Noise Canceling and Channel Equalization

Bernard Widrow and Michael A. Lehr

Introduction

The fields of adaptive signal processing and adaptive neural networks have been developing independently but have the adaptive linear combiner (ALC) in common. With its inputs connected to a tapped delay line, the ALC becomes a key component of an adaptive filter. With its output connected to a quantizer, the ALC becomes an adaptive threshold element or adaptive neuron.

Adaptive filters have enjoyed great commercial success in the signal processing field. All high-speed modems now use adaptive equalization filters. Long-distance telephone and satellite communications links are being equipped with adaptive echo cancelers to filter out echo, allowing simultaneous two-way communications. Other applications include noise canceling and signal prediction.

Adaptive threshold elements, however, are the building blocks of neural networks. Today neural nets are the focus of widespread research interest. Although neural network systems have not yet had the commercial impact of adaptive filtering, they are already being used widely in industry, business, and science to solve problems in control, pattern recognition, prediction, and financial analysis.

The commonality of the ALC to adaptive signal processing and adaptive neural networks suggests that the two fields have much to share with each other. This article describes the manner in which the ALC can be used in practical adaptive noise canceling and channel equalization.

The Adaptive Linear Filter

The *adaptive linear combiner* is the basic building block for most adaptive systems. Its output is a linear combination of its inputs. At each sample time, this element receives an input signal vector or input pattern vector $\mathbf{X} = [x_0, x_1, x_2, \ldots, x_n]^T$, and a desired response *d*, a special input used to effect learning. The components of the input vector are weighted by a set of coefficients, the weight vector $\mathbf{W} = [w_0, w_1, w_2, \ldots, w_n]^T$. The sum of the weighted inputs is then computed, producing a linear output, the inner product $y = \mathbf{X}^T\mathbf{W}$. The output signal y is compared with desired response d, and the difference is the error signal, ε. To optimize performance, the ALC's weights are generally adjusted to minimize the mean square of the error signal. Of the many adaptive algorithms to adjust the weights automatically, the most popular is the Widrow-Hoff LMS (least mean square) algorithm devised in 1959 (Widrow and Hoff, 1960). For the weight update occurring at sample time k, this algorithm is given simply by

$$\mathbf{W}_{k+1} = \mathbf{W}_k + 2\mu\varepsilon_k\mathbf{X}_k, \tag{1}$$

where μ is a small constant which determines stability and learning speed. The LMS algorithm represents an efficient implementation of the method of gradient descent on the mean-square-error surface in weight space (Widrow and Stearns, 1985).

Digital signals used by *adaptive filters* generally originate from sampling continuous input signals by analog-to-digital conversion. Digital signals are often filtered by means of a tapped delay line or transversal filter, as shown in Figure 1. The sampled input signal is applied to a string of delay elements (denoted by z^{-1}), each delaying the signal by one sampling period. An ALC is seen connected to the taps between the delay elements. The filtered output is a linear combination of the current and past input signal samples. By varying the weights, the impulse response from input to output is directly controllable. Since the frequency response is the Fourier transform of the impulse response, controlling the impulse response controls the frequency response. The weights are usually adjusted so that the output signal provides the best least-squares match over time to the desired-response input signal.

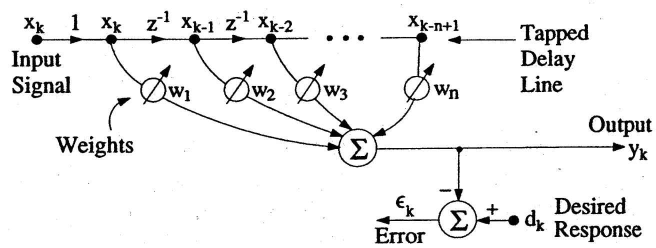

Figure 1. Linear adaptive transversal filter.

The literature reports many other forms of adaptive filters (Widrow and Stearns, 1985; Haykin, 1991). Some filters contain a second tapped delay line which feeds the output of the filter back to the input through a second set of weights. This feedback results in a transfer function which contains both poles and zeros. Because it uses no signal feedback, the filter of Figure 1 realizes only zeros. Another variation is adaptive filters based on ladderlike architectures called lattice structures which achieve more rapid convergence under certain conditions. The simplest, most robust, and most widely used filter, however, is that of Figure 1, adapted by the LMS algorithm.

The adaptive filter of Figure 1 has an input signal and produces an output signal. The desired response is supplied during training. A question naturally arises: If the desired response were known and available, why would one need the adaptive filter? Put another way, how would one obtain the desired response in a practical application? There is no general answer to these questions, but studying successful examples provides some insight.

Noise Canceling

Separating a signal from additive noise is a common problem in signal processing. Figure 2*A* shows a classical approach to this problem using optimal Wiener or Kalman filtering. The purpose of the optimal filter is to pass the signal s without distortion while stopping the noise n_0. In general, this cannot be done perfectly. Even with the best filter, the signal is distorted, and some noise goes through to the output.

Figure 2*B* shows another approach to the problem using adaptive filtering. This approach is viable only when an additional "reference input" is available containing noise n_1, which is correlated with the original corrupting noise n_0. In Figure 2*B*, the adaptive filter receives the reference noise, filters it, and subtracts the result from the noisy "primary input," $s + n_0$. For this adaptive filter, the noisy input $s + n_0$ acts as the desired response. The "system output" acts as the error for the adaptive filter. Adaptive noise canceling generally performs

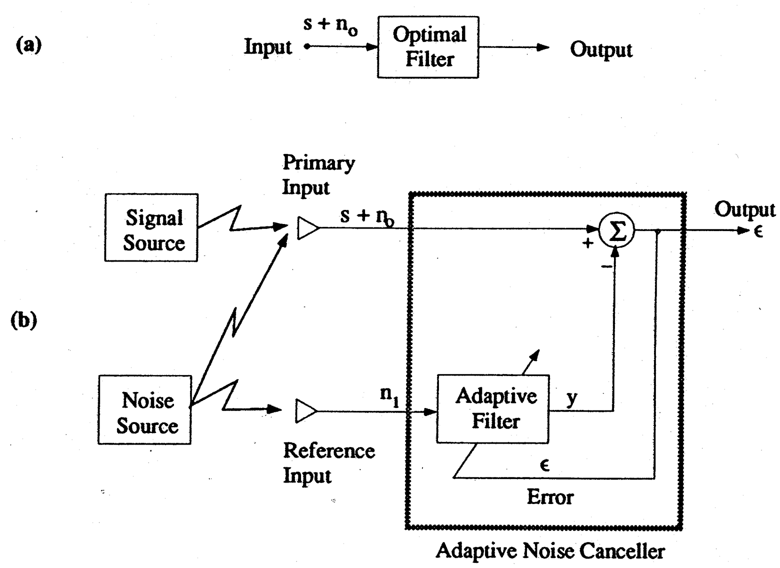

Figure 2. Separation of signal and noise. *A*, Classical approach. *B*, Adaptive noise-canceling approach.

much better than the classical approach, since the noise is subtracted out rather than filtered out.

One might think that some prior knowledge of the signal s or of the noises n_0 and n_1 would be necessary before the filter could adapt to produce the noise-canceling signal y. A simple argument shows, however, that little or no prior knowledge of s, n_0, n_1, or their interrelationships is required.

Assume that s, n_0, n_1, and y are statistically stationary and have zero means. Assume that s is uncorrelated with n_0 and n_1, and suppose that n_1 is correlated with n_0. The output is

$$\varepsilon = s + n_0 - y \tag{2}$$

Squaring, one obtains

$$\varepsilon^2 = s^2 + (n_0 - y)^2 + 2s(n_0 - y) \tag{3}$$

Taking expectations of both sides of Equation 3, and realizing that s is uncorrelated with n_0 and with y, yields

$$\begin{aligned} E[\varepsilon^2] &= E[s^2] + E[(n_0 - y)^2] + 2E[s(n_0 - y)] \\ &= E[s^2] + E[(n_0 - y)^2] \end{aligned} \tag{4}$$

Adapting the filter to minimize $E[\varepsilon^2]$ does not affect the signal power $E[s^2]$. Accordingly, the minimum output power is

$$E_{min}[\varepsilon^2] = E[s^2] + E_{min}[(n_0 - y)^2] \tag{5}$$

When the filter is adjusted so that $E[\varepsilon^2]$ is minimized, $E[(n_0 - y)^2]$ is therefore also minimized. The filter output y is then a best least-squares estimate of the primary noise n_0. Moreover, when $E[(n_0 - y)^2]$ is minimized, $E[(\varepsilon - s)^2]$ is also minimized, since, from Equation 2,

$$(\varepsilon - s) = (n_0 - y) \tag{6}$$

Adjusting or adapting the filter to minimize the total output power is tantamount to causing the output ε to be a best least-squares estimate of the signal s for the given structure and adjustability of the adaptive filter and for the given reference input.

There are many practical applications for adaptive noise canceling techniques. One involves canceling interference from the mother's heart when attempting to record clear fetal electrocardiograms (ECG). Figure 3 shows the location of the fetal and maternal hearts and the placement of the input leads. The abdominal leads provide the primary input (containing fetal ECG and interfering maternal ECG signals), and the chest leads provide the reference input (containing pure interference, the maternal ECG). Figure 4 shows the results. The maternal ECG from the chest leads was adaptively filtered and subtracted from the abdominal signal, leaving the fetal ECG.

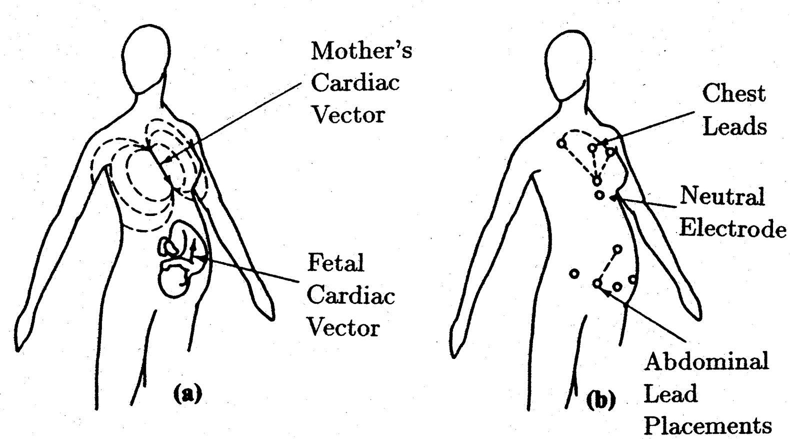

Figure 3. Canceling maternal heartbeat in fetal electrocardiography. *A*, Cardiac electric field vectors of mother and fetus. *B*, Placement of leads.

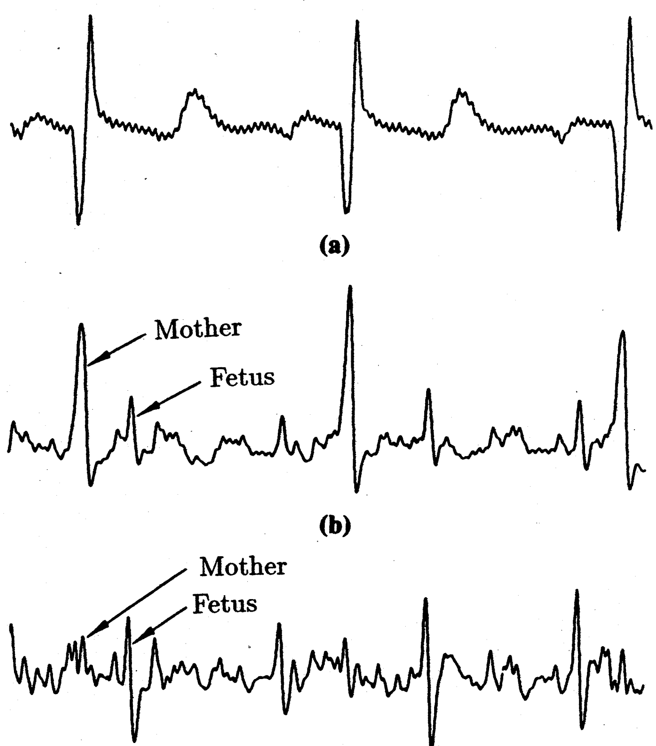

Figure 4. Result of fetal ECG experiment. *A*, Reference input (chest lead). *B*, Primary input (abdominal lead). *C*, Noise canceler output —the fetal ECG signal with maternal interference removed.

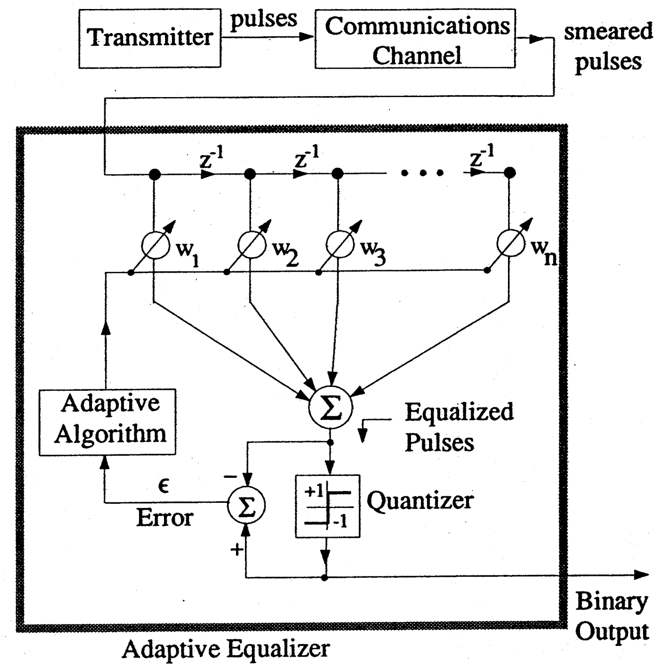

Figure 5. Adaptive channel equalizer with decision-directed learning.

Channel Equalization

Telephone channels, radio channels, and even fiber-optic channels can have nonflat frequency responses and nonlinear phase responses in the signal passband. Sending digital data at high speed through these channels often results in a phenomenon called "intersymbol interference," caused by signal pulse smearing in the dispersive medium. Equalization in data modems combats this phenomenon by filtering incoming signals. A modem's adaptive filter, by adapting itself to become a channel inverse, can compensate for the irregularities in channel magnitude and phase response.

The *adaptive equalizer* in Figure 5 consists of a tapped delay line with an adaptive linear combiner connected to the taps. Deconvolved signal pulses appear at the weighted sum, which is quantized to provide a binary output corresponding to the original binary data transmitted through the channel. Any least-squares algorithm can adapt the weights, but the telecommunications industry uses the LMS algorithm almost exclusively.

In operation, the weight at a central tap is generally fixed at unit value. Initially, all other weights are set to zero so that the equalizer has a flat frequency response and a linear phase response. Without equalization, telephone channels can provide quantized binary outputs that reproduce the transmitted data stream with error rates of 10^{-1} or less. As such, the quantized binary output can be used as the desired response to train the neuron. It is a noisy desired response initially. Sporadic errors cause adaptation in the wrong direction, but on average, adaptation proceeds correctly. As the neuron learns, noise in the desired response diminishes. Once the adaptive equalizer converges, the error rate typically is 10^{-6} or less. The method, called "decision-directed" learning, was invented by Robert W. Lucky of AT&T Bell Labs.

Using a modem with an adaptive equalizer enables transmitting approximately four times as much data through the same channel with the same reliability as without equalization.

Discussion

The simple concept of adapting the weights of a linear combiner to cause its output to approximate a desired response is the basis for the field of adaptive signal processing. In a large number of practical cases, it is possible to exploit this idea to solve difficult signal-processing problems with surprising accuracy, even when the statistics of the involved signals are unknown. Adaptive noise canceling and adaptive channel equalization are two examples which indicate the power of this approach. The burgeoning fields of neural networks, adaptive inverse control, and active noise control provide an indication of the generality and importance of methods based on this approach.

Acknowledgments. This work was sponsored by NSF under grant IRI 91–12531, by ONR under contract no N00014–92–J–1787, by EPRI under contract RP 8010–13, and by the U.S. Army under contract DAAK70–92–K–0003.

Road Map: Applications of Neural Networks
Related Reading: Adaptive Filtering; Adaptive Signal Processing; Perceptrons, Adalines, and Backpropagation

References

Haykin, S., 1991, *Adaptive Filter Theory*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall.

Widrow, B., and Hoff, M., 1960, Adaptive switching circuits, in *1960 IRE WESCON Convention Record*, Part 4, New York: IRE, pp. 96–104.

Widrow, B., and Stearns, S., 1985, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall. ◆

The Handbook of Brain Theory and Neural Networks

EDITED BY
Michael A. Arbib

A Bradford Book
THE MIT PRESS
Cambridge, Massachusetts
London, England