Stochastic Parallel Model Adaptation: Theory and Applications to Active Noise Cancelling, Feedforward Control, IIR Filtering and Identification*

Wei Ren[†] and P. R. Kumar[‡]

Abstract

We consider general stochastic parallel model adaptation problems which consist of an unknown linear time invariant system and a partially or wholly tunable system connected in parallel, with a common input. The goal of adaptation is to tune the partially tunable system so that its output matches that of the unknown system, despite the presence of any disturbance which is stochastically uncorrelated with the input.

Our general formulation of stochastic parallel adaptation schemes allows applications to adaptive feedforward control and adaptive active noise canceling with input contamination, in addition to output error identification and adaptive IIR filtering.

We show that in all the applications, the goal of adaptation is met, whenever a matching condition and a positive real condition are satisfied. A special case of our results therefore resolves the long standing problem of the convergence and the unbiasedness of output error identification scheme in the presence of colored noise. We also develop a simple general technique for analyzing the strong consistency of parameter estimation with projection.

1 Introduction

We consider general stochastic "parallel model" adaptation schemes which consist of an unknown linear time invariant system and a tunable system (or partly tunable and partly unknown system), connected in parallel, with a common input. The goal of adaptation is to

*The research reported here has been supported by the U.S.A.R.O. under Contract Nos. DAAL03-91-G-0182 and DAAL-03-88-K-0046, the Joint Services Electronics Program under Grant No. N00014-90-J-1270, and by an International Paper Company Fellowship for the first author.

[†]Dept. of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720.

[‡]Dept. of Electrical and Computer Engineering and Coordinated Science Laboratory, 1101 West Springfield Avenue, Urbana, IL 61801.

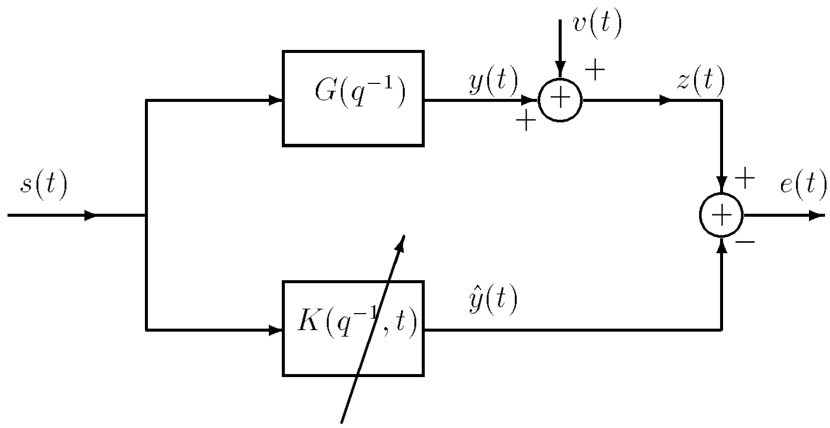

Figure 1: Problem A: Output error identification and adaptive IIR filtering.

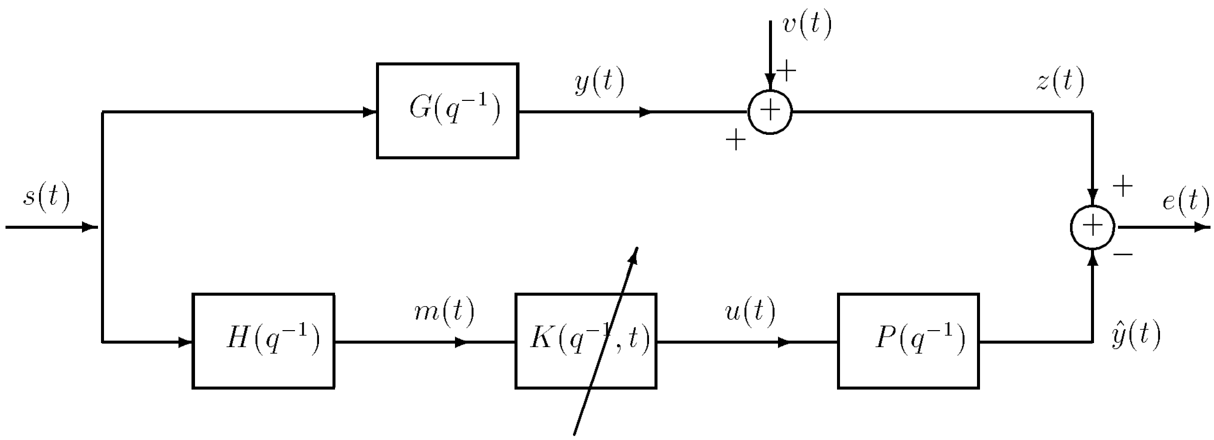

Figure 2: Problem B: Adaptive feedforward control.

tune the (partially) tunable system so that its output matches that of the unknown system, despite the presence of any disturbance which is stochastically uncorrelated with the input.

We have been motivated to study such stochastic "parallel model" adaptation schemes due to their applicability to the practical problems illustrated in Figures 1–3. All the fixed systems in Figures 1–3 are unknown but stable, linear and time invariant. In all the three figures, $G(q^{-1})$ is the system whose output $y(t)$ is to be matched by $\hat{y}(t)$, the output of the (partially) tunable system in the lower channels of the figures. The signal $s(t)$ is the common input to both upper and lower channel, while $v(t)$ is the disturbance.

Problem A, the simplest of the three, and depicted in Figure 1, encompasses output

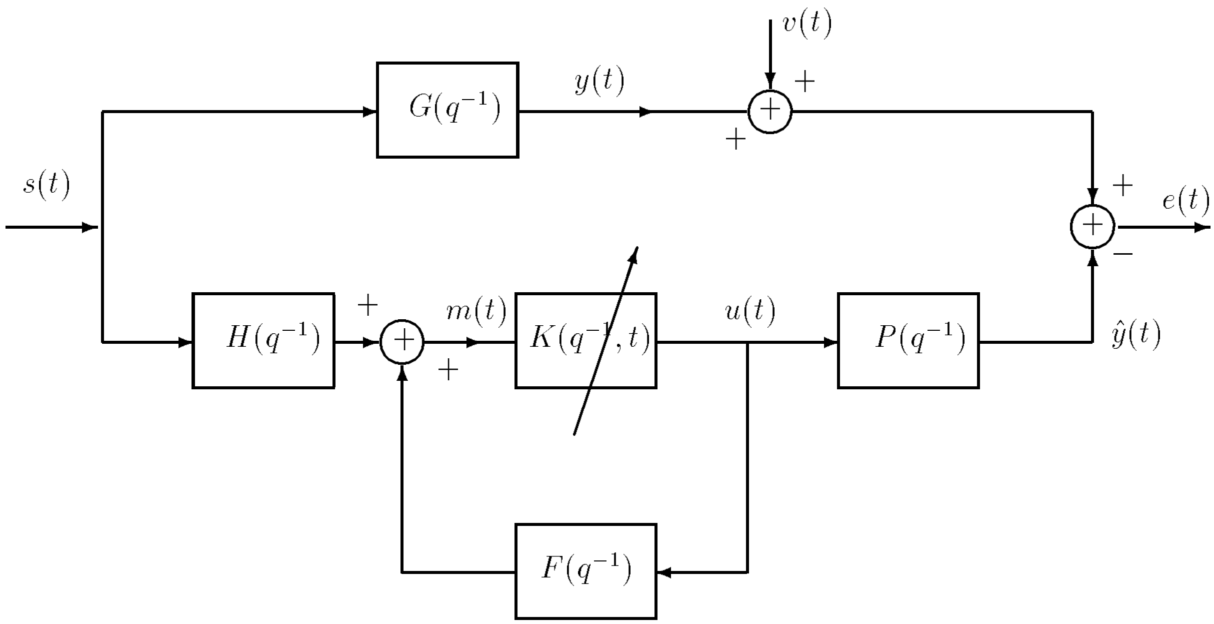

Figure 3: Problem C: Adaptive active noise canceling with input contamination.

error identification as well as adaptive IIR filtering. In the *identification* context, $G(q^{-1})$ is an unknown stable dynamical system to be identified, whose input $s(t)$, and output $z(t)$, corrupted by the noise $v(t)$, are measured. For this purpose we tune a "parallel system" $K(q^{-1}, t)$ so that its output $\widehat{y}(t)$ closely matches $z(t)$. The noise $v(t)$ can be arbitrary (for example, nonstationary), except for the reasonable and practical assumption that $\{v(t)\}$ and $\{s(t)\}$ are stochastically uncorrelated. It has been an open question for more than a decade now to show that such an adaptation scheme does indeed converge and identifiy $G(q^{-1})$ asymptotically, or at least "matches" its output "optimally."

Figure 1 is also applicable to the problem of *adaptive IIR filtering.* Here, $v(t)$ is the signal to be extracted, while $y(t)$ is the noise corrupting the signal $v(t)$. The signal $s(t)$ can be regarded as the source of the noise, while $G(q^{-1})$ is an unknown filter. The goal now is to tune the "parallel system" $K(q^{-1}, t)$ so that $\widehat{y}(t)$ closely matches $y(t)$, hence canceling the noise. If so, $e(t)$ can be regarded as an estimate of the signal $v(t)$. The attraction of this scheme is that the assumptions placed on the unknown signal $v(t)$ are minimal, allowing, for example, nonstationarity as in speech signals.

Problem B, depicted in Figure 2, is a generalization of Problem A, which arises in practice as an *adaptive feedforward control scheme for disturbance attenuation.* $P(q^{-1})$ is an unknown "plant," for which $u(t)$ is the input, and $e(t)$ is the output. The output $e(t)$ is subject to additive disturbances $v(t)$ and $y(t)$ which are mutually uncorrelated. A signal $m(t)$ related to $s(t)$, the source of the disturbance $y(t)$, is measured. The goal of adapting the *feedforward controller* $K(q^{-1}, t)$ is to eliminate from the output $e(t)$ that part which is due to $y(t)$, thus leaving only the uncancellable disturbance $v(t)$.

Problem C, depicted in Figure 3, incorporates a further generalization that is inspired by practical applications in the areas of active acoustic noise and vibration control [1]. Consider the (simplified) situation represented in Figure 4. Let A be a "source" of a noise, while D is the location at which the noise is to be cancelled. For this purpose we use a detection microphone to pick up a signal B related to the noise A. This signal is filtered and fed to a loudspeaker C. Our goal is to have this loudspeaker emit an "anti-noise" to cancel the noise at the location D. An error microphone picking up the uncancelled noise at D yields an error signal, which is used to tune the filter driving the loudspeaker.

This situation gives rise to the system shown in Figure 3, where $s(t)$ is the (fictitious) source signal for the noise, and $G(q^{-1})$ is the (unknown) stable transfer function from A to D. $H(q^{-1})$ represents the transfer function from A to B, with $m(t)$ as the output of the detection microphone. $K(q^{-1}, t)$ is the filter to be tuned, whose output $u(t)$ is the current in the loudspeaker coil. $P(q^{-1})$, which is unknown, but stable, represents the transfer function from the loudspeaker coil current to the error microphone output.

There are two additional features. First, there may be a signal $v(t)$, a part of the ambient sound uncorrelated with $s(t)$, which we do not wish to cancel, e.g., a speech or a warning signal. Thus, the goal is to have $e(t)$ replicate $v(t)$. Second, we have the additional complication that the detection microphone signal is generally contaminated by the anti-noise. This effect, called acoustic feedback, is modeled by $F(q^{-1})$. More generally, it has been termed as *input contamination* in [1].

Background

Extensive studies have been devoted to the special case of Problem A when the tunable

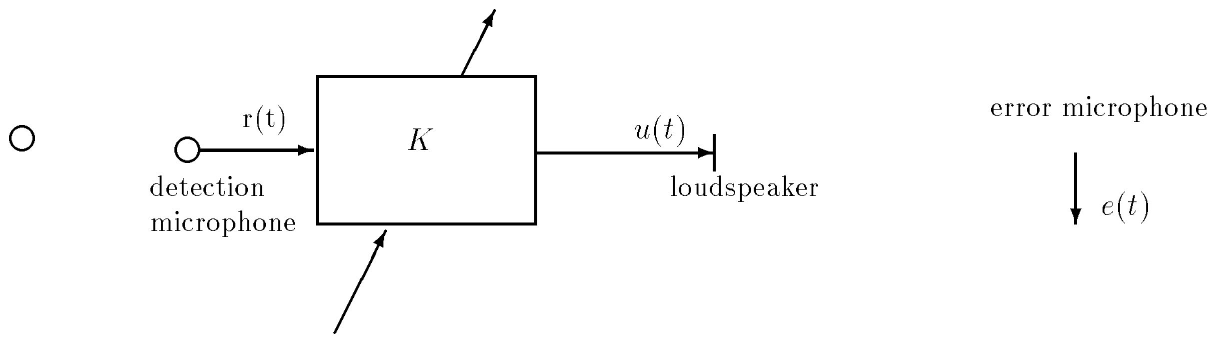

Figure 4: Active acoustic noise canceling

system $K(q^{-1},t)$ is restricted to be an all-zero (or FIR) model and a gradient-type (e.g., LMS) adaptive algorithm is used, see [2]. In this paper, we are concerned with the tuning of a general pole-zero model, i.e., the more general IIR case.

Let $K(q^{-1},t) = \frac{\hat{B}(q^{-1},t)}{\hat{A}(q^{-1},t)}$, where $\hat{A}(q^{-1},t) = 1 + \hat{a}(t)q^{-1} + \cdots + \hat{a}_n(t)q^{-n}$, and $\hat{B}(q^{-1},t) = \hat{b}_0(t) + \hat{b}_1(t)q^{-1} + \cdots + \hat{b}_m(t)q^{-m}$. The following algorithm for adapting $\hat{A}$ and $\hat{B}$, and its least squares and nonvanishing adaptation gain counterparts, are proposed first in Landau [3, 4] for identification, and applied in Johnson [5, 6] for filtering:

$$\begin{aligned} \theta(t) &= \theta(t-1) + \frac{\phi(t-1)}{r(t-1)}e(t), \\ r(t-1) &= r(t-2) + \|\phi(t-1)\|^2, \end{aligned} \tag{1}$$

where

$$\theta(t) := [\hat{a}_1(t), \ldots, \hat{a}_n(t), \hat{b}_0(t), \ldots, \hat{b}_m(t)]^{\mathrm{T}}, \tag{2}$$

$$e(t) := z(t) - \hat{y}(t),$$

$$\hat{y}(t) := K(q^{-1}, t-1)s(t) \equiv \phi^{\mathrm{T}}(t-1)\theta(t-1),$$

$$\phi(t-1) := [-\hat{y}(t-1), \ldots, -\hat{y}(t-n), s(t), \ldots, s(t-m)]^{\mathrm{T}}. \tag{3}$$

It has been conjectured by Landau [3, 4] that the above algorithm provides asymptotically unbiased parameter estimates, provided that there exists a $K(q^{-1})$ which results in a match between the upper and lower channels, i.e. $K(q^{-1})s(t) \equiv G(q^{-1})s(t)$, and a certain positive real condition is satisfied. In the filtering context, the signal $v(t)$ is estimated by $e(t)$. Therefore the corresponding conjecture in filtering is that $|v(t) - e(t)| \to 0$, or, in a long-term average sense,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} [e(t) - v(t)]^2 = 0 \qquad \text{a.s.} \tag{4}$$

The algorithm will be said to be *self-optimizing* or to have *optimal performance* if the above holds.

It is worthwhile emphasizing that no attempt is made in (1) to identify the noise model of $v(t)$; hence the algorithm is potentially applicable to the nonstationary noise case, as we exhibit in this paper. This is in contrast to identification schemes based on an ARMAX model parameterization where the unbiasedness of the parameter estimates of the dynamical part of the system crucially depends on the identification of the noise model. This ability to handle nonstationary noise is important in identification, since the dynamical part of the system is usually relatively stationary compared with the noise spectrum. The treatment of nonstationary noise is also important in filtering, since signals are generically nonstationary.

The important conjecture that the parallel model adaptation scheme (1) provides unbiased parameter estimates and undistorted signal extraction in the presence of general nonstationary colored noise has never been shown rigorously, although it appears plausible in view of the ODE analysis in Ljung [7, 8] and Dugard and Landau [9].

When $\{v(t)\}$ is restricted to be a white noise (or a martingale difference sequence), the techniques of Solo [10] can be extended easily to prove convergence. However, the more useful case when $v(t)$ is colored has defied any solution so far. The difficultly in analysis stems from the fact that the regression vector $\phi(t-1)$ is *not* uncorrelated with the error signal $e(t)$ unless the parameter estimates are held constant in the stable region.

We provide below an answer to this challenge, slightly modifying the algorithm to project the parameter estimates at each time step onto a compact convex set containing the true parameter, by choosing the closest point in the compact convex set whenever the estimate leaves the set. Such a projection has been shown to be useful for enhancing the robustness of adaptive systems [11]. The key ideas of the proof consist of the use of martingale limit theory, a decomposition of the disturbance $v(t)$, a "backward" recursion, and finally, a technique for establishing the strong consistency of the projected parameter estimates. These ideas are found to be useful in analyzing other adaptive systems too [12].

Adaptation schemes for Problems B and C which guarantee global convergence have never been considered formally and with rigor to the best of our knowledge. So all the results here appear to be new for adaptive feedforward control and active noise cancelling. In the latter case, it is interesting that our algorithm for tuning $K(q^{-1},t)$ succeeds in *simultaneously* adapting to all four unknown transfer functions, $G(q^{-1}), H(q^{-1}), P(q^{-1}),$ and$F(q^{-1})$.

Throughout this paper, we consider the ideal case when at least the following *weak matching condition* is satisfied.

Definition. *Given $\{s(t)\}$ bounded, and stable transfer functions G, P, F and H, the general parallel model system of Figure 3 is said to be weakly matchable if there exists a causal, finite dimensional linear time invariant controller $K(q^{-1})$ such that the entire system is internally stable, and*

$$\frac{1}{N}\sum_{t=1}^{N}[e(t)-v(t)]^2 = O\left(\frac{1}{\log^{1+\delta} N}\right), \text{ for some } \delta > 0.$$

We note that the weak matching condition is "signal-dependent," and is weaker than simply assuming that $e(t)-v(t) \equiv 0$. This allows us to exploit the important special case when $s(t)$ is periodic, as is often the case in noise canceling. For the output matching to be achieved for *all* sequences $s(t)$, one needs $K(q^{-1}) = \frac{G}{HP-GF}$. It is easy to show that the necessary and sufficient condition for the above controller to be causal, and to result in the internal stability of the entire system is the following *strong* matching condition.

Definition. *The parallel model system of Figure 3 is said to be strongly matchable if the non-minimum phase zeros of H and P are also the zeros of G, and the number of pure delays of HP is less than or equal to that of G.*

For clarity and ease of understanding, we take a bottom-up approach in this paper, i.e., simpler problems are studied first in detail, and then generalized. Section 2 establishes the optimal performance of the output error identification scheme in the presence of a fairly general class of colored noise. The strong consistency of the parameter estimates is established in Section 3. These convergence results are obtained for both the stochastic gradient (SG) algorithm and a modified least squares (MLS) algorithm. Although our definition of MLS algorithm includes SG algorihm as a special case, its analysis is more complicated and hence mostly relegated to the Appendix. Section 4 addresses the generalizations required for Problems B and C, adaptive feedforward control and active noise cancelling.

2 OPTIMAL PERFORMANCE OF OUTPUT ERROR RECURSION

In this section, we establish the self-optimizing property of SG and MLS algorithms for the Problem A, in the long term average sense, i.e., in the sense of (4).

2.1 Assumptions

Throughout this section, we will make the following assumptions on the system of Figure 1.

Assumptions.

A1) The disturbance is of the form $v(t) = \sum_{i=0}^{t} c_i(t) w(t-i)$, *where* $\{w(t)\}$ *is a martingale difference sequence with respect to the increasing sequence of* σ*-fields* $\mathcal{F}_t$ *generated by* $(w(0), \ldots, w(t), \{u(k)\}_0^\infty)$, *and* $\{c_i(t)\}$ *is deterministic and satisfies* $|c_i(t)| \leq K_c \alpha^{-i}$, $\forall t \geq 0$, *for some* $\alpha > 1$ *and* $K_c < \infty$.

A2) The sequences $\{|y(t)|\}$, $\{|s(t)|\}$, $\{|v(t)|\}$, *and* $\{|w(t)|\}$, *are uniformly bounded by a finite number* Γ, *almost surely.*

A3) The system is weakly matchable almost surely, i.e., there exists a stable finite dimensional transfer function $K_0(q^{-1})$ *such that*

$$\epsilon(t) := [G(q^{-1}) - K_0(q^{-1})]s(t) \tag{5}$$

satisfies

$$\frac{1}{N}\sum_{t=1}^{N}\epsilon^2(t) = O\left(\frac{1}{\log^{1+\delta} N}\right), \text{ a.s., } \textit{for some } \delta > 0. \tag{6}$$

A4) Let $K_0(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}$ *be one of the transfer functions satisfying (A3), where*

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_1} q^{-n_1},$$

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_{m_1} q^{-m_1}.$$

Assume $n_1 \le n$, $m_1 \le m$. *Further, let* $\theta_0 := [a_1, \ldots, a_{n_1}, 0, \ldots, 0, b_0, \ldots, b_{m_1}, 0, \ldots, 0]^{\mathrm{T}}$, *where the zeros are padded to match* $\theta(t)$ *defined in (2). We assume that a compact convex set* $\mathcal{M}$ *is known a-priori such that* $\theta_0 \in \mathcal{M}$.

A5) $\lim_{N\to\infty} \inf \frac{1}{N}\sum_{t=1}^{N} s^2(t) > 0.$ *a.s.*

Assumption (A1) states precisely what we mean by the requirement that $\{s(t)\}$ and $\{v(t)\}$ are "stochastically uncorrelated." It also assumes that the signal $v(t)$ has a geometrically decaying autocorrelation. It is worth pointing out that if the signal $v(t)$ has a *non-decaying* autocorrelation, say a random sinusoid, then the adaptive IIR filter may cancel $v(t)$ by simply tuning two of its poles on the unit circle to coincide with the frequency of the sinusoid, since that would result in a smaller mean square value of $e(t)$, thus resulting in poor performance.

Assumption (A2) is made for simplicity of analysis, and since it is quite reasonable in practice.

As mentioned in the previous section, the weak matching assumption (A3) allows us to exploit a smaller dimensional parameterization when $s(t)$ is periodic. Moreover, the effects of the initial conditions can also be lumped into $\epsilon(t)$. Hence, without loss of generality, we will assume hereafter that all dynamical systems have zero initial conditions.

2.2 Projected SG Algorithm With A Posteriori Estimates

We first consider the following SG algorithm for updating $K(q^{-1},t) = \frac{\hat{B}(q^{-1},t)}{\hat{A}(q^{-1},t)}$ of the Problem A. Let $\theta(t)$ be as defined in (2). It is updated by

$$\theta'(t) = \theta(t-1) + \frac{\phi(t-1)}{r(t-1)}e(t), \quad \theta(0) \in \mathcal{M}, \tag{7}$$

$$\theta(t) = F[\theta'(t)], \tag{8}$$

where $F[\cdot]$ denotes the projection onto $\mathcal{M}$, (i.e. $F(\theta) \in \mathcal{M}$, and $\|\theta - F(\theta)\| \leq \|\theta - \theta'\|$ for all $\theta' \in \mathcal{M}$),

$$\begin{aligned} r(t-1) &= r(t-2) + \|\phi(t-1)\|^2, \quad r(-1) = 1, \\ e(t) &:= z(t) - \phi^{\mathrm{T}}(t-1)\theta(t-1), \\ \phi(t-1) &:= [-y'(t-1), \ldots, -y'(t-n), s(t), \ldots, s(t-m)], \\ y'(t) &= \phi^{\mathrm{T}}(t-1)\theta'(t). \end{aligned} \tag{9}$$

The algorithm defined by (7-9) is slightly different from the algorithm (1) in that it employs a parameter estimate projection as well as "a posterior" estimates y' in the regression vector, in contrast to the use of a "a-priori" estimates $\hat{y}$ in (3).

For convenience, we define

$$\begin{aligned} \Delta(t) &:= \theta(t) - \theta(t-1), & \Delta'(t) &:= \theta'(t) - \theta(t-1), \\ \tilde{\theta}(t) &:= \theta(t) - \theta_0, & \tilde{\theta}'(t) &:= \theta(t) - \theta_0, \end{aligned}$$

$$v'(t) := z(t) - \phi^{\mathrm{T}}(t-1)\theta'(t). \tag{10}$$

To simplify the notation, we use K_i, C_i, $i = 1, 2, \ldots$, to denote generic finite positive numbers, related only to the bound Γ on $\{s(t), v(t), y(t)\}$, the radius of the compact set $\mathcal{M}$, K_c, α and the orders n and m. The symbol $\|\cdot\|$ will denote the Euclidean norm of vectors and the compatible norm for matrices.

The following lemma collects together some miscellaneous facts needed in the sequel.

Lemma 1.

i)

$$\|\Delta(t)\| \leq \|\Delta'(t)\| < C_1 < \infty, \qquad \forall t \geq 0, \quad \text{a.s.} \tag{11}$$

ii) For some number $K_\theta < \infty$, $\|\theta(t)\| \leq K_\theta, \quad \|\theta'(t)\| \leq K_\theta, \quad \forall t \geq 0, \quad$ a.s.

iii)

$$\|\phi(t)\| \leq (1 + K_\theta)\|\phi(t-1)\| + \Gamma, \quad \forall t \geq 0, \qquad \text{a.s.} \tag{12}$$

$$\|\phi(t)\|^2 \leq K_2\|\phi(t-1)\|^2 + 2\Gamma^2, \quad \forall t \geq 0, \quad \text{a.s., where } K_2 := 2(1 + K_\theta)^2. \tag{13}$$

iv) For each k, $\dfrac{r(t)}{r(t-k)}$ *is uniformly bounded in* t.

v)

$$\sum_{t=1}^{\infty} \frac{\|\phi(t)\|^2}{r(t-i)r(t-j)^{\delta}} < \infty \quad \text{a.s.}, \forall \delta > 0, \textit{and } i, j \textit{ finite.} \tag{14}$$

vi)

$$v'(t) = \frac{r(t-2)}{r(t-1)} e(t), \quad \forall t \geq 0. \tag{15}$$

vii)

$$A(q^{-1})[v'(t) - v(t) - \epsilon(t)] = -\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t), \quad \forall t \geq 0. \tag{16}$$

viii)

$$\sum_{t=1}^{\infty} \frac{\epsilon^2(t)}{t} < \infty \quad \text{a.s.} \tag{17}$$

Proof:

i) $\{\Delta'(t-1)\}$ is bounded since $\|\Delta'(t)\| = \frac{|z(t)-\phi^{\mathrm{T}}(t-1)\theta(t-1)|\ \|\phi(t-1)\|}{r(t-1)}$, and $z(t)$ and $\theta(t-1)$ are bounded while $\|\phi(t-1)\|^2 \leq r(t-1)$. Also, $\|\Delta(t)\| \leq \|\Delta'(t)\|$ because $\theta(t)$ is the projection of $\theta'(t)$ onto the convex set $\mathcal{M}$ and $\theta(t-1) \in \mathcal{M}$.

ii) The sequence $\{\theta(t)\}$ is bounded by definition, because of the projection onto the compact set $\mathcal{M}$. The boundedness of $\{\theta'(t)\}$ then follows from the boundedness of $\{\Delta'(t)\}$.

iii) By noting the "shift structure" of $\phi(t)$, we can write

$$\phi(t) = S\phi(t-1) - [1, 0, \ldots, 0]^{\mathrm{T}} \theta'^{\mathrm{T}}(t)\phi(t-1) + Ds(t+1), \tag{18}$$

where S is the appropriate "shift" matrix and $D := [0, \ldots, 0, 1, 0, \ldots, 0]^{\mathrm{T}}$. The claim then follows from the uniform boundedness of $\theta'(t)$ and $s(t)$.

iv) This follows easily from iii).

v) We have

$$\sum_{t=1}^{\infty} \frac{\|\phi(t)\|^2}{r(t-i)r(t-j)^{\delta}} = \sum_{t=1}^{\infty} \frac{\|\phi(t)\|^2}{r(t)r(t-1)^{\delta}}\ \frac{r(t)r(t-1)^{\delta}}{r(t-i)r(t-j)^{\delta}},$$

Since $\frac{r(t)r(t-1)^{\delta}}{r(t-i)r(t-j)^{\delta}}$ is bounded uniformly by (iv), the bound (14) then follows from the Pringsheim theorem ([13], page 326).

vi) This follows by multiplying both sides of (7) by $\phi^{\mathrm{T}}(t-1)$, and then subtracting them from $z(t)$.

vii) From (5) and (A4), we have

$$A(q^{-1})[y(t)-\epsilon(t)] = B(q^{-1})s(t).$$

Hence

$$\begin{aligned}
&A(q^{-1})[v'(t)-v(t)-\epsilon(t)] \\
&\quad = A(q^{-1})[z(t)-\phi^{\mathrm{T}}(t-1)\theta'(t)-v(t)-\epsilon(t)] \\
&\quad = B(q^{-1})s(t)-A(q^{-1})[\phi^{\mathrm{T}}(t-1)\theta'(t)] \\
&\quad = B(q^{-1})s(t)-(A(q^{-1})-1)[\phi^{\mathrm{T}}(t-1)\theta'(t)]-\phi^{\mathrm{T}}(t-1)\theta'(t) \\
&\quad = -\phi^{\mathrm{T}}(t-1)(\theta'(t)-\theta_0).
\end{aligned}$$

viii) Let $X(N) := \frac{1}{N}\sum_{t=0}^{N}\epsilon^2(t)$. From (A3),

$$X(t) = O\left(\frac{1}{\log^{1+\delta} t}\right).$$

Now,

$$\begin{aligned}
\sum_{t=1}^{N}\frac{\epsilon^2(t)}{t} &= \sum_{t=1}^{N}\frac{tX(t)-(t-1)X(t-1)}{t} \\
&= \sum_{t=1}^{N}(X(t)-X(t-1)) + \sum_{t=1}^{N}\frac{X(t-1)}{t} \\
&= X(N)-X(0)+\sum_{t=1}^{N}\frac{X(t-1)}{t} \\
&= O\left(\sum_{t=1}^{N}\frac{1}{t\log^{1+\delta} t}\right)+O(1) < \infty\,.
\end{aligned}$$

□

For convenience in the sequel, define

$$H(\theta) := S - [1, 0, \ldots, 0]^{\mathrm{T}}\theta \tag{19}$$

and rewrite (18) as

$$\phi(t-1) = H(\theta'(t-1))\phi(t-2) + Ds(t). \tag{20}$$

For ease of reference, the martingale limit theorem which we use is given below.

Lemma 2 [14]. *Let $w(t)$ be a bounded martingale difference sequence with respect to an increasing sequence of σ-algebras $\{\mathcal{F}_t\}$, and let f_t be a $\mathcal{F}_{t-1}$-measurable sequence. Then,* $\sum_{t=1}^{N} f_{t-1}w(t) = o\left(\sum_{t=1}^{N} f_{t-1}^2\right) + O(1)$ a.s. .

2.3 The Optimal Performance of SG Algorithm for Output Error Identification and Adaptive IIR Filtering

The following theorem establishes the self-optimizing property of the SG algorithm (7–9) in the long term average sense, as well as some other useful properties.

Theorem 1. *Let the assumptions (A1) to (A5) hold. Suppose that the following SPR condition holds,*

$$\mathrm{Re}[A(e^{j\omega})] > 0, \quad \forall\omega. \tag{21}$$

Then the algorithm (7–9) is self-optimizing, i.e.,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}[e(t) - v(t)]^2 = 0 \qquad \text{a.s.}. \tag{22}$$

Also,

$$\sum_{t=d}^{\infty} \|\theta(t) - \theta(t-1)\|^2 < \infty \quad \text{a.s.}, \tag{23}$$

$$r(N) \sim N \text{ a.s.}, \tag{24}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}[\phi^{\mathrm{T}}(t)\tilde{\theta}(t)]^2 = 0 \qquad \text{a.s.}, \tag{25}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}[v(t) - v'(t)]^2 = 0 \qquad \text{a.s.}. \tag{26}$$

Proof: Using (15), we can rewrite (7) as,

$$\tilde{\theta}'(t) - \frac{\phi(t-1)}{r(t-2)} v'(t) = \tilde{\theta}(t-1). \tag{27}$$

Taking squares,

$$\|\tilde{\theta}'(t)\|^2 + \frac{\|\phi(t-1)\|^2}{r(t-2)^2} v'(t)^2 - \frac{2\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)v'(t)}{r(t-2)} = \|\tilde{\theta}(t-1)\|^2.$$

Let $V(N) := \|\tilde{\theta}(N)\|^2$. Summing the above and using (5), we have

$$\begin{aligned} V(N) + \sum_{t=1}^{N} \|\Delta'(t)\|^2 + \sum_{t=1}^{N} \frac{2[A(q^{-1})(v'(t) - v(t) - \epsilon(t))][v'(t) - v(t)]}{r(t-2)} \\ \leq V(0) + \sum_{t=1}^{N} \frac{2\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)v(t)}{r(t-2)}. \end{aligned} \tag{28}$$

All subsequent analysis is sample pathwise, and inequalities hold almost surely.

From the assumption (21), and summation by parts, it is easy to show that for some finite positive numbers δ_1 and δ_2,

$$\sum_{t=1}^{N} \frac{[A(q^{-1})(v'(t) - v(t))][v'(t) - v(t)] - \delta_1[v'(t) - v(t)]^2 - \delta_2[\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)]^2}{r(t-2)} + O(1) \geq 0. \tag{29}$$

From (A5) and (17), we have

$$\begin{aligned} \left| \sum_{t=1}^{N} \frac{2[A(q^{-1})\epsilon(t)][v'(t) - v(t)]}{r(t-2)} \right| &\leq \frac{1}{\delta_1} \sum_{t=1}^{N} \frac{[A(q^{-1})\epsilon(t)]^2}{r(t-2)} + \delta_1 \sum_{t=1}^{N} \frac{[v'(t) - v(t)]^2}{r(t-2)} \\ &= \delta_1 \sum_{t=1}^{N} \frac{[v'(t) - v(t)]^2}{r(t-2)} + O(1)\,. \end{aligned} \tag{30}$$

Suppose that we have shown that $\forall \epsilon > 0$,

$$\left| \sum_{t=1}^{N} \frac{\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)v(t)}{r(t-2)} \right| \leq O(1) + o\left(\sum_{t=1}^{N} \left(\frac{\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)}{r(t-2)} \right)^2 \right) + \epsilon \sum_{t=1}^{N} \|\Delta'(t)\|^2. \tag{31}$$

Then by substituting (29), (30) and (31) into (28), we can obtain

$$(1-\epsilon)\sum_{t=1}^{N} \|\Delta'(t)\|^2 + \delta_1 \sum_{t=1}^{N} \frac{[v'(t) - v(t)]^2}{r(t-2)} + 2\delta_2 \sum_{t=1}^{N} \frac{[\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)]^2}{r(t-2)} = O(1).$$

Choosing $\epsilon < 1$, we can then conclude that

$$\sum_{t=d}^{\infty} \frac{[v'(t) - v(t)]^2}{r(t-2)} < \infty \quad \text{a.s.}, \tag{32}$$

$$\sum_{t=1}^{N} \frac{[\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)]^2}{r(t-2)} < \infty \quad \text{a.s.}, \tag{33}$$

and

$$\sum_{t=d}^{\infty} \|\Delta'(t)\|^2 < \infty \quad \text{a.s..} \tag{34}$$

Then, (23) follows from (34), (24) from (32), and (26) from (24), (32) and Kronecker's Lemma. Also,

$$\frac{[e(t)-v(t)]^2}{r(t-1)} = \frac{[v'(t)-v(t)+\phi^{\mathrm{T}}(t-1)\Delta'(t)]^2}{r(t-1)} \le \frac{2[v'(t)-v(t)]^2}{r(t-1)} + \frac{2\|\phi(t-1)\|^2 \ \|\Delta'(t)\|^2}{r(t-1)}.$$

So from (32) and (34), we have

$$\sum_{i=1}^{\infty} \frac{[e(t)-v(t)]^2}{r(t-1)} < \infty.$$

The self-optimality (22) then follows from the above , (24) and Kronecker's Lemma.

It now only remains to show (31). We decompose $v(t)$ as

$$v(t) = v_1(t) + v_2(t) \tag{35}$$

where

$$\begin{aligned} v_1(t) &= \begin{cases} \sum_{i=0}^{d(t)-1} c_i(t)w(t-i), & d(t) \ge 1, \\ 0\,, & d(t) = 0, \end{cases} \\ v_2(t) &= \sum_{i=d(t)}^{+t} c_i(t)w(t-i)\,, \end{aligned}$$

$d(t) := \max\{d \in Z^+ | \epsilon_d \log_\alpha r(t-d) \ge d\}$, and ϵ_d is a sufficiently small positive number such that some $\epsilon_1, \epsilon_2 > 0$ satisfy,

$$0 < \epsilon_d < \epsilon_d \log_\alpha K_2 < \epsilon_1 < 2\epsilon_1 < \epsilon_2 < \frac{1}{2}, \tag{36}$$

Because of the monotonicity of $\log_\alpha r(t)$, $d(t)$ is clearly uniquely defined and depends only on $(r(0), \ldots, r(t-d(t)))$, hence is $\mathcal{F}_{t-d(t)}$-measurable. Moreover, it can be easily verified that $d(t)$ satisfies:

i) $d(t) \uparrow \infty$;

ii) $0 \le d(t) \le t$;

iii)

$$\epsilon_d \log_\alpha r(t-d(t)-1) - 1 \le d(t) \le \epsilon_d \log_\alpha r(t-d(t)) \le \epsilon_d \log_\alpha r(t). \tag{37}$$

The basic idea for proving (31) is to bound $\left|\sum_{t=1}^N \frac{\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)v_2(t)}{r(t-2)}\right|$ by exploiting the vanishing magnitude of $|v_2(t)|$, and $\left|\sum_{t=1}^N \frac{\phi(t-1)\widetilde{\theta}'(t)v_1(t)}{r(t-2)}\right|$ by using a backward recursion developed in [15], as well as the martingale limit theorem given in Lemma 2.

From Lemma 1 (iii), we have

$$\|\phi(t-d(t)+k)\|^2 \le K_2^k \|\phi(t-d(t))\|^2 + C_2 K_2^k.$$

So,

$$\begin{aligned} r(t) - r(t-d(t)) &= \sum_{k=1}^{d(t)} \|\phi(t-d(t)-k)\|^2 \\ &\le \sum_{k=1}^{d(t)} (K_2^k \|\phi(t-d(t))\|^2 + C_2 K_2^k) \\ &\le d(t) K_2^{d(t)} (\|\phi(t-d(t))\|^2 + C_2) \\ &\le K_3 r(t)^{\epsilon_1} (\|\phi(t-d(t))\|^2 + C_2). \end{aligned} \tag{38}$$

Hence,

$$\frac{r(t)}{r(t-d(t))} \le 1 + \frac{r(t)-r(t-d(t))}{r(t-d(t))} \le 1 + K_3 r(t)^{\epsilon_1}(1+C_2) \le K_4 r(t)^{\epsilon_1},$$

So

$$r(t-d(t)) \ge \frac{1}{K_4} \cdot r(t)^{1-\epsilon_1} \tag{39}$$

From (A1), (37), (39), and Lemma 1 (iv), we have

$$|v_2(t)| \le \sum_{i=d(t)}^{t} K_c \Gamma \cdot \alpha^{-i} \le \frac{K_c \Gamma}{1-\alpha^{-1}} \alpha^{-d(t)} \le \frac{K_c \Gamma \alpha^{-1}}{1-\alpha^{-1}} r(t-d(t)-1)^{-\epsilon_d} \le K_5 r(t-2)^{-\epsilon_3},$$

where $\epsilon_3 := (1-\epsilon_1)\epsilon_d$. Hence,

$$\begin{aligned} \left|\sum_{t=1}^{\infty} \frac{\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)v_2(t)}{r(t-2)}\right| &\le \sum_{t=1}^{\infty} \frac{K_5 \|\phi(t-1)\|}{r(t-2)^{1+\epsilon_3}} \\ &\le \sum_{t=1}^{\infty} \frac{K_5 \|\phi(t-1)\|^2}{r(t-2)^{1+\epsilon_3}} + \sum_{t=1}^{\infty} \frac{K_5}{r(t-2)^{1+\epsilon_3}} < \infty, \end{aligned} \tag{40}$$

where the last inequality follows from Lemma 1 (v) and the assumption (A5).

We now work with the more difficult term $\frac{\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)v_1(t)}{r(t-2)}$. We decompose it as

$$\frac{\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)v_1(t)}{r(t-2)} = \frac{f^{\mathrm{T}}(t-1,t-d(t))\widetilde{\theta}'(t-d(t))v_1(t)}{r(t-d(t))} + [\zeta_1(t)+\zeta_2(t)+\zeta_3(t)]v_1(t), \quad (41)$$

where

$$\zeta_1(t) = \frac{\phi^{\mathrm{T}}(t-1)[\theta'(t)-\theta'(t-d(t))]}{r(t-2)},$$

$$\zeta_2(t) = \frac{[\phi(t-1)-f(t-1,t-d(t))]^{\mathrm{T}}\widetilde{\theta}'(t-d(t))}{r(t-2)},$$

$$\zeta_3(t) = \frac{f^{\mathrm{T}}(t-1,t-d(t))\widetilde{\theta}'(t-d(t))[r(t-2)-r(t-d(t))]}{r(t-2)r(t-d(t))},$$

and $f(t-1,t-d(t))$ is defined recursively by

$$f(t-1,t-d(t)) = H(\theta'(t-d(t)))f(t-2,t-d(t)) + Ds(t), \quad (42)$$

where $f(t-d(t),t-d(t)) := \phi(t-d(t))$. For simplicity, we will suppress the second argument of $f(\cdot,\cdot)$, i.e., write $f(t-k,t-d(t))$ as $f(t-k)$. Note that the vector $f(t-1)$ may be regarded as an approximation of $\phi(t-1)$ with the parameter estimate fixed at $\theta'(t-d(t))$ from time $t-d(t)$ to $t-1$.

For convenience, define $h(t) := \frac{f^{\mathrm{T}}(t-1)\widetilde{\theta}'(t-d(t))}{r(t-d(t))}$. By noting that $h(t)$ is $\mathcal{F}_{t-d(t)}$-measurable and using Lemma 2, we show next that

$$\sum_{t=1}^{N} h(t)v_1(t) = o\left(\sum_{t=1}^{N} h^2(t)\right) + O(1). \quad (43)$$

Let

$$\ell_i(t) := \begin{cases} c_i(t) & 0 \le i \le d(t)-1 \\ 0 & d(t) \le i \le t \end{cases}$$

So,

$$
\begin{aligned}
\sum_{t=0}^{N} h(t)v_1(t) &= \sum_{t=0}^{N} h(t) \sum_{i=0}^{d(t)-1} c_i(t)w(t-i) \\
&= \sum_{t=0}^{N}\sum_{i=0}^{t} \ell_i(t)h(t)w(t-i) \\
&= \sum_{t=0}^{N} w(t)\sum_{i=t}^{N} h(i)\ell_{i-t}(i) && \text{(changing the summation limits)} \\
&= o\left(\sum_{t=0}^{N}\left(\sum_{i=t}^{N} h(i)\ell_{i-t}(i)\right)^2\right) + O(1) && \text{(since } \sum_{i=t}^{N} h(i)\ell_{i-1}(i) \text{ is } \mathcal{F}_{t-1}\text{-measurable)} \\
&= o\left(\sum_{t=0}^{N}\left(\sum_{i=t}^{N} h^2(i)|\ell_{i-t}(i)|\right)\left(\sum_{i=t}^{N}|\ell_{i-t}(i)|\right)\right) + O(1) && \text{(from the Cauchy-Schwarz inequality)} \\
&= o\left(\sum_{t=0}^{N}\sum_{i=t}^{N} h^2(i)|\ell_{i-t}(i)|\right) + O(1) && \text{(since } \ell_i \text{ is geometrically decaying)} \\
&= o\left(\sum_{i=0}^{N} h^2(i)\sum_{t=0}^{i}|\ell_{i-t}(i)|\right) + O(1) \\
&= o\left(\sum_{i=0}^{N} h^2(i)\right) + O(1).
\end{aligned}
$$

From (43) and the decomposition (41), we have

$$
\begin{aligned}
\left|\sum_{t=1}^{N}\frac{\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)v_1(t)}{r(t-2)}\right| \quad \leq \quad & o\left(\sum_{t=1}^{N}\left(\frac{\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)}{r(t-2)}\right)^2\right) + o\left(\sum_{t=1}^{N}\left(\zeta_1^2(t)+\zeta_2^2(t)+\zeta_3^2(t)\right)\right) \\
& + K_6\sum_{t=1}^{N}\left(|\zeta_1(t)|+|\zeta_2(t)|+|\zeta_3(t)|\right) + O(1). \qquad (44)
\end{aligned}
$$

We will show that $|\zeta_i(t)|$, $i = 1, 2, 3$, are uniformly bounded. Then it will clearly follow that

$$
\sum_{t=1}^{N}\zeta_i^2(t) = O\left(\sum_{t=1}^{N}|\zeta_i(t)|\right).
$$

Hence, because of (40) and (44), to complete the proof of (31), it is enough to show that for $i = 1, 2, 3$ and $\forall\epsilon > 0$,

$$
\sum_{t=1}^{N}|\zeta_i(t)| \leq O(1) + \epsilon\sum_{t=1}^{N}\|\Delta'(t)\|^2. \qquad (45)
$$

It is clear that $|\zeta_1(t)|$ is uniformly bounded.

From (20) and (42), we have

$$
\phi(t-1) - f(t-1) = H(\theta'(t-d(t)))[\phi(t-2)-f(t-2)] - \phi^{\mathrm{T}}(t-2)[\theta'(t-1)-\theta'(t-d(t))]. \qquad (46)
$$

Noting that $\|H(\theta'(t))\| \leq 1 + K_\theta$, $\forall t \geq 0$, we have

$$\|\phi(t-1) - f(t-1)\| \leq \sum_{k=1}^{d(t)} (1+K_\theta)^k \|\phi(t-k-1)\| \; \|\theta'(t-k) - \theta'(t-d(t))\|. \tag{47}$$

So,

$$\frac{\|\phi(t-1) - f(t-1)\|}{r(t-2)} \leq 2K_\theta d(t)(1+K_\theta)^{d(t)} r(t-2)^{-\frac{1}{2}} \leq K_7 r(t)^{-\frac{1}{2}+\epsilon_1}.$$

Since $\epsilon_1 < \frac{1}{2}$, $|\zeta_2(t)|$ is uniformly bounded. From (38) and (47), we have

$$\begin{aligned} |\zeta_3(t)| &\leq \frac{K_\theta(\|\phi(t-1) - f(t-1)\| + \|\phi(t-1)\|)(r(t-2) - r(t-d(t)))}{r(t-2)r(t-d(t))} \\ &\leq K_{11} \frac{\|\phi(t-1) - f(t-1)\|}{r(t-2)^{1-\epsilon_1}} + \frac{K_\theta K_3 \|\phi(t-1)\| r(t)^{\epsilon_1}(1+C_2)}{r(t-2)}. \end{aligned}$$

Hence $|\zeta_3(t)|$ is also uniformly bounded. The following lemma will be needed for proving (45).

Lemma 3. *Let $x(t) \geq 0$, $\forall t \geq 0$. Then for $\forall \epsilon > 0$,*

$$\sum_{t=1}^{N} \sum_{k=0}^{d(t)} \frac{x(t-k)}{r(t)^\epsilon} \leq K_8 \sum_{t=1}^{N} x(t). \tag{48}$$

Proof: Let $X(t) := \sum_{k=0}^{t} x(k)$, and $\delta(k) := \min\{i \in Z^+ | d(i) = k\}$. Since $d(\delta(k)) = k$, it follows from (37) that $\epsilon_d \log_\alpha r(\delta(k)) \geq k$. Hence $r(\delta(k)) \geq \alpha^{\frac{k}{\epsilon_d}}$. So,

$$\begin{aligned} \sum_{t=1}^{N}\sum_{k=0}^{d(t)} \frac{x(t-k)}{r(t-2)^\epsilon} &= \sum_{k=0}^{d(N)} \sum_{t=\delta(k)}^{N} \frac{X(t-k) - X(t-k-1)}{r(t)^\epsilon} \\ &\leq \sum_{k=0}^{d(N)} r(\delta(k))^{-\epsilon} \sum_{t=\delta(k)}^{N} [X(t-k) - X(t-k-1)] \\ &\leq \sum_{k=0}^{d(N)} \alpha^{-\frac{\epsilon}{\epsilon_d}k} X(N-k) \leq X(N) \sum_{k=0}^{d(N)} \alpha^{-\frac{\epsilon}{\epsilon_d}k}, \end{aligned}$$

which yields (48). □

We return to show (45). For $i = 1$, we have $\forall \epsilon > 0$,

$$\begin{aligned} \sum_{t=1}^{N} |\zeta_1(t)| &\leq \frac{2}{\epsilon} \sum_{t=1}^{N} \frac{\|\phi(t-1)\|^2}{r(t-2)^{\frac{3}{2}}} + 2\epsilon \sum_{t=1}^{N} \frac{\|\theta'(t) - \theta'(t-d(t))\|^2}{r(t-2)^{\frac{1}{2}}} \\ &\leq O(1) + 2\epsilon \sum_{t=1}^{N} \frac{d(t) \sum_{k=0}^{d(t)} \|\Delta'(t-k)\|^2}{r(t-2)^{\frac{1}{2}}}. \end{aligned}$$

Hence (45) with $i = 1$ then follows from (48), (37) and Lemma 1 (iv). For $i = 2$, from (47), we have $\forall \epsilon > 0$.

$$\begin{aligned}
\sum_{t=1}^{N} |\zeta_2(t)| &\leq \sum_{t=1}^{N} \frac{K_\theta \sum_{k=1}^{d(t)} (1+K_\theta)^k \|\phi(t-k-1)\| \|\theta'(t-k) - \theta'(t-d(t))\|}{r(t-2)} \\
&\leq K_9 \sum_{t=1}^{d(t)} \frac{\sum_{t=1}^{N} \|\phi(t-k-1)\| \; \|\theta'(t-k) - \theta'(t-d(t))\|}{r(t)^{1-\epsilon_1}} \\
&\leq \frac{K_9}{\epsilon} \sum_{t=1}^{N} \frac{\sum_{k=0}^{d(t)} \|\phi(t-k-1)\|^2}{r(t)^{2-\epsilon_2-\epsilon_1} d^{-1}(t)} + K_9 \epsilon \sum_{t=1}^{N} \frac{\sum_{k=0}^{d(t)} \|\Delta'(t)\|^2}{r(t)^{\epsilon_2-\epsilon_1} d^{-1}(t)}.
\end{aligned}$$

Then (45) with $i = 2$ follows from Lemma 3 and Lemma 1 (v):

We now consider $\zeta_3(t)$. From (38) and (39), we have

$$\begin{aligned}
\sum_{t=1}^{N} |\zeta_3(t)| &\leq K_\theta \sum_{t=1}^{N} \frac{\|\phi(t-1) - f(t-1)\| [r(t-2) - r(t-d(t))]}{r(t-2) r(t-d(t))} \\
&\quad + K_\theta \sum_{t=1}^{N} \frac{\|\phi(t-1)\| [r(t-2) - r(t-d(t))]}{r(t-2) r(t-d(t))} \\
&\leq K_{10} \sum_{t=1}^{N} \frac{\|\phi(t-1) - f(t-1)\|}{r(t)^{1-\epsilon_1}} + K_{10} \sum_{t=1}^{N} \frac{\|\phi(t-1)\| \sum_{k=1}^{d(t)} \|\phi(t-k)\|^2}{r(t)^{2-\epsilon_1}}.
\end{aligned}$$

The first term on the RHS of the last inequality above can be bounded as for $\zeta_2(t)$, while the second term can be shown to be bounded by using Lemma 3 and Lemma 1 (iv). This completes the proof for (45), and hence (31).

□

Remark. Note that the controller $K_0(q^{-1})$ satisfying (A3) may be non-unique, either because its orders are overestimated, or because $s(t)$ is not sufficiently exciting. Thus, Theorem 1 only requires that there be *one* $K_0(q^{-1})$ satisfying (A3) whose denominator satisfies the SPR condition (21). This allows the relaxation of the SPR condition by overparameterization when $s(t)$ is sufficiently exciting. When $s(t)$ is the sum of finite number of sinusoids, there always exists an all-zero controller $K_0(q^{-1})$, with order twice the number of sinusoids in $s(t)$, which satisfies (A3), in which case the SPR condition (21) is always satisfied. In this connection we note that a "signal-dependent" SPR condition given in [16] is for local convergence only.

2.4 Least Squares Output Error Recursion

In this subsection we consider the least squares counterpart of the SG output error recursions. To prevent the condition number of the data covariance matrix from becoming unbounded, we consider the following modified least squares (MLS) algorithm incorporating a condition number bounding mechanism,

$$\theta'(t) = \theta(t-1) + P(t-1)\phi(t-1)[y(t) - \phi^{\mathrm{T}}(t-1)\theta(t-1)], \tag{49}$$

$$P^{-1}(t-1) = Q^{-1}(t-2) + \phi(t-1)\phi^{\mathrm{T}}(t-1), \tag{50}$$

$$r(t-1) = r(t-2) + \|\phi(t-1)\|^2, \tag{51}$$

$$Q(t-1) = \begin{cases} P(t-1), & r(t-1)\lambda_{\max}(P(t-1)) \le K, \\ \dfrac{r(t-2)}{r(t-1)}Q(t-2), & \text{otherwise,} \end{cases}$$

$$\theta(t) = F_t[\theta'(t)], \tag{52}$$

where $K \in [n+m+1, \infty)$, and $F_t[\cdot]$ represents a projection onto a compact convex set $\mathcal{M}_t$ under the weighted Euclidean norm $\|x\|_t^2 := x^{\mathrm{T}}Q^{-1}(t-1)x$. For details of the projection, see page 91 of [13]. The algorithm is initialized so as to satisfy

$$r(0) = tr(Q^{-1}(0)), \quad \text{and} \quad r(0)\lambda_{\max}(Q(0)) \le K\,. \tag{53}$$

Remarks: i) While the condition number bounding mechanism is introduced here mainly for analytical tractability, it may have the practical advantage of preventing numerical ill-conditioning and improving the robustness of the algorithm with respect to unmodelled effects. ii) When the condition number bounding constant $K = dim(P(t))$, the above MLS algorithm degenerates to the SG algorithm.

Theorem 2. *Let all the assumptions of Theorem 1 hold, except that the SPR condition (21) is replaced by the following stronger version,*

$$\mathrm{Re}[A^{-1}(e^{j\omega}) - \frac{1}{2} + \frac{1}{2K}] > 0, \quad \forall\omega. \tag{54}$$

Then all the conclusions of Theorem 1 hold for the MLS algorithm.

Proof: See the Appendix.

Remark: It is interesting to note that as the condition number bounding constant K gets larger, while the algorithm can converge "potentially" faster, the SPR condition (54) becomes more stringent.

3 CONVERGENCE TO THE TRUE PARAMETER IN OUTPUT ERROR IDENTIFICATION

In this section we establish the strong consistency of the parameter estimates of the output error recursion with projection, when the input $s(t)$ satisfies certain persistency of excitation conditions.

Theorem 3. *In addition to the assumptions of Theorem 1, suppose that $A(q^{-1})$ and $B(q^{-1})$ do not have a common factor, and that θ_0 is an interior point of the set $\mathcal{M}$. If $\{s(t)\}$ satisfies one of the following two conditions,*

i) $\{s(t)\}$ is sufficiently rich of order greater than or equal to $n+m+1$, i.e., $\exists$ an $\epsilon > 0$, an integer $T < \infty$, such that

$$\sum_{i=t}^{t+T} [s(t), \ldots, s(t-n-m)][s(t), \ldots, s(t-n-m)]^{\mathrm{T}} \geq \epsilon I_{n+m+1} \,,$$

for all t sufficiently large; or

ii) $s(t)$ is an ARMA process, i.e., $s(t) = C(q^{-1})\tau(t)$, where $C(q^{-1})$ is a stable rational transfer function, and $\tau(t)$ is a martingale difference sequence with finite but nonzero condition variance,

then, $\theta(t) \to \theta_0$ a.s..

Proof: Multiplying (7) by $R(t-1) := I + \sum_{i=1}^{t-1} \phi(i)\phi^{\mathrm{T}}(i)$, we have

$$\begin{aligned} R(t-1)\tilde{\theta}(t) &= R(t-2)\tilde{\theta}(t-1) + \phi(t-1)\phi^{\mathrm{T}}(t-1)\tilde{\theta}(t-1) \\ &\quad + R(t-1)\frac{\phi(t-1)[y(t) - \phi^{\mathrm{T}}(t-1)\theta(t-1)]}{r(t-1)} + R(t-1)\left[\theta'(t) - \theta(t)\right]. \end{aligned}$$

Summing from 1 to N and dividing by N, we obtain

$$\begin{aligned}
&\frac{1}{N}\left\|R(N-1)\widetilde{\theta}(N)\right\| \\
&\le \frac{1}{N}\left\|\sum_{t=1}^{N}\phi(t-1)\phi(t-1)^{\mathrm{T}}\widetilde{\theta}(t-1)\right\| + \frac{1}{N}\left\|\sum_{t=1}^{N}\frac{R(t-1)}{r(t-1)}\phi(t-1)[e(t)-v(t)]\right\| \\
&\quad + \frac{1}{N}\left\|\sum_{t=1}^{N}\frac{R(t-1)}{r(t-1)}\phi(t-1)v(t)\right\| + \frac{1}{N}\sum_{t=1}^{N}\|R(t-1)\|\ \|\theta'(t)-\theta(t)\| + o(1)\,.
\end{aligned}$$

The first two terms on the RHS above converge to zero by the Schwarz inequality, (25) and the self-optimality (22). The third term can be shown to converge to zero by the techniques in the proof of Theorem 1, i.e., the decomposition of $v(t)$, the backward recursion for $\phi(t-1)$, and Lemma 2.

We turn next to the fourth term on the RHS above. Let $\overline{I}_D(x)$ denote the indicator function of the complement of a set D. Clearly

$$\|\theta'(t)-\theta(t)\| \le \overline{I}_{\mathcal{M}}(\theta'(t))\|\Delta'(t)\|\,.$$

Hence,

$$\begin{aligned}
&\frac{1}{N}\sum_{t=1}^{N}\|R(t-1)\|\ \|\theta'(t)-\theta(t)\| \\
&\le \frac{1}{N}\sum_{t=1}^{N}\|\phi(t-1)e(t)\|\,\overline{I}_{\mathcal{M}}(\theta'(t)) \\
&\le \frac{1}{N}\sum_{t=1}^{N}\|\phi(t-1)[e(t)-v(t)]\| + \frac{1}{N}\sum_{t=1}^{N}\|\phi(t-1)v(t)\|\,\overline{I}_{\mathcal{M}}(\theta'(t)) \\
&\le o(1) + \Gamma\,\frac{1}{N}\left(\sum_{t=1}^{N}\|\phi(t-1)\|^2\right)^{\frac{1}{2}}\left(\sum_{t=1}^{N}\overline{I}_{\mathcal{M}}(\theta'(t))\right)^{\frac{1}{2}}.
\end{aligned}$$

From either condition i) or ii) of the theorem, by using the techniques of [17, 18], one can establish, based on (23) and (25), that $\theta(t)$ converges to θ_0 in the Cesaro sense, i.e., for any open set D containing θ_0, $\frac{1}{N}\sum_{t=1}^{N}\overline{I}_D(\theta(t)) \to 0$ a.s.. Since θ_0 is an interior point of $\mathcal{M}$, and $\|\theta'(t)-\theta(t)\| \to 0$, it follows that $\frac{1}{N}\sum_{t=1}^{N}\overline{I}_{\mathcal{M}}(\theta'(t)) \to 0$ a.s.. Therefore, $\frac{1}{N}\|R(N-1)\widetilde{\theta}(N)\| \to 0$ a.s.. From either i) or ii), and (26), it can be shown that $\lim\inf_{N\to\infty}\frac{1}{N}\lambda_{\min}(R(N)) > 0$ a.s..

Hence, $\|\tilde{\theta}(t)\| \to 0$ a.s.. □

The extension of the above consistency result to the MLS algorithm is straightforward, and is hence omitted.

Remark: When the parameter estimate converges almost surely to the true parameter, as established in Theorem 3 under certain persistency of excitation condition, the long term average optimality (22) can be strengthened to

$$e(t) - v(t) \to 0 \quad a.s..$$

4 ADAPTIVE FEEDFORWARD CONTROL

We now turn to Problems B and C depicted in Figs. 2 and 3. Clearly,

$$e(t) = G(q^{-1})s(t) - P(q^{-1})u(t) + v(t)\,, \tag{55}$$

and

$$m(t) = H(q^{-1})s(t) + F(q^{-1})u(t)\,, \tag{56}$$

For simplicity we first consider the case when the following *strong matching condition* holds.

Assumption.

A6) $G(q^{-1}), P(q^{-1}), H(q^{-1})$ and $F(q^{-1})$ are stable rational transfer functions, and $P(q^{-1})$ is minimum phase. The non-minimum phase zeros of H are also the zeros of G, and the number of pure delays of HP is less than or equal to that of G.

The model (55, 56) can be equivalently represented by

$$A(q^{-1})[e(t) - v(t)] = q^{-d}B(q^{-1})u(t) + q^{-k}D(q^{-1})m(t)\,, \tag{57}$$

where $A(q^{-1}), B(q^{-1})$ and $D(q^{-1})$ are polynomials. From (A6), we have that $d \leq k$, and that $A(q^{-1})$ is Hurwitz since its poles consist of the poles of $G(q^{-1})H^{-1}(q^{-1})$ and $P(q^{-1})$.

With the model parameterized as in (57), it is clear that the minimum variance feedforward control is given by

$$B(q^{-1})u(t) + q^{-(k-d)}D(q^{-1})m(t) = 0\,.$$

Let us now consider the adaptation procedure. Let $B(q^{-1}) = b_0 + \cdots + b_\ell q^{-\ell}, D(q^{-1}) = d_0 + \cdots + d_n q^{-n}, \theta_0 = [b_0, \ldots, b_\ell, d_0, \ldots, d_n]$, and

$$\phi(t) := [u(t), \ldots, u(t-\ell-1), m(t-k+d), \ldots, m(t-k+d-n-1)]^{\mathrm{T}}.$$

Then

$$A(q^{-1})[e(t) - v(t)] = \phi^{\mathrm{T}}(t-d)\theta_0\,.$$

We propose the following algorithm for adaptive noise cancelling as well as adaptive feedforward control. The control $u(t)$ is generated so as to satisfy

$$\phi^{\mathrm{T}}(t)\theta(t) = 0\,,$$

and $\theta(t)$, the estimate of θ_0, is updated by

$$\begin{aligned} \theta(t) &= F\left(\theta(t-1) + \frac{a\phi(t-d)}{r(t-d)}e(t)\right), \quad a > 0 \\ r(t-d) &= r(t-d-1) + \|\phi(t-d)\|^2, \quad r(-d) = 1\,, \end{aligned}$$

where $F(\cdot)$ denotes the projection onto a convex compact set $\mathcal{M}$. The following theorem establishes the optimality of the above adaptive control law.

Theorem 4. *Let the following conditions hold:*

i) $\{v(t)\}$ and $\{s(t)\}$ satisfy the assumptions (A1), (A2) and (A5).

ii) Assumption (A6) holds.

iii)

$$\mathrm{Re}[A(e^{j\omega})] > \left(d - \frac{1}{2}\right)a, \quad \forall\omega\,. \tag{58}$$

iv) $\theta_0 \in \mathcal{M}$, *and* $\widehat{b}_0 \neq 0$, $\forall \widehat{\theta} \in \mathcal{M}$.

Then

$$\lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}[e(t)-v(t)]^2 = 0 \quad \text{a.s.}. \tag{59}$$

Proof: Let us denote $V(t) := \|\widetilde{\theta}(t)\|^2$, and $\Delta(t) := \frac{\|a\phi(t-d)e(t)\|}{r(t-d)}$. Then $V(t)$ satisfies the recursion

$$\begin{aligned} V(N) &\leq V(0) + \sum_{t=1}^{N}\Delta^2(t) + 2a\sum_{t=1}^{N}\frac{\phi^{\mathrm{T}}(t-d)\widetilde{\theta}(t-d)e(t)}{r(t-d)} \\ &\quad + 2a\sum_{t=1}^{N}\frac{\phi^{\mathrm{T}}(t-d)[\widetilde{\theta}(t-1)-\widetilde{\theta}(t-d)]e(t)}{r(t-d)}. \end{aligned} \tag{60}$$

The last term on the RHS above can be bounded as follows,

$$\begin{aligned} 2a\sum_{t=1}^{N}\frac{\phi^{\mathrm{T}}(t-d)[\widetilde{\theta}(t-1)-\widetilde{\theta}(t-d)]e(t)}{r(t-d)} &\leq 2\sum_{t=1}^{N}\Delta(t)\sum_{i=1}^{d-1}\Delta(t-i) \\ &\leq \sum_{t=1}^{N}\sum_{i=1}^{d-1}[\Delta^2(t)+\Delta^2(t-i)] \\ &\leq 2(d-1)\sum_{t=1}^{N}\Delta^2(t) + O(1). \end{aligned} \tag{61}$$

As in the proof of Theorem 1, we can show that

$$\begin{aligned} &\left|\sum_{t=1}^{N}\frac{\phi^{\mathrm{T}}(t-d)\widetilde{\theta}(t-d)v(t)}{r(t-d)}\right| \\ &\leq O(1) + o\left(\sum_{t=1}^{N}\left(\frac{\phi^{\mathrm{T}}(t-d)\widetilde{\theta}(t-d)}{r(t-d)}\right)^2\right) + \epsilon\sum_{t=1}^{N}\Delta^2(t), \end{aligned} \tag{62}$$

$\forall \epsilon > 0$. Substituting (61) and (62) into (60), we obtain

$$\begin{aligned} V(N) &\leq (2d-1+\epsilon)\sum_{t=1}^{N}\Delta^2(t) + 2a\sum_{t=1}^{N}\frac{\phi^{\mathrm{T}}(t-d)\widetilde{\theta}(t-d)[e(t)-v(t)]}{r(t-d)} \\ &\quad + o\left(\sum_{t=1}^{N}\left(\frac{\phi^{\mathrm{T}}(t-d)\widetilde{\theta}(t-d)}{r(t-d)}\right)^2\right) + O(1). \end{aligned} \tag{63}$$

Now,

$$\begin{aligned} \sum_{t=1}^{N}\Delta^2(t) &= a^2\sum_{t=1}^{N}\frac{\|\phi(t-d)\|^2[[e(t)-v(t)]^2+v^2(t)+2e(t)v(t)]}{r^2(t-d)} \\ &\leq a^2\sum_{t=1}^{N}\frac{[e(t)-v(t)]^2}{r(t-d)} + O(1). \end{aligned}$$

Substituting the above into (63), noting that

$$A(q^{-1})[e(t)-v(t)] = -\phi^{\mathrm{T}}(t-d)\widetilde{\theta}(t-d)\,,$$

and using (58), we obtain

$$\sum_{t=1}^{\infty} \frac{[e(t)-v(t)]^2}{r(t-1)} < \infty. \quad \text{a.s.}.$$

It follows from Kronecker's Lemma that

$$\lim_{N\to\infty} \frac{1}{r(N-1)} \sum_{t=1}^{N} [e(t)-v(t)]^2 = 0 \qquad \text{a.s.}. \tag{64}$$

Finally, since $F(q^{-1})$ is stable and $P(q^{-1})$ is minimum phase, we have

$$\begin{aligned} r(N-1) &= O\left(\textstyle\sum_{t=1}^{N}[m^2(t)+u^2(t)]\right) &&= O\left(\textstyle\sum_{t=1}^{N} u^2(t)\right) + O(N) \\ &= O\left(N + \textstyle\sum_{t=1}^{N}[e(t)-v(t)]^2\right) &&= O(N) + o(r(N)) \end{aligned}$$

Hence, $r(N) \sim N$. The optimality of the adaptive scheme, (59), then follows from (64). □

Remarks: i) It is perhaps somewhat surprising that the presence of the input contamination $F(q^{-1})$ *does not* require any algorithmic modification. ii) Unlike the output error recursion for filtering or identification, it is *not* possible to use a posteriori estimates in the regression vector for adaptive feedforward control. iii) It can be shown by an ODE analysis that the weaker SPR condition (21), instead of (58), is sufficient for local convergence .

We now consider the case when the following *weak matching assumption* holds.

Assumption.

A6') There exists a controller $K_0(q^{-1}) = -\frac{D_0(q^{-1})}{B_0(q^{-1})}$, *where* $D_0(q^{-1})$ *and* $B_0(q^{-1})$ *are polynomials, such that*

i) $(1-K_0F)^{-1}$ *is stable,*

ii) With $u(t) = K_0(q^{-1})m(t)$, $\epsilon(t) := e(t)-v(t)$ *satisfies*

$$\frac{1}{N}\sum_{t=1}^{N} \epsilon^2(t) = O\left(\frac{1}{\log^{1+\delta} N}\right), \ \textit{for some } \delta > 0. \tag{65}$$

Further assume that $P(q^{-1})$ is minimum phase.

Assuming (A6'), it can be shown, after some algebraic manipulations, that

$$A_0(q^{-1})[e(t) - v(t) - \epsilon(t)] = B_0(q^{-1})u(t-d) + D_0(q^{-1})m(t-d)\,, \tag{66}$$

where d is the pure delay of $P(q^{-1})$, and

$$A_0(q^{-1}) := [B_0(q^{-1}) - D_0(q^{-1})F(q^{-1})]P^{-1}(q^{-1})q^d\,.$$

Then it is clear that Theorem 4 continues to hold if (A6) is replaced by (A6'), and if the SPR condition (58) is imposed on $A_0(q^{-1})$, instead of $A(q^{-1})$.

5 CONCLUSIONS

Stochastic parallel model adaptation problems which have applications in identification, adaptive filtering, adaptive feedforward control, and adaptive active noise and vibration control, have been considered in this paper. We have given a relatively complete solution to these problems in the ideal case when the system is "matchable." As a special case of our results, the long standing open problem of the convergence of the output error identification scheme in the presence of colored noise is resolved.

We have employed two modifications to the (standard) SG and LS algorithms: parameter estimate "projection", and condition number bounding for the data covariance matrix. These modifications are known to enhance robustness and to improve numerical conditioning. However, it is of interest to see if such modifications are really necessary for convergence in the ideal case considered here. It is also of interest to extend the analysis to the case where the noise coefficients $\{c_i(t)\}$ are not geometrically decaying in i.

The case when the matching condition does not hold is important and requires further investigation.

A APPENDIX

Proof of Theorem 2: From (53), it can be shown by induction that $\forall t \geq 1$,

$$\begin{aligned} trace(P^{-1}(t)) &= trace(Q^{-1}(t)) = r(t), \quad (67) \\ \frac{\lambda_{\min}(Q^{-1}(t))}{r(t)} &\geq \frac{1}{K}. \quad (68) \end{aligned}$$

The parameter update recursion (49) can be again expressed in terms of a posteriori error $v'(t)$, defined in (10), as follows,

$$\theta'(t) = \theta(t-1) + Q(t-2)\phi(t-1)v'(t). \quad (69)$$

It is then easy to show that

$$\begin{aligned} &\tilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-2)\tilde{\theta}'(t) + \phi^{\mathrm{T}}(t-1)Q(t-2)\phi(t-1)v'^2(t) - 2\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)v'(t) \\ &\qquad = \tilde{\theta}(t-1)Q(t-2)^{-1}\tilde{\theta}(t-1). \quad (70) \end{aligned}$$

Define the "stochastic Lyapunov function"

$$V(t) := \frac{\tilde{\theta}^{\mathrm{T}}(t)Q^{-1}(t-1)\tilde{\theta}(t)}{r(t-1)}.$$

Dividing (70) by $r(t-2)$ and using the above definition, we obtain

$$\begin{aligned} \frac{\tilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-2)\tilde{\theta}'(t)}{r(t-2)} &- \frac{2\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)\,[v'(t)-v(t)]}{r(t-2)} + \frac{\phi^{\mathrm{T}}(t-1)Q(t-2)\phi(t-1)v'^2(t)}{r(t-2)} \\ &\leq V(t-1) + \frac{2\phi^{\mathrm{T}}(t-1)\tilde{\theta}'(t)v(t)}{r(t-2)} \quad (71) \end{aligned}$$

The projection (52) has the property that,

$$\tilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-1)\tilde{\theta}'(t) \geq \tilde{\theta}(t)Q(t-1)^{-1}\tilde{\theta}(t).$$

Hence, when $r(t-1)\lambda_{\max}(P(t-1)) > K$, we have

$$\frac{\tilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-2)\tilde{\theta}'(t)}{r(t-2)} = \frac{\tilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-1)\tilde{\theta}'(t)}{r(t-1)} \geq V(t). \quad (72)$$

When $r(t-1)\lambda_{\max}(P(t-1)) \leq K$, we have, using (67, 68),

$$\begin{aligned} & \frac{\widetilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-2)\widetilde{\theta}'(t)}{r(t-2)} \\ = & \quad \frac{\widetilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-1)\widetilde{\theta}'(t)}{r(t-1)} + \frac{\widetilde{\theta}'^{\mathrm{T}}(t)Q^{-1}(t-1)\widetilde{\theta}'(t)\|\phi(t-1)\|^2}{r(t-1)r(t-2)} - \frac{[\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)]^2}{r(t-2)} \\ \geq & \quad V(t) + \left(\frac{1}{K} - 1\right)\frac{[\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)]^2}{r(t-2)}. \end{aligned} \tag{73}$$

Substituting (72) or (73) into (71) yields,

$$\begin{aligned} V(t) & - \frac{2\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)\left[\left(\frac{1}{2} - \frac{1}{2K}\right)\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t) + v'(t) - v(t)\right]}{r(t-2)} \\ & + \frac{\phi^{\mathrm{T}}(t-1)Q(t-2)\phi(t-1)v'^2(t)}{r(t-2)} \leq V(t-1) + \frac{2\phi^{\mathrm{T}}(t-1)\widetilde{\theta}'(t)v(t)}{r(t-2)}. \end{aligned}$$

Since the above equation is similiar to (28), the rest of the proof follows that of Theorem 1. □

References

[1] W. Ren and P. R. Kumar, "Adaptive active noise control: Structures, algorithms and convergence analysis," in *Proceedings of the Inter-Noise 89*, (Newport Beach, CA), pp. 435–440, December 4-6 1989.

[2] "Special Issues on Adaptive Filtering." *IEEE Trans. on Information Theory*, vol. 34,, March 1984.

[3] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Transactions on Automatic Control*, vol. AC-21, pp. 194–202, 1976.

[4] I. D. Landau, "Addendum to unbiased recursive identification using model reference adaptive techniques," *IEEE Transactions to Automatic Control*, vol. AC-23, pp. 97–99, 1978.

[5] C. R. Johnson, Jr., "A stable family of adaptive IIR filters," in *Proceedings of the 1980 IEEE International Conference on ASSP*, pp. 1001–1004, 1980.

[6] C. R. Johnson, Jr., "Adaptive IIR filtering, Current results and open issues," *IEEE Transactions on Information Theory*, vol. IT-30, pp. 237–250, 1984.

[7] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Transactions on Automatic Control*, vol. AC-22, pp. 551–575, 1977.

[8] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Transactions on Automatic Control*, vol. AC-22, pp. 539–551, 1977.

[9] L. Dugard and L. D. Landau, "Recursive output error identification algorithms: Theory and evaluation," *Automatica*, vol. 16, pp. 443–462, 1980.

[10] V. Solo, "The convergence of AML," *IEEE Transactions on Automatic Control*, vol. AC-24, pp. 958–962, 1979.

[11] B. E. Ydstie, "Stability of discrete MRAC-revisited," *Systems and Control Letters*, vol. 13, pp. 429–439, 1989.

[12] W. Ren and P. R. Kumar, "Direct stochastic minimum variance control with non-interlaced algorithms." To appear in *Proceedings of the 1991 American Control Conference*, 1990.

[13] G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control.* Englewood Cliffs, NJ: Prentice-Hall, 1984.

[14] T. L. Lai and C. Z. Wei, "Least squares estimate in stochastic regression with applications to identification and control of dynamic systems," *Annals of Mathematical Statistics*, vol. 10, pp. 154–166, 1982.

[15] W. Ren and P. R. Kumar, "The convergence of output error identification and adaptive IIR filtering algorithms in the presence of colored noise," in *Proceedings of the 29th Conference on Decision and Control*, pp. 3534–3539, December 5-7 1990. Honolulu, HI.

[16] B. Riedle, L. Praly, and P. Kokotović, "Examination of the SPR condition in output error parameter estimation," *Automatica*, vol. 22, no. 4, pp. 495–498, 1986.

[17] A. Becker, P. R. Kumar, and C. Z. Wei, "Adaptive control with the stochastic approximation algorithm: Geometry and convergence," *IEEE Transactions on Automatic Control*, vol. AC-30, pp. 330–338, April 1985.

[18] P. R. Kumar and L. Praly, "Self-tuning trackers," *SIAM Journal on Control and Optimization*, vol. 25, pp. 1053–1071, July 1987.

Contents