

# Adaptive IIR Filtering Algorithms for System Identification: A General Framework

Sergio L. Netto, Paulo S. R. Diniz, *Senior Member, IEEE*, and Panajotis Agathoklis, *Senior Member, IEEE*

**Abstract**— Adaptive IIR (infinite impulse response) filters are particularly beneficial in modeling real systems because they require lower computational complexity and can model sharp resonances more efficiently as compared to the FIR (finite impulse response) counterparts. Unfortunately, a number of drawbacks are associated with adaptive IIR filtering algorithms that have prevented their widespread use, such as: Convergence to biased or local minimum solutions, requirement of stability monitoring, and slow convergence. Most of the recent research effort on this field is aimed at overcoming some of the above mentioned drawbacks. In this paper, a number of known adaptive IIR filtering algorithms are presented using a unifying framework that is useful to interrelate the algorithms and to derive their properties. Special attention is given to issues such as the motivation to derive each algorithm and the properties of the solution after convergence. Several computer simulations are included in order to verify the predicted performance of the algorithms.

**Index Terms**— adaptive filters, adaptive algorithms.

## I. INTRODUCTION

IN the last decades, substantial research effort has been spent to turn adaptive IIR<sup>1</sup> filtering techniques into a reliable alternative to traditional adaptive FIR filters. The main advantages of IIR filters are that they are more suitable to model physical systems, due to the pole-zero structure, and also require much less parameters to achieve the same performance level of FIR filters. Unfortunately, these good characteristics come along with some possible drawbacks inherent to adaptive filters with recursive structure such as algorithm instability, convergence to biased and/or local minimum solutions, and slow convergence. Consequently, several new algorithms for adaptive IIR filtering have been proposed in the literature attempting to overcome these problems. Extensive research on this subject, however, seems to suggest that no general purpose

Manuscript received September 23, 1992; revised November 1, 1993. This work was supported by CAPES-Ministry of Education (Brazil) and Micronet (Canada).

S. L. Netto was at the Programa de Engenharia Elétrica at COPPE/EE/Federal University of Rio de Janeiro, Caixa Postal 68504, Rio de Janeiro, RJ, 21945, Brazil. He is now with the Department of Electrical and Computer Engineering at the University of Victoria, P.O. Box 3055, Victoria, B.C., V8W 3P6, Canada.

P. S. R. Diniz is with COPEE/EE/Federal University of Rio de Janeiro, Caixa Postal 68504, Rio de Janeiro, RJ, 21945, Brazil.

P. Agathoklis is with the Department of Electrical and Computer Engineering at the University of Victoria, P.O. Box 3055, Victoria, B.C., V8W 3P6, Canada.

IEEE Log Number 9407859.

<sup>1</sup>The acronyms FIR and IIR filters are commonly used in time-invariant digital filter theory to indicate, respectively, the finite or infinite impulse response characteristic of these devices. FIR filters are usually implemented by nonrecursive (all-zero) realizations and IIR filters by recursive (zero-pole) realizations. The same nomenclature also applies to adaptive filter theory.

optimal algorithm exists. In fact, all available information must be considered when applying adaptive IIR filtering, in order to determine the most appropriate algorithm for the given problem.

The main objective of this paper is to present the characteristics of the most commonly used algorithms for IIR adaptive filtering, when applied to system identification applications, in a simple and unified framework. There is a plethora of system identification techniques in the literature [2], [16], [41], [57]. This paper deals with simple on-line algorithms that are being used for adaptive IIR filtering. Earlier general papers on this topic were presented by Johnson [33], Shynk [55], and Gee and Rupp [17]. In [33], Johnson presents a tutorial on adaptive IIR filtering techniques highlighting the common theoretical basis between adaptive filtering and system identification. This work was the first attempt to unify the concepts and the terminology used in the fields of adaptive control and adaptive filtering. Later, in 1989, Shynk [55] published a tutorial on adaptive IIR filtering that deals with different algorithms, error formulations, and realizations. Due to its general content, however, this paper addresses only a few algorithms. Moreover, several new techniques were proposed after the publication of these papers motivating additional work on this topic.

The organization of the present paper is as follows: In Section II, the basic concepts of adaptive signal processing are discussed and a brief introduction to the system identification application is presented, providing the necessary background to study the characteristics of the several adaptive filtering algorithms based on different error definitions. Section III presents a detailed analysis of the Equation Error (EE) [41], Output Error (OE) [61], [69], Modified Output Error (MOE) [14], [36], SHARF [36], [39], Steiglitz and McBride (SM) [8], [63], Bias-Remedy Equation Error (BRLE) [40], Composite Regressor (CR) [34], and Composite Error (CE) [50] algorithms, including their properties of stability, solution characteristics, computational complexity, robustness *etc.*. The advantages/disadvantages of each algorithm are also emphasized. In Section IV, some simulation results are provided to illustrate some of the properties discussed in Section III.

## II. ADAPTIVE SIGNAL PROCESSING

### A. Basic Concepts

Fig. 1 depicts the basic block diagram of a general adaptive system in practice. At each time interval, an input signal sample  $x(n)$  is processed by a time-varying filter generating the output  $\hat{y}(n)$ . This signal is compared to a reference  $y(n)$ ,

TABLE I  
CHARACTERISTICS OF THE BASIC ADAPTIVE IIR REALIZATIONS

Filter Realization	Gradient Implementation	Stability Monitoring	Convergence Speed	Global Multimodality
Direct	simple	nonpractical if $\hat{n}_a > 2$	slow	no
Lattice	complex	extremely simple	good	no
Cascade	complex	simple	slow	yes
Parallel	simple	simple	slow	yes

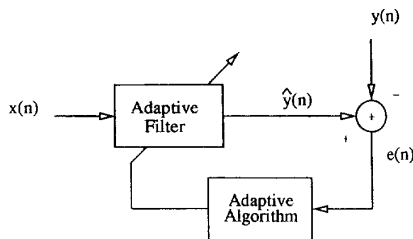


Fig. 1. Block diagram of a general adaptive system.

also called desired output, to generate an error signal  $e(n)$ . Finally, this error signal is used on an algorithm to adjust the adaptive filter coefficients in order to minimize a given performance criterion.

The specification of a complete adaptive system as shown in Fig. 1 consists of three items:

1) Application: The type of application is defined by the choice of the signals acquired from the environment to be the input and desired output signals. The number of different applications in which adaptive techniques are being successfully used increased enormously during the last decade. Some examples are echo cancellation, equalization of dispersive channels, system identification, and control. The study of different applications, however, is out of the scope of this paper. Good sources of information about adaptive filtering applications are the references [21], [66], [71].

2) Adaptive Filter Structure: The choice of the structure can influence the computational complexity (amount of arithmetic operations per iteration) of the process and also the necessary number of iterations to achieve a desired performance level. Basically, there are two classes of adaptive digital<sup>2</sup> filter realizations:

- Adaptive FIR filter realizations: The most widely used adaptive FIR filter structure is the transversal filter, also called tapped delay line (TDL), that implements an all-zero transfer function with a canonic direct form realization without feedback. For this realization, the output signal  $\hat{y}(n)$  is a linear combination of the filter coefficients, what yields a quadratic mean-square-error (MSE

<sup>2</sup>Despite of the fact that some adaptive filters can also be implemented with continuous-time techniques, general results had shown that this type of realization still faces many practical implementation problems [25], [44], [67]. As a consequence, this paper will focus on discrete-time implementations of adaptive systems.

$= E[e^2(n)]$  function with a unique optimal operation point [71]. Other alternative adaptive FIR realizations are also used in order to obtain improvements as compared to the transversal filter structure in terms of computational complexity [7], [15], speed of convergence [43], [45], [46], and finite word length properties [21].

- Adaptive IIR filter realizations: An early attempt to implement an adaptive IIR filter was made by White [69] in 1975 and since then a large number of papers have been published in this area. Initially, most of the works on adaptive IIR filters made use of the canonic direct-form realization due to its simple implementation and analysis. However, due to some inherent problems of recursive adaptive filters that are also structure dependent such as continuous poles monitoring requirement and slow speed of convergence, different realizations were studied attempting to overcome the limitations of the direct form structure. Among these alternative structures, the cascade [6], lattice [52], and parallel [54] realizations can be considered by their unique features. The most important characteristics of these recursive filter structures are summarized in Table I. From Table I, it can be easily concluded that each of these structures has some specific advantages when compared to the others, what seems to indicate that in practice there is no general optimal structure. The study of alternative realizations is a research direction that has been vastly explored by many authors, specially during the most recent years [5], [6], [52]-[54], [68].

3) Algorithm: The algorithm is the procedure used to adjust the adaptive filter coefficients in order to minimize a prescribed criterion. The algorithm is determined by defining the search method (or minimization algorithm), the objective function and the error signal nature. The choice of the algorithm determines several crucial aspects of the overall adaptive process, such as existence of suboptimal solutions, biased optimal solution, and computational complexity.

The main objective of this paper is to analyze a number of known algorithms used in adaptive IIR signal processing. In order to present a simple framework, all the analysis shown in this work will be based on the system identification application and on the direct-form IIR structure. However, all results discussed can be easily extended for other applications and realizations following the studies of Johnson [28], [32] and Nayeri [47], respectively.

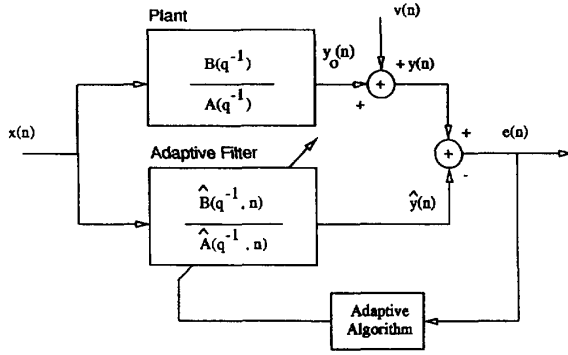


Fig. 2. Block diagram of an adaptive system identifier.

### B. System Identification with IIR Direct-Form Realization

In the system identification configuration, the adaptive algorithm searches for the adaptive filter such that its input/output relationship matches as close as possible that of the unknown system.

Fig. 2 depicts the general block diagram of an adaptive system identifier where the unknown system or plant is described by

$$\left. \begin{aligned} y_o(n) &= \left[ \frac{B(q^{-1})}{A(q^{-1})} \right] x(n) \\ y(n) &= y_o(n) + v(n) \end{aligned} \right\} y(n) = \left[ \frac{B(q^{-1})}{A(q^{-1})} \right] x(n) + v(n) \quad (1)$$

where  $A(q^{-1}) = 1 - \sum_{i=1}^{n_a} a_i q^{-i}$  and  $B(q^{-1}) = \sum_{j=0}^{n_b} b_j q^{-j}$  are coprime polynomials of the unit delay operator  $q^{-1}$ , and  $x(n)$  and  $v(n)$  are the input signal and the additive perturbation noise, respectively. The adaptive filter is implemented with the direct-form structure described by

$$\hat{y}(n) = \left[ \frac{\hat{B}(q^{-1}, n)}{\hat{A}(q^{-1}, n)} \right] x(n) \quad (2)$$

where  $\hat{A}(q^{-1}, n) = 1 - \sum_{i=1}^{\hat{n}_a} \hat{a}_i(n) q^{-i}$  and  $\hat{B}(q^{-1}, n) = \sum_{j=0}^{\hat{n}_b} \hat{b}_j(n) q^{-j}$ .

Another way to represent the adaptive identification process depicted in Fig. 2 can be obtained by defining the following vectors:

$$\begin{aligned} \theta &= [a_1 \dots a_{n_a} b_0 \dots b_{n_b}]^T = [a_k | b_l]^T \quad (3) \\ \phi(n) &= [y_o(n-k) | x(n-l)]^T, \\ &\text{for } k = 1, \dots, n_a; l = 0, \dots, n_b \\ \hat{\theta}(n) &= [\hat{a}_i(n) | \hat{b}_j(n)]^T \\ \hat{\phi}_{MOE}(n) &= [\hat{y}(n-i) | x(n-j)]^T, \\ &\text{for } i = 1, \dots, \hat{n}_a; j = 0, \dots, \hat{n}_b \end{aligned}$$

where  $\theta$  is the plant parameter vector,  $\phi(n)$  is the plant information vector,  $\hat{\theta}(n)$  is the adaptive filter parameter vec-

tor, and  $\hat{\phi}_{MOE}(n)$  is the adaptive filter information vector, respectively.<sup>3,4</sup>

With the above definitions, (1) and (2) can be respectively rewritten in the forms

$$y(n) = \phi^T(n) \theta + v(n) \quad (4)$$

$$\hat{y}(n) = \hat{\phi}_{MOE}^T(n) \hat{\theta}(n) \quad (5)$$

The physical meaning of a signal is more clear when using the delay operator polynomial notation. At the same time, the vectorial notation is also quite useful, since it greatly simplifies the adaptive algorithm representation, as will be seen later.

In order to present the adaptive IIR filtering algorithms in a structured form, it is useful to classify the identification problem by combining three of the following features, one in each item.

1) Classification with respect to the adaptive filter order:

- Feature (a) - insufficient order:  $n^* < 0$ ;
- Feature (b) - strictly sufficient order:  $n^* = 0$ ;
- Feature (c) - more than sufficient order:  $n^* > 0$ ,

where  $n^* = \min[(n_a - \hat{n}_a); (n_b - \hat{n}_b)]$ . In many cases, features (b) and (c) are grouped in one class, called sufficient order, where  $n^* \geq 0$ .

2) Classification with respect to the input signal properties:

- Feature (d) - persistent exciting input signal;
- Feature (e) - nonpersistent exciting input signal.

Basically, the persistence of excitation concept [2], [56] can be associated to the amount of information carried by the external signals  $x(n)$  and  $y(n)$  of the adaptive process.

3) Classification with respect to the disturbance signal properties:

- Feature (f) - without perturbation;
- Feature (g) - with perturbation correlated with the input signal;
- Feature (h) - with perturbation uncorrelated with the input signal.

Processes with feature (e) may lead to situations where it is not possible to identify the system parameters and therefore they are not widely studied in the literature. Also, feature (g) can be considered a special case of feature (a). All the other cases will be considered in this paper.

### C. Introduction to Adaptive Algorithms

The basic objective of the adaptive filter in a system identification problem is to set the parameters  $\hat{\theta}(n)$  in such way that it describes in an *equivalent* form the unknown system input-output relationship, i.e., the mapping of  $x(n)$  into  $y(n)$ . Usually, system equivalence [2] is determined by an objective function  $\mathcal{W}$  of the input, available plant output, and adaptive filter output signals, i.e.,  $\mathcal{W} = \mathcal{W}[x(n), y(n), \hat{y}(n)]$ . Two systems, S1 and S2, are considered equivalent if, for the

<sup>3</sup>In this shorter notation, the vertical bar delimiter '|' emphasizes the fact that the respective information vector is formed by subvectors of the indicated variables. This notation will be used throughout the paper in order to obtain a simple presentation.

<sup>4</sup>The index intervals for  $i$  and  $j$  will remain valid for all equations in this paper, and from now on they will be omitted for the sake of simplification.

same signals  $x(n)$  and  $y(n)$ , the objective function assumes the same value for these systems:  $\mathcal{W}[x(n), y(n), \hat{y}_1(n)] = \mathcal{W}[x(n), y(n), \hat{y}_2(n)]$ . It is important to notice that in order to have a consistent definition the objective function must satisfy the following properties:

- Nonnegativity:  $\mathcal{W}[x(n), y(n), \hat{y}(n)] \geq 0, \forall \hat{y}(n)$ ;
- Optimality:  $\mathcal{W}[x(n), y(n), y(n)] = 0$ .

Based on the concepts presented above, we may understand that in an adaptive process the adaptive algorithm attempts to minimize the functional  $\mathcal{W}$  in such a way that  $\hat{y}(n)$  approximates  $y(n)$  and, as a consequence,  $\hat{\theta}(n)$  converges to  $\theta$ , or to the best possible approximation of  $\theta$ .

Another way to interpret the objective function is to consider it a direct function of a generic error signal  $e(n)$ , which in turn is a function of the signals  $x(n)$ ,  $y(n)$ , and  $\hat{y}(n)$ , *i.e.*,  $\mathcal{W} = \mathcal{W}[e(n)] = \mathcal{W}[e(x(n), y(n), \hat{y}(n))]$ . Using this approach, we can consider that an adaptive algorithm is composed of three basic items: Definition of the minimization algorithm, definition of the objective function form, and definition of the error signal. These items are following discussed:

1) Definition of the minimization algorithm for the functional  $\mathcal{W}$ : This item is the main subject of the Optimization Theory and it essentially affects the adaptive process speed of convergence. The most commonly used optimization methods in the adaptive signal processing field are:

- Newton method: This method seeks the minimum of a second-order approximation of the objective function using an iterative updating formula for the parameter vector given by

$$\hat{\theta}(n+1) = \hat{\theta}(n) - \mu \mathcal{H}_{\hat{\theta}}^{-1} \{ \mathcal{W}[e(n)] \} \nabla_{\hat{\theta}} \{ \mathcal{W}[e(n)] \} \quad (6)$$

where  $\mu$  is a factor that controls the step size of the algorithm,  $\mathcal{H}_{\hat{\theta}} \{ \mathcal{W}[e(n)] \}$  is the Hessian matrix of the objective function, and  $\nabla_{\hat{\theta}} \{ \mathcal{W}[e(n)] \}$  is the gradient of the objective function with respect to the adaptive filter coefficients;

- *Quasi-Newton* methods: This class of algorithms is a simplified version of the method described above, as it attempts to minimize the objective function using a recursively calculated estimate of the inverse of the Hessian matrix, *i.e.*,

$$\hat{\theta}(n+1) = \hat{\theta}(n) - \mu P(n) \nabla_{\hat{\theta}} \{ \mathcal{W}[e(n)] \} \quad (7)$$

where  $P(n)$  is an estimate of  $\mathcal{H}_{\hat{\theta}}^{-1} \{ \mathcal{W}[e(n)] \}$  such that  $\lim_{n \rightarrow \infty} P(n) = \mathcal{H}_{\theta}^{-1} \{ \mathcal{W}[e(n)] \}$ . A usual form to implement this approximation is through the matrix inversion lemma (see for example [42]). Also, the gradient vector is usually replaced by a computationally efficient estimate;

- Gradient method: This type of algorithm searches the objective function minimum point following the opposite direction of the gradient vector of this function. Consequently, the updating equation assumes the form

$$\hat{\theta}(n+1) = \hat{\theta}(n) - \mu \nabla_{\hat{\theta}} \{ \mathcal{W}[e(n)] \} \quad (8)$$

In general, gradient methods are easier to implement, but on the other hand, the Newton method usually requires a smaller number of iterations to reach a neighborhood of the minimum point. In many cases, *Quasi-Newton* methods can be considered a good compromise between the computational efficiency of the gradient methods and the fast convergence of the Newton method. However, the latter class of algorithms are susceptible to instability problems due to the recursive form used to generate the estimate of the inverse Hessian matrix. A detailed study of the most widely used minimization algorithms can be found in [42].

It should be pointed out that with any minimization method, the convergence factor  $\mu$  controls the stability, speed of convergence, and misadjustment [71] of the overall adaptive algorithm. Usually, an appropriate choice of this parameter requires a reasonable amount of knowledge of the specific adaptive problem of interest. Consequently, there is no general solution to accomplish this task. In practice, computational simulations play an important role and are, in fact, the most used tool to address the problem.

2) Definition of the objective function  $\mathcal{W}[e(n)]$ : There are many ways to define an objective function that satisfies the optimality and nonnegativity properties formerly described. This definition directly affects the complexity of the gradient vector (and/or the Hessian matrix) calculation. Using the algorithm computational complexity as a criterion, we can list the following forms for the objective function as the most commonly used in the derivation of an adaptive algorithm:

- Mean Squared Error (MSE):  $\mathcal{W}[e(n)] = E[e^2(n)]$ ;
- Least Squares (LS):  $\mathcal{W}[e(n)] = \frac{1}{N+1} \sum_{i=0}^N e^2(n-i)$ ;
- Instantaneous Squared Value (ISV):  $\mathcal{W}[e(n)] = e^2(n)$ .

The MSE, in a strict sense, is of theoretical value since it requires an infinite amount of information to be measured. In practice, this ideal objective function can be approximated by the other two listed. The LS and ISV differ in the implementation complexity and in the convergence behavior characteristics; in general, the ISV is easier to implement but presents noisy convergence properties as it represents a greatly simplified objective function.

3) Definition of the error signal  $e(n)$ : The choice of the error signal is crucial for the algorithm definition since it can affect several characteristics of the overall algorithm including computational complexity, speed of convergence, robustness, and most importantly, the occurrence of biased and multiple solutions. Several examples of error signals are presented in detail in the following section.

The minimization algorithm, the objective function, and the error signal as presented give us a structured and simple way to interpret, analyze, and study an adaptive algorithm. In fact, almost all known adaptive algorithms can be visualized in this form, or in a slight variation of this organization. In the next section, using this framework we present a detailed review of the best known adaptive algorithms applicable to adaptive IIR filtering. It may be observed that the minimization algorithm and the objective function mainly affect the convergence speed of the adaptive process. Actually, the most important task for the definition of an adaptive algorithm definition seems to be

the choice of the error signal, since this task exercises direct influence in many aspects of the overall convergence process. Therefore, in order to concentrate efforts on the analysis of the influence of the error signal and to present a simple study, we will keep the minimization algorithm and the objective function fixed by using a gradient search method to minimize the instantaneous squared value of the error signal.

### III. ADAPTIVE IIR ALGORITHMS

This section presents an analysis of the Equation Error (EE), Output Error (OE), Modified Output Error (MOE), SHARF, Modified SHARF (MSHARF), Steiglitz-McBride (SM), Bias-Remedy Equation Error (BRLE), Composite Regressor (CR), and Composite Error (CE) algorithms. Specifically, are discussed properties of stability, solution characteristics, computational complexity, and robustness, highlighting the advantages/disadvantages of each algorithm when related to the others.

#### A. The Equation Error Algorithm

The simplest way to model an unknown system is to use the input-output relationship described by a linear difference equation as

$$y(n) = \hat{a}_1(n)y(n-1) + \dots + \hat{a}_{\hat{n}_a}(n)y(n-\hat{n}_a) + \hat{b}_0(n)x(n) + \dots + \hat{b}_{\hat{n}_b}(n)x(n-\hat{n}_b) + e_{EE}(n) \quad (9)$$

where  $\hat{a}_i(n)$  and  $\hat{b}_j(n)$  are the adaptive parameters, and  $e_{EE}(n)$  is a residual error, called *equation error*. Equation (9) can be rewritten using the delay operator polynomial form as

$$y(n) = \left[ \frac{\hat{B}(q^{-1}, n)}{\hat{A}(q^{-1}, n)} \right] x(n) + \left[ \frac{1}{\hat{A}(q^{-1}, n)} \right] e_{EE}(n) \quad (10)$$

or in a vectorial form as

$$y(n) = \hat{\phi}_{EE}^T(n) \hat{\theta}(n) + e_{EE}(n) \quad (11)$$

with  $\hat{\phi}_{EE}(n) = [y(n-i) \mid x(n-j)]^T$ .

From equations above, it is easy to verify that the adaptive algorithm that attempts to minimize the equation error squared value using a gradient search method is given by

$$\begin{aligned} \hat{\theta}(n+1) &= \hat{\theta}(n) - \mu \nabla_{\hat{\theta}} [e_{EE}^2(n)] \\ &= \hat{\theta}(n) - \mu \nabla_{\hat{\theta}} [e_{EE}(n)] e_{EE}(n) \\ &= \hat{\theta}(n) + \mu \hat{\phi}_{EE}(n) e_{EE}(n) \end{aligned} \quad (12)$$

with  $\mu = 2\mu'$ . This algorithm is characterized by the following properties [19]:

**Property 1:** The Euclidean square-norm of the error parameter vector defined by  $s(n) = \|\hat{\theta}(n)\|^2 = \|\hat{\theta}(n) - \theta\|^2$  is a convergent sequence if  $n^* \geq 0$  and  $\mu$  satisfies

$$0 \leq \mu \leq \frac{2}{\|\hat{\phi}_{EE}(n)\|^2} \quad (13)$$

The equation error  $e_{EE}(n)$  is a convergent sequence if  $n^* \geq 0$  and  $\mu$  satisfies

$$0 < \mu < \frac{2}{\|\hat{\phi}_{EE}(n)\|^2} \quad (14)$$

This first property establishes the upper bound of  $\mu$  to guarantee the stability of the equation error algorithm. Although this property asserts that  $s(n)$  and  $e_{EE}(n)$  are convergent sequences, it is not clear to what value  $s(n)$  tends at convergence. This point is clarified by the following statement:

**Property 2:** The characteristic of the equation error solution depends on the system identification process type as follows:

In a sufficient order case ( $n^* \geq 0$ ), if the perturbation noise is zero ( $v(n) \equiv 0$ ), all global minimum points of the mean-square equation error (MSEE) performance surface are described by

$$\begin{cases} \hat{A}^*(q^{-1}) = A(q^{-1})C(q^{-1}) \\ \hat{B}^*(q^{-1}) = B(q^{-1})C(q^{-1}) \end{cases} \quad (15)$$

with  $C(q^{-1}) = \sum_{k=0}^{n^*} c_k q^{-k}$ . It means that all global minimum solutions have included the polynomials describing the unknown system plus a common factor  $C(q^{-1})$  present in the numerator and denominator polynomials of the adaptive filter. On the other hand, if the perturbation noise is present, the final solution is biased with the degree of bias being a function of the variance of the disturbance signal.

In an insufficient order case ( $n^* < 0$ ), the solution is always biased and the degree of bias is a function of the plant and the input signal characteristics.

The main characteristic of the equation error algorithm is the unimodality of the MSEE performance surface due to the linear relationship existent between the equation error signal and the adaptive filter coefficients. This property, however, comes along with the drawback of biased solution in the presence of a disturbance signal. The following items show some algorithms that attempt to overcome this significant problem.

#### B. The Output Error Algorithm

The output error algorithm attempts to minimize the mean squared value of the output error signal, where the output error is given by the difference between the plant and the adaptive filter output signals, i.e.,

$$\begin{aligned} e_{OE}(n) &= \left[ \frac{B(q^{-1})}{A(q^{-1})} - \frac{\hat{B}(q^{-1}, n)}{\hat{A}(q^{-1}, n)} \right] x(n) + v(n) \\ &= \phi^T(n) \theta - \hat{\phi}_{MOE}^T(n) \hat{\theta}(n) + v(n) \end{aligned} \quad (16)$$

where  $\phi(n)$  and  $\hat{\phi}_{MOE}(n)$  were defined in (3). By finding the gradient of an estimate of the objective function given by  $\mathcal{W}[e_{OE}(n)] = E[e_{OE}^2(n)]$  with respect to the adaptive filter coefficients, we obtain

$$\begin{aligned} \nabla_{\hat{\theta}} [e_{OE}^2(n)] &= 2e_{OE}(n) \nabla_{\hat{\theta}} [e_{OE}(n)] \\ &= -2e_{OE}(n) \nabla_{\hat{\theta}} [\hat{y}(n)] \end{aligned} \quad (17)$$

with

$$\begin{aligned} \nabla_{\hat{\theta}} [\hat{y}(n)] &= \left[ \begin{array}{l} \frac{\partial \hat{y}(n)}{\partial \hat{a}_i} \Big|_{\hat{a}_i = \hat{a}_i(n)} \\ \frac{\partial \hat{y}(n)}{\partial \hat{b}_j} \Big|_{\hat{b}_j = \hat{b}_j(n)} \end{array} \right] \\ &= \left[ \begin{array}{l} \hat{y}(n-i) + \sum_{k=1}^{\hat{n}_a} \hat{a}_k(n) \frac{\partial \hat{y}(n-k)}{\partial \hat{a}_i} \Big|_{\hat{a}_i = \hat{a}_i(n)} \\ x(n-j) + \sum_{k=1}^{\hat{n}_b} \hat{a}_k(n) \frac{\partial \hat{y}(n-k)}{\partial \hat{b}_j} \Big|_{\hat{b}_j = \hat{b}_j(n)} \end{array} \right] \end{aligned} \quad (18)$$

From equations above, it follows that the gradient calculation requires at each iteration the partial derivatives values of past samples of signal  $\hat{y}(n)$  with respect to the variables of  $\hat{\theta}$  at the present moment  $n$ . This requires a relatively high number of memory devices to store data. In practice, this problem is overcome by an assumption called *small step approximation* [4], [22], [23] that considers the adaptive filter coefficients slowly varying. The resulting gradient vector is calculated as

$$\begin{aligned} \nabla_{\hat{\theta}}[y(n)] &\approx \begin{bmatrix} \hat{y}(n-i) + \sum_{k=1}^{\hat{n}_a} \hat{a}_k(n) \frac{\partial \hat{y}(n-k)}{\partial \hat{a}_k} \Big|_{\hat{a}_k = \hat{a}_k(n-k)} \\ x(n-j) + \sum_{k=1}^{\hat{n}_b} \hat{b}_k(n) \frac{\partial \hat{y}(n-k)}{\partial \hat{b}_k} \Big|_{\hat{b}_k = \hat{b}_k(n-k)} \end{bmatrix} \\ &= \begin{bmatrix} \hat{y}(n-i) + \sum_{k=1}^{\hat{n}_a} \hat{a}_k(n) q^{-k} \frac{\partial \hat{y}(n)}{\partial \hat{a}_k} \Big|_{\hat{a}_k = \hat{a}_k(n)} \\ x(n-j) + \sum_{k=1}^{\hat{n}_b} \hat{b}_k(n) q^{-k} \frac{\partial \hat{y}(n)}{\partial \hat{b}_k} \Big|_{\hat{b}_k = \hat{b}_k(n)} \end{bmatrix} \\ &= \begin{bmatrix} \left[ \frac{1}{\hat{A}(q^{-1}, n)} \right] \hat{y}(n-i) \\ \left[ \frac{1}{\hat{A}(q^{-1}, n)} \right] x(n-j) \end{bmatrix} \\ &\approx \begin{bmatrix} \left[ \frac{1}{\hat{A}(q^{-1}, n-i)} \right] \hat{y}(n-i) \\ \left[ \frac{1}{\hat{A}(q^{-1}, n-j)} \right] x(n-j) \end{bmatrix} \\ &= [\hat{y}^f(n-i) \mid x^f(n-j)]^T = \hat{\phi}_{OE}(n) \end{aligned} \quad (19)$$

The computation of gradient vector requires an all pole filtering of the signals  $\hat{y}(n-1)$  and  $x(n)$  in order to obtain  $\hat{y}^f(n-1)$  and  $x^f(n)$ , respectively. Therefore, using the small step approximation, the output error algorithm is described by

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu v_{OE}(n) \hat{\phi}_{OE}(n) \quad (20)$$

The convergence properties of the output error algorithm are dependent on the form of the mean-square output error (MSOE) performance surface, which in turn is characterized by some important concepts described below.

**Property 3:** *The stationary points [42] of the MSOE performance surface defined by  $W[v_{OE}(n)] = E[v_{OE}^2(n)]$  are given by [3], [56]*

$$E \left\{ \left[ \frac{\hat{A}(q^{-1}, n)B(q^{-1}) - A(q^{-1})\hat{B}(q^{-1}, n)}{A(q^{-1})\hat{A}(q^{-1}, n)} \right] x(n) \right\} \cdot \left\{ \left[ \frac{\hat{B}(q^{-1}, n)}{\hat{A}^2(q^{-1}, n)} \right] x(n-i) \right\} = 0 \quad (21)$$

$$E \left\{ \left[ \frac{\hat{A}(q^{-1}, n)B(q^{-1}) - A(q^{-1})\hat{B}(q^{-1}, n)}{A(q^{-1})\hat{A}(q^{-1}, n)} \right] x(n) \right\} \cdot \left\{ \left[ \frac{-1}{\hat{A}(q^{-1}, n)} \right] x(n-j) \right\} = 0 \quad (22)$$

*In practice, only the stable stationary points, so called equilibrium points, are of interest and usually these points are classified as [49]*

- *Degenerated points: The degenerated points are the equilibrium points where*

$$\begin{cases} \hat{B}(q^{-1}, n) \equiv 0 : \hat{n}_b < \hat{n}_a \\ \hat{B}(q^{-1}, n) \equiv L(q^{-1})\hat{A}(q^{-1}, n) : \hat{n}_b \geq \hat{n}_a \end{cases} \quad (23)$$

where  $L(q^{-1}) = \sum_{k=0}^{\hat{n}_b - \hat{n}_a} l_k q^{-k}$ .

- *Nondegenerated points: All the equilibrium points that are not degenerated points.*

The next properties define how the equilibrium points influence the form of the performance surface associated to the output error adaptive algorithm.

**Property 4:** *If  $n^* \geq 0$ , all global minima of the MSOE performance surface are given by [3], [56]*

$$\begin{cases} \hat{A}^*(q^{-1}) = A(q^{-1})C(q^{-1}) \\ \hat{B}^*(q^{-1}) = B(q^{-1})C(q^{-1}) \end{cases} \quad (24)$$

with  $C(q^{-1}) = \sum_{k=0}^n c_k q^{-k}$ . It means that all global minimum solutions have included the polynomials describing the unknown system plus a common factor  $C(q^{-1})$  present in the numerator and denominator polynomials of the adaptive filter.

**Property 5:** *If  $n^* \geq 0$ , all equilibrium points that satisfy the strictly positive realness condition*

$$\operatorname{Re} \left[ \frac{\hat{A}^*(z^{-1})}{\hat{A}(z^{-1})} \right] > 0 : |z| = 1 \quad (25)$$

are global minima [59].

**Property 6:** *Let the input signal  $x(n)$  be given by  $x(n) = \left[ \frac{F(q^{-1})}{G(q^{-1})} \right] w(n)$ , where  $F(q^{-1}) = \sum_{k=0}^{n_f} f_k q^{-k}$  and  $G(q^{-1}) = 1 - \sum_{k=1}^{n_g} g_k q^{-k}$  are coprime polynomials, and  $w(n)$  is a white noise. Then if*

$$\begin{cases} n^* \geq n_f \\ \hat{n}_b - n_a + 1 \geq n_g \end{cases} \quad (26)$$

all equilibrium points are global minima [59].

This later property is indeed the most general result about the unimodality of the MSOE performance surface in cases of sufficient order identification and it has two important consequences:

**Consequence 1** *If  $\hat{n}_a = n_a = 1$  and  $\hat{n}_b \geq n_b \geq 1$ , then there is only one equilibrium point, which is the global minimum.*

**Consequence 2** *If  $x(n)$  is a white noise ( $n_f = n_g = 0$ ), the orders of the adaptive filter are strictly sufficient ( $\hat{n}_a = n_a$  and  $\hat{n}_b = n_b$ ), and  $\hat{n}_b - n_a + 1 \geq 0$ , then there is only one equilibrium point, which is the global minimum.*

The case analyzed by this last statement was further investigated by Nayeri in [48] where it was obtained a less restrictive sufficient condition to guarantee unimodality of the output error algorithm when the input signal is a white noise and the orders of the adaptive filter exactly match the unknown system. This result is given by

**Property 7:** *If  $x(n)$  is a white noise ( $n_f = n_g = 0$ ), the orders of the adaptive filter are strictly sufficient, ( $\hat{n}_a = n_a$  and  $\hat{n}_b = n_b$ ), and  $\hat{n}_b - n_a + 2 \geq 0$ , then there is only one equilibrium point, which is the global minimum [48].*

Using numerical examples, Fan and Nayeri [11] showed that this last condition is the least restrictive sufficient condition that assures unimodality of the adaptive process for the corresponding adaptive system identification case. Another important property is

**Property 8:** *All degenerated equilibrium points are saddle points and their existence implies multimodality (existence of stable local minimum) of the performance surface if either  $\hat{n}_a > \hat{n}_b = 0$  or  $\hat{n}_a = 1$  [49].*

Notice that this last property is independent of the value of  $n^*$  and, as a consequence, is also valid for insufficient order cases.

Besides all properties previously listed, another interesting statement related to the output error algorithm characteristics was made by Stearns [62] who in 1981 conjectured that: *If  $n^* \geq 0$  and  $x(n)$  is a white noise input signal, then the performance surface defined by the MSOE objective function is unimodal.* This conjecture was supported by innumerable numerical examples, and remained considered valid until 1989, when Fan and Nayeri published a numerical counterexample [11] for it.

Basically, the most important characteristics of the output error algorithm are the possible existence of multiple local minima, what can affect the overall convergence of the adaptive algorithm, and the existence of an unbiased global minimum solution even in presence of perturbation noise in the unknown system output signal. Other important aspect related to the output error algorithm is the stability checking requirement during the adaptive process. Although this checking process can be efficiently performed by choosing an appropriate adaptive filter realization, some research effort was spent in order to avoid this requirement, as detailed in the next items.

### C. The Modified Output Error Algorithm

Another adaptive algorithm based on the output error signal can be obtained using the following simplification on the derivation of the gradient vector

$$\begin{aligned} \nabla_{\hat{\theta}}[e_{OE}^2(n)] &= 2e_{OE}(n)\nabla_{\hat{\theta}}[e_{OE}(n)] \\ &= -2e_{OE}(n)\nabla_{\hat{\theta}}[\hat{\phi}_{MOE}^T(n)\hat{\theta}(n)] \\ &\approx 2e_{OE}(n)\hat{\phi}_{MOE}(n) \end{aligned} \quad (27)$$

leading to the modified output error (MOE) algorithm described by [14]

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_{OE}(n)\hat{\phi}_{MOE}(n) \quad (28)$$

with  $\hat{\phi}_{MOE}(n)$  defined in (3). We can interpret the approximation shown in (27) as a linearization of the relationship between the output error and the adaptive coefficient vector. Since this relationship is nonlinear, where the nonlinearity is inherent to the definition of the vector  $\hat{\phi}_{MOE}(n)$ , the MOE algorithm is also called *pseudo-linear regression algorithm* [41]. The MOE algorithm has the following global convergence property:

**Property 9:** *In cases of sufficient order identification ( $n^* \geq 0$ ), the MOE algorithm may not converge to the global minimum of the MSOE performance surface, if the plant transfer function denominator polynomial does not satisfy the following strictly positive realness condition*

$$\operatorname{Re}\left[\frac{1}{A(z^{-1})}\right] > 0; |z| = 1 \quad (29)$$

Property 9 implies that the poles of the unknown system must lie inside the hyperstability region defined by (29). In general, the hyperstability region is always a subset of the

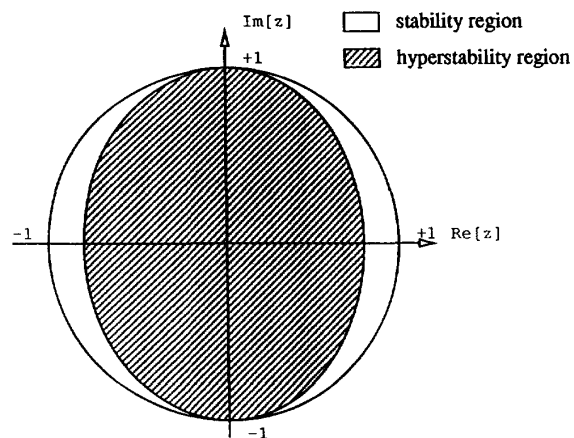


Fig. 3. Hyperstability region for the modified output error algorithm.

stability region of the complex plane  $Z$ . For example, the hyperstability region of a second order unknown system is shown in Fig. 3. Property 9 also emphasizes the fact that the MOE algorithm may converge in some cases to the optimal solution of the MSOE surface [14]. This global convergence, however, can not be assured if the plant does not satisfy (29) [26], [70]. Moreover, it must be noticed that Property 9 has limited practical use, since the unknown system denominator polynomial is not available in general. This fact constitutes a major drawback for the MOE algorithm.

### D. The SHARF Algorithm

From Fig. 3, it can be inferred that the global convergence of the MOE algorithm can not be guaranteed when the second order unknown system has poles in the neighborhood of  $z = \pm 1$ . In order to solve this general problem and make the adaptive algorithm more robust in terms of global convergence, an additional moving average filtering can be performed on the output error signal in (28), generating an error signal given by

$$e_{SHARF}(n) = [D(q^{-1})]e_{OE}(n) \quad (30)$$

with  $D(q^{-1}) = \sum_{k=1}^{n_d} d_k q^{-k}$ . The resulting algorithm is described by

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_{SHARF}(n)\hat{\phi}_{MOE}(n) \quad (31)$$

The adaptive algorithm described by this equation is commonly called SHARF, as a short name for *simple hyperstable algorithm for recursive filters*. In 1976, Landau [36] developed an algorithm for off-line system identification, based on the hyperstability theory [39], that can be considered the origin of the SHARF algorithm. In [39], some numerical examples show how the additional processing of the error signal can change the allowed region for the poles, where the algorithm global convergence is guaranteed. Basically, the SHARF algorithm has the following convergence properties [27], [36], [37]:

**Property 10:** *In cases of sufficient order identification ( $n^* \geq 0$ ), the SHARF algorithm may not converge to the global minimum of the MSOE performance surface if the plant transfer*

function denominator polynomial does not satisfy the following strictly positive realness condition

$$\operatorname{Re} \left[ \frac{D(z^{-1})}{A(z^{-1})} \right] > 0; |z| = 1 \quad (32)$$

**Property 11:** In cases of insufficient order identification ( $n^* < 0$ ), the adaptive filter output signal  $\hat{y}(n)$  and the adaptive filter coefficients vector  $\hat{\theta}$  are stable sequences, provided the input signal is sufficiently persistent exciting [1], [31].

From (30) and (31), it can be concluded that the MOE algorithm can be interpreted as a special case of the SHARF algorithm when  $D(q^{-1}) = 1$ . In this case, Property 10 becomes identical to the Property 9 associated to the MOE algorithm and Property 11 is also valid for the MOE algorithm.

The main problem of the SHARF algorithm seems to be the nonexistence of a robust practical procedure to define the moving average filter  $D(q^{-1})$  in order to guarantee the global convergence of the algorithm. This is a consequence of the fact that the condition in (32) depends on the plant denominator characteristics, that in practice are unknown.

#### E. The Modified SHARF Algorithm

In order to guarantee global convergence for the SHARF algorithm independently of the plant characteristics, Landau [38] proposed the application of a time-varying moving average filtering to the output error signal. Using Landau's approach, the modified SHARF (MSHARF) algorithm can be described by

$$e_{MSHARF}(n) = [D(q^{-1}, n)] e_{OE}(n); \quad (33)$$

with  $D(q^{-1}, n) = \sum_{k=0}^{n_d} d_k q^{-k}$

$$d_k(n+1) = d_k(n) + \mu e_{MSHARF}(n) e_{OE}(n-k); \quad (34)$$

$k = 0, \dots, n_d$

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_{MSHARF}(n) \hat{\phi}_{MOE}(n) \quad (35)$$

Another interesting interpretation of the MSHARF algorithm can be found in [35]. The next statement describes the convergence properties of the MSHARF algorithm.

**Property 12:** The MSHARF error signal  $e_{MSHARF}(n)$  is a sequence that converges to zero in the mean sense if  $n^* \geq 0$  and  $\mu$  satisfies

$$0 < \mu < \frac{1}{\|\hat{\phi}_{MSHARF}(n)\|^2} \quad (36)$$

where  $\hat{\phi}_{MSHARF}(n)$  is the extended information vector defined as

$$\hat{\phi}_{MSHARF}(n) = [\hat{y}(n-i) | x(n-j) | e_{MSHARF}(n-k)]^T \quad (37)$$

It should be mentioned that if the signal  $e_{MSHARF}(n)$  tends to zero, the output error  $e_{OE}(n)$  signal does not necessarily tend to zero. In fact, it was shown in [29] that the minimum phase condition of  $D(q^{-1}, n)$  must also be satisfied in order to guarantee that  $e_{OE}(n)$  converges to zero in the mean sense. This additional condition implies that a continuous minimum phase monitoring should be performed in the polynomial

$D(q^{-1}, n)$  to assure global convergence of the MSHARF algorithm. This fact prevents the general use of the MSHARF algorithm in practice.

It is also important to mention that although the members of the SHARF family of adaptive algorithms, that includes the MOE, SHARF, and MSHARF algorithms, attempt to minimize the output error signal, these algorithms do not present a gradient descent convergence characteristic, since they were derived by using a convergence concept from the hyperstability theory.

#### F. The Steiglitz and McBride Algorithm

In [63], Steiglitz and McBride developed an adaptive algorithm attempting to combine the good characteristics of the output error and equation error algorithms, namely unbiased and unique global solution, respectively. In order to achieve these properties, the so called Steiglitz and McBride (SM) algorithm is based on an error signal  $e_{SM}(n)$  that is a linear function of the adaptive filter coefficients, yielding a unimodal performance surface, and has a physical interpretation similar to the output error signal, leading to an unbiased global solution. The  $e_{SM}(n)$  error signal is given by

$$\begin{aligned} e_{SM}(n) &= \left[ \frac{\hat{A}(q^{-1}, n)}{\hat{A}(q^{-1}, n-1)} \right] e_{OE}(n) \\ &= \left[ \frac{1}{\hat{A}(q^{-1}, n-1)} \right] e_{EE}(n) \end{aligned} \quad (38)$$

The gradient vector associated to this error signal is

$$\begin{aligned} \nabla_{\hat{\theta}} [e_{SM}^2(n)] &= 2e_{SM}(n) \nabla_{\hat{\theta}} [e_{SM}(n)] \\ &= -2e_{SM}(n) \left[ \frac{1}{\hat{A}(q^{-1}, n-1)} \right] \hat{\phi}_{EE}(n) \end{aligned} \quad (39)$$

Using the small step approximation, we can write

$$\begin{aligned} \left[ \frac{1}{\hat{A}(q^{-1}, n-1)} \right] \hat{\phi}_{EE}(n) &= \left[ \frac{1}{\hat{A}(q^{-1}, n-1)} y(n-i) \right] \\ &= \left[ \frac{1}{\hat{A}(q^{-1}, n-1)} x(n-j) \right] \\ &\approx \left[ \frac{1}{\hat{A}(q^{-1}, n-i)} y(n-i) \right] \\ &= \left[ \frac{1}{\hat{A}(q^{-1}, n-j)} x(n-j) \right] \\ &= \begin{bmatrix} y^f(n-i) \\ x^f(n-j) \end{bmatrix} = \hat{\phi}_{SM}(n) \end{aligned} \quad (40)$$

leading to the updating equation

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_{SM}(n) \hat{\phi}_{SM}(n) \quad (41)$$

In [8], Fan and Jenkins presented a complete family of adaptive algorithms based on the original SM method. In fact, all members of this family are asymptotically equivalent, *i.e.*, they have identical steady state characteristics, whereas the transient properties of the convergence process may vary. The most relevant properties of the SM algorithm are:



**Property 13:** In a sufficient order case ( $n^* \geq 0$ ), if the perturbation signal  $v(n)$  is a white noise, all global minimum points of the performance surface associated to the SM algorithm are described by [64]

$$\begin{cases} \hat{A}^*(q^{-1}) = A(q^{-1})C(q^{-1}) \\ \hat{B}^*(q^{-1}) = B(q^{-1})C(q^{-1}) \end{cases} \quad (42)$$

with  $C(q^{-1}) = \sum_{k=0}^{n^*} c_k q^{-k}$ . It means that all global minimum solutions have included the polynomials describing the unknown system plus a common factor  $C(q^{-1})$  present in the numerator and denominator polynomials of the adaptive filter. On the other hand, if the perturbation signal is a colored noise, the final solution is biased when compared to the OE global solution, with the degree of bias being a function of the characteristics of the disturbance signal.

In an insufficient order case ( $n^* < 0$ ), the solution of the SM algorithm is always biased compared to the OE algorithm global solution [60]. Also, when  $n^* < 0$ , the SM algorithm may present multiple solutions [13].

It should be mentioned that when  $n^* < 0$ , although the solutions of the SM and OE algorithms are different, there are some cases where their distance is not easily measurable even by a numerical computer [13]. However, these cases are completely unpredictable and this assumption of equality can not be generalized.

**Property 14:** If the adaptive filter initial condition  $\frac{\hat{B}(q^{-1}, 0)}{\hat{A}(q^{-1}, 0)}$  is stable and the perturbation signal is a white noise, then the SM algorithm is globally stable if at least one of these conditions is satisfied [13], [58], [64]: -  $\hat{n}_a = 1$ . -  $n^* < 0$  and the signal/noise ratio  $\frac{E[y_{\hat{O}}^2(n)]}{E[v^2(n)]}$  is sufficiently small; -  $n^* \geq 0$  and the signal/noise ratio  $\frac{E[y_{\hat{O}}^2(n)]}{E[v^2(n)]}$  is sufficiently large;

This result shows that in some cases the SM algorithm does not require a pole monitoring procedure during the adaptation process to maintain stability. However, since these are specific cases, the stability monitoring is necessary in general.

#### G. The Bias Remedy Equation Error Algorithm

Analyzing the convergence of the equation error (EE) algorithm, it can be concluded that the presence of bias on the algorithm solution is due to the definition of the equation error information vector  $\hat{\phi}_{EE}(n)$  that includes past samples of the unknown system output  $y(n)$ . This signal  $y(n)$  includes information related to the perturbation signal  $v(n)$ . One way to avoid this bias on the global solution could be obtained by subtracting the perturbation signal from the EE information vector. However, in practice this additive noise signal is not directly available. An alternative solution in this case is to use the output error signal, since this signal can be considered a good estimate for the perturbation noise as the adaptive process converges. Hence, using this technique, the information vector can be expressed as

$$\hat{\phi}_{BRLE}(n) = \hat{\phi}_{EE}(n) - \tau \mathbf{e}_{OE}(n) \quad (43)$$

where  $\mathbf{e}_{OE}(n) = [e_{OE}(n-i) \mid \underbrace{0 \dots 0}_{(n_b+1)}]^T$ , and the parameter  $\tau$  is used to control the amount of bias that is eliminated and the stability of the adaptive algorithm.

Using the approach above described, Lin and Unbehauen [40] developed the so called *bias remedy least-mean-square equation error* (BRLE) algorithm described by

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_{EE}(n) \hat{\phi}_{BRLE}(n) \quad (44)$$

The properties of the BRLE algorithm are [40]:

**Property 15:** The BRLE algorithm is globally stable if all the conditions below are satisfied -  $0 \leq \tau \leq \min\left(k \frac{\|\hat{\phi}_{BRLE}(n)\|}{\|\mathbf{e}_{OE}(n)\|}, 1\right)$ , with a finite  $k > 0$ ; -  $0 < \mu < \min\left(\frac{2}{\lambda_{max}}, \sigma\right)$ , where  $\lambda_{max}$  is the largest eigenvalue of  $E[\hat{\phi}_{BRLE}(n) \hat{\phi}_{BRLE}^T(n)]$ , and  $\sigma$  is a sufficiently small positive number.

The first condition establishes an upper limit for  $\tau$ , less or equal to one, necessary to guarantee the stability of the BRLE algorithm. In practice, there is a trade-off between bias and stability of the BRLE algorithm: The larger  $\tau$  is, less biased is the algorithm, however the more unstable the convergence tends to be. The second stability condition presents the range of the variable  $\mu$  that guarantees a stable global convergence of the BRLE algorithm.

**Property 16:** In a sufficient order case ( $n^* \geq 0$ ), if  $\tau = 1$  and the conditions of the previous property are satisfied, then the BRLE algorithm converges to a solution described by

$$\begin{cases} \hat{A}^*(q^{-1}) = A(q^{-1})C(q^{-1}) \\ \hat{B}^*(q^{-1}) = B(q^{-1})C(q^{-1}) \end{cases} \quad (45)$$

with  $C(q^{-1}) = \sum_{k=0}^{n^*} c_k q^{-k}$ . It means that all solutions have included the polynomials describing unknown system plus a common factor  $C(q^{-1})$  present in the numerator and denominator polynomials of the adaptive filter.

#### H. The Composite Regressor Algorithm

The Steiglitz and McBride (SM) algorithm previously described can be considered the first attempt to combine the characteristics of two distinct algorithms, namely equation error and output error. In the SM algorithm, the composite characteristic is implicit in the definition of the  $e_{SM}(n)$  error signal. In [34], Kenney and Rohrs proposed the *composite regressor* (CR) algorithm based on the idea of combining two elementary adaptive algorithms, the equation error and modified output error. This method, however, uses a weighting parameter that allows a better control of the final characteristics of the adaptive algorithm. The CR algorithm is described by

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_{CR}(n) \hat{\phi}_{CR}(n) \quad (46)$$

where  $e_{CR}(n)$  and  $\hat{\phi}_{CR}(n)$  are given by weighting of the respective error signals and information vectors of the EE and MOE algorithms as follows

$$e_{CR}(n) = \left[ \frac{\hat{A}(q^{-1}, n)}{\gamma + \hat{A}(q^{-1}, n)(1 - \gamma)} \right] e_{OE}(n) \quad (47)$$

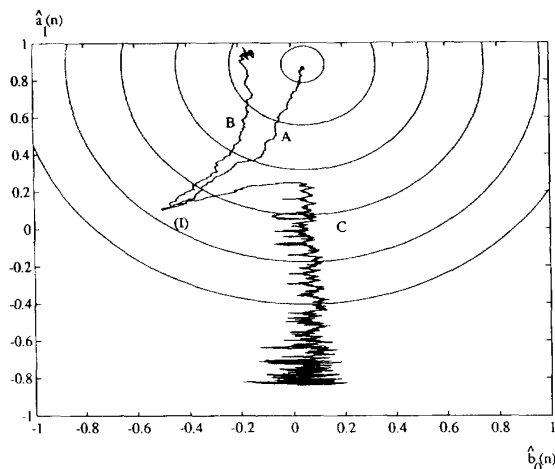


Fig. 4. Trajectories of the adaptive algorithms with initial point (I) =  $[-0.5; 0.1]$  on the MSE performance surface: A - EE ( $\mu = 0.001$ ), B - CE ( $\mu = 0.001$ ;  $\beta = 0.8$ ), and C - BRLE ( $\mu = 0.001$ ;  $\tau = 1$ ).

$$\hat{\phi}_{CR}(n) = \gamma \hat{\phi}_{EE}(n) + (1 - \gamma) \hat{\phi}_{MOE}(n) \quad (48)$$

with  $0 \leq \gamma \leq 1$ .

In terms of convergence characteristics, clearly the CR algorithm must present intermediate properties between its two basic algorithms, EE ( $\gamma = 1$ ) and MOE ( $\gamma = 0$ ), as is shown in [34].

#### I. The Composite Error Algorithm

Another algorithm for IIR adaptive filtering that applies Kenney's technique of explicit combination was presented in [50] using the equation error and the original output error algorithms. The so called *composite error* (CE) algorithm is described by

$$e_{CE}(n) = \beta e_{EE}(n) + (1 - \beta) e_{OE}(n) \quad (49)$$

$$\hat{\phi}_{CE}(n) = \beta \hat{\phi}_{EE}(n) + (1 - \beta) \hat{\phi}_{OE}(n) \quad (50)$$

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \mu e_{CE}(n) \hat{\phi}_{CE}(n) \quad (51)$$

The main advantage of combining the EE and OE algorithms is to obtain a gradient descent algorithm that can present good performance even in cases of insufficient order identification. In fact, the CE algorithm as presented in (49)-(51) can be interpreted as an on-line version of a graduated nonconvexity (GNC) method [65] if the weighting parameter  $\beta$  is made time-varying with value decreasing from  $\beta = 1$  to  $\beta = 0$ . In [18], Gerald *et al.* present a slightly modified version of the CE algorithm for the echo cancellation configuration. In [20], Hall and Hughes present another special case of the CE algorithm where the weighting parameter  $\beta$  repeatedly switches its value from  $\beta = 1$  to  $\beta = 0$  and *vice-versa*. Although these distinct versions may be visualized as algorithms with a common central idea, it must be noticed that each method has individual characteristics of implementation and convergence behavior that are presented in the respective literature.

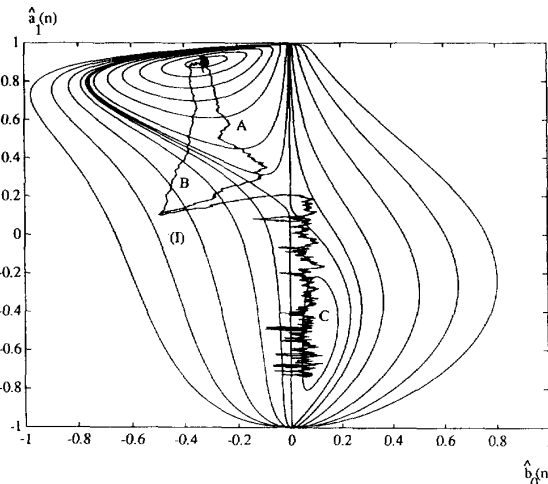


Fig. 5. Trajectories of the adaptive algorithms with initial point (I) =  $[-0.5; 0.1]$  on the MSOE performance surface: A - OE ( $\mu = 0.001$ ), B - SM ( $\mu = 0.0005$ ), and C - MOE ( $\mu = 0.003$ ).

#### IV. SIMULATIONS

Although there are many important works [9], [10], [12], [30] on the convergence of adaptive algorithms, the most common form to analyze the convergence process characteristics is through the use of computational examples. In this section, some computational simulations are presented in order to illustrate some of the properties associated to the algorithms used in adaptive IIR filtering. The example chosen consists of the identification of a second-order plant with transfer function [26], [51]

$$H(z) = \frac{0.05 - 0.4z^{-1}}{1 - 1.1314z^{-1} + 0.25z^{-2}} \quad (52)$$

with an adaptive filter described by

$$\hat{H}(z, n) = \frac{\hat{b}_0(n)}{1 - \hat{a}_1(n)z^{-1}} \quad (53)$$

leading to an insufficient order identification problem. The input signal is a Gaussian white noise with zero mean and unity variance.

A summary of the results for the different algorithms is given in Table II. The characteristics of the solution obtained by each algorithm are following described:

The EE algorithm converged to the optimum point of the MSE surface, as illustrated in Fig. 4. The solution is biased with respect to the MSOE global optimum due to the insufficient order nature of the problem.

The MSOE performance surface is multimodal, what makes the convergence of the OE algorithm to the global optimum of the adaptive filter conditioned to the initial point (see Fig. 5 and Fig. 6).

The MOE algorithm presented an unacceptable behavior due to the insufficient order nature of the identification process, converging to an apparently meaningless stationary point. Observe in Fig. 5 that the MOE algorithm does not even

TABLE II  
SIMULATION CHARACTERISTICS

Adaptive Algorithm	Algorithm Parameters	Initial Point	Number of Iterations	Final Point
EE	$\mu = 0.001$	$[-0.5; +0.1]$	$\approx 6000$	$[+0.050; +0.884]$
OE	$\mu = 0.001$	$[-0.5; +0.1]$	$\approx 4000$	$[-0.311; +0.906]$
OE	$\mu = 0.003$	$[+0.5; -0.2]$	$\approx 10000$	$[+0.114; -0.519]$
MOE	$\mu = 0.003$	$[-0.5; +0.1]$	$\approx 40000$	$[+0.050; -0.852]$
SHARF	$\mu = 0.003; D(q^{-1}) = 1 - q^{-1}$	$[-0.5; +0.1]$	$\approx 40000$	$[+0.050; -0.852]$
MSHARF	$\mu = 0.003; n_d = 1$	$[-0.5; +0.1]$	$\approx 40000$	$[+0.050; -0.852]$
SM	$\mu = 0.0005$	$[-0.5; +0.1]$	$\approx 3000$	$[-0.312; +0.905]$
SM	$\mu = 0.0005$	$[+0.5; -0.2]$	$\approx 4000$	$[-0.312; +0.905]$
BRLE	$\mu = 0.001; \tau = 0.5$	$[-0.5; +0.1]$	$\approx 8000$	$[+0.050; +0.827]$
BRLE	$\mu = 0.003; \tau = 1.0$	$[-0.5; +0.1]$	$\approx 80000$	$[+0.049; -0.826]$
CE	$\mu = 0.001; \beta = 0.04$	$[-0.5; +0.1]$	$\approx 4000$	$[-0.306; +0.907]$
CE	$\mu = 0.001; \beta = 0.60$	$[-0.5; +0.1]$	$\approx 3000$	$[-0.146; +0.942]$
CE	$\mu = 0.003; \beta = 0.04$	$[+0.5; -0.2]$	$\approx 22000$	$[-0.306; +0.907]$

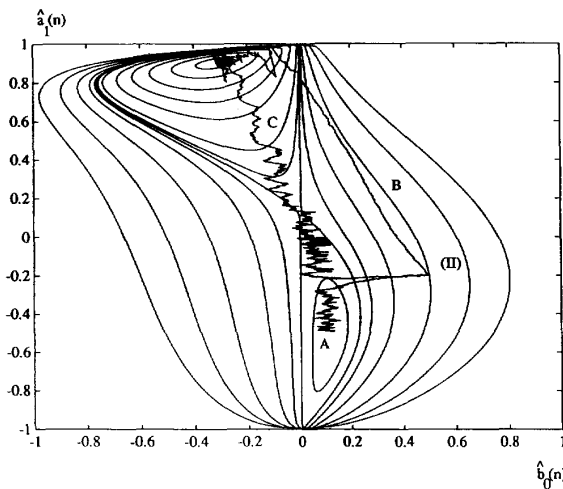


Fig. 6. Trajectories of the adaptive algorithms with initial point (II) =  $[0.5; -0.2]$  on the MSOE performance surface: A - OE ( $\mu = 0.003$ ), B - SM ( $\mu = 0.0005$ ), and C - CE ( $\mu = 0.001; \beta = 0.04$ ).

converge to the local minimum of the MSOE performance surface.

The SHARF and MSHARF convergence properties are dependent on the additional MA filtering characteristics. In this set of simulations, the order of the MA filter was chosen equals  $n_d = 1$  for both algorithms. For the SHARF algorithm, where the filter is time-invariant, the additional polynomial was set to  $D(q^{-1}) = 1 - q^{-1}$ . Since  $n^* < 0$  the consistency of the SHARF algorithm can not be guaranteed. In fact, as can be observed from Table II, both algorithms SHARF and MSHARF presented a poor convergence behavior similar to the MOE algorithm.

Despite the insufficient order adaptive filter, the SM method converged to a point extremely close to the MSOE global minimum point, independent of the initial point, as can be

verified in Table II. However, as was pointed out before in this paper, this excellent behavior does not occur in general and in fact can not be predicted in practice.

The BRLE algorithm presented a convergence behavior similar to the EE algorithm when the bias-remedy parameter  $\tau$  was made smaller than 0.5 with  $\mu = 0.001$ . However, with  $\tau \geq 0.6$  the algorithm became extremely slow and with  $\tau \approx 1$  it started converging to a meaningless point, as shown in Fig. 4.

The simulations with the CR algorithm were not included here, since both EE and MOE algorithms presented convergence problems. The same problems are expected to occur with the CR algorithm.

The CE algorithm presented excellent properties when the weighting parameter was kept in the interval  $0.04 \leq \beta < 1$ . In fact, within those limits the performance surface associated to the algorithm is unimodal and the bias between its minimum and the MSOE global minimum is negligible, as can be inferred from Table II. However, with  $\beta \approx 1$  the convergence point presents a significant bias with respect to the MSOE global minimum (see Fig. 6).

From the previous example, the reader can not conclude that a given algorithm is the best choice in general, since in slightly distinct situations the answer could be different. In fact, the example presented is not meant to be conclusive. In practice, the choice of the most appropriate adaptive IIR algorithm is not an easy task. The properties of the algorithms as analyzed in this paper are certainly useful tools and can be used to choose the most adequate algorithm.

## V. CONCLUSION

The purpose of this review paper has been to outline some of the issues involved in the choice of an adaptive IIR filtering algorithm. Several well known algorithms have been presented in a unified form. Emphasis was placed in providing a simple and general framework that enables easy understanding of the interrelationships and convergence properties of the

algorithms. Simulations were included to illustrate some of the results surveyed.

#### ACKNOWLEDGMENT

The authors are grateful to Mr. Juan E. Cousseau at COPPE - Federal University of Rio de Janeiro, Brazil, for his many interesting discussions during the preparation of the present work.

#### REFERENCES

- [1] B. D. O. Anderson and C. R. Johnson, Jr., "On reduced-order adaptive output error identification and adaptive IIR filtering," *IEEE Trans. Automat. Cont.*, vol. AC-27, no. 4, pp. 927-933, Aug. 1982.
- [2] K. J. Åström and P. Eykhoff, "System identification—A survey," *Automatica*, vol. 7, no. 2, pp. 123-162, Mar. 1971.
- [3] K. J. Åström and T. Söderström, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model," *IEEE Trans. Automat. Cont.*, vol. AC-19, no. 6, pp. 769-773, Dec. 1974.
- [4] J. A. Cadzow, "Recursive digital filter synthesis via gradient based algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, no. 5, pp. 349-355, Oct. 1976.
- [5] J. E. Cousseau and P. S. R. Diniz, "Alternative parallel realization for adaptive IIR filters," in *Proc. Int. Symp. Circuits Syst.*, 1990, pp. 1927-1930.
- [6] R. A. David, "A modified cascade structure for IIR adaptive algorithms," in *Proc. 15th Asilomar Conf. Circuits, Syst., Comput.*, 1981, pp. 175-179.
- [7] M. Dentino, J. M. McCool, and B. Widrow, "Adaptive filtering in the frequency domain," *Proc. IEEE*, vol. 66, no. 12, pp. 1658-1659, Dec. 1978.
- [8] H. Fan and W. K. Jenkins, "A new adaptive IIR filter," *IEEE Trans. Circuits Syst.*, vol. CAS-33, no. 10, pp. 939-947, Oct. 1986.
- [9] H. Fan, "A unified analysis of direct form adaptive IIR filters using an ODE method," in *Proc. 21st Asilomar Conf. Circuits, Syst., Comput.*, 1987, pp. 139-143.
- [10] ———, "Application of Benveniste's convergence results in a study of adaptive IIR filtering algorithms," *IEEE Trans. Inform. Theory*, vol. 34, no. 4, pp. 692-709, July 1988.
- [11] H. Fan and M. Nayeri, "On error surfaces of sufficient order adaptive IIR filters: Proofs and counterexamples to a unimodality conjecture," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 9, pp. 1436-1442, Sept. 1989.
- [12] H. Fan and Y. Yang, "Analysis of a frequency-domain adaptive IIR filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 4, pp. 864-870, May 1990.
- [13] H. Fan and M. Nayeri, "On reduced order identification; revisiting 'On some system identification techniques for adaptive filtering'," *IEEE Trans. Circuits Syst.*, vol. 37, no. 9, pp. 1144-1151, Sept. 1990.
- [14] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, no. 11, pp. 1622-1624, Nov. 1976.
- [15] E. R. Ferrara, "Fast implementation of LMS adaptive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 4, pp. 474-475, Aug. 1980.
- [16] B. Friedlander, "System identification techniques for adaptive signal processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 2, pp. 240-246, Apr. 1982.
- [17] S. Gee and M. Rupp, "A comparison of adaptive IIR echo canceller hybrids," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1991, pp. 1541-1544.
- [18] J. A. B. Gerald, N. L. Esteves, and M. M. Silva, "An adaptive IIR echo canceller combining output-error and equation-error criteria," in *Proc. Int. Symp. Circuits Syst.*, 1990, pp. 779-783.
- [19] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [20] M. C. Hall and P. M. Hughes, "The master-slave IIR filter adaptation algorithm," in *Proc. Int. Symp. Circuits Syst.*, 1988, pp. 2145-2148.
- [21] S. S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [22] S. Horvath, Jr., "A new adaptive recursive LMS filter," *Digital Signal Processing*, V. Cappellini and A. G. Constantinides, Eds. San Diego, CA: Academic Press, 1980, pp. 21-26.
- [23] T. C. Hsia, "A simplified adaptive recursive filter design," *Proc. IEEE*, vol. 69, no. 9, pp. 1153-1155, Sept. 1981.
- [24] A. W. Hull and W. K. Jenkins, "A comparison of two IIR filter structures for adaptive digital filtering," in *Proc. Int. Symp. Circuits Syst.*, 1988, pp. 527-530.
- [25] D. A. Johns, W. M. Snelgrove, and A. S. Sedra, "Nonideal effects in continuous-time adaptive recursive filters," in *Proc. 32nd. Midwest Symp. Circuits Syst.*, 1990, pp. 594-587.
- [26] C. R. Johnson Jr., and M. G. Larimore, "Comments on and additions to 'An adaptive recursive LMS filter'," *Proc. IEEE*, vol. 65, no. 9, pp. 1399-1402, Sept. 1977.
- [27] C. R. Johnson Jr., "A convergence proof for a hyperstable adaptive recursive filter," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 6, pp. 745-749, Nov. 1979.
- [28] ———, "An output error identification interpretation of model reference adaptive control," *Automatica*, vol. 16, pp. 419-421, July 1980.
- [29] C. R. Johnson Jr., and T. Taylor, "Failure of a parallel adaptive identifier with adaptive error filtering," *IEEE Trans. Automat. Cont.*, vol. AC-25, no. 6, pp. 1248-1250, Dec. 1980.
- [30] 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, no. 3, pp. 659-670, June 1981.
- [31] C. R. Johnson, Jr., and B. D. O. Anderson, "Sufficient excitation and stable reduced-order adaptive IIR filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. 6, pp. 1212-1215, Dec. 1981.
- [32] C. R. Johnson Jr., "The common parameter estimation basis of adaptive filtering, identification, and control," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, no. 4, pp. 587-595, Aug. 1982.
- [33] ———, "Adaptive IIR filtering: Current results and open issues," *IEEE Trans. Inform. Theory*, vol. IT-30, no. 2, pp. 237-250, Mar. 1984.
- [34] J. B. Kenney and C. E. Rohrs, "The composite regressor algorithm for IIR adaptive systems," *Trans. Signal Processing*, vol. 41, no. 2, pp. 617-628, Feb. 1993.
- [35] K. Kurosawa and S. Tsujii, "An IIR parallel-type adaptive algorithm using the fast least squares method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 8, pp. 1226-1230, Aug. 1989.
- [36] I. D. Landau, "Unbiased recursive identification using model reference adaptive techniques," *IEEE Trans. Automat. Cont.*, vol. AC-21, no. 2, pp. 194-202, Apr. 1976.
- [37] I. D. Landau, "An addendum to 'Unbiased recursive identification using model reference adaptive techniques'," *IEEE Trans. Automat. Cont.*, vol. AC-23, no. 1, pp. 97-99, Feb. 1978.
- [38] I. D. Landau, "Elimination of the real positivity condition in the design of parallel MRAS," *IEEE Trans. Automat. Cont.*, vol. AC-23, no. 6, pp. 1015-1020, Dec. 1978.
- [39] 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, no. 4, pp. 428-440, Aug. 1980.
- [40] J.-N. Lin and R. Unbehauen, "Bias-remedy least mean square equation error algorithm for IIR parameter recursive estimation," *IEEE Trans. Signal Processing*, vol. 40, no. 1, pp. 62-69, Jan. 1992.
- [41] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [42] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1973.
- [43] D. F. Marshall, W. K. Jenkins, and J. J. Murphy, "The use of orthogonal transforms for improving performance of adaptive filters," *IEEE Trans. Circuits Syst.*, vol. 36, no. 4, pp. 474-484, Apr. 1989.
- [44] U. Menzi and G. S. Moschytz, "Adaptive switched capacitor filters based on the LMS algorithm," *Proc. Int. Symp. Circuits Syst.*, 1990, pp. 2220-2223.
- [45] S. S. Narayan and A. M. Peterson, "Frequency domain least-mean-square algorithm," *Proc. IEEE*, vol. 69, no. 1, pp. 124-126, Jan. 1981.
- [46] S. S. Narayan, A. M. Peterson, and N. J. Narasimha, "Transform domain LMS algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, no. 3, pp. 609-615, June 1983.
- [47] M. Nayeri and W. K. Jenkins, "Alternate realizations to adaptive IIR filters and properties of their performance surfaces," *IEEE Trans. Circuits Syst.*, vol. 36, no. 4, pp. 485-496, Apr. 1989.
- [48] M. Nayeri, "Uniqueness of MSOE estimates in IIR adaptive filtering; a search for necessary conditions," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1989, pp. 1047-1050, 1989.
- [49] M. Nayeri, H. Fan, and W. K. Jenkins, "Some characteristics of error surfaces for insufficient order adaptive IIR filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 7, pp. 1222-1227, July 1990.
- [50] S. L. Netto and P. S. R. Diniz, "Composite algorithms for adaptive IIR filtering," *IEE Electron. Lett.*, vol. 28, no. 9, pp. 886-888, Apr. 1992.
- [51] D. Parikh and N. Ahmed, "On an adaptive algorithm for IIR filters," *Proc. IEEE*, vol. 66, no. 5, pp. 585-588, May 1978.

- [52] P. A. Regalia, "Stable and efficient lattice algorithms for adaptive IIR filtering," *IEEE Trans. Signal Processing*, vol. 40, no. 2, pp. 375-388, Feb. 1992.
- [53] J. J. Shynk and R. P. Gooch, "Frequency-domain adaptive pole-zero filtering," *Proc. IEEE*, vol. 73, no. 10, pp. 1526-1528, Oct. 1985.
- [54] J. J. Shynk, "Adaptive IIR filtering using parallel-form realizations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 4, pp. 519-533, Apr. 1989.
- [55] J. J. Shynk, "Adaptive IIR filtering," *IEEE Acoust., Speech, Signal Processing Mag.*, vol. 6, no. 2, pp. 4-21, Apr. 1989.
- [56] T. Söderström, "On the uniqueness of maximum likelihood identification," *Automatica*, vol. 11, no. 2, pp. 193-197, Mar. 1975.
- [57] T. Söderström, L. Ljung, and I. Gustavsson, "A theoretical analysis of recursive identification methods," *Automatica*, vol. 14, no. 3, pp. 231-244, May 1978.
- [58] T. Söderström and P. Stoica, "On the stability of dynamic models obtained by least-squares identification," *IEEE Trans. Automat. Cont.*, vol. AC-26, no. 2, pp. 575-577, Apr. 1981.
- [59] ———, "Some properties of the output error method," *Automatica*, vol. 18, no. 1, pp. 93-99, Jan. 1982.
- [60] ———, "On some system identification techniques for adaptive filtering," *IEEE Trans. Circuits Syst.*, vol. 35, no. 4, pp. 457-461, Apr. 1988.
- [61] S. D. Stearns, G. R. Elliott, and N. Ahmed, "On adaptive recursive filtering," in *Proc. 10th Asilomar Conf. Circuits, Syst., Comput.*, 1976, pp. 5-11.
- [62] 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.
- [63] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Automat. Cont.*, vol. AC-10, no. 4, pp. 461-464, Oct. 1965.
- [64] P. Stoica and T. Söderström, "The Steiglitz-McBride identification algorithm revisited—Convergence analysis and accuracy aspects," *IEEE Trans. Automat. Cont.*, vol. AC-26, no. 3, pp. 712-717, June 1981.
- [65] V. L. Stonick and S. T. Alexander, "Numerical methods for globally optimal adaptive IIR filtering," in *Proc. 22th Asilomar Conf. Circuits, Syst., Comput.*, 1990, pp. 745-749.
- [66] J. R. Treichler, C. R. Jonhson Jr., and M. G. Larimore, *Theory and Design of Adaptive Filters*. New York: Wiley, 1987.
- [67] C.-P. J. Tzeng, "An adaptive offset cancellation technique for adaptive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 5, pp. 799-803, May 1990.
- [68] B. E. Usevitch and W. K. Jenkins, "A cascade implementation of a new IIR adaptive digital filter with global convergence and improved convergence rates," in *Proc. Int. Symp. Circuits Syst.*, 1989, pp. 2140-2143.
- [69] S. A. White, "An adaptive recursive digital filter," in *Proc. 9th Asilomar Conf. Circuits, Syst., Comput.*, 1975, pp. 21-25, 1975.
- [70] B. Widrow and J. M. McCool, "Comments on 'An adaptive recursive LMS filter'," *Proc. IEEE*, vol. 65, no. 9, pp. 1402-1404, Sept. 1977.
- [71] B. Widrow and S. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.



**Sergio L. Netto** received the B.Sc. degree in electronic engineering from the Universidade Federal do Rio de Janeiro (UFRJ) and the M.Sc. degree in electrical engineering from COPPE/UFRJ, Brazil, in 1991 and 1992, respectively.

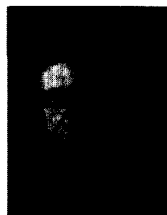
He is presently working towards the Ph.D. degree in electrical engineering at the University of Victoria, B.C., Canada.

His research interests are in IIR and 2-D adaptive filtering.



**Paulo S. R. Diniz** (S'80, M'84, SM'92) received the B.Sc. degree from the Universidade Federal do Rio de Janeiro (UFRJ) in 1979, the M.Sc. degree from COPPE/UFRJ in 1981, and the Ph.D. degree from Concordia University, Montreal, P.Q., Canada, in 1984, all in electrical engineering.

Since 1979 he has been with the Department of Electronic Engineering, UFRJ. He has also been with the Program of Electrical Engineering, COPPE/UFRJ, since 1984, where he is currently a Professor. From January 1991 to July 1992, he was a visiting Research Associate in the Department of Electrical and Computer Engineering at the University of Victoria, B.C., Canada. His teaching and research interests are in analog and digital signal processing, electronic circuits and adaptive signal processing.



**Panajotis Agathoklis** (M'81, SM'89) received the Dipl. Ing. degree in electrical engineering and the Dr.Sc.Tech. degree from the Swiss Federal Institute of Technology, Zurich, Switzerland, in 1975 and 1980, respectively.

From 1981 until 1983, he was with the University of Calgary as a Post-Doctoral Fellow and part-time instructor. Since 1983, he has been with the Department of Electrical and Computer Engineering, University of Victoria, B.C., Canada, where he is currently a Professor. His fields of interest are

multidimensional systems, control systems, signal processing and computer graphics.

He has received a NSERC University Research Fellowship (1984-1986) and Visiting Fellowships from the Swiss Federal Institute of Technology (1982, 1984, 1986 and 1993) and from the Australian National University (1987). He has been Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS since 1990.