Adaptive Combining of Wideband Array Data for Optimal Reception

RICHARD T. LACOSS

Abstract—**A study of several closely related adaptive processors for array data has been completed. Processors were designed to converge approximately to a minimum-variance linear unbiased estimator of an unknown signal common to all elements of the array. Any spatial structure of the background noise is used by such a system to enhance the output signal-to-noise ratio. Possible areas of application include sonar, underwater communication, space communication, and seismology.**

The basic linear adaptive processor has variable coefficients adjusted by a rule similar to that for the minimization by the projection gradient method of a quadratic form which is subject to a linear constraint. Modifications of the basic adjustment procedure have been introduced to reduce system sensitivity to data anomalies, decrease computational requirements, and decrease memory requirements.

Experimental evaluation of the adaptive array processors has been completed using data from an actual array of seismometers. Both transient recovery from initially poor processor coefficients and steady-state operation have been found quite satisfactory.

I. Introduction

A LARGE APERTURE Seismic Array (LASA) has been in operation in the United States since mid-1965. Fig. 1 shows the geometrical arrangement of the array. There are 525 short-period vertical seismometers located in 21 clusters of 25 instruments. Smaller arrays of seismometers have been in operation in other nations as well as in the United States for a number of years.[1] Of course, the concept of using an array to increase signal-to-noise ratio is well known. For example, it has supplied the motivation for the installation of hydrophone arrays and arrays of electromagnetic antennas, as well as many other kinds of arrays.

Although arrays of all kinds have been designed and continue to be designed under the assumption that unwanted noise is spatially disorganized, it has become clear that this basic assumption is often untenable. For example, biological or man-made noise can be highly directional as seen by a hydrophone array. A radio telescope may be subjected to highly directional noise from an active region of the sky or even from the sun. Seismic arrays will see directional noise produced by weather conditions, local cultural activity, and other sources.

Data from an array can be processed to receive signals and to optimally reject directional noise.[2],[3] Fig. 2 shows the structure of one simple processor which can

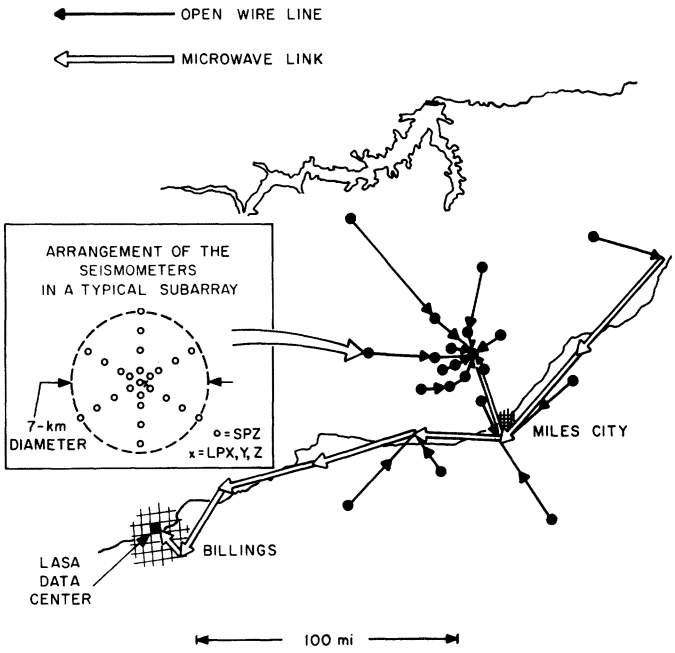

Fig. 1. Large aperture seismic array.

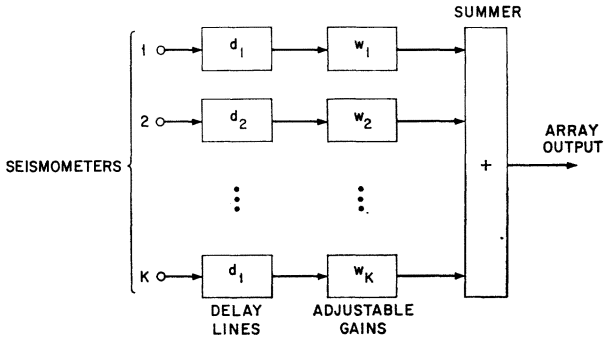

Fig. 2. Simple array data processor.

be used to process wideband data from a seismic array. Appropriate delays are introduced to line up the desired signal on each seismogram. The weights applied to each channel are chosen to use the directional structure of the noise to minimize noise power in the final output. Under stationary noise conditions, lengthy observations of noise can be made and, with the expenditure of considerable computational effort, best values for weights can be determined. Such a procedure is costly, off-line, and may not work if the noise structure is time variable. These considerations led to the adaptive methods for processing seismic data which are described in the main sections of this paper.

Manuscript received August 4, 1967.

The author is with the M.I.T. Lincoln Laboratory, Lexington, Mass. (Operated with support from the U. S. Advanced Research Projects Agency.)

The initial work with active adaptive arrays was not concerned primarily with directional noise but with the distortion of the wavefronts of the desired signal.[4] This kind of adaptive system usually requires that test signals be used to learn the shape of the wavefront. Large seismic arrays also require corrections for nonplanar wavefronts. The appropriate corrections, when required, are normally obtained from observations of many different earthquakes.[5] Other adaptive methods for adjusting delays before summing seismic signals are under consideration.[6] The 7.0-km subarrays of LASA are sufficiently small so that signals from earthquakes can be treated as plane waves. Since we have only attempted to combine signals from a single subarray, the adaptive adjustment of delays has not been considered.

Array processors which adapt to noise structure rather than to signal structure are not very common. The filtering methods of Capon *et al.*[2] are adaptive insofar as the noise structure is estimated shortly before filters are to be used. However, the methods are quite impractical for most on-line operations. This is especially true of a time variable situation where filters must be updated very often. Some adaptive array processors which are more similar to those developed in the sequel have been described in the literature. Shor[7] has proposed a gradient adaptive method for narrowband hydrophone arrays when the autocorrelation function of signals is known. Adams[8] has indicated a possible on-line adaptive structure for a ground station receiving signals from a deep-space probe.

The contents of the present paper are arranged as follows. In Section II, the design of a processor is mathematically formalized. Section III describes a linear adaptive method for filter design which is related to stochastic approximation methods and is computationally feasible for on-line operation. These evolved methods reduced sensitivity to anomalous noise bursts and reduced memory requirements. The final algorithm is appropriate for a fast special-purpose device. Section VI containing experimental results using actual seismic data, has been included in order to verify that the adaptive design methods do operate as asserted. Section VII contains the summary, discussion, and suggestions for future investigation.

II. Stochastic Approximation Design of an Optimal Processor for Stationary Noise

Consider a time variable vector whose components at time t are the outputs of the delay lines indicated in Fig. 2. The delays are assumed correct to line up a desired signal $s(t)$. The observed vector is then

$$x_k(t) = s(t) + n_k(t), \qquad k = 1, \cdots, K \tag{1}$$

where K is the dimension of the observation vector, and the $n_k(t)$ are noise waveforms as seen at the output of the delay lines. Assume that the $n_k(t)$ are zero mean and wide-sense stationary with correlation functions

$$E\{n_k(t)n_l(t+\tau)\} = R_{kl}(\tau). \tag{2}$$

The output waveform from the processor is

$$\begin{aligned} y(t) &= \sum_{k=1}^{K} w_k x_k(t) \\ &= \left[\sum_{k=1}^{K} w_k\right] s(t) + \sum_{k=1}^{K} w_k n_k(t). \end{aligned} \tag{3}$$

Thus, if $y(t)$ is to be an unbiased estimator of $s(t)$, the w_k must satisfy

$$\sum_{k=1}^{K} w_k = 1. \tag{4}$$

The optimum choice of the w_k is that which minimizes the variance of $y(t)$,

$$\text{var } y(t) = \sum_{k,l=1}^{K} w_k R_{kl}(0) w_l. \tag{5}$$

The use of Lagrange multipliers immediately gives the optimum w_k by solving

$$\left[\begin{array}{c|c} R(0) & \mathbf{1} \\ \hline \mathbf{1}^T & 0 \end{array}\right]\left[\begin{array}{c} \boldsymbol{w} \\ \hline \lambda \end{array}\right] = \left[\begin{array}{c} \mathbf{0} \\ \hline 1 \end{array}\right]$$

where $R(0)$ is a $K \times K$ matrix with elements $R_{kl}(0)$,

$$\boldsymbol{w} = \text{col } (w_1, \cdots, w_k)$$
$$\mathbf{0} = \text{col } (0, \cdots, 0)$$
$$\mathbf{1} = \text{col } (1, \cdots, 1)$$

and a superscript T indicates transpose. The Lagrange multiplier is λ.[2]

In order to derive a practical iterative method to optimize the array processor, it is more fruitful to consider a gradient projection method of solving for optimum weights. If $\boldsymbol{w}$ satisfies (4) and is optimum, then the gradient of var $\{y(t)\}$ must be collinear with $\mathbf{1}$. That is, the projection of the gradient into the $K-1$ dimensional subspace defined by (4) must be zero. Thus,

$$\mathcal{P}R(0)\boldsymbol{w} = \mathbf{0} \tag{6}$$

and (4) define the optimum weights where $R(0)\boldsymbol{w}$ is the gradient of (5) with respect to $\boldsymbol{w}$ and

$$\mathcal{P} = I - \frac{1}{K}\mathbf{1}\,\mathbf{1}^T \tag{7}$$

is the projection operator.

Let $\boldsymbol{w}(k)$ be a sequence of weight vectors given by

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - a(k)\mathcal{P}R(0)\boldsymbol{w}(k) \tag{8}$$

where $a(k)$ is smaller than one divided by the largest eigenvalue of $R(0)$,

$$\lim_{N\to\infty} \sum_{k=1}^{N} a(k) = \infty \tag{9}$$

and $\boldsymbol{w}(0)$ satisfies (4). It is demonstrated in Appendix I that

$$\lim_{k\to\infty} \boldsymbol{w}(k) = \boldsymbol{w}^*$$

where $\boldsymbol{w}^*$ is the optimum weight vector. Weaker constraints upon $a(k)$ are possible but complicate the issue without adding significantly.

It is a small step from (8) to a variety of stochastic approximation[9] methods which can be used to solve the stationary problem described in this section. Suppose that $R(0)$ is not known a priori, but that a sequence of estimates $\hat{R}(k)$ of $R(0)$ is available. That suggests using

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - a(k)\mathcal{P}\hat{R}(k)w(k) \tag{10}$$

in order to try to converge to $\boldsymbol{w}^*$. One possibility is to pick $\hat{R}(k)$ which converge to $R(0)$ as k gets large. This tends to be computationally wasteful and has not been pursued. We have considered

$$\hat{R}(k) = \frac{1}{L}\sum_{j=1}^{L} \mathbf{x}(kL+j)\mathbf{x}^T(kL+j) \tag{11}$$

where L is a given integer and, for convenience, time is assumed to take on only integer values. Changes in the weighting vector take place after every L observations of data. This is one of many choices which could be made.

For seismic data, the signal is an earthquake so that $\hat{R}(k)$ is usually computed with no signal present. The expected value of $\hat{R}(k)$ is then $R(0)$. In the presence of signal, $\hat{R}(k)$ has a different average value but, due to the projection operator, the $\hat{R}(k)$ can still be used effectively. One way to see this is to note that in the presence of signal

$$E\{\hat{R}(k)\} = R(0) + \sigma_s^2(k)\mathbf{1}\,\mathbf{1}^T$$

where $\sigma_s^2(k)$ is the rms value of signal over the set of L points used to compute $\hat{R}(k)$. Now consider

$$E\mathcal{P}\hat{R}(k) = \mathcal{P}R(0) + \sigma_s^2(k)\mathcal{P}\mathbf{1}\,\mathbf{1}^T.$$

Substituting (7) for $\mathcal{P}$ in the second term on the right gives

$$\sigma_s^2(k)\mathcal{P}\mathbf{1}\,\mathbf{1}^T = \sigma_s^2(k)\{\mathbf{1}\,\mathbf{1}^T - \mathbf{1}\,\mathbf{1}^T\}.$$

Thus,

$$E\mathcal{P}\hat{R}(k) = \mathcal{P}R(0).$$

On the average, the projected gradient is correct even when signals are present. Only the situation with no signal present is considered in most of what follows.

By introducing a few additional constraints upon the $a(k)$ and $R(\tau)$, it is possible to show that $\boldsymbol{w}(k)$ given by (10) converges in mean square to $\boldsymbol{w}^*$. If

$$\sum_{k=1}^{\infty} a^2(k) < \infty$$

and $R(\tau)$ goes to zero sufficiently quickly with $|\tau|$, then convergence is assured. A proof is outlined in Appendix II for the case $R(\tau)=0$, if $\tau\neq 0$. Weaker bounds on $R(\tau)$, which admit most situations of real interest, require significantly more sophisticated treatment but can be handled by methods similar to those used by Sakrison.[10]

III. Linear Iterative Design with Constant Gain

Although the iterations described in Section II are interesting, they do not seem particularly relevant for continuing on-line operation of an adaptive array processor. For that purpose, consider a rule given by (10) and (11), but for which $a(k)=a$ is a positive constant. It is demonstrated in Appendix III that the asymptotic mean-square error between $\boldsymbol{w}(k)$ and the optimum weight is bounded by $aC_1/(1-aC_2)$ where C_1 and C_2 are positive constants. This verifies the intuitive notion that a linear iterative process with constant gain a can be made satisfactory by using a sufficiently small gain.

The linear iteration rule should not be applied by first computing $\hat{R}(k)$ from (11) and then using (10). Using (3) as well as (10) and (11), the iteration law can be written as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \frac{a}{L}\mathcal{P}\left[\sum_{j=1}^{L} \mathbf{x}(kL+j)y(kL+j)\right]. \tag{12}$$

The scalar $y(kL+j)$ requires K MULTIPLYs and ADDs (MADs) to compute. If optimum processing is being done, this is the minimum of computation for each output point, even when the optimum weights are known. The vector

$$\sum_{j=1}^{L} \mathbf{x}(kL+j)y(kL+j)$$

is thus formed by $2K\times L$ MADs. This is just double the MADs used to apply optimum weights to L samples of data. Suppose that (10) and (11) had been used directly to find $\boldsymbol{w}(k+1)$. The same vector $L\hat{R}(k)\boldsymbol{w}(k)$ would then require $K^2\times(L+1)$ MADs to compute. If $K>2$, the computational advantage of using (12) is clear.

It is of interest to compare the computation required by (12) for $k=0, 1, \cdots, N-1$ with that required when directly solving (5) with $R(0)$ replaced by

$$\tilde{R}(0) = \frac{1}{N}\sum_{k=0}^{N-1} \hat{R}(k).$$

The application of (12) requires $2KL+2K$ MADs per iteration or $2NK(L+1)$ MADs total. Computation of $\tilde{R}(0)$ will require NLK^2 MADs. We shall neglect any additional computations required to solve (5) using $\tilde{R}(0)$ in place of $R(0)$. The ratio of the number of MADs is thus $2(1+1/L)/K$.

Suppose that experience shows that N is a good value to use in computing $\tilde{R}(0)$ in the sense that averaging fewer terms to get $\tilde{R}(0)$ results in unsatisfactory values of $\boldsymbol{w}$ computed from (5). If the iterative rule converges in less than about $KN/2$ steps, then the iterative scheme will be at least as efficient as the direct method. Experiments with modifications of (12) indicate that convergence can be obtained in about $3N$ steps or less for the seismic data which have been used. Since $K=24$ was used, the linear iterative mode would require about 0.25 of the computation required by the direct solution. In fact, one of the modified methods considered requires fewer computations than the linear mode and can be implemented with special-purpose equipment. Thus, this estimate of time saving is somewhat pessimistic. Of course, in addition to saving time, the iterative methods are well adapted to on-line operation and, as is discussed in Section V, can reduce computer memory requirements.

IV. Clipped Gradient Iterative Method

One shortcoming of the linear constant gain technique described in Section III is its sensitivity to large atypical noise bursts. A variety of tests can be applied to check for significantly anomalous data and decisions made concerning their disposition. If the norm of the vector

$$\hat{\boldsymbol{g}}(k) = \frac{1}{L}\sum_{j=1}^{L} \mathbf{x}(kL+j)y(kL+j)$$

is unusually large, it might be discarded and no adjustment made to the weights. An alternative which has been implemented is to normalize $\hat{\boldsymbol{g}}(k)$ for each k. The resulting recursion relation for $\boldsymbol{w}(k)$ is

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \frac{a}{\|\hat{\boldsymbol{g}}(k)\|}\mathcal{P}\hat{\boldsymbol{g}}(k). \tag{13}$$

Note that it is the estimated gradient vector $\hat{\boldsymbol{g}}(k)$ which is infinitely clipped, not the projection of $\hat{\boldsymbol{g}}(k)$. Thus, if L is large and $\boldsymbol{w}(k)=\boldsymbol{w}^*$, there will tend to be no correction because of the projection operator. If $\boldsymbol{w}(k)$ is quite different from $\boldsymbol{w}^*$, the $\hat{\boldsymbol{g}}(k)$ will no longer tend to be in the null space of $\mathcal{P}$ and the norm of the correction vector can be as large as a, but no larger.

Although it is intuitively clear, for large L and small a, that for large k the deviation between $\boldsymbol{w}(k)$ and $\boldsymbol{w}^*$ will be small, no explicit demonstrations have been developed. However, the assertion has been verified experimentally. The results, presented in Section VI of this paper, indicated that satisfactory convergence could be obtained. The particular choices of a and M gave fast convergence and good stability.

V. One-Bit Correlation Iterative Method

More stable operation under a variety of noise conditions was obtained using the normalized gradient algorithm than by using the linear gradient. However, the former required additional computations to normalize the gradient estimates before each adjustment of weights. The algorithm given below was evolved in order to reduce computational requirements without sacrificing the ability to operate in the presence of anomalous transient noise conditions. Much of this algorithm is well suited for implementation by a special-purpose machine, and requires significantly less computer memory than the preceding algorithms.

Assume that σ_i^2 is known and is the expected squared power in the ith channel of the array in the absence of signal. Let σ^2 denote the average of the σ_i^2. Let $\boldsymbol{h}(k)$ be a vector defined by

$$h_i(k) = \frac{1}{L}\sum_{j=1}^{L} \operatorname{sgn}\,[x_i(kL+j)y(kL+j)], \qquad i = 1, 2, \cdots, K \tag{14}$$

where sgn $(\cdot)$ is ± 1, depending upon the sign of its argument. Finally, define $\tilde{\boldsymbol{g}}(k)$ by

$$\tilde{g}_i(k) = \frac{\sigma_i}{\sigma}\sin\left(\frac{\pi}{2}h_i(k)\right). \tag{15}$$

The adjustment rule for weights which is to be considered can now be written as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - a\mathcal{P}\tilde{\boldsymbol{g}}(k). \tag{16}$$

This modification of the linear gradient iterative scheme can be understood as follows. Define a vector $\boldsymbol{G}$ by

$$G_i = \frac{\sigma_i}{\sigma}\sin\left\{\frac{\pi}{2}E[\operatorname{sgn}\,(x_i(kL)y(kL))]\right\}, \qquad i = 1, 2, \cdots, K \tag{17}$$

and consider

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - a\mathcal{P}\boldsymbol{G}. \tag{18}$$

If $x_i(kL)$ and $y(kL)$ are Gaussian data with zero mean, then[11]

$$\sin\left\{\frac{\pi}{2}E[\operatorname{sgn}\,(x_i(kL)y(kL)]\right\} = \frac{E[x_i(kL)y(kL)]}{\sqrt{E[x_i^2(kL)]E[y^2(kL)]}}. \tag{19}$$

It follows that

$$\boldsymbol{G} = \frac{1}{\sigma\sqrt{E[y^2(kL)]}}R(0)\boldsymbol{w}(k). \tag{20}$$

Equation (18) can be rewritten as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) - \frac{a}{\sigma\sqrt{Ey^2(kL)}}\mathcal{P}R(0)\boldsymbol{w}(k). \tag{21}$$

Equation (21) is essentially the same as (8). It is clear then that the $\boldsymbol{w}(k)$ given by (18) would converge to the optimum $\boldsymbol{w}^*$. The iteration defined by (14) and (15) is the same as (18), except that the estimate $h_i(k)$ is used

in place of $E\{\text{sgn } [x_i(kL)] \text{ sgn } y(k)]\}$. It is clear that for large L this should be a successful method of finding $\boldsymbol{w}^*$. No error analysis has been completed, but experimental data indicate success for very modest L.

This adaptive scheme which uses one-bit correlation can be computationally more efficient than the infinitely clipped gradient method of Section IV. Calculation of $L\boldsymbol{h}(k)$ requires $L \times K$ MADs. Those are simply operations required to obtain a processed array output for L values of time. The remaining logic and counting operations required to obtain $L\boldsymbol{h}(k)$ should add, at worst, only a small amount to computation time. In fact, they can be trivially accomplished in parallel with forming $y(k)$, the array output at time k. Obtaining sin $[(\pi/2)h_i(k)]$ can be done by a simple table look-up. Forming $w_i(k+1)$ from $w_i(k)$ and sin $[(\pi/2)h_i(k)]$ requires a MULTIPLY and ADD. This amounts to $(L+1)K$ MADs per iteration.

If the σ_i must be updated slowly or periodically, that would add more operations. The upper bound on additional operations is about $L \times K$ MADs required to estimate the σ_i or update their estimates with L new observations. If such continuous updating of estimates of σ_i is required, the algorithm described in this section is comparable in computational requirements to that of Section IV. If σ_i changes very slowly, as would be the case if variations were due to equipment variations for different sensors, or several output traces are to be formed so the values of σ_i could be used for several outputs, then the number of operations for estimating the σ_i can be greatly reduced. For seismic data, the ratios σ_i/σ_j, which are really what must be constant, tend to be stable for long times so that the estimation of the σ_i would require only a small amount of computation in an adaptive system.

As long as the number of different optimum beams to be formed is considerably less than the total number of sensors, the use of one-bit correlation introduces another possible saving. The memory required in a general or special-purpose digital device can be reduced by an order of magnitude by the use of one-bit correlation. Up to this point, it has been assumed that signals are lined up in time as indicated by (1). Suppose the delay lines indicated in Fig. 2 are all removed so that (1) must be replaced by

$$x_k(t) = s(t - d_k) + n_k(t), \qquad k = 1, \cdots, K$$

where the d_k are constants indicating the delay between some central element of the array and the kth element. In this case,

$$y(t) = \sum_{k=1}^{K} w_k x_k(t + d_k)$$

is the output of the array processor designed to pass $s(t)$.

Let

$$T_1 - 1 = \max_i d_i - \min_i d_i$$

and

$$T_2 - 1 = \sum_{k=1}^{K} [d_k - \min_i d_i].$$

The sum box method[12] described in Appendix IV for computing $y(t)$ will require only T_1 words of memory for each beam formed. This is much less than the T_2 required, independent of the number of beams, if the x_k are passed through delay lines to line up signals $s(t)$. However, except for the one-bit correlation iterative method, the methods we have considered require the buffer of length T_2 in order to obtain $w(k+1)$ from $w(k)$ and the data. If B is the length in bits of a computer word, then storage locations can be reduced from T_2 to $T_1+(T_2/B)$ by use of one-bit correlation, which requires that only the sign bit of data be passed through a delay line, in conjunction with the sum box method.

VI. Experimental Results

Background seismic noise from a collection of 24 seismometers spread over a 7-km aperture in Montana has been used to test the iteration rules previously described. A PDP-7 computer was used to make data tapes compatible with an IBM 360/65 system which was used as the principal computational tool in the experiments. Early experience with the linear gradient method of adjustment indicated that, for a given rate of convergence, the method was too sensitive to atypical seismic noise bursts, bit dropouts in digital data, and a variety of equipment failures. Only results from the hard-limited gradient and one-bit correlation methods have been included here. The noise was a 900-second length of noise sampled every 0.05 second. The total bandwidth of the data was from about 0.1 Hz to 5.0 Hz, although most of the noise was in the portion below 1.0 Hz. A detailed discussion of optimum processing of such short-period seismic data can be found in a report by Capon *et al.*[13] Only data relating specifically to the operation of adaptive processors are included in the present paper.

Figs. 3 and 4 show the recovery of the iterative schemes from an initial condition in which each of the $K=24$ seismometers has a weight $1/K$. Results are shown for a variety of gain constants, $c=a$. Adjustments of weights were made every $L=25$ samples. The power plotted at any time t is the average which would be observed over the entire 900 seconds if the weights were set constant over that interval at the value which they had at time t. The optimum value indicated is that which results from picking constant weights to minimize the mean-square array output over the 900-second interval. The mean-square value over 900 seconds which would have been obtained with weights $1/K$, is 6541 digital levels squared. The mean-square value of the 24 individual channels was 10 895 digital levels squared during the 900 seconds.

Figs. 5 and 6 show the observed behavior of iterative array processors during the last 75 of 900 seconds of a

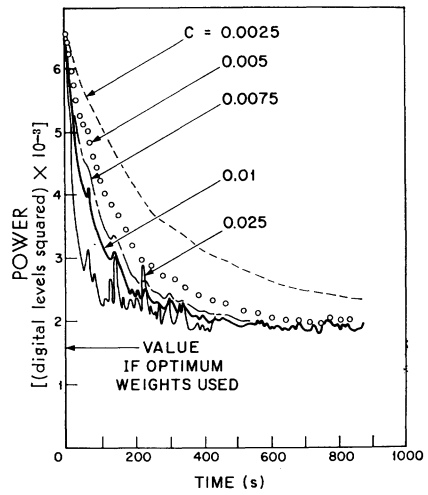

Fig. 3. Transient recovery of one-bit correlator adaptive array processor, $K=24$, $L=25$.

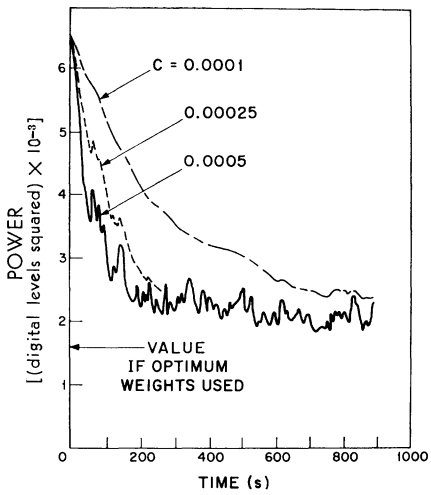

Fig. 4. Transient recovery of clipped gradient adaptive array processor, $K=24$, $L=25$.

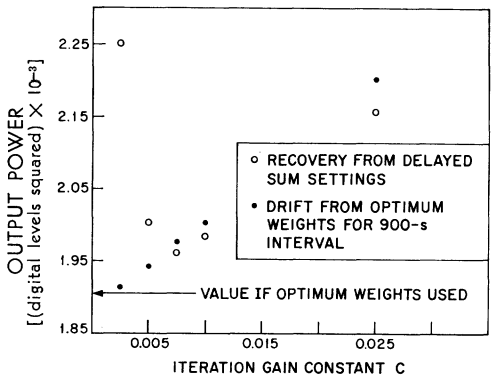

Fig. 5. Behavior of one-bit correlator adaptive array processor after 900 seconds, $K=24$, $L=25$.

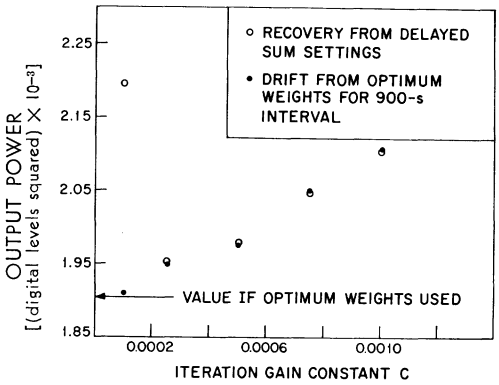

Fig. 6. Behavior of clipped gradient adaptive array processor after 900 seconds, $K=24$, $L=25$

data interval used for experimentation. Recovery from initial $1/K$ weights is shown as a function of c. As expected, the mean-square output power decreases as c decreases. This holds as long as c is large enough so that the adaptive processor has reached steady state. Drift away from initially optimum weights is also shown in Figs. 5 and 6. For large values of c, it is clear that steady state has been reached since the points are close to those resulting from recovery from $1/K$ weights.

The results shown in Figs. 3 through 6 are typical of those obtained using other samples of data. The choice of gain constants c has not been sensitive to the particular sample of noise used. Note that a change in the general noise level cannot affect the adjustment of weights.

It is generally recognized that the percentage deviation of a quadratic form from its minimum value may be very different from the deviation of the vector involved from that which minimizes the quadratic form. This has been experimentally observed during the iterative array design experiments. For example, while using one-bit correlation iterations, an array output which was only 2.5 percent larger than the optimum was observed while the error in the weights was 41 percent. In the case of clipped gradient operation, 1.5 percent and 42 percent were observed. This simply underlines the fact that, for optimization problems, it is the quantity minimized, not the parameters manipulated, which is of prime importance.

VII. Comments

Experimental results using seismic data have verified that practical adaptive array processors can perform nearly as well as optimum processors in a stationary environment. In addition, the adaptive processors can compensate for nonstationary trends in the noise statistics. It should be possible to adapt similar iterative processors to sonar and electromagnetic arrays which operate in a directional noise environment. It might be possible to minimize reverberation as well as ambient noise in systems where reverberation is significant. Extensive modifications might be required for any operational application of adaptive array processors.

Computational and memory requirements of the adaptive processors have been emphasized. Their efficiency in these respects would be of great importance in any operational system embodying the basic adaptive ideas. The fact that adaptive operators can also be robust would also be significant.

In addition to possible applications, the adaptive processors suggest more theoretical work is needed. In a stationary noise field, the asymptotic deviation of the average array output power from the optimum minimum value could further be investigated theoretically. In a nonstationary field, the optimum constant values of gain constants would be of interest. These are, of course, the same general kinds of problems which are relevant to most adaptive systems.

Appendix I

Although the convergence of $\boldsymbol{w}(k)$ to $\boldsymbol{w}^*$ given by (8) is intuitively clear, it requires some effort to explicitly demonstrate convergence. It is convenient for this purpose to introduce a change in coordinate system. Let ψ be an orthonormal matrix which transforms $1/K\ \mathbf{1}$ into col $(0, \cdots, 0, \sqrt{1/K})$. If $\boldsymbol{w}$ satisfies (4), then define

$$\psi \boldsymbol{w} = \operatorname{col}\left[p_1, \cdots, p_{K-1}, \sqrt{\frac{1}{K}}\right] = \begin{bmatrix} \boldsymbol{p} \\ \hline \sqrt{1/K} \end{bmatrix} = \boldsymbol{P}.$$

Clearly any $\boldsymbol{w} = \psi^T \boldsymbol{P}$ satisfies (4) for any $\boldsymbol{p}$. Equation (6) can be written as

$$\psi \mathcal{P} R(0) \psi^T \begin{bmatrix} \boldsymbol{p} \\ \hline \sqrt{1/K} \end{bmatrix} = \mathbf{0}. \tag{22}$$

Since $\mathcal{P}$ projects into a space orthogonal to the last basis vector for the new coordinate system, the Kth element in (22) is identically zero independent of $\boldsymbol{p}$. Define a matrix M by

$$M_{kl} = [\psi \mathcal{P} R(0) \psi^T]_{kl}, \quad 1 \leq k,l \leq K-1$$

and a vector $\boldsymbol{v}$ by

$$v_k = \sqrt{\frac{1}{K}}\,[\psi \mathcal{P} R(0) \psi^T]_{kK}, \quad 1 \leq k \leq K-1.$$

Equation (22) can then be written

$$M\boldsymbol{p} + \boldsymbol{v} = 0.$$

The solution $\boldsymbol{p}^*$ of this equation can clearly be used to obtain $\boldsymbol{w}^*$. The matrix M is nonsingular as long as $R(0)$ is nonsingular.

A little algebra shows that (8) and the constraints on $\boldsymbol{w}(0)$ are equivalent to

$$\boldsymbol{p}(k+1) = \boldsymbol{p}(k) - a(k)[M\boldsymbol{p}(k) + \boldsymbol{v}] \tag{23}$$

with no constraints on $\boldsymbol{p}(0)$. It remains to show that

$$\lim_{k \to \infty} \boldsymbol{p}(k) = \boldsymbol{p}^*.$$

From (23) and the definition of $\boldsymbol{p}^*$ one obtains

$$\boldsymbol{\varepsilon}(k+1) = \boldsymbol{\varepsilon}(k) - a(k) M \boldsymbol{\varepsilon}(k)$$

where

$$\boldsymbol{\varepsilon}(k) = \boldsymbol{p}(k) - \boldsymbol{p}^*.$$

Let μ and λ be the minimum and maximum eigenvalues of M. The relation of M to $R(0)$ forces the relationships

$$\delta \leq 1 - a(k)\lambda \leq 1 - a(k)\mu$$

where δ is some positive number. It should be clear that $\boldsymbol{\varepsilon}(k)$ converges to zero if

$$e(k+1) = [1 - a(k)\mu] e(k) \tag{24}$$

converges to zero where

$$e(0) = \sum_{k=1}^{K} [\varepsilon_i^2(0)]^{1/2} = \|\boldsymbol{\varepsilon}(0)\|.$$

Since

$$e(k+1) = \prod_{i=0}^{k} [1 - a(k)\mu] e(0)$$

it is sufficient to show that

$$\prod_{i=0}^{k} [1 - a(k)\mu]$$

converges to zero. This is done using the relationship

$$\log [1 - a(k)\mu] \leq - a(k)\mu$$

which is valid if $a(k)\mu$ is positive and less than 1. Thus,

$$\log \prod_{i=0}^{k} [1 - a(k)\mu] \leq - \mu \sum_{i=0}^{k} a(k).$$

The sum of the $a(k)$ converges to $+\infty$ which completes the proof that $e(k)$ converges to zero.

Appendix II

Continuing in the coordinate system used in Appendix I, the stochastic approximation method specified by (10) and (11) is equivalent to

$$\boldsymbol{p}(k+1) = \boldsymbol{p}(k) - a(k)[\hat{M}(k)\boldsymbol{p}(k) + \hat{\boldsymbol{v}}(k)] \tag{25}$$

where $\hat{M}(k)$ and $\hat{\boldsymbol{v}}(k)$ are obtained from $\hat{R}(k)$ in the same way as M and $\boldsymbol{v}$ were obtained from $R(0)$. One can express $\hat{M}(k)$ and $\hat{\boldsymbol{v}}(k)$ as

$$\hat{M}(k) = M + M(k)$$
$$\hat{\boldsymbol{v}}(k) = \boldsymbol{v} + \boldsymbol{v}(k)$$

where $M(k)$ and $\boldsymbol{v}(k)$ are zero mean, but not independent. However, if $R(\tau) = 0$ for $\tau \neq 0$, then $M(k)$ and $\boldsymbol{v}(k)$ are independent of $M(l)$ and $v(l)$ for $k \neq l$. Equation (25) can now be written as

$$\boldsymbol{p}(k+1) = \boldsymbol{p}(k) - a(k)[M\boldsymbol{p}(k) + \boldsymbol{v}] - a(k)[M(k)\boldsymbol{p}(k) + \boldsymbol{v}(k)].$$

If $\boldsymbol{\varepsilon}(k) = \boldsymbol{p}(k) - \boldsymbol{p}^*$ as in Appendix I, then

$$\boldsymbol{\varepsilon}(k+1) = [I - a(k)M]\boldsymbol{\varepsilon}(k) - a(k)[M(k)\boldsymbol{p}(k) + \boldsymbol{v}(k)]. \tag{26}$$

Let E denote statistical expectation. Then

$$E\boldsymbol{\varepsilon}^T(k+1)\boldsymbol{\varepsilon}(k+1) = E\|\boldsymbol{\varepsilon}(k+1)\|^2 = E\|[I - a(k)M]\boldsymbol{\varepsilon}(k)\|^2 + a^2(k)E\|M(k)\boldsymbol{p}(k) + \boldsymbol{v}(k)\|^2.$$

The expectation of the cross term does not appear. It is zero, since its expectation conditioned on $p(k)$ is zero. Substitute for $\boldsymbol{p}(k)$ to get

$$E\|\boldsymbol{\varepsilon}(k+1)\|^2 = E\|[I - a(k)M]\boldsymbol{\varepsilon}(k)\|^2 + a^2(k)E\|M(k)\boldsymbol{\varepsilon}(k) + M(k)\boldsymbol{p}^* + \boldsymbol{v}(k)\|^2.$$

Assume all elements of $m(k)$, $v(k)$, and $\boldsymbol{p}^*$ are bounded. Thus

$$E\|\boldsymbol{\varepsilon}(k+1)\|^2 \leq (1 - a(k)\mu)^2E\|\boldsymbol{\varepsilon}(k)\|^2 + a^2(k)[B_1E\|\boldsymbol{\varepsilon}(k)\|^2 + B_2]$$

where B_1 and B_2 are finite bounds and μ is the minimum eigenvalue of M. Rearranging terms gives

$$E\|\boldsymbol{\varepsilon}(k+1)\|^2 \leq (1 - 2a(k)\mu + a^2(k)[\mu^2 + B_1])E\|\boldsymbol{\varepsilon}(k)\|^2 + a^2(k)B_2. \quad (27)$$

In the same way that it has been shown in Appendix I that

$$\prod_{i=0}^{k}[1 - a(k)\mu]$$

converges to zero, it can be shown that

$$\prod_{i=0}^{k}(1 - 2a(k)\mu + a^2(k)[\mu^2 + B_1])$$

converges to zero. The proof that $E\|\boldsymbol{\epsilon}(k)\|^2$ converges to zero can now be completed in exactly the same way that Dvoretzky[9] has treated a special case in his work.

Appendix III

Consider the iteration given by (25) of Appendix II. Suppose that $a(k)$ is set equal to a constant a. Define a scalar $V^2(k)$ by

$$V^2(0) = E\|\boldsymbol{\varepsilon}0\|$$

and

$$V^2(k+1) = (1 - 2a\mu + a^2[\mu^2 + B_1])V^2(k) + a^2B_2.$$

Notation is the same as in Appendix II. Assume that a is sufficiently small so that $(1-2a\mu+a^2[\mu^2+B_1])<0$. Using induction, it is clear from (27) that $V^2(k) \geq E\|\boldsymbol{\varepsilon}(k)\|^2$. $V^2(k)$ is a bound on $E\|\boldsymbol{\varepsilon}(k)\|^2$. The difference equation generating $V^2(k)$ is stable with a stationary point given by

$$0 = (-2a\mu + a^2[\mu^2 + B_1])V^2(\infty) + a^2B_2.$$

Thus

$$V^2(\infty) = \frac{a[B_2/2\mu]}{1 - a\dfrac{[\mu^2 + B_1]}{2\mu}} \geq E\|\boldsymbol{\varepsilon}(\infty)\|^2.$$

This is a bound on steady-state mean-square deviation of weights from optimum values.

It is of some interest that

$$\lim_{k\to\infty} E\boldsymbol{\varepsilon}(k) = 0.$$

To see this, it is necessary only to take the expectation of (26) and require that $E\boldsymbol{\varepsilon}(\infty+1)=E\boldsymbol{\varepsilon}(\infty)$. That gives

$$aME\boldsymbol{\varepsilon}(\infty) = 0$$

which implies $\boldsymbol{\varepsilon}(\infty)=0$. This and the preceding result depend upon independent estimates of $R(0)$.

Appendix IV

The sum box method of array processing is most easily explained by a simple example. Suppose three seismometer outputs x_1, x_2, and x_3 are to be combined after delays of 0, 1, and 3 sample points, respectively. Let $b_1, \cdots, b_4$ be words of memory which have the contents

$$b_1(t) = w_1x_1(t) + w_2x_2(t-1) + w_3x_3(t-3)$$
$$b_2(t) = w_2x_2(t) + w_3x_3(t-2)$$
$$b_3(t) = w_3x_3(t-1)$$
$$b_4(t) = w_3x_3(t)$$

at time t. Note that $b_1(t)=y(t)$, the weighted delayed sum of the signals. At time $t+1$, the memory can be modified so that

$$b_1(t+1) = w_3x_3(t+1)$$
$$b_2(t+1) = b_2(t) + w_1x_1(t+1)$$
$$b_3(t+1) = b_3(t) + w_2x_2(t+1)$$
$$b_4(t+1) = b_4(t).$$

Three MULTIPLYs and ADDs have been done and $b_2(t+1)=y(t+1)$. The process can be continued to get $b_3(t+2)=y(t+2)$. The memory contains

$$b_1(t+2) = b_1(t+1)$$
$$b_2(t+2) = w_3x_3(t+2)$$
$$b_3(t+2) = b_3(t+1) + w_1x_1(t+2)$$
$$b_4(t+2) = b_4(t+1) + w_2x_2(t+2)$$

at time $t+2$. Continuing in a similar manner generates array outputs efficiently with the use of only a few memory locations.

The sum box processing method can be generalized to allow for convolutional filtering of each of the input traces before summing. The technique is exactly the same as that described previously by example. The number of memory boxes required is increased by the number of filter points in the convolutional filters.

ACKNOWLEDGMENT

The author wishes to thank Dr. R. J. Greenfield and other members of Group 64 at the M.I.T. Lincoln Laboratory, Lexington, Mass., for their participation in several discussions concerning adaptive processing. He also appreciates the criticisms of Dr. J. Capon and Dr. F. C. Schweppe, who read a draft of this paper.

REFERENCES

[1] Special Issue on Nuclear Test Detection, *Proc. IEEE*, vol. 53, December 1965.
[2] J. Capon, R. J. Greenfield, and R. J. Kolker, "Multidimensional maximum-likelihood processing of a large aperture seismic array," *Proc. IEEE*, vol. 55, pp. 192–211, February 1967.
[3] J. P. Burg, "Three-dimensional filtering with an array of seismometers," *Geophys.*, vol. 29, October 1964.
[4] Special Issue on Active and Adaptive Antennas, *IEEE Trans. Antennas and Propagation*, vol. AP-12, March 1964.
[5] "Travel time anomalies at LASA," Rept. LL-1, prepared for M.I.T. Lincoln Lab., Lexington Mass., by Earth Sciences, a Teledyne Company, November 2, 1966.
[6] A. Gangi, private communication.
[7] S. W. W. Shor, "Adaptive technique to discriminate against coherent noise in a narrow-band system," *J. Acoust. Soc. Amer.*, vol. 39, January 1966.
[8] R. T. Adams, "An adaptive antenna for maximizing signal-to-noise ratio," *WESCON Conv. Rec.*, pt. 5, 1966.
[9] A. Dvoretzky, "On stochastic approximation," *3rd Berkeley Symp. on Math. Statistics and Probability*. Berkeley, Calif.: University of California Press, 1956.
[10] D. J. Sakrison, "Application of stochastic approximation methods to system optimization," M.I.T. Research Lab. of Electronics, Cambridge, Mass., Tech. Rept. 391, July 10, 1962.
[11] J. A. McFadden, "The correlation function of a sine wave plus noise after extreme clipping," *IRE Trans. Information Theory*, vol. IT-2, pp. 82–83, June 1956.
[12] L. Fleck, private communication.
[13] J. Capon, R. J. Greenfield, and R. T. Lacoss, "Off-line signal processing results for the large aperture seismic array," M.I.T. Lincoln Lab., Lexington, Mass., Tech. Note 1966-37, July 11, 1966, DDC 637016.

An Iterative Technique for Determining Inverse Filters

E. ORAN BRIGHAM, HAROLD W. SMITH, SENIOR MEMBER, IEEE, FRANCIS X. BOSTICK, JR. MEMBER, IEEE, AND WILLIAM C. DUESTERHOEFT, JR., SENIOR MEMBER, IEEE

***Abstract*—It is the objective of much geophysical research to increase the resolution of signals recorded on a sluggish measuring device. The approach genreally followed is to use inverse digital filters. This treatment presents an iterative technique for obtaining "stable inverse digital filters." A stable inverse filter is one whose impulse response decays to zero with increasing time.**

An optimum inverse filter $R(\omega)$ is defined here as $R(\omega)=1/H(\omega)$, where $H(\omega)\neq 0$; $R(\omega)=0$, where $H(\omega)=0$. It is shown that one can converge to this solution by operating in the time domain using the method of successive substitution. This approach to inverse filtering is unique in that the inverse filter is obtained by an iterative technique, thereby eliminating the dependence on computer limitations, as indicated in some reported techniques. In addition, a method of handling the zero crossings of $H(\omega)$ is posed. A smoothing technique to modify these filters for inverse filtering in the presence of noise is also presented.

I. INTRODUCTION

THE PROBLEM of the lost resolution resulting from a sluggish measuring device is the subject of much geophysical work. To this aim, considerable attention has been given to the design of high-resolution inverse digital filters.[1]–[10] In most of these treatments, the basis for analysis stems from the fact that the frequency response of the perfect inverse filter is given by $1/H(\omega)$, where $H(\omega)$ is the frequency response of the measuring instrument. The impulse response of the inverse filter is then found by forming the inverse Fourier transform of $1/H(\omega)$. Because problems arise in forming this transformation numerically, $H(\omega)$ may have zero crossings, the majority of the technical literature deals with various approximations used to obtain a "stable inverse filter." Here, a stable filter is one whose impulse response decays to zero with increasing time.

This paper presents an iterative technique for obtaining a stable inverse digital filter. It is shown that one can converge to the solution previously defined by operating in the time domain using the method of successive substitution. This approach to inverse filtering is new from the standpoint that the inverse filter is obtained by an iterative technique. Such an approach eliminates the problem of large-scale matrix inversions, as indicated in some reported techniques. This paper also poses a technique to improve those inverse filters determined for system functions which have zero crossings. Signal restoration in the presence of noise is considered under the assumption that no statistics are known for either the signal or the additive noise. A priori knowledge of the impulse response of the measuring device is

Manuscript received August 7, 1967; revised January 3, 1968. This work is based on a Ph.D. dissertation submitted to the University of Texas, January 1967.

E. O. Brigham was with the Dept. of Elec. Engrg., University of Texas, Austin, Tex. He is now with LTV Electrosystems, Inc., Greenville, Tex.

H. W. Smith, F. X. Bostick, Jr., and W. C. Duesterhoeft, Jr. are with the Dept. of Elec. Engrg., University of Texas, Austin, Tex., and are Consultants to Lane Wells, Inc.