Moore (1987). Of course, it is immediate that $C(z)$ and R_w can be uniquely determined from $D(z)$ and $R_{\bar{w}}$ via (2.15) to within an all-pass factor. Without loss of generality we can take $C(z)$ minimum phase. In this case $C(z)$ is uniquely determined from $D(z)$, $R_{\bar{w}}$.

Remark 4

The convergence rates of the estimates $\hat{\theta}(t)$ are virtually the same as that given in earlier theory for multivariable ARMAX models with $v(t)$ zero and (2.23) holding (Chen and Guo 1986). Of course, the covariance of the different error terms is inevitably higher because of the added noise, but this need not be the case with the additional processing proposed by Moore (1982).

Remark 5

The requirement that $w(t)$, $v(t)$ be gaussian is a technical one required by the particular martingale convergence theorem employed in subsequent analysis. It appears that by defining martingales in terms of orthogonal projections rather than in terms of conditional expectations could relax this requirement. Details are not explored here. Certainly simulations suggest that the gaussian assumption is overly strong.

3. Preliminary theory

Let $\{z(t), F_t\}$ and $\{\psi(t), F_t\}$ be two sequences of adapted random vectors (not necessarily defined by (2.6) and (2.9)). Consider the following general prediction error algorithm based on a predictor $\hat{z}(t, \theta) \triangleq \hat{z}\{t, \theta, [z(0), \ldots, z(t-1)]\} \in F_{t-1}$

$$\hat{\theta}(t+1) = \hat{\theta}(t) + P(t)\psi(t)\{z^\tau(t+1) - z^\tau[t+1, \hat{\theta}(t)]\} \tag{3.1 a}$$

$$P(t) = P(t-1) - \frac{P(t-1)\psi(t)\psi^\tau(t)P(t-1)}{1 + \psi^\tau(t)P(t-1)\psi(t)}, \quad P(0) > 0 \tag{3.1 b}$$

Set

$$a(t) \triangleq [1 + \psi^\tau(t)P(t-1)\psi(t)]^{-1} \tag{3.2}$$

$$\xi(t) \triangleq a(t-1)[z(t) - \hat{z}(t, \hat{\theta}(t-1))] - \{z(t) - E[z(t) \mid F_{t-1}]\} \tag{3.3}$$

$$S_t(\theta, \alpha) \triangleq \sum_{i=1}^{t} \left[\xi^\tau(i+1)\tilde{\theta}^\tau(i+1)\psi(i) - \frac{1+\alpha}{2} \|\tilde{\theta}^\tau(i+1)\psi(i)\|^2 \right], \quad \alpha > 0 \tag{3.4}$$

where $\tilde{\theta}(t) \triangleq \theta - \hat{\theta}(t)$ and θ is an arbitrary matrix of appropriate dimensions.

Lemma 3.1

Suppose that the adapted sequence $\{z(t), F_t\}$ satisfies

$$\sup_{t \geq 0} E[\|z(t) - E[z(t) \mid F_{t-1}]\|^\beta \mid F_{t-1}] < \infty, \quad \text{a.s.} \tag{3.5}$$

for some $\beta \geq 2$. Then, for any θ and any $\alpha > 0$, the estimate $\hat{\theta}(t)$ given by (3.1) satisfies the following relation

$$\|\hat{\theta}(t+1)-\theta\|^2 \leqslant \frac{-2S_t(\theta,\alpha)}{\lambda_{\min}(t)} + 0\left(\frac{\log^{[1+\alpha\delta(\beta-2)]}\lambda_{\max}(t)}{\lambda_{\min}(t)}\right), \quad \text{a.s.} \tag{3.6}$$

where $\lambda_{\max}(t)=\lambda_{\max}[P^{-1}(t)]$ and $S_t(\theta,\alpha)$ is defined in (3.4), and where $\delta(x)\triangleq 0$ for $x>0$ and $\delta(x)\triangleq 1$ for $x=0$.

Proof

See the Appendix.

Remark 6

The proof techniques follow closely those of Chen and Guo (1986), but the result is in fact more general than theirs. Here (3.1) is a more general prediction error scheme than that of Chen and Guo (1986), which is an extended least-squares scheme with $z[t+1,\hat{\theta}(t)]=\hat{\theta}^\tau(t)\psi(t)$.

Lemma 3.2

Consider that the conditions of Lemma 3.1 apply. Consider also that at some point θ, $E[z(t+1)\,|\,F_t]$ can be expressed by

$$E[z(t+1)\,|\,F_t]=\hat{z}[t+1,\hat{\theta}(t)]+\tilde{\theta}^\tau(t)\psi(t)+[H(z)-I]\tilde{\theta}^\tau(t+1)\psi(t)+\delta(t) \tag{3.7}$$

where $\tilde{\theta}(t)=\theta-\hat{\theta}(t)$, and $\delta(t)$ is an F_t-measurable random vector. Then, if the transfer matrix $H(z)-[(1+\alpha_0)/2]I$, $\alpha_0>1$, is positive-real, then the following expansion holds

(i) $$\|\hat{\theta}(t+1)-\theta\|^2 \leqslant 0\left[\frac{\log^{[1+\alpha\delta(\beta-2)]}\lambda_{\max}(t)}{\lambda_{\min}(t)}\right]+\frac{2}{\alpha_0-\alpha}\left[\frac{\sum_{i=1}^{t}\|\delta(i)\|^2}{\lambda_{\min}(t)}\right] \tag{3.8}$$

(ii) $$\sum_{i=1}^{t}\|\tilde{\theta}^\tau(i+1)\psi(i)\|^2 \leqslant 0\{\log^{[1+\alpha\delta(\beta-2)]}\lambda_{\max}(t)\}+\frac{4}{(\alpha_0-\alpha)^2}\sum_{i=1}^{t}\|\delta(i)\|^2 \tag{3.9}$$

for any $\alpha\in(0,\alpha_0)$.

Proof

By (A 1) and (A 4) in the Appendix and (3.7) we see that $\xi(t+1)$ defined by (3.3) can be rewritten as

$$\begin{aligned}\xi(t+1)&=a(t)\{e(t+1)+E[z(t+1)\,|\,F_t]-\hat{z}(t+1,\hat{\theta}(t)\}-e(t+1)\\&=a(t)\{e(t+1)+\tilde{\theta}^\tau(t)\psi(t)+[H(z)-I]\tilde{\theta}^\tau(t+1)\psi(t)+\delta(t)\}-e(t+1)\\&=a(t)\{e(t+1)+[\tilde{\theta}^\tau t+1)+P(t-1)\psi(t)[\xi^\tau(t+1)+e^\tau(t+1)]]^\tau\psi(t)\\&\qquad+[H(z)-I]\tilde{\theta}^\tau(t+1)\psi(t)+\delta(t)\}-e(t+1)\\&=a(t)\{[1+\psi^\tau(t)P(t-1)\psi(t)]e(t+1)+\xi(t+1)\psi^\tau(t)P(t-1)\psi(t)\\&\qquad+H(z)\tilde{\theta}^\tau(t+1)\psi(t)+\delta(t)\}-e(t+1)\end{aligned}$$

From here we immediately obtain

$$\begin{aligned} a(t)\xi(t+1) &= [1-\psi^\tau(t)P(t-1)\psi(t)a(t)]\xi(t+1) \\ &= a(t)[H(z)\tilde{\theta}^\tau(t+1)\psi(t)+\delta(t)] \end{aligned}$$

and consequently

$$\xi(t+1) = H(z)\tilde{\theta}^\tau(t+1)\psi(t)+\delta(t) \tag{3.10}$$

Since $H(z)-[(1+\alpha_0)/2]I$ is positive-real, there exists constants K_0 so that for all $\alpha\in(0,\alpha_0)$, from (3.4) and (3.10)

$$\begin{aligned} S_t(\theta,\alpha) &= \sum_{i=1}^{t}\Big\{[H(z)\tilde{\theta}^\tau(i+1)\psi(i)+\delta(i)]^\tau\tilde{\theta}^\tau(i+1)\psi(i) \\ &\qquad -\frac{1+\alpha}{2}\|\tilde{\theta}^\tau(i+1)\psi(i)\|^2\Big\} \\ &= \sum_{i=1}^{t}\left[\left(H(z)-\frac{1+\alpha_0}{2}\right)\tilde{\theta}^\tau(i+1)\psi(i)\right]^\tau\cdot\tilde{\theta}^\tau(i+1)\psi(i)+K_0 \\ &\quad +\sum_{i=1}^{t}\delta^\tau(i)\tilde{\theta}^\tau(i+1)\psi(i)+\frac{\alpha_0-\alpha}{2}\sum_{i=1}^{t}\|\tilde{\theta}^\tau(i+1)\psi(i)\|^2-K_0 \\ &\geqslant \sum_{i=1}^{t}\delta^\tau(i)\tilde{\theta}^\tau(i+1)\psi(i)+\frac{\alpha_0-\alpha}{2}\sum_{i=1}^{t}\|\tilde{\theta}^\tau(i+1)\psi(i)\|^2-K_0 \end{aligned}$$

By the elementary inequality

$$\|a^\tau b\|\leqslant\frac{1}{2\varepsilon}\|a\|^2+\frac{\varepsilon}{2}\|b\|^2,\quad\forall\epsilon>0$$

we see that for any $\alpha\in(0,\alpha_0)$

$$\sum_{i=1}^{t}|\delta^\tau(i)\tilde{\theta}^\tau(i+1)\psi(i)| = \frac{1}{\alpha_0-\alpha}\sum_{i=1}^{t}\|\delta(i)\|^2+\frac{\alpha_0-\alpha}{4}\sum_{i=1}^{t}\|\tilde{\theta}^\tau(i+1)\psi(i)\|^2 \tag{3.12}$$

Finally, by (3.11), (3.12) and Lemma 3.1, it follows that for any $\alpha\in(0,\alpha_0)$

$$\begin{aligned} \|\hat{\theta}(t+1)-\theta\|^2 &\leqslant \frac{-2\sum_{i=1}^{t}\delta^\tau(i)\tilde{\theta}^\tau(i+1)\psi(i)}{\lambda_{\min}(t)}-\frac{(\alpha_0-\alpha)\sum_{i=1}^{t}\|\tilde{\theta}^\tau(i+1)\psi(i)\|^2}{\lambda_{\min}(t)} \\ &\quad +0\left(\frac{\log^{[1+\alpha\delta(\beta-2)]}\lambda_{\max}(t)}{\lambda_{\min}(t)}\right) \\ &\leqslant \frac{2}{\alpha_0-\alpha}\frac{\sum_{i=1}^{t}\|\delta(i)\|^2}{\lambda_{\min}(t)}-\frac{\alpha_0-\alpha}{2}\frac{\sum_{i=1}^{t}\|\tilde{\theta}^\tau(i+1)\psi(i)\|^2}{\lambda_{\min}(t)} \\ &\quad +0\left(\frac{\log^{[1+\alpha\beta(\beta-2)]}\lambda_{\max}(t)}{\lambda_{\min}(t)}\right) \end{aligned}$$

and then the conclusions of the lemma follow immediately.

Remark 7

In Lemmas 3.1 and 3.2, no models are pre-postulated for $\{z(t), F_t\}$ and so the process $\{z(t)\}$ can be generated from an ARMAX model, since the restriction $x(t) \in F_{t-1}$ of Theorems 2.1 and 2.2 is not imposed here. When $\{z(t), F_t\}$ with $F_t = \sigma\{z_i, i \leqslant t\}$ is generated from an ARMAX model and $\hat{\theta}(t)$ is given by the standard extended least-squares algorithm, the process $\{\delta(t)\}$ appearing in Lemma 3.2 is zero. However, when there are unmodelled dynamics and time variations of the coefficients, then $\{\delta(t)\}$ is no longer zero (Chen and Guo 1987 a). Lemma 3.2 provides a unified approach to the convergence/robustness analysis of general prediction error algorithms such as pseudo-linear regression with or without filtering (Goodwin and Sin 1984). It is worth noting that the robustness properties of the algorithm are closely related to the passivity margin of the transfer function concerned (see (3.8)). Other applications of Lemma 3.2 will be noted elsewhere.

Lemma 3.3

Let $C(z)$ be defined as in (2.16), and $\{w(t)\}$ and $\{v(t)\}$ be defined as in (2.3) and (2.5), then there exists a gaussian martingale difference sequence $\{\bar{w}(t), G_t^0\}$ with

$$E\bar{w}(t)\bar{w}^\tau(t) \xrightarrow[t\to\infty]{} R_{\bar{w}}, \quad \text{exponentially fast}$$

and a matrix polynomial $D(z)$

$$D(z) = I + D_1 z + \ldots + D_r z^r \tag{3.13}$$

such that

$$C(z)w(t) + v(t) = D(z)\bar{w}(t) + \eta(t) \tag{3.14}$$

where $\eta(t)$ is G_{t-1}^0-measurable and exponentially tending to zero as $t \to \infty$. Moreover, for any $\alpha_0 \in [0, 1)$, if

$$\sigma_v^2 \geqslant r\left(\frac{1+\alpha_0}{1-\alpha_0}\right)^2 \|R_w\| \cdot \|[C_1, \ldots, C_r]\|^2 - \lambda_{\min}(R_w) \tag{3.15}$$

then

$$D^{-1}(z) - \frac{1+\alpha_0}{2} I \tag{3.16}$$

is positive-real.

Proof

Define $\zeta(t)$ as in § 2.4

$$\zeta(t) = C(z)w(t) + v(t) \tag{3.17}$$

and set

$$A = \underbrace{\begin{bmatrix} 0 & I_m & \ldots & 0 \\ & & \ddots & \\ & & & I_m \\ 0 & & \ldots & 0 \end{bmatrix}}_{m \times (r+1)}, \quad C = \begin{bmatrix} I_m \\ C_1 \\ \vdots \\ C_r \end{bmatrix}, \quad H = \underbrace{[I_m \quad 0 \quad \ldots \quad 0]}_{m \times (r+1)} \tag{3.18}$$

then $\zeta(t)$ can be expressed by

$$x^*(t+1) = Ax^*(t) + Cw(t+1), \quad \zeta(t) = Hx^*(t) + v(t)$$

According to the Kalman filtering theory $\zeta(t)$ can be generated by the following innovation model (Anderson and Moore 1979)

$$\hat{x}^*(t+1) = A\hat{x}^*(t) + K(t)\bar{w}(t), \quad \zeta(t) = H\hat{x}^*(t) + \bar{w}(t) \tag{3.19}$$

where $\hat{x}(t)$ is the estimator for $x(t)$ and $K(t)$ is the filter gain given by

$$K(t) = AP(t)H^\tau[HP(t)H^\tau + \sigma_v^2 I]^{-1} \tag{3.20}$$

$$P(t+1) = AP(t)A^\tau - AP(t)H^\tau[HP(t)H^\tau + \sigma_v^2 I]^{-1}HP(t)A^\tau + CR_wC^\tau \tag{3.21}$$

and where the innovation process $\{\bar{w}(t), G_t^0\}$ is a gaussian martingale difference sequence with

$$E\bar{w}(t)\bar{w}^\tau(t) = HP(t)H^\tau + \sigma_v^2 I_m \tag{3.22}$$

By (3.17) and (3.19) we see that

$$C(z)w(t) + v(t) = H(I - Az)^{-1}K(t)\bar{w}(t-1) + \bar{w}(t) \tag{3.23}$$

Note that (3.19) is asymptotically stable, and hence (Anderson and Moore 1979, Goodwin and Sin 1984)

$$P(t) \to P, \quad K(t) \to K, \quad \text{exponentially fast} \tag{3.24}$$

where P and K are defined by

$$P = APA^\tau - APH^\tau(HPH^\tau + \sigma_v^2)^{-1}HPA^\tau + CR_wC^\tau \tag{3.25}$$

$$K = APH^\tau(HPH^\tau + \sigma_v^2)^{-1} \tag{3.26}$$

Since $E\bar{w}(t)\bar{w}^\tau(t) \to HPH^\tau + \sigma_v^2 I_m$, by the Borel–Cantelli Lemma it is easy to see that

$$\eta(t) \triangleq H(I - Az)^{-1}[K(t) - K]\bar{w}(t-1) \xrightarrow[t\to\infty]{} 0, \quad \text{a.s.} \quad \text{exponentially fast} \tag{3.27}$$

Thus, by (3.25) and (3.27) we have

$$C(z)w(t) + v(t) = [H(I - Az)^{-1}Kz + I]\bar{w}(t) + \eta(t) \tag{3.28}$$

We write K as

$$K = [K_1^\tau \quad \dots \quad K_{r+1}^\tau]^\tau$$

from (3.18) and (3.26) it can be seen that $K_{r+1} = 0$.

Set

$$D(z) = I + K_1 z + \dots + K_r z^r \tag{3.29}$$

Then by (3.18) it can be verified that

$$D(z) = [H(I - Az)^{-1}Kz + I]$$

Therefore, by (3.28), we see that (3.14) is proved. We now proceed to prove (3.16).

By (3.14) it is easy to see that

$$\sum_{i=1}^{r} C_i R_w C_i^\tau + R_w + \sigma_v^2 I = \sum_{i=1}^{r} D_i R_{\bar{w}} D_i^\tau + R_{\bar{w}} \tag{3.30}$$

By (3.25) and (3.26) it is not difficult to see that

$$P = [A - KH]P[A - KH]^\tau + K\sigma_v^2 K^\tau + CR_w C^\tau$$

From this and (3.22), we immediately obtain

$$R_{\tilde{w}} = HPH^\tau + \sigma_v^2 I_m \geqslant HCR_w C^\tau H^\tau + \sigma_v^2 I_m = R_w + \sigma_v^2 I_m \tag{3.31}$$

Consequently, by (3.30) and (3.31) we have

$$\sum_{i=1}^r C_i R_w C_i^\tau = \sum_{i=1}^r D_i R_{\tilde{w}} D_i^\tau + R_{\tilde{w}} - R_w - \sigma_v^2 I \geqslant \sum_{i=1}^r D_i R_{\tilde{w}} D_i^\tau$$
$$\geqslant [\lambda_{\min}(R_w) + \sigma_v^2] \sum_{i=1}^r D_i D_i^\tau$$

From here it follows that

$$\begin{aligned}
\|[D_1 \quad \dots \quad D_r]\|^2 &= \lambda_{\max}\left([D_1 \quad \dots \quad D_r]\begin{bmatrix} D_1^\tau \\ D_r^\tau \end{bmatrix}\right) \\
&\leqslant \frac{\lambda_{\max}(R_w)}{\lambda_{\min}(R_w) + \sigma_v^2}\lambda_{\max}\left([C_1 \quad \dots \quad C_r]\begin{bmatrix} C_1^\tau \\ C_r^\tau \end{bmatrix}\right) \\
&= \frac{\|R_w\|}{\lambda_{\min}(R_w) + \sigma_v^2}\|[C_1 \quad \dots \quad C_r]\|^2
\end{aligned} \tag{3.32}$$

and therefore, by (3.32) and (3.15) we see that

$$\|[D_1 \quad \dots \quad D_r]\|^2 \leqslant \frac{1}{r}\left(\frac{1-\alpha_0}{1+\alpha_0}\right)^2 \tag{3.33}$$

It is easy to see that

$$\begin{aligned}
&\|D_1 \exp(i\theta) + D_2 \exp(2i\theta) + \dots + D_r \exp(ir\theta)\|^2 \\
&= \lambda_{\max}\left([D_1 \quad \dots \quad D_r]\begin{bmatrix} \exp(i\theta) I_m \\ \vdots \\ \exp(ir\theta) I_m \end{bmatrix}[\exp(-i\theta) I_m \quad \dots \quad \exp(-ir\theta) I_m]\begin{bmatrix} D_1^\tau \\ \vdots \\ D_r^\tau \end{bmatrix}\right) \\
&\leqslant \lambda_{\max}\left([D_1 \quad \dots \quad D_r]\begin{bmatrix} D_1^\tau \\ \vdots \\ D_r^\tau \end{bmatrix}\right) \cdot \lambda_{\max}\left(\begin{bmatrix} \exp(i\theta) I_m \\ \vdots \\ \exp(ir\theta) I_m \end{bmatrix}[\exp(-i\theta) I_m \quad \dots \quad \exp(-ir\theta) I_m]\right) \\
&= \|[D_1 \quad \dots \quad D_r\|^2 \cdot \lambda_{\max}\left([\exp(-r\theta) I_m \quad \dots \quad \exp(-ir\theta) I_m]\begin{bmatrix} \exp(i\theta) I_m \\ \vdots \\ \exp(ir\theta) I_m \end{bmatrix}\right) \\
&= \|[D_1 \quad \dots \quad D_r]\|^2 \cdot \lambda_{\max}(r \cdot I_m) \\
&= \|[D_1 \quad \dots \quad D_r]\|^2 \cdot r
\end{aligned}$$

From here and (3.33) we immediately have

$$\|D(\exp(i\theta)) - I\| \leqslant \frac{1-\alpha_0}{1+\alpha_0} \tag{3.34}$$

By (3.34) it follows that for any $\theta \in [0, 2\pi]$

$$\begin{aligned}
&\|\alpha_0[D^\tau(\exp(-i\theta)) - I] + \alpha_0[D(\exp(i\theta) - I)] \\
&\quad + (1+\alpha_0)[D\exp(i\theta) - I][D^\tau \exp(-i\theta) - I]\| \\
&\qquad \leqslant 2\alpha_0 \|D(\exp(i\theta)) - I\| + (1+\alpha_0)\|D(\exp(i\theta)) - I\|^2 \\
&\qquad \leqslant 2\alpha_0 \frac{1-\alpha_0}{1+\alpha_0} + \frac{(1-\alpha_0)^2}{1+\alpha_0} = 1 - \alpha_0
\end{aligned}$$

Consequently, for any $\theta \in [0, 2\pi]$

$$\begin{aligned}
&D^\tau(\exp(-i\theta)) + D(\exp(i\theta)) - (1+\alpha_0)D(\exp(i\theta))D^\tau(\exp(-i\theta)) \\
&\quad = (1-\alpha_0)I - \{\alpha_0[D^\tau(\exp(-i\theta)) - I] + \alpha_0[D(\exp(i\theta)) - I] \\
&\qquad + (1+\alpha_0)[D(\exp(i\theta)) - I][D^\tau(\exp(-i\theta)) - I]\} \geqslant 0
\end{aligned}$$

Finally, we obtain

$$\begin{aligned}
&D^{-1}(\exp(i\theta)) + D^{-\tau}\exp((-i\theta)) - (1+\alpha_0)I \\
&\quad = D^{-1}(\exp(i\theta))[D^\tau(\exp(-i\theta)) + D(\exp(i\theta)) - (1+\alpha_0)D(\exp(i\theta))D^\tau(\exp(-i\theta))] \\
&\qquad \times D^{-\tau}(\exp(-i\theta)) \\
&\quad \geqslant 0, \quad \forall\, \theta \in [0, 2\pi]
\end{aligned}$$

This proves the positive-realness of $D^{-1}(z) - [(1+\alpha_0)/2]I$, and the proof of Lemma 3.3 is complete.

Remark 8

The lower bound of σ_v^2 can be improved using the following result (Anderson and Moore 1979). With initial condition $P(0) = 0$ in (3.21), $P(t)$ increases monotonically (exponentially fast) to P. Thus, the inequality $HPH^\tau \geqslant HP(1)H^\tau = HCR_wC^\tau H^\tau = R_w$, which is the essence of (3.31), can be strengthened as

$$\begin{aligned}
HPH^\tau \geqslant HP(2)H^\tau &= C_1 R_w C_1^\tau - C_1 R_w (R_w + \sigma_v^2 I)^{-1} R_w C_1^\tau + R_w \\
&= \sigma_v^2 C_1 R_w^{1/2} (R_w + \sigma_v^2 I)^{-1} R_w^{1/2} C_1^\tau + R_w
\end{aligned}$$

and likewise as $HPH^\tau \geqslant HP(t)H^\tau$ for higher t. Since convergence of $P(t)$ is exponentially fast to P, with a time constant linked to that of the Kalman filter, there are diminishing returns from taking t larger than (say) the dominant time constant of the Kalman filter. We do not explore this aspect of the results further here.

Remark 9

With appropriate initial conditions in the signal model and Kalman filter, $\eta(t)$ can be taken as zero. The term is left in our analysis to indicate a certain robustness in the noise modelling. The term $\eta(t)$ in (3.14) needs only to be square-summable for the proofs of Theorems 2.1 and 2.2 to apply.

Remark 10

G_t^0 defined in § 2.4 can be expressed by

$$G_t^0 = \sigma\{\bar{w}(i), i \leqslant t\}$$

because $\{\bar{w}(i)\}$ is the innovation sequence. Further, since $\{\bar{w}(t), G_t^0\}$ is a gaussian martingale difference sequence, it then follows that $\{\bar{w}(t)\}$ is an independent sequence. Since $\bar{w}(t+1) \in G_{t+1}^0$ and G_{t+1}^0 is independent of G_t^1, it is clear that

$$\begin{aligned} E[\bar{w}(t+1) \mid F_t] &= E[\bar{w}(t+1) \mid \sigma\{G_t^0 \cup G_t^1\}] = E[\bar{w}(t+1) \mid G_t^0] \\ &= 0, \quad \forall t \geqslant 0 \end{aligned}$$

This means that $\{\bar{w}(t), F_t\}$ is also a martingale difference sequence. All of these facts will be used in the sequel without explanations.

4. Proofs of theorems

Proof of Theorem 2.1, Part (i)

Here, we prove the first conclusion of Theorem 2.1; the proof for Part (ii) is given in the Appendix. To prove (2.12), we need to verify the conditions of Lemma 3.2.

Note that in the present case

$$z(t) = y(t) + v(t), \quad \hat{z}[t+1, \hat{\theta}(t)] = \hat{\theta}^\tau(t)\psi(t) \tag{4.1}$$

so, by Lemma 3.3 we can rewrite (2.1) in the following form

$$\begin{aligned} z(t+1) &= y(t+1) + v(t+1) \\ &= \theta_0 x(t+1) + C(z)w(t+1) + v(t+1) \\ &= \theta_0 x(t+1) + D(z)\bar{w}(t+1) + \eta(t+1) \end{aligned} \tag{4.2}$$

and so by (4.2) it follows that

$$\begin{aligned} \sup_t E\{\|z(t) - E[z(t) \mid F_{t-1}]\|^3 \mid F_{t-1}\} &= \sup_t E[\|\bar{w}(t)\|^3 \mid F_{t-1}] \\ &= \sup_t E[\|\bar{w}(t)\|^3 \mid G_{t-1}^0] \\ &= \sup_t E[\|\bar{w}(t)\|^3 < \infty \end{aligned} \tag{4.3}$$

since $\{\bar{w}(t)\}$ is gaussian random sequence with uniformly bounded covariance (see (3.22)).

By (4.2) we have the following expansion for $E[z(t+1) \mid F_t]$ at point $\bar{\theta}$

$$\begin{aligned} E[z(t+1) \mid F_t] &= \theta_0 x(t+1) + [D(z) - I]\bar{w}(t+1) + \eta(t+1) \\ &= \bar{\theta}^\tau\psi(t) + [D(z) - I][\bar{w}(t+1) - z(t+1) + \hat{\theta}^\tau(t+1)\psi(t)] + \eta(t+1) \\ &= \bar{\theta}^\tau\psi(t) + [D(z) - I]D^{-1}(z) \\ &\quad \times \{D(z)[\bar{w}(t+1) - z(t+1) + \hat{\theta}^\psi(t+1)\psi(t)]\} + \eta(t+1) \\ &= \bar{\theta}^\tau\psi(t) + [I - D^{-1}(z)] \\ &\quad \times \{z(t+1) - \theta_0 x(t+1) - \eta(t+1) - D(z)[z(t+1) - \hat{\theta}^\tau(t+1)\psi(t)]\} \\ &\quad + \eta(t+1) \end{aligned}$$

$$= \bar{\theta}^{\tau}\psi(t) + [I - D^{-1}(z)]$$
$$\times \{-\bar{\theta}^{\tau}\psi(t) + z(t+1) - \eta(t+1) - z(t+1) + \hat{\theta}^{\tau}(t+1)\psi(t)\}$$
$$+ \eta(t+1)$$
$$= \hat{\theta}^{\tau}(t)\psi(t) + \tilde{\theta}^{\tau}(t)\psi(t) + [D^{-1}(z) - I][\tilde{\theta}^{T}(t+1)\psi(t)]$$
$$+ D^{-1}(z)\eta(t+1) \tag{4.4}$$

where $\tilde{\theta}(t) = \bar{\theta} - \hat{\theta}(t)$. Now, since

$$\sigma_v^2 > r\|R_w\| \cdot \|[C_1 \quad \ldots \quad C_r]\|^2 - \lambda_{\min}(R_w)$$

there exists $\alpha_0 > 0$ such that

$$\sigma_v^2 > r\left(\frac{1+\alpha_0}{1-\alpha_0}\right)^2 \|R_w\| \cdot \|[C_1 \quad \ldots \quad C_r]\|^2 - \lambda_{\min}(R_w)$$

and hence by Lemma 3.3 we know that

$$D^{-1}(z) - \frac{1+\alpha_0}{2} I \tag{4.5}$$

is positive-real, so by (4.3)–(4.5) we know that Lemma 3.2 is applicable, and we then have for $0 < \alpha < \alpha_0$

$$\|\hat{\theta}(t+1) - \bar{\theta}\|^2 \leqslant 0\left(\frac{\log \lambda_{\max}(t)}{\lambda_{\min}(t)}\right) + \frac{2}{\alpha_0 - \alpha} \sum_{i=1}^{t} \frac{\|D^{-1}(z)\eta(i+1)\|^2}{\lambda_{\min}(t)}$$
$$= 0\left(\frac{\log \lambda_{\max}(t)}{\lambda_{\min}(t)}\right)$$

since $\eta(t) \to 0$ is exponentially fast. This proves the first conclusion of Theorem 2.1.

Proof of Theorem 2.2

Set

$$\xi(t) = z(t) - \hat{\theta}^{\tau}(t)\psi(t-1) - \bar{w}(t) \tag{4.6}$$
$$\psi^{\xi}(t) = [0 \quad \xi^{\tau}(t) \quad \ldots \quad \xi^{\tau}(t-r+1)]^{\tau} \tag{4.7}$$

By a similar argument as used in the proof of (4.4) we know that

$$D(z)\xi(t) = \tilde{\theta}^{\tau}(t)\psi(t-1) + \eta(t) \tag{4.8}$$

Since $D(z)$ is strictly positive-real it must be stable, by (4.4), (4.7) and (4.8) and Lemma 3.2, we have

$$\sum_{i=1}^{t} \|\psi^{\xi}(i+1)\|^2 = 0\left(\sum_{i=1}^{t} \|\tilde{\theta}^{\tau}(i+1)\psi(i)\|^2\right) + 0(1)$$
$$= 0[\log \lambda_{\max}(t)] \tag{4.9}$$

Note that

$$\psi(t) = \psi^{0}(t) + \psi^{\xi}(t) \tag{4.10}$$

and hence by use of (4.9) and (4.10) similar to the proof of Theorem 2 of Chen and

Guo (1986: p. 1465) we know that

$$\lambda_{\max}(t) = O[\lambda^0_{\max}(t)], \quad \lambda^0_{\min}(t) = O[\lambda_{\min}(t)], \quad \text{a.s.}$$

This result together with Theorem 2.1 yields the desired results immediately and Theorem 2.2 is now established.

5. Conclusions

This paper has shown that modifying standard ELS algorithms for linear regression model identification can obviate the need for a positive-real condition on the coloured noise model. Estimates of the regression part parameter matrix θ_0 in (2.1) and those of the modified noise model D_i are achieved without any compromise on convergence rates. The recovery of the original noise model parameters C_i by an on-line spectral factorization is studied in a companion paper (Guo and Moore 1987). A method to remove an estimator error variance increase by additional processing is currently under study. The methods and theory of the paper fall short of giving precise results for avoiding the positive-real condition for general ARMAX models.

Appendix

Proof of bound (2.24)

With

$$C(z) = 1 + \sum_{i=1}^{r} C_i z^i \triangleq \prod_{i=1}^{r} (1 + \alpha_i z)$$

in the scalar case, the minimum phase condition is that $|\alpha_i| < 1$ for all i. Denoting the binomial coefficients $\binom{r}{i}$ as

$$\binom{r}{i} = \frac{r!}{i!(r-i)!}, \quad 0 \leqslant i \leqslant r$$

comparing the coefficients of z^i in the above identity and noting $|\alpha_i| < 1$, we know that

$$|C_i| < \binom{r}{i}, \quad 1 \leqslant i \leqslant r$$

Consequently, it follows that

$$1 + \sum_{i=1}^{r} C_i^2 < \sum_{i=0}^{r} \binom{r}{i}^2$$

but by comparing the coefficients of z^r in the following identity

$$(1+z)^r(1+z)^r = (1+z)^{2r}$$

it is easy to know

$$\sum_{i=0}^{r} \binom{r}{i}^2 = \binom{2r}{r}$$

and hence

$$1+\sum_{i=1}^{r} C_i^2<(2r)!(r!)^{-2}$$

which is tantamount to (2.24).

Remark A.1

The bound (2.24) is sharp for all r.

Remark A.2

A similar bound is not available for the multidimensional case, unless extra conditions in addition to the minimum phase assumption on $C(z)$ are imposed.

Proof of Lemma 3.1

Set

$$e(t) \triangleq z(t)-E[z(t) \mid F_{t-1}] \tag{A 1}$$

we see that $\{e(t), F_t\}$ is a martingale difference sequence, and satisfies

$$\sup_{t \geqslant 0} E[\|e(t)\|^\beta \mid F_{t-1}]<\infty, \quad \text{a.s.} \quad \beta \geqslant 2 \tag{A 2}$$

By (3.1 b) and (3.2) it is easy to see that

$$\begin{aligned} P(t)\psi(t) &= [P(t-1)-a(t)P(t-1)\psi(t)\psi^\tau(t)P(t-1)]\psi(t) \\ &= a(t)P(t-1)\psi(t) \end{aligned} \tag{A 3}$$

and then by (3.3) and (A 1) we know

$$P(t)\psi(t)[z^\tau(t+1)-\hat{z}^\tau(t+1, \hat{\theta}(t))]=P(t-1)\psi(t)[\zeta^\tau(t+1)+e^\tau(t+1)]$$

Therefore, we can rewrite (3.1 a) as

$$\tilde{\theta}(t+1)=\tilde{\theta}(t)-P(t-1)\psi(t)[\xi^\tau(t+1)+e^\tau(t+1)] \tag{A 4}$$

with $\tilde{\theta}(t)=\theta-\hat{\theta}(t)$, for any θ.

From (3.1 b) it is known that

$$P(t)=\left[\sum_{i=1}^{t} \psi(i)\psi^\tau(i)+P^{-1}(0)\right]^{-1}, \quad t \geqslant 0 \tag{A 5}$$

We now prove our results along the lines of the proof of Theorem 1 of Chen and Guo (1986). By (A 4) and (A 5) a similar treatment used as in the proof of (19) of Chen and Guo (1986: p. 1462) leads to

$$\begin{aligned} \operatorname{tr} \tilde{\theta}^\tau(t+1)P^{-1}(t)\tilde{\theta}(t+1) \leqslant{} & \operatorname{tr} \theta^\tau(t)P^{-1}(t-1)\tilde{\theta}(t) \\ & +\|\tilde{\theta}^\tau(t+1)\psi(t)\|^2-2\xi^\tau(t+1)\tilde{\theta}^\tau(t)\psi(t) \\ & -2e^\tau(t+1)\tilde{\theta}^\tau(t+1)\psi(t), \quad \forall\,\theta \\ ={} & \operatorname{tr} \tilde{\theta}^\tau(t)P^{-1}(t-1)\tilde{\theta}(t) \\ & -2\left[\xi^\tau(t+1)\tilde{\theta}^\tau(t+1)\psi(t)-\frac{1+\alpha}{2}\|\tilde{\theta}^\tau(t+1)\psi(t)\|^2\right] \\ & -\alpha\|\tilde{\theta}^\tau(t+1)\psi(t)\|^2-2e^\tau(t+1)\tilde{\theta}^\tau(t+1)\psi(t), \quad \forall\,\theta, \quad \forall\,\alpha>0 \end{aligned} \tag{A 6}$$

Summing both sides of (A 6) and using (3.4) we obtain

$$\operatorname{tr} \tilde{\theta}^{\tau}(t+1)P^{-1}(t)\tilde{\theta}(t+1) \leqslant \operatorname{tr} \tilde{\theta}^{\tau}(1)P^{-1}(0)\tilde{\theta}(1) - 2S_t(\theta, \alpha) - \alpha \sum_{i=1}^{t} \|\tilde{\theta}^{\tau}(i+1)\psi(i)\|^2 - 2\sum_{i=1}^{t} e^{\tau}(i+1)\tilde{\theta}^{\tau}(i+1)\psi(i), \quad \forall \theta, \quad \forall \alpha > 0 \tag{A 7}$$

We now estimate the last term on the right-hand side of (A 7). Since $\{e(t), F_t\}$ is a martingale difference sequence and satisfies (A 2), by Lemma 2 of Chen and Guo (1987 b), we know that for any F_t-measurable matrix $M(t)$

$$\sum_{i=0}^{t} M(i)e(i+1) = 0\left(\sum_{i=0}^{t} \|M(i)\|^2\right)^{1/2+\eta}, \quad \text{a.s.} \quad \forall \eta > 0 \tag{A 8}$$

Set

$$\eta(t) = E[z(t+1) \mid F_t] - \hat{z}[t+1, \hat{\theta}(t)] \tag{A 9}$$

Obviously, $\eta(t)$ is F_t-measurable, and by (A 1), (A 9) it follows from (3.1) that

$$\tilde{\theta}(t+1) = \tilde{\theta}(t) - P(t)\psi(t)[e^{\tau}(t+1) + \eta^{\tau}(t)] \tag{A 10}$$

for any θ.

Then by using (A 8) and (A 10) similar to the proof of (22) of Chen and Guo (1986: p. 1463), we have

$$\left|\sum_{i=1}^{t} e^{\tau}(i+1)\tilde{\theta}^{\tau}(i+1)\psi(i)\right| = 0\left(\sum_{i=1}^{t} \|\tilde{\theta}^{\tau}(i+1)\psi(i)\|^2\right)^{1/2+\eta} + 0\left(\sum_{i=1}^{t} \psi^{\tau}(i)P(i)\psi(i)\,\|e(i+1)\|^2\right) \tag{A 11}$$

for any $\eta > 0$.

However, by (A 3) and both (29) and (30) of Chen and Guo (1986: p. 1465) we know that

$$\sum_{i=1}^{t} \psi^{\tau}(i)P(i)\psi(i)\,\|e(i+1)\|^2 = 0 \log^{[1+\alpha\delta(\beta-2)]}\lambda_{\max}(t) \tag{A 12}$$

Finally, putting (A 11) and (A 12) into (A 7) and taking $\eta < \frac{1}{2}$, we see that for any $\alpha > 0$ and any θ

$$\operatorname{tr} \tilde{\theta}^{\tau}(t+1)P^{-1}(t)\tilde{\theta}(t+1) \leqslant 0(1) - S_t(\theta, \alpha) + 0 \log^{[1+\alpha\delta(\beta-2)]}\lambda_{\max}(t)$$

and the desired result follows from here immediately and therefore the proof of Lemma 3.1 is completed.

Proof of Theorem 2.1 (ii)

We now prove the second result (2.15) of Theorem 2.1. Multiplying $P^{-1}(t)$ on both sides of (2.9) and using (A 5) we have

$$\begin{aligned} P^{-1}(t)\hat{\theta}(t+1) &= [P^{-1}(t) - \psi(t)\psi^{\tau}(t)]\hat{\theta}(t) + \psi(t)z^{\tau}(t+1) \\ &= P^{-1}(t-1)\hat{\theta}(t) + \psi(t)z^{\tau}(t+1) \end{aligned}$$

Consequently, from this and (4.2) and (4.10) we know

$$
\begin{aligned}
P^{-1}(t)\hat{\theta}(t+1) &= P^{-1}(0)\hat{\theta}(1) + \sum_{i=1}^{t} \psi(i)z^{\tau}(i+1) \\
&= P^{-1}(0)\hat{\theta}(1) + \sum_{i=1}^{t} \psi(i)\{[\psi^{\tau}(i) - \psi^{\xi\tau}(i)]\bar{\theta} + \bar{w}^{\tau}(i+1) + \eta^{\tau}(i+1)\} \\
&= P^{-1}(0)\hat{\theta}(1) + \sum_{i=1}^{t} \psi(i)\psi^{\tau}(i)\bar{\theta} \\
&\quad - \sum_{i=1}^{t} \psi(t)[\psi^{\xi\tau}(i)\bar{\theta} - \eta^{\tau}(i+1)] + \sum_{i=1}^{t} \psi(i)\bar{w}^{\tau}(i+1)
\end{aligned}
$$

and hence

$$
\begin{aligned}
\tilde{\theta}(t+1) &= \left[\sum_{i=1}^{t} \psi(i)\psi^{\tau}(i) + P^{-1}(0)\right]^{-1} \\
&\quad \times \left\{P^{-1}(0)\tilde{\theta}(1) - \sum_{i=1}^{t} \psi(i)\bar{w}^{\tau}(i+1) + \sum_{i=1}^{t} \psi(i)[\psi^{\xi\tau}(i)\bar{\theta} - \eta^{\tau}(i-1)]\right\} \qquad \text{(A 13)}
\end{aligned}
$$

Again by (4.2) and (4.10) from (2.12) we know

$$
\begin{aligned}
t\hat{R}_{\bar{w}}(t) &= \sum_{i=0}^{t-1} [\bar{\theta}^{\tau}\psi^{0}(i) + \bar{w}(i+1) + \eta(i+1) - \hat{\theta}^{\tau}(t)\psi(i)] \\
&\quad \times [\bar{\theta}^{\tau}\psi^{0}(i) + \bar{w}(i+1) + \eta(i+1) - \hat{\theta}^{\tau}(t)\psi(i)]^{\tau} \\
&= \sum_{i=0}^{t-1} [\tilde{\theta}^{\tau}(t)\psi(i) - \bar{\theta}^{\tau}\psi^{\xi}(i) + \bar{w}(i+1) \\
&\quad + \eta(i+1)][\tilde{\theta}^{\tau}(t)\psi(i) - \bar{\theta}^{\tau}\psi^{\xi}(i) + w(i+1) + \eta(i+1)]^{\tau} \\
&= \tilde{\theta}^{\tau}(t)\sum_{i=0}^{t-1} \psi(i)\psi^{\tau}(i)\tilde{\theta}(t) + \sum_{i=0}^{t-1} \\
&\quad \times [-\bar{\theta}^{\tau}\psi^{\xi}(i) + w(i+1) + \eta(i+1)]\psi^{\tau}(i)\tilde{\theta}(t) \\
&\quad + \tilde{\theta}^{\tau}(t)\sum_{i=1}^{t-1} \psi(i)[-\bar{\theta}^{\tau}\psi^{\xi}(i) + \bar{w}(i+1) + \eta(i+1)]^{\tau} \\
&\quad + \sum_{i=0}^{t-1} [\bar{\theta}^{\tau}\psi^{\xi}(i) - \bar{w}(i+1) - \eta(i+1)] \\
&\quad \times [\bar{\theta}^{\tau}\psi^{\xi}(i) - \bar{w}(i+1) - \eta(i+1)]^{\tau} \\
&\triangleq S_1(t) + S_2(t) + S_2^{\tau}(t) + S_3(t) \qquad \text{(A 14)}
\end{aligned}
$$

However, by the Schwarz inequality and (4.9) we obtain

$$
\begin{aligned}
&\left\|\left[\sum_{j=0}^{t-1} \psi(j)\psi^{\tau}(j) + P^{-1}(0)\right]^{-1/2} \sum_{i=0}^{t-1} \psi(i)\psi^{\xi\tau}(i)\right\|^{2} \\
&\qquad \leqslant \sum_{i=0}^{t-1} \left\|\left[\sum_{j=0}^{t-1} \psi(j)\psi^{\tau}(j) + P^{-1}(0)\right]^{-1/2} \psi(i)\right\|^{2} \sum_{i=0}^{t-1} \|\psi^{\xi}(i)\|^{2} \\
&\qquad = \operatorname{tr}\left\{\sum_{i=0}^{t-1} \psi^{\tau}(i)\left[\sum_{j=0}^{t-1} \psi(j)\psi^{\tau}(j) + P^{-1}(0)\right]^{-1} \psi(i)\right\} \sum_{i=0}^{t-1} \|\psi^{\xi}(i)\|^{2}
\end{aligned}
$$

$$= 0(1)\sum_{i=0}^{t-1} \|\psi^{\xi}(i)\|^2$$

$$= 0[\log \lambda_{\max}(t-1)]$$

and hence

$$\left\|\left[\sum_{i=0}^{t-1} \psi(i)\psi^{\tau}(i) + P^{-1}(0)\right]^{-1/2} \sum_{i=0}^{t-1} \psi(i)\psi^{\xi\tau}(i)\right\| = 0([\log \lambda_{\max}(t)]^{1/2}) \tag{A 15}$$

We need the following estimates for weighted-sum of martingale difference sequences (Lai and Wei 1986 b). Let $\{X_t, F_t\}$ be an adapted vector sequence and $\{e_t, F_t\}$ be martingale difference sequence with

$$\sup_t \mathrm{E}[\|e_t\|^{2+\delta} | F_{t-1}] < \infty, \quad \delta > 0$$

Then, for $\varepsilon > 0$

$$\left\|\left(\sum_{i=0}^{t} X_i X_i^{\tau} + \varepsilon I\right)^{-1/2} \sum_{i=0}^{t} X_i e_{i+1}^{\tau}\right\| = 0\left(\log \lambda_{\max}\left(\sum_{i=1}^{t} X_i X_i^{\tau} + \varepsilon I\right)\right)^{1/2} \tag{A 16}$$

Applying (A 15) and (A 16) to $S_1(t)$ and $S_2(t)$ defined in (A 14) and noting (A 13), it is not difficult to show that

$$S_1(t) = 0[\log \lambda_{\max}(t)], \quad \text{a.s.} \tag{A 17}$$

$$S_2(t) = 0[\log \lambda_{\max}(t)], \quad \text{a.s.} \tag{A 18}$$

By (4.9) and the inequality (A 8) we see that the last term in (A 14) can be estimated by

$$S_3(t) = \sum_{i=0}^{t-1} \bar{w}(i+1)\bar{w}^{\tau}(i+1) + 0[\log \lambda_{\max}(t)] \tag{A 19}$$

By Lemma 3.3, $\{\bar{w}(t), F_t\}$ is a gaussian martingale difference sequence and $E\bar{w}(t)\bar{w}^{\tau}(t) \to R_{\bar{w}}$ (exponentially fast), and hence by the laws of the iterated logarithm (Stout 1974) it is not difficult to convince oneself that

$$\frac{1}{t}\sum_{i=0}^{t-1} [\bar{w}(i+1)\bar{w}^{\tau}(i+1) - R_{\bar{w}}] = 0\left(\frac{\log \log t}{t}\right)^{1/2}, \quad \text{a.s.} \tag{A 20}$$

Finally, putting (A 17)–(A 20) into (A 14), the result (2.13), follows.

References

Anderson, B. D. O., and Moore, J. B., 1979, *Optimal Filtering* (Englewood Cliffs, NJ: Prentice Hall).

Chen, H. F., and Guo, L., 1986, *Int. J. Control*, **44**, 1459; 1987 a, *Syst. Control Lett.*, **9**, 131; 1987 b, *SIAM Jl Control Optim.*, **25**, 845.

Goodwin, G. C., and Sin, K. S., 1984, *Adaptive Filtering, Prediction and Control* (Englewood Cliffs, NJ: Prentice Hall).

Guo, L., and Moore, 1987, Adaptive spectral factorization with applications to system identification. Technical report, Department of Systems Engineering, Australian National University.

Lai, T. L., and Wei, C. Z., 1986 a, *I.E.E.E. Trans. autom. Control*, **31**, 898; 1986 b, *Stochastics*, **16**, 227.

Ledwich, G., and Moore, J. B., 1977, Multivariable self-tuning filters, Differential Games and Control Theory, II. *Lecture Notes in Pure and Applied Mathematics* (New York: Marcel-Dekker), pp. 345–376.

LJUNG, L., 1977, *I.E.E.E. Trans. autom. Control*, **22**, 539.
MOORE, J. B., 1982, *Ric. Autom.*, **8**, 501.
STOUT, W. F., 1974, *Almost Sure Convergence* (New York: Academic Press).