ALGORITHMS AND STRUCTURES OF ADAPTIVE FILTERING: A REVIEW

Azzedine Zerguine†
Department of Physics

and

Maamar Bettayeb*
Department of Electrical Engineering

King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia

الخلاصـــة :

طُـبِّقـتْ تقنيـة المرشحات المتكيفة بنجاح في مـجالات مـهمـة وكثيرة، منها تنقيـة الصـوت من الصدى، والمسوِّيات في الاتصالات الرقمية، وموانع الضجيج.

ولقد أصبح التصميم المباشر للمرشحات المتكيفة ممكنا وذلك بفضل النمو السريع في تكنولوجيا الكمبيوتر، وتزايد اهتمام الباحثين لتطوير خوارزميات للمرشحات المتكيفة لأوضاع صعبة مثل الخبو الدوري القوي، والتشوه الكبير في الإشارات.

ونُقدم في هذه الورقة أحدث خوارزميات المرشحات المتكيفة، كما نعرض بالتفصيل الخوارزميات المعتمدة على طرق LMS و RLS. وخاصّةً خوارزميات المرشحات المتكيفة المعتمدة على طرق LMF والمعايير المختلفة. وقد تَمَّ منافشة متانة وعيوب هذه الخوارزميات، ومن ثم عرضت نتائج البحث المعتمدة على محاكاة بعض الخوارزميات المشهورة.

*Address for correspondence:
KFUPM Box 1730
King Fahd University of Petroleum & Minerals
Dhahran 31261
Saudi Arabia
e-mail:maamarab@kfupm.edu.sa

†e-mail:azzedine@kfupm.edu.sa

ABSTRACT

Adaptive filtering techniques have been successfully implemented in various important applications such as echo cancellation, equalization, noise cancellation, and others.

The rapid growth of computer technology made the on line implementation of adaptive filters possible and raised the interest of researchers in the development of new adaptive filtering algorithms for more challenging situations such as strong fading and large signal and noise distortions.

In this paper, the authors present the state of the art of adaptive filtering algorithms. Both LMS and RLS families of algorithms are detailed. However, emphasis is also given to recently proposed least mean fourth (LMF) and mixed norm criteria based adaptive filtering algorithms. The strengths and drawbacks of the above algorithms are also discussed. Simulation results are given for some popular algorithms.

ALGORITHMS AND STRUCTURES OF ADAPTIVE FILTERING: A REVIEW

1. INTRODUCTION

Adaptive systems are playing a vital role in the development of modern telecommunications. Also, adaptive systems proved to be extremely effective in achieving high efficiency, high quality, and high reliability of around-the-world ubiquitous telecommunication services.

The role of adaptive systems is widespread covering almost all aspects of telecommunication engineering, but perhaps most notable in the context [1] of ensuring high-quality signal transmission over unknown and time varying channels.

Interest in adaptive filters continues to grow as they find practical real-time applications in areas such as echo cancellation [2], channel equalization [3], noise cancellation [4, 5], and many other adaptive signal processing applications. This is due mainly to the recent advances in very large-scale integration (VLSI) technology.

The key to successful adaptive signal processing is understanding the fundamental properties of adaptation algorithms. These properties are stability, speed of convergence, misadjustement errors, robustness to both additive noise and signal conditioning (spectral coloration), numerical complexity, and round-off error analysis of adaptive filtering algorithms. However, some of these properties are often in direct conflict with each other, since consistently fast converging algorithms tend to be in general more complex and numerically sensitive. Also, the performance of any algorithm with respect to any of these criteria is entirely dependent on the choice of the adaptation update function, that is the cost function used in the minimization process. A compromise must be then reached among these conflicting factors to come up with the appropriate algorithm for the concerned application.

After presenting, in Section 2, the common adaptive system configurations using adaptive filters, Section 3 will deal with a more explicit development of adaptive filters. Performance evaluation of the resulting algorithms using the properties of the finite-duration impulse response (FIR) adaptive filter are also mentioned.

Section 4 reviews the theory of adaptive filtering algorithms used with these filters, including the least mean squares (LMS), the least mean fourth (LMF), the mixed-norm (MN), and the recursive least squares (RLS) algorithms. Also, the shortcomings of these algorithms are discussed.

In Section 5, the above algorithms are evaluated and compared. Areas of future research are discussed in Section 6. Finally, a summary is given in Section 7.

2. APPLICATIONS OF ADAPTIVE FILTERS

Adaptive filtering has been successfully applied in such diverse fields as communications, radar, sonar, and biomedical engineering. Although these applications are indeed quite different in nature, nevertheless, they have one basic common feature: an input signal and a desired response are used to compute the error, which is in turn used to control the values of a set of adjustable filter coefficients. However, the main difference among the various applications of adaptive filtering arises in the manner in which the desired response is extracted.

In this context, we may classify an adaptive filter into one of the four following categories:

2.1. System Identification

In this first application, depicted in Figure 1, the adaptive filter is used to provide a linear model that represents the best fit to the unknown system. Both the adaptive filter and the unknown system are driven by the same input. The error estimate is used to update the filter coefficients of the adaptive filter. After convergence, the adaptive filter output will approximate the output of the unknown system in an optimum sense. Provided that the order of the adaptive filter matches that of the unknown system and the input, $x(n)$, is broad band (flat spectrum) this will be achieved by convergence of adaptive filter coefficients to the same values as the unknown system.

The major practical use of this structure in telecommunications is for echo cancellation [2, 6, 7]. Typically, the input signal $x(n)$ will be either speech or data.

2.2. Inverse Modeling

In this second class of applications, the function of the adaptive filter is to provide an inverse model that represents the best fit to the unknown system. Thus, at convergence, the adaptive filter transfer function approximates the inverse of the transfer function of the unknown system. In practice, a delay may have to be introduced into the desired response path as shown in Figure 2, so as to ensure that the input to the adaptive filter is minimum phase and suitable for equalization by a linear structure.

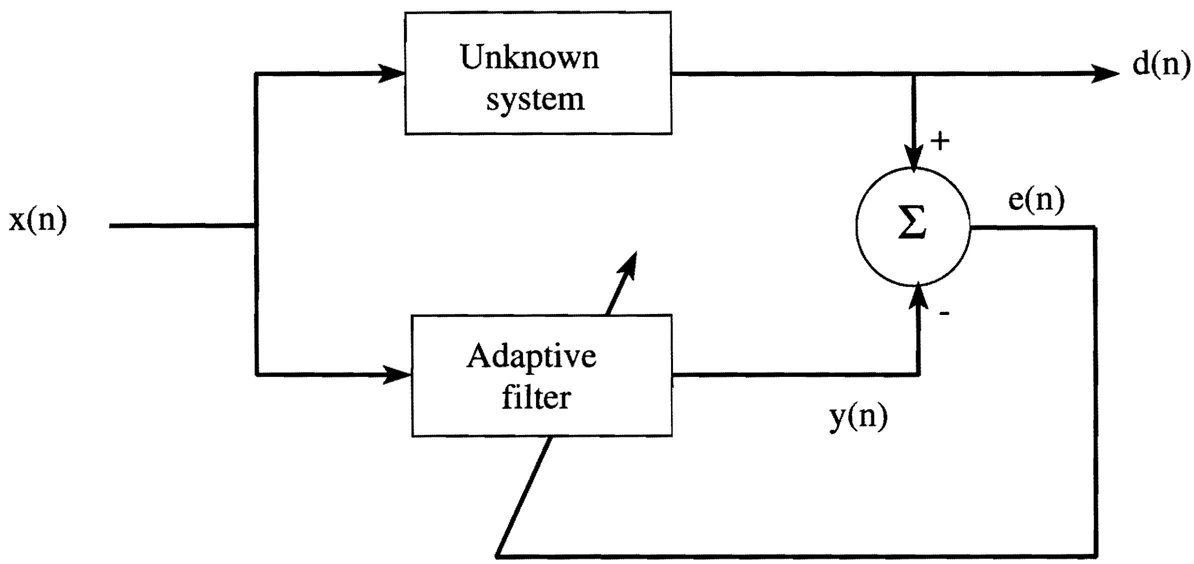

Figure 1. Direct System Modeling Configuration of an Adaptive Filter.

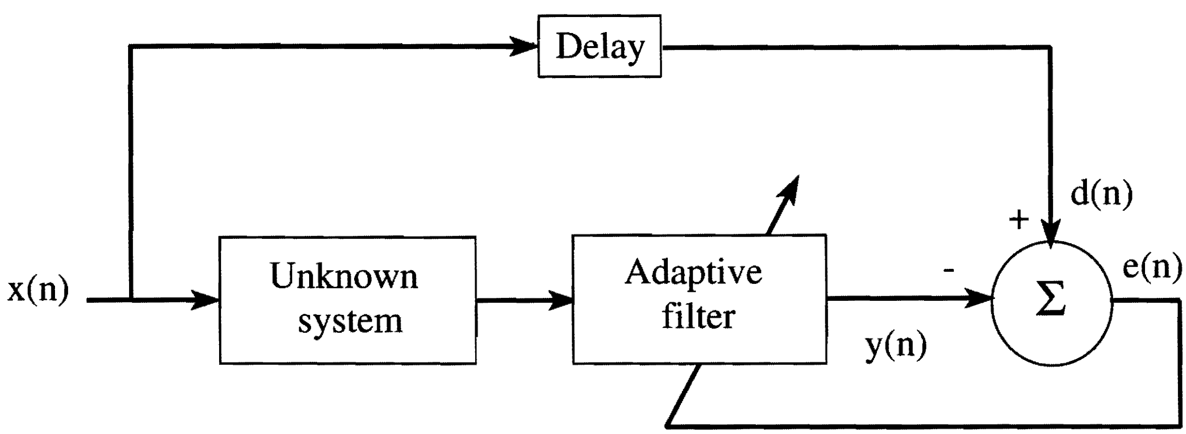

Figure 2. Inverse System Modeling Configuration of an Adaptive Filter.

The primary use of inverse system modeling is for reducing the effects of intersymbol interference (ISI) in digital receivers. This is achieved through the use of channel equalization for digital communications [3], [8–12] allowing faster data rates with an acceptably low probability of error.

2.3. Prediction

In this structure, the function of the adaptive filter is to provide the best prediction of the present value of the input signal from its previous values. The configuration shown in Figure 3 is used for this purpose, where the desired signal, $d(n)$, is the instantaneous value and the input to the adaptive filter is a delayed version of the same signal.

This application is widely used in linear predictive coding (LPC) of speech [13, 14] and in adaptive differential pulse-code modulation (DPCM) [15]. Another approach to prediction is given in [16].

2.4. Noise Cancellation

In this final class of applications, the adaptive filter is used to cancel unknown interference contained in a primary signal, as Figure 4 depicts it. The primary signal serves as the desired response of the adaptive filter.

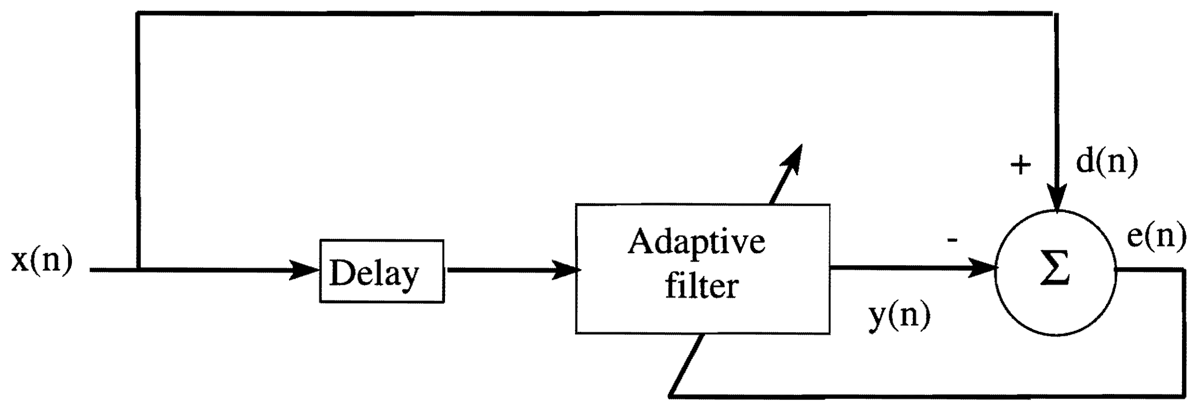

Figure 3. Configuration of an Adaptive Filter as a Predictor.

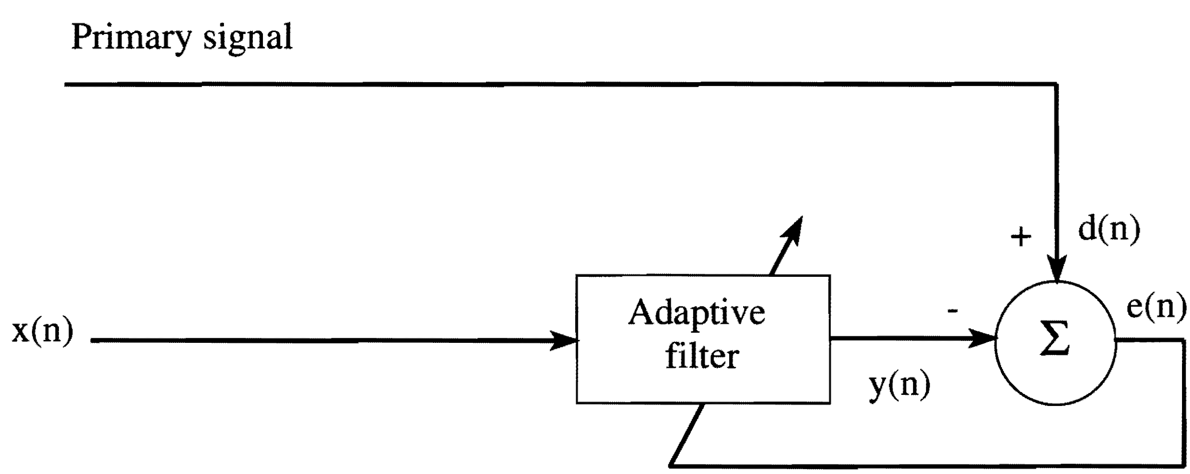

Figure 4. Configuration of an Adaptive Filter as a Noise Canceller.

function, and therefore has only one local minimum which is, of course, the global minimum. Hence, the use of a gradient based adaptation scheme for the convergence to the minimum can be applied.

Many other adaptive filtering algorithms based upon non-mean-square-error cost functions can be also defined to improve the adaptation performance in specific statistical environments [33, 36]. Approaches in [33–39] show that the use of the adaptive filtering algorithm based on a cost function with the error to the power lower than quadratic can be advantageous. They are defined by the following cost function

$$J(n) = E[|e(n)|^p]; \quad 1 < p < 2. \tag{8}$$

The above cost function can be shown to be a convex function [40], that is every minimum of the performance function is a global minimum.

Finally, before stating the possible linear structures used in implementing adaptive filters, it is worth mentioning the properties of the cost functions. All the functions presented in this section and others not mentioned in this work should be positive and monotonically increasing [41] for their corresponding algorithms to perform correctly.

3.2. Structures

A number of different structures for adaptive systems have been proposed. These may be divided into linear and non-linear structures. Linear digital filters may be further subdivided into finite and infinite-duration impulse structures. However, the implementation of the adaptive infinite-duration impulse response (IIR) filters is not straightforward as the poles of the filter can wonder onto or outside the unit circle of the z-plane and, in such a situation, instability can occur. The difficulty is that the adaptation algorithm will choose a set of coefficients which may place poles outside the unit circle in the z-plane and so provoke an unstable response. These difficulties, hence, make the IIR structure less attractive than the well established FIR one. Examples of the FIR filter are the linear transversal filter [42–44] and the lattice filter [45–47] depicted in Figure 6 and Figure 7, respectively. The structure of the IIR filter [6, 48–50] is shown in Figure 8.

Various non-linear digital filter structures have been suggested for adaptive filtering applications including a range of artificial neural networks [51–55], which model the filter on a simplified brain-like structure. An example

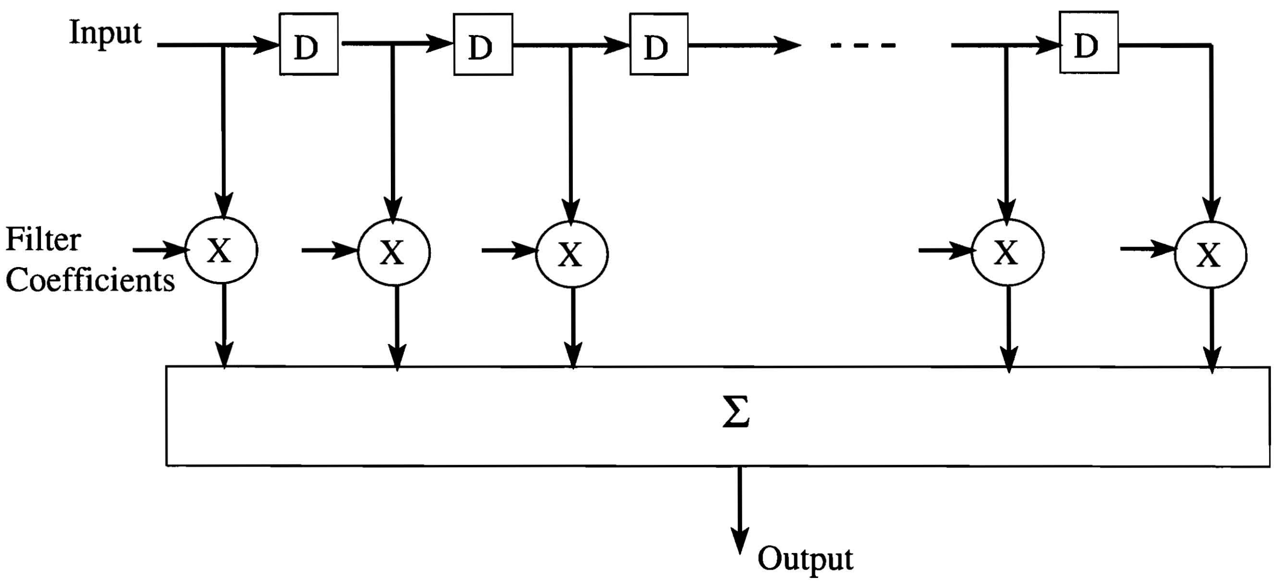

Figure 6. Structure of a Linear Transversal FIR Filter.

of a neural network is shown in Figure 9. While adaptive neural networks are an area of very active research [56–59], the theoretical aspects of non-linear structures are not nearly as well understood as the linear structures.

In all the above mentioned structures the tap spacing is equal to the reciprocal of the symbol rate and the corresponding structure is said to be synchronous. On the other hand, for example in a fractionally spaced equalizer (FSE) [60–63], the equalizer taps are spaced closer than the reciprocal of the symbol rate. Advantages of these structures are numerous [3] and among them the great capability of compensating for delay distortion much more effectively than the synchronous equalizer at the expense of relatively higher computations.

The work in this review is concentrated with the linear transversal filter structure and emphasis is made on using this well understood and often used structure for the study of the subsequent algorithms.

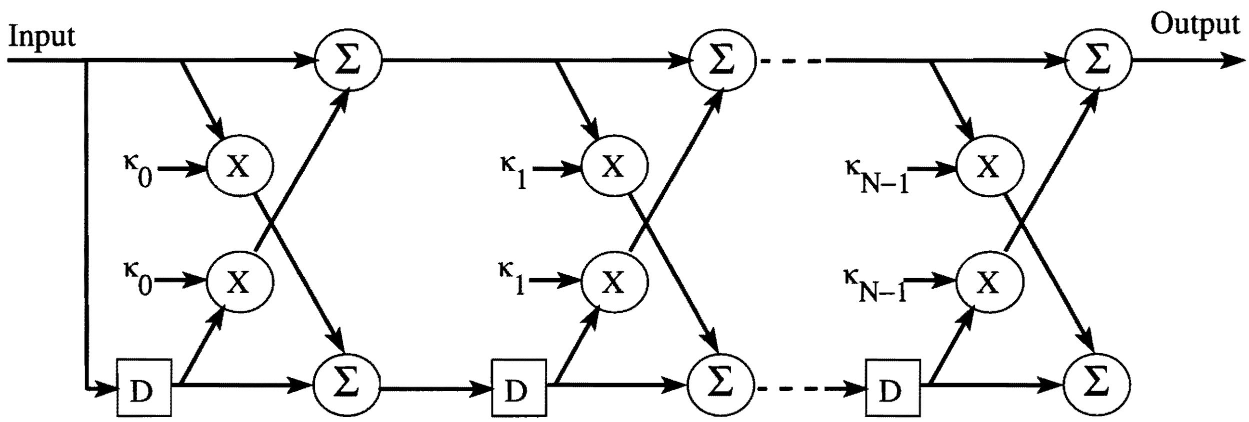

Figure 7. Structure of a FIR Lattice Filter.

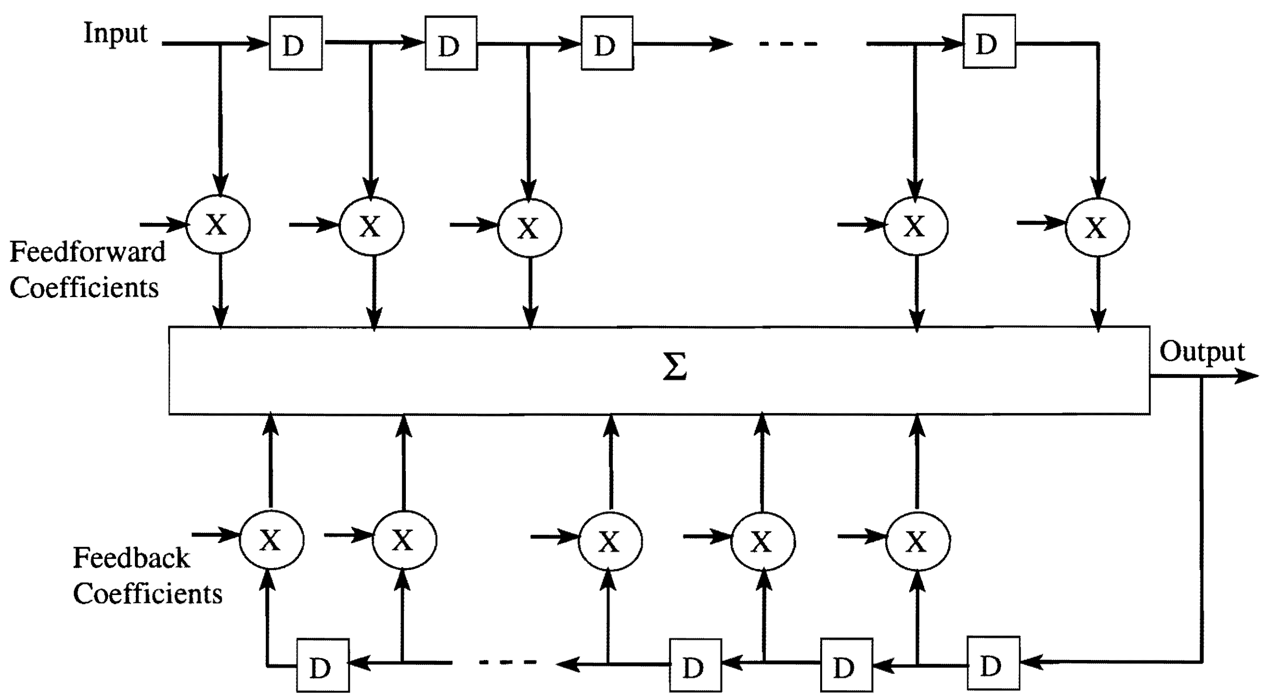

Figure 8. Structure of a Recursive IIR Filter.

3.3. The FIR Adaptive Filter

Assuming that the input sequence $\{x(n)\}$ and the desired sequence $\{d(n)\}$ are wide-sense stationary, the mean-square-error function, Equation (6), can be more conveniently expressed in terms of the input autocorrelation matrix, $\mathbf{R}$, and the crosscorrelation vector, $\mathbf{p}$, between the desired response and the input components, as follows:

$$J(n) = E[d^2(n)] - 2\mathbf{c}^T(n)\mathbf{p} + \mathbf{c}^T(n)\mathbf{R}\mathbf{c}(n), \tag{9}$$

where

$$\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^T(n)], \tag{10}$$

and

$$\mathbf{p} = E[\mathbf{x}(n)d(n)]. \tag{11}$$

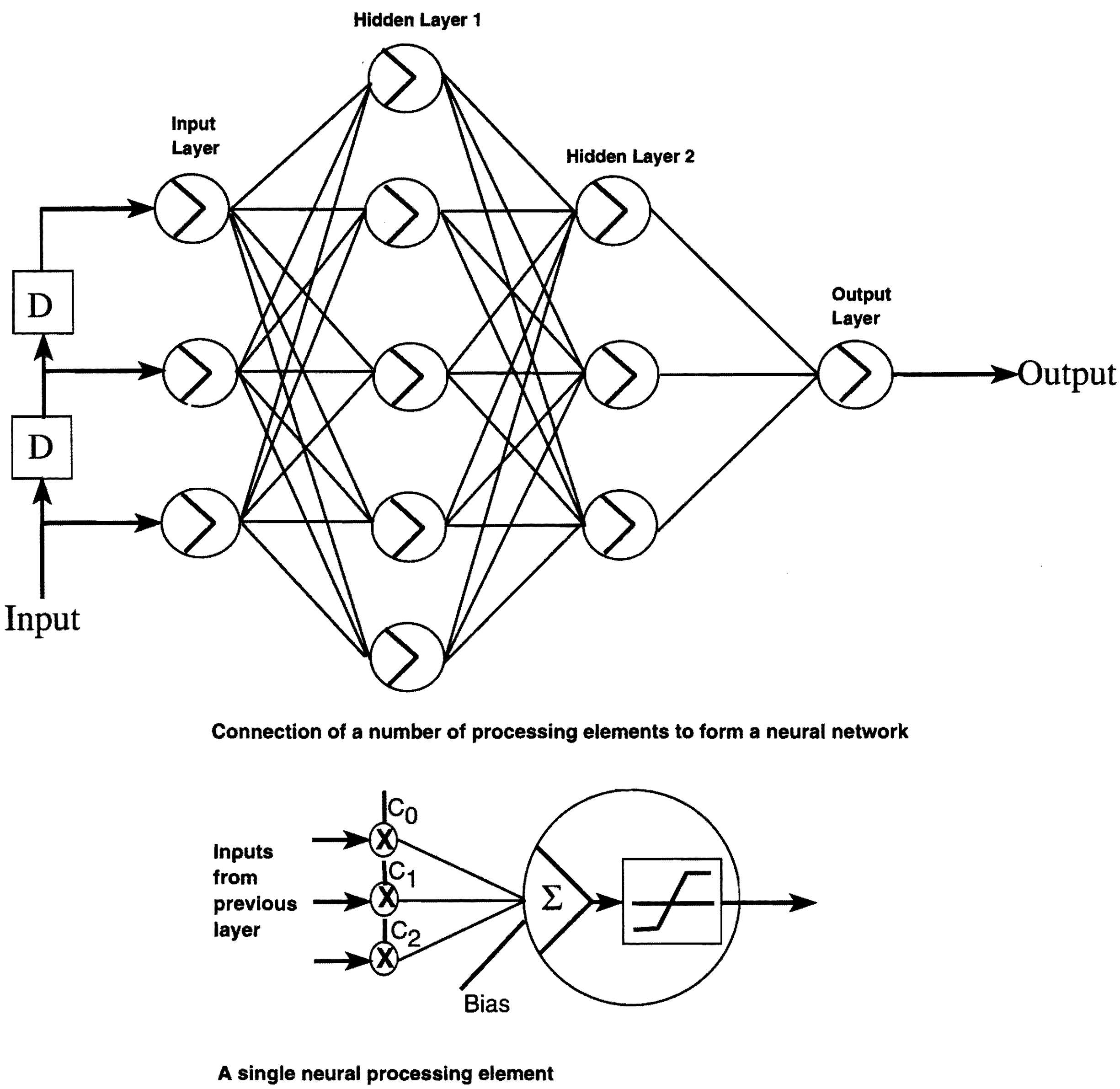

Figure 9. Structure of the Multi-Layer Perceptron which is One Class of Neural Network.

It is clear from expression (9) that the MSE is precisely a quadratic function of the components of the tap coefficients. Thus, the shape associated with this MSE is hyperboloid.

In general, for the linear transversal structure, the surface will be quadratic, when the MSE is used, with a single global minimum. The goal of an adaptive filtering algorithm is to set the filter coefficients so as to obtain an operating point at this minimum, where the filter gives optimum performance.

The point at the bottom of the performance surface corresponds to the optimal tap coefficients, $\mathbf{c}_{opt}$, or minimum MSE. The gradient method is used to cause the tap coefficients vector to seek the minimum of the performance surface. It is defined as

$$\begin{aligned}\nabla E[e^2(n)] &= \frac{\partial J(n)}{\partial \mathbf{c}(n)} \\ &= \left[\frac{\partial J(n)}{\partial c_0(n)}\ \frac{\partial J(n)}{\partial c_1(n)}\ \cdots\ \frac{\partial J(n)}{\partial c_{N-1}(n)}\right]^T \\ &= -2E[e(n)\mathbf{x}(n)] \\ &= 2\mathbf{R}\mathbf{c}(n) - 2\mathbf{p}. \end{aligned} \tag{12}$$

To obtain the minimum MSE, the tap-coefficients vector $\mathbf{c}(n)$ is set to its optimal value, $\mathbf{c}_{opt}$, where the gradient is zero, that is,

$$\nabla E[e^2(n)] = \mathbf{R}\mathbf{c}_{opt} - \mathbf{p} = \mathbf{0}. \tag{13}$$

Under this condition, the optimum value is given by:

$$\mathbf{c}_{opt} = \mathbf{R}^{-1}\mathbf{p}, \tag{14}$$

where this is obtained under the assumption that the autocorrelation matrix $\mathbf{R}$ of the input signal is positive definite and hence nonsingular. Properties of the autocorrelation matrix $\mathbf{R}$ of the input signal can be found in [25]. The minimum MSE, J_{min}, is hence obtained by substitution of (14) in (9), that is,

$$J_{min} = E[d^2(n)] - \mathbf{c}_{opt}^T\mathbf{p}. \tag{15}$$

The solution for $\mathbf{c}_{opt}$ involves inverting the input autocorrelation matrix $\mathbf{R}$, hence, requiring precise knowledge of the second order statistics of the data, *i.e.*, the autocorrelation matrix and the crosscorrelation vector. Unfortunately, it is the data sequences rather than their second order statistics that are available in practice. Alternatively, an iterative procedure may be used to determine $\mathbf{c}_{opt}$. This is the function of an adaptive FIR filter algorithm which has to find the optimum filter from available data rather than from the second statistics of the data [64]. Thus, an adaptive FIR filter can be defined as an algorithm which operates on the sequences $\{x(n)\}$ and $\{d(n)\}$ to form a time-varying impulse response $\mathbf{c}(n)$ which converges in the mean [65] to $\mathbf{c}_{opt}$ as the number of iterations becomes very large, that is:

$$\lim_{n\to\infty} E[\mathbf{c}(n)] = \mathbf{c}_{opt}. \tag{16}$$

4. ADAPTIVE FILTERING ALGORITHMS

In the previous section it was shown that the optimum tap coefficient vector for the adaptive FIR filter could be defined by the statistical properties of the input and desired signals. This implies that if these properties were known then the optimum tap coefficients could be obtained directly. However, it is unlikely to have an accurate measurement, they may be varying with time and the matrix inversion would require considerable amount of computations, specifically if there were a large number of coefficients. Practical adaptation algorithms usually involve iterative techniques. The following gives the two most widely used adaptive filtering algorithms suitable for practical real time applications. These are the least mean-squares (LMS) algorithm and the recursive least-squares (RLS) algorithm. The least mean fourth (LMF) algorithm is also highlighted.

4.1. The LMS Algorithm

Probably the simplest iterative procedure is the method of steepest descent defined according to the following relation [25]

$$\mathbf{c}(n+1) = \mathbf{c}(n) - \frac{1}{2}\mu\nabla E[e^2(n)], \tag{17}$$

where μ is a positive number chosen small enough to insure convergence of the iterative procedure.

Given that the gradient vector, $\nabla E[e^2(n)]$, depends on both the input autocorrelation matrix, $\mathbf{R}$, and the vector $\mathbf{p}$ of cross correlations, this makes the steepest descent difficult for determining the optimum tap coefficients. Instead, estimates of the gradient vector may be used. That is, the LMS algorithm for recursively adjusting the tap coefficients of the adaptive filter is expressed in the form

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \mu e(n)\mathbf{x}(n). \tag{18}$$

The convergence behavior of the LMS algorithm given in Equation (18) is governed by the step size parameter μ. For a larger value of μ, the convergence becomes faster, but it results in a larger residual error and is more prone to instability. Consequently, the tap coefficients will converge to their optimum values if μ satisfies the inequality [30], [66]

$$0 < \mu < \frac{2}{\lambda_{max}}, \tag{19}$$

where λ_{max} is the largest eigenvalue of $\mathbf{R}$. The convergence condition, (19), can be derived in the following manner. Subtracting $\mathbf{c}_{opt}$ from both sides of (18) and then taking the expected value of the result, gives

$$E[\hat{\mathbf{c}}(n+1)] = [\mathbf{I} - \mu\mathbf{R}]E[\hat{\mathbf{c}}(n)], \tag{20}$$

where this is obtained under the assumption that the vectors $\mathbf{x}(n)$ and the coefficient error vector $\hat{\mathbf{c}}(n)$, defined as

$$\hat{\mathbf{c}}(n) = \mathbf{c}(n) - \mathbf{c}_{opt}, \tag{21}$$

are independent [25, 67].

Equation (20) reveals that the algorithm will converge to the optimal value if all the eigenvalues of the matrix $(\mathbf{I} - \mu\mathbf{R})$ are less than unity, that is

$$|1 - \mu\lambda_i| < 1, \quad i = 0, 1, \cdots, N-1 \tag{22}$$

where it is assumed that the autocorrelation matrix, $\mathbf{R}$, is positive definite with eigenvalues, λ_i, hence, it can be factorized as

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \tag{23}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues

$$\mathbf{\Lambda} = diag[\lambda_0 \ \lambda_1 \ \cdots \ \lambda_{N-1}] \tag{24}$$

and $\mathbf{Q}$ is the orthonormal matrix whose ith column is the eigenvector of $\mathbf{R}$ associated with the ith eigenvalue. The convergence of the algorithm is then obtained, which is basically (19), and the time constant, τ_i, associated with the eigenvalue, λ_i, can be derived to give the approximated value:

$$\tau_i = \frac{1}{\mu\lambda_i}, \ i = 0, 1, \cdots, N-1. \tag{25}$$

Hence the longest time constant, τ_{max}, is associated with the smallest eigenvalue, λ_{min}, of the autocorrelation matrix $\mathbf{R}$, that is

$$\tau_{max} = \frac{1}{\mu\lambda_{min}}. \tag{26}$$

Equations (19) and (26) can be combined to give the following result in terms of the eigenvalue spread (condition number), $\frac{\lambda_{max}}{\lambda_{min}}$,

$$\tau_{max} > \frac{\lambda_{max}}{2\lambda_{min}}. \tag{27}$$

Hence, from the point of view of convergence speed, the ideal value of the condition number is unity; the larger the value, the slower will be the convergence of the LMS algorithm. It can be shown [25] that the eigenvalues of the autocorrelation matrix are bounded by the maximum and minimum values of the power spectral density of the input. Furthermore, as the order of the matrix N approaches infinity,

$$\lambda_{min} \rightarrow min[S(\omega)] \tag{28}$$

and

$$\lambda_{max} \rightarrow max[S(\omega)], \tag{29}$$

where $S(\omega)$ is the power spectral density of the input.

It is therefore concluded that the optimum signal for fastest convergence of the LMS algorithm is white noise, and that any form of coloring in the signal will increase the convergence time.

A useful measure for the cost of adaptability is provided by the misadjustment factor M, defined as the ratio of the excess mean-squared error, $J_{ex} = J(\infty) - J_{min}$, to the minimum mean-squared error J_{min}, that is $M = \frac{J_{ex}}{J_{min}}$. In view of the fact that, in steady-state, the weight error vectors are uncorrelated [30, 68], the misadjustment factor for the LMS algorithm can be expressed by [4]:

$$M_{LMS} = \mu \sum_{i=0}^{N-1} \lambda_i. \tag{30}$$

The relationship between the step size and the misadjustment is clearly observed in the above expression. Since speed of convergence and misadjustment lead to conflicting requirements on the step size a compromise must then be reached. In general, to ensure convergence of the iterative procedure and produce less misadjustment error a small step size is chosen. Finally, substituting (25) in (30) yields another form of the misadjustment factor:

$$M_{LMS} = \sum_{i=0}^{N-1} \frac{1}{\tau_i}. \tag{31}$$

Other developed adaptive schemes as well, all of which are LMS variants, *e.g.*, the sign LMS [6], the normalized LMS (NLMS) [69], the leaky LMS [70], the variable step (VS) size algorithm [71] and the fuzzy step-size (FSS) algorithm [72], the block LMS (BLMS) [73], and many others, have been studied to enhance more the performance of the LMS algorithm for the desired application. Table 1 lists the LMS algorithm and some of its derivatives.

Recently, a series of LMS–Newton adaptive filtering [74–77] with variable step size were developed and shown to have attractive convergence speed in nonstationary environments. Application of these algorithms to adaptive filtering in subbands is demonstrated.

The LMS algorithm can be regarded as being obtained from the general expression, (7), when the value $k = 1$. However, the least mean-fourth (LMF) algorithm [32], which is another modification to the general expression, is obtained when $k = 2$. This algorithm is presented next.

The algorithm for adjusting the tap coefficients, $\mathbf{c}(n)$, is given by the following recursion:

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \mu\{\gamma + 2(1-\gamma)e^2(n)\}e(n)\mathbf{x}(n), \tag{40}$$

where the step size μ can be shown to be [81]:

$$0 < \mu < \frac{2}{N[\gamma + 6(1-\gamma)E[w^2(n)]]\sigma_x^2}, \tag{41}$$

N is the length of the adaptive filter, $E[w^2(n)]$ is the measurement noise power, and σ_x^2 is the power of the input signal.

Ultimately, the misadjustment factor M for the MN algorithm can in this case be shown to be expressed by [79]:

$$M_{MN} = \frac{\mu N\{\gamma^2 E[w^2(n)] + 2\gamma(1-\gamma)E[w^4(n)] + (1-\gamma)^2 E[w^6(n)]\}}{2E[w^2(n)][\gamma + 3(1-\gamma)E[w^2(n)]]} E[x^2(n)]. \tag{42}$$

4.4. The RLS Algorithm

The LMS algorithm is widely used due to its comparatively easy implementation, lower order of complexity (only N operations (additions and multiplications) are required per update), and its well-established characteristics. However, the convergence is slow for highly correlated signals. The RLS algorithm, as it is discussed next, however, does not exhibit this dependence behavior.

The RLS algorithm determines the coefficients that minimize the squared error summed over time [6, 86], *i.e.*,

$$J(n) = \sum_{j=0}^{n} e^2(j). \tag{43}$$

Due to the fact that the values of the filter coefficients, that minimize the above cost function, are functions of all past inputs, the associated adapted algorithm will have an infinite memory. A more convenient way, to limit this infinite memory problem, is to introduce a weighting function, $\gamma(n)$, in the cost function so that recent data are given more weight than past data. The resulting new cost function, that will replace that of Equation (43), is defined as:

$$J(n) = \sum_{j=0}^{n} e^2(j)\gamma(n-j), \tag{44}$$

with weighting function $\gamma(n)$ taken as an example, as follows:

$$\gamma(n) = (1-\beta)^n, \quad 0 < \beta < 1. \tag{45}$$

The tap coefficients are adapted to minimize $J(n)$. Taking the derivative of $J(n)$ with respect to $c_i(n)$, $\{i = 0, 1, \cdots, N-1\}$, and setting it equal to zero, *i.e.*,

$$\frac{\partial J(n)}{\partial \mathbf{c}(n)} = \mathbf{0}, \tag{46}$$

the following vector of the adaptive filter is obtained [23]

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \beta \mathbf{R}^{-1}(n)\mathbf{x}(n)e(n), \tag{47}$$

where $\mathbf{R}(n)$, the autocorrelation matrix of the input signal vector $\mathbf{x}(n)$, is found recursively

$$\mathbf{R}(n) = (1-\beta)\mathbf{R}(n-1) + \beta\mathbf{x}(n)\mathbf{x}^T(n). \tag{48}$$

Table 2. The Computational Complexity of the Different Algorithms.

Algorithm	Multiplications	Additions/Subtractions	Divisions
RLS	$2.5N^2 + 4.5N$	$1.5N^2 + 4.5N$	1
Fast Kalman	$10N + 3$	$9N + 6$	2
FAEST	$7N + 10$	$7N + 8$	4
FTF	$7N + 14$	$7N + 7$	3

Comparing (47) with (18) we see that the simple scalar loop gain in LMS algorithm has been replaced with β times the inverse of $\mathbf{R}(n)$. The normalization with $\mathbf{R}^{-1}(n)$ offers more than the simple power normalization in LMS; it normalizes the adaptation in each eigenvector direction by the signal power in that direction. Thus the convergence becomes independent of both the signal type and power [87].

The solution to the above equations, (47) and (48), would require a large number of computations per update. This algorithm requires approximately $2.5N^2 + 4.5N$ multiplications and additions per update [6], N being the number of tap coefficients, significantly greater than the order of N for the LMS algorithm. Thus, as N increases, the number of operations increases in proportion to the square of the filter order. Hence, obtaining the optimum coefficient value involves computation of the inverse of the autocorrelation matrix and results in complex implementation. However, the advantage of this algorithm is fast convergence irrespective of the correlation characteristics of the input signal.

The high complexity of the RLS algorithm may be reduced by exploiting the shifting properties of the input sequence with time. This has resulted in several fast RLS algorithms such as the fast Kalman [26, 88, 89], the fast *a posteriori* error sequential technique (FAEST) [27, 90, 91] and the fast transversal filters (FTF) algorithms [28, 92, 93], all of which are characterized by a computational complexity which is directly proportional to the filter length N. Table 2 lists the computational complexity of the different algorithms.

In contrast to the good feature of fast convergence observed with the RLS based algorithms, their computational complexities are still not attractive as those of the LMS algorithm. Moreover, instability problems are still drawbacks for these algorithms which suffer from severe numerical instability [94] when implemented using either fixed or floating point digital arithmetic [95, 96]. They are highly sensitive to small numerical errors at each iteration and will often diverge suddenly from the correct least squares solution.

Also, recently, other recursive algorithms based on the minimization of the least-fourth cost function gave rise to the recursive least fourth (RLF) algorithm [97].

Ultimately, Figure 11 shows a tree structure of the families of adaptive filtering algorithms which have been suggested for all the cost functions presented in this work.

5. PERFORMANCE OF THE ALGORITHMS

In this section we evaluate and compare the performance of the RLS, the LMS, the LMF, and the MN algorithms. The basic digital transmission system considered for the adaptive system identification is shown in

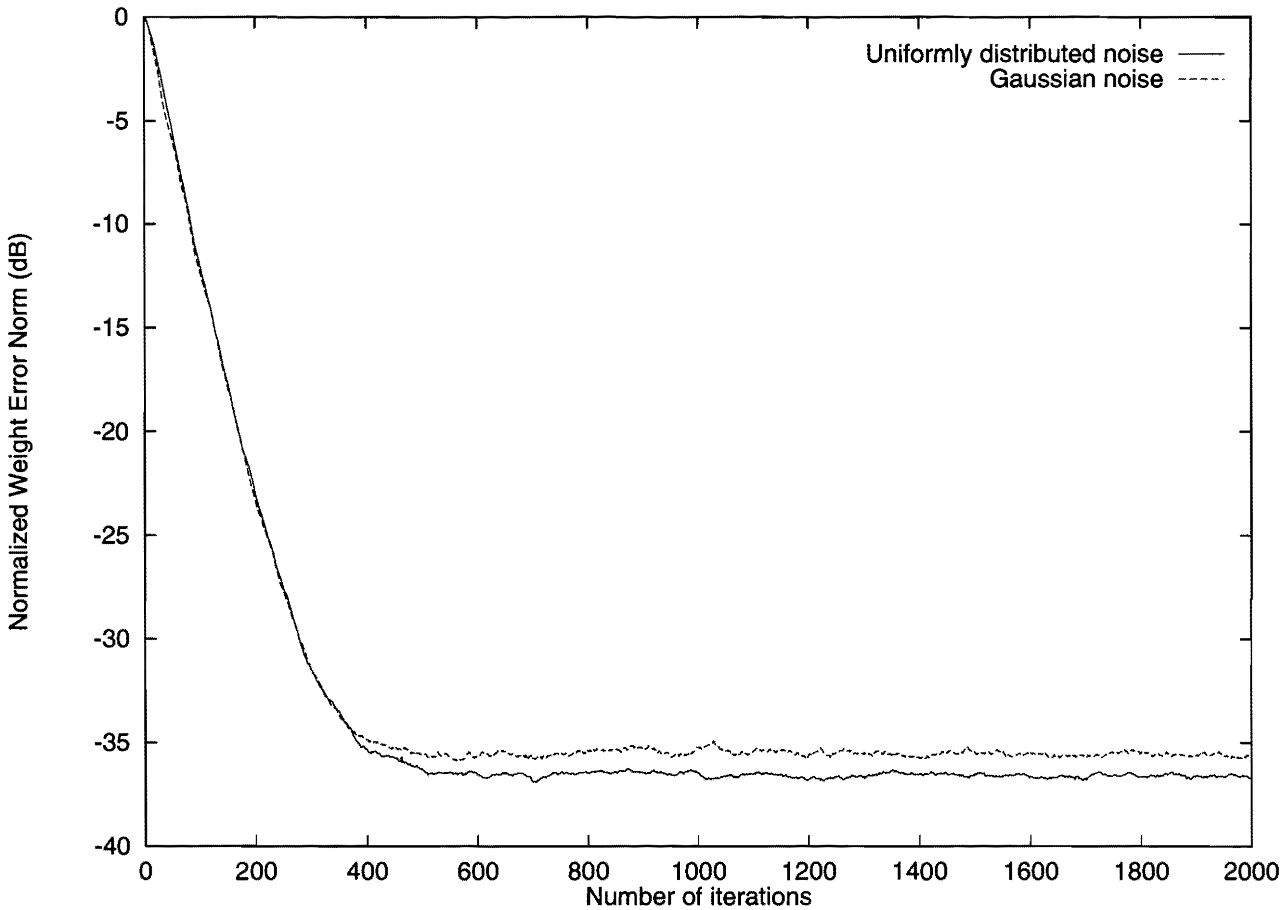

Figure 13. Effect of Noise Distribution on the Convergence Behavior of the MN Algorithm.

Recently, Sayed and Kailath [100] showed the exact relationship between the RLS algorithm and the Kalman filter algorithm. This relationship may be taken into advantage in the near future to exploit the well developed Kalman filter theory in order to improve adaptive filtering algorithms in nonstationary environments. Also, it is shown in [101] that the LMS algorithm is optimal under the H^{∞} criterion. This can take a new research direction where the robustness properties of the LMS algorithm are derived and exploited.

Several new developments in signal processing that emerged recently, including cyclostationarity, chaotic signals, wavelet representation, and fractional lower-order-moments, can be effectively used for the development of new adaptive filtering algorithms.

7. SUMMARY

In this work, we presented various common structures of FIR adaptive filters with their respective adaptive filtering algorithms. The issue of adaptive filtering is still and will remain a very active field of research for some considerable time. This is mainly due to the advances in the computing facilities that were not previously available and to the need for such algorithms.

The wide spread use of the least-demanding computing algorithm, *i.e.*, the LMS algorithm, is with no doubt due to its both simplicity and relative performance. The RLS algorithm, for example, gives very fast convergence to the algorithm at the expense of very heavy computational loads, irrespective of the input signal statistics. Moreover, the performance of the LMS algorithm is better or as good as that of the RLS algorithm in nonstationary environments [32].

Finally, the difference in performance for all algorithms in general is due to the fact that they operate under different minimization functions. Several of these cost functions have been mentioned in this review. A novelty of this paper is the unified treatment of the recently emerging non-square and mixed norm algorithms, together with the more traditional squared norm based algorithms. Lately other algorithms based on higher order statistics [102] are also emerging.

ACKNOWLEDGMENT

The authors acknowledge KFUPM for the support. They acknowledge the reviewers for their useful comments and suggestions as well.

REFERENCES

[1] K. Murano, "Adaptive Signal Processing Applied in Telecommunications", *IFAC Adaptive Systems in Control and Signal Processing*, 1992, pp. 431–441.

[2] M.M. Sondhi, "An Adaptive Echo Canceller", *Bell Syst. Tech. J.,* **46(3)** (1967), pp. 497–511,

[3] S.U.H. Qureshi, "Adaptive Equalization", *Proc. IEEE,* **73(9)** (1985), pp. 1349–1387.

[4] 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. Goodlin, "Adaptive Noise Cancelling: Principles and Applications", *Proc. IEEE,* **63(12)** (1975), pp. 1692–1716.

[5] C.F.N. Cowan, J. Mavor, and J.W. Arthur, "Noise Cancellation and Inverse Filtering using a Compact High-Performance C.C.D Adaptive Filter", *Electr. Lett.,* **15(1)** (1979), pp. 35–37.

[6] C.F.N. Cowan and P.M. Grant, *Adaptive Filters.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[7] M.M. Sondhi and D.A. Berkley, "Silencing Echoes on the Telephone Network", *Proc. IEEE,* **68(8)** (1980), pp. 948–963.

[8] S.U.H. Qureshi, "Adaptive Equalization", *IEEE Commun. Magazine,* **21(2)** (1982), pp. 9–16.

[9] C.F.N. Cowan, *Channel Equalization.* ed. N. Kalouptsidis and S. Theodoridis. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[10] R.W. Lucky, "Automatic Equalization for Digital Communication", *Bell Syst. Tech. J.,* **44(3)** (1965), pp. 547–588.

[11] A. Gersho, "Adaptive Equalization of Highly Dispersive Channels for Data Transmission", *Bell Syst. Tech. J.,* **48(1)** (1969), pp. 55–70.

[12] C.W. Nissen, "Automatic Channel Equalization Algorithms", *Proc. IEEE,* **55(5)** (1967), pp. 698.

[13] A.H. Gray and J.D. Markel, *Linear Prediction of Speech.* Berlin: Springer Verlag, 1976.

[14] L. Makhoul, "Linear Prediction a Tutorial Review", *Proc. IEEE,* **63(4)** (1975), pp. 561–580.

[15] J.G. Proakis, *Digital Communications.* Singapore: McGraw-Hill, 1983.

[16] T. Kailath, "A View of Three Decades of Linear Filtering Theory", *IEEE Trans. Inf. Theory,* **IT-20** (1974), pp. 146–181.

[17] B. Widrow, P.E. Mantey, L.J. Griffiths, and B.B. Goode, "Adaptive Antenna Systems Array", *Proc. IEEE,* **55** (1967), pp. 2143–2159.

[18] G. Leitmann, *Mathematics in Science and Engineering, vol. 5: Optimization Techniques.* New York: Academic Press, 1962.

[19] G. Leitmann, *Mathematics in Science and Engineering, vol. 31: Topics in Optimization*, New York: Academic Press, 1967.

[20] D. Luenberger, *Optimization by Vector Space Methods.* New York: John Wiley & Sons, 1969.

[21] N. Levinson, "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction", *J. Math. Phys.,* **25** (1947), pp. 261–278.

[22] H.W. Sorenson, "Least-Squares Estimation from Gauss to Kalman", *IEEE Spectrum,* **7** (1970), pp. 63–68.

[23] J.M. Mendel, *Discrete Techniques of Parameter Estimation.* New York: Marcel Dekker, 1973.

[24] T. Kailath, *Lectures on Linear Least-Squares Estimation.* New York: Springer-Verlag, 1981.

[25] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1991.

[26] D.D. Falconer and L. Ljung, "Application of Fast Kalman Estimation to Adaptive Equalization", *IEEE Trans. Commun.,* **COM-26** (1978), pp. 1439–1446.

[27] G. Carayannis, D.G. Manolakis, and N. Kalouptsidis, "A Fast Sequential Algorithm for Least-Squares Filtering and Prediction", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-31** (1983), pp. 1394–1402.

[28] J.M. Cioffi and T. Kailath, "Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-32** (1984), pp. 304–337.

[29] N. Wiener, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* New York: John Wiley & Sons, 1962.

[30] B. Widrow, J.M. McCool, M.G. Larimore, and C.R. Johnson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter", *Proc. IEEE,* **64(8)** (1976), pp. 1151–1162.

[31] C.F.N. Cowan, "Performance Comparison of Finite Linear Adaptive Filters", *IEE Proceedings Part F,* **134(3)** (1987), pp. 211–216.

[32] E. Walach and B. Widrow, "The Least Mean Fourth (LMF) Adaptive Algorithm and its Family", *IEEE Trans. Inf. Theory,* **IT-30** (1984), pp. 275–283.

[33] J.M. Chen, B.S. Chen, and W.S. Chang, "Parameter Estimation of Linear Systems with Input–Output Noise Data: A Generalized l_p Norm Approach", *Signal Processing,* **37** (1994), pp. 345–356.

[34] J. Schroeder, R. Yarlagadda, and J. Hershey, "L_p Normed Minimization with Applications to Linear Predictive Modeling for Sinusoidal Frequency Estimation", *Signal Processing,* **24** (1991), pp. 193–216.

[35] R. Yarlagadda, J. Bee Bednar, and T.L. Watt, "Fast Algorithms for l_p Deconvolution", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-33(1)** (1985), pp. 174–182.

[36] S. Siu and C.F.N. Cowan, "Performance Analysis of the l_p Norm Back Propagation Algorithm for Adaptive Equalisation", *IEE Proceedings-F,* **140** (1993), pp. 43–47.

[37] S.C. Pei and C.C. Tseng, "Least Mean P-Power Error Criterion for Adaptive FIR Filter", *IEEE Journal on Selected Areas in Comm.,* **12(9)** (1994), pp. 1540–1547.

[38] S.A.H. Shah and C.F.N. Cowan "Modified Stochastic Gradient Algorithm Using Nonquadratic Cost Functions for Data Echo Cancellation", *IEE Proc.-Vis. Image Signal Process.,* **142(3)** (1995), pp. 187–191.

[39] M. Shan and C.L. Nikias, "Signal Processing with Fractional Lower Order Moments: Stable Processes and Their Applications", *Proc. IEEE,* **81(7)** (1993), pp. 986–1010.

[40] D.G. Luenberger, *Linear and Nonlinear Programming, 2nd edn.* Reading, MA: Addison Wesley, 1984.

[41] B. Mulgrew and C.F.N. Cowan, *Adaptive Filters and Equalizers.* Norwell, MA: Kluwer Academic Publishers, 1988.

[42] A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1975.

[43] R.A. Roberts and C.T. Mullis, *Digital Signal Processing.* Reading, MA: Addison Wesley, 1987.

[44] J.G. Proakis and D.G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications, 2nd edn.* New York: Macmillan, 1992.

[45] E.H. Satorius and S.T. Alexander, "Channel Equalization Using Adaptive Lattice Algorithms", *IEEE Trans. Commun.,* **COM-27(9)** (1979), pp. 899–905.

[46] B. Friedlander, "Lattice Filters for Adaptive Processing", *Proc. IEEE,* **70** (1982), pp. 829–867.

[47] M.L. Honig and D.G. Messerschmitt, *Adaptive Filter: Structures, Algorithms, and Applications.* Hingham, MA: Kluwer Academic Publishers, 1984.

[48] M.G. Larimore, J.R. Treichler, and C.R. Johnson, "SHARF; An Algorithm for Adapting IIR Digital Filters", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-28** (1976), pp. 1622–1624.

[49] J.J. Shynk, "Adaptive IIR Filtering", *IEEE ASSP Magazine,* **28(1)** (1989), pp. 4–21.

[50] C.R. Johnson, Jr., "Adaptive IIR Filtering: Current Results and Open Issues", *IEEE Trans. Inform. Theory,* **IT-30(2)** (1984), pp. 237–250.

[51] D. Ruck, S.K. Rogers, M. Kabrinsky, M.E. Oxley, and B.W. Suter, "The Multilayer Perceptron as an Approximation to the Bayes Optimal Discriminant Function ", *IEEE Trans. Neural Net.,* **1** (1990), pp. 296–298.

[52] S.I. Gallant, "Perceptron Based Learning Algorithms", *IEEE Trans. Neural Net.,* **1** (1990), pp. 171–191.

[53] J. Shynk, "Performance Surface of a single Layer Perceptron", *IEEE Trans. Neural Net.,* **1** (1990), pp. 268–274.

[54] J.W. Watterson, "An Optimal Multilayer Perceptron Neural Receiver for Signal Detection", *IEEE Trans. Neural Net.,* **1** (1990), pp. 298–300.

[55] A. Barto, R. Sutton, and C. Anderson, "Neuron-Like Adaptive Elements that can Solve Difficult Learning Control Problems", *IEEE Trans. Syst. Man & Cybernet.,* **13** (1983), pp. 834–846.

[56] C.H. Chang, S. Siu, and C.H. Wei, "A Polynomial-Perceptron Based Decision Feedback Equalizer with a Robust Learning Algorithm", *Signal Processing,* **47** (1995), pp. 145–158.

[57] B. Widrow and M. Lehr, "30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation", *Proc. IEEE,* **78(9)** (1990), pp. 1415–1441.

[58] R.P. Lippmann, "An Introduction to Computing with Neural Nets", *IEEE Acoustics, Speech and Signal Processing Magazine,* **4(2)** (1987), pp. 4–22.

[59] D.R. Hush and B.G. Horne, "Progress in Supervised Neural Networks", *IEEE Acoustics, Speech and Signal Processing Magazine,* **10(1)** (1993), pp. 8–39.

[60] G. Ungerboeck, "Fractional Tap Spacing Equalizer and Consequences for Clock Recovery in Data Modems", *IEEE Trans. Commun.*, 1976, pp. 856–864.

[61] R.D. Gitlin and S.B. Weinstein, "Fractionally-Spaced Equalization an Improved Digital Transversal Equalizer", *Bell Syst. Tech. J.,* **60(2)** (1981), pp. 275–296.

[62] M.S. El-Hennaway, A. Zerguine, and E.E. Hassan, "Non-Equally Fractionally-Spaced Equalizers", *Electr. Lett.,* **26(16)** (1990), pp. 1254–1256.

[63] M. Bettayeb and A. Zerguine, "Fractionally-Spaced Equalization Using the Lattice Structure", *Computers Elect. Engng.,* **21(1)** (1995), pp. 47–55.

[64] B. Widrow, "Adaptive Filters", in *Aspects of Network and System Theory.* ed. R.E. Kalman and N. Declaris. New York: Holt, Rinehart & Winston, 1971.

[65] B. Widrow and S.D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[66] G. Ungerboeck, "Theory of the Speed of Convergence in Adaptive Equalizers for Digital Communications", *IBM J. Res. Dev.,* **16** (1972) pp. 546–552.

[67] J.E. Mazo, "On the Independence Theory of Equalizer Convergence", *Bell Syst. Tech. Journ.,* **58** (1979), pp. 963–993.

[68] W.A. Gardner, "Learning Characteristics of Stochastic-Gradient-Descent Algorithms: A General Study, Analysis and Critique", *Signal Processing,* **6** (1984), pp. 113–133.

[69] I.J. Nagumo and A. Noda, "A Learning Method for System Identification", *IEEE Trans. Automatic Control,* **AC-28** (1967), pp. 282–287.

[70] R.D. Gitlin, H.C. Meadors, and S.B. Weinstein, "The Tap-Leakage Algorithm: An Algorithm for the Stable Operation of a Digitally Implemented, Fractionally Spaced Adaptive Equalizer", *Bell Syst. Tech. J.,* **61** (1982), pp. 1817–1839.

[71] R.W. Harris, D.M. Chabries, and F.A. Bishop "A Variable Step (VS) Adaptive Filter Algorithm", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-34** (1986), pp. 309–316.

[72] W.S. Gan "Fuzzy Step-Size Adjustment for the LMS Algorithm", *Signal Processing,* **49** (1996), pp. 145–149.

[73] G.A. Clark, S.K. Mitra, and S.R. Parker, "Block Implementations of Adaptive Digital Filters", *IEEE Trans. Circuits and Systems,* **CAS-28** (1981), pp. 584–592.

[74] D.F. Marshall and W.K. Jenkins, "A Fast Quasi-Newton Adaptive Filtering Algorithm", *IEEE Trans. Signal Processing,* **40** (1992), pp. 1652–1662.

[75] P.S.R. Diniz and L.M.P. Biscainho, "Optimal Variable Step Size for the LMS–Newton Algorithm with Application to Subband Adaptive Filtring", *IEEE Trans. Signal Processing,* **40** (1992), pp. 2825–2829.

[76] M.L.R. de Campos, P.S.R. Diniz, and A. Antoniou, "Performance of LMS–Newton Adaptive Filtering Algorithms with Variable Convergence Factor in Nonstationary Environments", in *Proc. ICASSP, Mineapolis,* 1993, pp. 392–395.

[77] P.S.R. Diniz, M.L.R. de Campos, and A. Antoniou, "Analysis of LMS–Newton Adaptive Filtering Algorithms with Variable Convergence Factor", *IEEE Trans. Signal Processing,* **43** (1995), pp. 617–627.

[78] J.A. Chambers, O. Tanrikulu, and A.G. Constantinides, "Least Mean Mixed-Norm Adaptive Filtering", *Electr. Lett.,* **30(19)** (1994), pp. 1574–1575.

[79] O. Tanrikulu and J.A. Chambers, "Convergence and Steady-State Properties of the Least Mean Mixed-Norm (LMMN) Adaptive Filtering", *IEE Proc.-Vis. Image Signal Process.,* **143(3)** (1996), pp. 137–142.

[80] D.I. Pazaitis and A.G. Constantinides, "LMS+F Algorithm", *Electr. Lett.,* **31(17)** (1995), pp. 1423–1424.

[81] A. Zerguine, C.F.N. Cowan, and M. Bettayeb, "Novel LMS-LMF Adaptive Scheme for Echo Cancellation", *Electr. Lett.,* **32(19)** (1996), pp. 1776–1778.

[82] A. Zerguine, C.F.N. Cowan, and M. Bettayeb, "Adaptive Echo Cancellation Using Least Mean Mixed-Norm Algorithm", *IEEE Trans. Signal Processing,* **SP-45(5)** (1997), pp. 1340–1343.

[83] A. Zerguine, C.F.N. Cowan, and M. Bettayeb, "A Compound Near-Far End Least Mean Square-Fourth Error Minimization for Adaptive Echo Cancellation", *Thirtieth Annual Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California,* November 3–6, 1996.

[84] A. Zerguine, M. Bettayeb, and C.F.N. Cowan, "A Hybrid LMS-LMF Scheme for Echo Cancellation", in *Proc. ICASSP, Munich, Germany,* 1997, pp. 2313–2316.

[85] A. Zerguine and M. Bettayeb, "Decision Feedback Equalization Using Least Mean Mixed-Norm Algorithm", To be submitted to *IEE Proc.-Vis. Image Signal Processing.*

[86] M.L. Honig, "Echo Cancellation of Voiceband Data Signals Using Recursive Least Squares and Stochastic Gradient Algorithms", *IEEE Trans. Commun.,* **COM-33(1)** (1985), pp. 65–73.

[87] C.W.K. Gritton and D.W. Lin, "Echo Cancellation Algorithms", *IEEE ASSP Magazine,* April 1984, pp. 30–37.

[88] L. Ljung, M. Morf, and D. Falconer, "Fast Calculation of Gain Matrices for Recursive Estimation Schemes", *Int. J. Control,* **27** (1978), pp. 1–19.

[89] C. Samson, "A Unified Treatment of Fast Kalman Algorithms for Identification", *Int. J. Control,* **31** (1982), pp. 909–934.

[90] N. Kalouptsidis, G. Carayannis, and D.G. Manolakis"A Fast Covariance Algorithm for Sequential Least-Squares Filtering and Prediction", *IEEE Trans. Auto. Control,* **AC-29** (1984), pp. 752–755.

[91] G. Carayannis, N. Kalouptsidis, and D.G. Manolakis "Fast Recursive Algorithms for a Class of Linear Equations", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-30** (1982), pp. 227–239.

[92] J.M. Cioffi and T. Kailath, "Windowed Fast Transversal Filter Adaptive Algorithms with Normalization", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-33** (1985), pp. 607–625.

[93] J.M. Cioffi, "The Block Processing FTF Adaptive Algorithm", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-34** (1986), pp. 77–90.

[94] J.M. Cioffi, "Limited Precision Effects in Adaptive Filtering", *IEEE Trans. Circuits Syst.,* **CAS-34(7)** (1987), pp. 821–833.

[95] C.W. Barnes, B.N. Tran, and S.H. Leung, "On the Statistics of Fixed Point Roundoff Error", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-33** (1985), pp. 595–606.

[96] L.P. Mulcahy, "On Fixed Point Roundoff Error Analysis", *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-37** (1989), pp. 1623.

[97] C.L. Nikias and A.P. Petropulu, *High-Order Spectra Analysis: A Nonlinear Signal Processing Framework.* Englewood Cliffs, NJ: Prentice-Hall, 1993.

[98] O. Macchi, *The Least Mean Squares Approach with Applications in Transmission,* New York: John Wiley, 1994.

[99] Q. Ma and C.F.N. Cowan, "Genetic Algorithms Applied to the Adaptation of IIR Filters", *Signal Processing,* **48** (1996), pp. 155–163.

[100] A.H. Sayed and T. Kailath, "A State Space Approach to Adaptive RLS Filtering", *IEEE Signal Processing Magazine,* **11** (1994), pp. 18–60.

[101] B. Hassibi, A.H. Sayed, and T. Kailath, "H^∞ Optimality of the LMS Algorithm", *IEEE Trans. Signal Processing,* **44** (1996), pp. 267–280.

[102] O. Tanrikulu and A.G. Constantinides, "Least-Mean Kurtosis: A Novel Higher-Order Statistics based Adaptive Filtering Algorithm", *Electr. Lett.,* **30(3)** (1994), pp. 189–190.

Paper Received 13 October 1996; Revised 4 June 1997; Accepted 13 December 1997.