Adaptive IIR Filtering: Current Results and Open Issues

C. RICHARD JOHNSON, JR., SENIOR MEMBER, IEEE

Abstract—A tutorial-style framework is presented for understanding the current status of adaptive infinite-impulse-response (IIR) filters. The paper begins with a detailed discussion of the difference equation models that are useful as adaptive IIR filters. The particular form of the resulting prediction error generic to adaptive IIR filters is highlighted and the structures of provable convergent adaptive algorithms are derived. A brief summary of particular, currently known performance properties, drawn principally from the system identification literature, is followed by the formulation of three illustrative adaptive signal processing problems, to which these adaptive IIR filters are applicable. The concluding section discusses various open issues raised by the formulation of this framework.

I. Introduction

Adaptive infinite-impulse-response (IIR) filters are contemplated as replacements for adaptive finite-impulse-response (FIR) filters when the desired filter can be more economically modeled with poles and zeros than with the all-zero form of an FIR tapped-delay line. The possible benefits in reduced complexity and improved performance have spawned the adaptive IIR filter efforts of, e.g., [1]–[9]. A significant portion of these efforts has been spurred by the similarities of the adaptive IIR filtering problem and certain system identification problems [6]–[9]. The system identification literature has offered a rich trove of theoretical results, but, as we will see by the end of this paper, their translation to the adaptive IIR filtering problem reveals a number of important unanswered questions, several of which have not arisen in system identification studies.

This paper is intended to present a readily comprehendable progression from deterministic and stochastic identification-style modeling issues pertinent to the output error formulation underlying adaptive IIR filtering, through a broad categorization of algorithm forms and resulting convergence properties, to appropriate adaptive signal processing applications. This then permits the succinct statement of several open issues that form major future directions for adaptive IIR filter research. The hope is that those readers interested in adaptive signal processing will be encouraged to further beneficial examination of system identification and its rich literature. Conversely, hopefully those readers interested in system identification will recognize adaptive IIR filtering as an emerging source of numerous problems (and fixes) that challenge their skills. Such an objective requires promulgation of a cohesive, relatively tutorial perspective with just enough tantalizing detail to prompt the readers to continue their study with closer examination of the referenced sources.

The pedagogical structure developed in this paper relies on the detailed examination in the next section of the pertinent models and associated predictors yielding prediction errors suitable for use in adapting the predictor parameters. A prediction error structure generic to the output error formulation underlying adaptive IIR filters is carefully revealed as demonstrably different from the prediction error structure of the more familiar equation error formulation underlying adaptive FIR filters. The resulting adaptive output error algorithm modifications, relative to equation error forms, are produced in Section III from both minimization (via gradient descent) and stability theory viewpoints. Section IV briefly states a number of currently available convergence results for various modeling and adaptive algorithm combinations. Section V shows how these adaptive parameter estimation algorithms can be applied to provide adaptive IIR filters for three particular signal processing applications. (Those readers wishing to first examine the practical motivation of the parameter estimation basis of adaptive IIR filtering can turn directly to Section V for study of its representative applications and then return to Section II.) Section VI closes the paper with a discussion of various open issues raised by the preceding sections. Be aware that the originality of this paper lies as much in its unifying form as in its detailed substance and proceed accordingly.

II. Modeling and Prediction Error

This section focuses on autoregressive moving average with exogenous input (ARMAX) models and the resulting prediction errors, which underlie the adaptive parameter estimation view of adaptive filtering. The principal message is that the modeling choices that can be interpreted as useful as adaptive IIR filter structures share a particular prediction error characteristic that distinguishes them from the prediction errors attributable to models useful as adaptive FIR filters. Forming a meaningful succinct statement of this distinction and its adaptive algorithm formulation and performance consequences is difficult. In fact, the description of this distinction and its consequences is the major theme of this paper. In jargon that will be developed

Manuscript received May 24, 1983; revised September 7, 1983. This work was supported by the National Science Foundation under Grant NSF 8119312.

The author is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853.

0018-9448/84/0300-0237$01.00

in this and subsequent sections, this distinction can be described as the presence of an autoregressive (AR) filtering of the inner product of the parameter error and information vectors in the measurable prediction error associated with adaptive IIR filters. This AR filtering is absent in the prediction error of adaptive FIR filters. Following precedent terminology, models such as those for adaptive FIR filters which yield a prediction error that is an inner product of the parameter estimate error and information vectors (plus, possibly, a sample from an uncorrelated zero-mean sequence) are called equation error formulations. The output error formulation label is attached to models, such as those for adaptive IIR filters, which yield a measurable prediction error that is an AR filtered version of the inner product of the parameter estimate error and information vectors (again with the possible addition of an uncorrelated, zero-mean sequence sample). To make these statements understandable we will first review the modeling and subsequent prediction error of adaptive FIR filter forms and then expand this view to more general parameter estimation forms which we will discover are applicable as adaptive IIR filters.

The general linear process model within which we would like to estimate parameters will be restricted in this paper to the single-output, two-input difference equation, or ARMAX model,

$$y(k) = \sum_{i=1}^{n} [a_i y(k-i) + b_i u(k-i) + c_i w(k-i)] + w(k) \quad (2.1)$$

where y is the measurable scalar output, u the measurable scalar input, and w the unmeasurable scalar input typically considered to be a white, zero-mean sequence uncorrelated with $\{u(k)\}$. Note that the order n is the upper bound on the delay line lengths on the right of (2.1) and that various parameters can be zero for a particular process. A more compact form for (2.1) uses delay operator notation:

$$y(k) = A(q^{-1})y(k) + B(q^{-1})u(k) + C(q^{-1})w(k) + w(k), \quad (2.2)$$

where

$$A(q^{-1}) = a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n}, \quad (2.3)$$

$B(q^{-1})$ and $C(q^{-1})$ are similarly defined, and q^{-i} denotes a delay operation of i samples, i.e., $q^{-i}y(k) \equiv y(k-i)$. Note that each of these polynomial operators is defined without an undelayed term. For clarity this convention will be maintained throughout this paper. Use of this notational convention leads to the omission of a possible direct feedthrough term $b_0 u(k)$ on the right of (2.1). Such a term could be incorporated in the following with the concomitant notational complexity increase.

The process (2.2) is illustrated in block diagram form in Fig. 1 as the plant. The remainder of Fig. 1 illustrates a system identification style setup for estimating the plant parameters. A model is constructed, driven by the measurable signals, and parametrized, perhaps without $\hat{F}$ or $\hat{G}$ or $\hat{H}$ to reflect the model structural assumptions, such that the prediction error e is minimized. Given Fig. 1, minimizing e can be viewed as completely removing the portion of y due to the deterministic, measurable u and whitening the portion of y driven by the stochastic w. Thus e should equal w. From the block diagram with $\hat{F}$, $\hat{G}$, and $\hat{H}$ time-invariant, polynomial operators defined similarly to (2.3) we have that

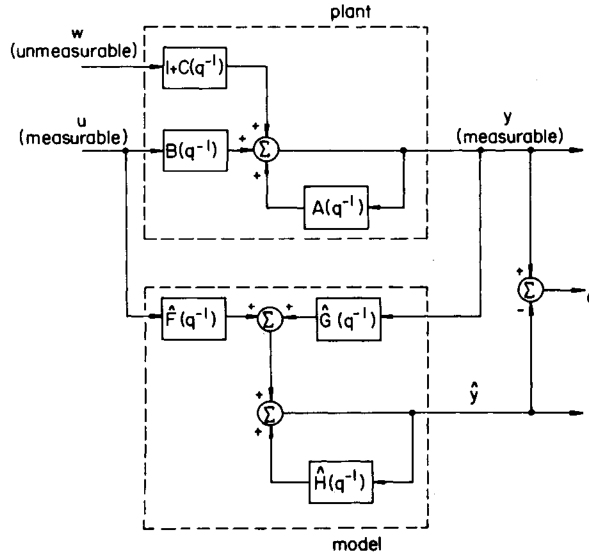

Fig. 1. General parameter estimation model.

$$e(k) = \left[\left\{\frac{B}{1-A}\right\}\left\{\frac{1-\hat{H}-\hat{G}}{1-\hat{H}}\right\} - \frac{\hat{F}}{1-\hat{H}}\right]u(k) + \left[\left\{\frac{1+C}{1-A}\right\}\left\{\frac{1-\hat{H}-\hat{G}}{1-\hat{H}}\right\}\right]w(k), \quad (2.4)$$

where the q^{-1} arguments of A, B, C, $\hat{F}$, $\hat{G}$, and $\hat{H}$ are suppressed for clarity. Thus $e(k) \equiv w(k)$ if $\hat{F} = B$, $\hat{G} = C + A$, and $\hat{H} = -C$. Note that if $w \equiv 0$, then $e \equiv 0$ if $\hat{F} = B$, $\hat{G} = A$, and $\hat{H} = 0$ or if $\hat{F} = B$, $\hat{G} = 0$, and $\hat{H} = A$. (These equivalence statements for e obviously assume appropriate or zero initial conditions. For arbitrary initial conditions $e \to w$. Stable cancellations are also assumed.) Thus different "structural constraints" in the model can lead to different parameterizations. Note that it is these various plausible designations of $\hat{F}$, $\hat{G}$, and $\hat{H}$ in Fig. 1 that keep us from replacing them with a single combination of $\hat{A}$, $\hat{B}$, and $\hat{C}$. Note that throughout this paper the carets will be used to denote estimated entities.

Parameter estimators use the value of the prediction error $e(k)$ in conjunction with the concurrent "information" in the available signal values $u(k)$, $y(k)$, and $\hat{y}(k)$ to improve the estimates $\hat{F}(q^{-1})$, $\hat{G}(q^{-1})$, and $\hat{H}(q^{-1})$. The form of the parameter estimate correction is strongly dependent on the functional relationship between the parameter estimate errors and the prediction error. Thus the remainder of this section will examine this relationship for various special cases of (2.4). This will be done algebraically rather than via manipulation of Fig. 1. However, reference to Fig. 1 will be made in each of the following special cases to emphasize the structural character of the underlying parameter estimation problem.

A. Equation Error Formulation

Consider the special case where $C(q^{-1}) \equiv 0$, which reduces (2.2) to

$$y(k) = A(q^{-1})y(k) + B(q^{-1})u(k) + w(k). \quad (2.5)$$

Since the white, unmeasurable w is unpredictable, (2.5) is "adequately" modeled by

$$\hat{y}(k) = \hat{A}(q^{-1})y(k) + \hat{B}(q^{-1})u(k) \quad (2.6)$$

with $\hat{A}$ and $\hat{B}$ dimensioned just as A and B as in (2.3), e.g.,

$$\hat{A}(q^{-1}) = \hat{a}_1 q^{-1} + \hat{a}_2 q^{-2} + \cdots + \hat{a}_n q^{-n}. \quad (2.7)$$

The model (2.6) is adequate in the sense that if $\hat{A} = A$ and $\hat{B} = B$ then the prediction error

$$e(k) = y(k) - \hat{y}(k) \quad (2.8)$$

is whitened and equals $w(k)$. Our interest is in the form of the prediction error of (2.8) which can be written as

$$e(k) = \tilde{A}(q^{-1})y(k) + \tilde{B}(q^{-1})u(k) + w(k) \quad (2.9)$$

where

$$\tilde{A}(q^{-1}) \triangleq A(q^{-1}) - \hat{A}(q^{-1}) \quad (2.10)$$

and $\tilde{B}$ is similarly defined. Throughout this paper the tilde will be used to denote the error of the estimated entities as in (2.10). For comparison with subsequent prediction error forms we will write (2.9) as

$$e(k) = [\theta - \hat{\theta}]^T X(k) + w(k) = \tilde{\theta}^T X(k) + w(k) \quad (2.11)$$

where

$$\theta = [a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n]^T \quad (2.12)$$

$$\hat{\theta} = [\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n, \hat{b}_1, \hat{b}_2, \cdots, \hat{b}_n]^T \quad (2.13)$$

and

$$X(k) = [y(k-1), y(k-2), \cdots, y(k-n), u(k-1), u(k-2), \cdots, u(k-n)]^T. \quad (2.14)$$

Note that the measurable prediction error in (2.11) is simply the inner product of the parameter estimate error vector $\tilde{\theta}$ and the information vector X plus a sample from an uncorrelated, zero-mean sequence, i.e., an equation error formulation.

Two special cases of (2.5) and (2.6) are of interest. The first is when A and thus $\hat{A}$ are constrained to zero and w is also absent. This clearly reduces (2.5) (and (2.6)) to a single-input (u), single-output (y), FIR form. Thus, if the parameter estimates $\hat{B}$ in (2.6) were recursively adapted, (2.6) would describe an adaptive FIR filter. With reference to Fig. 1, this special case uses $\hat{F} = \hat{B}$, $\hat{G} = 0$, and $\hat{H} = 0$.

The second special case of interest that also fits within the form of (2.5)–(2.6), and thus (2.11), is when $u \equiv 0$ and thus $B = \hat{B} = 0$. Note that this reduces (2.5) to an autoregressive (AR) filtering of white noise. Furthermore (2.6) becomes a single-input (y), single-output ($\hat{y}$) FIR filter. If $\hat{A}$ is recursively adapted an adaptive FIR filter results, the output of which, once $\hat{A}$ converges to A, can be subtracted from y to whiten the difference $y - \hat{y}$ to w. In Fig. 1 in this special case $\hat{G} = \hat{A}$, $\hat{F} = 0$, and $\hat{H} = 0$.

How to generate an adaptive algorithm for these two special cases or the two-input (y and u), single-output ($\hat{y}$) FIR model of (2.5) will be discussed in Section III. At this point it is sufficient to recognize that the prediction error of (2.11) suggests a least squares solution that will asymptotically provide unbiased parameter estimates due to the whiteness and zero-mean character of $\{w(k)\}$.

To relate (2.6) to Fig. 1 note that $\hat{F} \equiv \hat{B}$ and $\hat{G} \equiv \hat{A}$ while $\hat{H} \equiv 0$. Furthermore, since $\hat{H} = 0$, it is apparent from Fig. 1 that the model does not possess an IIR response. We shall see that the zeroing of $\hat{H}$ distinguishes the equation error formulation and its adaptive FIR filter applicability from the so-called output error formulation, (where $\hat{H} \neq 0$), and its adaptive IIR filter applicability.

B. Output Error Formulation

If instead, (2.5) is modeled via

$$\hat{y}(k) = \hat{A}(q^{-1})\hat{y}(k) + \hat{B}(q^{-1})u(k), \quad (2.15)$$

a single-input (u), single-output ($\hat{y}$) IIR model results. The prediction error (2.8) associated with (2.15) arises from subtracting (2.15) from (2.5):

$$\begin{aligned} e(k) = {} & A(q^{-1})[y(k) - \hat{y}(k)] \\ & + [A(q^{-1}) - \hat{A}(q^{-1})]\hat{y}(k) \\ & + [B(q^{-1}) - \hat{B}(q^{-1})]u(k) + w(k) \quad (2.16) \end{aligned}$$

or

$$[1 - A(q^{-1})]e(k) = \tilde{A}(q^{-1})\hat{y}(k) + \tilde{B}(q^{-1})u(k) + w(k) \quad (2.17)$$

or

$$e(k) = [1 - A(q^{-1})]^{-1}[\tilde{\theta}^T X(k) + w(k)] \quad (2.18)$$

where

$$\tilde{\theta} = [a_1 - \hat{a}_1, \cdots, a_n - \hat{a}_n, b_1 - \hat{b}_1, \cdots, b_n - \hat{b}_n]^T \quad (2.19)$$

and

$$X(k) = [\hat{y}(k-1), \cdots, \hat{y}(k-n), u(k-1), \cdots, u(k-n)]^T. \quad (2.20)$$

Compare (2.18) with (2.11). Note that, while the prediction error of (2.11), associated with the two-input, single-output model of (2.5) in (2.6), is the inner product of the associated parameter estimate error and information vectors plus white, zero-mean noise sample w, the prediction error of (2.18), associated with the single-input, single-output model of (2.5) in (2.15), is an AR filtered version of the inner product of the associated parameter estimate error and information vectors plus a similar w. The notation of (2.18) indicates that the prediction error e is the "output" of a system with transfer function $[1 - A]^{-1}$, which is purely AR, driven by the "input" $\tilde{\theta}^T X + w$. Note that the poles of this transfer function are those of the actual process in (2.5), which we are trying to estimate. Finally, note that the information vectors of (2.11) and (2.18) differ. For (2.11) the information vector of (2.14) includes past y and u; for (2.18) the information vector of (2.20) is composed from past $\hat{y}$ and u. With reference to Fig. 1, (2.15) associates $\hat{F}$ with $\hat{B}$ and $\hat{H}$ with $\hat{A}$ while $\hat{G}$ is constrained to zero.

In modeling the full ARMAX process in (2.2) we cannot measure w in order to append the term $\hat{C}(q^{-1})w(k)$ to (2.5) or (2.15). Note that if we assume that a model, to be defined, generating $\hat{y}$ can be refined (or adapted) such that $e = y - \hat{y} = w$, we would be inclined to add instead the term $\hat{C}(q^{-1})e(k)$ as in

$$\hat{y}(k) = \hat{A}(q^{-1})y(k) + \hat{B}(q^{-1})u(k) + \hat{C}(q^{-1})e(k). \tag{2.21}$$

The validity of this use of e, in a sense for w, is provided by the ability, discussed in Section IV, to adapt the estimates in (2.21) such that $e \to w$. The prediction error from subtracting (2.21) from (2.2) is

$$e(k) = \tilde{A}(q^{-1})y(k) + \tilde{B}(q^{-1})u(k) + \tilde{C}(q^{-1})e(k) - C(q^{-1})[e(k) - w(k)] + w(k) \tag{2.22}$$

or

$$e(k) = [1 + C(q^{-1})]^{-1}[\tilde{\theta}^T X(k)] + w(k) \tag{2.23}$$

where

$$\tilde{\theta} = [a_1 - \hat{a}_1, \cdots, a_n - \hat{a}_n, b_1 - \hat{b}_1, \cdots, b_n - \hat{b}_n, c_1 - \hat{c}_1, \cdots, c_n - \hat{c}_n]^T \tag{2.24}$$

and

$$X(k) = [y(k-1), \cdots, y(k-n), u(k-1), \cdots, u(k-n), e(k-1), \cdots, e(k-n)]^T. \tag{2.25}$$

The IIR form of (2.21) is implicit. Since $e = y - \hat{y}$, (2.21) can be rewritten as

$$\hat{y}(k) = -\hat{C}(q^{-1})\hat{y}(k) + \hat{B}(q^{-1})u(k) + [\hat{C}(q^{-1}) + \hat{A}(q^{-1})]y(k), \tag{2.26}$$

which is a two-input (u and y), single-output ($\hat{y}$) IIR form with characteristic polynomial $1 + C$. This characteristic polynomial, which is the to-be-estimated noise MA polynomial in (2.2), is also the polynomial in (2.23) autoregressively filtering the parameter error vector $\tilde{\theta}$ inner product with the information vector X. This mimics the similar pattern of (2.18). With reference to Fig. 1, (2.21), or equivalently (2.26), associates $\hat{F}$ with $\hat{B}$, $\hat{G}$ with $\hat{C} + \hat{A}$, and $\hat{H}$ with $-\hat{C}$. Note that both (2.15) and (2.21) result in $\hat{H} \neq 0$, which clearly provides IIR modeling.

A special case of interest here is when $u \equiv 0$ and thus $\hat{B} \equiv 0$. In this case (2.21) becomes

$$\hat{y}(k) = \hat{A}(q^{-1})y(k) + \hat{C}(q^{-1})e(k) \tag{2.27}$$

which, given that $e = y - \hat{y}$ can also be written as

$$\hat{y}(k) = \hat{A}(q^{-1})[\hat{y}(k) + e(k)] + \hat{C}(q^{-1})e(k). \tag{2.28}$$

These two special cases of (2.21), both of which result in (2.23) with the $\tilde{b}_i$ removed from $\tilde{\theta}$ and the $u(k-i)$ removed from $X(k)$, will prove directly useful in the applications of Section V. However, the most significant observation at this point is of the addition of AR filtering in the measurable prediction error of the IIR models relative to the prediction errors of FIR models.

III. Algorithm Forms

The intent of this section is to illustrate convincingly the additions to more well-known equation error based parameter estimation algorithms, useful for establishing adaptive FIR filter algorithms, in order to yield less widely known output error based parameter estimation algorithms, useful for establishing adaptive IIR filter algorithms. Understanding the subsequent alterations in adaptive parameter estimation algorithms from the equation error to the output error case requires reference to the equation error solution. Thus the equation error problem will be addressed via each of two generic approaches to deterministic ($w \equiv 0$) algorithm development, each followed by extension to the output error case. The reason for considering only two of the many approaches to algorithm establishment is that, in the output error case, these two approaches will establish the generic structure of the two most widely studied solutions to the output error formulation. This bifurcation is somewhat surprising, since in the equation error case both approaches yield the same basic algorithm form. As we shall see, it consists of correcting the old parameter estimates to the new parameter estimates by addition of a term composed of the product of a bounded step-size term (possibly a time-varying matrix), the information vector, and the scalar prediction error. The two modifications in the output error case are essentially either a fixed moving-average (MA) filtering of the prediction error or a time-varying AR filtering of the information vector. Both of these "fixes" can (and will) be interpreted as attempts to counter the AR filtering present in the output error form prediction error but absent in the equation error form. Once trivialized support for these additions is provided, a number of references will be cited for more complete development of these generic results.

A. Minimization Approach

Consider the equation error formulation of (2.5)–(2.14) with $w \equiv 0$. Thus

$$y(k) = \theta^T X(k), \tag{3.1}$$

$$\hat{y}(k) = \hat{\theta}^T(k-1)X(k), \tag{3.2}$$

$$\hat{\theta}^T(k) = [\hat{a}_1(k), \cdots, \hat{a}_n(k), \hat{b}_1(k), \cdots, \hat{b}_n(k)], \tag{3.3}$$

and θ and $X(k)$ are as in (2.12) and (2.14), respectively. The measurable *a priori* prediction error is

$$e(k) = \tilde{\theta}^T(k-1)X(k). \tag{3.4}$$

This prediction error between y and $\hat{y}$ at time k is termed *a priori* since the parameter estimate vector of the previous time instant $\hat{\theta}(k-1)$ is used to form $\hat{y}(k)$.

A common approach to adaptive parameter estimation algorithm establishment is to invoke a strategy that minimizes the squared prediction error $e^2(k)$. Admittedly for this deterministic problem, with the accumulation of enough data an exact solution exists based on matrix inversion. However, our purpose here is instructive, so we will consider a simple gradient descent based solution

$$\hat{\theta}(k) = \hat{\theta}(k-1) - \mu \frac{\delta e^2(k)/2}{\delta \hat{\theta}(k-1)}, \tag{3.5}$$

where μ is a small positive step size to be taken in the correction of $\hat{\theta}$ in the direction opposite to the upward (positive) partial derivative of e^2 with respect to the current $\hat{\theta}$. Recall that $e(k) = y(k) - \hat{y}(k)$ and that y and u and therefore X are not functions of $\hat{\theta}$. Thus

$$\frac{\delta e(k)}{\delta \hat{\theta}(k-1)} = \frac{-\delta \hat{y}(k)}{\delta \hat{\theta}(k-1)} = -X(k) \tag{3.6}$$

from (3.2). Therefore (3.5) becomes the familiar LMS-type adaptive algorithm [10]

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \mu X(k) e(k). \tag{3.7}$$

Note the form of the correction term in (3.7) as the product of a step-size μ, the information vector X, and the prediction error e. If we had pursued a recursive solution of the matrix formulation of the least squares problem, we would have generated the same structure as in (3.7) except that the scalar μ would be a time-varying matrix [11].

Now consider repeating this gradient descent approach for the corresponding output error formulation with $w \equiv 0$ of (2.5), (2.15)–(2.20) where (3.2) and (3.3) apply but $X(k)$ is given by (2.20). The prediction error is as in (2.18) with $w \equiv 0$

$$e(k) = \left[1 - A(q^{-1})\right]^{-1}\left[\tilde{\theta}^T(k-1)X(k)\right]. \tag{3.8}$$

As before neither y nor u are functions of $\hat{\theta}$ so

$$-\frac{\delta e(k)}{\delta \hat{\theta}(k-1)} = \frac{\delta \hat{y}(k)}{\delta \hat{\theta}(k-1)} = \frac{\delta\left[\hat{\theta}^T(k-1)X(k)\right]}{\delta \hat{\theta}(k-1)}. \tag{3.9}$$

Now however, since $X(k)$ includes past $\hat{y}$ which are dependent on past $\hat{\theta}$ which are used to form new $\hat{\theta}$, a portion of $X(k)$ is not independent of $\hat{\theta}$. So the transpose of the right equation in (3.9) is

$$\begin{aligned}&\left[\frac{\delta \hat{y}(k)}{\delta \hat{a}_1(k-1)}, \cdots, \frac{\delta \hat{y}(k)}{\delta \hat{a}_n(k-1)}, \frac{\delta \hat{y}(k)}{\delta \hat{b}_1(k-1)}, \cdots, \frac{\delta \hat{y}(k)}{\delta \hat{b}_n(k-1)}\right] \\ &= X^T(k) + \sum_{i=1}^{n} \hat{a}_i(k-1) \\ &\cdot \left[\frac{\delta \hat{y}(k-i)}{\delta \hat{a}_1(k-1)}, \cdots, \frac{\delta \hat{y}(k-i)}{\delta \hat{a}_n(k-1)}, \frac{\delta \hat{y}(k-i)}{\delta \hat{b}_1(k-1)}, \cdots, \frac{\delta \hat{y}(k-i)}{\delta \hat{b}_n(k-1)}\right]. \end{aligned} \tag{3.10}$$

Note that (3.10) indicates the necessity of reevaluation of the derivatives of past $\hat{y}$ with respect to current parameter estimates. A simplifying assumption commonly made in the adaptive IIR filtering literature [2], [4], [5] is that μ is sufficiently small such that $\hat{\theta}(k-1) \cong \hat{\theta}(k-2) \cong \cdots \cong \hat{\theta}(k-n-1)$, which suggests replacement of (3.10) by

$$\frac{\delta \hat{y}(k)}{\delta \hat{\theta}(k-1)} \cong X(k) + \sum_{i=1}^{n} \hat{a}_i(k-1)\left[\frac{\delta \hat{y}(k-i)}{\delta \hat{\theta}(k-i-1)}\right] \tag{3.11}$$

or

$$\frac{\delta \hat{y}(k)}{\delta \hat{\theta}(k-1)} - \sum_{i=1}^{n} \hat{a}_i(k-1)\left[\frac{\delta \hat{y}(k-i)}{\delta \hat{\theta}(k-i-1)}\right] \cong X(k). \tag{3.12}$$

Note that (3.12) provides a recursive method of approximating $\delta \hat{y}(k)/\delta \hat{\theta}(k-1)$. Defining the operator

$$\hat{A}(q^{-1}, k-1) \triangleq \hat{a}_1(k-1)q^{-1} + \hat{a}_2(k-1)q^{-2} + \cdots + \hat{a}_n(k-1)q^{-n}, \tag{3.13}$$

similarly to $A(q^{-1})$ in (2.3) converts (3.12) to $[1 - \hat{A}(q^{-1}, k-1)][\delta \hat{y}(k)/\delta \hat{\theta}(k-1)] \cong X(k)$ and suggests the definition of

$$\psi(k) \triangleq \left[1 - \hat{A}(q^{-1}, k-1)\right]^{-1} X(k), \tag{3.14}$$

where $\psi(k)$ can be considered an approximation of $\delta \hat{y}(k)/\delta \hat{\theta}(k-1)$. As an aside, note that $[1 - \hat{A}(q^{-1}, k-1)]^{-1}$ provides a shorthand operator notation for the difference equation form of (3.14), which is similar to that of (3.12); $\hat{A}(q^{-1}, k-1)$ does not indicate a z transform, which is why the time operator q^{-1} is used rather than the transform operator z^{-1}. The form of ψ in (3.14) requires a significant computational burden since n past values of the full $2n \times 1$ ψ vector must be stored and each element of ψ updated independently via an nth order AR. Recognizing that $X(k)$ in (2.20), which drives the propagation of ψ in (3.14), is composed of successively delayed versions of $\hat{y}$ and u suggests [5], [12] a more computationally efficient approximation for $\delta \hat{y}(k)/\delta \hat{\theta}(k-1)$. This alternate approximation uses filtered versions of $\hat{y}$ and u via

$$\hat{y}^F(k) = \left[1 - \hat{A}(q^{-1}, k-1)\right]^{-1} \hat{y}(k) \tag{3.15}$$

$$u^F(k) = \left[1 - \hat{A}(q^{-1}, k-1)\right]^{-1} u(k) \tag{3.16}$$

and composes $\bar{\psi}$ from past values

$$\bar{\psi}(k) \triangleq \left[\hat{y}^F(k-1), \hat{y}^F(k-2), \cdots, \hat{y}^F(k-n), u^F(k-1), \cdots, u^F(k-n)\right]^T. \tag{3.17}$$

That (3.17) is not equivalent to (3.14) can be noted by comparing the second entry ψ_2 in (3.14)

$$\psi_2(k) = \hat{y}(k-2) + \sum_{i=1}^{n} \hat{a}_i(k-1)\psi_2(k-i) \tag{3.18}$$

and the second entry in (3.17)

$$\hat{y}^F(k-2) = \hat{y}(k-2) + \sum_{i=1}^{n} \hat{a}_i(k-2)\hat{y}^F(k-i-2). \tag{3.19}$$

Even if the past values $\psi_2(k-i)$ and $\hat{y}^F(k-i-2)$ were equal, $\psi_2(k)$ need not equal $\hat{y}^F(k-2)$ due to the difference in the time indices on the $\hat{a}_i$. However, under the assumption used to generate (3.12), i.e., the $\hat{\theta}$ and thus $\hat{a}_i$ are slowly time-varying, these two approximations ψ and $\bar{\psi}$ for $\delta\hat{y}(k)/\delta\hat{\theta}(k-1)$ are essentially interchangeable.

From (3.5) the adaptive algorithm is

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \mu\psi(k)e(k). \qquad (3.20)$$

Admittedly, the construction of (3.14) and (3.20) does not constitute a proof of its convergence, but it does succinctly capture one of the generic additions suggested in the output error case, i.e., the AR filtering of the information vector by the current estimate of the AR appearing in the associated prediction error, (3.8) in this case. Note that in the equation error case, represented by (3.4) and (3.7),

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \mu\{X(k)\}\{X^T(k)\tilde{\theta}(k-1)\}; \qquad (3.21)$$

whereas, in the output error case, represented here by (3.8), (3.14), and (3.20),

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \mu\left\{\left[1-\hat{A}(q^{-1},k-1)\right]^{-1}X(k)\right\} \cdot\left\{\left[1-A(q^{-1})\right]^{-1}\left[X^T(k)\tilde{\theta}(k-1)\right]\right\}. \qquad (3.22)$$

Thus the time-varying AR filtering of the information vector in (3.22) can be viewed relative to (3.21) as "compensation" for the fixed AR filtering of $\tilde{\theta}^T X$ in the measurable prediction error. This fix is more common in the stochastic setting of (2.21) and (2.25), where the time-varying AR filtering of the information vector is constructed from the estimate of the noise MA polynomial $[1+C]$ that appears as the AR filtering of $\tilde{\theta}^T X$ in forming the prediction error in (2.23). The minimization approximation procedure resulting in (3.14) and (3.20) with $[1+\hat{A}]^{-1}$ replaced by $[1+\hat{C}]^{-1}$ and X appropriately redefined as in (2.25) is much more complicated in this full ARMAX case. Refer to [12]–[13]. Convergence proof is commonly based on the ODE method of [12]–[14]. Some of the proven properties of this solution are discussed in Section IV. The most critical is that in order to retain parameter estimator stability, the roots of $1+\hat{C}$ (or $1-\hat{A}$ in the deterministic case of (3.14)) must be constrained within the unit circle. This is needed to insure that ψ does not become undesirably unbounded. This requires an on-line stability check of $[1+\hat{C}]^{-1}$ (or $[1-\hat{A}]^{-1}$ in the deterministic case) and, when instability is encountered, a projection of $\hat{C}$ (or $\hat{A}$) back into the region of the parameter space where $[1+\hat{C}]^{-1}$ (or $[1-\hat{A}]^{-1}$) is stable. This obviously adds an undesirable computational burden.

B. Stability Theory Approach

An alternate approach to algorithm development arises from a stability theory interpretation. Reconsider the equation error formulation of (3.1)–(3.4). Subtracting both sides of the adaptive algorithm in (3.7) from a time-invariant θ yields, given (3.4),

$$\tilde{\theta}(k) = \left[I - \mu X(k)X^T(k)\right]\tilde{\theta}(k-1). \qquad (3.23)$$

The stability theory approach to adaptive parameter estimation is based on recognizing that (3.23) represents an unforced, time-varying, possibly nonlinear (when X is a function of $\hat{\theta}$ and therefore $\tilde{\theta}$) system, the zero state stability of which represents the desired convergence property of (3.7). In other words, showing that $\hat{\theta} \to \theta$, or equivalently that $\tilde{\theta} \to \mathbf{0}$, for any finite $\tilde{\theta}(0)$ can be done by proving (3.23) to be globally, asymptotically stable. A similar stability theory formulation exists in expectation when $\{X\}$ is stochastic rather than deterministic. Applicable stability theorems abound under the labels of Lyapunov stability theory [15], [16] and hyperstability [6], [17] in the deterministic case and ODE analysis [12], [18], [19] and martingale convergence theory [20] in the stochastic case.

For the error system of (3.23) consider the Lyapunov function candidate of summed squared parameter estimate errors

$$V(k) = \tilde{\theta}^T(k)\tilde{\theta}(k). \qquad (3.24)$$

If

$$\Delta V(k) = V(k) - V(k-1) \le 0 \qquad (3.25)$$

for all k and $V(0)$ is finite, then $\Delta V \to 0$ [15]. Evaluating (3.25) given (3.4), (3.23) and (3.24) yields [15]

$$\Delta V(k) = -\mu e^2(k)\left[2 - \mu X^T(k)X(k)\right]. \qquad (3.26)$$

Since μ is defined as positive, if

$$0 < \mu < \frac{2-\sigma}{X^T(k)X(k)}, \qquad \text{for all } k \text{ and some } \sigma\epsilon(0,2). \qquad (3.27)$$

then (3.25) is satisfied and $\Delta V \to 0$ implies that $\mu e^2(k) \to 0$ or

$$e(k) = \tilde{\theta}^T(k-1)X(k) \to 0. \qquad (3.28)$$

Note that (3.27) implicitly assumes that $X(k)$ is bounded or that u is bounded and $[1-A]^{-1}$ is stable. Furthermore note that (3.28) does not imply that $\tilde{\theta} \to 0$ unless $\{X(k)\}$ is sufficiently rich such that a nonzero $\tilde{\theta}$ is not orthogonal to $X(k)$ for all $k > \bar{k}$. The information vector X can be interpreted as being sufficiently rich to excite every mode of the plant such that errors in identifying any of these models are observable in the prediction error. This nonorthogonality condition can be shown [21] to be guaranteed if the smallest eigenvalue of $\sum_{k=j}^{j+S} X(k)X^T(k)$ is bounded away from zero for all j and some S greater than the dimension of X, i.e., for some positive scalar ρ

$$\sum_{k=j}^{j+S} X(k)X^T(k) \ge \rho I > 0, \qquad \text{for all } j \text{ and some } S > \dim(X). \qquad (3.29)$$

Similar conditions exist in expectation for stochastic $\{X\}$ [22]. The "persistent" excitation condition of (3.29) requires that over any time window of S consecutive samples enough information exists in $\sum XX^T$ to solve by bulk pseudoinversion the matrix reformulation of the parameter estimation task. This persistency will prove critical in allowing parameter tracking with unpredictable plant parameter changes. Given (3.29), (3.23) is globally asymptotically stable and $\tilde{\theta}(k) \to \mathbf{0}$.

In an attempt to guarantee (3.27) consider replacing the constant μ in (3.7) by the time-varying $\alpha/[1 + \gamma X^T(k)X(k)]$. Since

$$\frac{\alpha}{\gamma^{-1} + X^T(k)X(k)} < \frac{2}{X^T(k)X(k)} \tag{3.30}$$

with $\infty > \gamma > 0$ and $0 < \alpha < 2$,

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{\alpha\gamma X(k)e(k)}{1 + \gamma X^T(k)X(k)} \tag{3.31}$$

has the property from (3.26), where $\alpha/[\gamma^{-1} + X^T(k)X(k)]$ replaces μ, that

$$\frac{\alpha e^2(k)}{\gamma^{-1} + X^T(k)X(k)} \to 0, \tag{3.32}$$

since the bracketed term in (3.26) is always positive. If X is bounded or can grow no faster than e [23], then $e \to 0$, and, if in addition (3.29) is satisfied, $\tilde{\theta} \to \mathbf{0}$. Note that given (3.4) and (3.31) the *a posteriori* prediction error $\epsilon(k)$, which uses the current parameter estimate error vector $\tilde{\theta}(k)$ rather than its previous value $\tilde{\theta}(k-1)$ as in the *a priori* prediction error in (3.4), becomes

$$\begin{aligned} \epsilon(k) &\triangleq X^T(k)\tilde{\theta}(k) \\ &= X^T(k)\tilde{\theta}(k-1) - \frac{\alpha\gamma X^T(k)X(k)e(k)}{1+\gamma X^T(k)X(k)} \\ &= \left[1 + \gamma(1-\alpha)X^T(k)X(k)\right] \\ &\quad \cdot \left[1 + \gamma X^T(k)X(k)\right]^{-1} e(k). \end{aligned} \tag{3.33}$$

Thus with $\alpha \equiv 1$, (3.31) can be written as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \gamma X(k)\epsilon(k). \tag{3.34}$$

Note that (3.34) is an implicit formula for $\hat{\theta}(k)$ since computation of $\epsilon(k)$ from (3.33) as

$$\epsilon(k) = y(k) - \hat{\theta}^T(k)X(k) \tag{3.35}$$

requires $\hat{\theta}(k)$. However, the form of (3.34) will prove useful for extension to the output error case.

The pertinent stability theory result for the output error case from hyperstability theory [6] or Lyapunov stability theory [16] is that

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \gamma X(k)v(k), \tag{3.36}$$

where

$$v(k) = \mathscr{H}(q^{-1})\left[\tilde{\theta}^T(k)X(k)\right], \tag{3.37}$$

is stable if the transfer function $\mathscr{H}(q^{-1})$ is strictly positive real (SPR), i.e.,

$$\operatorname{Re}\left[\mathscr{H}(e^{-j\omega})\right] > 0, \qquad \text{for all } \omega \in [0, 2\pi]. \tag{3.38}$$

Furthermore given the boundedness of u, the stability of $[1 - A]^{-1}$ and (3.29), $\tilde{\theta} \to \mathbf{0}$. For the equation error case of (3.34) $\mathscr{H} \equiv 1$, which is trivially SPR. For the output error formulation, the *a posteriori* prediction error, similar to (3.8), is

$$\epsilon(k) = \left[1 - A(q^{-1})\right]^{-1}\left[\tilde{\theta}^T(k)X(k)\right] \tag{3.39}$$

and

$$\mathscr{H}(q^{-1}) = \left[1 - A(q^{-1})\right]^{-1}. \tag{3.40}$$

Thus, if (3.40) is SPR, the output error formulation of (3.34) will be stable. As shown in [7], that (3.40) is SPR is unlikely even in simple low-order cases. This raises the possibility of the second fix. Consider the adaptive parameter estimator of (3.36) with

$$\begin{aligned} v(k) &= \left[y(k) - \hat{\theta}^T(k)X(k)\right] \\ &\quad - \sum_{i=1}^{n} d_i\left[y(k-i) - \hat{\theta}^T(k-i)X(k-i)\right] \\ &= \left[1 - D(q^{-1})\right]\epsilon(k) \\ &= \frac{\left[1 - D(q^{-1}\right]}{\left[1 - A(q^{-1})\right]}\left[\tilde{\theta}^T(k)X(k)\right]. \end{aligned} \tag{3.41}$$

In this case, with reference to (3.37),

$$\mathscr{H}(q^{-1}) = \frac{\left[1 - D(q^{-1})\right]}{\left[1 - A(q^{-1})\right]}, \tag{3.42}$$

which is SPR if D is sufficiently close to A. If (3.42) is SPR, then using v from (3.41) in (3.36) will result in $\tilde{\theta} \to \mathbf{0}$ given (3.29) as stated for the equivalent (3.36)–(3.38). The implicit form of (3.36) using (3.41) can be made explicit by noting that

$$\begin{aligned} v(k) &= y(k) - \left[\hat{\theta}^T(k-1) + \gamma X^T(k)v(k)\right]X(k) \\ &\quad - \sum_{i=1}^{n} d_i\epsilon(k-i) \\ &= \left[1 + \gamma X^T(k)X(k)\right]^{-1}\left[e(k) - \sum_{i=1}^{n} d_i\epsilon(k-i)\right], \end{aligned} \tag{3.43}$$

with the usual *a priori* output estimate (or prediction) error $e(k) = y(k) - \hat{\theta}^T(k-1)X(k)$ and the *a posteriori* output estimate (or prediction) error $\epsilon(k)$ in (3.35), and using (3.43) for (3.41).

Obviously choosing a satisfactory D presents a practical difficulty due to its dependence on the unknown A. But, *if* D is satisfactorily selected such that (3.42) is SPR, no stability check is required as for the first fix as illustrated in (3.22). Comparing the equation error solution of (3.34) rewritten as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \gamma\{X(k)\}\{X^T(k)\tilde{\theta}(k)\} \tag{3.44}$$

and the corresponding output error solution of (3.36) rewritten as

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \gamma\{X(k)\} \cdot \left\{\frac{\left[1 - D(q^{-1})\right]}{\left[1 - A(q^{-1})\right]}\left[X^T(k)\tilde{\theta}(k)\right]\right\} \tag{3.45}$$

clearly reveals the character of the present "compensation." A complete convergence proof for (3.45) when applied to the deterministic identification problem of (2.5) (with $w \equiv 0$), (3.2), (3.3), and (2.20) appears in [6] and [17]. If, rather than a scalar constant γ as in (3.45), a time-varying, possibly matrix, step size is used, the SPR condition on (3.42) is varied such that $\mathscr{H} - \lambda$ must be SPR [24]. The positive constant λ is dependent on various choices of designer-selected parameters within the step-size formula [24]. Note that with $\lambda > 0$, requiring $\mathscr{H} - \lambda$ to be SPR is more severe than requiring $\mathscr{H}$ alone to be SPR.

Reliance on the SPR of the filtering of $X^T(k)\tilde{\theta}(k)$ in the prediction error in (2.23) used in the update term has also appeared in the stochastic setting. For example, in extended least squares [12] or approximate maximum likelihood [20], where no *a posteriori* error smoothing is added and vanishing, ergo time-varying, step sizes are used, $[1 + C(q^{-1})]^{-1} - \lambda$ is required to be SPR. In modeling these stochastic identification problems X includes past output estimate errors. (Refer to (2.25).) These convergence proofs require the use of past *a posteriori* estimates, rather than *a priori*, in X. The generic observation is that the denominator of $\mathscr{H}$ in (3.37) for the output error formulation always includes the unknown AR filtering of $\tilde{\theta}^T X$ forming the prediction error $y - \hat{y}$ as detailed in Section II.

Thus we can characterize solutions to the output error problem, where the measurable prediction error is an AR filtered version of the parameter estimate error and information vector inner product possibly plus a white, zero-mean signal, by their use of either a) AR filtering, by the estimate of the AR in the prediction error, of the information vector or b) MA filtering, satisfying an SPR condition, of the *a posteriori* prediction (or output) error in the general algorithm form

$$\begin{bmatrix}\text{new}\\ \text{parameter}\\ \text{estimate}\end{bmatrix} = \begin{bmatrix}\text{old}\\ \text{parameter}\\ \text{estimate}\end{bmatrix} + \begin{bmatrix}\text{bounded}\\ \text{step}\\ \text{size}\end{bmatrix} \cdot \begin{bmatrix}\text{function}\\ \text{of the}\\ \text{information}\\ \text{vector}\end{bmatrix} \cdot \begin{bmatrix}\text{function}\\ \text{of the}\\ \text{prediction}\\ \text{error}\end{bmatrix}. \quad (3.46)$$

IV. Algorithm Convergence Properties

The categorizations arising only from the distinctions based on the various models from Section II and the two algorithm "compensation" schemes of the preceding section are numerous. Add to this the various techniques for selecting the step size as fixed or time-varying, as scalar or matrix, the prediction errors used as a combination of *a priori* and past *a posteriori* prediction errors, etc., and the bewildering array of "different" adaptive algorithms populating the literature is explained. Attempting to list the convergence properties of even the subset applicable to what we have termed an output error type problem becomes monumental. Instead we will highlight only a few broad classes of interest that have been widely studied and have been shown to possess distinctive attractive convergence properties. This summary should help us identify remaining issues in adaptive IIR filter theory research in the last section.

TABLE I
Adaptive IIR Filter Categorizations

CATEGORY LABEL	ONE TYPE	OTHER TYPE
Plant	Deterministic	Stochastic
Model	With Noise Model	Without Noise Model
Algorithm Filtering	Of Prediction Error	Of Information Vector
Step-Size	Positive (bounded away from zero)	Vanishing

TABLE II
Adaptive IIR Filter Classes for Which Convergence Properties Are Discussed

CLASS NO.	CATEGORIZATION			
	PLANT	MODEL	ALGORITHM FILTERING	STEP-SIZE
1	Deterministic	Without Noise Model	Information Vector	Positive
2	Deterministic	Without Noise Model	Prediction Error	Positive
3	Stochastic	Without Noise Model	Prediction Error	Vanishing
4	Stochastic	With Noise Model	None	Vanishing
5	Stochastic	With Noise Model	Information Vector	Vanishing

Refer to Table I, which identifies four broad binary categorizations of adaptive IIR filter types. The first descriptor differentiates between cases where $w \equiv 0$ and $w \neq 0$. The second division considers the predictor as without noise modeling, i.e., $\hat{C} \equiv 0$ as in (2.15), or with noise modeling, $\hat{C} \neq 0$ as in (2.21). The third label is based on the filtering "fix" used: a time-varying AR information vector filtering, as in (3.22), or a time-invariant MA prediction error filtering, as in (3.45). The final categorization simply distinguishes between adaptive parameter estimators with positive step sizes guaranteed to be bounded away from zero and those with step sizes that purposefully vanish with time. We need not consider all sixteen combinations, since some have not received significant study. In fact, we will limit ourselves to the five classes indicated in Table II.

A. Class 1

The squared prediction error minimization interpretation of the output error problem has been exploited in, e.g., [1], [2], [4], and [5] with the solutions falling within this class. One concern is the modality of the error surface being descended. It was shown [25]–[26] to be possibly multimodal in reduced-order use, i.e., when the plant order n in (2.1) exceeds that of the adaptive IIR filter. This reveals a limitation to local minimization in a reduced-order application. However, the major drawback of this class is the need for on-line stability monitoring of the time-varying, AR information vector filter $[1 - \hat{A}]^{-1}$. This issue will be discussed further in Section VI.

B. Class 2

The initial proof of the convergence of an algorithm in this class, i.e., deterministic plant, no noise modeling, prediction error smoothing, and nonvanishing step size, appeared in [17], was based on hyperstability, and was intended for use as an identifier. This algorithm was subsequently translated [6] and simplified [7] for adaptive IIR filter use. The major practical drawback is the associated SPR condition, e.g., on (3.42), guaranteed satisfaction of which relies on unknown plant parameters. This will be discussed further in Section VI.

Given the persistent excitation of the information vector, as in (3.29), this class was shown to posses an exponential convergence rate [22]. Recognition that the persistent excitation requirement in (3.29) was unacceptably dependent on the practically unpredictable character of the $\hat{y}$ in X spurred translation of (3.29) to a similar condition on $\{u(k)\}$ alone [21]. The advantage of an exponential convergence rate in ideal usage is that it supports the proof of local stability in the presence of nonideal application such as reduced-order modeling (or bounded disturbance accommodation) [27] and time-varying desired parameter tracking [28]. These results are based on the recognition that such nonidealities result in the addition of bounded disturbances to the homogeneous error system. The exponential stability of a homogeneous, time-varying, nonlinear system is widely used in the stability theory literature, e.g., [29], [30], to establish a bounded-input, bounded-state property when this system is forced with small signals. Such robustness properties have important practical significance.

The behavior of the parameter error trajectories in the parameter error space has also been considered [31] to discern their relationship to the squared output error surface steepest descent character of Class 1. A consistently biased descent interpretation may be possible. However, this biased descent does not necessarily take place on the squared output error surface and, as shown in [25], does not result in squared output error minimization in the reduced-order case as do the algorithms of Class 1. In fact the particular prediction error smoothing used can be shown to effect the location of the local convergence point(s). Therefore, if squared output error minimization is the actual objective of the adaptive IIR filter, this class can be considered deficient.

C. Class 3

The convergent parameter estimate unbiasedness of this class given zero-mean white output measurement noise, $C = -A$ in (2.2), was the principal focus in its development in [17]. Note that with this choice for C in (2.2), the prediction error with (2.15) becomes

$$e(k) = \left[1 - A(q^{-1})\right]^{-1}\tilde{\theta}^T X(k) + w(k). \quad (4.1)$$

This unbiasedness was extended to general ARMAX plants in [32]. A principal requirement for such results is the vanishing step size. This can be heuristically justified as follows. If the parameter estimates in (2.15) are unbiased when applied to (2.2), then the prediction error is the filtered noise sequence with zero mean. For convergence to a point in the parameter space this random prediction error must not be allowed to continually perturb the parameter estimates via (3.46). This is accomplished with the decay of the step-size to zero. However, as the step size vanishes, the ability of the parameter estimator in (3.46) to react to nonzero prediction error due to the time-variation of the plant parameters in (2.2) also vanishes. Since a central purpose of adaptive parameter estimators, or adaptive filters, is to track such time-varying parameters, this class and the two following ones should, strictly speaking, not be considered adaptive.

Unfortunately, the use of nondisappearing step size gains for tracking capability, instead of the vanishing gains of this class, appears from our simulation experience to result in a bias in the parameter estimates that is proportionally related to the strictly positive lower bound on the step size. Roughly speaking, the smaller the step size the smaller the parameter estimate bias.

D. Class 4

Algorithms in this class rely on the strict positive reality of $[1 + C]^{-1} - \lambda$ and do not use any prediction error smoothing, e.g., [12], [20]. Again, as with Class 3, the convergence proofs of asymptotically unbiased behavior rely critically on the vanishing step size. It is conjectured that, as with Class 3 algorithms, nonvanishing step sizes could create biased parameter estimates.

E. Class 5

Of the stochastic Classes 3–5, this last one has received the most thorough study due to its asymptotic maximum likelihood behavior even in nonideal applications [12]–[14]. In addition to such asymptotic consistency, the asymptotic efficiency, i.e., the convergence of the estimation error to the Cramer–Rao lower bound, and the asymptotic normality, i.e., the convergence of the estimation error distribution to a Gaussian distribution, of this class have been proven. Again these properties are predicated on the vanishing step size. They are likely to be practically retained for a sufficiently small nonvanishing step size.

V. Signal Processing Applications

The preceding development of an understanding of the adaptive parameter estimation basis of adaptive IIR filters would be hollow without some indication as to how these model-algorithm combinations could be used to provide adaptive IIR filters useful in signal processing applications. A number of applications for adaptive IIR filters, including noise canceling [7], [8], multipath cancellation [7], sinusoid detection [9], time delay estimation [8], line-enhancing [33], adaptive differential pulse code modulation (ADPCM) [34], and air pollution prediction [35], have been suggested. For illustrative purposes we will only consider

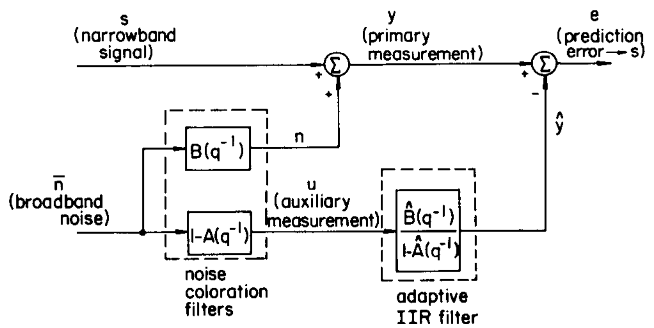

Fig. 2. Adaptive noise canceling.

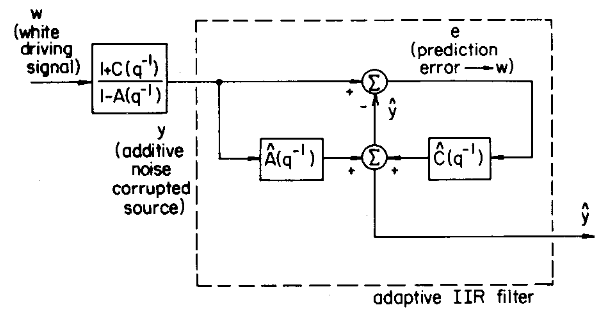

Fig. 3. Adaptive line enhancing.

three applications, each of which falls within one of the last three model-algorithm classes of Section IV: adaptive noise canceling, adaptive line-enhancing, and ADPCM.

A. Adaptive Noise-Canceling

The format of adaptive noise-canceling, introduced in [36], is illustrated in Fig. 2. The physical setup presumes that two measurements are available: one the signal (s) to be extracted plus noise (n) and the other a measurement (u) correlated with the noise (n). Fig. 2 adds the presumption that the additive noise corrupting the signal is a filtered version of a white sequence $\{\bar{n}(k)\}$ and that the measured correlated noise is this same $\bar{n}$ passed through a different coloration filter. For notational simplicity and easy correlation with the earlier sections of this paper both of these coloration filters are presumed to be FIR (or tapped-delay line) filters. The critical characteristic is that the relationship between u and n is described by an IIR transfer function, i.e.,

$$n(k) = B(q^{-1})\bar{n}(k) \tag{5.1}$$

and

$$u(k) = [1 - A(q^{-1})]\bar{n}(k) \tag{5.2}$$

so

$$n(k) = A(q^{-1})n(k) + B(q^{-1})u(k). \tag{5.3}$$

Now since

$$y(k) = s(k) + n(k), \tag{5.4}$$

(5.3) can be written as

$$y(k) = A(q^{-1})y(k) + B(q^{-1})u(k) - A(q^{-1})s(k) + s(k). \tag{5.5}$$

The adaptive noise-canceling filter in Fig. 2 has the form of (2.15). Forming the prediction error to be used in updating $\hat{A}$ and $\hat{B}$ yields a form similar to (4.1) with w replaced by s. Clearly if $\tilde{A} \triangleq A - \hat{A} \equiv 0$ and $\tilde{B} \triangleq B - \hat{B} \equiv 0$ then $e(k) = s(k)$ as desired, i.e., the output of the adaptive noise-canceling structure is the uncorrupted signal.

Note that if $s(k)$ were itself white (or colored) noise with a zero mean this problem would fall within Class 3 of the preceding section where an ARMAX process is "identified" without a "noise" model. In such a case, with disappearing step-size gains and appropriate error smoothing, $\tilde{A} \to 0$ and $\tilde{B} \to 0$ can be guaranteed. In practice to track time-varying A and B the step size cannot be allowed to disappear. Insufficient theory exists to explain the resulting misbehavior of Class 3 through 5 algorithms when step sizes are not allowed to decay to zero. Simulations [7] have shown that, even when s is a zero-mean highly correlated signal rather than white, but remains uncorrelated with u, and the step sizes are small but nonzero, significant SNR improvement is available when using e as an estimate of s over the SNR of $y = s + n$.

B. Adaptive Line-Enhancing

As with adaptive noise-canceling, adaptive line-enhancement was first suggested as an application of adaptive FIR filtering [36], [37]. This adaptive FIR filter usage is commonly predicated on the assumption of a white noise driven AR model with its to-be-recovered output signal corrupted by additive white measurement noise. As shown in [33], this generates the measurable signal model of (2.2) with $u \equiv 0$ and w an appropriate unmeasurable white composite of the uncorrelated driving and measurement noises. Furthermore, the optimal predictor [33] for the underlying process output is to pass y through (2.27) with $\hat{A} = A$ and $\hat{C} = C$. The structure for this adaptive IIR line enhancer from (2.27) is shown in Fig. 3.

If an AR model driven by white noise accurately describes the source to be recovered such that (2.2) with $u \equiv 0$ is an accurate model for the additive noise corrupted source, then a Class 4 or 5 adaptive algorithm could cause the adaptive IIR line enhancer in Fig. 3 to converge to the optimal predictor. Several successful simulations of a Class 5 algorithm are presented in [33]. As noted earlier, the theory requires vanishing step sizes for exact parameter convergence. However, as noted in the preceding subsection, the near convergence chatter and subsequent performance degradation due to small nonvanishing step sizes can be tolerable. However, insufficient theory exists to strengthen such a robustness statement.

Note that $1 + C$ should be satisfactorily damped to avoid frequent use of the Class 5 adaptive algorithm stability projection facility for $1 + \hat{C}$ once convergence is neared. Also as the roots of $1 + C$ are nearer the unit circle and as its order increases the SPR condition on $[1 + C]^{-1} - \lambda$ of Class 4 is more likely to be violated.

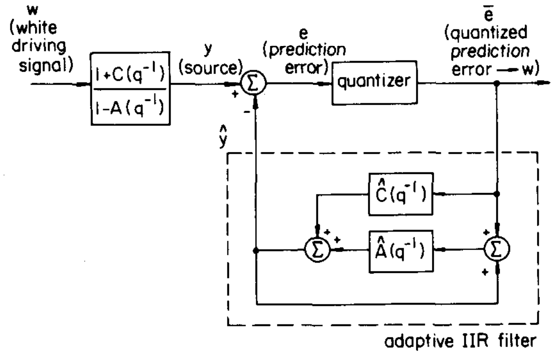

Fig. 4. Adaptive differential pulse code modulation encoder.

C. Adaptive Differential Pulse Code Modulation

The ADPCM problem [38] is one that only recently [34] has been seen as a possible application for adaptive IIR filters. The configuration for ADPCM suggesting adaptive IIR filter algorithm use is shown in Fig. 5.3. The source signal to be encoded is assumed to be the output of a white, zero-mean noise driven ARMA process as in (2.2) with $u \equiv 0$. The feedback predictor from Fig. 4 is similar to (2.28)

$$\hat{y}(k) = \hat{A}(q^{-1})[\hat{y}(k) + \bar{e}(k)] + \hat{C}(q^{-1})\bar{e}(k). \quad (5.6)$$

We will define the quantization error $n(k)$ as the difference between the prediction error $e(k) = y(k) - \hat{y}(k)$ and the quantized prediction error $\bar{e}(k)$

$$e(k) - \bar{e}(k) = n(k). \quad (5.7)$$

Thus, similar to the development of (2.23),

$$\begin{aligned}\bar{e}(k) &= y(k) - \hat{y}(k) - n(k)\\ &= [1 + C(q^{-1})]^{-1}\\ &\quad \cdot\{[A(q^{-1}) - \hat{A}(q^{-1})][\hat{y}(k) + \bar{e}(k)]\\ &\quad + [C(q^{-1}) - \hat{C}(q^{-1})]\bar{e}(k)\}\\ &\quad + w(k) - [1 + C(q^{-1})]^{-1}[1 - A(q^{-1})]n(k).\end{aligned} \quad (5.8)$$

Note that if $n(k) \equiv 0$, i.e., the quantization error is negligible, then (5.8) reduces to (2.23) where $\tilde{\theta}$ and X are as in (2.24) and (2.25) but with the $b_i - \hat{b}_i$ and $u(k-i)$ terms removed since $u \equiv 0$. Therefore, a Class 5 algorithm results in $e(=\bar{e}) \rightarrow w$. Clearly n is not zero in actual application. Note in (5.8) that even if parameter convergence occurred such that $\hat{A} = A$ and $\hat{C} = C$, $\bar{e}$ would not be whitened and would not equal w. Furthermore the presence of the colored perturbation by filtered $\{n(k)\}$ will lead to biased parameter estimates in an attempt to adaptively whiten $\bar{e}$ by selection of $\hat{A}$ and $\hat{C}$. The objective of ADPCM, however, is not to achieve exact parameter identification but rather to lower the variance of e relative to that of y so that a coarser quantizer can be used on e, than y, with the same fidelity. This quantizer bit reduction is usually quantified by the SNR of y to e where each 6 dB of improvement corresponds to an allowable removal of one quantizer bit. Preliminary simulation studies using real speech data have shown over 12 dB improvement.

VI. Open Issues

The preceding sections summarizing the modeling, algorithmic, convergence, and application aspects of adaptive IIR filters are purposefully bereft of cloying detail. The intent was to capture in a readable, unified framework the generic aspects of adaptive IIR filtering, heretofore undocumented in a single source. This rather tutorial presentation also serves another purpose: to permit a discussion of several open issues with regard to adaptive IIR filters. The ones chosen for discussion in the concluding section of this paper are a) methods for SPR condition satisfaction, b) stability projection techniques, c) alternative adaptive algorithm forms, d) the alteration of "noise" and "signal" definitions relative to identification use, and e) prediction error quantization effects. This list omits other relevant issues, such as convergence rate, the effects of nondisappearing step-size gains in a stochastic setting, lattice versus tapped-delay line implementation, and the investigation of additional applications, all of which are undeniably important. Those issues omitted from discussion here are already widely recognized as important in the adaptive filtering literature; while the issues to be discussed represent some that emerge from this particular study and have not received adequate advertisement in the adaptive filtering literature.

A. SPR Condition

As noted in Section III-B one of the major "compensations" for the AR filtering of the parameter estimate error and information vectors inner product present in the output error formulation is MA filtering of the prediction error prior to its use in the adaptive algorithm. The ratio of this MA filtering to the prediction error autoregression must satisfy an SPR condition, as in (3.38) and (3.42), to ensure stable convergence for all possible input sequences. One tendency of the *a priori* information needed for SPR satisfaction of, e.g., (3.42) should be noted again: as the roots of the prediction error AR approach the unit circle, i.e., become more oscillatory, the tendency is for the compensating prediction error smoothing polynomial to be required to have its roots closer and closer to the "unknown" roots of the AR.

The need for satisfaction of such an SPR condition to guarantee convergence for all possible input sequences does *not* mean that for certain input sequences convergence cannot occur without satisfaction of this SPR condition. In fact it has been shown [39] that if the information vector has a sufficiently large inner product with itself (i.e., the sum of the squares of its entries) at *every* time instant, which differs from the outer product persistent excitation condition of (3.29), no MA error smoothing is required. As shown in [39] convergence is retained by effectively modifying the SPR condition. Though not noted in [39], this inner product condition can be satisfied by a similar condition on the $\{u\}$ portion alone of the information vector. This sufficiently positive inner product condition has the disadvantage of disallowing even temporary periods of too many successive near-zero values in the input sequence. An open issue is whether or not this inner product condition, if satisfied, and the outer product condition of (3.29), translated to $\{u\}$ alone as in [21], can assure exponential convergence and the accompanying robustness properties. Furthermore, the possibility of translating this inner product constraint to the stochastic case has not been investigated.

Since satisfaction of the SPR condition has been considered critical for proper algorithm behavior, simultaneous adaption of these error smoothing coefficients has been suggested [40]. Unfortunately the desired stability of the resulting algorithms cannot be guaranteed without the same stability check and projection facility needed by the alternate "compensation" technique of time-varying AR information vector filtering of Section III-A [41]. Thus the desire to use a fixed prediction error smoothing, and thereby avoid this costly stability check and projection facility, remains as a severe practical limitation of such schemes. Needed is a thorough study of the practical ability to obtain the information allowing SPR condition satisfaction in a real application. Admittedly this needed information is short of precisely knowing the "unknown" AR coefficients in the prediction error, though not much less severe in certain applications.

B. Stability Projection Techniques

The need to provide a real-time stability check for the algorithms using a time-varying AR information vector (or time-varying, MA prediction error) filtering is a significant computational burden in many situations. Of more concern, however, is the associated projection facility which resituates the time-varying filter coefficients such that its instantaneous roots are always within the unit circle. As might be suspected, simulations have verified that when the desired roots for this time-varying compensator are near the unit circle, more frequent projection is required. Current theoretical results cannot prove that this projection facility is needed only finitely often in every situation, which would imply stable convergence. Rather, the algorithm convergence theorems, e.g., [14], state that either the parameter vector stably converges to the desired value or converges to a boundary point of the projection region. This implies that the time-varying AR filtering of the information vector has roots that inappropriately cluster near the unit circle (or the boundary, within the unit circle, of the region to which instantaneously unstable parameterizations are projected). The open issue is to find a projection facility that can be proven to be needed only finitely often and requires only a reasonable amount of computation.

One simple-minded projection facility is to ignore those updatings of the parameter estimate vector that would lead to the instability of the information vector filtering polynomial. Recall that in all of the algorithms using a time-varying AR filtering of the information vector, these filter coefficients are a portion of the full parameter estimate vector. (Refer to (3.22) and the comments following it.) This simplistic strategy of ignoring what could be called "unstable updates" has been seen to lock-up in numerous simulations, i.e., roots of the time-varying AR information vector filter cluster and freeze near the unit circle. An important attribute suggested in [18] for the projection scheme is that it be to a decidedly interior point of the stable region to avoid such lock-up. An alternate projection facility repeatedly shrinks the adaptive step size until the downscaled correction term does not result in instability of the new parameter estimate. This technique appears more robust, but not infallible, in simulation. Another scheme that has been studied [42] is based on the reflection of any roots found to be unstable. This technique has a somewhat logical basis since it maintains the magnitude of the frequency response of the nonminimum phase polynomial while altering only its phase characteristic. Admittedly the numerical complexity of such a technique is considerable. But even this approach has not yet been proven to *always* be successful. Development and evaluation of further projection candidates are needed.

C. Alternate Adaptive Algorithm Forms

Section III is devoted to introducing the two currently widely studied "compensations" for the AR filtering of the parameter estimate error and information vectors inner product in the prediction error of the output error formulation. As noted in the two preceding subsections both techniques for augmenting the simple correlation form of the correction term of the simpler equation error form have limitations. One uses a prediction error smoothing that must satisfy an SPR condition in undesirable conjunction with an "unknown" polynomial. The other requires a stability check on the added, time-varying AR information vector filter, for which there is presently no universally acceptable projection scheme. As noted in Section VI-A an algorithm form has been suggested to adapt the MA error smoothing coefficients of the first fix, but this then requires the stability check and projection of the second approach. A fourth possibility is a fixed AR information vector filtering. One might expect an SPR requirement to arise, as suggested in [13]. Our preliminary simulation studies of such a scheme have indicated some desirable attributes, especially with regard to deterministic reduced-order use. From Sections IV-A and IV-B arises the expectation that AR information vector filtering leads to squared output error minimization while MA prediction error smoothing does not. The desirable minimization characteristic appears to be retained with fixed AR information vector filtering, if the fixed filtering matches the optimum value of the respective portion of the parameter estimate vector. Admittedly such a fortuitous choice is practically impossible in a real application. More surprising was the graceful degradation in the convergent squared output error as this fixed filter is suboptimally chosen. Apparently this fourth choice of fixed AR information vector filtering deserves further study.

Two algorithmic modifications particular to the output error form, that were motivated by study of a time-varying AR information vector filtering (class 5) algorithm, were introduced in [43]. The first is the use of a "pull" factor that essentially divides the delay operator by a positive factor greater than one, e.g.,

$$1 + \hat{C}(\alpha q^{-1}) = 1 + \hat{c}_1 \alpha q^{-1} + \hat{c}_2 \alpha^2 q^{-2} + \cdots + \hat{c}_n \alpha^n q^{-n} \tag{6.1}$$

where $0 \leq \alpha \leq 1$. If this device is used in the AR information vector filter it will cause this filter to be more stable than otherwise and therefore lessen the need for use of the projection facility. The second modification could arise from consideration of (2.18) and (2.23), which can be associated with Classes 3 and 5, respectively, in Section IV. The suggestion is rather than use $[1 + \hat{C}]^{-1}$ for information vector filtering in conjunction with (2.23) use $[1 - \hat{A}]^{-1}$, which more naturally arises with (2.18) or a weighted combination of these two. In certain applications, $C \equiv -A$. For example, in the line enhancer of Section V-B as the corrupted source SNR increases $C \to -A$. In such cases this second modification exploits the fact that with Class 5 algorithms $\hat{A}$ tends to converge faster than $\hat{C}$. As noted in [43], these two modification have been studied mostly via simulation and require analytical evaluation.

D. "Noise" and "Signal" Definitions

One peculiarity of adaptive IIR filtering application in comparison to identification application of output error schemes is a reversal of the normally stochastic and deterministic definitions, respectively, of "noise" and "signal." In the identification format of Section II, which is the most common setting in which output error algorithms have been analyzed, the desired convergent prediction error is the unmeasurable, stochastic, sequentially uncorrelated signal w. Refer to (2.23). Contrast this with the adaptive noise canceling application of Section V-A where the desired convergent prediction error is the unmeasurable, possibly deterministic, likely highly correlated signal s. Refer to the remarks below (5.5). Presently, theoretical convergence results are available only for the identification format. The adaptive IIR filtering format requires further attention to prove stochastic convergence properties with this stochastic to deterministic conversion of the character of the unmeasurable input in (2.2). From a signal processing viewpoint what is being exploited in the noise-canceling application of Section V-A is the separability of uncorrelated signals with insignificantly overlapping spectra. Perhaps such a distinction will require incorporation in a subsequent theory handling this issue.

E. Prediction Error Quantization Effects

The last issue to be discussed also arises from consideration of an adaptive signal processing application, specifically ADPCM of Section V-C. In Section V-C it was noted that the negligibility of the quantization error was needed to justify the unbiased convergence of the predictor parameters. This assumption clashes with the ADPCM objective of withstanding as coarse quantization as possible. The simulations cited in Section V-C included a quantizer, yet performed quite well. Thus reasonable behavior can be expected when this quantization error is small relative to w. However, a thorough analysis of its effects is warranted. Other adaptive IIR filtering applications can be expected to provide similar "new" issues.

Acknowledgment

Hopefully this paper provides an attractive framework on which to hang one's understanding of adaptive IIR filters. If successful, the credit is to be shared with several colleagues who have been instrumental in forcing the author to sharpen this viewpoint, including D. A. Lawrence and J. R. Treichler of Cornell University, M. G. Larimore of ARGO Systems, Inc., B. D. O. Anderson and R. R. Bitmead of the Australian National University, and B. Q. Friedlander of Systems Control Technology, Inc. The author would also like to thank D. R. Morgan of General Electric Co., S. D. Stearns and R. A. David of Sandia Laboratories, C. Heegard of Cornell University, and an anonymous reviewer for their careful reading and helpful evaluation of the original manuscript.

References

[1] S. A. White, "An adaptive recursive filter," in *Proc. 9th Asilomar Conf. Circuits, Syst., Comput.*, Pacific Grove, CA, Nov. 1975, pp. 21–25.

[2] S. D. Stearns, G. R. Elliott, and N. Ahmed, "On adaptive recursive filtering," in *Proc. 10th Asilomar Conf. Circuits, Syst., Comput.*, Pacific Grove, CA, Nov. 1976, pp. 5–10.

[3] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, pp. 1622–1624, Nov. 1976.

[4] D. Parikh and N. Ahmed, "On an adaptive algorithm for IIR filters," *Proc. IEEE*, vol. 65, pp. 585–587, May 1978.

[5] S. Horvath, Jr., "A new adaptive recursive LMS filter," in *Digital Signal Processing*, V. Cappellini and A. G. Constantinides, Eds. New York: Academic, 1980, pp. 21–26.

[6] C. R. Johnson, Jr., "A convergence proof for a hyperstable adaptive recursive filter," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 745–749, Nov. 1979.

[7] M. G. Larimore, J. R. Treichler, and C. R. Johnson, Jr., "SHARF: An algorithm for adapting IIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 428–440, Aug. 1980.

[8] B. Friedlander, "System identification techniques for adaptive signal processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 240–246, Apr. 1982.

[9] I. D. Landau, L. Dugard, and S. Cabrera, "Applications of output error recursive estimation algorithms for adaptive signal processing," in *Proc. 1982 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Paris, France, May 1982, pp. 639–642.

[10] 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*, vol. 64, pp. 1151–1162, Aug. 1976.

[11] G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Analysis*. New York: Academic, 1977.

[12] T. Söderström, L. Ljung, and I. Gustavsson, "A theoretical analysis of recursive identification methods," *Automatica*, vol. 14, pp. 231–244, May 1978.

[13] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT, 1983.

[14] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 551–575, Aug. 1977.

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

[16] C. R. Johnson, Jr., "Another use of the Lin–Narendra error model: HARF," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 985–988, Oct. 1980.

[17] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 194–202, Apr. 1976.

[18] L. Ljung, "On positive real functions and the convergence of some recursive schemes," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 539–551, Aug. 1977.

[19] L. Dugard and I. D. Landau, "Recursive output error identification algorithms—theory and evaluation," *Automatica*, vol. 16, pp. 443–462, Sept. 1980.

[20] V. Solo, "The convergence of AML," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 958–962, Dec. 1979.

[21] B. D. O. Anderson and C. R. Johnson, Jr., "Exponential convergence of adaptive identification and control algorithms," *Automatica*, vol. 18, pp. 1–13, Jan. 1982.

[22] R. R. Bitmead and B. D. O. Anderson, "Performance of adaptive estimation algorithms in dependent random environments," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 782–794, Aug. 1980.

[23] G. C. Goodwin, P. J. Ramadge, and P. E. Caines, "Discrete-time multivariable adaptive control," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 449–456, June 1980.

[24] D. A. Lawrence, C. R. Johnson, Jr., and B. D. O. Anderson, "Exponential convergence and robustness of persistently excited recursive-least-squares-with-forgetting output error identification," in *Proc. 17th Conf. on Inform. Sci. and Syst.*, Baltimore, MD, Mar. 1983, pp. 293–298.

[25] C. R. Johnson, Jr., and M. G. Larimore, "Comments on and additions to 'An adaptive recursive LMS filter,'" *Proc. IEEE*, vol. 65, pp. 1399–1401, Sept. 1977.

[26] S. D. Stearns, "Error surfaces of recursive adaptive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 3, pp. 763–766, June 1981.

[27] B. D. O. Anderson and C. R. Johnson, Jr., "On reduced-order adaptive output error identification and adaptive IIR filtering," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 927–933, Aug. 1982.

[28] B. D. O. Anderson and R. M. Johnstone, "Adaptive systems and time varying plants," *Int. J. Contr.*, vol. 37, pp. 367–377, 1983.

[29] J. L. Willems, *Stability Theory of Dynamical Systems*. London: Nelson, 1970.

[30] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*. New York: Academic, 1975.

[31] C. R. Johnson, Jr., M. G. Larimore, J. R. Treichler, and B. D. O. Anderson, "SHARF: Convergence properties," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 659–670, June 1981.

[32] J. B. Moore, "Global convergence of output error recursions in colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 1189–1199, Dec. 1982.

[33] B. Friedlander, "A recursive maximum likelihood algorithm for ARMA line enhancement," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 651–657, Aug. 1982.

[34] C. R. Johnson, Jr., J. P. Lyons, Jr., and C. Heegard, "A new adaptive parameter estimation structure applicable to ADPCM," in *Proc. 1983 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Boston, MA, April 1983, pp. 1–4.

[35] A. Kikuchi, S. Omatu, and T. Soeda, "Applications of adaptive digital filtering to the data processing for the environmental system," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 790–803, Dec. 1979.

[36] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., and R. C. Goodwin, "Adaptive noise cancelling: principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975.

[37] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 53–62, Feb. 1979.

[38] J. D. Gibson, "Adaptive prediction in speech differential encoding systems," *Proc. IEEE*, vol. 68, pp. 488–525, Apr. 1980.

[39] M. Tomizuka, "Parallel MRAS without compensation block," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 505–509, Apr. 1982.

[40] I. D. Landau, "Elimination of the real positivity condition in the design of parallel MRAS," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 1015–1020, Dec. 1978.

[41] C. R. Johnson, Jr., and T. Taylor, "Failure of a parallel adaptive identifier with adaptive error filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 1248–1250, Dec. 1980.

[42] Y. Z. Tsypkin, E. D. Avedyan, and O. V. Gulinskiy, "On convergence of the recursive identification algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 1009–1017, Oct. 1981.

[43] B. Friedlander, "A modified prefilter for some recursive parameter estimation algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 232–235, Feb. 1982.