10. Conclusion

We wish to re-emphasize the motivation that backs our promotion of chattering systems. These systems model instantaneous oscillations, a useful tool when encountering systems where control is by some sort of modulation of a highly oscillating input. The applicability is even more prominent when, at the design stage, errors and uncertainties of the oscillations are expected. Then we would like to be able to do the analysis with a robust model. We observe that the relevant data are the time densities of the control coefficients, and the limit case is modeled by the chattering systems. We identified a convergence mode for the chattering systems, with respect to which the performance of the systems and the structure of the control policies are robust. Thus, the solutions in the chattering model can be used in the highly oscillating case. Since the chattering systems can be analyzed in a rather straightforward way, as is shown in the paper with both abstract and concrete examples, we obtain a convenient tool for the analysis of rapidly oscillating systems.

References

[AM] B. D. O. Anderson and J. B. Moore, *Linear Optimal Control*, Prentice-Hall, Englewood Cliffs, NJ, 1971.
[A1] Z. Artstein, Discrete and continuous bang-bang and facial spaces, or: Look for the extreme points, *SIAM Rev.*, **22** (1980), 172–185.
[A2] Z. Artstein, Stability, observability and invariance, *J. Differential Equations*, **44** (1982), 224–248.
[A3] Z. Artstein, Uniform controllability via the limiting systems, *Applied Math. Optim.*, **9** (1982), 111–131.
[A4] Z. Artstein, A variational convergence that yields chattering systems, *Ann. Inst. Henri Poincaré Anal. Non Linéare* (to appear).
[A5] H. Attouch, *Variational Convergence for Functions and Operators*, Applicable Mathematics Series, Pitman, London, 1984.
[B1] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
[B2] R. W. Brockett, *Finite-Dimensional Linear Systems*, Wiley, New York, 1970.
[CV] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Mathematics, vol. 580, Springer-Verlag, Berlin, 1977.
[K1] R. E. Kalman, Contributions to the theory of optimal control, *Bol. Soc. Mat. Mexicana*, **5** (1960), 102–119.
[K2] B. C. Kuo, *Automatic Control Systems*, 2nd edn., Prentice-Hall, Englewood Cliffs, NJ, 1967.
[KV] J. Kurzweil and Z. Vorel, Continuous dependence of solutions of differential equations on parameters, *Czechoslovak Math. J.*, **7** (1957), 568–583.
[KP] W. H. Kwon and A. E. Pearson, A modified quadratic cost problem and feedback stabilization of linear systems, *IEEE Trans. Automat. Control*, **22** (1977), 838–842.
[LM] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, Wiley, New York, 1967.
[R] D. Russel, *Mathematics of Finite-Dimensional Control Systems*, Marcel-Dekker, New York, 1979.
[T] L. Tartar, Compensated compactness and applications to partial differential equations, in *Nonlinear Analysis and Mechanics* (R. J. Knap, ed.), pp. 136–211, Heriot-Watt Symposium, Vol. 4, Pitman, London, 1975.
[W] J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972
[Y1] T. Yoshizawa, *Stability Theory and the Existence of Periodic Solutions and Almost Periodic Solutions*, Springer-Verlag, New York, 1975.
[Y2] L. C. Young, *Calculus of Variations and Optimal Control Theory*, Saunders, Philadelphia, PA, 1969.

Math. Control Signals Systems (1989) 2: 381–403

Mathematics of Control, Signals, and Systems

Adaptive Algorithms with Filtered Regressor and Filtered Error*

W. A. Sethares,† B. D. O. Anderson,‡ and C. R. Johnson, Jr.§

Abstract. This paper presents a unified framework for the analysis of several discrete time adaptive parameter estimation algorithms, including RML with nonvanishing stepsize, several ARMAX identifiers, the Landau-style output error algorithms, and certain others for which no stability proof has yet appeared. A general algorithmic form is defined, incorporating a linear time-varying regressor filter and a linear time-varying error filter. Local convergence of the parameters in nonideal (or noisy) environments is shown via averaging theory under suitable assumptions of persistence of excitation, small stepsize, and passivity. The excitation conditions can often be transferred to conditions on external signals, and a small stepsize is appropriate in a wide range of applications. The required passivity is demonstrated for several special cases of the general algorithm.

Key words. Adaptive estimation, Convergence, Averaging, Passivity.

1. Introduction

The LMS (least mean square) adaptive algorithm has been studied extensively over the past several years, and its convergence and stability properties are well known [B], [WMLJ]. LMS can be viewed as an algorithm for identifying the parameters of an unknown linear system when only its inputs and outputs can be measured. An error signal, which is equal (in the ideal case) to the inner product of the regressor and the parameter error, is used to drive the LMS parameter updates. In many filtering, identification, and control applications, the measured error signal is a *filtered* version of this inner product, plus some small nonidealities [L1], [J2]. It is natural to attempt to compensate for this filtering in order to regain the desirable stability properties of the LMS algorithm.

This paper examines two such methods of compensation: filtering of the error,

* Date received: February 17, 1987. Date revised: May 20, 1988. The first and third authors were supported by NSF Grants ECS-8506149, INT-8513400, and MIP-8608787.
† Research done while at the School of Electrical Engineering, Cornell University, Ithaca, New York 14853, U.S.A. Currently with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.
‡ Department of Systems Engineering, Australian National University, Canberra, ACT 2601, Australia.
§ School of Electrical Engineering, Cornell University, Ithaca, New York 14853, U.S.A.

and filtering of the regressor. These lead to the generic parameter update form

$$\left\{\begin{matrix}\text{new}\\ \text{parameter}\\ \text{estimate}\end{matrix}\right\} = \left\{\begin{matrix}\text{old}\\ \text{parameter}\\ \text{estimate}\end{matrix}\right\} + \left\{\text{stepsize}\right\}\left\{\begin{matrix}\text{filtered}\\ \text{version of}\\ \text{regressor}\end{matrix}\right\}\left\{\begin{matrix}\text{filtered}\\ \text{version of}\\ \text{error}\end{matrix}\right\}$$

in which the filters represent linear, possibly time-varying, rational operators. Many such algorithms have been proposed, and several have been satisfactorily analyzed (see [L5] and Table 1 for references). This paper presents a unified approach that proves the local stability of this entire class of algorithms in nonideal (or noisy) situations under appropriate persistence of excitation, passivity, and small stepsize assumptions.

Motivation

Consider a linear plant parametrized by an unknown constant vector θ^* which maps a bounded scalar input u_k to a scalar output $y_k = X_k^T\theta^*$, where $X_k^T = (u_k, \ldots, u_{k-n+1})$ is the regressor vector. Let $\hat{\theta}_k$ be an estimate of θ^*, and form the estimated output $\hat{y}_k = X_k^T\hat{\theta}_k$. The error $e_k = y_k - \hat{y}_k$ between the measured output and the estimated output can be used to improve the estimates of the parameter vector. Using gradient descent techniques [WMLJ] or L_2 minimization ideas [L6] leads to the parameter update scheme

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu X_k e_k, \tag{1.1}$$

where μ is a small positive stepsize. This is called the LMS adaptive algorithm [WMLJ] or the equation error algorithm [M] depending on the exact structure of θ^* and the specific problem for which it is utilized. For purposes of analysis, it is more convenient to work with the parameter error update

$$\theta_{k+1} = \theta_k - \mu X_k\{X_k^T\theta_k\}, \tag{1.2}$$

where $\theta_k = \theta^* - \hat{\theta}_k$. The error system (1.2) is Lyapunov stable [M], and the equilibrium $\theta = 0$ is exponentially asymptotically stable (EAS) if X_k is persistently spanning [B], that is, if there is a finite time window m such that for every j, $\sum_{k=j}^{j+m} X_k X_k^T$ is uniformly positive definite. The exponential asymptotic stability of (1.2) implies that $\hat{\theta}$ of (1.1) is exponentially convergent to θ^*.

In actual implementation (the "nonideal" case) the output y_k may be corrupted by disturbances such as measurement noise or unmodeled dynamics. The prediction error is then

$$e_k = y_k - \hat{y}_k + \eta_k, \tag{1.3}$$

where η_k represents the disturbance. The exponential character of the convergence imparts a robustness to the algorithm and guarantees that stability is maintained for suitably small η_k [BA].

When the X_k sequence *is not* a function of the θ_k sequence, then (1.1) and (1.2) are linear and the EAS is global; when the X_k sequence *is* a function of the θ_k, then the EAS is, in general, only local. Local exponential stability argues strongly for good performance of the system once it is near its operating point even when (suitably small) disturbances are present. Global stability, however, is of dubious value unless also accompanied by a local exponential result since adaptive systems can be globally stable but locally unstable [RPK], leading to bounded parameter estimates but poor performance. This paper therefore focuses on local exponential stability.

In certain applications [J2], the measured prediction error e_k is not precisely of the form $X_k^T\theta_k$. Table 1 describes two model structures (a transfer function model and an ARMAX, or autoregressive moving average with exogenous input model), which, when combined with an appropriate identifier input–output form, yield a prediction error that can be expressed as a transfer function operating on the inner product of the regressor vector and the parameter error vector. A simple illustrative case is when e_k contains a known filtering of $X_k^T\theta_k$ by a fixed stable rational operator $F(q^{-1})$, that is, $e_k = F(q^{-1})\{X_k^T\theta_k\}$. If $F(q^{-1})$ is stably invertible, then it is natural to consider an algorithmic form which filters the prediction error, as in

$$\theta_{k+1} = \theta_k - \mu X_k F^{-1}(q^{-1})\{e_k\} = \theta_k - \mu X_k\{X_k^T\theta_k\}, \tag{1.4}$$

to regain the known stability properties of (1.2).

An alternative approach is to recognize that, when the stepsize is small, the dynamics of θ_k are much slower than the dynamics of X_k. The prediction error e_k is then approximately equal to $\mathbf{F}(q^{-1})\{X_k^T\}\theta_k$ where $\mathbf{F}(q^{-1})$ operates on each component of the vector X_k^T in the same way that $F(q^{-1})$ acts on the scalar $X_k^T\theta_k$ (see Section 2 for details). Thus

$$e_k = F(q^{-1})\{X_k^T\theta_k\} = \mathbf{F}(q^{-1})\{X_k^T\}\theta_k + O(\mu). \tag{1.5}$$

This development leads to an alternative algorithmic form which filters the regressor vector X_k to recapture the desirable stability properties of (1.2). This is

$$\theta_{k+1} = \theta_k - \mu\mathbf{F}(q^{-1})\{X_k\}e_k, \tag{1.6}$$

where $\mathbf{F}(q^{-1})\{X_k\}$ plays the same role in (1.6) that X_k plays in (1.2), and the $O(\mu)$ perturbation of (1.5) plays the same role as the disturbance η_k of (1.3). Thus, it is suspected that if the vector sequence $\mathbf{F}(q^{-1})\{X_k\}$ is persistently spanning (as made precise in Section 3), and the stepsize is suitably small, then algorithm (1.6) will be EAS.

Both algorithms (1.4) and (1.6) attempt to recapture the desirable properties of the LMS form (1.2) in a modified problem setting. Incorporation of a known filter $F(q^{-1})$ in the prediction error is somewhat contrived, but it motivates two possible modifications to the basic algorithm (1.2): filtering of the error and filtering of the regressor.

A more realistic situation is when the filter $F(q^{-1})$ is fixed but unknown, in which case algorithms (1.4) and (1.6) cannot be implemented. In the output error identification problem [L3], [J2], for instance, $F(q^{-1})$ is the denominator (autoregressive) portion of the plant θ^*. Although θ^* is unknown, $\hat{\theta}_k$, an estimate of θ^*, is available, and it is reasonable to estimate $F(q^{-1})$ by $\hat{F}(q^{-1}, k)$, defined to be the denominator (autoregressive) portion of $\hat{\theta}_k$. By direct analogy with (1.4) and (1.6), it is natural to consider the filtered error algorithm

$$\theta_{k+1} = \theta_k - \mu X_k\hat{F}^{-1}(q^{-1}, k)\{e_k\} \tag{1.7}$$

Table 1

Prediction error for a given model and identifier combination	Some special choices for L and M: $L(q^{-1}, k)$	$M(q^{-1}, k)$	Some relevant literature
A transfer function model			
$e_k = \text{error} = d_k - \hat{y}_k = \dfrac{1}{1 - A^*(q^{-1})}\{X_k^T(\theta^* - \hat{\theta}_k)\} + w_k$	$\dfrac{1}{1 - \hat{A}(q^{-1}, k)}$	$\dfrac{1}{1 - A^*(q^{-1})}$	Stearns algorithm [F]
d_k = desired signal, $\hat{y}_k$ = estimate of $d_k = X_k^T\hat{\theta}_k$; w_k = unmodeled disturbances and measurement noise	$\dfrac{1}{1 - F(q^{-1})}$	$\dfrac{1}{1 - A^*(q^{-1})}$	See [L4], [L5], and [LJ]
X_k = regressor $= (\hat{y}_{k-1}, \ldots, \hat{y}_{k-n}, u_{k-1}, \ldots, u_{k-m})^T$	1	$\dfrac{1 - F(q^{-1})}{1 - A^*(q^{-1})}$	Landau-style algorithm, see [L1]
$\hat{\theta}_k$ = parameter estimate $= (\hat{a}_{1,k}, \ldots, \hat{a}_{n,k}, \hat{b}_{1,k}, \ldots, \hat{b}_{m,k})^T$; θ^* = "true" parameter $= (a_1, \ldots, a_n, b_1, \ldots, b_m)^T$	1	$\dfrac{1 - \hat{A}(q^{-1}, k)}{1 - A^*(q^{-1})}$	Proposed in [L5]. See also [L2], [JT], and [DG] for discussion of stability
$A^*(q^{-1}) = \sum_{i=1}^{n} a_i q^{-i}$			
An ARMAX model			
$e_k = \text{error} = d_k - \hat{y}_k = \dfrac{1}{1 + C^*(q^{-1})}\{X_k^T(\theta^* - \hat{\theta}_k)\} + w_k$	$\dfrac{1}{1 + \hat{C}(q^{-1}, k)}$	$\dfrac{1}{1 + C^*(q^{-1})}$	Similar to RML [L2] but with constant stepsize
d_k = desired signal, $\hat{y}_k$ = estimate of $d_k = X_k^T\hat{\theta}_k$	$\dfrac{1}{1 + F(q^{-1})}$	$\dfrac{1}{1 + C^*(q^{-1})}$	
X_k = regressor $= (d_{k-1}, \ldots, d_{k-n}, u_{k-1}, \ldots, u_{k-m}, e_{k-1}, \ldots, e_{k-p})^T$; $\hat{\theta}_k$ = parameter estimate $= (\hat{a}_{1,k}, \ldots, \hat{a}_{n,k}, \hat{b}_{1,k}, \ldots, \hat{b}_{m,k}, \hat{c}_{1,k}, \ldots, \hat{c}_{p,k})^T$	1	$\dfrac{1 + F(q^{-1})}{1 + C^*(q^{-1})}$	Generalization of extended least squares or approx. maximum likelihood [S]
θ^* = "true" parameter $= (a_1, \ldots, a_n, b_1, \ldots, b_m, c_1, \ldots, c_p)^T$; $C^*(q^{-1}) = \sum_{i=1}^{p} c_i q^{-i}$	1	$\dfrac{1 + \hat{C}(q^{-1}, k)}{1 + C^*(q^{-1})}$	

and the filtered regressor algorithm

$$\theta_{k+1} = \theta_k - \mu \hat{\mathbf{F}}(q^{-1}, k)\{X_k\} e_k. \tag{1.8}$$

If $\hat{F}$ is close to F, μ is small, and an appropriate excitation condition is fulfilled, algorithms (1.7) and (1.8) should retain (at least locally) the basic stability properties of (1.2). It is also easy to imagine algorithms which compensate for the presence of $F(q^{-1})$ in the prediction error by filtering both the error and the regressor vector in an appropriate manner.

All of the above algorithms are special cases of the general algorithm form

$$\theta_{k+1} = \theta_k - \mu \mathbf{L}(q^{-1}, k)\{X_k\} M(q^{-1}, k)\{X_k^T \theta_k\}, \tag{1.9}$$

where $\mathbf{L}(q^{-1}, k)$ filters the regressor vector and $M(q^{-1}, k)\{X_k^T \theta_k\}$ is the filtered prediction error. This paper finds conditions on μ, $\mathbf{L}$, M, and X_k under which (1.9) is EAS. The general algorithm form (1.9) may be viewed as a synthesis of many popular adaptive algorithms, such as those detailed in Table 1. For instance, using the definitions of the leftmost column of the table and a few lines of algebra (as in [J2]), $\mathbf{L}$ and M can be derived. A readable derivation of several such cases can be found in Chapter 6 of [TJL]. An equivalent error system for adaptive control is derived in [AB] and several possible algorithms, with various $\mathbf{L}$ and M, are discussed.

Some adaptive schemes replace the a priori error $X_k^T \theta_k$ in (1.9) with an a posteriori error $X_k^T \theta_{k+1}$. Appendix A shows that an a posteriori version of (1.9) is implemented as a normalized a priori algorithm. The behavior of the two schemes is virtually identical when μ is suitably small. The present analysis focuses on the small stepsize, a priori scheme, although it easily extends to include the small stepsize, a posteriori case. Equivalently, the analysis focuses on the small stepsize unnormalized scheme, but extends to the small stepsize normalized versions.

Preview

This paper examines the general algorithmic form (1.9), and in particular the eight adaptive algorithms listed in Table 1. The first four estimate the parameters of a transfer function model and the last four estimate the parameters of an ARMAX process. Some have been successfully analyzed previously; others have not. The local convergence properties of all eight algorithms are analyzed simultaneously by studying the local stability of the error equation (1.9), even though the implementable forms of the algorithms (involving updates of $\hat{\theta}_k$ using measured quantities) differ substantially. This averaging analysis is conceptually similar to the unified ODE (ordinary differential equation) approach of [LS], which approximates the discrete algorithm by a continuous-time differential equation. The major advantage of the present approach is that the analysis is not asymptotic in μ.

Section 2 provides the definitions used in the subsequent analysis, which proceeds in four steps. The first step (Section 3) finds sufficient conditions on $M(q^{-1}, k)$ and μ so that the stability of

$$\theta_{k+1} = \theta_k - \mu \mathbf{L}(q^{-1}, k)\{X_k\} \mathbf{M}(q^{-1}, k)\{X_k^T\} \theta_k \tag{1.10}$$

implies the stability of (1.9). Note that in (1.9), the scalar error $X_k^T\theta_k$ is filtered by $M(q^{-1}, k)$ while in (1.10) the regressor vector X_k^T is filtered by $\mathbf{M}(q^{-1}, k)$, where $\mathbf{M}$ represents the operator which acts on each component of the vector in the same way that M acts on $X_k^T\theta_k$ (see Section 2). Algorithm (1.10) is in a form to which averaging theory can be applied.

The second step (Section 4) recalls an appropriate averaging result (Theorem 1) which gives conditions on the regressor X_k and the time-varying filters $\mathbf{L}$ and $\mathbf{M}$ under which (1.10) and hence (1.9) are exponentially asymptotically stable. This exponential stability in the ideal case (no measurement noise or unmodeled dynamics) implies a certain degree of robustness in the nonideal case. The third step (Section 5) develops the machinery to translate the conditions of Sections 3 and 4, which involve stability and passivity of time-varying operators, to conditions on related time-invariant operators whose stability and passivity properties can be more readily determined.

The fourth and last step (summarized in Theorem 2) is the derivation in Section 6, where several sets of sufficient conditions are given for stability of the general algorithm form (1.9). These are then interpreted in terms of the eight algorithms of Table 1. These eight are illustrative of a variety of possible algorithms which filter the regressor and/or the error sequence as in (1.9), and the stability/convergence properties of such variants can often be determined by Theorem 2. This establishes a framework for the analysis and development of a wide variety of adaptive algorithms.

2. Notations and Definitions

A *rational operator* $N(q^{-1}, k)$ is defined to be the input–output mapping of a single input, single output finite-dimensional linear system

$$
\begin{aligned}
X_{k+1} &= A_k X_k + B_k u_k, \\
y_k &= C_k X_k + d_k u_k,
\end{aligned}
\tag{2.1}
$$

where all matrices are bounded, u_k is zero for all $k < 0$, and $X_0 = 0$. $N(q^{-1}, k)$ is said to be *exponentially stable* if there is a K and $0 < \alpha < 1$ such that $\|\Pi_{i=k}^{k+l} A_i\| \le K\alpha^l$ for all k. The *impulse response* at time $l + k$ of $N(q^{-1}, k)$ to an impulse at time k is

$$I_N(k + l, k) = C_{l+k}\left\{\prod_{i=k}^{k+l-1} A_i\right\} B_k.$$

Consequently, the sequence η_k filtered by $N(q^{-1}, k)$ is expressible as

$$N(q^{-1}, k)\{\eta_k\} = \sum_{l=-\infty}^{k} I_N(k, l)\eta_l.$$

If $N(q^{-1}, k)$ is exponentially stable, then the impulse response decays exponentially *independent of* k, that is, there is a K and $0 < \beta < 1$ such that $|I_N(l + k, k)| \le K\beta^l$ for all k. Note that $N(q^{-1}, k + l)\{u_k\}$ denotes the value of the image function at time $k + l$, that is, the value y_{k+l}. The expression $N(q^{-1}, k)\{u_{k-j}\}$ denotes the value of the image function at time k when the operator acts on the sequence v_k where $v_k = u_{k-j}$.

We often wish to consider a rational operator acting on a vector U_k. We reserve the bold notation to denote diagonal operators constructed as multiple copies of a scalar operator, that is, $\mathbf{N}(q^{-1}, k) = \mathrm{diag}[N(q^{-1}, k), N(q^{-1}, k), \ldots, N(q^{-1}, k)]$. Thus, $\mathbf{N}(q^{-1}, k)\{U_k\}$ operates on each component of the vector U_k in the same way that $N(q^{-1}, k)$ operates on a scalar sequence u_k. With a slight abuse of notation, let $\mathbf{N}(q^{-1}, k)\{U_k^T\}$ be the row vector obtained by filtering the jth component of U_k^T by $N(q^{-1}, k)$. For example, $\mathbf{M}(q^{-1}, k)$ of (1.10) acts on each component of the vector X_k^T exactly as $M(q^{-1}, k)$ of (1.9) acts on the scalar sequence $X_k^T\theta_k$.

The time-invariant operator $N(q^{-1})$ can be associated with the transfer function $N(z) = d + C(zI - A)^{-1}B$. The transfer function $N(z)$ (or equivalently, the quadruple $\{A, B, C, d\}$ realizing $N(z)$) is *strictly positive real* (SPR) if Re $N(e^{j\omega}) > 0$ for every ω and if all poles of $N(z)$ lie in $|z| < \alpha < 1$ (which is implied by, and in the minimal case equivalent to, $|\lambda_i(A)| < \alpha < 1$ for every eigenvalue of A). If $N(z)$ is SPR, then there are $\rho_1, \rho_2 \in (0, 1)$ such that $N(\rho_2 z) - \rho_1$ is SPR (equivalently, $\{\rho_2^{-1}A, B, \rho_2^{-1}C, d - \rho_1\}$ is SPR), since

$$(d - \rho_1) + \rho_2^{-1}C(zI - \rho_2^{-1}A)^{-1}B = N(\rho_2 z) - \rho_1.$$

This idea can be generalized to the time-varying case. The rational operator $N(q^{-1}, k)$ with associated time-varying linear system (2.1) is *strictly passive* if $N(q^{-1}, k)$ is exponentially stable, and if the following input-output inequality holds for some ρ:

$$\sum_{i=m}^{l} y_i u_i \ge \rho \sum_{i=m}^{l} u_i^2$$

for all $l \ge m$, for every input sequence $\{u_i\}$ with support in $i \ge m$, and for $X_m = 0$ (i.e., zero initial conditions).

A necessary and sufficient condition for a minimal realization $\{A, B, C, d\}$ to be SPR is the existence of matrices $P > 0$, $Q > 0$, L, and scalars $\rho > 0$ and n such that

$$
\begin{aligned}
A^T P A - P &= -LL^T - Q, \\
B^T P A + nL^T &= C, \\
n^2 &= 2d - 2\rho - B^T P B.
\end{aligned}
\tag{2.2}
$$

The positive real lemma for time-varying systems (see Appendix B of [L3]) shows that $N(q^{-1}, k)$ with associated time-varying linear system (2.1) is strictly passive if there exist matrices $P_k > 0$, $Q_k > 0$, L_k, and scalars n_k, $\rho > 0$ with P_k, P_k^{-1}, Q_k, Q_k^{-1}, L_k, and n_k bounded and such that

$$
\begin{aligned}
A_k^T P_k A_k - P_k &= -L_k L_k^T - Q_k, \\
B_k^T P_k A_k + n_k L_k^T &= C_k, \\
n_k^2 &= 2d_k - 2\rho - B_k^T P_k B_k.
\end{aligned}
\tag{2.3}
$$

Intuitively, a passive system is one which does not generate energy. The magnitudes of the eigenvalues of Q_k provide a measure of the amount of energy dissipated at each timestep.

3. Approximation of Filtered Error

This section shows that the exponential stability of (1.10) implies the exponential stability of (1.9), the error equation governing the behavior of the adaptive algorithm. This allows averaging theory to be applied to (1.10) to derive sufficient conditions for the exponential stability of the error system, and consequently of the adaptive system. We follow a course similar to that in [AB].

One condition that is virtually necessary for (1.9) to be stable for all bounded regressor sequences X_k is that both the regressor filter $\mathbf{L}(q^{-1}, k)$ and the error filter $M(q^{-1}, k)$ be exponentially stable. We therefore assume the exponential stability of these time-varying operators. Later (Section 5), we examine the question of how to guarantee the stability of these time-varying operators from the stability of the associated frozen operators, defined from $\mathbf{L}$ and M at each fixed time k. In practice, stability of the frozen operators can often be guaranteed, or can be easily verified.

Suppose $M(q^{-1}, k)$ is exponentially stable and let $I_M(l + k, k)$ be the impulse response. Then there is a K and $0 < \alpha < 1$ such that $|I_M(l + k, k)| < K\alpha^l$ for all k. In terms of I_M, the action of $M(q^{-1}, k)$ on the sequence $X_k^T\theta_k$ is described by the convolution sum

$$M(q^{-1}, k)\{X_k^T\theta_k\} = \sum_{l=-\infty}^{k} I_M(k, l)X_l^T\theta_l. \tag{3.1}$$

The first lemma compares (3.1) with the action of $\mathbf{M}(q^{-1}, k)$ on the sequence X_k, then multiplied by θ_k

$$\mathbf{M}(q^{-1}, k)\{X_k^T\}\theta_k = \left\{\sum_{l=-\infty}^{k} I_M(k, l)X_l^T\right\}\theta_k \tag{3.2}$$

and shows that the difference between (3.1) and (3.2) can be bounded in terms of the norm of X, the exponential decay rate of the operator M, and the successive differences of the parameter errors.

Lemma 1.

$$\|M(q^{-1}, k)\{X_k^T\theta_k\} - \mathbf{M}(q^{-1}, k)\{X_k^T\}\theta_k\| \le \|X\|_\infty \frac{K\alpha}{(1-\alpha)^2} \sup_{i\le k} \|\theta_i - \theta_{i-1}\|_\infty. \tag{3.3}$$

Proof. Combining (3.1) and (3.2) gives

$$\begin{aligned} M(q^{-1}, k)\{X_k^T\theta_k\} - \mathbf{M}(q^{-1}, k)\{X_k^T\}\theta_k &= \sum_{l=-\infty}^{k} I_M(k, l)X_l^T\theta_l - \left\{\sum_{l=-\infty}^{k} I_M(k, l)X_l^T\right\}\theta_k \\ &= \sum_{l=-\infty}^{k-1} I_M(k, l)X_l^T \sum_{m=1}^{k-l} (\theta_{l+m-1} - \theta_{l+m}). \end{aligned}$$

Since I_M is exponentially decaying, the norm can be bounded

$$\begin{aligned} &\|M(q^{-1}, k)\{X_k^T\theta_k\} - \mathbf{M}(q^{-1}, k)\{X_k^T\}\theta_k\|_\infty \\ &\le \|X\|_\infty \left\| \sum_{l=-\infty}^{k-1} K\alpha^{k-l} \sum_{m=1}^{k-l} (\theta_{l+m-1} - \theta_{l+m}) \right\|_\infty \\ &\le \|X\|_\infty \left\| \sum_{l=-\infty}^{k-1} K\alpha^{k-l}(k-l) \right\|_\infty \sup_{i\le k} \|\theta_i - \theta_{i-1}\|_\infty \\ &\le \|X\|_\infty \frac{K\alpha}{(1-\alpha)^2} \sup_{i\le k} \|\theta_i - \theta_{i-1}\|_\infty. \end{aligned}$$

■

In order to make this estimate useful, it is necessary to bound the sup term in (3.3). Let $\|\mathbf{L}X\|_\infty$ denote the norm of the vector sequence $\mathbf{L}(q^{-1}, k)\{X_k\}$. This quantity exists since X_k is assumed to be a bounded sequence (see Section 6) and $\mathbf{L}$ is exponentially stable.

Lemma 2. *With quantities as above,*

$$\sup_{i\le k} \|\theta_i - \theta_{i-1}\|_\infty \le \frac{\mu K}{1-\alpha} \|\mathbf{L}X\|_\infty \|X\|_\infty \sup_{i\le k} \|\theta_i\|_\infty. \tag{3.4}$$

Proof. Observe that

$$\begin{aligned} \|M(q^{-1}, k)\{X_k^T\theta_k\}\| &\le \sum_{l=-\infty}^{k} \|I_M(k, l)\|_\infty \|X_l\|_\infty \|\theta_l\|_\infty \\ &\le \frac{K}{1-\alpha} \|X\|_\infty \sup_{i\le k} \|\theta_i\|_\infty. \end{aligned}$$

Then (3.4) follows immediately from (1.9). ■

The two bounds (3.3) and (3.4) can now be combined to give

$$\|M(q^{-1}, k)\{X_k^T\theta_k\} - \mathbf{M}(q^{-1}, k)\{X_k^T\}\theta_k\|_\infty \le \mu \frac{K\alpha}{(1-\alpha)^2} \|\mathbf{L}X\|_\infty \|X\|_\infty^2 \sup_{i\le k} \|\theta_i\|_\infty$$

which shows that the difference between $M(q^{-1}, k)\{X_k^T\theta_k\}$ and $\mathbf{M}(q^{-1}, k)\{X_k^T\}\theta_k$ is $O(\mu)$. Equation (1.9) can then be rewritten as

$$\theta_{k+1} = \theta_k - \mu\mathbf{L}(q^{-1}, k)\{X_k\}\mathbf{M}(q^{-1}, k)\{X_k^T\}\theta_k + \Delta(q^{-1}, k)\{\theta_k\}, \tag{3.5}$$

where Δ is a bounded operator with gain proportional to μ^2. The stepsize μ can then be chosen small enough so that the dominant part of (3.5) is

$$\theta_{k+1} = (I - \mu\mathbf{L}(q^{-1}, k)\{X_k\}\mathbf{M}(q^{-1}, k)\{X_k^T\})\theta_k \tag{3.6}$$

which is precisely (1.10). Thus, exponential stability of (1.10) implies exponential stability of the adaptive system (1.9), provided the stepsize is chosen suitably small. It is not surprising that the upper bound on μ depends on $\mathbf{L}$, M, and $\|X\|_\infty$.

In the above analysis we have assumed that the action of $M(q^{-1}, k)$ on the sequence $X_k^T\theta_k$ (and the action of $\mathbf{M}(q^{-1}, k)$ on X_k^T) occurs with zero initial conditions. More precisely, adopting a state variable representations of the form (2.1) for M, $\mathbf{M}$, and $\mathbf{L}$, the analysis assumes zero state at time zero. Nonzero initial states for (2.1) would correspond to exponentially decaying terms added to $M(q^{-1}, k)\{X_k^T\theta_k\}$, $\mathbf{M}(q^{-1}, k)\{X_k^T\}$, and $\mathbf{L}(q^{-1}, k)\{X_k\}$. Equation (3.6) would then be perturbed by an additive term that was decaying exponentially. The basic equivalence of the exponentially stability of (1.10) and (1.9) would remain valid, but the upper bound on μ would also depend on the magnitude of the initial states in $\mathbf{L}$ and M. Note that for fixed degree $\mathbf{L}$ and M, the initial condition effects become negligible as $\mu \to 0$. The details of a rigorous analysis that includes such initial condition effects can be carried out as in [AB] or [KAM].

4. Averaging and Persistence of Excitation

This section recalls an averaging theorem and defines generalized persistence of excitation conditions that guarantee exponential asymptotic stability of the adaptive error system. These ideas are then illustrated with two simple examples. The following is well known [AB], [BJ], [SV]; see [BJ] for a proof.

Theorem 1. *Consider the system*

$$X_{k+1} = (I - \mu A_k)X_k, \tag{4.1}$$

where $A_k \in \mathbf{R}^{n\times n}$ is a sequence of bounded matrices. Define the sliding average $\bar{A}_k(m) = (1/m)\sum_{i=1}^{m} A_{k+i-1}$. Suppose that for some positive definite matrix P there is an integer m and an $\alpha > 0$ such that for all k and for each eigenvalue λ_i

$$\lambda_i\{P\bar{A}_k(m) + \bar{A}_k^T(m)P\} \geq \alpha. \tag{4.2}$$

Then there is a μ^* such that (4.1) *is uniformly EAS for every* $0 < \mu < \mu^*$.

This theorem says that difference equations with sufficiently small stepsizes are stable whenever the averaged equation $X_{k+1}^{av} = (I - \mu\bar{A}_k(m))X_k^{av}$ has a certain degree of stability, determined by α.

To analyze the adaptive system (1.10), let $A_k = \mathbf{L}(q^{-1}, k)\{X_k\}\mathbf{M}(q^{-1}, k)\{X_k^T\}$. Then the sliding average is

$$\bar{A}_k(m) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{L}(q^{-1}, k+i-1)\{X_k\}\mathbf{M}(q^{-1}, k+i-1)\{X_k^T\}. \tag{4.3}$$

In order to streamline subsequent discussion, we propose the following.

Definitions. Consider a particular algorithm with an error system of the form (1.9) with given operators $\mathbf{L}(q^{-1}, k)$ and $M(q^{-1}, k)$. Let $\bar{A}_k(m)$ be defined as in (4.3) where $\mathbf{M}(q^{-1}, k)$ is the vector version of $M(q^{-1}, k)$. Then, if there exists a $P > 0$, an $\alpha > 0$, and an m such that (4.2) holds for every k and for every eigenvalue λ_i, the regressor sequence X_k will be called *persistently exciting for this algorithm*. If X_k fulfills (4.2) with $\mathbf{L}$ and $\mathbf{M}$ identity operators and $P = \frac{1}{2}I$, then X_k will be said to be *persistently spanning*.

The averaging theorem can then be restated concisely. If the input to an adaptive algorithm is persistently exciting (for that algorithm) and the stepsize is small enough, then the error system associated with the algorithm is EAS. In general, the class of signals that persistently excites an algorithm with filters $\mathbf{L}_1$ and $\mathbf{M}_1$ will differ from the class of signals that persistently excites an algorithm with filters $\mathbf{L}_2$ and $\mathbf{M}_2$. Thus *persistence of excitation conditions must always be linked to a particular algorithm.*

As a simple example, consider the "equation error" algorithm [M] which estimates the parameters of a (linear time-invariant) transfer function. In the notation of Table 1, $X_k^T = (y_{k-1}, \ldots, y_{k-n}, u_{k-1}, \ldots, u_{k-m})$ and $\hat{\theta}_k^T = (\hat{a}_{1,k}, \ldots, \hat{a}_{n,k}, \hat{b}_{1,k}, \ldots, \hat{b}_{m,k})$, $\mathbf{L}(q^{-1}, k)$ and $\mathbf{M}(q^{-1}, k)$ are identity operators. Taking $P = \frac{1}{2}I$ in (4.2), a condition for exponential stability of the equation error algorithm is that there exists an $\alpha > 0$ and an $m > 0$ such that

$$\lambda_j\left\{\frac{1}{m}\sum_{i=1}^{m} X_{k+i-1}X_{k+i-1}^T\right\} > \alpha \qquad \text{for all } j. \tag{4.4}$$

This is precisely the standard persistence of excitation requirement [B] and may be interpreted as a condition on the spanning properties of the regressor [AJ] or as a condition on the frequency content of the input [BS].

Another example is furnished by the "output error" algorithm [J2], which also estimates the parameters of a transfer function. For this algorithm, $X_k^T = (\hat{y}_{k-1}, \ldots, \hat{y}_{k-n}, u_{k-1}, \ldots, u_{k-m})$, $\hat{\theta}_k$ and $\mathbf{L}(q^{-1}, k)$ are as above, and $M(q^{-1}, k) = 1/(1 - A^*(q^{-1}))$ where $A^*(q^{-1})$ represents the fixed but unknown autoregressive part of θ^* as in Table 1. Again, taking $P = \frac{1}{2}I$ in (4.2), a condition for exponential stability of the output error algorithm is that there exists an m and an $\alpha > 0$ such that

$$\lambda_j\left\{\frac{1}{m}\sum_{i=1}^{m} X_{k+i-1}\mathbf{M}(q^{-1}, k)\{X_{k+i-1}^T\}\right\} > \alpha \qquad \text{for all } j. \tag{4.5}$$

This condition, which is the persistence of excitation condition for the output error algorithm, will be satisfied if X_k is persistently spanning and if $M(q^{-1}, k)$ (and hence $\mathbf{M}(q^{-1}, k)$) is strictly passive. Recall that for a time-invariant operator, strict passivity is equivalent to SPR. (This special case of Theorem 2 has been shown in [AB].)

Condition (4.5) is an average positivity condition and does not require the transfer function associated with $\mathbf{M}(q^{-1}, k)$ to be SPR for every frequency. Suppose that this transfer function fails to be positive for some interval of frequencies $[\omega_0, \omega_1]$. If the regressor X_k has most of its energy outside of $[\omega_0, \omega_1]$, then $\mathbf{M}$ will still act, on the average, as a passive operator and (4.6) will still hold, implying exponential stability of the output error algorithm. This idea has been exploited in [RPK]. This emphasizes again that persistence of excitation conditions are algorithm-dependent.

These two examples are particularly simple because the operators $\mathbf{L}$ and $\mathbf{M}$ are time invariant. In the more general case, it will be useful to relate the stability and passivity of slowly time-varying systems to the stability and passivity of related time-invariant systems. These will provide the last link necessary to analyze excitation conditions for the more general algorithms of Table 1.

5. Stability and Passivity of Slowly Varying Systems

The variation in θ_k, and in the filters **L** and **M** (which are typically parameterized by θ) is slow compared with the variation in the input $\mathbf{X}_k$ due to the small stepsize μ. This section exploits the slowness in three ways. First, lemma 3 relates the EAS of a slowly time-varying system to the EAS of related "frozen" (time-invariant) systems. Next, Lemma 4 relates the passivity of a slowly varying system to the passivity of the frozen systems. Lemma 5 and its corollary then construct a family of strictly passive (slowly) time-varying operators. This family includes operators which are "near" the identity, and can be used to demonstrate that $\mathbf{ML}^{-1}$ of (4.3) is passive for **M** and **L** which are nearly equal.

All three results depend critically on the time-scale separation (slowness of variation). The stability and passivity of the frozen systems, which are relatively easy to determine, are used in the next section to provide conditions for the EAS of adaptive systems with time-varying regressor and error filters.

Lemma 3. *Consider the time-varying system*

$$X_{k+1} = A_k X_k \tag{5.1}$$

and assume that the related frozen systems

$$X_{k+1} = A_p X_k$$

are exponentially stable for every integer p, uniformly in p. If $\|A_p\| < \alpha_1$ *for all p and if* $\sup_{p \geq p_0} \|A_{p+1} - A_p\|$ *is small enough for some finite* p_0, *then systems* (5.1) *are exponentially asymptotically stable.*

Proof. This proof introduces some ideas which will be used in succeeding lemmas; another proof of this result is given in [D]. Let P_p satisfy

$$P_p - A_p^T P_p A_p = I, \tag{5.2}$$

where $0 < \alpha_2 I \leq P_p \leq \alpha_3 I$ for some α_2 and α_3. This is always possible by the discrete-time Lyapunov stability theorem and the uniformity of the theorem hypothesis. Let $V(X_k, k) = X_{k+1}^T P_k X_{k+1}$ be a candidate Lyapunov function for (5.1). It is clearly globally positive definite and decrescent. Also,

$$\begin{aligned} V(X_{k+1}, k+1) - V(X_k, k) &= X_k^T A_k^T P_k A_k X_k - X_k^T P_{k-1} X_k \\ &= X_k^T (P_k - P_{k-1}) X_k - X_k^T X_k. \end{aligned} \tag{5.3}$$

The solution of (5.2) can be obtained from

$$\operatorname{vec}(P_p) = [I - A_p^T \otimes A_p^T]^{-1} \operatorname{vec}(I),$$

where $\otimes$ represents the Kronecker product and $\operatorname{vec}(P_p)$ indicates a single column vector which is a concatenation of the columns of P_p. The uniform positive definiteness of P_p guarantees that A_p has all its eigenvalues uniformly less than 1 in magnitude. Since the eigenvalues of $A \otimes B$ are $\lambda_i \mu_j$ where λ_i are the eigenvalues of A and μ_j are the eigenvalues of B, and since the determinant is the product of the eigenvalues,

$$\det|I - A_p^T \otimes A_p| = \prod_{i=1}^{n} \prod_{j=1}^{n} (1 - \lambda_i(A)\lambda_j(A))$$

which is bounded away from zero. Hence $\operatorname{vec}(P_p)$ depends continuously on A_p for $p \geq p_0$, and so given $\varepsilon > 0$, there is a δ such that $\|P_{k+1} - P_k\| < \varepsilon$ whenever $\|A_{k+1} - A_k\| < \delta(\varepsilon)$ for all $k \geq p_0$. Thus, for $\varepsilon = \frac{1}{2}$, (5.3) is negative definite and system (5.1) is exponentially asymptotically stable. ■

The SPR of certain the time-invariant operators is relevant in securing the stability of some adaptive algorithms [L5]. The analog for time-varying operators is strict passivity. The next result parallels the above stability analysis, in which the stability of a time-varying system was related to the stability of a collection of time-invariant systems, by relating the strict passivity of a time-varying system to the strict passivity of a family of time-invariant systems. For the algorithms of interest, strict passivity of this family may be guaranteed or easily checked. As with the stability result, *slow* time variation is crucial.

Recall (from Section 2) that if $N(z)$ is SPR, then there are $\rho_1, \rho_2 \in (0, 1)$ such that $\{\rho_2^{-1} A, B, \rho_2^{-1} C, d - \rho_1\}$ is SPR. A collection of time-invariant operators $N(q^{-1}, p)$, $p = 0, 1, 2, \ldots$, is said to be strictly passive for every p *uniformly in* p if ρ_1, ρ_2 can be chosen independent of p so that $\{\rho_2^{-1} A_p, B_p, \rho_2^{-1} C_p, d_p - \rho_1\}$ is SPR for all p. Given any time-varying rational operator $N(q^{-1}, k)$ we can associate with it time-invariant operators $N(q^{-1}, p)$ by freezing at each time instant the defining equation (2.1) for $N(q^{-1}, k)$. This leads to:

Lemma 4. *Consider the discrete time-varying operator* $N(q^{-1}, k)$ *and the related frozen (time-invariant) operators* $N(q^{-1}, p)$. *Assume that the quadruples* $\{A_p, B_p, C_p, d_p\}$ *of* (2.1) *defining* $N(q^{-1}, p)$ *are minimal. If* $N(q^{-1}, p)$ *is strictly passive for every p, uniformly in p, and if the time variation in* $N(q^{-1}, k)$ *is slow enough in the sense that* $\|A_{p+1} - A_p\|$, $\|B_{p+1} - B_p\|$, $\|C_{p+1} - C_p\|$, $\|d_{p+1} - d_p\|$ *are suitably small uniformly in p, then* $N(q^{-1}, k)$ *is strictly passive.*

Proof. See Appendix B. ■

Lemma 4 shows that the property of strict passivity of an operator is robust to slight perturbations in A_k, B_k, C_k, and d_k of (2.1).

The next task is to develop a class of operators which are time varying and strictly passive. Towards this end we construct a family of passive time-invariant operators. The idea is simple; if two operators N_1 and N_2 are "similar," then $N_1 N_2^{-1}$ and $N_2 N_1^{-1}$ are "close" to the identity, which is passive. This will be useful in examining persistence of excitation conditions where the two operators **L** and **M** (of equation (1.10)) are approximately equal.

Lemma 5. *Suppose that* $N(q^{-1}, \theta)$ *has a state variable realization defined by* $\{A(\theta), B(\theta), C(\theta), d(\theta)\}$ *with the constituent matrices continuously dependent on the parameter* θ *for all* $\theta \in \Theta$. *Suppose that for any* $\theta \in \Theta$, *the operators* $N(q^{-1}, \theta)$ *and*

$N^{-1}(q^{-1}, \theta)$ are exponentially stable, uniformly in θ. Fix θ. Then there exists β such that for all $\psi \in \Theta$, $\|\psi - \theta\| \le \beta$ implies that the operators $N(q^{-1}, \psi)N^{-1}(q^{-1}, \theta)$ and $N^{-1}(q^{-1}, \psi)N(q^{-1}, \theta)$ are strictly passive.

Proof. Exponential stability of the operators is guaranteed by assumption. When $\psi = \theta$, the transfer function of each operator takes the value 1 on $|z| = 1$. By continuity, the transfer functions of $N(q^{-1}, \psi)N^{-1}(q^{-1}, \theta)$ and $N^{-1}(q^{-1}, \psi)N(q^{-1}, \theta)$ will have strictly positive real parts for $\|\psi - \theta\| \le \beta$. ■

This idea can be extended to time-varying operators by combining the last two lemmas. Suppose that a rational operator $N(q^{-1}, \theta_k)$ is dependent on a parameter θ taking the value θ_k at time k, i.e., there exist $A_k = A(\theta_k)$, $B_k = B(\theta_k)$, $C_k = C(\theta_k)$, and $d_k = d(\theta_k)$ that describe the operator. Denote the inverse operator by $N^{-1}(q^{-1}, \theta_k)$.

Corollary. *Adopt the hypotheses of Lemma 5. Assume that the sequence $\theta_k \in \Theta$ is slowly varying, that is, $\|\theta_{k+1} - \theta_k\| \le \varepsilon$. Consider the sequence $\psi_k \in \Theta$ with $\|\psi_{k+1} - \psi_k\| \le \varepsilon$, and $\|\theta_k - \psi_k\| \le \beta$ for all k. Then for suitably small ε and β, the operators $N(q^{-1}, \psi_k)N^{-1}(q^{-1}, \theta_k)$ and $N^{-1}(q^{-1}, \psi_k)N(q^{-1}, \theta_k)$ are strictly passive.*

Proof. Combine Lemmas 4 and 5. ■

Thus, it is possible to determine the stability and passivity properties of slowly time-varying operators from stability and passivity properties of the related frozen operators.

6. Interpretation of Excitation Conditions

The error system associated with each of the algorithms of Table 1 is in the form of equation (1.9), where each algorithm is specified by a given pair of filters **L** and M. This section gathers together the previous analyses to show that the exponential asymptotic stability of the error system can be guaranteed if the regressor is persistently spanning, if the filter $\mathbf{ML}^{-1}$ is strictly passive, and if the stepsize is small. This is accomplished in two steps. Lemma 6 shows that if the *filtered* regressor is persistently spanning, then it persistently excites the algorithm. Lemma 7 translates this to a spanning condition on the unfiltered regressor. Theorem 2 presents the main result.

In each of the algorithms of the table, $\mathbf{L}(q^{-1}, k)$ is an autoregression, which implies that $\mathbf{L}^{-1}(q^{-1}, k)$ exists and is a moving average (and therefore exponentially stable). Define the filtered regressor vector sequence $Z_k = \mathbf{L}(q^{-1}, k)\{X_k\}$. With invertibility of **L**, this can be written as $X_k = \mathbf{L}^{-1}(q^{-1}, k)\{Z_k\}$ and the matrix $\bar{A}_k$ of (4.3) can be rewritten in terms of the filtered regressor as

$$\bar{A}_k(m) = \frac{1}{m}\sum_{i=1}^{m} Z_{k+i-1}\mathbf{M}(q^{-1}, k+i-1)\{\mathbf{L}^{-1}(q^{-1}, k+i-1)\{Z_{k+i-1}^T\}\}.$$

Theorem 1 showed that the error system (1.9) is exponentially stable when $\bar{A}_k(m) + \bar{A}_k^T(m)$ is positive definite, that is, when the algorithm is persistently excited. The next lemma shows that if $\mathbf{M}(q^{-1}, k)\{\mathbf{L}^{-1}(q^{-1}, k)\}$ is strictly passive and Z_k is persistently spanning, then Z_k persistently excites the algorithm and hence leads to exponential stability of the error system (1.9).

Lemma 6. *Suppose there is an $\alpha > 0$ and an $m > 0$ such that for all j, $\sum_{k=j}^{j+m} Z_k Z_k^T > \alpha I$. Suppose that $\mathbf{N}(q^{-1}, k) = \mathbf{M}(q^{-1}, k)\{\mathbf{L}^{-1}(q^{-1}, k)\}$ is strictly passive. Then there is some $\rho > 0$ such that*

$$\bar{A}_k(m) + \bar{A}_k^T(m) = \sum_{k=j}^{j+m} Z_k\mathbf{N}(q^{-1}, k)\{Z_k^T\} + \left\{\sum_{k=j}^{j+m} Z_k\mathbf{N}(q^{-1}, k)\{Z_k^T\}\right\}^T > \rho I \quad \textit{for all } j.$$

Proof. Let W be an arbitrary nonzero vector with the same dimension as Z_k. Neglecting initial conditions (which die away exponentially),

$$W^T\left\{\sum_{k=j}^{j+m} Z_k\mathbf{N}(q^{-1}, k)\{Z_k^T\}\right\}W + W^T\left\{\sum_{k=j}^{j+m} Z_k\mathbf{N}(q^{-1}, k)\{Z_k^T\}\right\}^T W$$

$$= 2W^T\left\{\sum_{k=j}^{j+m} Z_k\mathbf{N}(q^{-1}, k)\{Z_k^T\}\right\}W = 2\sum_{k=j}^{j+m} v_k N(q^{-1}, k)\{v_k\},$$

where $v_k = Z_k^T W$. Since N is strictly passive, there is some $\rho > 0$ such that this expression is bounded below by

$$2\rho\sum_{k=j}^{j+m} v_k^2 = 2\rho W^T \sum_{k=j}^{j+m} Z_k Z_k^T W$$

which gives the desired inequality. ■

If the initial conditions are not neglected, then the lower bound is $2\rho\sum_{k=j}^{j+m} v_k^2 + IC$ where IC represents initial condition effects; α must then be assumed large enough to overcome these initial condition effects. This point is discussed in [KAM].

Lemma 6 relates the persistency of excitation condition to a spanning property of the filtered regressor Z_k and the passivity of $\mathbf{ML}^{-1}$. This can now be translated to a condition on the regressor vector itself.

Lemma 7. *Let $X_k = \mathbf{L}^{-1}(q^{-1}, k)\{Z_k\}$ where $\mathbf{L}^{-1}(q^{-1}, k) = I + \mathbf{F}(q^{-1}, k)$, **F** is a time-varying polynomial in q^{-1}, and $\mathbf{L}^{-1}$ is exponentially stable. If the rate of variation of the coefficients of **F** is slow enough, and if X_k is persistently spanning, then Z_k is persistently spanning.*

Proof. In outline, the proof goes as follows. For **F** time invariant, the result follows by combining Theorems 2.2 and 2.4 of [AJ]. The time-varying result follows by modifying the above proof utilizing the slowness of the time variation as in the previous lemmas. See also [AG]. ■

Notice that in Table 1, all of the **L** operators have the form $(I + \mathbf{F}(q^{-1}, k))^{-1}$ as required by the lemma.

This result, combined with Theorem 1, shows that if **M** and **L** are exponentially stable, if $\mathbf{ML}^{-1}$ is strictly passive, and if the regressor is persistently spanning, then the algorithm is persistently excited and hence exponentially stable. Lemma 3 showed that the exponential stability of slowly time-varying operators $\mathbf{L}(q^{-1}, k)$ and $M(q^{-1}, k)$ can be inferred from the exponential stability of the related frozen operators $\mathbf{L}(q^{-1}, p)$ and $M(q^{-1}, p)$. Moreover, the strict passivity of $\mathbf{ML}^{-1}$ can be deduced from the strict passivity of $\mathbf{M}(q^{-1}, p)\{\mathbf{L}^{-1}(q^{-1}, p)\}$ for all p and the slowness of variation of the operators. Gathering these results together gives the main result.

Theorem 2. *Consider the error system*

$$\theta_{k+1} = \theta_k - \mu \mathbf{L}(q^{-1}, k)\{X_k\} M(q^{-1}, k)\{X_k^T \theta_k\} \tag{6.1}$$

associated with an adaptive algorithm with regressor filter $\mathbf{L}(q^{-1}, k)$ *and error filter* $M(q^{-1}, k)$. *If the regressor sequence* X_k *is persistently exciting for this algorithm, the stepsize is small, the initial error* θ_0 *is small, and the initial states of* M *and* **L** *are small, then the error system is locally exponentially asymptotically stable. Persistency of excitation of the algorithm is guaranteed if*:

(1) $M(q^{-1}, k)$ *and* $\mathbf{L}(q^{-1}, k)$ *are exponentially stable.*
(2) $\mathbf{M}(q^{-1}, k)\{\mathbf{L}^{-1}(q^{-1}, k)\}$ *is strictly passive.*
(3) *The regressor* X_k *is persistently spanning.*

In turn, condition (1) *is true whenever*

(1a) *the frozen systems* $M(q^{-1}, p)$ *and* $\mathbf{L}(q^{-1}, p)$ *are uniformly exponentially stable for all* p, *and*
(1b) $M(q^{-1}, k)$ *and* $\mathbf{L}(q^{-1}, k)$ *are slowly varying.*

Condition (2) *is true whenever*

(2a) $\mathbf{M}(q^{-1}, p)\{\mathbf{L}^{-1}(q^{-1}, p)\}$ *is uniformly strictly passive for all* p, *and*
(2b) $M(q^{-1}, k)$ *and* $\mathbf{L}(q^{-1}, k)$ *are slowly varying.*

It should be noted that (1) and (2) are not necessary to have persistence of excitation, nor are (1a) and (1b) necessary to have (1), nor are (2a) and (2b) necessary for (2). This theorem, then, provides several possible combinations of sufficient conditions for the exponential stability of the error system (6.1).

The strict passivity condition (2) is a generalization of the familiar SPR condition that appears when **L** and M are time invariant. In the time-invariant case, exponential stability can sometimes be retained even if the SPR condition is violated, by restricting the frequency content of the regressor. Similarly, in the time-varying case, exponential stability can be maintained even if the strict passivity condition is violated, as long as the persistence of excitation condition holds.

Although it is difficult to make sense of "frequency content" in the context of a time-varying system, the requirement that M and **L** vary slowly can be interpreted as a time-scale separation. For small stepsizes, the dynamics of θ, M, and **L** are much slower than the dynamics of the regressor, and it is reasonable to interpret the lack of strict passivity of $\mathbf{ML}^{-1}$ using frequency domain intuition. For instance, $\mathbf{ML}^{-1}$ may be capable of generating energy at certain frequencies. If $\mathbf{ML}^{-1}$ dissipates even more energy at other frequencies, or if the regressor never excites these modes, then $\mathbf{ML}^{-1}$ may act, on the average, as a passive operator, and the persistence of excitation condition may still be fulfilled. Thus, although $\mathbf{ML}^{-1}$ may not be passive at every time step k for all regressors X_k, if it is, *on the average*, strictly passive, stability can be assured.

Theorem 2 is only a local result, that is, θ_k is guaranteed to converge only if the initial value of θ_0 is not too large and if the initial states of M and **L** are small. This is a consequence of the linearization in Lemmas 1 and 2, and of the assumption (used in Section 2) that X_k is bounded. This boundedness is not guaranteed a priori since X_k may contain signals such as estimated outputs which can diverge if the algorithm is unstable. If, however, the initial magnitude θ_0 is small, then the difference between the first desired output d_1 and the first estimated output $\hat{y}_1$ will also be small. Exponential stability then guarantees that this difference remains small as time evolves, which implies that $\hat{y}_k$ (and hence X_k) remain bounded. At first glance, this appears to be a circular argument, and if this were an attempt to demonstrate *global* (in θ) stability, it would indeed be circular. The presumption here, however, is that θ_0 is initialized near θ^*, implying that the initial prediction errors are small. When the algorithm is persistently excited, small errors remain small, and hence $\|X_k\|$ is finite. Said another way, the local exponential stability is a local contraction [H]. Once the trajectories are within the grip of the contraction, they cannot escape, and Theorem 2 applies. Outside of the contractive region, nothing has been said by our analysis.

Although it would be desirable to quantify the adjectives "large" and "small" here and in Theorem 2, such quantification is difficult. The general trends, however, are apparent. Larger (smaller) eigenvalues of the excitation matrix (4.3) allow larger (smaller) errors in the initial estimates, and allow the algorithm to retain stability in environments with larger (smaller) disturbances. Smaller stepsizes (typically) imply slower variation in the filters **L** and M, and tend to "average" disturbances more effectively. The requirement on the initial states of **L** and M is a technical condition with little impact on algorithm design. Some quantified results are available for time-invariant operators in [AB].

A major reason for focusing on the exponential (as opposed to bounded input bounded output) stability of the adaptive error system is that exponential convergence guarantees a certain robustness in the presence of nonidealities such as unmodeled dynamics or measurement noise and allows consideration of the situation in which θ^* is itself varying slowly. Suppose, for instance, that θ^* varies on a timescale of $1/\mu^2$ or slower. Then the analysis of the previous sections is unchanged except for an added $O(\mu^2)$ perturbation which can be easily incorporated in equation (3.5) as part of the $\Delta(q^{-1}, k)$ operator. A persistently excited (and hence exponentially stable) algorithm near its equilibrium continues to operate well in the presence of suitably small disturbances and is robust to slow variations in the "true" parametrization θ^*.

Perhaps the most serious limitation of Theorem 2 is that it is, in general, a nontrivial problem to translate the persistence of excitation condition on the regressor vector X_k and the filters **L** and M, to a condition on the signals which can be manipulated in any given problem context. In certain applications such as identification, it may be possible to directly or indirectly manipulate the regressor vector to achieve persistence of excitation. In some applications the regressor contains estimated quantities which approach desired quantities as θ_k approaches zero. Near the operating point, then, the spanning properties of the regressor closely match the spanning properties of a vector of desired quantities which can often be shown (or manipulated) to achieve a persistently spanning property.

Theorem 2 can be applied to any of the algorithms of Table 1. If the regressor is persistently excited, if the stepsize is small enough, and if the initial error $\|\hat{\theta}_0 - \theta^*\|$ is not large, then the parameter estimates $\hat{\theta}_k$ converge to (and remain in) a small ball about the true parametrization θ^*. Conditions (1a), (1b), (2a), and (2b) then give several possible combinations of sufficient conditions for guaranteeing persistence of excitation which can be applied, as appropriate, to the various algorithms.

An important observation about the local character of Theorem 2 is that for several of the algorithms (numbers 1, 4, 5, and 8), small θ_0 implies that $\mathbf{ML}^{-1}$ is close to being strictly passive. If θ_0 were actually zero, then the operator $\mathbf{ML}^{-1}$ would be the identity. Lemma 5 shows that for small perturbations around the equilibrium $\theta_0 = 0$, $\mathbf{ML}^{-1}$ remains passive.

In other algorithms (2, 3, 6, and 7), the size of θ_k does not influence the passivity of $\mathbf{ML}^{-1}$. Instead, a fixed filter $F(q^{-1}, k)$ is chosen to make $\mathbf{ML}^{-1}$ passive. It is, however, difficult to choose an appropriate filter without some a priori knowledge of A^* or C^*. Some recent results in this area may be found in [DB].

Another implementation issue (especially algorithms 1 and 5) is that stability of the regressor filter **L** must be maintained. This requires "projection" which monitors the stability of $\mathbf{L}(q^{-1}, p)$ at each timestep p. Lemma 3 assures that if each frozen $\mathbf{L}(q^{-1}, p)$ is EAS (and the stepsize is small), then the time-varying $\mathbf{L}(q^{-1}, k)$ will be EAS. The projection facility is undesirable because of its complexity and because of potential lock-up problems. See [LS].

7. Conclusion and Extensions

In the generic parameter update form (1.9) of Section 1, convergence (on average) of the parameter estimates occurs if the correction term is (on average) zero. With nonvanishing stepsizes, this occurs when the average of the product of the filtered regressor and the filtered prediction error is zero. The objective of this paper has been to find conditions that guarantee local stability about this solution point. This is akin to the objective (typical of adaptive filtering analysis) of proving the boundedness of the variance of the parameter estimate excursions about this average (or mean) solution point. These excursions do not vanish unless there is a parametrization that exactly zeros the filtered version of the prediction error, a situation which is unlikely to occur in any practical (nonideal) setting. Thus, proof of local stability about this solution point is one way to demonstrate desirable performance of the adaptive algorithm.

Average convergence of the parameter estimates in LMS in (1.1) occurs when

$$\operatorname{avg}[X_k X_k^T \theta_k] = 0, \tag{7.1}$$

where "avg" represents an averaging operation similar to the expectation operator used in the stochastic analysis of adaptive filters. A geometrical viewpoint interprets (7.1) as an average orthogonality condition on the parameter error θ_k and the regressor vector X_k. Equation (7.1) can also be interpreted as an implicit description of the desired average "solution" of the adaptive algorithm.

Actually solving (7.1) is nontrivial, especially when (7.1) is nonlinear, i.e., when the regressor is a function of θ_k. The incorporation of regressor filtering **L** and/or error filtering M changes the LMS form from (1.1) to the more general adaptive form (1.9). This can be viewed as altering (7.1) and thus changing the average solution sought by the adaptive algorithm. This may be beneficial since different applications may benefit from different solutions. Using the general form (1.9) changes the solution (7.1) to

$$\operatorname{avg}[\mathbf{L}(q^{-1}, k)\{X_k\} M(q^{-1}, k)\{X_k^T \theta_k\}] = 0. \tag{7.2}$$

This suggests two areas of investigation: (i) confirmation of the attraction and local stability properties of the solution implied by particular versions of (7.2) and (ii) connecting the various practical problems best solved by (7.2) with particular combinations of **L** and M. This paper falls in the first area by dealing with the local stability issue for a generic update term that encompasses a variety of adaptive algorithms, including LMS [WMLJ], SHARF [LTJ], Stearn's algorithm [F], RML [F], AML [S], and certain forms of recursive instrumental variables schemes [LS].

The local exponential stability of these adaptive algorithms was proven simultaneously by considering a generalized algorithmic framework with (time-varying) regressor filters and (time-varying) error filters. Several possible sets of sufficient conditions for stability were given in terms of persistence of excitaton, and the stability and passivity of certain frozen (time-invariant) filters.

This unified algorithmic framework may also be useful to facilitate generation of new algorithms in new application environments, and to analyze other algorithms which can be viewed as containing filtering of the regressor and error sequences. The basic results, for instance, retain their validity for nonlinear filters $\mathbf{L}(q^{-1}, k)$ and $M(q^{-1}, k)$ which are Lipschitz continuous, and for error sequences which consist of a sum of terms, each passed through a different filter. Investigation of these ideas is underway.

In ideal circumstances, each algorithm converges (under appropriate conditions) to the parameter value for which the prediction error is zero, and to a small ball in nonideal environments. Though Theorem 2 demonstrates the exponential stability of the various algorithms, it *does not show* that different algorithms converge to the same average value in nonideal use. The comments regarding (7.1) indicate that these convergent averages can be quite different, and the effect of various regressor and error filters on the convergent ball is an important area for further study.

Appendix A

Consider two different estimates of the output y_k, the a posteriori predicted output $z_{k+1} = \hat{\theta}_{k+1}^T X_k$ and the a priori predicted output $\hat{y}_{k+1} = \hat{\theta}_k^T X_k$. Let the a posteriori prediction error be $e_{k+1} = y_{k+1} - z_{k+1}$ and let v_{k+1} be a filtered version of e_{k+1},

$$v_{k+1} = (1 + N(q^{-1}, k))\{e_{k+1}\},$$

where $N(q^{-1}, k)$ represents the strictly causal part of the filtering. The a posteriori algorithm form can then be written

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu \mathbf{L}(q^{-1}, k)\{X_k\} v_{k+1}. \tag{A.1}$$

Note that z_{k+1}, and hence e_{k+1} and v_{k+1}, contain $\hat{\theta}_{k+1}$, and so (A.1) is an implicit equation in $\hat{\theta}_{k+1}$. This appendix shows that the unnormalized (and noncausal) a posteriori form (A.1) is the same as a normalized (and implementable) a priori update form (A.4). The development is a generalization of the approach in [J1]. Let

$$\begin{aligned}\bar{v}_k &= v_{k+1}(1 + \mu \mathbf{L}(q^{-1}, k)\{X_k^T\} X_k) \\ &= v_{k+1} + \mu \mathbf{L}(q^{-1}, k)\{X_k^T\} v_{k+1} X_k. \end{aligned} \tag{A.2}$$

Using (A.1), this becomes

$$= v_{k+1} + (\hat{\theta}_{k+1}^T - \hat{\theta}_k^T) X_k.$$

From the definitions of $\hat{y}_k$, z_k, e_k, and v_k, this is

$$\begin{aligned}&= (1 + N(q^{-1}, k))\{e_{k+1}\} + z_{k+1} - \hat{y}_{k+1} \\ &= y_{k+1} - \hat{y}_{k+1} + N(q^{-1}, k)\{e_{k+1}\}. \end{aligned} \tag{A.3}$$

Note that y_{k+1} is measurable *as*, and $\hat{y}_{k+1}$ is computable *before*, the parameter updates at time $k+1$ occur. Although $e_{k+1} = y_{k+1} - \hat{\theta}_{k+1}^T X_k$ is not available, the past values e_{k-i}, $i = 0, 1, 2, \ldots, n-1$, can be constructed. Since $N(q^{-1}, k)$ contains no direct feedthrough, $\bar{v}_k$ can be calculated before the parameter updates at time $k+1$. Equation (A.2) shows that $v_{k+1} = \bar{v}_k/(1 + \mu \mathbf{L}(q^{-1}, k)\{X_k^T\} X_k)$ and so the update (A.1) is equivalent to the implementable form

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \frac{\mu \mathbf{L}(q^{-1}, k)\{X_k\}}{1 + \mu \mathbf{L}(q^{-1}, k)\{X_k^T\} X_k} \bar{v}_k, \tag{A.4}$$

where $\bar{v}_k$ is defined by (A.3). To relate algorithm (A.1) to (1.9), recall that the error sequence e_{k+1} may equal $\hat{\theta}_{k+1}^T X_k$ (as in LMS or the equation error algorithm), in which case the filter $M(q^{-1}, k)$ of (1.9) is equal to $1 + N(q^{-1}, k)$. Often, however, the problem setup dictates that the measured error sequence contains a rational filtering $e_{k+1} = (1 + F(q^{-1}, k))\{\theta_{k+1}^T X_k\}$ as in output error or ARMAX problems. For this case, $M(q^{-1}, k)$ of (1.9) is equal to $(1 + N(q^{-1}, k))(1 + F(q^{-1}, k))$.

One potential problem with the a posteriori scheme (A.1) is evident in the normalization term of the equivalent implementable form (A.4). Since $\mathbf{L}(q^{-1}, k)\{X_k^T\} X_k$ can be negative, there is the possibility of division by zero. If a bound on X_k is known, then choosing

$$0 < \mu < \frac{1}{\|\mathbf{L}X\|_\infty \|X\|_\infty} - \varepsilon$$

for some positive ε guarantees that the update term is always bounded. Thus a small stepsize (where "small" is a function of the norm of the regressor and the norm of L) is required for the general a priori and a posteriori forms. This paper concentrates on the small stepsize a priori forms since they are more easily implemented, although the analysis extends to small stepsize a posteriori forms without difficulty.

To outline this extension, consider the a posteriori versions of Lemmas 1 and 2 which bound the difference between $M(q^{-1}, k)\{X_k^T \theta_{k+1}\}$ and $M(q^{-1}, k)\{X_k^T\}\theta_{k+1}$. The a posteriori version of (3.5) then has θ_{k+1} on the right-hand side which can be replaced (using Lemma 2) by $\theta_k + \Delta_2\{\theta_k\}$ where Δ_2 is proportional to μ. Stability of (3.6) then implies stability of the a posteriori version of (1.9). The rest of the analysis proceeds unchanged.

Appendix B. Proof of Lemma 4

Let $N_p(z) = D_p + C_p(zI - A_p)^{-1} B_p$ be associated with $N(q^{-1}, p)$ and assume that the minimal quadruples $\{\rho_2^{-1} A_p, B_p, \rho_2^{-1} C_p, d_p - \rho_1\}$ which define

$$\tilde{N}_p(z) = \bar{N}_p(\rho_2 z) - \rho_1 = (d_p - \rho_1) + \rho_2^{-1} C_p(zI - \rho_2^{-1} A_p)^{-1} B_p \tag{B.1}$$

are SPR for all p, with all eigenvalues of A_p in $|z| < \rho_2$. Associated with $N_p(z)$ is a unique minimum phase spectral factor

$$\tilde{W}_p(z) = 1 + \tilde{L}_p^T(zI - A_p)^{-1} B_p \tag{B.2}$$

and a unique positive $\tilde{Q}_p$ for which

$$\tilde{N}_p(z) + \tilde{N}_p(z^{-1}) = \tilde{W}_p(z^{-1}) \tilde{Q}_p \tilde{W}_p(z). \tag{B.3}$$

$\tilde{L}_p^T$ is unique since (A_p, B_p) is reachable. From the positive real lemma [L3] there exists a positive definite $\tilde{P}_p$ such that

$$\begin{aligned}\tilde{P}_p - \rho_2^{-2} A_p^T \tilde{P}_p A_p &= \tilde{L}_p \tilde{Q}_p \tilde{L}_p^T, \\ \rho_2^{-1} B_p^T \tilde{P}_p A_p + \tilde{Q}_p \tilde{L}_p^T &= C_p, \\ \tilde{Q}_p &= 2(d_p - \rho_1) - B_p^T \tilde{P}_p B_p. \end{aligned} \tag{B.4}$$

Note that there are many solution triples $\tilde{P}_p$, $\tilde{L}_p$, $\tilde{Q}_p$ of (B.4), but only one that is associated with the minimum phase spectral factor $\tilde{W}_p(z)$. Moreover, the $\tilde{P}_p$ satisfying (B.4) associated with $\tilde{W}_p(z)$ is minimal, see [FCG].

It is shown in [AG] that the minimum phase $\tilde{W}_p(z)$ with $\tilde{W}_p(\infty) = 1$ satisfying (B.3) obeys a continuity property: small $L_\infty[0, 2\pi]$ adjustments in $N(e^{j\omega})$ produce small $L_2[0, 2\pi]$ adjustments in $\tilde{W}_p(e^{j\omega})$ and small adjustments in $\tilde{Q}_p$. If small variations of $N(e^{j\omega})$ occur as a result of small variations in A_p, B_p, C_p, d_p, then the effect is to produce a small adjustment in $\tilde{L}_p$ as well as $\tilde{Q}_p$. Thus, $\tilde{L}_p$ and $\tilde{Q}_p$ depend continuously on A_p, B_p, C_p, d_p.

Because all the poles of $\tilde{N}_p(z)$ lie in $|z| < \rho_2$, all eigenvalues of $\rho_2^{-1}A_p$ lie in $|z| < \rho_2 < 1$. Accordingly, by reasoning similar to that used in studying (5.2), $\tilde{P}_p$ depends continuously on A_p, B_p, C_p, d_p. Now define

$$P_{k+1} = \tilde{P}_k, \qquad K_k = \tilde{Q}_k^{1/2}, \quad \text{and} \quad L_k^T = \rho_2 \tilde{Q}_k^{1/2} \tilde{L}_k^T. \tag{B.5}$$

Equations (B.4) then yield

$$\begin{aligned} P_k - A_k^T P_{k+1} A_k &= (1 - \rho_2^{-2}) P_k + (P_k - P_{k+1}) + L_k L_k^T, \\ B_k^T P_{k+1} A_k + K_k L_k^T &= C_k, \\ K_k^T K_k &= 2(d_k - \rho_1) - B_k^T P_{k+1} B_k. \end{aligned} \tag{B.6}$$

Since $\tilde{P}_p$ depends continuously on A_p, B_p, C_p, d_p, if there is some ε such that

$$\sup_k \{\|A_{k+1} - A_k\|, \|B_{k+1} - B_k\|, \|C_{k+1} - C_k\|, \|d_{k+1} - d_k\|\} < \varepsilon,$$

then

$$(1 - \rho_2^2) P_k + (P_k - P_{k+1}) = Q_k \geq 0.$$

By (2.3), this means that $N(q^{-1}, k) - \rho_1 I$ is passive, and consequently that $N(q^{-1}, k)$ is strictly passive.

References

[AB] B. D. O. Anderson, R. R. Bitmead, C. R. Johnson, Jr., P. V. Kokotovic, R. L. Kosut, I. M. Y. Mareels, L. Praly, and B. D. Riedle, *Stability of Adaptive Systems: Passivity and Averaging Analysis*, MIT Press, Cambridge, MA, 1986.

[AG] B. D. O. Anderson and M. Gevers, Identifiability of linear stochastic systems operating under linear feedback, *Automatica*, **18** (1982), 195–213.

[AJ] B. D. O. Anderson and C. R. Johnson, Jr., Exponential convergence of adaptive identification and control algorithms, *Automatica*, **18** (1982), 1–13.

[B] R. R. Bitmead, Persistence of excitation conditions and the convergence of adaptive schemes, *IEEE Trans. Inform. Theory*, **30** (1984), 183–191.

[BA] R. R. Bitmead and B. D. O. Anderson, Performance of adaptive estimation algorithms in dependent random environments, *IEEE Trans. Automat. Control*, **25** (1980), 788–794.

[BJ] R. R. Bitmead and C. R. Johnson, Jr., Discrete averaging principles and robust adaptive identification, in *Control and Dynamic Systems: Advances in Theory and Applications* (C. T. Leondes, ed.), Vol. 25, pp. 237–271, Academic Press, Orlando, FL, 1987.

[BS] S. Boyd and S. S. Sastry, On parameter convergence in adaptive control, *Systems Control Lett.*, **3** (1983), 311–319.

[DB] S. Dasgupta and A. S. Bhagwat, Conditions for designing strictly positive real transfer functions for adaptive output error identification, *IEEE Trans. Circuits and Systems*, **34** (1987), 731–736.

[D] C. A. Desoer, Slowly varying discrete system $x_{i+1} = A_i x_i$, *Electrons. Lett.*, **6** (1970), 339–340.

[DG] L. Dugard and G. C. Goodwin, Global convergence of Landau's "Output error with adjustable compensator" adaptive algorithm, *IEEE Trans. Automat. Control*, **30** (1985), 593–595.

[FCG] P. Faurre, M. Clerget, and F. Germain, *Operateurs Rationnels Positifs*, Dunod, Paris, 1979.

[F] B. Friedlander, System identification techniques for adaptive signal processing, *IEEE Trans. Acoust. Speech Signal Process.*, **30** (1982), 240–246.

[H] W. Hahn, *Stability of Motion*, Springer-Verlag, Berlin, 1967.

[J1] C. R. Johnson, Jr., A convergence proof for a hyperstable adaptive recursive filter, *IEEE Trans. Inform. Theory*, **25** (1979), 745–749.

[J2] C. R. Johnson, Jr., Adaptive IIR filtering: current results and open issues, *IEEE Trans. Inform. Theory*, **30** (1984), 237–250.

[JT] C. R. Johnson, Jr., and T. Taylor, Failure of a parallel adaptive identifier with adaptive error filtering, *IEEE Trans. Automat. Control*, **25** (1980), 1248–1250.

[KAM] R. L. Kosut, B. D. O. Anderson, and I. M. Y. Mareels, Stability theory for adaptive systems: methods of averaging and persistence of excitation, *Proceedings of the 24th IEEE Conference on Decision and Control*, Fort Lauderdale, FL, 1985, pp. 478–483.

[L1] I. D. Landau, Unbiased recursive identification using model reference adaptive techniques, *IEEE Trans. Automat. Control*, **21** (1976), 194–202.

[L2] I. D. Landau, Elimination of the real positivity condition in the design of parallel MRAS, *IEEE Trans. Automat. Control*, **23** (1978), 1015–1020.

[L3] I. D. Landau *Adaptive Control: The Model Reference Approach*, Marcel Dekker, New York, 1979.

[LTJ] M. G. Larimore, J. R. Treichler, and C. R. Johsnon, Jr., SHARF: An algorithm for adapting IIR digital filters, *IEEE Trans. Acoust. Speech Signal Process.*, **28** (1980), 428–440.

[L4] D. A. Lawrence, Adaptive System Stability Analysis via Energy Exchange, Ph.D. Thesis, Cornell University, June 1985.

[LJ] D. A. Lawrence and C. R. Johnson Jr., Recursive parameter identification algorithm stability analysis via π-sharing, *IEEE Trans. Automat. Control*, **31** (1986), 16–25.

[L5] L. Ljung, On positive real transfer functions and the convergence of some recursive schemes, *IEEE Trans. Automat. Control*, **22** (1977), 539–551.

[LS] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.

[L6] D. G. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1968.

[M] J. M. Mendel, *Discrete Techniques of Parameter Estimation: The Equation Error Formulation*, Marcel Dekker, New York, 1973.

[RPK] B. Riedle, L. Praly, and P. V. Kokotovic, Examination of the SPR condition in output error parameter estimation *Automatica*, **22** (1986), 495–498.

[SV] J. A. Sanders and F. Verhulst, *Averaging Methods in Nonlinear Dynamical Systems*, Springer-Verlag, New York, 1985.

[S] V. Solo, The convergence of AML, *IEEE Trans. Automat. Control*, **24** (1979), 958–962.

[TJL] J. R. Treichler, C. R. Johnson, Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters*, Wiley-Interscience, New York, 1987.

[V] M. Vidyasagar, *Nonlinear Systems Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1978.

[WMLJ] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., Stationary and nonstationary learning characteristics of the LMS adaptive filter, *Proc. IEEE*, **64** (1976), 1151–1162.