STOCHASTIC SYSTEM IDENTIFICATION VIA ADAPTIVE SPECTRAL FACTORIZATION

Guo Lei (郭 雷)

(Institute of Systems Science, Academia Sinica, Beijing 100080, China)

John B. Moore

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

Abstract. We consider in this paper the problem of recursive identification for stochastic systems when the noise model does not satisfy the positive real condition associated with convergence of standard algorithms. To avoid the positive real condition, adaptive spectral factorization techniques are exploited on the basis of a class of non-standard time-varying recursive Riccati equations. The asymptotic properties of the Riccati equations are studied as a crucial step to the convergence results of the paper.

Key words. Adaptive, parameter estimation, spectral factorization.

1. Introduction

Consider a random process $\{z(t)\}$, which is the output process of a linear, asymptotically stable, finite-dimensional system driven by zero mean, stationary white noise $\{v(t)\}$ commencing in the infinitely remote past. Then, $\{z(t)\}$ is a stationary process with power spectrum matrix

$$\Phi(z)=\sum_{t=-\infty}^{\infty}E[z(t)z^{\tau}(0)]z^{-t}.$$

It is known that, for some matrices F, H, M and L, $\Phi(z)$ can be expressed by (e.g. [1]):

$$\Phi(z)=L+H^{\tau}(zI-F^{\tau})^{-1}M+M^{\tau}(z^{-1}I-F)^{-1}H, \tag{1.1}$$

based on which one then can find $W(z)$ and Ω such that

$$\Phi(z)=W(z)\Omega W^{\tau}(z^{-1}), \tag{1.2}$$

where $W(z)$ is the transfer function of the system generating $\{z(t)\}$ and Ω is the covariance of the driven noise $\{v(t)\}$. This task is known as spectral factorization.

Usually, the matrices F, H, M, L are not available; one then uses estimates $F(t)$, $H(t)$, $M(t)$ and $L(t)$ constructed by use of the observation data up to time t to replace them, and solve the corresponding spectral factorization problem to get an estimate $W_t(z)$ for $W(z)$. When $W_t(z)$ is computed on-line and recursively, the task is termed adaptive spectral factorization.

In this paper, we propose an on-line recursive algorithm for computing $W_t(z)$. We give sufficient conditions for convergence of $W_t(z)$ to $W(z)$ and rates of

Received August 29,1990.

convergence in terms of those of $F(t)$, $H(t)$, $\cdots$ to F, H, $\cdots$. Then, we apply the obtained results to the linear regression model and moving average process identification.

An initial motivation for the present work was the challenge of avoiding a strict positive real condition for convergence of recursive ARMAX model identification. Preliminary and incomplete results for this are in an earlier work [2]. This paper and a companion paper [3] are an attempt to strengthen the insights of [2] and place them in a rigorous mathematical framework. During the course of this study, there has emerged the connection that adaptive spectral factorization algorithms should be considered as a fundamental building block for recursive on-line identification just as are recursive least squares algorithms.

2. Adaptive Spectral Factorization Algorithm and Its Convergence

To be precise, let us assume that $\{F(t),\ H(t),\ M(t),\ L(t)\}$ are the same dimensional estimates of $\{F,\ H,\ M,\ L\}$ with $L(t)$ symmetric for all $t\geq 0$. Define the estimates $\Phi_t(z)$ for the spectrum matrix $\Phi(z)$ as follows:

$$\Phi_t(z)=L(t)+H^\tau(t)[zI-F^\tau(t)]^{-1}M(t)+M^\tau(t)[z^{-1}I-F(t)]^{-1}H(t). \quad (2.1)$$

Let us introduce the following recursive Riccati equation for computing matrices $\{\Sigma_t(s),\ 0\leq s\leq[\log^2 t],\ t\geq 1\}$:

$$\Sigma_t(s+1)=F^\tau(t)\Sigma_t(s)F(t)-[F^\tau(t)\Sigma_t(s)H(t)+M(t)][H^\tau(t)\Sigma_t(s)H(t)+L(t)]^\dagger$$
$$\cdot[F^\tau(t)\Sigma_t(s)H(t)+M(t)]^\tau,\ 0\leq s\leq[\log^2 t],\ t\geq 1\},\ \Sigma_t(0)=0, \quad (2.2)$$

where $X^\dagger$ denotes the Moore-Penrose pseudo-inverse of a matrix X, and where $[\log^2 t]$ denotes the integer part of a real number $\log^2 t$.

Rather than set an estimate $\hat{\Sigma}(t)=\Sigma_t(s)$ with $\Sigma_t(0)=\Sigma_{t-1}(s)$ for some positive integer s, as is conventional, here we set

$$\hat{\Sigma}(t)\triangleq\Sigma_t([\log^2 t]),\quad t\geq 1. \quad (2.3)$$

The estimates $\hat{\Omega}(t)$ and $\hat{W}_t(z)$ for Ω and $W(z)$ are then defined by

$$\hat{\Omega}(t)=H^\tau(t)\hat{\Sigma}(t)H(t)+L(t), \quad (2.4)$$

$$\hat{W}_t(z)=I+H^\tau(t)[zI-F^\tau(t)]^{-1}\hat{K}(t), \quad (2.5)$$

where

$$\hat{K}(t)=[F^\tau(t)\hat{\Sigma}(t)H(t)+M(t)]\hat{\Omega}^\dagger(t). \quad (2.6)$$

Let us introduce the following conditions for the spectrum matrix as defined in (1.1):

(A1) $\Phi(z)$ is positive definite hermitian on the unit circle, i.e., $\Phi(z)>0$, $|z|=1$.

(A2) The matrix quadruple $\{F,\ H,\ M,\ L/2\}$ is minimal. i. e., $[F,\ H]$ and $[F,\ M]$ are respectively controllable and observable.

(A3) All eigenvalues of F lie in the unit circle, i.e., $|\lambda_i(F)|<1,\ 1\leq i\leq m$ (dimension of F).

(A4) The matrix $(\det F)L-M^\tau[\operatorname{Adj} F]H$ is nonsingular.

We now make some comments on the above conditions. Conditions (A1) and (A2) imply that the transfer function

$$Z_d(z) = H^\tau (zI - F^\tau)^{-1} M$$

is strictly discrete positive real. Condition (A3) is the stationary condition for the process $\{z(t)\}$ mentioned in Section 1, and condition (A4) is the nonsingularity requirement of $\Phi(\infty)$ when $\det F \neq 0$ since $(\det F) I_m = F[\mathrm{Adj}\, F]$.

Recall that a matrix function $W(z)$ is termed (asymptotically) stable and minimum phase if the following two conditions are satisfied:

(i) $W(z)$ has constant rank in $|z|>1$,

(ii) $W(z)$ is analytic in $|z|\geqslant 1$,

and further, if in (i) $|z|>1$ is replaced by $|z|\geqslant 1$, then the term (asymptotically) stable strictly minimum phase applies.

Before proceeding with new results, let us first review some theory on matrix spectral factorization summarized in the following lemma.

Lemma 2.1. *Consider the power spectrum matrix $\Phi(z)$ defined in* (1.1). *Assume that conditions* (A1)—(A3) *are satisfied. Then*

(i) *There exists a factorization of $\Phi(z)$ as*

$$\Phi(z) = W(z)\Omega W^\tau(z^{-1}), \tag{2.7}$$

where $W(z)$ is a square, real, rational, asymptotically stable strictly minimum phase transfer function matrix $W(\infty) = I$, and $\Omega > 0$.

Moreover, $W(z)$ with $W(\infty) = I$ and Ω are unique within the class of asymptotically stable, minimum phase transfer function matrices and the class of positive definite matrices respectively.

(ii) *The factorization $W(z)$ and Ω can be expressed by*

$$W(z) = I + H^\tau (zI - F^\tau)^{-1} K, \tag{2.8}$$

$$\Omega = H^\tau \Sigma H + L, \tag{2.9}$$

where

$$K \triangleq (F^\tau \Sigma H + M)\Omega^{-1} \tag{2.10}$$

and Σ is the limiting solution of the following equation (i.e., $\Sigma = \lim_{k\to\infty} \Sigma_k$):

$$\Sigma_{k+1} = F^\tau \Sigma_k F - [F^\tau \Sigma_k H + M][H^\tau \Sigma_k H + L]^+ [F^\tau \Sigma_k H + M]^\tau$$

with $\Sigma_0 = 0$, and satisfies the following algebraic Riccati equation:

$$\Sigma = F^\tau \Sigma F - [F^\tau \Sigma H + M][H^\tau \Sigma H + L]^{-1} [F^\tau \Sigma H + M]^\tau. \tag{2.11}$$

(iii) $|\lambda_i(F - HK^\tau)| < 1$, *for all i, $1 \leqslant i \leqslant m$.*

Proof. The first conclusion of the lemma follows directly from Theorem 4.1 and its remarks in [1] (p. 240—241). For the second conclusion, we note that $H^\tau(zI - F^\tau)^{-1}M + (L/2)$ is a strictly positive real transfer functson with minimal realization $\{F, H, M, L/2\}$, and hence a stable minimum phase spectral factor $W_d(z)$ of $\Phi(z)$ (i.e. $\Phi(z) = W_d(z) W_d^\tau(z^{-1})$) can be constructed by ([4], p. 746):

$$W_d(z) = [I + H^\tau (zI - F^\tau)^{-1} (F^\tau \Sigma H + M)(N^\tau N)^+] N^\tau,$$

where N is such that $N^\tau N = H^\tau \Sigma H + L$. Comparing this with (2.7) and using the uniqueness of factorization, we obtain conclusion (ii). Finally, conclusion (iii) is provided in [4] (p. 747).

Remark. We note that the Riccati equation defined by (2.11) corresponds to a special variational problem, the solution of which may not be unique [4]. It is different from the usual one considered in LQ control problems and Kalman filtering problems. Hence the analysis of the resulting adaptive Riccati equation (2.2) is completely different from those in e.g. [7].

We now proceed to state the convergence results of the adaptive spectral factorization algorithm defined by (2.1)—(2.6).

As usual, for a complex-valued matrix function $G(z)$ which is defined and bounded on the unit circle $|z|=1$, we denote its norm by

$$\|G(z)\|_\infty=\sup_{|z|=1}[\lambda_{\max}(G(z)G^*(z))]^{\frac{1}{2}},$$

where "$*$" denotes the transpose complex conjugate. When G is a constant matrix, we shall write

$$\|G\|=[\lambda_{\max}(GG^*)]^{\frac{1}{2}}.$$

Let us denote the estimation error for the realization $\{F, H, M, L\}$ by $\Delta(t)$, i.e.,

$$\Delta(t)=\|[F(t), H(t), M(t)]-[F, H, M]\|+\|L(t)-L\|. \tag{2.12}$$

It is straightforward to show [see next section (3.6)] that if $\Delta(t)\to 0$, then for $\Phi(z)$ and $\Phi_t(z)$ defined by (1.1) and (2.1), there holds

$$\|\Phi_t(z)-\Phi(z)\|_\infty=O(\Delta(t)), \quad \text{as } t\to\infty.$$

We now proceed with the main purpose of this paper, which is to establish convergence rates of the adaptive spectral factors $\hat{W}_t(z)$ to $W(z)$.

Theorem 2.1. *Consider the adaptive spectral factorization algorithm defined by (2.1)—(2.6). Assume that the power spectrum matrix $\Phi(z)$ defined by (1.1) satisfies conditions (A1)—(A4). If $\Delta(t)\to 0$, then the following convergence rates hold:*

$$\|\hat{W}_t(z)-W(z)\|_\infty=O(\Delta(t))+O(t^{-\alpha\log t}) \tag{2.13}$$

and

$$\|\hat{\Omega}(t)-\Omega\|=O(\Delta(t))+O(t^{-\alpha\log t}) \tag{2.14}$$

for some positive constant $\alpha>0$, where $\{\Omega, W(z)\}$ are given from the unique spectral factorization specified in Lemma 2.1, and where $\Delta(t)$ is defined by (2.12).

Remarks. (i) We remark that $\hat{W}_t(z)$ is generally not a spectral factor of $\Phi_t(z)$ defined by (2.1), hence the existing results on continuity of spectral factorizations (e. g. [8]) can not be directly used in the present case.

(ii) Actually, there are many ways to define $\hat{\Sigma}(t)$, for example, if $\hat{\Sigma}(t)$ is defined as

$$\hat{\Sigma}(t)=\Sigma_t([d(t)]) \quad t\geq 0,$$

where $d(t)>0$ is any sequence such that $d(t+1)\geq d(t)$ and $d(t)\xrightarrow[t\to\infty]{}\infty$, then the resulting convergence rate is now $O(\Delta(t))+O(e^{-\alpha d(t)})$ for some $\alpha>0$.

3. Convergence Analysis of the Algorithm

For the proof of Theorem 2.1, we need to establish the following results.

Lemma 3.1. *Consider the spectrum estimates $\Phi_t(z)$ defined as in (2.1). Consider also that the conditions of Theorem 2.1 apply. Then for appropriately large t (say $t \geq t_0$), $\Phi_t(z)$ has the following unique factorization:*

$$\Phi_t(z) = W_t(z)\Omega(t)W_t^\tau(z^{-1}), \tag{3.1}$$

where $W_t(z)$ is the (asymptotically) stable, strictly minimum phase spectral factor with $W_t(\infty) = I$ and $\Omega(t) > 0$. Further, $W_t(z)$ and $\Omega(t)$ can be expressed by

$$W_t(z) = I + H^\tau(t)[zI - F^\tau(t)]^{-1}K(t), \tag{3.2}$$

$$\Omega(t) = H^\tau(t)\Sigma(t)H(t) + L(t), \tag{3.3}$$

where

$$K(t) = [F^\tau(t)\Sigma(t)H(t) + M(t)]\Omega^{-1}(t)$$

and $\Sigma(t)$ satisfies the following algebraic equation:

$$\Sigma(t) = F^\tau(t)\Sigma(t)F(t) - [F^\tau(t)\Sigma(t)H(t) + M(t)] \cdot [H^\tau(t)\Sigma(t)H(t) + L(t)]^{-1}[F^\tau(t)\Sigma(t)H(t) + M(t)]^\tau. \tag{3.4}$$

Also, $|\lambda_i[F(t) - H(t)K^\tau(t)]| < 1$ for all i and all $t \geq t_0$, and $\Sigma(t)$ has the following convergence rates:

$$\|\Sigma(t) - \Sigma\| = O(\Delta(t)), \quad t \to \infty, \tag{3.5}$$

where Σ is given in Lemma 2.1, and $\Delta(t)$ is defined by (2.12).

Proof. We first establish the following estimation property mentioned in Section 2

$$\|\Phi_t(z) - \Phi(z)\|_\infty = O(\Delta(t)), \quad t \to \infty. \tag{3.6}$$

For this, we need only to show that

$$\|M^\tau(t)[z^{-1}I - F(t)]^{-1}H(t) - M^\tau[z^{-1}I - F]^{-1}H\|_\infty = O(\Delta(t)). \tag{3.7}$$

Note that

$$\begin{aligned}&\|[z^{-1}I - F(t)]^{-1} - (z^{-1}I - F)^{-1}\|_\infty \\ &= \|[z^{-1}I - F(t)]^{-1}[F(t) - F][z^{-1}I - F]^{-1}\|_\infty \\ &\leq \|[z^{-1}I - F(t)]^{-1}\|_\infty \cdot \|F(t) - F\| \cdot \|[z^{-1}I - F]^{-1}\|_\infty.\end{aligned} \tag{3.8}$$

Since $F(t) \to F$, for appropriately large t we know that $|\lambda_i[F(t)]| < 1$ by condition (A3). Consequently, we have the following expansion on $|z| = 1$:

$$[z^{-1}I - F(t)]^{-1} = z\sum_{i=0}^{\infty}[zF(t)]^i. \tag{3.9}$$

We need the following fact: if matrices G_k converge to a matrix G with $|\lambda_i(G)| < 1$, then there are constants $\lambda \in (0, 1)$, $k_0 > 0$, and $c > 0$ such that ([5], p. 191):

$$\|[G_k]^i\| \leq c\lambda^i, \quad \forall k \geq k_0, \quad \forall i \geq 0. \tag{3.10}$$

By (3.10) from (3.9) it is evident that for large t

$$\|[z^{-1}I - F(t)]^{-1}\|_\infty \text{ is uniformly bounded in } t. \tag{3.11}$$

So, by (3.8) it is clear that as $t\to\infty$,

$$\|[z^{-1}I-F(t)]^{-1}-(z^{-1}I-F)^{-1}\|_\infty=O(\Delta(t)). \tag{3.12}$$

By (3.11) and (3.12) with some simple manipulations, it is readily shown that (3.7) and hence (3.6) hold.

Note that $\Phi(z)$ is analytic on $|z|=1$, so that by (3.6) and condition (A1) it is evident that on $|z|=1$,

$$\begin{aligned}\Phi_t(z)&=\Phi(z)+[\Phi_t(z)-\Phi(z)]\\&\geqslant\min_{|z|=1}\lambda_{\min}[\Phi(z)]\cdot I-\|\Phi_t(z)-\Phi(z)\|_\infty\cdot I\\&\geqslant\frac{1}{2}\min_{|z|=1}\lambda_{\min}[\Phi(z)]I>0\end{aligned}$$

provided that t is appropriately large. Therefore, from here and $\Delta(t)\to 0$ we know that there exists some $t_0>0$ such that for all $t\geqslant t_0$, $\Phi_t(z)$ satisfies conditions (A1)—(A4) with $\Phi(z)$ replaced by $\Phi_t(z)$. Hence, Lemma 2.1 is applicable to $\Phi_t(z)$ for $t\geqslant t_0$, and then all conclusions of the lemma except (3.5) follow immediatley.

We now proceed to prove the last conclusion (3.5). First let us establish convergence.

Substituting (3.2) into (3.1) and applying (2.1) and (3.9), we find that for $t\geqslant t_0$,

$$\begin{aligned}L(t)&=\frac{1}{2\pi j}\oint\Phi_t(z)z^{-1}dz\\&=\Omega(t)+\frac{1}{2\pi j}\oint H^\tau(t)[zI-F^\tau(t)]^{-1}K(t)\Omega(t)K^\tau(t)\\&\quad\cdot[z^{-1}I-F(t)]^{-1}H(t)z^{-1}dz\\&=\Omega(t)+\sum_{i=0}^{\infty}H^\tau(t)[F^\tau(t)]^iK(t)\Omega(t)K^\tau(t)[F(t)]^iH(t),\end{aligned}$$

where $j\triangleq\sqrt{-1}$ and the integral is around the unit circle $|z|=1$. From here we have

$$\begin{aligned}\operatorname{tr}L(t)&\geqslant\operatorname{tr}\Omega(t)+\operatorname{tr}\sum_{i=0}^{m-1}\Omega^{\frac{1}{2}}(t)K^\tau(t)[F(t)]^iH(t)H^\tau(t)[F^\tau(t)]^iK(t)\Omega^{\frac{1}{2}}(t)\\&\geqslant\operatorname{tr}\Omega(t)+[\operatorname{tr}K(t)\Omega(t)K^\tau(t)]\cdot\lambda_{\min}\Big(\sum_{i=0}^{m-1}[F(t)]^iH(t)H^\tau(t)[F^\tau(t)]^i\Big)\end{aligned} \tag{3.13}$$

Since $[F(t),\ H(t)]\to[F,\ H]$ and $[F,\ H]$ is controllable, it is evident that

$$\liminf_{t\to\infty}\lambda_{\min}\Big(\sum_{i=0}^{m-1}[F(t)]^iH(t)H^\tau(t)[F^\tau(t)]^i\Big)>0.$$

So from here and (3.13) it is known that both the sequence $\{\Omega(t)\}$ and the sequence $\{K(t)\Omega(t)K^\tau(t)\}$ are bounded. We now show that $\{K(t)\}$ is also a bounded sequence. For this, by (3.13) it suffices to show that

$$\inf_{t\geqslant t_0}\lambda_{\min}[\Omega(t)]>0. \tag{3.14}$$

Again, substituting (3.2) into (3.1) and noting (2.1) we have the following identity.

$$L(t)+H^{\tau}(t)[zI-F^{\tau}(t)]^{-1}M(t)+M^{\tau}(t)[z^{-1}I-F(t)]^{-1}H(t)$$
$$=\{I+H^{\tau}(t)[zI-F^{\tau}(t)]^{-1}K(t)\}\Omega(t)\{I+K^{\tau}(t)[z^{-1}I-F(t)]^{-1}H(t)\}.$$

From here by use of the following relation

$$[zI-F^{\tau}(t)]^{-1}=\mathrm{Adj}[zI-F^{\tau}(t)]/\det[zI-F^{\tau}(t)]$$

it then follows that

$$L(t)\det[zI-F^{\tau}(t)][z^{-1}I-F(t)]+H^{\tau}(t)\mathrm{Adj}[zI-F^{\tau}(t)]M(t)\det[z^{-1}I-F(t)]$$
$$+M^{\tau}(t)\mathrm{Adj}[z^{-1}I-F(t)]H(t)\det[zI-F^{\tau}(t)]$$
$$=\{\det[zI-F^{\tau}(t)]I+H^{\tau}(t)\mathrm{Adj}[zI-F^{\tau}(t)]K(t)\}\Omega(t)\{\det[z^{-1}I-F(t)]I$$
$$+K^{\tau}(t)\mathrm{Adj}[z^{-1}I-F(t)]H(t)\}.$$

Recall that $F(t)$ is an $m\times m$ matrix. Comparing the coefficients of z^m on both sides of the preceding identity it is not difficult to show that

$$L(t)[\det F(t)]\cdot(-1)^m+M^{\tau}(t)[\mathrm{Adj}F(t)]H(t)(-1)^{m-1}$$
$$=\Omega(t)\{(-1)^m\det F(t)I+(-1)^{m-1}K^{\tau}(t)[\mathrm{Adj}F(t)]H(t)\}.$$

Then it follows that

$$\Omega^{\frac{1}{2}}(t)\{\Omega^{\frac{1}{2}}(t)\det F(t)-\Omega^{\frac{1}{2}}(t)K^{\tau}(t)[\mathrm{Adj}F(t)]H(t)\}$$
$$=[\det F(t)]L(t)-M^{\tau}(t)[\mathrm{Adj}F(t)]H(t)\xrightarrow[t\to\infty]{}[\det F]L-M^{\tau}[\mathrm{Adj}\,F]H \tag{3.15}$$

which is nonsingular by condition (A4).

Since both $\{\Omega^{\frac{1}{2}}(t)\}$ and $\{\Omega^{\frac{1}{2}}(t)K^{\tau}(t)\}$ are bounded sequences, from (3.15) and the nonsingularity of its limit, it is evident that any limit points of $\{\Omega^{\frac{1}{2}}(t)\}$ are nonsingular. Consequently, we conclude that any limit points of $\{\Omega(t)\}$ are nonsingular. From this, it is easy to conclude (3.14) and hence the boundedness of $\{K(t)\}$.

Next, we show that

$$K(t)\to K \text{ and } \Omega(t)\to\Omega, \tag{3.16}$$

where K and Ω are given in Lemma 2.1.

Let $\{K',\Omega'\}$ be any limit point of $\{K(t),\Omega(t)\}$. It is known from the above that $\Omega'>0$.

Without loss of generality, we assume

$$\lim_{t\to\infty}K(t)=K',\quad \lim_{t\to\infty}\Omega(t)=\Omega'.$$

Taking limits on both sides of (3.1), (3.2) and noting (2.7), (3.7) and (3.6) we have

$$W(z)\Omega W^{\tau}(z^{-1})=W'(z)\Omega'W'^{\tau}(z^{-1}),$$

where

$$W'(z)=I+H^{\tau}(zI-F^{\tau})^{-1}K'. \tag{3.17}$$

Note that for all $t\geq t_0$, $W_t(z)$ is asymptotically stable and strictly minimum phase with $W_t(\infty)=I,\forall t\geq t_0$. Evidently, its limit, $W'(z)$ with $W'(\infty)=I$, is asymptotically stable minimum phase. Hence by the uniqueness of factorizations we know that

$$\Omega'=\Omega, \quad W'(z)=W(z). \tag{3.18}$$

We now show that $K'=K$. By (2.8), (3.17) and (3.18) it follows that

$$\sum_{i=0}^{\infty} H^{\tau}(F^{\tau})^{i}Kz^{i+1}=\sum_{i=0}^{\infty} H^{\tau}(F^{\tau})^{i}K'z^{i+1}, \quad |z|=1.$$

This implies that

$$\begin{bmatrix} H^{\tau} \\ H^{\tau}F^{\tau} \\ \vdots \\ H^{\tau}(F^{\tau})^{m-1}\end{bmatrix} K=\begin{bmatrix} H^{\tau} \\ H^{\tau}F^{\tau} \\ \vdots \\ H^{\tau}(F^{\tau})^{m-1}\end{bmatrix} K'$$

and consequently $K=K'$ by the controllability assumption of $[F, H]$.

Thus, we have shown that any limit point of $\{K(t), \Omega(t)\}$ concides with $\{K, \Omega\}$. This proves (3.16).

Now, by the definition of $K(t)$ and $\Omega(t)$ we see that $\Sigma(t)$ defined by (3.4) can be rewritten as

$$\Sigma(t)=F^{\tau}(t)\Sigma(t)F(t)-K(t)\Omega(t)K^{\tau}(t)$$

or

$$\Sigma(t)=-\sum_{i=0}^{\infty}[F^{\tau}(t)]^{i}K(t)\Omega(t)K^{\tau}(t)[F(t)]^{i}, \quad t\geqslant t_0. \tag{3.19}$$

Now, by use of (3.10) and the boundedness of $\{K(t), \Omega(t)\}$ it is readily shown that the series in (3.19) is uniformly convergent in $t\geqslant t_0$, so by taking the limit we get

$$\Sigma(t)\xrightarrow[t\to\infty]{} -\sum_{i=0}^{\infty}(F^{\tau})^{i}K\Omega K^{\tau}F^{i}=\Sigma, \tag{3.20}$$

where Σ is given by (2.11).

We are now in a position to establish the convergence rate (3.5).

Since $\Sigma(t)\to\Sigma$ and $\Delta(t)\to 0$, from (3.4), with some manipulations it is easy to see that

$$\Sigma(t)=F^{\tau}(t)\Sigma(t)F(t)-\overline{K}(t)\Omega(t)\overline{K}^{\tau}(t)+O(\Delta(t)), \tag{3.21}$$

where

$$\overline{K}(t)=[F^{\tau}\Sigma(t)H+M]\Omega^{-1}(t), \quad \Omega(t)=H^{\tau}\Sigma(t)H+L. \tag{3.22}$$

Consequently, by (2.9)—(2.11) and (3.21)—(3.22), it follows that

$$\begin{aligned}\Sigma(t)-\Sigma&=F^{\tau}[\Sigma(t)-\Sigma]F-\overline{K}(t)\Omega(t)\overline{K}^{\tau}(t)+K\Omega K^{\tau}+O(\Delta(t))\\ &=F^{\tau}[\Sigma(t)-\Sigma]F-\overline{K}(t)\Omega(t)\overline{K}^{\tau}(t)+K\Omega K^{\tau}+K[\Omega-\Omega(t)]\overline{K}^{\tau}(t)\\ &\quad+KH^{\tau}[\Sigma(t)-\Sigma]H\overline{K}^{\tau}(t)+O(\Delta(t))\\ &=F^{\tau}[\Sigma(t)-\Sigma]F-[\overline{K}(t)\Omega(t)-K\Omega]\overline{K}^{\tau}(t)-K[\Omega(t)\overline{K}^{\tau}(t)-\Omega K^{\tau}]\\ &\quad+KH^{\tau}[\Sigma(t)-\Sigma]H\overline{K}^{\tau}(t)+O(\Delta(t))\\ &=(F-HK^{\tau})^{\tau}[\Sigma(t)-\Sigma][F-H\overline{K}^{\tau}(t)]+O(\Delta(t)).\end{aligned} \tag{3.23}$$

Now by (3.20) and (3.22) we know that

$$F-H\overline{K}^{\tau}(t)\xrightarrow[t\to\infty]{} F-HK^{\tau}$$

which satisfies $|\lambda_i(F-HK^\tau)|<1$ by Lemma 2.1. So from (3.23) and inequality (3.10) we see that for sufficiently large t,

$$\Sigma(t)-\Sigma=\sum_{i=0}^{\infty}[F^\tau-KH^\tau]^i O(\Delta(t))[F-H\overline{K}^\tau(t)]^i=O(\Delta(t)).$$

This proves the last conclusion of Lemma 3.1 and the proof of the lemma is completed.

We are now able to prove Theorem 2.1.

Proof of Theorem 2.1. Since for sufficiently large t, $\{F(t),\ H(t),\ M(t),\ L(t)/2\}$ is a minimal realization of $M^\tau(t)[z^{-1}I-F(t)]^{-1}H(t)+L(t)/2$, which is a discrete positive real transfer function (see Lemma 3.1), for any sufficiently large but fixed t applying a result in [4, p. 745] we know that $\Sigma_t(s)$ defined by (2.2) monotonically decreases in s and $\Sigma_t(s)\geqslant\Sigma(t)$, for any $s\geqslant 0$, where $\Sigma(t)$ is given in Lemma 3.1. Consequently, for appropriately large t, $[H^\tau\Sigma_t(s)H+L(t)]$ is nonsingular for any $s\geqslant 0$. By (2.2) and (3.4), a similar treatment as used in the derivation of (3.23) leads to

$$\Sigma_t(s+1)-\Sigma(t)=[F(t)-H(t)K^\tau(t)]^\tau[\Sigma_t(s)-\Sigma(t)][F(t)-H(t)K_t^\tau(s)], \tag{3.24}$$

for sufficiently large t and any $s\geqslant 0$, where

$$K_t(s)=[F^\tau(t)\Sigma_t(t)H(t)+M(t)]\Omega_t^{-1}(s),$$
$$\Omega_t(s)=H^\tau(t)\Sigma_t(s)H(t)+L(t).$$

Now, note that

$$\begin{aligned}F(t)-H(t)K_t^\tau(s)&=F(t)-H(t)\Omega_t^{-1}(s)[H^\tau(t)\Sigma_t(s)F(t)+M^\tau(t)]\\&=F(t)-H(t)\Omega_t^{-1}(s)H^\tau(t)[\Sigma_t(s)-\Sigma(t)]F(t)\\&\quad-H(t)\Omega_t^{-1}(s)[H^\tau(t)\Sigma(t)F(t)+M^\tau(t)]\\&=\{I-H(t)\Omega_t^{-1}(s)H(t)[\Sigma_t(s)-\Sigma(t)\}\cdot F(t)\\&\quad-H(t)\{I-\Omega_t^{-1}(s)H^\tau(t)[\Sigma_t(s)-\Sigma(t)]H(t)\}K^\tau(t)\\&=\{I-H(t)\Omega_t^{-1}(s)H^\tau(t)[\Sigma_t(s)-\Sigma(t)]\}\\&\quad\cdot[F(t)-H(t)K^\tau(t)].\end{aligned}$$

Substituting this identity into (3.24) and noting the fact that for sufficiently large t, $\Sigma_t(s)-\Sigma(t)\geqslant 0$, $\forall s\geqslant 0$, and $\Omega_t(s)\geqslant\Omega(t)>0$, $\forall s\geqslant 0$, we then have

$$\begin{aligned}\Sigma_t(s+1)-\Sigma(t)&\leqslant[F(t)-H(t)K^\tau(t)]^\tau[\Sigma_t(s)-\Sigma(t)][F(t)-H(t)K^\tau(t)]\\&\leqslant[F^\tau(t)-K(t)H^\tau(t)]^{s+1}[\Sigma_t(0)-\Sigma(t)][F(t)-H(t)K^\tau(t)]^{s+1}.\end{aligned}$$

Again by the convergence of $K(t)\to K$, $\Delta(t)\to 0$ and $|\lambda_i(F-HK^\tau)|<1$ and inequality (3.10), from here it follows that

$$\begin{aligned}\hat{\Sigma}(t)-\Sigma(t)&=\Sigma_t([\log^2 t])-\Sigma(t)\\&=O(\lambda^2\log^2 t),\quad \lambda\in(0,1),\\&=O(t^{-\alpha\log t}),\quad \alpha=\log\frac{1}{\lambda^2}>0.\end{aligned}$$

Consequently, by Lemma 3.1 we see that

$$\hat{\Sigma}(t)-\Sigma=[\Sigma(t)-\Sigma(t)]+[\Sigma(t)-\Sigma]=O(\Delta(t))+O(t^{-\alpha\log t}). \tag{3.25}$$

To complete the proof of Theorem 2.1 we need to establish similar convergence rates for $\hat{W}_t(z)$ and $\hat{\Omega}(t)$, i.e., (2.13) and (2.14).

By the definition (2.4) for $\hat{\Omega}(t)$ and (3.25) it is obvious that the conclusion (2.14) holds.

By the definition (2.6) for $\hat{K}(t)$, (2.14), and (3.25) it is evident that after some simple manipulations,

$$\|\hat{K}(t)-K\|=O(\Delta(t))+O(t^{-\alpha\log t}),$$

where K is given in Lemma 2.1. From here and (3.12) and the expressions (2.5) and (2.8) for $\hat{W}_t(z)$ and $W(z)$, it is clear that

$$\|\hat{W}_t(z)-W(z)\|_\infty=O(\Delta(t))+O(t^{-\alpha\log t}).$$

This completes the proof of Theorem 2.1.

4. Application to Linear Regression Model Identification

Let us consider the following l-dimensional linear regression model:

$$y(t)=\theta_0 x(t)+s(t), \quad t\geqslant 0, \tag{4.1}$$

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

Assume that the system noise $s(t)$ is a moving average process

$$s(t)=w(t)+C_1 w(t-1)+\cdots+C_r w(t-r) \tag{4.2}$$

with unknown matrix coefficient C_j, $1\leqslant j\leqslant r$, where the driven noise $\{w(t)\}$ is assumed to be a Gaussian white noise sequence and

$$Ew(t)=0, \quad Ew(t)w^\tau(t)=R_w>0, \quad t\geqslant 0. \tag{4.3}$$

Let us denote

$$C(z)=I+C_1 z^{-1}+\cdots+C_r z^{-r}. \tag{4.4}$$

To identify the unknown parameter θ_0 and the unknown noise model $C(z)$ and R_w, consider the introduction of a Gaussian white noise sequence $\{v(t)\}$ which is independent of $\{w(t)\}$ with properties

$$Ev(t)=0, \quad Ev(t)v^\tau(t)=\sigma_v^2 I_l, \quad \sigma_v^2>0. \tag{4.5}$$

Define the "pre-whitening" process $\{z(t)\}$ as

$$z(t)=y(t)+v(t) \tag{4.6}$$

and consider the following prediction error algorithm:

$$\hat{\theta}(t+1)=\hat{\theta}(t)+P(t)\psi(t)[z^\tau(t+1)-\psi^\tau(t)\hat{\theta}(t)], \tag{4.7}$$

$$P(t)=P(t-1)-\frac{P(t-1)\psi(t)\psi^\tau(t)P(t-1)}{1+\psi^\tau(t)P(t-1)\psi(t)}, \quad P(0)>0, \tag{4.8}$$

$$\psi(t)=[x^\tau(t),\ z^\tau(t)-\psi^\tau(t-1)\hat{\theta}(t)\cdots,\ z^\tau(t-r+1)-\psi^\tau(t-r)\hat{\theta}(t-r+1)]^\tau \tag{4.9}$$

together with the following estimates for the covariance of the prediction errors:

$$\hat{R}_{\overline{w}}(t)=\frac{1}{t}\sum_{i=0}^{t-1}[z(i+1)-\hat{\theta}^{\tau}(t)\psi(i)][z(i+1)-\hat{\theta}^{\tau}(t)\psi(i)]^{\tau} \tag{4.10}$$

(there is a recursive form for $\hat{R}_{\overline{w}}(t)$ also).

Let us denote

$$G_t^0=\sigma\{s(i)+v(i),\quad i\leq t\} \tag{4.11}$$

and assume that the regression vector $\{x(t), F_{t-1}\}$ is any adapted random sequence, where F_t is defined by

$$F_t=\sigma\{G_t^0\cup G_t^1\} \tag{4.12}$$

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 $t\geq 0$.

In [3] we have established the following convergence results.

Lemma 4.1. *For the model and algorithm described by* (4.1)—(4.12), *if in the pre-whitening of* (4.5)—(4.6), σ_v^2 *is chosen to satisfy*

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

then the following convergence rates hold:

(i)
$$\|\hat{\theta}(t+1)-\overline{\theta}\|=O\left[\sqrt{\frac{\log\lambda_{\max}(t)}{\lambda_{\min}(t)}}\right],\quad \text{a.s. } t\to\infty, \tag{4.13}$$

(ii)
$$\|\hat{R}_{\overline{w}}(t)-R_{\overline{w}}\|=O\left[\sqrt{\frac{\log\lambda_{\max}(t)}{\lambda_{\min}(t)}}\right]+O\left[\frac{\log\lambda_{\max}(t)}{\lambda_{\min}(t)}\right]. \tag{4.14}$$

Here

$$\overline{\theta}=[\theta_0,\ D_1,\ \cdots,\ D_r]^{\tau} \tag{4.15}$$

and $\{D_i,\ 1\leq i\leq r,\ R_{\overline{w}}\}$ *satisfies*

$$D(z)R_{\overline{w}}D^{\tau}(z^{-1})=C(z)R_wC^{\tau}(z^{-1})+\sigma_v^2I \tag{4.16}$$

with

$$D(z)\triangleq I+D_1z^{-1}+\cdots+D_rz^{-r}. \tag{4.17}$$

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

$$\sum_{i=0}^{t}\psi(i)\psi^{\tau}(i)+sI,\quad s>0.$$

Now we use the results of the present paper to show how to recover the original noise model $C(z)$ and R_w.

For this purpose, let us denote

$$\Phi(z)=C(z)R_wC^{\tau}(z^{-1}) \tag{4.18}$$
$$=L+M_1z^{-1}+\cdots+M_rz^{-r}+M_1^{\tau}z+\cdots+M_r^{\tau}z^r. \tag{4.19}$$

Set

$$F=\begin{bmatrix}0 & I & & 0\\ & \cdot & \cdot & \\ & & \cdot & I\\ 0 & & & 0\end{bmatrix},\quad H=\begin{bmatrix}0\\ \cdot\\ \cdot\\ \cdot\\ I\end{bmatrix},\quad M=\begin{bmatrix}M_r\\ \cdot\\ \cdot\\ \cdot\\ M_1\end{bmatrix}. \tag{4.20}$$

Then $\Phi(z)$ can be expressed by

$$\Phi(z)=L+H^{\tau}(zI-F^{\tau})^{-1}M+M^{\tau}(z^{-1}I-F)^{-1}H. \tag{4.21}$$

In order to find estimates $\Phi_t(z)$ for $\Phi(z)$, let us write $\hat{\theta}(t)$ defined by (4.7) in its component form

$$\hat{\theta}(t)=[\hat{\theta}_0(t),\ \hat{D}_1(t)\cdots\hat{D}_r(t)]^\tau \tag{4.22}$$

and set

$$\hat{D}_t(z)=I+\hat{D}_1(t)z^{-1}+\cdots+\hat{D}_r(t)z^{-r}. \tag{4.23}$$

Then we can formulate $\Phi_t(z)$ as

$$\begin{aligned}\Phi_t(z)&=\hat{D}_t(z)\hat{R}_w(t)\hat{D}_t^\tau(z^{-1})-\sigma_v^2 I\\&\triangleq L(t)+M_1(t)z^{-1}+\cdots+M_r(t)z^{-r}+M_1^\tau(t)z+\cdots+M_r^\tau(t)z^r\end{aligned} \tag{4.24}$$

which also can be expressed by

$$\Phi_t(z)=L(t)+H^\tau(zI-F^\tau)^{-1}M(t)+M^\tau(t)(z^{-1}I-F)^{-1}H \tag{4.25}$$

with

$$M(t)=[M_r^\tau(t)\cdots M_1^\tau(t)]^\tau. \tag{4.26}$$

We now have the following results:

Theorem 4.1. *Assume that the conditions of Lemma 4.1 are applied, that in the adaptive spectral factorization algorithm (2.1)—(2.6), $F(t)\equiv F$, $H(t)\equiv H$ and F, H, $L(t)$ and $M(t)$ are specified in (4.20), (4.24) and (4.26), and that the following conditions are satisfied:*

(i) *$C(z)$ is strictly minimum phase and* $\det C_r\neq 0$,

(ii) $\log\lambda_{\max}(t)/\lambda_{\min}(t)\xrightarrow[t\to\infty]{}0$, $\log\lambda_{\max}(t)/t\to 0$.

Then,

$$\|\hat{W}_t(z)-C(z)\|_\infty=O\left(\left[\frac{\log\lambda_{\max}(t)}{\lambda_{\min}(t)}\right]^{\frac{1}{2}}\right)+O\left(\frac{\log\lambda_{\max}(t)}{t}\right),$$

$$\|\hat{\Omega}(t)-R_w\|=O\left(\left[\frac{\log\lambda_{\max}(t)}{\lambda_{\min}(t)}\right]^{\frac{1}{2}}\right)+O\left(\frac{\log\lambda_{\max}(t)}{t}\right),$$

where $\hat{W}_t(z)$ and $\hat{\Omega}(t)$ are given respectively by (2.5) and (2.4).

Proof. Lemma 4.1 guarantees that

$$\Delta(t)=O\left(\left[\frac{\log\lambda_{\max}(t)}{\lambda_{\min}(t)}\right]^{\frac{1}{2}}\right)+O\left(\frac{\log\lambda_{\max}(t)}{t}\right).$$

Hence for Theorem 2.1 to apply we need only to verify that conditions (A1)—(A4) are satisfied in the present case. But this is straightforward since $\det C_r\neq 0$ ensures $\det M_r\neq 0$, which in turn ensures the observability of $[F,\ M]$ and condition (A4).

5. Application to Moving Average Model Identification

Consider the following moving average model $MA(r)$:

$$y(t)=w(t)+C_1 w(t-1)+\cdots+C_r w(t-r) \tag{5.1}$$

with unknown coefficients C_i, $1\leqslant i\leqslant r$, and noise covariance R_w, and without loss of generality take $C(z)$ to be minimum phase.

Of course, the model (5.1) is a specialization of the linear regression model

(4.1). But, here we shall show that in this case there is no need to add white noise $\{v(t)\}$ into the algorithm, the conditions imposed on $\{w(t)\}$ are weaker, and the results are stronger.

To be precise, let us assume that $\{w(t), F_t\}$ is a stationary ergodic martingale difference sequence with

$$E[w(t)w^{\tau}(t)\,|\,F_{t-1}]=R_w>0 \tag{5.2}$$

and

$$E[\|w(t)\|^4\,|\,F_{t-1}]\leqslant \text{constant}<\infty. \tag{5.3}$$

Let $\Phi(z)$ be defined also by (4.18)—(4.21). In contrast with the preceding section, we now formulate the estimates $L(t)$ and $M_i(t)$ for L and M_i as follows.

$$L(t)=\frac{1}{t}\sum_{s=1}^{t}y(s)y^{\tau}(s), \tag{5.4}$$

$$M_i(t)=\frac{1}{t}\sum_{s=1}^{t}y(s)y^{\tau}(s-i),\quad 1\leqslant i\leqslant r, \tag{5.5}$$

and set

$$M(t)=[M_r^{\tau}(t)\cdots M_1^{\tau}(t)]^{\tau}. \tag{5.6}$$

Theorem 5.1. *Consider the $MA(r)$ model as described above. Consider also that in the adaptive spectral factorization algorithm (2.1)—(2.6), $F(t)\equiv F$, $H(t)\equiv H$ and F, H, $L(t)$ and $M(t)$ are specified in (4.20) and (5.4)—(5.6). If $C(z)$ defined by (4.4) is strictly minimum phase and* $\det C_r\neq 0$, *then*

$$\|\hat{W}_t(z)-C(z)\|_{\infty}=O\left(\left[\frac{\log\log t}{t}\right]^{\frac{1}{2}}\right),\quad \text{a.s.}$$

and

$$\|\hat{\Omega}(t)-R_w\|=O\left(\left[\frac{\log\log t}{t}\right]^{\frac{1}{2}}\right),\quad \text{a.s.}$$

Proof. For Theorem 2.1 to apply, we need only to show that $\Delta(t)$ defined by (2.12) satisfies

$$\Delta(t)=O\left(\left[\frac{\log\log t}{t}\right]^{\frac{1}{2}}\right),\quad \text{a.s.} \tag{5.7}$$

Note that with $\{y(t)\}$ given by (5.1) and L, M_i defined in (4.19),

$$L=E[y(t)y^{\tau}(t)],\quad M_i=E[y(t)y^{\tau}(t-i)].$$

Also, with $L(t)$, $M_i(t)$ defined by (5.4)—(5.5), by the laws of the iterated logarithm for stationary ergodic martingales ([6]) it is readily seen that

$$\|L(t)-L\|=O\left(\left[\frac{\log\log t}{t}\right]^{\frac{1}{2}}\right),\quad \text{a.s.}$$

$$\|M_i(t)-M_i\|=O\left(\left[\frac{\log\log t}{t}\right]^{\frac{1}{2}}\right),\quad \text{a.s.}$$

which imply (5.7), and the proof is completed.

6. Conclusions

In stochastic system identification, extended least squares algorithms converge under a positive real condition on the coloured noise model. Here, algorithms based on adaptive (recursive, on-line) spectral factorization are studied which converge whether or not the positive real condition is satisfied. By setting up a specific form for the adaptive spectral factorization, it is shown that there is no compromise on convergence rates using this approach. There is, however, a computational cost as the time index t becomes large, being of the order of $\log^2 t$. In algorithms with computational cost invariant of t, there is a mild convergence rate cost.

References

[1] Anderson, B. D. O. and J. B. Moore, Optimal Filtering, Prentice Hall, Inc., Englewood Cliffs, N. J., 1979.

[2] Moore, J. B., Side-stepping the positive real restriction for stochastic adaptive schemes, *Richerche Di Automatica*, **8** (1982), 501—523.

[3] Guo, L., L. Xia and J. B. Moore, Robust recursive identification of multidimensional linear regression models, *International J. Control*, **48** (1988), 961—979.

[4] Anderson, B. D. O., K. L. Hitz and N. D. Diem, Recursive algorithm for spectral factorization, *IEEE Trans. Circuits and Systems*, CAS-21 (1974), 742—750.

[5] Chen, H. F., Recursive Estimation and Control for Stochastic Systems, John Wiley, New York, 1985.

[6] Stout, W. F., The Hartman-Wintner laws of the iterated logarithm for martingales, *Ann. Math. Statistics*, **41** (1970), 2158—2160.

[7] Chen, H. F. and L. Guo, Optimal adaptive control and consistent parameter estimates for ARMAX model with quadratic cost, *SIAM J. on Control and Optimisation*, **25** (1987), 845—867.

[8] Clancey, K. and I. Gohberg, Factorization of Matrix Functions and Singular Integral Operators, Birkhauser Verlag, 1981.