Adaptive Filters

Simon Haykin
McMaster University
Hamilton, Ontario, Canada L8S 4K1

1. Introduction

An *adaptive filter* is defined as a self-designing system that relies for its operation on a recursive algorithm, which makes it possible for the filter to perform satisfactorily in an environment where knowledge of the relevant statistics is not available.

Adaptive filters are classified into two main groups: linear, and non linear. *Linear adaptive filters* compute an estimate of a desired response by using a linear combination of the available set of observables applied to the input of the filter. Otherwise, the adaptive filter is said to be <u>nonlinear</u>. Adaptive filters may also be classified into:

(i) *Supervised adaptive filters*, which require the availability of a training sequence that provides different realizations of a desired response for a specified input signal vector. The desired response is compared against the actual response of the filter due to the input signal vector, and the resulting error signal is used to adjust the free parameters of the filter. The process of parameter adjustments is continued in a step-by-step fashion until a steady-state condition is established.

(ii) *Unsupervised adaptive filters*, which performs adjustments of its free parameters without the need for a desired response. For the filter to perform its function, its design includes a set of rules that enable it to compute an input-output mapping with specific desirable properties. In the signal-processing literature, unsupervised adaptive filtering is often referred to as blind deconvolution or blind adaptation.

Gabor [1] was the first to conceive the idea of a nonlinear adaptive filter in 1954 using a Volterra series. The first algorithm used to design a linear adaptive filter is the ubiquitous least-mean-square (LMS) algorithm developed by Widrow and Hoff [2]. The LMS algorithm is often referred to as the *Widrow-Hoff rule*; it was originally derived by Widrow and Hoff in 1959 in their study of a pattern recognition system known as the adaptive linear element (Adaline). The LMS algorithm is closely related to Rosenblatt's *perceptron* [3] in that they are both built on error-correction learning. They both emerged about the same time in the late 1950s during the formative years of neural networks. The importance of Rosenblatt's perceptron is largely historical today. On the other hand, the LMS algorithm has survived the test of time.

Adaptive filters find applications in highly diverse fields: channel equalization, system identification, predictive deconvolution, spectral analysis, signal detection, noise cancellation, and beamforming.

Several books have been written on linear adaptive filters, the most prominent ones of which are Haykin [4] and Widrow and Stearns [5].These two books also cover their applications.

2. Least-Mean-Square (LMS) Algorithm

The LMS algorithm has established itself as the workhorse of adaptive signal processing for two primary reasons:

- Simplicity of implementation and a computational efficiency that is linear in the number of adjustable parameters.
- Robust performance

Hassibi et al. [6] have shown that a single realization of the LMS algorithm is optimal in the H^∞ (i.e., minimax) sense. This result explains the robust behavior of the LMS algorithm.

Basically, the LMS algorithm is a stochastic gradient algorithm, which means that the gradient of the error performance surface with respect to the free parameter vector changes randomly from one iteration to the next. This stochasticity, combined with the presence of nonlinear feedback, is responsible for making a detailed convergence analysis of the LMS algorithm a difficult mathematical task. Indeed, it has attracted research attention for over 25 years [7-10].

The LMS algorithm has two major drawbacks: slow rate of convergence, and sensitivity to the eigenvalue spread (i.e., the ratio of the largest eigenvalue to the smallest eigenvalue) of the correlation matrix of the input signal vector. One way of overcoming these limitations is to use projections of the input signal on an orthogonal basis. This desirable objective can be attained by means of transform-domain adaptive algorithms, so called because the adaptation is performed in the frequency domain rather than the original time domain [11-16]. For a more refined method, we may use a multi-rate adaptive filter which provides better trade-offs between performance improvement, computational complexity and transmission delay [17].

3. Recursive Least-Squares (RLS) Algorithm

The LMS algorithm attains simplicity of implementation by using instantaneous estimates of (1) the autocorrelation matrix of the input signal vector, and (2) the cross-correlation vector between the input vector and the desired response. In contrast, the *recursive least-squares (RLS) algorithm* utilizes continuously updated estimates of these two quantities, which go back to the beginning of the adaptive process. Accordingly, the RLS algorithm exhibits the following properties:

- Rate of convergence that is typically an order of magnitude faster than the LMS algorithm.
- Rate of convergence that is invariant to the eigenvalue spread of the correlation matrix of the input vector.

These desirable properties are however attained at the cost of increased computational complexity [4].

The standard RLS algorithm is built around an FIR filter. For its derivation we may use the matrix inversion lemma, or exploit the correspondence that exists between the variables characterizing the RLS algorithm and those characterizing the celebrated Kalman filter [18]. This latter approach is highly attractive as it provides the basis for deriving the many variants of the RLS algorithm, which include the following:

- *Square-root adaptive filters,* based on the numerically stable QR-decomposition procedure; this class of adaptive filters includes the QR-RLS [19], extended QR-RLS [20,21], and inverse QR-RLS filters [22], all of which can be implemented using systolic arrays.

- *Order-recursive adaptive filters*, the most important form of which is the QRD-LSL (QR decomposition-based least-squares lattice) filter [23,4]. An important property of this class of adaptive filtering algorithms is that their computational complexity is linear in the number of adjustable parameters as in the LMS algorithm.

4. Tracking of Time-Varying Systems

When an adaptive filter operates in a nonstationary environment the requirement is not only to converge to the minimum point of the error performance surface but also to continually track the statistical variations of the input signal. Tracking is a steady-state phenomenon, whereas convergence is a transient phenomenon. This means that an adaptive filter must pass from the transient mode to the steady-state mode before it can start tracking. Moreover, the rate of convergence and tracking are two distinct properties, which means that an algorithm with good convergence properties does not necessarily possess a fast tracking capability.

In general, the LMS algorithm exhibits a more robust tracking behavior than the standard RLS algorithm [24,25]. This limitation of the RLS algorithm may be overcome by incorporating prior knowledge into the design of the algorithm so as to minimize the mismatch between the state-space model of the algorithm and the mathematical model for the task at hand [26,4].

5. Unsupervised Adaptive Filters

The need for unsupervised adaptive filtering arises in such applications as channel equalization, system identification, and separation of independent signal sources, where there is no access to a desired response. We may identify three broadly defined families of unsupervised adaptive filtering algorithms:

(i) *Higher-order statistics (HOS)-based algorithms*, which may be divided into two sub-groups:

- *Implicit HOS-based algorithms*, which exploit higher-order statistics of the input signal in an implicit sense. They include the Sato algorithm [27] and the Godard algorithm or the constant modulus algorithm (CMA) [28,29]. A detailed overview of the CMA is presented in [30].

- *Explicit HOS-based algorithms*, which explicitly use higher-order statistics or their discrete-time Fourier transforms known as polyspectra [31]. The disadvantage of these algorithms is that they can be computationally demanding.

(ii) *Cyclostationary statistics-based algorithms*, which exploit the second-order cyclostationary statistics of the input signal [32.33]. The property of cyclostationarity is known to arise in a modulated signal that results from varying the amplitude, phase or frequency of a sinusoidal carrier, which is basic to the electrical communication process [34].

(iii) *Information-theoretic algorithms*, whose derivations invoke the notion of likelihood function [35], entropy [36] or Kullbeck-Leibler divergence [37].

Approach (iii) is rooted in the Infomax (maximum mutual information) principle [38]; it provides a powerful approach to unsupervised adaptive filtering [39].

6. Neural Networks

A discussion of adaptive filters would be incomplete without a brief mention of (artificial) *neural networks*, for one important class of which we offer the following definition:

A neural network is a massively parallel distributed processor made up of simple processing units (known as neurons), which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the human brain in two respects:

- *Knowledge is acquired by the network from its environment through a learning process.*
- *Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge.*

The point to take from this definition is that learning is in reality a nonlinear form of statistical signal processing [40]. Most importantly, a neural network permits a set of data to speak for itself. Indeed, so long as the data set is representative of the physical environment of interest, we can build a supervised neural network that can capture the underlying dynamics of the environment, whether the environment is stationary or nonstationary. This is indeed a powerful statement on nonlinear adaptive filtering with profound practical implications. For a detailed and up-to-date book on neural networks, the reader is referred to Haykin [41].

References

[1] Gabor, D., 1954. "Communication theory and cybernetics", *IRE Transactions on Circuit Theory*, Vol. CT-1, pp. 19-31.

[2] Widrow, B., and M.E. Hoff, Jr., 1960. "Adaptive switching circuits", *IRE WESCON Convention Record*, pp. 96-104.

[3] Rosenblatt, F., 1958, "The Perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, Vol. 65, pp. 386-408.

[4] Haykin, S., 1996. *Adaptive Filter Theory*, Third Edition, Englewood Cliffs, NJ: Prentice-Hall.

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

[6] Hassibi, B., A.H. Sayed, and T. Kailath, 1996. "The H^{∞} optimality of the LMS algorithm", *IEEE Transactions on Signal Processing*, Vol. 44, pp. 267-280.

[7] Widrow, B., et al., 1976. "Stationary and nonstationary learning characteristics of the LMS adaptive filter", *Proc. IEEE*, Vol. 64, pp. 1151-1162.

[8] Solo, V., and X. Kong, 1995. *Adaptive Signal Processing Algorithms*, Englewood Cliffs, NJ: Prentice-Hall.

[9] Butterweck, H.J., 1995. "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption", *ICASSP*, Detroit, Michigan, pp. 1404-1407.

[10] Quirk, K., J. Zeidler, and L. Milstein, 1998. "Bounding the performance of the LMS estimator for cases where performance exceeds that of the finite Wiener filter", *ICASSP*, Seattle, Vol. III, pp. 1417.

[11] Clark, G.A., S.K. Mitra, and S.R. Parker, 1981. "Block implementation of adaptive digital filters", *IEEE Trans. Circuits Syst.*, Vol. CAS-28, pp. 584-592.

[12] Ferrara, E.R., Jr., 1980. "Fast implementation of LMS adaptive filters", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-28, pp. 474-475.

[13] Sommen, P.C.W., and J.A.K.S. Jayasinghe, 1988. "On frequency-domain adaptive filters using the overlap-add method", in *Proc. IEEE Int. Symp. Circuits Systems*, Espoo, Finland, pp. 27-30.

[14] Shynk, J.J., 1992. "Frequency-domain and multirate adaptive filtering", *IEEE Signal Processing Magazine*, Vol. 9, No. 1, pp. 14-37.

[15] Narayan, S.S., A.M. Peterson, and M.J. Narashima, 1983. "Transform domain LMS algorithm", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-31, pp. 609-615.

[16] Beaufays, F., 1995. "Transform-domain adaptive filters: an analytical approach", *IEEE*

Transactions on Signal Processing, Vol. 43, pp. 422-431.

[17] de Courville, M., and P. Duhamel, 1998. "Adaptive filtering in subbands using a weighted criterion", *IEEE Transactions on Signal Processing*, Vol. 46, pp. 2359-2371.

[18] Sayed, A.H., and T. Kailath, 1994. "A state-space approach to adaptive RLS filtering", *IEEE Signal Processing Magazine*, Vol. 11, pp. 18-60.

[19] Gentleman, W.M., and H.T. Kung, 1981. "Matrix triangularization by systolic arrays", in *Proceedings SPIE*, Vol. 298, *Real Time Signal Processing IV*, pp. 298-303.

[20] Hudson, J.E., and T.J. Shepherd, 1989. "Parallel weight extraction by a systolic least squares algorithm", in *Proceedings SPIE, Advanced Algorithms and Architectures for Signal Processing IV*, Vol. 1152, pp. 68-77.

[21] Yang, V., and J.F. Böhme, 1992. "Rotation-based RLS algorithms: unified derivations, numerical properties and parallel implementation", *IEEE Transactions on Signal Processing,* Vol, 40, pp. 1151-1167.

[22] Alexander, S.T., and A.L. Ghirnikar, 1993. "A method for recursive least-squares filtering based upon an inverse QR decomposition", *IEEE Transactions on Signal Processing*, Vol. 41, pp. 20-30.

[23] Proudler, I.K., J.G. McWhirter, and T.J. Shepherd, 1991. "Computationally efficient, QR decomposition approach to least squares adaptive filtering", *IEE Proceedings (London)*, Part F, Vol. 138, pp. 341-353.

[24] Macchi, O., and N.J. Bershad, 1991. "Adaptive recovery of a chirped sinusoid in noise, Part I: Performance of the RLS algorithm", *IEEE Trans. Acoust. Speech Signal Process.,* Vol. 39, pp. 583-494.

[25] Bershad, N.J., and O. Macchi, 1991. "Adaptive recovery of a chirped sinusoid in noise, Part 2: Performance of the LMS algorithm", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 39, pp. 595-602.

[26] Haykin, S., A.H. Sayed, J.R. Zeidler, P. Yee, and P.C. Wei, 1997. "Adaptive tracking of linear time-variant systems by extended RLS algorithm", *IEEE Transactions on Signal Processing*, Vol. 45, pp. 1118-1128.

[27] Sato, Y., 1975. "Two extensional applications of the zero-forcing equalization method", *IEEE Transactions on Communications*, Vol. COM-23, pp. 684-687.

[28] Godard, D.N., 1980. "Self-recovering equalization and carrier tracking in a two-dimensional data communication system", *IEEE Trans. Acoust. Speech Signal Process.,*Vol. COM-28, pp. 1867-1875.

[29] Treichler, J.R., and B.G. Agee, 1983. "A new approach to multipath correction of constant modulus signals", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-31, pp. 459-471.

[30] Johnson, C.R., Jr., P. Schniter, I. Fijalkow, L. Tong, J.D. Bohm, M.G. Larimore, D.R. Brown, R.A. Casas, T.J. Endres, S. Lambotharan, A. Touzni, H.H. Zeng, M. Green, and J.R. Treichler, 1991. "The Core of FSE-CMA Behavior Theory", In S. Haykin, editor, "Unsupervised Adaptive Filtering", New York, NY: Wiley.

[31] Hatzinakos, D., and C.L. Nikias, 1991. "Blind equalization using a tricepstrum based algorithm", *IEEE Transactions on Communications*, Vol. COM-39, pp. 669-682.

[32] Tong, L., G. Xu, and T. Kailath, 1994. "Blind identification and equalization based on second-order statistics: a time-domain approach", *IEEE Transactions on Information Theory,* Vol. 40, pp. 340-349.

[33] Moulines, E., P. Duhamel, J.-F. Cardoso, and S. Mayrargue, 1995. "Subspace methods for blind identification of multichannel FIR filters", *IEEE Transactions on Signal Processing*, Vol. 43, pp. 516-525.

[34] Gardner, W.A., 1991. "A new method of channel identification", *IEEE Transactions on Communications*, Vol. COM-39, pp. 813-817.

[35] Pham, D.T., and P. Garrat, 1997. "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach", *IEEE Transactions on Signal Processing*, Vol. 45, pp.

1712-1725.
[36] Bell, A.J., and T.J. Sejnowski, 1995. “An information-maximization approach to blind separation and blind deconvolution”, *Neural Computation*, Vol. 6, pp. 1129-1159.
[37] Amari, S., A. Cichoki, and H.H. Yang, 1996. “A new learning algorithm for blind signal separation”, *Advances in Neural Information Processing Systems*, Vol. 8, pp, 757-763, Cambridge, MA: MIT Press.
[38] Linsker, R., 1988. “Self-organization in a perceptual network”, *Computer*, Vol. 21, pp. 105-117.
[39] Haykin, S., editor, 1999. *Unsupervised Adaptive Filtering*, New York, NY: John Wiley.
[40] Vapnik, V.N., 1998. *Statistical Learning Theory*, New York, NY: John Wiley.
[41] Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation*, Second Edition, Englewood Cliffs, NJ: Prentice-Hall.