AD-A081 388 SYRACUSE UNIV NY F/G 9/5
RESEARCH ON LARGE ADAPTIVE ARRAYS.(U)
DEC 79 B WIDROW, R CHESTEK, H MESIWALA F30602-78-C-0083
UNCLASSIFIED RADC-TR-79-308 NL

1 OF 2
AD A 081388

(12) LEVEL II

RADC-TR-79-308
Final Technical Report
December 1979

RESEARCH ON LARGE ADAPTIVE ARRAYS

Stanford University

B. Widrow
R. Chestek
H. Mesiwala
K. Duvall

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: RADC-TR-79-308	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): RESEARCH ON LARGE ADAPTIVE ARRAYS	5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report August 78 — June 79
6. PERFORMING ORG. REPORT NUMBER: N/A	
7. AUTHOR(s): B. Widrow, H. Mesiwala, R. Chestek, K. Duvall	8. CONTRACT OR GRANT NUMBER(s): F30602-78-C-0083
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Stanford University, Stanford CA 94305	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F, 2305J8P1
11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (DCCR), Griffiss AFB NY 13441	12. REPORT DATE: December 1979
	13. NUMBER OF PAGES: 100
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: John A. Graniero (DCCR) This work was sponsored in part by the Naval Air Systems Command under Contract N00019-78-C-0276.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Adaptive Antennas
Adaptive Arrays
Adaptive Filters
LMS Algorithm
Null Steering

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Most adaptive filters are designed to achieve minimization of estimation error. Other design goals could include satisfaction of constraints as well as minimization of mean square error. The paper proposes a particular set of performance criteria, involving "soft constraints". A "soft-constraint LMS algorithm" is derived from a performance function which is the sum of mean square error and weighted squared constraint violation errors. The soft constraint algorithm is applied to adaptive antenna arrays as an example, which includes a demonstration of the effects of varying the softness of the (cont'd on reverse)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A079604

constraints. Also, convergence properties of the algorithm are presented. A relation between the output power of a signal from a converged soft constraint LMS adaptive filter and the signal's input power is derived, which demonstrates some unexpected behavior.

TABLE OF CONTENTS

LIST OF FIGURES

Figures (Continued)

LIST OF TABLES

EVALUATION

This effort is a part of Research Area 7, Electronics, Sub-Area 3 Communications. The objective is to define and assess adaptive nulling algorithms compatible with adaptive antennas with large numbers of antenna elements or weights. This research work supports RADC TPO 4A, C^3 Survivability, Thrust - Communications ECCM. The overall objective is to advance the state-of-the-art in adaptive array antennas to provide an Electronic Counter Countermeasure (ECCM) capability for Air Force Communications Systems.

In this research effort, a new adaptive nulling algorithm was formulated and modeled by computer simulation. The algorithm is based on a newly developed generalized performance function that allows specification of the directivity pattern of the antenna. The new performance function permits constraints on the antenna gain in desired directions while minimizing interfering signals. The strength of the constraint can be varied such that deviations from it can be controlled, i.e., important locations can be stiffened so that deviation from it remain small while constraints in less important directions can be made softer so that larger vairations are permitted. This algorithm is considered to be a significant advance over the conventional least mean square error (LMS) algorithm, allowing use of excess degrees of freedom to specify the antenna pattern. It also will allow use of direction-of-arrival (DOA) desired signal discriminants in adaptive arrays in applications where DOA information is not sufficiently accurate for conventional algorithms. The next step should be to investigate hardware implementation of the algorithm for specific communications applications.

John A. Graniero

JOHN A. GRANIERO
Project Engineer

1. INTRODUCTION

In the field of linear estimation, a common goal for the optimum filter is to minimize the mean square error. The Widrow-Hoff Least Mean Square (LMS) algorithm [1,2] is a well known algorithm used with adaptive filters to approach the optimum filter. Subsequent researchers have proposed various modifications to the LMS algorithm. These modifications have been introduced when the adaptive filter does not perform satisfactorily with respect to other criteria in which the researcher is interested. Once modifications are introduced to the algorithm, the adaptive filter is no longer trying to minimize the mean square error, but is instead optimizing some other (often unstated) performance function.

This paper proposes the explicit addition of terms to the performance function reflecting the designer's additional criteria. A specific modification is studied: the addition of "soft" constraints. With a soft constraint, some constraint error results because the weights do not exactly solve a specified set of linear equalities. The optimum filter tries to minimize this constraint error simultaneously with the error incurred by not performing perfect least mean square estimation. The "soft-constraint LMS algorithm", closely related to the LMS algorithm, is derived which causes the adaptive transversal filter to approach the optimum filter. Convergence properties of this algorithm are studied. A relation with unexpected properties between the output power of a signal from the optimum filter and the signal's input power is derived. An application in the area of adaptive antenna arrays is presented as an example of a use of the proposed performance

function and the corresponding adaptation algorithm. The relationship between the soft-constraint LMS algorithm and other versions of the LMS algorithm is discussed.

II. PREVIOUS WORK IN MODIFIED LMS ALGORITHMS

Adaptive filters using the LMS algorithm have been proposed for many applications [3-7]. However, in some situations it has been necessary or desirable to modify the algorithm [8-12]. Frost [9] proposed forcing the weights to exactly satisfy a set of linear equalities, which are called here a set of "hard constraints." This modification of the LMS algorithm has been applied to adaptive antenna arrays, to force the gain of the array to be exactly unity in a specified direction, while attenuating signals arriving from other directions.

Another modification of the LMS algorithm is the "leaky" LMS algorithm. This algorithm has a leak factor, so that in the absence of inputs the weights decay to zero. This form has been proposed independently by several researchers [11-14]. Using the property that the leak is equivalent to introducing a white noise in the input of the filter, Treichler [11] proposed using the algorithm to modify the characteristics of an adaptive line enhancer in a desirable manner. Ahmed et al [12] used the leak effect to reduce numerical instabilities occurring in their application. White [13] showed that the leak could reduce inaccuracies caused by imperfect hardware multipliers.

Zahm [14] used the leaky LMS algorithm with adaptive antenna arrays to suppress strong "jammers" in the presence of weaker signals. However, using the leaky LMS algorithm alone resulted in the undesirable characteristic that the array rejected all signals (and jammers) after a period of time. To counteract this effect, Zahm introduced a set of "steering" weights into the algorithm, so that the weights of the adaptive array

converge to the steering weights in the absence of any jammers or signals. These steering weights prevent the adaptive antenna array from turning itself off. Also, the steering weights Zahm chose as an example introduced desirable effects in the directivity pattern of the array.

Extending Zahm's work and the work on the leaky LMS algorithm results in the modification to the LMS algorithm discussed here.

III. DEFINITIONS AND TERMINOLOGY FOR THE ADAPTIVE TRANSVERSAL FILTER

Although applicable to any linear combiner, this work assumes for ease of discussion that the performance function and the soft-constraint LMS algorithm developed later are used with an adaptive transversal filter, as illustrated in Figure 3-1. Definitions and terminology for the adaptive filter follow.

A sampled time sequence u(j) is the input to an n-1 delay transversal filter, where j is the time index of samples taken. The n weights $w_0, w_1, \ldots, w_{n-1}$ can be adjusted by the adaptation algorithm as time progresses. The filter output y(j) is compared against a time sequence d(j), which is called the desired signal. (The source and nature of the desired signal varies with the application.) The purpose of tne filter is to provide an estimate y(j) of the desired signal d(j). The difference between d(j) and y(j) is called the error signal e(j).

The input sequence u(j) may contain one or all of three types of signals. A signal may be noise; it may be a deliberately produced sequence but of no use in forming the estimate (an interferer or jammer); or it may be a sequence relevant to estimating d(j).

The values at the taps of the transversal filter at time j are denoted by the data vector X(j):

$$X(j) \triangleq [u(j)\ u(j-1)\ \ldots\ u(j-n+1)]^T \quad . \tag{3-1}$$

The set of n weights is written in vector form as:

$$W \triangleq [w_0\ w_1\ \ldots\ w_{n-1}]^T \quad . \tag{3-2}$$

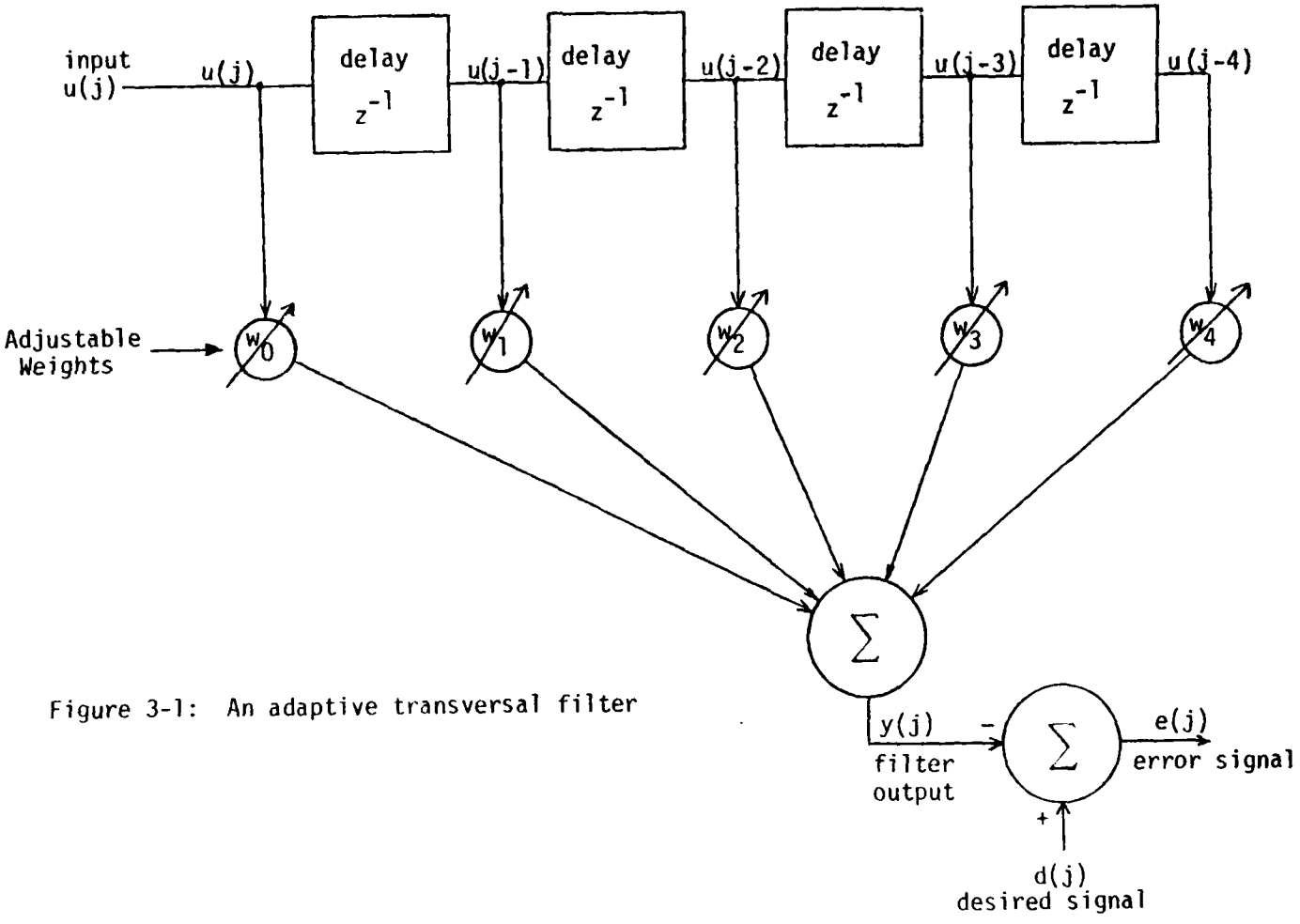


Figure 3-1: An adaptive transversal filter

Then the filter output y(j), the estimate of d(j), is expressed as the inner product of two vectors:

$$y(j) = X^T(j)W = W^TX(j) \quad . \tag{3-3}$$

The error signal is simply

$$e(j) = d(j) - y(j) \quad . \tag{3-4}$$

IV. THE PERFORMANCE FUNCTION

Any function measuring the performance of an adaptive filter must reflect the concerns of the filter designer. The basic consideration is often simply that estimation is being performed. The performance of an adaptive filter in estimating an unknown signal is often measured by the mean square error, a widely used criterion [15-17].

But the designer may recognize additional considerations in some applications. Using an adaptive antenna array as an example, it is sometimes desirable to specify the array gain in a particular direction, provided this requirement does not increase the estimation error (mean square error) excessively. But if the estimation error does increase too much, it may be possible to decrease it significantly while still staying close to (though not exactly meeting) the gain specification.

The array example shows that the performance function must do two things. First, it must measure the extent to which the array gain specification is violated by the chosen filter, as well as measure the estimation error. Second, the performance function must weight the relative contributions of the estimation error and of the gain specification error, so that a balance may be struck between the two sources of error.

A performance function p(j) satisfying these two considerations is:

$$p(j) \triangleq E\{e^2(j)\} + \sum_{i=1}^{m} b_i (e_{ci})^2 \quad , \tag{4-1}$$

where E{a} denotes the expected value of a. The first term, $E\{e^2(j)\}$,

is the mean square error (the estimation error); the second term is the weighted contribution of the specification ("constraint") errors. Definition (4-1) assumes that the designer specifies m constraints, and that the error resulting from not exactly meeting constraint i is e_{ci}. The designer selects the non-negative constant b_i to specify the relative importance of the i^{th} constraint error compared to the estimation error. The greater b_i is, the more the error e_{ci} affects the value of the performance function.

The relation between the constraint errors e_{ci} and the adaptive filter's weights is still unspecified. The filter designer is free to choose any function. Different selections will produce different adaptation algorithms. The form for constraint error e_{ci} studied in this paper is a linear function of the weight vector.

As an example, return to the adaptive antenna array. Let the i^{th} constraint specify the desired gain in a particular direction at a specified frequency. The actual gain of the array in this direction (at the specified frequency) is calculated by a linear expression:

$$\text{gain} = A_i^T W \quad , \tag{4-2}$$

where A_i is a constant vector with n components. (Section X contains details for constructing A_i.) Thus, if the desired gain is the scalar h_i, the constraint error is:

$$e_{ci} \triangleq A_i^T W - h_i \quad . \tag{4-3}$$

Using this form for the constraint error in (4-1) results in the performance function studied here:

$$p(j) = E\{e^2(j)\} + \sum_{i=1}^{m} b_i (A_i^T W - h_i)^2 \quad . \tag{4-4}$$

This performance function is written in matrix form as:

$$p(j) = E\{e^2(j)\} + (\underline{A}W - H)^T \underline{B}(\underline{A}W - H) \quad , \tag{4-5}$$

where $\underline{A}$ is the mxn matrix composed of the vectors A_i:

$$\underline{A} \triangleq [A_1 \; A_2 \; \ldots \; A_i \; \ldots \; A_m]^T \quad ; \tag{4-6}$$

B is the mxm diagonal matrix with diagonal elements b_i:

$$\underline{B} \triangleq \text{diag}[b_1, b_2, \ldots, b_i, \ldots, b_m] \quad ; \tag{4-7}$$

and H is an m dimensional vector composed of the individual desired constraint values h_i:

$$H \triangleq [h_1 \; h_2 \; \ldots \; h_i \; \ldots \; h_m]^T \quad . \tag{4-8}$$

This performance function (4-5) will be called a "soft-constraint least mean square error performance criterion." The constraints are called soft because, unlike constraints in most optimization problems, they can be violated (not satisfied exactly).

The goal of the adaptation algorithm which is developed in section VIII is to find the weight vector that minimizes the performance function p(j) in (4-5).

The dependence of p(j) on the weight vector W is important. The absence of non-global minima is desired, since this absence helps prevent an adaptation algorithm from settling to an incorrect weight vector (i.e. finding a local optimum). The dependence of p(j) on W is

obtained by expanding the mean square error term in (4-5) and using (3-3):

$$\begin{aligned} E\{e^2(j)\} &= E\{[d(j)-y(j)]^2\} \\ &= E\{d^2(j) - 2d(j)X^T(j)W + W^TX(j)X^T(j)W\} \\ &= E\{d^2(j)\} - 2P^T(j)W + W^T\underline{R}(j)W \quad , \end{aligned} \tag{4-9}$$

where the cross correlation between the data vector X(j) and the desired signal d(j) is denoted by P(j):

$$P(j) \triangleq E\{d(j)X(j)\} \quad ; \tag{4-10}$$

and $\underline{R}(j)$ denotes the autocorrelation matrix of the data vector:

$$\underline{R}(j) \triangleq E\{X(j)X^T(j)\} \quad . \tag{4-11}$$

Substituting (4-9) into (4-5) expresses the performance function directly in terms of the weight vector W:

$$p(j) = E\{d^2(j)\} - 2P^T(j)W + W^T\underline{R}(j)W + (\underline{A}W - H)^T\underline{B}(\underline{A}W - H) \quad . \tag{4-12}$$

Clearly, p(j) is a quadratic function of the weights. Because it is a sum of squared quantities, it cannot be negative. Thus one of two situations exists. The first possibility is that there is exactly one minimum to the performance function, and only one weight vector achieves this minimum. This situation may be visualized as a parabolic bowl in a hyperspace of dimension n. The second possibility is that the performance function attains the same minimum value for a whole set of weight vectors. In this case the set of weight vectors forms a connected space so that all minima of the performance function are adjacent to one

another; there are no isolated minima. This situation may be visualized in an n-dimensional hyperspace as a trough, equally deep at all points along the bottom of the trough, and with parabolic sides to the trough.

This paper's primary interest is on the first case, where the weight vector yielding the optimum (minimum) value of the performance function is unique. The analysis can, if desired, be extended to the second situation, by essentially considering a smaller hyperspace which contains a unique minimum.

V. THE OPTIMUM FILTER

This section derives an expression for the optimum weight vector, defined here as the unique weight vector specifying the filter which has the optimum (minimum) performance p(j). The condition under which the minimum p(j) occurs with a non-unique weight vector is also determined.

Any weight vector W minimizing the performance function p(j) forces the gradient of p(j) to zero. From (4-5), the overall gradient of p(j) with respect to W is:

$$\nabla_W p(j) = \nabla_W E\{e^2(j)\} + \nabla_W[(\underline{A}W - H)^T \underline{B}(\underline{A}W - H)] \quad . \tag{5-1}$$

Analyzing the first term by taking the gradient of (4-9) yields:

$$\nabla_W E\{e^2(j)\} = -2[P(j) - \underline{R}(j)W]^T \quad . \tag{5-2}$$

This first term of the overall gradient comes from the mean square error (estimation error) term of the performance criterion. This is the same gradient used to develop the LMS algorithm.

Analyzing the second term of (5-1) yields:

$$\nabla_W[(\underline{A}W - H)^T \underline{B}(\underline{A}W - H)] = 2[\underline{A}^T \underline{B}(\underline{A}W - H)]^T \quad . \tag{5-3}$$

This second term is due entirely to the soft constraints imposed by the filter designer.

From (5-1) the overall gradient is the sum of (5-2) and (5-3):

$$\nabla_W p(j) = -2[P(j) - \underline{R}(j)W]^T + 2[\underline{A}^T \underline{B}(\underline{A}W - H)]^T \quad . \tag{5-4}$$

The optimum value for W occurs when the gradient (5-4) is set equal

to zero, yielding:

$$[\underline{R}(j) + \underline{A}^T\underline{B}\underline{A}]W = P(j) + \underline{A}^T\underline{B}H \ . \tag{5-5}$$

Thus the necessary condition for the optimum (minimum) performance to occur at a unique weight vector is that the matrix $\underline{R}(j)+\underline{A}^T\underline{B}\underline{A}$ be nonsingular. Under this condition, the unique optimum weight vector, denoted $W_{opt}(j)$, is:

$$W_{opt}(j) = [\underline{R}(j)+\underline{A}^T\underline{B}\underline{A}]^{-1}[P(j)+\underline{A}^T\underline{B}H] \ . \tag{5-6}$$

Note that it is not necessary for either $\underline{R}(j)$ or $\underline{A}^T\underline{B}\underline{A}$ to be nonsingular. In fact, one use for the soft-constraint LMS algorithm arises when the data vector autocorrelation matrix $\underline{R}(j)$ is indeed singular (or possibly just ill-conditioned). In such a case a set of soft constraints can be generated to yield a unique optimum weight vector, as has been done with the leaky LMS algorithm [12].

VI. AN ASSUMPTION OF STATIONARITY

The remainder of this work assumes that the signals d(j) and u(j) re generated by stationary stochastic processes. Thus, the statistics (j) and $\underline{R}(j)$, the performance criterion p(j), and the optimum weight ector $W_{opt}(j)$ are constant, and are now denoted by P, $\underline{R}$, p, and W_{opt} espectively, dropping the time index j.

Using this assumption of stationarity, the performance function is now written from (4-5) as:

$$p = E\{e^2(j)\} + (\underline{A}W - H)^T\underline{B}(\underline{A}W - H) , \tag{6-1}$$

the gradient of p from (5-4) is:

$$\nabla_W p = -2[P - \underline{R}W]^T + 2[\underline{A}^T\underline{B}(\underline{A}W - H)]^T , \tag{6-2}$$

and the optimum weight vector W_{opt} is written from (5-6) as:

$$W_{opt} = (\underline{R}+\underline{A}^T\underline{B}\underline{A})^{-1}(P+\underline{A}^T\underline{B}H) . \tag{6-3}$$

VII. DETERMINATION OF THE OPTIMUM WEIGHT VECTOR BY GRADIENT SEARCH

Calculating the optimum weight vector using (6-3) is not always feasible, even when all quantities are known. This may be due to the size of the filter, or to numerical difficulties caused by properties of the matrices. Thus alternative approaches have been devised. A common technique is to make successive approximations to the optimum weight vector. Given one estimate of the weight vector, denoted by W(j), the next estimate, W(j+1), is iteratively generated from W(j), governed by how well W(j) satisfies (5-5). The time index j denotes sequential estimates since it is assumed here that one update occurs at each time instant.

The technique of successive approximation used in this research is called gradient search [18]. The gradient of the performance surface is calculated for the current value of the weight vector, W(j). The gradient specifies the direction of weight vector change which will increase the performance function p most rapidly; but since the goal is to reduce the performance function, the next estimate of the optimum weight vector is obtained by moving from the current estimate in the direction opposite to that of the gradient, through a distance proportional to the magnitude of the gradient:

$$W(j+1) = W(j) - \mu \nabla_W^T p \ , \qquad (7\text{-}1)$$

where μ is a positive constant chosen by the filter designer.

Using (6-2) for $\nabla_W p$ in (7-1) gives the update equation:

$$W(j+1) = W(j) + 2\mu[P - \underline{R}W(j)] - 2\mu \underline{A}^T \underline{B}[\underline{A}W(j) - H] \ . \qquad (7\text{-}2)$$

Repeatedly using this update equation causes the estimate of the optimum weight vector to approach the actual optimum W_{opt} of (6-3), provided μ is small enough (see section IX).

VIII. THE SOFT-CONSTRAINT LMS ALGORITHM

The algorithm (7-2) for approaching the optimum weight vector W_{opt} is applicable only if all quantities are known. Such knowledge is generally not available in practice. If the statistics of the input signal u(j) are unknown, then P and $\underline{R}$ are unknown. This can occur when a known signal is subject to additive noise, is passed through a filter whose characteristics are not perfectly known, or is distorted. Nevertheless, it is still possible to perform signal estimation subject to soft constraints. To do this, the update equation (7-2) is modified by replacing P and $\underline{R}$ with estimates. The estimates chosen must depend upon the input signal u(j), so that these estimates are based on data statistics, rather than on a priori guesses. The estimates chosen are:

$$\begin{aligned} \hat{P} &= d(j)X(j) \ , \\ \hat{\underline{R}} &= X(j)X^T(j) \ . \end{aligned} \tag{8-1}$$

It is easily shown that these estimates are unbiased:

$$\begin{aligned} E\{\hat{P}\} &= E\{d(j)X(j)\} = P \ , \\ E\{\hat{\underline{R}}\} &= E\{X(j)X^T(j)\} = \underline{R} \ . \end{aligned} \tag{8-2}$$

Thus the gradient search algorithm (7-1) is replaced by:

$$W(j+1) = W(j) - \mu\hat{\nabla}_W^T p \ , \tag{8-3}$$

where an estimate of the gradient is used in place of the true gradient. This means that the estimates $\hat{P}$ and $\hat{\underline{R}}$ replace the true values of P and $\underline{R}$

in (6-2), yielding:

$$\hat{\nabla}_W p = -2[\hat{P}-\hat{R}W(j)]^T + 2\{\underline{A}^T\underline{B}[\underline{A}W(j)-H]\}^T \quad . \qquad (8\text{-}4)$$

Substituting (8-1) into (8-4) results in:

$$\hat{\nabla}_W p = -2[d(j)X(j)-X(j)X^T(j)W(j)]^T + 2\{\underline{A}^T\underline{B}[\underline{A}W(j)-H]\}^T \quad . \qquad (8\text{-}5)$$

Using definitions (3-3) and (3-4) in (8-5) and rearranging yields:

$$\hat{\nabla}_W p = -2e(j)X^T(j) + 2\{\underline{A}^T\underline{B}[\underline{A}W(j)-H]\}^T \quad . \qquad (8\text{-}6)$$

Substituting (8-6) into the update equation (8-3) results in:

$$W(j+1) = W(j) + \underbrace{2\mu e(j)X(j)}_{\text{gradient due to estimation error}} - \underbrace{2\mu\underline{A}^T\underline{B}[\underline{A}W(j)-H]}_{\text{gradient due to constraint errors}} , \qquad (8\text{-}7)$$

or;

$$W(j+1) = (\underline{I}-2\mu\underline{A}^T\underline{B}\underline{A})W(j) + 2\mu e(j)X(j) + 2\mu\underline{A}^T\underline{B}H \qquad (8\text{-}8)$$

Equations (8-7) and (8-8) are alternate forms of what is defined here as the "soft-constraint LMS algorithm."

IX. STATISTICAL PROPERTIES OF THE SOFT-CONSTRAINT LMS ALGORITHM WEIGHT VECTOR

Random Noise in the Weight Vector

If the steepest descent update algorithm in (7-2) is used to adapt the weight vector, the resulting weight vector sequence depends only on P, $\underline{R}$, and the weight vector's initial value. In this ideal case, P and $\underline{R}$ are known *a priori*. Hence, a specific sequence is generated, regardless of which ensemble member of the stochastic processes generating u(j) and d(j) occurs. This is not true, however, with the soft constraint LMS algorithm (8-7) or (8-8). Although the gradient term in (8-4) due to the soft constraints is calculated perfectly from the designer's specification of the soft constraints and knowledge of the current weight vector, the term due to the mean square error is only an estimate, since $\hat{P}$ and $\hat{\underline{R}}$ are estimates of P and $\underline{R}$. The estimate chosen results in a random quantity, since it depends on the actual sequences u(j) and d(j). This results in an ensemble of weight vector sequences. The ensemble can be pictured as arising from a bank of adaptive filters, all beginning with the same initial weight vector, but all receiving different ensemble members for u(j) and d(j). This section discusses the statistical properties of this ensemble of weight vectors.

Convergence of the Mean Weight Vector

Theorem 1: Convergence of the Mean Weight Vector

If 1) The soft-constraint LMS algorithm (8-7) or (8-8) produces a weight vector sequence $W(j)$ from a data vector sequence $X(j)$ and a desired signal sequence $d(j)$, and if

2) $W(j)$ and $X(j)$ are statistically independent, and if

3) The matrix $\underline{R}+\underline{A}^T\underline{B}\underline{A}$ is nonsingular, and if

4) $0 < \mu < \frac{1}{\lambda_{max}}$,

where $\lambda_{max} \triangleq \max\{\text{eig}\{\underline{R}+\underline{A}^T\underline{B}\underline{A}\}\}$,

$\text{eig}\{\underline{Y}\}$ is the set of eigenvalues of matrix $\underline{Y}$,

and $\max\{\text{a set}\}$ is the maximum value of the set,

Then in the limit the mean weight vector converges to the optimum weight vector:

$$\lim_{j\to\infty} E\{W(j)\} = W_{opt} , \tag{9-1}$$

where $E\{W(j)\}$ is the expected value (mean) taken over the ensemble of weight vectors at time j.

The proof of this theorem is contained in Appendix A.

The definition of convergence used in Theorem 1 is weaker than that used with stochastic approximation methods [19]. The latter require that in addition to the mean weight vector converging to the optimum value as given in (9-1), the weight vector's covariance must go to zero; meaning that every member of the ensemble of weight vector sequences must approach the optimum. However, stochastic approximation methods suffer from the disadvantage that if the signal statistics vary slowly (are not strictly stationary), the weight vector cannot track the time-

varying optimum value. By contrast, the soft-constraint LMS algorithm can follow a slowly moving optimum weight vector; the exact characteristics in this environment are a subject for further study.

The second condition of Theorem 1, that for convergence W(j) and X(j) must be statistically independent, is not met when the soft-constraint LMS algorithm is applied to a transversal filter (Figure 3-1). Due to the nature of the algorithm, W(j) is a function of all past data vectors up to X(j-1). And because of the operation of a tapped delay line, X(j) is a vector consisting of exactly n-1 elements of the vector X(j-1). Thus, since W(j) and X(j) are both functions of X(j-1), they cannot be statistically independent of each other. However, when the adaptation constant μ is small, W(j) depends only weakly on X(j-1); hence the cross-correlation between W(j) and X(j) is small, yielding a close approximation to the assumption of independence. The effect of violating this assumption has been studied for the LMS algorithm [20-22], with the conclusion that the weight vector mean converges to a value which is biased away from the optimum weight vector; but a value as close as desired to the optimum can be attained by making μ small. It is expected that the same behavior can be proved for the soft-constraint LMS algorithm, due to the close similarity to the standard LMS algorithm. Experience with the algorithm supports this expectation.

Note that the maximum value for μ permitting convergence of the mean weight vector (condition 4 of Theorem 1) depends on λ_{max}, which is unknown when $\underline{R}$ is unknown *a priori*. However, an upper bound on λ_{max} which is easy to compute in an actual problem is:

$$\lambda_{max} = \max\{\text{eig}\{\underline{R}+\underline{A}^T\underline{B}\underline{A}\}\}$$

$$\leq \max\{\text{eig}\{\underline{R}\}\} + \max\{\text{eig}\{\underline{A}^T\underline{B}\underline{A}\}\} \; . \qquad (9\text{-}2)$$

This is true since all eigenvalues of $\underline{R}$ and $\underline{A}^T\underline{B}\underline{A}$ are non-negative, so the maximum eigenvalue of $(\underline{R}+\underline{A}^T\underline{B}\underline{A})$ cannot be larger than the sum of the maximum eigenvalues of $\underline{R}$ and $\underline{A}^T\underline{B}\underline{A}$. Now the maximum eigenvalue of $\underline{A}^T\underline{B}\underline{A}$ is available, since these matrices are predetermined by the filter designer. And since the trace of an autocorrelation matrix is the sum of its (non-negative) eigenvalues, the maximum eigenvalue of $\underline{R}$ must be less than or equal to the trace of $\underline{R}$, which is just n times the input power $E\{u^2(j)\}$ to the filter. Thus

$$\lambda_{max} \leq \text{Tr}[\underline{R}] + \max\{\text{eig}\{\underline{A}^T\underline{B}\underline{A}\}\}$$

$$\leq n[E\{u^2(j)\}] + \max\{\text{eig}\{\underline{A}^T\underline{B}\underline{A}\}\} \; ; \qquad (9\text{-}3)$$

so that a sufficient condition on μ to satisfy assumption 4 of Theorem 1 is:

$$0 < \mu < \frac{1}{n[E\{u^2(j)\}] + \max\{\text{eig}\{\underline{A}^T\underline{B}\underline{A}\}\}} \; , \qquad (9\text{-}4)$$

which can be calculated without *a priori* knowledge of $\underline{R}$. It is generally more restrictive than the bound of Theorem 1.

The proof of Theorem 1 in Appendix A points out the interesting fact that the mean weight vector follows exactly the same trajectory that the weight vector would follow if perfect gradient measurements were available. Thus the approximation to the gradient (inclusion of "gradient noise") does not change the convergence rate of the mean

weight vector.

Weight Vector Covariance

The weight vector covariance measures how much individual members of the ensemble of weight vector sequences vary from the mean weight vector. Since the mean weight vector converges in the limit to the optimum weight vector, the greater the weight vector covariance is, the further individual members of the ensemble of weight vector sequences are from the optimum in the limit. This variation implies poor performance.

The weight vector covariance matrix $\underline{C}_{WW}(j)$ is defined by:

$$\underline{C}_{WW}(j) \triangleq E\{[W(j)-\overline{W}(j)][W(j)-\overline{W}(j)]^T\}$$

$$= E\{W(j)W^T(j)\} - \overline{W}(j)\overline{W}^T(j) \ , \qquad (9\text{-}5)$$

where $E\{W(j)\}$, the mean weight vector, is written as $\overline{W}(j)$ to simplify notation.

Theorem 2: Weight Vector Covariance

If 1) Theorem 1 holds, and if

2) W(j) and d(j) are statistically independent, and if

3) d(j) and u(j) are gaussianly distributed,

then the recursion equation for the weight vector covariance is:

$$\underline{C}_{WW}(j+1) = [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\underline{C}_{WW}(j)[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]$$

$$+ 4\mu^2\Big\{\underline{R}\underline{C}_{WW}(j)\underline{R} + \underline{R}\mathrm{Tr}[\underline{C}_{WW}(j)\underline{R}] + \underline{R}E\{e^2(j)|_{W=\overline{W}(j)}\}$$

$$+ [P-\underline{R}\overline{W}(j)][P-\underline{R}\overline{W}(j)]^T\Big\} \qquad (9\text{-}6)$$

The proof of this theorem is contained in Appendix B.

It is important to know when the weight vector covariance matrix remains bounded, since on occasion the mean weight vector will converge to the optimum value, while the weight vector covariance matrix grows without bound. This means that the individual members of the weight vector sequence ensemble vary around the proper solution, but the variations grow larger and larger. Such a situation is undesirable. The following analysis finds conditions where the weight vector covariance matrix is guaranteed to remain finite, and finds other conditions where the weight vector covariance matrix is guaranteed to grow without bound. The behavior of the weight vector covariance matrix is undetermined for the remaining cases.

The trace of the weight vector covariance matrix measures its "magnitude". All off-diagonal terms of a covariance matrix are less than or equal in magnitude to the largest of the diagonal terms[†]; and the diagonal terms are all positive; so the trace upper bounds the magnitude of every element of the covariance matrix. Applying the trace (a linear operator) to each term of (9-6) yields the recursion equation for the trace of the weight vector covariance matrix:

$$\begin{aligned} \mathrm{Tr}[\underline{C}_{WW}(j+1)] = \mathrm{Tr}\{&[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\underline{C}_{WW}(j)[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\} \\ &+ 4\mu^2\Big\{\mathrm{Tr}[\underline{R}\underline{C}_{WW}(j)\underline{R}] + \mathrm{Tr}[\underline{R}]\mathrm{Tr}[\underline{C}_{WW}(j)\underline{R}] \\ &\quad + \mathrm{Tr}[\underline{R}]E\{e^2(j)|_{W=\overline{W}(j)}\} \\ &\quad + [P-\underline{R}\overline{W}(j)]^T[P-\underline{R}\overline{W}(j)]\Big\} . \end{aligned} \tag{9-7}$$

[†]This results from applying Schwarz's Inequality to the autocorrelation function of a stationary process.

Theorem 3: Sufficient conditions for boundedness of the trace of the weight vector covariance matrix

1) If Theorem 2 holds, and

2) if a) $\gamma_{max}^2 + \gamma_{max} Tr[\underline{R}] \leq \lambda_{max}\lambda_{min}$

and

$$0 \leq \mu < \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{max}^2 + \gamma_{max} Tr[\underline{R}]} ;$$

or if b) $\gamma_{max}^2 + \gamma_{max} Tr[\underline{R}] \geq \lambda_{max}\lambda_{min}$

and

$$0 \leq \mu < \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{max}^2 + \gamma_{max} Tr[\underline{R}]} ;$$

where

$$\gamma_{max} \triangleq \max\{eig\{\underline{R}\}\}$$

$$\lambda_{max} \triangleq \max\{eig\{\underline{R}+\underline{A}^T\underline{B}\underline{A}\}\}$$

$$\lambda_{min} \triangleq \min\{eig\{\underline{R}+\underline{A}^T\underline{B}\underline{A}\}\}$$

Then $Tr[\underline{C}_{WW}(j)]$ will be bounded for all time.

The proof of this theorem is contained in Appendix C.

Theorem 3 presents conditions on μ which are sufficient to guarantee that $Tr[\underline{C}_{WW}(j)]$ is a bounded sequence. Next, necessary conditions for $Tr[\underline{C}_{WW}(j)]$ to be a bounded sequence are determined; however, these are not sufficient conditions.

Theorem 4: Necessary conditions for boundedness of the trace of the weight vector covariance matrix

For $Tr[\underline{C}_{WW}(j)]$ to be a bounded sequence, it is necessary that

1) Theorem 2 hold, and also that

2) a) When $\gamma_{min}^2 + \gamma_{min} Tr[\underline{R}] \geq \lambda_{max}^2$,

$$\text{that } 0 \leq \mu < \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{min}^2 + \gamma_{min} Tr[\underline{R}]} ;$$

b) When $\lambda_{min}^2 \leq \gamma_{min}^2 + \gamma_{min} Tr[\underline{R}] \leq \lambda_{max}^2$,

$$\text{that } 0 \leq \mu < \frac{1}{\sqrt{\gamma_{min}^2 + \gamma_{min} Tr[\underline{R}]}} ;$$

c) When $\gamma_{min}^2 + \gamma_{min} Tr[\underline{R}] \leq \lambda_{min}^2$,

$$\text{that } 0 \leq \mu < \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{min}^2 + \gamma_{min} Tr[\underline{R}]} .$$

The proof of this theorem is contained in Appendix D.

Figure 9-1 demonstrates some of the interrelationships among the bounds on μ presented in Theorems 1, 3, and 4. It will be seen that satisfying the bound on μ is not always adequate to obtain good performance. The figure is an example, obtained by plotting the various bounds on μ as a function of the power of a signal, α^2, in a particular environment†.

†Figures 9-1 and 9-2 were obtained by assuming that a six tap filter is being used, receiving a signal of power α^2 and a noise of power 1, with γ_{max} receiving half the input power, and the rest of the power distributed evenly among the remaining eigenvalues, and assuming that the eigenvalues of $\underline{A}^T\underline{B}\underline{A}$ all have a value of 15.

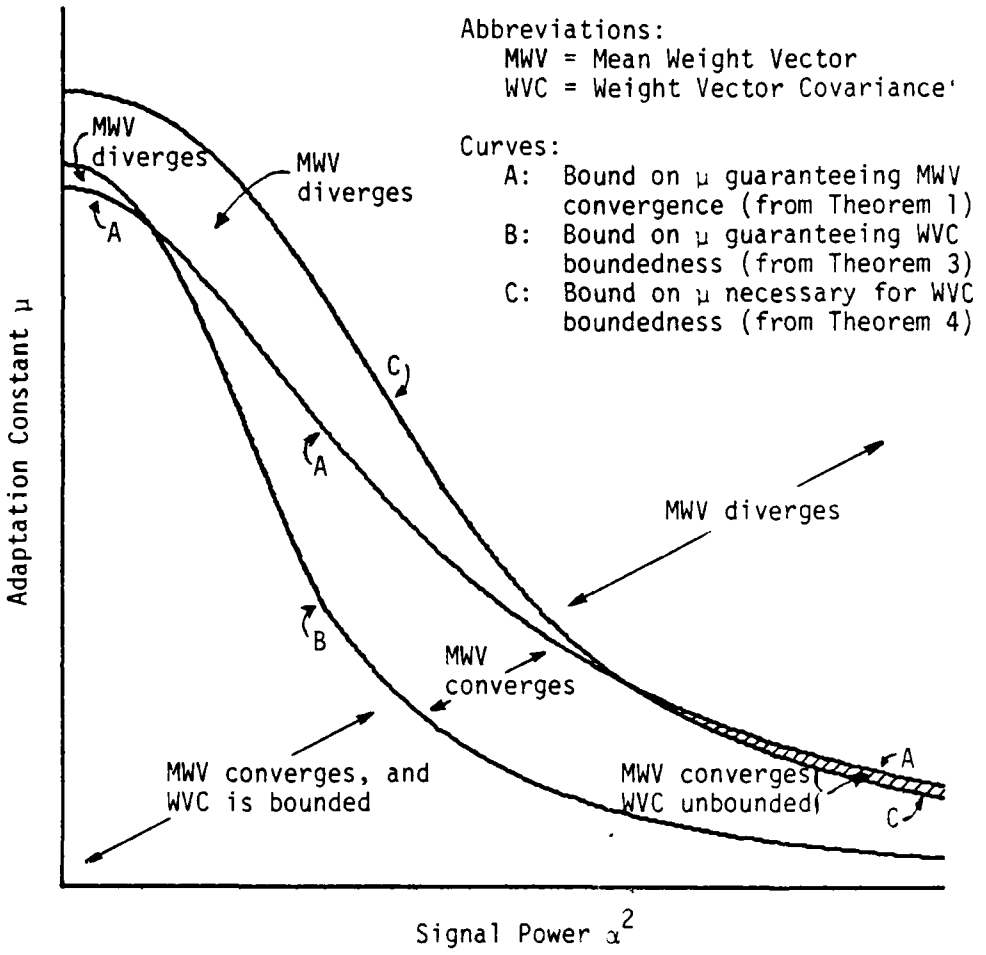


Figure 9-1: Sample relationships of bounds on μ from Theorems 1, 3, and 4 (assuming all other Theorem conditions met)

Assuming all other conditions of Theorems 1 through 4 are met, if μ lies below curve A, then the mean weight vector will converge. If μ lies above A, then the mean weight vector will diverge, and the conditions on Theorems 2, 3, and 4 are not satisfied, so behavior of the weight vector covariance matrix is unknown. If μ lies below both curves A and B, the weight covariance matrix is guaranteed to be bounded. The snaded area at the lower right of Figure 9-1 where curve A lies above curve C is particularly interesting. If μ lies within this shaded region, the mean weight vector converges because μ is below curve A, but the weight vector covariance matrix is guaranteed to grow without bound. This performance is unacceptable even though the mean weight vector converges.

Contrasting with the above is the area at the extreme left of the figure (low signal power) where curve A lies below curve B, the bound on μ which guarantees that the trace of the weight vector covariance matrix remains bounded. In this case, and in this case only, satisfying the bound on μ to guarantee convergence of the mean weight vector also guarantees that the weight vector covariance will remain finite.

There are other areas of the figure where the mean weight vector is guaranteed to converge, but it is unknown if the weight covariance matrix will remain finite or not. Thus Figure 9-1 demonstrates that the bound on μ from Theorem 1 by itself is insufficient to guarantee desired behavior; the bounds from Theorems 3 and 4 must also be considered.

A single upper bound for μ is desired which guarantees that the mean weight vector converges, and also guarantees that the weight

vector covariance matrix remains bounded. A useful bound should be calculable from prior knowledge and input signal power only, and not depend upon knowledge about the data vector autocorrelation matrix $\underline{R}$. It can be shown[†] that the bound in (9-4), which meets the criteria of calculability and guarantees that the mean weight vector converges, does not guarantee boundedness of the weight vector covariance matrix. However, a bound satisfying these conditions is:

$$0 < \mu < \frac{1}{3\mathrm{Tr}[\underline{R}] + \mathrm{Tr}[\underline{A}^T\underline{B}\underline{A}]} . \qquad (9\text{-}8)$$

Appendix E demonstrates that this bound satisfies the criteria listed above.

The bound (9-8) is plotted as curve D in Fig. 9-2, along with the bounds for convergence of the mean weight vector and the bound guaranteeing boundedness of the weight vector covariance matrix. This figure confirms that the bound (9-8) lies below the other bounds. It also demonstrates that the bound (9-8) can be overly restrictive, since it lies so far below curves A and B. The distance between the bound (9-8) and the curves A and B depends partially on how closely the traces of the matrices are related to the maximum eigenvalues; the closer the trace is to the maximum eigenvalue, the closer (9-8) will be to curves A and B.

[†]For example, consider the scalar case with $\gamma_{max}=\gamma_{min}=1$, $\mathrm{Tr}[\underline{R}]=1$, $\max\{\mathrm{eig}\{A^TBA\}\}=1$ (which implies $\lambda_{max}=2$).

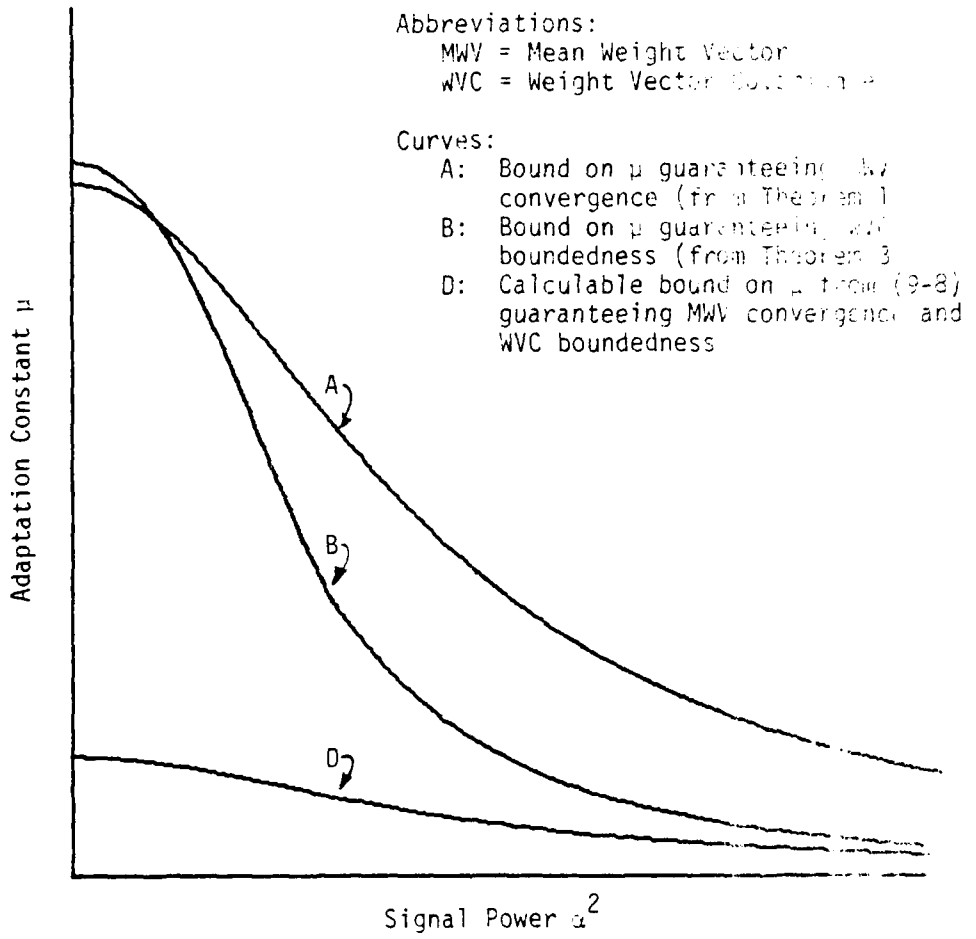


Figure 9-2: A calculable bound on μ guaranteeing mean weight vector convergence and weight vector covariance boundedness (assuming all other Theorem conditions met)

X. AN APPLICATION TO ADAPTIVE ANTENNA ARRAYS

This section demonstrates an application of the soft-constraint LMS algorithm to adaptive antenna arrays. The soft constraints are used to affect the shape of the antenna array's directivity pattern.

Fig. 10-1 shows the adaptive antenna array system. Each of the six antenna elements is omnidirectional. The output of each antenna element, s_1 through s_6, is fed to its own two-tap (and two-weight) adaptive transversal filter (TF_1 through TF_6). The summed outputs of the filters form the system output $y(j)$. It is assumed that no desired signal $d(j)$ is available, so the error signal $e(j)$ is the array's output $y(j)$. The objective of the algorithm with this system is to minimize the output power; simultaneously trying to keep the array's gain in certain directions at certain frequencies close to values specified by the designer.

The weight vector of the antenna array system is constructed by stacking the six weight vectors of the individual adaptive filters. Denote the weight vector of transversal filter k at time j by the two dimensional vector $W_k(j)$; the weight vector $W(j)$ of the entire system is then a twelve dimensional vector:

$$W(j) = [W_1^T(j)\ W_2^T(j)\ \cdots\ W_k^T(j)\ \cdots\ W_6^T(j)]. \qquad (10\text{-}1)$$

Construct the data vector $X(j)$ for the entire system similarly.

The soft constraints will be used to specify desirable antenna gains in a particular direction at a specified frequency.

Imagine the antenna array receiving a sinusoid of power C_r^2 at frequency ω_r from direction θ_r. Denote the sinusoid at the input to

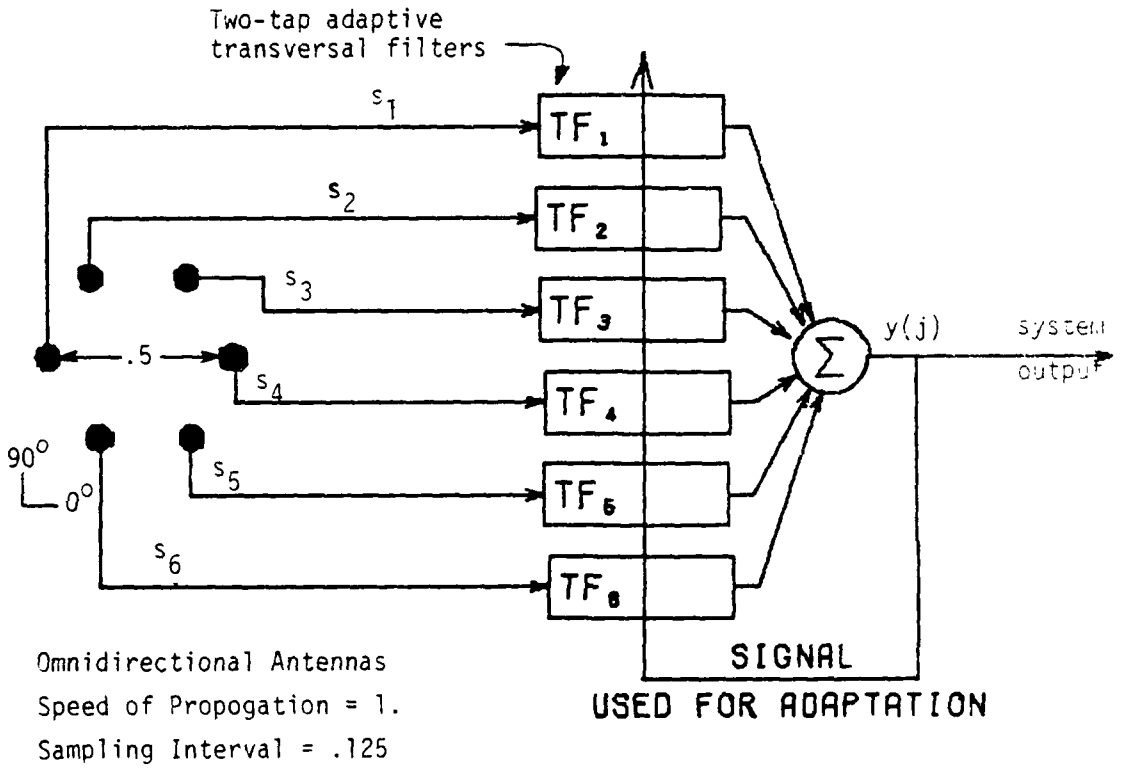


Figure 10-1. Adaptive Antenna Array used in Figures 10-2 through 10-12.

transversal filter TF_k at time j by the phasor $C_r\exp\{i[\omega_r jT+\phi_{rk}]\}$, where ϕ_{rk} is the signal's phase difference between sensor k and some arbitrary reference point. ϕ_{rk} is a function of the angle of arrival of the signal (θ_r) and of the antenna geometry. In this case the data vector is:

$$X(j) = \begin{bmatrix} C_r\exp\{i[\omega_r jT+\phi_{r1}]\} & \text{(data in} \\ C_r\exp\{i[\omega_r (j-1)T+\phi_{r1}]\} & TF_1) \\ \vdots & \\ C_r\exp\{i[\omega_r jT+\phi_{r6}]\} & \text{(data in} \\ C_r\exp\{i[\omega_r (j-1)T+\phi_{r6}]\} & TF_6) \end{bmatrix} \qquad (10\text{-}2)$$

Since the array output is $X^T(j)W(j)$, the array gain to this signal is $X^T(j)W(j)/C_r\exp\{i\omega_r jT\} = [X^T(j)/C_r\exp\{i\omega_r jT\}]W(j)$. Define a vector A_r by $X(j)/C_r\exp\{i\omega_r jT\}$;

$$A_r \triangleq \begin{bmatrix} \exp\{i\phi_{r1}\} \\ \exp\{i(\phi_{r1}-\omega_r T)\} \\ \vdots \\ \exp\{i\phi_{r6}\} \\ \exp\{i(\phi_{r6}-\omega_r T)\} \end{bmatrix} \qquad (10\text{-}3)$$

The array gain to signal r at time j is $A_r^T W(j)$. Suppose it is desirable that the array gain to this signal be $D_r\exp\{i\eta_r\}$. Then the

constraint is written as

$$A_r^T W(j) = D_r \exp\{i\eta_r\} , \qquad (10\text{-}4)$$

which is made a soft constraint. But W(j) is a set of real weights, while A_r and $D_r \exp\{i\eta_r\}$ are complex quantities. The proposed constraint can still be specified with purely real values by separating it into the real and imaginary parts:

$$\mathrm{Re}\{A_r^T\} W(j) = \mathrm{Re}\{D_r \exp\{i\eta_r\}\} , \qquad (10\text{-}5)$$

$$\mathrm{Im}\{A_r\} W(j) = \mathrm{Im}\{D_r \exp\{i\eta r\}\} . \qquad (10\text{-}6)$$

This yields two constraints which are used as soft constraints. Thus the antenna array attempts to keep a complex gain of $D_r \exp\{i\eta_r\}$ in direction θ_r at frequency ω_r, but since the constraints are soft, the gain can vary from the specification ($D_r \exp\{i\eta_r\}$).

This procedure can be followed for several different sinusoids, at the same or different frequencies, yielding a set of constraints. Form the set of constraint vectors (A_r) into a matrix $\underline{A}$; stack the specified gains into a corresponding vector H. Then the set of soft constraints is:

$$\underline{A}W = H . \qquad (10\text{-}7)$$

Weight the soft constraints by constants b_r, which compose the diagonal weighting matrix $\underline{B}$ used in the algorithm (8-8) or (8-7).

Three different sets of constraints, derived as shown, are listed in Tables 1, 2, and 3. The features and effects of each set of constraints are now studied.

The Effect of Table 1 Constraints

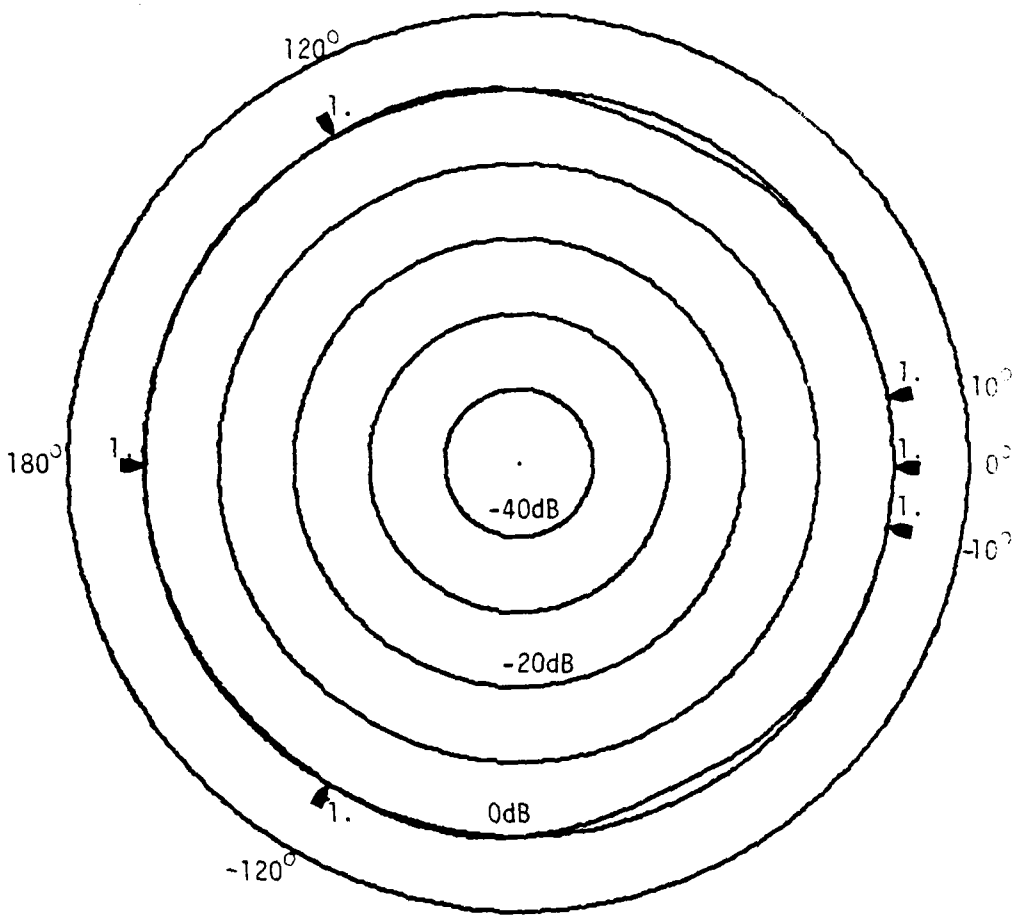
Figures 10-2 through 10-6 show the first example of the use of soft constraints. The soft constraints specify unity power gain at frequency 2, in the directions 0, 10, 120, 180, -120, -10 (in degrees). The constraints are weighted equally at 1. Table 1 summarizes these constraints.

Direction of Constraint (degrees) θ_r	Amplitude of Desired Gain at Frequency 2. D_r	Phase of Desired Gain (degrees) η_r	Constraint Weighting Factor b_r
0.	1.	180.0	1.
10	1.	177.3	1.
120	1.	-90.0	1.
180	1.	-180.0	1.
-120	1.	-90.0	1.
-10	1.	177.3	1.

Table 1 - Constraint Set 1

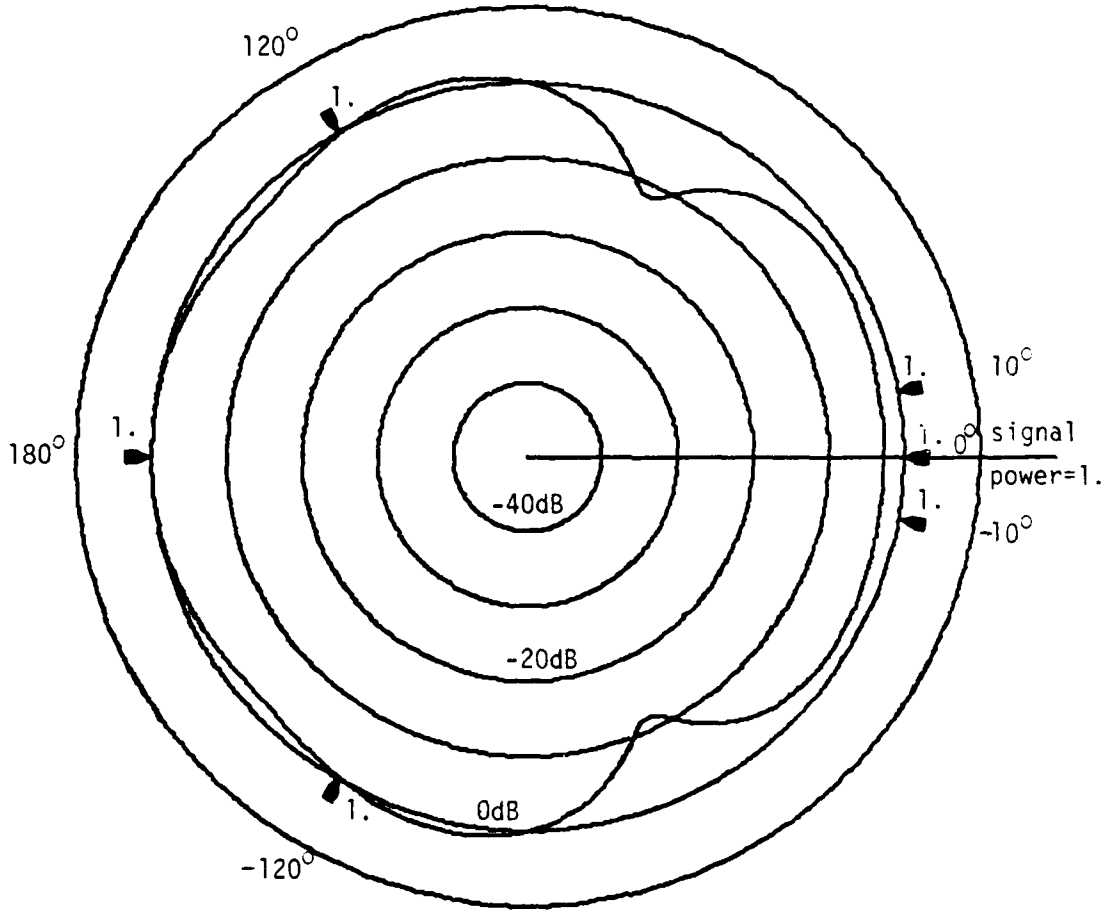
Figure 10-2 shows the antenna directivity pattern that results when the constraint equations (10-7) are solved for the weight vector which satisfies them exactly. This figure is a plot of the power gain that a signal at a frequency of 2 receives, as a function of the arrival direction of the signal, when the weight vector is the solution to the constraint equations.

Figure 10-3 shows the antenna directivity pattern resulting when a unity power sinusoid at frequency 2 is received from 0 degrees, when the antenna array system has adapted to the point of convergence. This example (and all others in this section) also has an isotropic white



◀ - Constraint (figures next to constraint points are constraint weighting factors)

Figure 10-2. Antenna array directivity pattern determined by soft constraints listed in Table 1. (Weight vector frozen at the solution to the constraint equations.)



◀ - Constraint (figures next to constraint points are constraint weighting factors)

Figure 10-3. Antenna array directivity pattern after adaptation by the soft-constraint LMS algorithm (8-8) with:

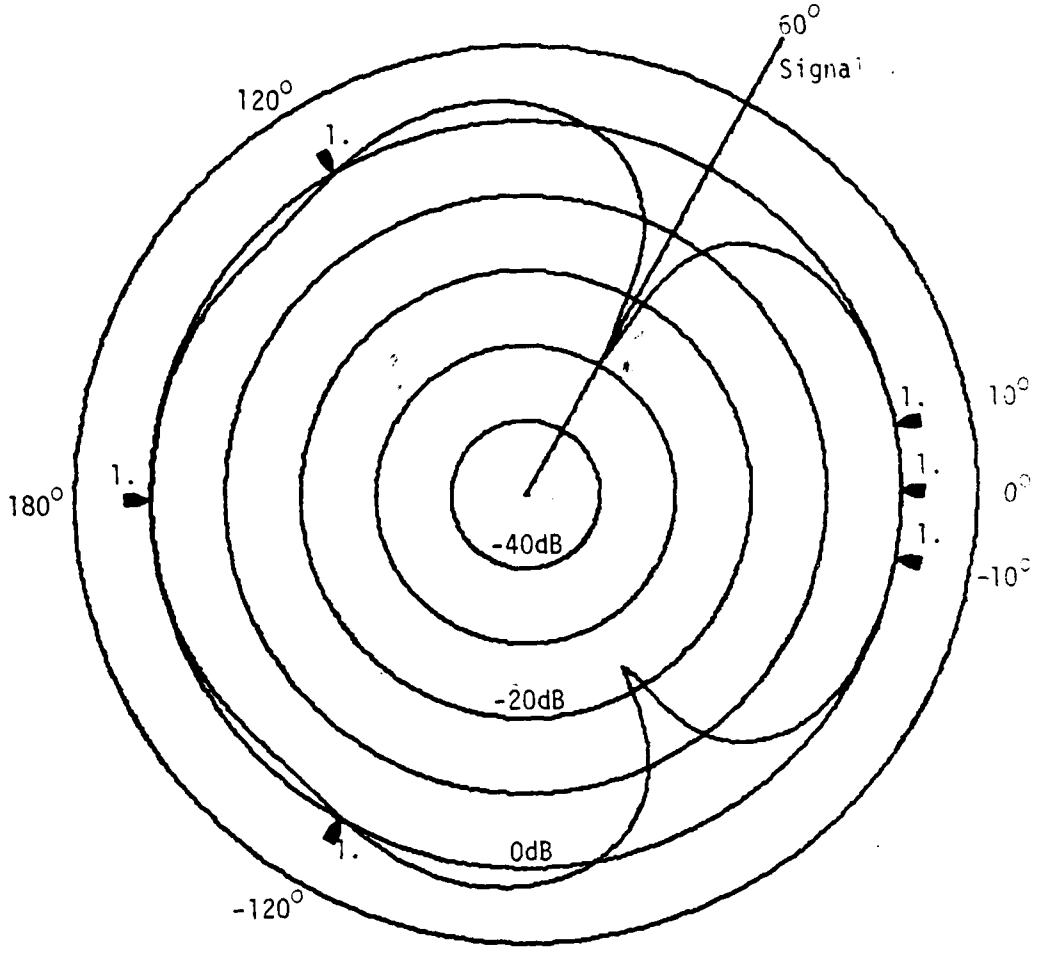
1) A sinusoid of power 1, frequency 2, from 0°
2) Soft constraints of Table 1
3) Isotropic noise of power 0.1

noise field impinging on the antenna array. The noise power at each antenna element is 0.1. Recall that the goal is to minimize the output power while trying to keep the constraint errors small (i.e. keep the gain in the constraint direction close to the constraint values). It can be seen that in the signal's direction the array gain has decreased slightly from that of Fig. 10-2. But as the gain in the signal direction has decreased, the constraint error in that direction has grown (as the constraint errors in the 10 and -10 degree directions have also). Thus the soft constraints result in the gain in the signal direction remaining high, keeping the constraint errors low. The constraint errors at 120, -120 and 180 degrees are kept small without increasing the system output power significantly.

Figure 10-4 shows the converged antenna array directivity pattern for the same constraints, when a unity power sinusoid at a frequency of 2 is received from 60 degrees, an angle not near any of the constraints. When the adaptive filters reach convergence, the signal is attenuated by 30dB, while the constraint error remains small.

Figure 10-5 shows the antenna array directivity pattern for the same set of constraints (Table 1) when the unity power sinusoid at a frequency of 2 is received from 120 degrees, coincident with a constraint. The attenuation in the signal direction is small compared to that of Figure 10-4, but is greater than that of Figure 10-3, in which the signal was arriving close to three constraints, instead of only a single constraint.

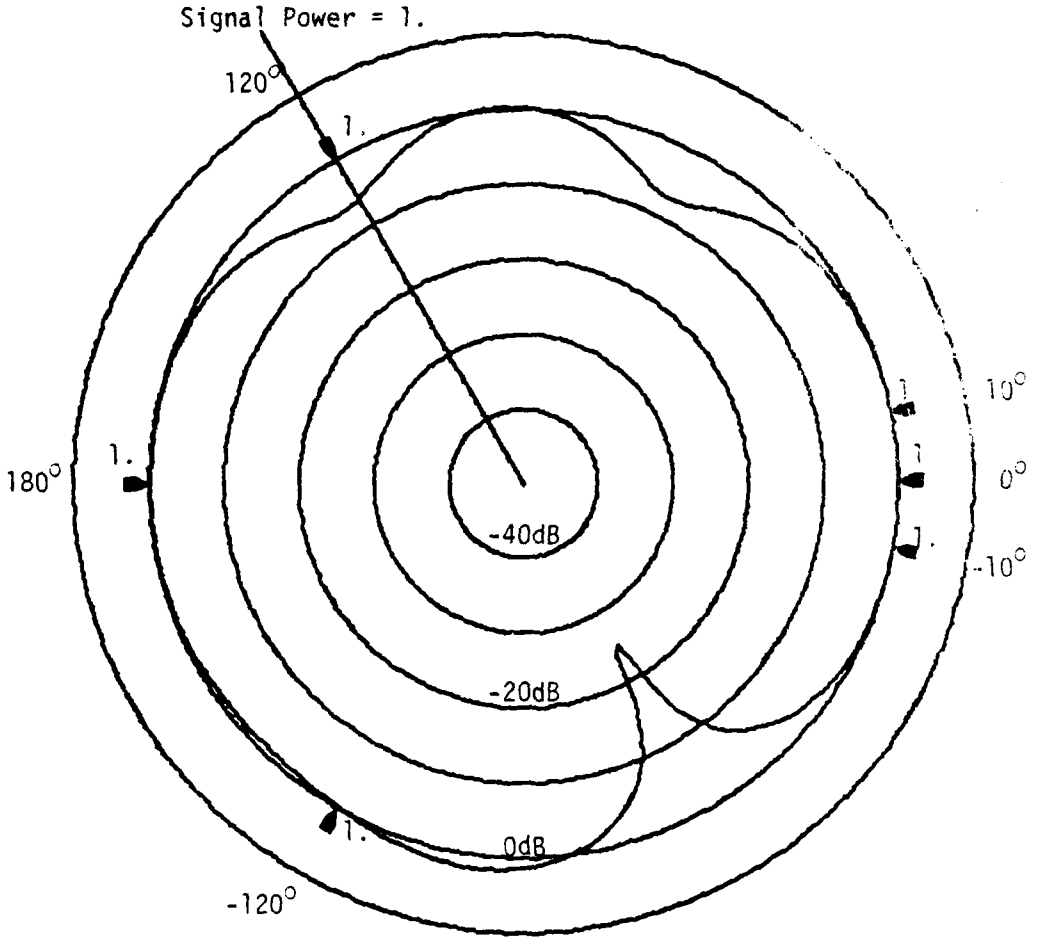
Figure 10-6 is a plot of the converged array gain in the signal direction, for all possible signal arrival directions. This plot is



◀ - Constraint (figures next to constraint points are constraint weighting factors)

Figure 10-4. Antenna directivity pattern after adaptation by the soft-constraint LMS algorithm (8-8) with:

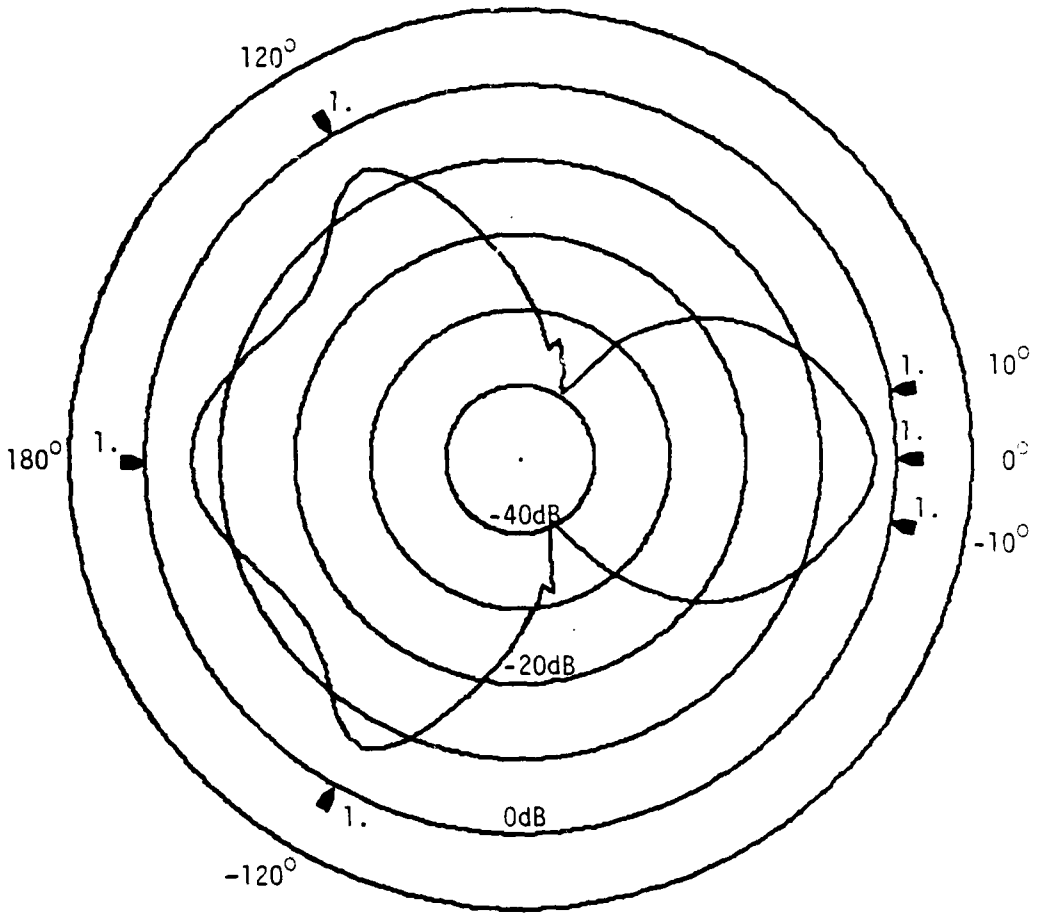
1) A sinusoid of power 1, frequency 2, from 60°
2) Soft constraints of Table 1
3) Isotropic noise of power 0.1



◀ - Constraint (figures next to constraints points are constraint weighting factors)

Figure 10-5. Antenna array directivity pattern after adaptation by the soft-constraint LMS algorithm (8-8) with:

1) A sinusoid of power 1, frequency 2, from 120°
2) Soft constraints as listed in Table 1
3) Isotropic noise of power 0.1



◀ - Constraint (figures next to constraint points are constraint weighting factors)

Figure 10-6. Converged array gain in the signal direction for all possible directions of signal arrival.

Conditions:
1) Soft constraints of Table 1
2) Adaptation by soft-constraint LMS algorithm (8-8)
3) Sinusoid power = 1, frequency = 2
4) Isotropic noise of power 0.1

obtained by placing the unity power signal at a specified direction, calculating the optimum weight vector for this signal configuration, using this optimum weight vector to calculate the gain in the signal direction, and plotting this gain as a single point in Figure 10-6. For example, in Figure 10-5 the gain in the signal direction (120 degrees) is approximately -6dB. This same gain is plotted on Figure 10-6 in the 120 degree position. Figure 10-6 demonstrates that for the constraints specified in Table 1 the gain remains high in directions close to constraints, but signals are more strongly attenuated when not close to constraints.

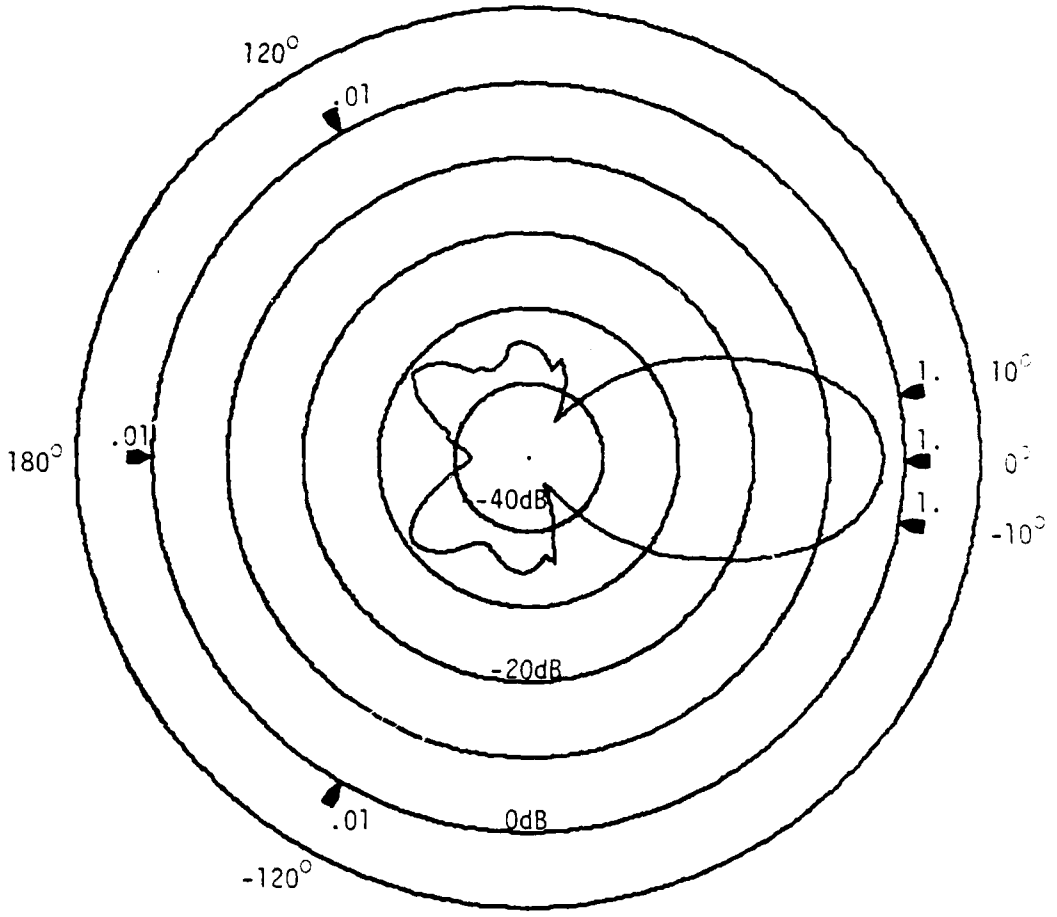
Effect of Table 2 Constraints

Figure 10-7 shows the converged array gain in the signal direction,

Direction of Constraint (degrees) θ_r	Amplitude of Desired Gain at Frequency 2 D_r	Phase of Desired Gain (degrees) η_r	Constraint Weighting Factor b_r
0	1.	180.0	1.
10	1.	177.3	1.
120	1.	-90.0	.01
180	1.	-180.0	.01
-120	1.	-90.0	.01
-10	1.	177.3	1.

Table 2 - Constraint Set 2

for all possible directions of signal arrival, for the set of constraints in Table 2. These constraints differ from the previous constraints in that the weightings in the 120, -120, and 180 degree positions are decreased by a factor of 100. Figure 10-7 shows the effect: the array's gain to signals arriving from directions close to the weak constraints



◀ - Constraint (figures next to constraint points are constraint weighting factors

Figure 10-7. Converged array gain in the signal direction for all possible directions of signal arrival.

Conditions:
1) Soft constraints of Table 2
2) Adaptation by soft-constraint LMS algorithm (8-8)
3) Sinusoid power = 1, frequency = 2
4) Isotropic noise of power 0.1

is greatly reduced from the previous case (Fig. 10-6). This occurs because the output power is significantly reduced by decreasing the array gain in the signal arrival direction, while incurring only small constraint errors due to the low weighting coefficient. Thus the weighting coefficients control the "softness" of the constraint. A large weighting coefficient implies that the decrease in output power must be large to allow a small deviation from the constraint; a small weighting coefficient implies that greater deviation from the constraint is allowed with little penalty, so the algorithm can decrease the output power significantly.

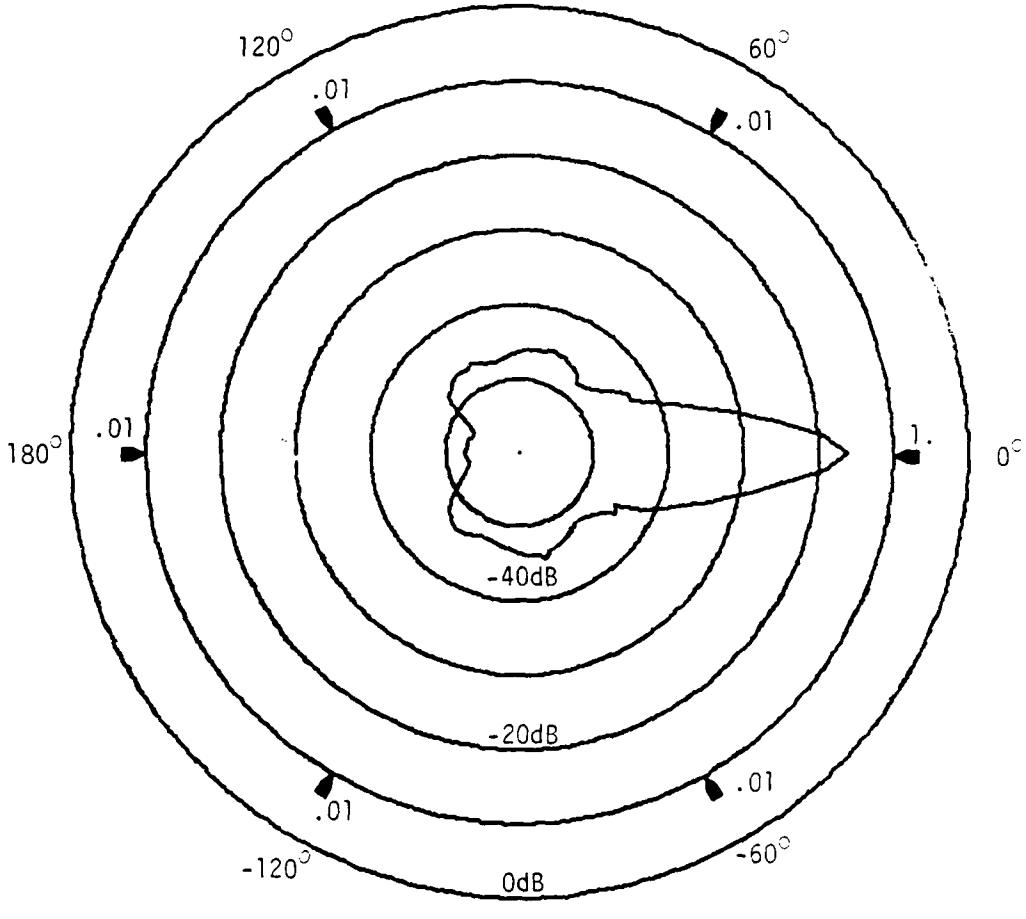
Effect of Table 3 Constraints

Figure 10-8 demonstrates the effect on the large lobe of Figure 10-7 when the two constraints at 10 and -10 degrees are moved to 60 and -60 degrees and simultaneously weakened by a factor of 100. Table 3 presents

Direction of Constraint (degrees) θ_r	Amplitude of Desired Gain at Frequency 2 D_r	Phase of Desired Gain (degrees) η_r	Constraint Weighting Factor b_r
0.	1.	180.0	1.
60	1.	177.3	.01
120	1.	-90.0	.01
180	1.	-180.0	.01
-120	1.	-90.0	.01
-60	1.	177.3	.01

Table 3 - Constraint Set 3

this set of constraints. Comparing Figures 10-7 and 10-8 shows that when two strong constraints are at the 10 and -10 degree positions as in



◀ - Constraint (figures next to constraint points are constraint weighting factors)

Figure 10-8. Converged array gain in the signal direction for all possible directions of signal arrival.

Conditions:
1) Soft constraints of Table 3
2) Adaptation by soft-constraint LMS algorithm (8-8)
3) Sinusoid power = 1, frequency = 2
4) Isotropic noise of power 0.1

Figure 10-7, the angular sector over which a signal is received without significant attenuation is much broader than when only a single strong constraint is present, as in Figure 10-8.

Antenna Array Gain in the Constraint Directions

Figure 10-9 shows the gain in the 0 degree direction maintained by the soft constraint, for all possible arrival directions of a unit power signal with a frequency of 2, for the constraints of Table 2. The plot is calculated by placing the signal at a given direction, calculating the converged weight vector, calculating the resulting gain at 0 degrees (frequency of 2), and plotting it on Figure 10-9. The gain in the 0 degree position remains close to the unity gain specification, decreasing only when the signal is also close to 0 degrees. When the signal arrives from close to 0 degrees, the array gain in the 0 degree direction drops slightly to reduce the system output power, but cannot drop significantly without causing large constraint errors.

Figure 10-10 shows the array gain in the direction of the much weaker constraint at 180 degrees for all possible arrival directions of a unit power signal with a frequency of 2 (again using the constraints of Table 2). Since the gain in this direction can vary greatly without incurring large constraint error (no strong constraints in this region), the adaptive array concentrates on minimizing output power rather than on maintaining the constraint, as seen by the wide variation in the array's gain in this direction.

A Two-Signal Case

Figures 10-11 and 10-12 show a two signal case. Figure 10-11 shows

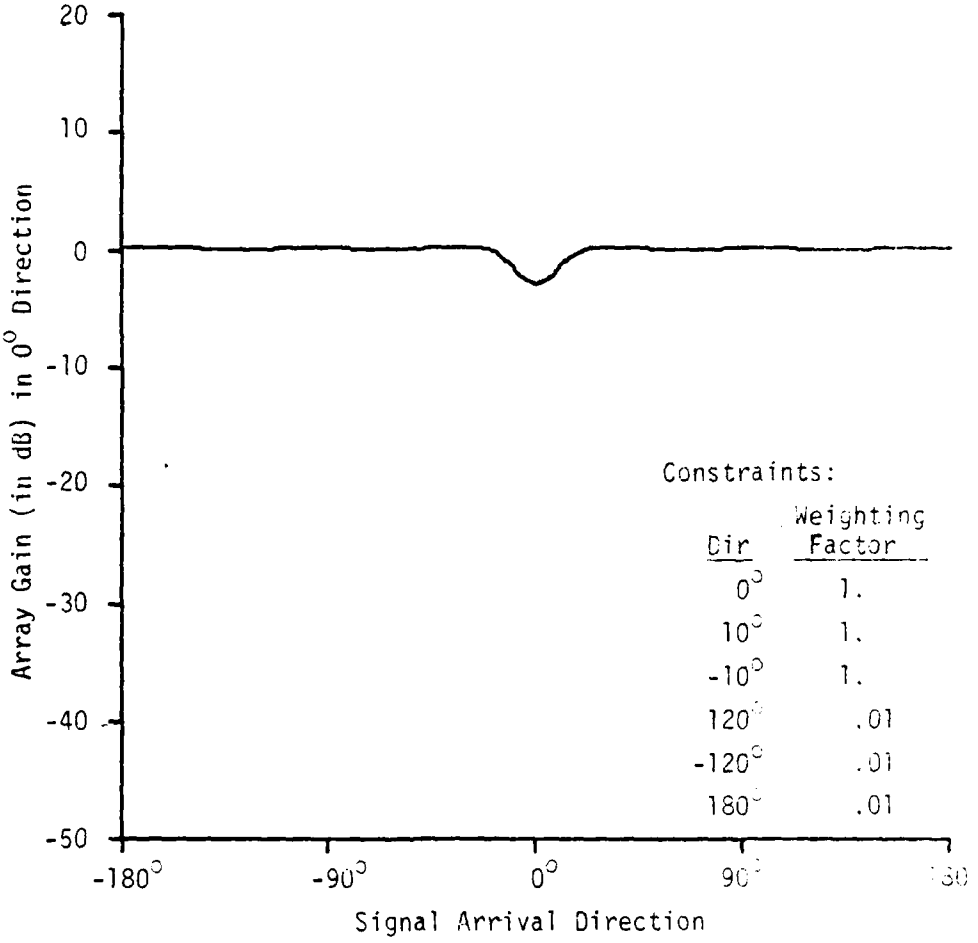


Figure 10-9. Array gain in the zero degree direction maintained by the soft constraint, when array has been adapted to convergence on a signal arriving from the direction specified along the abscissa.

Conditions:
1) Soft constraints of Table 2, as listed
2) Adaptation by soft-constraint algorithm (8-8)
3) Sinusoid power = 1, frequency = 2
4) Isotropic noise of power 0.1

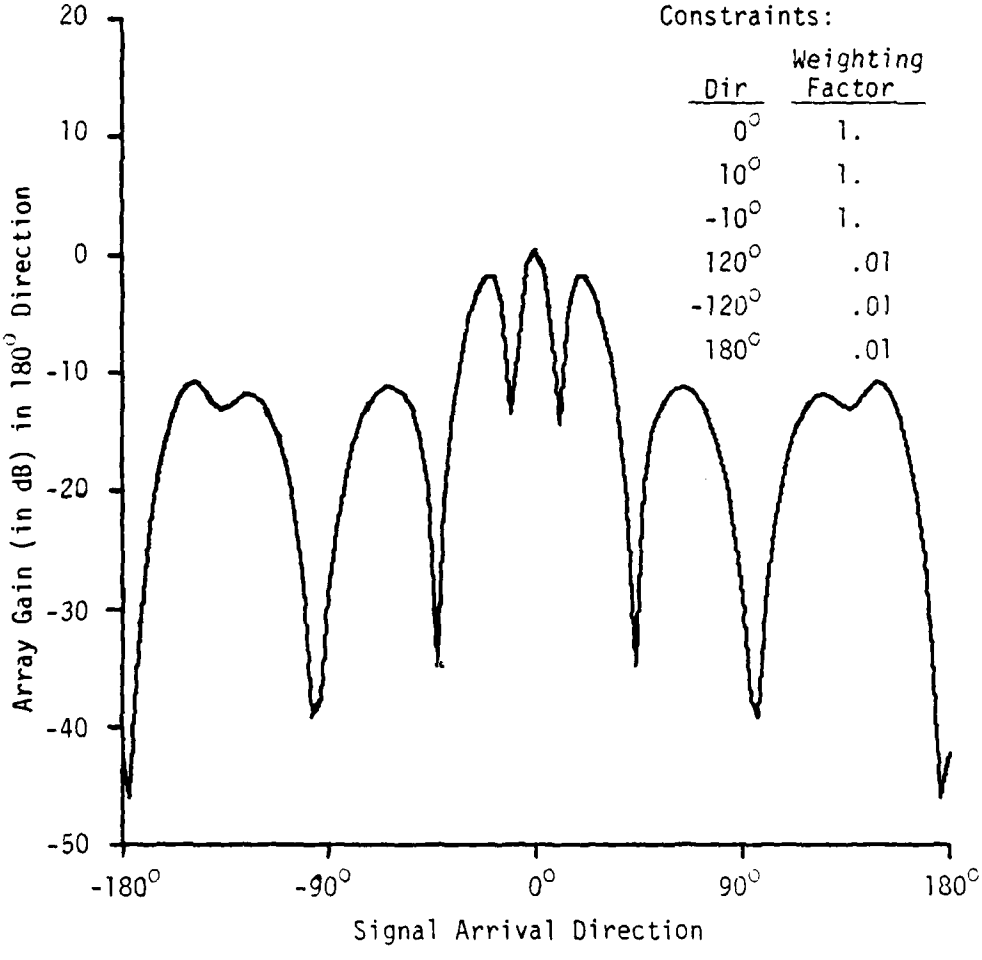
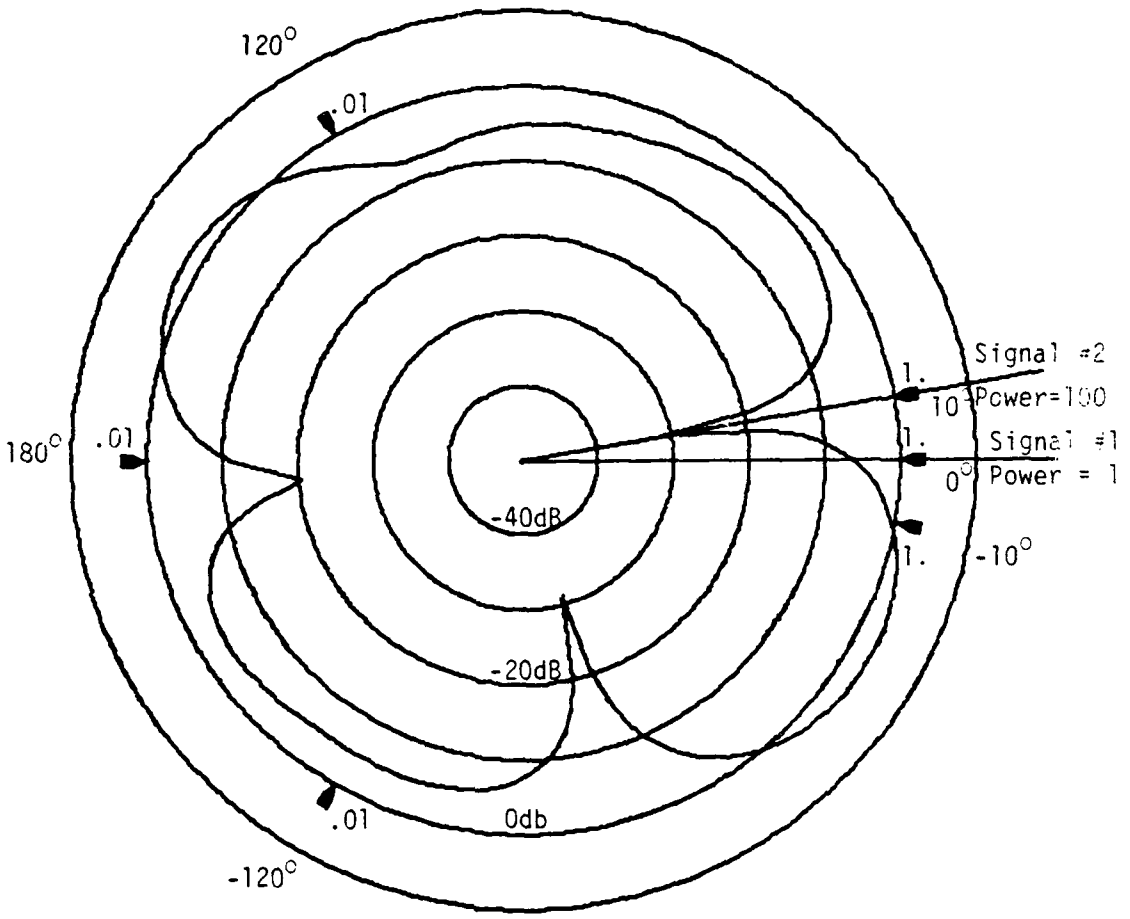


Figure 10-10. Array gain in the 180 degree direction maintained by the soft constraint, when array has been adapted to convergence on a signal arriving from the direction specified along the abscissa

Conditions:

1) Soft constraints of Table 2, as listed
2) Adaptation by soft-constraint algorithm (8-8)
3) Sinusoid power = 1, frequency 2
4) Isotropic noise of power 0.1



◄ - Constraint (figures next to constraint points are constraint weighting factors)

Figure 10-11: Antenna array directivity pattern after adaptation by the soft-constraint LMS algorithm (8-8) with two signals

Conditions:
1) Signal #1 power = 1, frequency = 2, from 0°
2) Signal #2 power = 100, frequency = 2, from 10°
3) Soft constraints as listed in Table 2
4) Isotropic noise of power 0.1

the antenna array directivity pattern for a particular signal configuration, with the Table 2 constraints. The signal (#1) arriving from 0 degrees has a power of 1; the signal (#2) arriving from 10 degrees has a power of 100. Both signals have a frequency of 2. The figure shows that the strong signal (#2) is greatly attenuated even though it is arriving from a direction where a constraint is located. This is because the signal is very strong compared to the constraint in this direction; the array concentrates on attenuating the signal rather than on satisfying the constraint. The weak signal (#1) arriving from 0 degrees is only slightly attenuated, because it is weak in comparison with any of the constraints in the neighborhood. Note that the presence of the strong signal has had little or no effect on the array's gain to the weak signal (compare the gain with Figure 10-3).

Figure 10-12 shows the array's gain to the weak signal (#1) at 0 degrees as a function of the arrival direction of the strong signal (#2). This figure shows that the array's gain to the weak signal is only affected by the strong signal when the strong signal is arriving from a direction very close to that of the weak signal. When the strong signal arrives from 0 degrees the two signals are inseparable; the array acts as if there were only one strong signal. Since this composite signal is strong compared to the constraint at 0 degrees, the array concentrates on attenuating the composite signal. As the strong signal moves away from the weak signal, the array is better able to resolve the two signals, and continues to attenuate the strong signal, while allowing the gain in the direction of the constraint at 0 degrees to increase again.

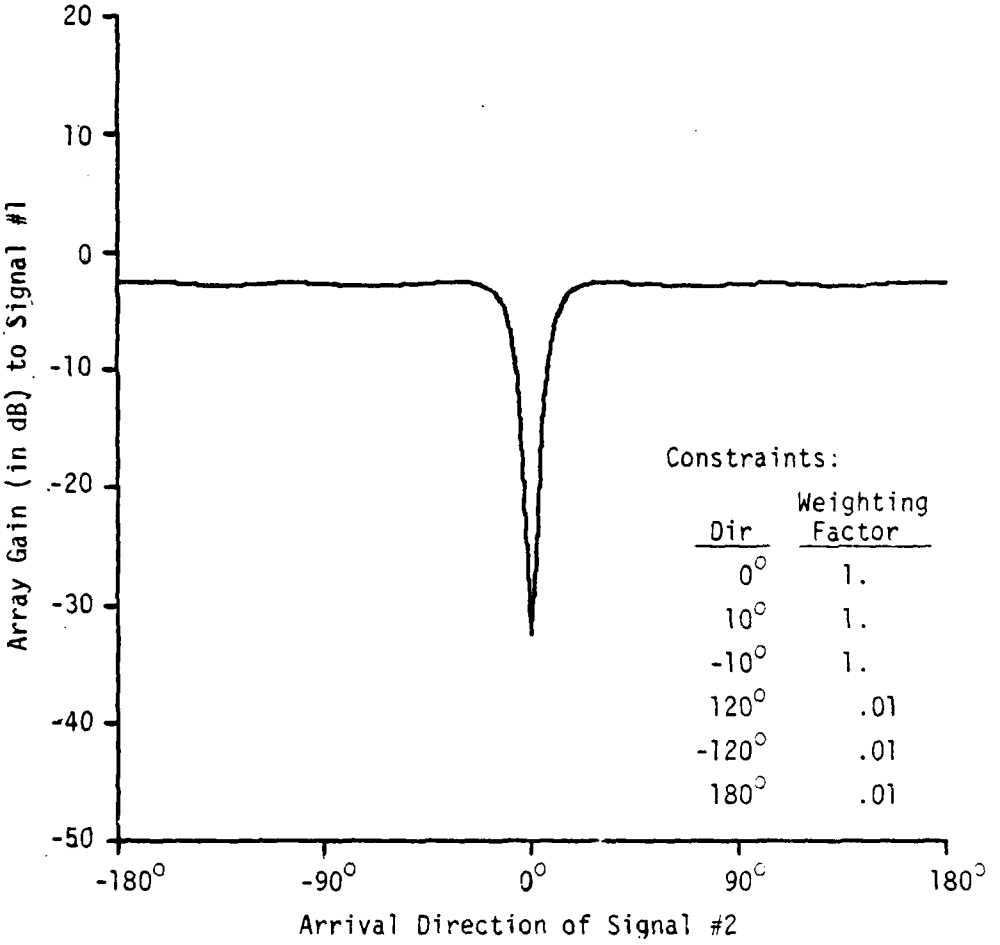


Figure 10-12. Array gain in direction of Signal #1 maintained by a soft constraint, when array has been adapted to convergence with Signal #2 arriving from the direction specified along the abscissa

Conditions:
1) Soft constraints of Table 2, as listed
2) Adaptation by soft-constraint LMS algorithm (8-8)
3) Sinusoid #1 power = 1, frequency = 2, from 0°
4) Sinusoid #2 power = 100, frequency = 2
5) Isotropic noise of power 0.1

Summary

This section has demonstrated the use of soft constraints for adaptive antenna arrays. It has shown that soft constraints can maintain array gain in the presence of signals which are weak compared to the constraints; but strong signals are attenuated. It was also seen that the strength of constraints could be varied, and placing constraints closely together could expand the angular sector over which the array gain is maintained.

XI. OUTPUT POWER DUE TO A SIGNAL AS A FUNCTION OF ITS INPUT POWER

This section investigates the power at the output of a converged soft-constraint LMS adaptive filter, due to a particular input signal; relating the output power to the input power.

When fixed (non-adapting) filters are used for signal processing, an increase in the input power of a signal always means a corresponding increase in the output power of the filter. But this is not necessarily true for adaptive filters. An increase in the power of any signal can change the optimum filter. And the new optimum filter might attenuate the signal more strongly than the previous optimum filter. It is possible that the increase in attenuation is so great that the signal's increase of input power is more than cancelled; so it is possible that the output power due to the signal is actually less than before. Thus, when a signal increases its power at the input of an adaptive filter, the output power due to the signal can actually decrease. This phenomenon is studied in this section.

Assume that the time sequence u(j), the input to the adaptive filter, consists of a stationary signal to be studied, denoted s(j); and that all other signals in u(j) are also stationary, and when summed together are denoted by n(j):

$$u(j) = s(j) + n(j) \quad . \tag{11-1}$$

Denote the autocorrelation of s(j) when it is in the tapped delay line by $\alpha^2 \underline{R}_{ss}$, where α^2 is the power of s(j). Denote the cross-correlation

between s(j) and the desired signal d(j) by $f(\alpha)P_{ds}$, where $f(\alpha)$ is a function of the input power of s(j). (Several expressions for $f(\alpha)$ are studied later in this section.) Assume that s(j) and n(j) are uncorrelated:

$$E\{s(j)n(j)\} = 0 \ . \tag{11-2}$$

Denote the autocorrelation of n(j) when in the tapped delay line by $\underline{R}_{nn}$, and the cross correlation between n(j) and d(j) by P_{dn}.

With these definitions the complete input autocorrelation matrix is $\alpha^2\underline{R}_{ss}+\underline{R}_{nn}$, and the complete input cross-correlation with the desired signal is $f(\alpha)P_{ds}+P_{dn}$.

Using (6-3) the optimum weight vector is:

$$W_{opt} = (\alpha^2\underline{R}_{ss}+\underline{R}_{nn}+\underline{A}^T\underline{B}\underline{A})^{-1}[f(\alpha)P_{ds}+P_{dn}+\underline{A}^T\underline{B}H] \ . \tag{11-3}$$

For ease of notation, denote $\underline{R}_{nn}+\underline{A}^T\underline{B}\underline{A}$ by $\underline{U}$, and $P_{dn}+\underline{A}^T\underline{B}H$ by V; also assume that $\underline{U}$ is a matrix of full rank. This yields:

$$W_{opt} = (\alpha^2\underline{R}_{ss}+\underline{U})^{-1}[f(\alpha)P_{ds}+V] \ . \tag{11-4}$$

The output power of the adaptive filter due to the input signal under study s(j) is:

$$P_{out} = W^T\alpha^2\underline{R}_{ss}W \ . \tag{11-5}$$

Using the optimum weight vector of (11-4) yields:

$$P_{out} = [f(\alpha)P_{ds}+V]^T(\alpha^2\underline{R}_{ss}+\underline{U})^{-1}\alpha^2\underline{R}_{ss}(\alpha^2\underline{R}_{ss}+\underline{U})^{-1}[f(\alpha)P_{ds}+V]. \tag{11-6}$$

Appendix F demonstrates that since $\underline{R}_{ss}$ and $\underline{U}$ are both hermitian matrices and $\underline{U}$ is nonsingular, a matrix $\underline{S}$ can be found such that:

$$\underline{S}^T\underline{R}_{ss}\underline{S} = \underline{\Psi} \ , \qquad (11\text{-}7)$$

$$\underline{S}^T\underline{U}\underline{S} = \underline{I} \ , \qquad (11\text{-}8)$$

where $\underline{\Psi}$ is a diagonal matrix, and $\underline{I}$ is the identity matrix. ($\underline{S}$ will be a purely real matrix, since both $\underline{R}_{ss}$ and $\underline{U}$ are purely real. Thus $\underline{S}^T = \underline{S}^+$.) Rearranging (11-7) and (11-8) yields:

$$\underline{R}_{ss} = \underline{S}^{-T}\underline{\Psi}\underline{S}^{-1} \ , \qquad (11\text{-}9)$$

$$\underline{U} = \underline{S}^{-T}\underline{I}\underline{S}^{-1} \ , \qquad (11\text{-}10)$$

where $(\underline{S}^T)^{-1}$ is abbreviated to $\underline{S}^{-T}$. Substituting (11-9) and (11-10) into (11-6) yields:

$$\begin{aligned} P_{out} = \quad & [f(\alpha)P_{ds}+V]^T(\alpha^2\underline{S}^{-T}\underline{\Psi}\underline{S}^{-1}+\underline{S}^{-T}\underline{I}\underline{S}^{-1})^{-1} \\ & \cdot \alpha^2\underline{S}^{-T}\underline{\Psi}\underline{S}^{-1} \ \cdot \\ & (\alpha^2\underline{S}^{-T}\underline{\Psi}\underline{S}^{-1}+\underline{S}^{-T}\underline{I}\underline{S}^{-1})^{-1}[f(\alpha)P_{ds}+V] \\ = \ & [f(\alpha)P_{ds}+V]^T\underline{S}(\alpha^2\underline{\Psi}+\underline{I})^{-1}\underline{S}^T\alpha^2\underline{S}^{-T}\underline{\Psi}\underline{S}^{-1}\underline{S}(\alpha^2\underline{\Psi}+\underline{I})^{-1}\underline{S}^T[f(\alpha)P_{ds}+V] \\ = \ & [f(\alpha)P_{ds}+V]^T\underline{S}(\alpha^2\underline{\Psi}+\underline{I})^{-1}\alpha^2\underline{\Psi}(\alpha^2\underline{\Psi}+\underline{I})^{-1}\underline{S}^T[f(\alpha)P_{ds}+V] \ . \end{aligned} \qquad (11\text{-}11)$$

Now since $\underline{\Psi}$ is a diagonal matrix ($\underline{\Psi} = \text{diag}(\psi_i)$), P_{out} can be written in terms of individual components as:

$$P_{out} = \sum_{i=1}^{n} P_{outi} = \sum_{i=1}^{n} \frac{\alpha^2\psi_i}{(\alpha^2\psi_i+1)^2}\{\underline{S}^T[f(\alpha)P_{ds}+V]\}_i^2 \ . \qquad (11\text{-}12)$$

where $\{\underline{S}^T[f(\alpha)P_{ds}+V]\}_i$ denotes the i^{th} element of the vector $\underline{S}^T[f(\alpha)P_{ds}+V]$.

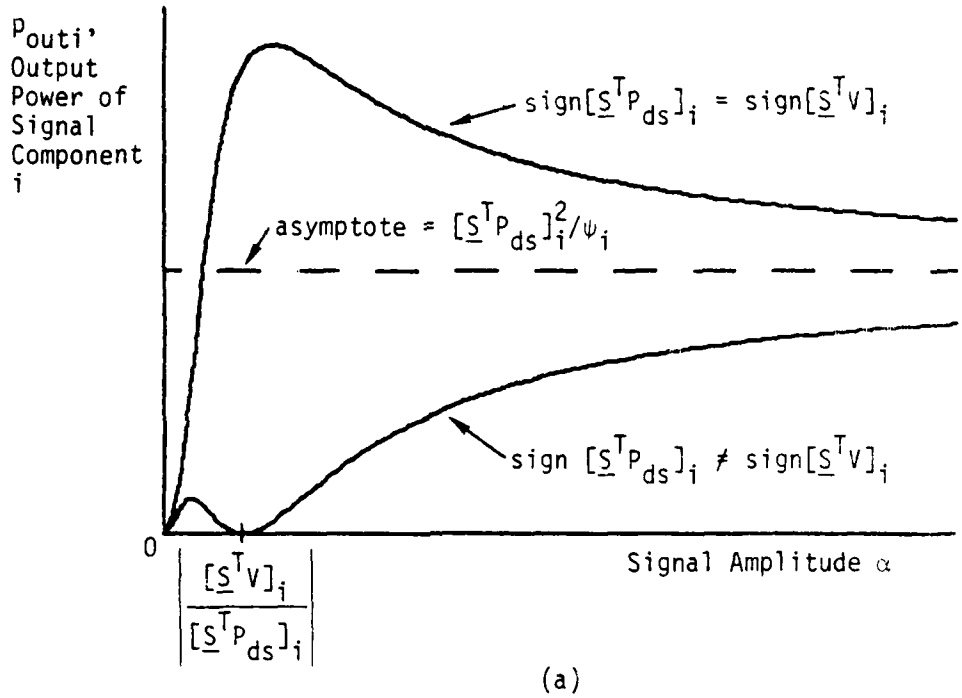
Thus the output power is the sum of a set of components p_{outi} which vary individually as the input power α^2 of the signal under study is varied. The value of an individual component p_{outi} of the output power p_{out} may be plotted as a function of the input power α^2 once $f(\alpha)$ is known. Three cases are of particular interest:

Case 1: $f(\alpha) = \alpha$.

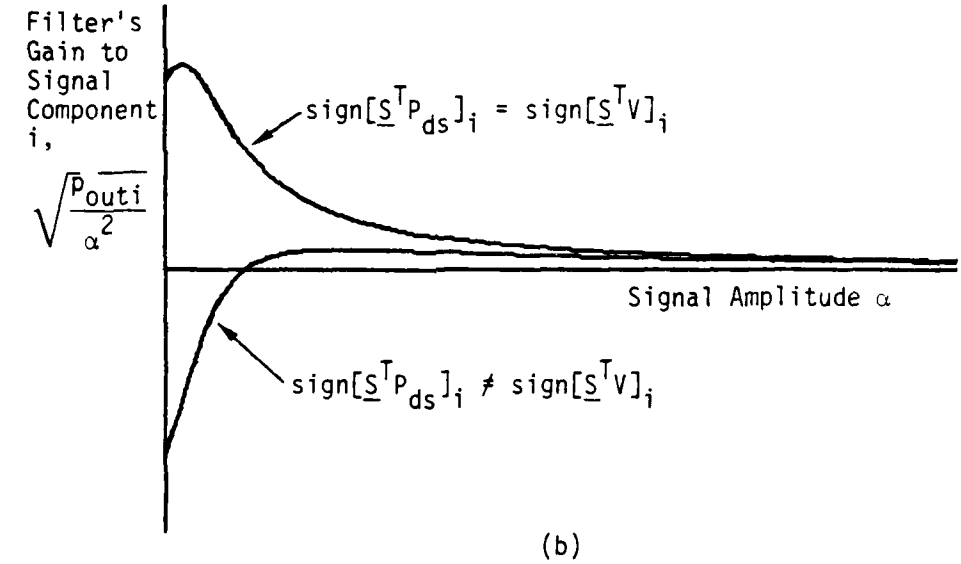
The signal under study s(j) is correlated with the desired signal d(j), but the power of d(j) remains constant even when the input power of s(j) increases. This case can occur when d(j) is generated separately from s(j). For this case, a component of (11-12) has the form

$$p_{outi} = \frac{\alpha^2 \psi_i}{(\alpha^2 \psi_i + 1)^2} \{\underline{S}^T [\alpha P_{ds} + V]\}_i^2 . \qquad (11\text{-}13)$$

The shape of this function is shown in Figure 11-1a. The figure shows that the curve can have one of two forms, depending on whether or not $[\underline{S}^T P_{ds}]_i$ and $[\underline{S}^T V]_i$ have the same sign. When the signs are the same, the gain of the filter to s(j) increases slightly at first, then decreases toward zero (as seen in Fig. 11-1b.) However, the rate of decrease of gain compared to the rate of increase of input power is such that the output power approaches an asymptotic value of $[\underline{S}^T P_{ds}]_i / \psi_i$ (as seen in Fig. 11-1a). The decrease occurs because s(j) begins to dominate n(j) and overwhelm the soft constraints, so the adaptive filter begins to do power equalization to make the power of filter output y(j) match that of d(j). For opposite signs, Fig. 11-1a shows that the output power due to s(j) can increase, then decrease to zero again, and finally increase to the asymptotic value. The reason for the decrease is that initially the weight vector is dominated by



(a)



(b)

Figure 11-1. Soft-constraint LMS adaptive filter gain to a signal component, and the corresponding signal component output power, as a function of the signal component input power. Case 1: $f(\alpha) = \alpha$.

n(j) or the soft constraints, and the estimate of d(j) has the wrong sign compared to d(j). Fig. 11-1b shows that the filter's gain begins with the wrong sign. As s(j) grows, it has more effect on the weight vector. For a good estimate, the sign of the estimate and the filter gain must change, causing the output power to go through zero at some point.

Case 2: $f(\alpha) = \alpha^2$.

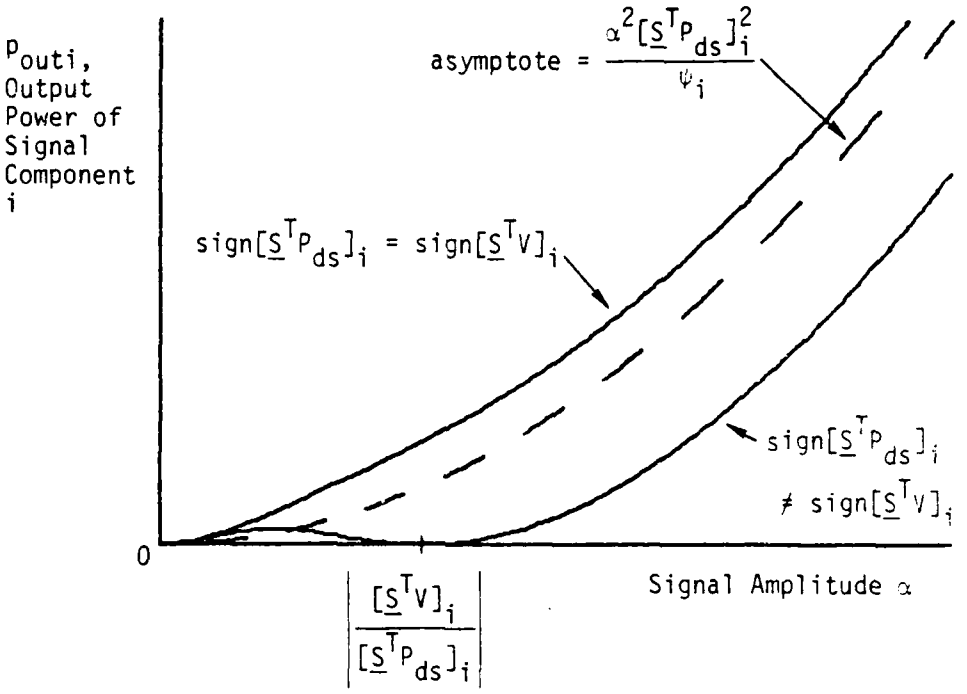
The signal under study s(j) is correlated with the desired signal d(j), and d(j) is derived from s(j), so that the power of d(j) increases linearly with an increase in the power of s(j). An example of this relationship is the line enhancer configuration [3,4,11]. In this case a component of (11-12) has the form

$$P_{outi} = \frac{\alpha^2 \psi_i}{(\alpha^2 \psi_i + 1)^2} \{\underline{S}^T[\alpha^2 P_{ds} + V]\}_i^2 \quad . \tag{11-14}$$

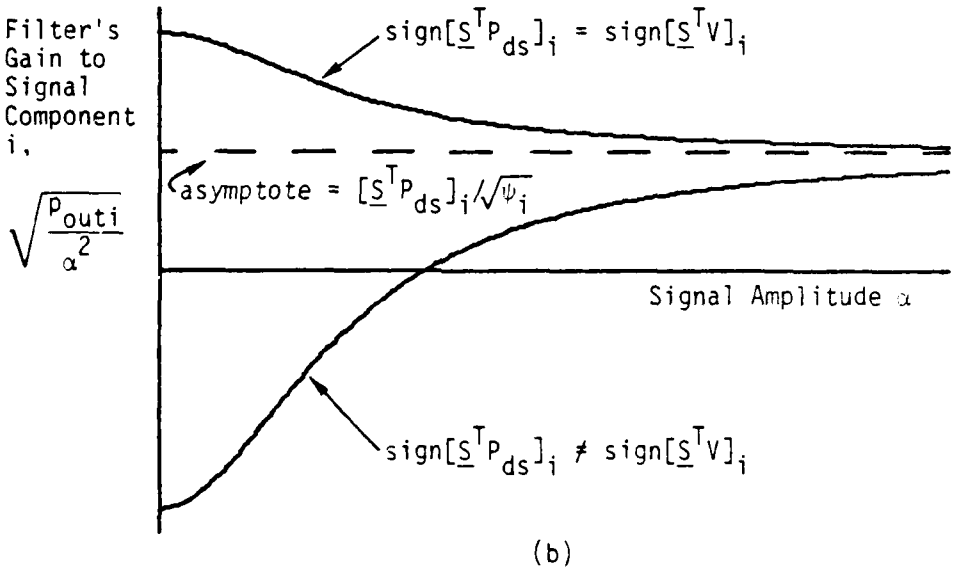
Figure 11-2a shows the shape of this curve. The asymptote of the output power curves (Fig. 11-2a) is a parabola. When $[\underline{S}^T P_{ds}]_i$ and $[\underline{S}^T V]_i$ have the same sign, the output power curve essentially follows the parabola; the soft constraints (and/or n(j)) cause the weight to be of the proper sign but larger than necessary, so the output power curve is above the asymptote. For opposite signs, the weight must again change signs, causing the dip to zero output power as seen, then increase once the proper sign is obtained. Fig. 11-2b shows that in either case the filter gain approaches a constant.

Case 3: $f(\alpha) = 0$.

The signal under study s(j) is uncorrelated with the desired signal d(j). This occurs when s(j) is noise or interference. In this case,



(a)



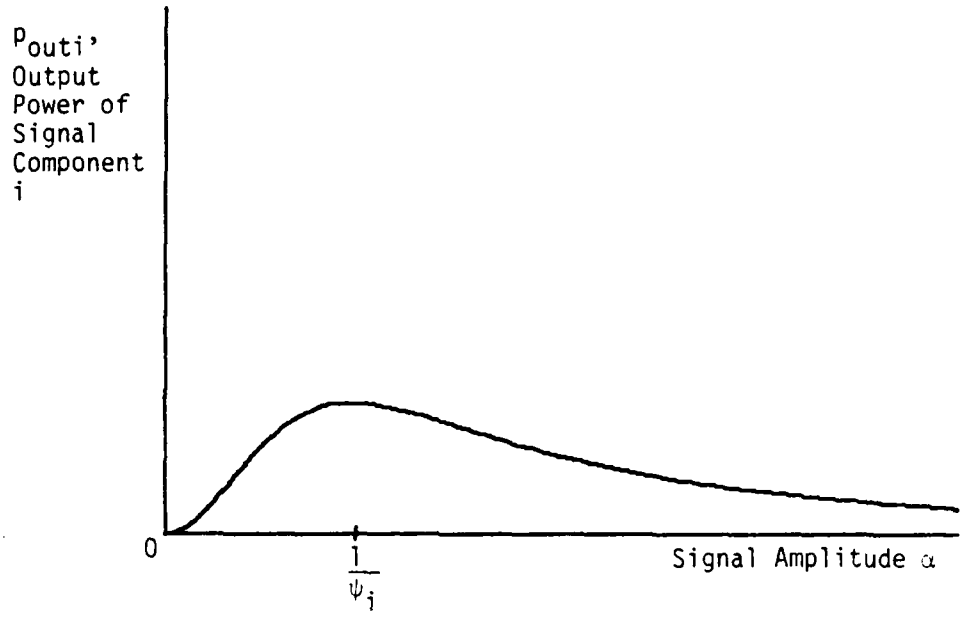
(b)

Figure 11-2. Soft-constraint LMS adaptive filter gain to a signal component, and the corresponding signal component output power, as a function of the signal component input power. Case 2: $f(\alpha) = \alpha^2$.

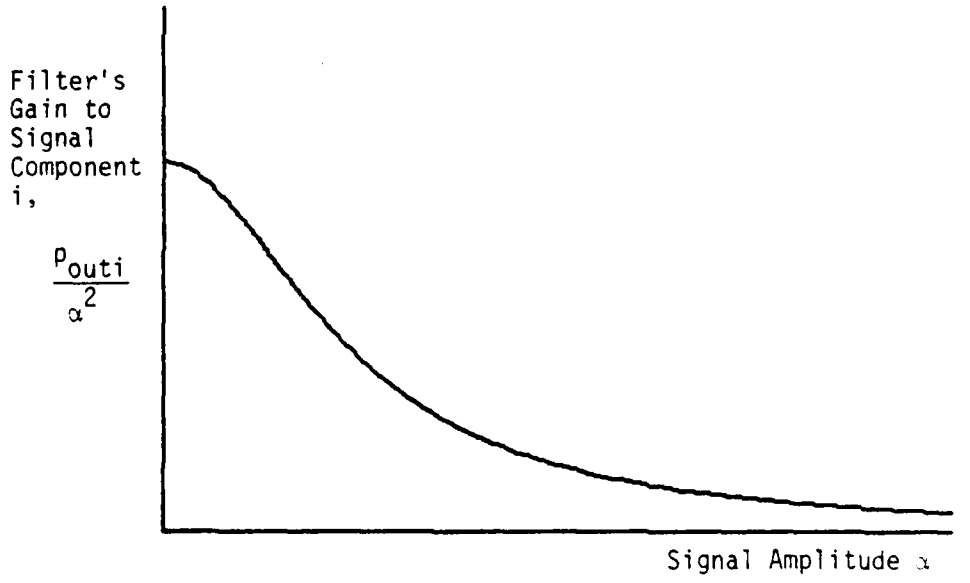
a component of (11-12) has the form

$$P_{outi} = \frac{\alpha^2 \psi_i}{(\alpha^2 \psi_i + 1)^2} \{\underline{S}^T V\}_i^2 \ . \qquad (11\text{-}15)$$

Figure 11-3a shows the shape of this curve. Here, the filter begins to turn itself off as s(j) begins to dominate n(j) and the soft constraints. This curve is of strong interest because it shows that strong signals can be attenuated much more than weak signals. This phenomenon could be used to create adaptive filters which pass weak signals but attenuate strong signals, effectively filtering signals based on their strength. The quantity ψ_i determining where the peak of this curve occurs is under some control by the filter designer, since selection of the soft constraints affects ψ_i.



(a)



(b)

Figure 11-3. Soft-constraint LMS adaptive filter gain to a signal component, and the corresponding signal component output power, as a function of the signal component input power. Case 3: $f(\alpha) = 0$.

XII. RELATION OF THE SOFT-CONSTRAINT LMS ALGORITHM TO OTHER VERSIONS OF THE LMS ALGORITHM

The LMS Algorithm

The LMS algorithm defined in references [1-3] is:

$$W(j+1) = W(j) + 2\mu e(j)X(j) \quad . \tag{12-1}$$

By comparing (12-1) with the soft-constraint LMS algorithm (8-7) it is seen that the LMS algorithm is a special case of the soft-constraint LMS algorithm, since setting $\underline{B} = 0$ in (8-7) yields (12-1). The effect of setting $\underline{B} = 0$ is that all of the soft constraints are turned off.

The Leaky LMS Algorithms

By examining (8-8) it can be seen that the "leaky" LMS algorithm [11-13]:

$$W(j+1) = \nu W(j) + 2\mu e(j)X(j) \tag{12-2}$$

is also a special case of the soft-constraint LMS algorithm. The leaky LMS algorithm has a multiplier ν on the $W(j)$ term which is a positive scalar less than one. The soft-constraint LMS algorithm has a corresponding term $\underline{I}-2\mu\underline{A}^T\underline{B}\underline{A}$ which is a matrix. However, if $\underline{A}$ is chosen to be an identity matrix, and $\underline{B}$ is diagonal with the diagonal elements all equal to a scalar γ, then the multiplier in the soft-constraint LMS algorithm (8-8) reduces to the scalar $1-2\mu\gamma$.

The other difference between the leaky LMS algorithm (12-2) and the soft-constraint LMS algorithm (8-8) is the presence of the driving term $2\mu\underline{A}^T\underline{B}H$ in the latter. However, by choosing H to be zero, this term disappears. Thus, the leaky LMS algorithm is seen to be a special case of the soft-constraint LMS algorithm, by choosing the constraints in the latter to constrain each of the weights to zero, with identical weighting on each of the constraints.

Zahm's Algorithm

Zahm's algorithm [14] is:

$$W(j+1) = \nu W(j) - 2\mu y(j)X(j) + V \qquad (12\text{-}3)$$

where V is a constant vector.

Zahm's algorithm is also a special case of the soft-constraint LMS algorithm (8-8), when the matrices $\underline{A}$ and $\underline{B}$ in the latter are chosen in the same manner as for the Leaky LMS algorithm, and a non-zero constraint vector H is selected such that $V=2\mu\underline{A}^T\underline{B}H$, and with no desired signal available.

Frost's Hard Constraint LMS Algorithm

Frost's algorithm [9] is extremely similar to the soft-constraint LMS algorithm (8-8) because the constraints are identical. The only difference is that Frost requires exact solution of the constraints at all times. Intuitively, one would feel that as the soft constraints are stiffened, the soft-constraint LMS algorithm's solution would approach that of the hard-constraint LMS algorithm. This is true, and can be stated as follows: denote the optimum weight vector for Frost's hard constraint problem by W_{hc}. Now consider letting the weighting matrix $\underline{B}$

on a set of soft constraints for (8-8) be multiplied by a scalar γ, so that the true weighting is $\gamma\underline{B}$. Then for the optimum weight vector W_{opt} in (8-8):

$$\lim_{\gamma \to \infty} W_{opt} = W_{hc} \quad . \tag{12-4}$$

This relation is proved as Theorem 5 in Appendix G.

Thus the optimum weight vector of the soft constraint LMS algorithm approaches the optimum weight vector of Frost's hard constraint LMS algorithm in the limit as the hardness (weighting) of the soft constraints goes to infinity.

XIII. CONCLUSIONS AND DISCUSSION

The designers of adaptation algorithms usually derive the algorithms to minimize estimation error. But often the designer has additional criteria for the algorithm to satisfy, which requires modification of the algorithm. The underlying concept of this paper is that the adaptation algorithm should be derived from a function which explicitly includes terms involving all of the design criteria.

This paper has demonstrated the principle by combining a set of soft linear constraints with a mean square error criterion. Once the performance function was so defined, the soft-constraint LMS algorithm was directly obtained.

It has been proved that in a stationary environment and when certain conditions are satisfied, the soft-constraint LMS algorithm causes the filter to converge, minimizing the performance function. It was also shown that setting the adaptation constant to obey the conditions for convergence in the mean was not always sufficient to obtain good behavior; in some cases more restrictive conditions must be observed. Since these conditions depend upon considerable *a priori* knowledge of the environment, which is generally not available, an even more restrictive condition was proposed which has the advantage of depending on the environment only in that the total input power (a measurable quantity) must be known.

The usefulness of the soft-constraint LMS algorithm has been shown by applying it to an adaptive antenna array (section X). The constraints were changed in strength, with resulting changes in the

directivity pattern of the array and in its response to incoming sinusoidal signals. This example demonstrated the effect of varying the "stiffness" of a constraint. A constraint in an important location (direction of possible signal arrival) can be stiffened so that deviations from it remain small, while constraints in less important locations can be made softer so that larger variations are permitted. This may in some cases prove advantageous as a "trade-off" in return for maintaining a close approach to minimization of estimation error. This flexibility of the soft-constraint LMS algorithm could be an advantage in building adaptive antenna arrays which attenuate strong jammers while maintaining reception in directions where desirable signals are expected to appear.

This paper has also derived a relation between the output power of a signal from a converged soft constraint LMS adaptive filter and the signal input power. This relation demonstrated the unexpected behavior that in some cases, although the input power is increasing monotonically, the output power could increase, then decrease to zero, and then increase again. Another interesting case was shown where the output power increases to a peak, and then decreases monotonically, while the input power is increasing monotonically. It is possible that useful applications of these output power phenomena exist; this is an area for future research. In addition, no physical interpretation for the matrix Ψ used in the development has been presented; some properties of Ψ are given at the end of appendix F. This remains an area for further study.

It was shown that the LMS, leaky LMS, and Zahm's algorithms are all special cases of the soft-constraint LMS algorithm. Thus, results for the soft-constraint LMS algorithm also hold for these previous algorithms. It was also shown that the optimum solution to a soft constraint problem approaches the optimum solution of a hard constraint problem as the stiffness of the soft constraints goes to infinity.

Thus the soft-constraint LMS algorithm is a generalization of several existing algorithms. It has potential usefulness in a number of areas.

APPENDIX A

PROOF OF THEOREM 1: CONVERGENCE OF THE MEAN WEIGHT VECTOR OF THE SOFT-CONSTRAINT LMS ALGORITHM

The theorem statement is:

Theorem 1: Convergence of the Mean Weight Vector

If 1) The soft-constraint LMS algorithm (8-7) or (8-8) produces a weight vector sequence $W(j)$ from a data vector sequence $X(j)$ and a desired signal sequence $d(j)$, and if

2) $W(j)$ and $X(j)$ are statistically independent, and if

3) The matrix $\underline{R}+\underline{A}^T\underline{B}\underline{A}$ is nonsingular, and if

4) $0 < \mu < \frac{1}{\lambda_{max}}$,

Then the mean weight vector converges to the optimum weight vector:

$$\lim_{j\to\infty} E\{W(j)\} = W_{opt} \quad . \tag{A-1}$$

The proof requires an expression for $E\{W(j)\}$. The update algorithm for $W(j)$ is the soft-constraint LMS algorithm (8-8):

$$W(j+1) = (\underline{I}-2\mu\underline{A}^T\underline{B}\underline{A})W(j) + 2\mu e(j)X(j) + 2\mu\underline{A}^T\underline{B}H \quad . \tag{A-2}$$

Expanding $e(j)$ using (3-4) and regrouping terms yields:

$$W(j+1) = \{\underline{I}-2\mu[X(j)X^T(j)+\underline{A}^T\underline{B}\underline{A}]\}W(j) + 2\mu[d(j)X(j)+\underline{A}^T\underline{B}H] \quad . \tag{A-3}$$

Taking the expectation of this update equation, and using the assumption that X(j) and W(j) are independent random processes yields:

$$E\{W(j+1)\} = [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]E\{W(j)\} + 2\mu(P+\underline{A}^T\underline{B}H) \quad . \qquad (A\text{-}4)$$

This recursion equation is identical (after regrouping of terms) to the recursion equation which is obtained for the weight vector when perfect gradient measurements are available (7-2). Thus the mean of the weight vector follows the trajectory that is obtained when perfect gradient measurements are available.

Iterating (A-4) yields the relation:

$$E\{W(j)\} = [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]^j W(0)+2\mu\{\sum_{t=0}^{j-1}[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]^t\}(P+\underline{A}^T\underline{B}H) \quad . \qquad (A\text{-}5)$$

The summation can be replaced by use of the matrix identity:

$$\sum_{t=0}^{j-1}\underline{M}^t = (\underline{I}-\underline{M}^j)(\underline{I}-\underline{M})^{-1} \quad . \qquad (A\text{-}6)$$

Using this identity in (A-5) results in:

$$\begin{aligned} E\{W(j)\} &= [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]^j W(0) \\ &\quad + 2\mu\{\underline{I}-[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]^j\}\{\underline{I}-[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\}^{-1}(P+\underline{A}^T\underline{B}H) \\ &= [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]^j[W(0)-(\underline{R}+\underline{A}^T\underline{B}\underline{A})^{-1}(P+\underline{A}^T\underline{B}H)] \\ &\quad + (\underline{R}+\underline{A}^T\underline{B}\underline{A})^{-1}(P+\underline{A}^T\underline{B}H) \quad . \qquad (A\text{-}7) \end{aligned}$$

Recalling the expression for the optimum weight vector (6-3) and substituting this relationship in (A-7) yields:

$$E\{W(j)\} = [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]^j[W(0)-W_{opt}] + W_{opt} \quad . \qquad (A\text{-}8)$$

Then (A-1) will be true for all W(0) if and only if

$$\lim_{j\to\infty} [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]^j = \underline{0} \quad . \qquad (A\text{-}9)$$

This is true if and only if the magnitude of every eigenvalue of the matrix $[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]$ is less than one (by the assumptior of nonsingularity, every eigenvalue is nonzero). This condition is written as:

$$|1-2\mu\lambda_t| < 1 \qquad t=1, \ldots, n \qquad (A\text{-}10)$$

where λ_t is the t^{th} eigenvalue of the matrix $\underline{R}+\underline{A}^T\underline{B}\underline{A}$. This condition will be satisfied if and only if

$$0 < \mu < \frac{1}{\lambda_t} \qquad \text{for all } t,\ t=1, \ldots, n \quad . \qquad (A\text{-}11)$$

Since

$$\frac{1}{\lambda_{max}} \leq \frac{1}{\lambda_t} \qquad \text{for all } t=1, \ldots, n \qquad (A\text{-}12)$$

the required condition is

$$0 < \mu < \frac{1}{\lambda_{max}} \quad . \qquad (A\text{-}13)$$

When this condition is satisfied, (A-9) is true, so that under the stated assumptions the conclusion (A-1) of Theorem 1 is true.

APPENDIX B

PROOF OF THEOREM 2: WEIGHT VECTOR COVARIANCE MATRIX RECURSION

The statement of Theorem 2 is:

Theorem 2: Weight Vector Covariance

If

1) Theorem 1 holds, and if

2) W(j) and d(j) are statistically independent, and if

3) d(j) and u(j) are gaussianly distributed,

then the recursion equation for the weight vector covariance is:

$$\underline{C}_{WW}(j+1) = [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\underline{C}_{WW}(j)[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]$$
$$+ 4\mu^2\Big\{\underline{R}\underline{C}_{WW}(j)\underline{R} + \underline{R}\mathrm{Tr}[\underline{C}_{WW}(j)\underline{R}] + \underline{R}E\{e^2(j)|_{W=\overline{W}(j)}\}$$
$$+ [P-\underline{R}\overline{W}(j)][P-\underline{R}\overline{W}(j)]^T\Big\} . \qquad (B\text{-}1)$$

The proof begins by recalling the recursion expressions for the weight vector W(j) and the mean weight vector $\overline{W}(j)$. The expression for W(j) from (8-7) is:

$$W(j+1) = W(j) + 2\mu e(j)X(j) - 2\mu\underline{A}^T\underline{B}[\underline{A}W(j)-H] . \qquad (B\text{-}2)$$

The expression for $\overline{W}(j)$ from rearranging (A-4) is:

$$\overline{W}(j+1) = \overline{W}(j) + 2\mu P - 2\mu\underline{R}\overline{W}(j) - 2\mu\underline{A}^T\underline{B}\underline{A}\overline{W}(j) + 2\mu\underline{A}^T\underline{B}H. \qquad (B\text{-}3)$$

Define the difference between the weight vector and its mean by:

$$\Delta W(j) = W(j) - \overline{W}(j) \qquad (B\text{-}4)$$

Combining (B-2) and (B-3) results in a recursion equation for $\Delta W(j)$:

$$\Delta W(j+1) = \Delta W(j) + 2\mu[d(j)X(j)-P]$$

$$- 2\mu[X(j)X^T(j)W(j)-\underline{R}\bar{W}(j)] - 2\mu\underline{A}^T\underline{BA}\Delta W(j) \quad . \quad (B-5)$$

Using recursion (B-5) in the definition of a covariance matrix yields:

$$C_{WW}(j+1) = E\{[W(j+1)-\bar{W}(j+1)][W(j+1)-\bar{W}(j+1)]^T\}$$

$$= E\{\Delta W(j+1)\Delta W^T(j+1)\}$$

$$= E\{\Delta W(j)\Delta W^T(j) + 2\mu\Delta W(j)[d(j)X(j)-P]^T$$

$$- 2\mu\Delta W(j)[W^T(j)X(j)X^T(j)-\bar{W}^T(j)\underline{R}] - 2\mu\Delta W(j)\Delta W^T(j)\underline{A}^T\underline{BA}$$

$$+ 2\mu[d(j)X(j)-P]\Delta W^T(j) + 4\mu^2[d(j)X(j)-P][d(j)X(j)-P]^T$$

$$- 4\mu^2[d(j)X(j)-P][W^T(j)X(j)X^T(j)-\bar{W}^T(j)\underline{R}]$$

$$- 4\mu^2[d(j)X(j)-P]\Delta W^T(j)\underline{A}^T\underline{BA}$$

$$- 2\mu[X(j)X^T(j)W(j)-\underline{R}\bar{W}(j)]\Delta W^T(j)$$

$$- 4\mu^2[X(j)X^T(j)W(j)-\underline{R}\bar{W}(j)][d(j)X(j)-P]^T$$

$$+ 4\mu^2[X(j)X^T(j)W(j)-\underline{R}\bar{W}(j)][W^T(j)X(j)X^T(j)-\bar{W}^T(j)\underline{R}]$$

$$+ 4\mu^2[X(j)X^T(j)W(j)-\underline{R}\bar{W}(j)]\Delta W^T(j)\underline{A}^T\underline{BA}$$

$$- 2\mu\underline{A}^T\underline{BA}\Delta W(j)\Delta W^T(j) - 4\mu^2\underline{A}^T\underline{BA}\Delta W(j)[d(j)X(j)-P]^T$$

$$- 4\mu^2\underline{A}^T\underline{BA}\Delta W(j)[W^T(j)X(j)X^T(j)-\bar{W}^T(j)\underline{R}]$$

$$- 4\mu^2\underline{A}^T\underline{BA}\Delta W(j)\Delta W^T(j)\underline{A}^T\underline{BA}\} \quad . \quad (B-6)$$

Now using the assumptions of independence, and noting that $E\{\Delta W(j)\}=0$, terms 2, 5, 8, and 14 become zero. Regrouping terms then yields:

$$\begin{aligned}\underline{C}_{WW}(j+1) = {} & [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\underline{C}_{WW}(j)[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})] - 4\mu^2\underline{R}\underline{C}_{WW}(j)\underline{R} \\ & + E\{4\mu^2[d^2(j)X(j)X^T(j)-PP^T] \\ & - 4\mu^2[d(j)X(j)-P][W^T(j)X(j)X^T(j)-\overline{W}^T(j)\underline{R}] \\ & - 4\mu^2[X(j)X^T(j)W(j)-\underline{R}\overline{W}(j)][d(j)X(j)-P]^T \\ & + 4\mu^2[X(j)X^T(j)W(j)W^T(j)X(j)X^T(j)-\underline{R}\overline{W}(j)\overline{W}^T(j)\underline{R}]\} \quad \text{(B-7)}\end{aligned}$$

Now it can be shown that for $W(j)$, $d(j)$, and $X(j)$ assumed gaussian:

$$E\{d^2(j)X(j)X^T(j)\} = E\{d^2(j)\}\underline{R} + 2PP^T \quad , \qquad \text{(B-8)}$$

$$E\{d(j)X(j)W^T(j)X(j)X^T(j)\} = P\overline{W}^T(j)\underline{R} + \underline{R}\overline{W}(j)P^T + \underline{R}P^T\overline{W}(j) \quad , \qquad \text{(B-9)}$$

$$\begin{aligned}E\{X(j)X^T(j)W(j)W^T(j)X(j)X^T(j)\} = {} & 2\underline{R}\underline{C}_{WW}(j)\underline{R} \\ & + 2\underline{R}\overline{W}(j)\overline{W}^T(j)\underline{R} \\ & + \underline{R}\,\mathrm{Tr}[\underline{C}_{WW}(j)\underline{R}] \\ & + \underline{R}\,\mathrm{Tr}[\overline{W}(j)\overline{W}^T(j)\underline{R}] \quad . \qquad \text{(B-10)}\end{aligned}$$

These relations are shown by expanding each element of the matrices on the left hand sides individually by using the summations implied by the matrix notation on the right hand sides, applying the expression for the expectation of 4 jointly distributed gaussian random variables (eq. 7.2-15 of [23]), and reconstructing the matrices.

Now, applying relations (B-8) through (B-10) to equation (B-7), cancelling terms, and regrouping yields:

$$\begin{aligned}\underline{C}_{WW}(j+1) = {} & [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\underline{C}_{WW}(j)[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})] \\ & + 4\mu^2\{\underline{R}\underline{C}_{WW}(j)\underline{R} + \underline{R}\mathrm{Tr}[\underline{C}_{WW}(j)\underline{R}] \\ & \quad + \underline{R}[E\{d^2(j)\} - 2P^T\bar{W}(j) + \bar{W}^T(j)\underline{R}\bar{W}(j)] \\ & \quad + [PP^T-\underline{R}\bar{W}(j)P^T-P\bar{W}^T(j)\underline{R}+\underline{R}\bar{W}^T(j)\underline{R}]\} \quad . \qquad (B\text{-}11)\end{aligned}$$

Now, the mean square error evaluated at the mean weight vector is $E\{d^2(j)\}-2P^T\bar{W}(j)+\bar{W}^T(j)\underline{R}\bar{W}(j)$. Substituting this relation yields the theorem's conclusion:

$$\begin{aligned}\underline{C}_{WW}(j+1) = {} & [\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\underline{C}_{WW}(j)[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})] \\ & + 4\mu^2\{\underline{R}\underline{C}_{WW}(j)\underline{R} + \underline{R}\mathrm{Tr}[\underline{C}_{WW}(j)\underline{R}] \\ & \quad + \underline{R}E\{e^2(j)|_{W=W(j)}\} \\ & \quad + [P-\underline{R}\bar{W}(j)][P-\underline{R}\bar{W}(j)]^T\} \quad . \qquad (B\text{-}12)\end{aligned}$$

This concludes the proof of the theorem.

APPENDIX C

PROOF OF THEOREM 3: SUFFICIENT CONDITIONS FOR BOUNDEDNESS OF THE TRACE OF THE WEIGHT COVARIANCE MATRIX

The statement of Theorem 3 is:

Theorem 3: Sufficient conditions for boundedness of the trace of the trace of the weight vector covariance matrix

1) If Theorem 2 holds, and

2) if a) $\gamma_{max}^2 + \gamma_{max} Tr[\underline{R}] \leq \lambda_{max}\lambda_{min}$

and

$$0 < \mu < \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{max}^2 + \gamma_{max} Tr[\underline{R}]} \quad ;$$

or if b) $\gamma_{max}^2 + \gamma_{max} Tr[\underline{R}] \geq \lambda_{max}\lambda_{min}$

and

$$0 < \mu < \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{max}^2 + \gamma_{max} Tr[\underline{R}]} \quad ;$$

Then $Tr[\underline{C}_{WW}(j)]$ will be bounded for all time.

To determine conditions under which the weight vector covariance matrix is guaranteed to remain finite, consider the recursion of the trace of the weight covariance matrix (Eq. 9-7).

First, for any matrices $\underline{E}$, $\underline{F}$, and $\underline{G}$:

$$\mathrm{Tr}[\underline{EFG}] = \mathrm{Tr}[\underline{GEF}] \quad . \tag{C-1}$$

Applying this relation to (9-7) yields:

$$\begin{aligned}\mathrm{Tr}[\underline{C}_{WW}(j+1)] = {} & \mathrm{Tr}\{[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})][\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]\underline{C}_{WW}(j)\} \\ & + 4\mu^2\{\mathrm{Tr}[\underline{R}\underline{R}\underline{C}_{WW}(j)] + \mathrm{Tr}[\underline{R}]\mathrm{Tr}[\underline{R}\underline{C}_{WW}(j)] \\ & + \mathrm{Tr}[\underline{R}]E\{e^2(j)|_{W=\overline{W}(j)}\} \\ & + [P-\underline{R}\overline{W}(j)]^T[P-\underline{R}\overline{W}(j)]\} \quad .\end{aligned} \tag{C-2}$$

Now Moschner showed (relation 2.10 in [24]) that for $\underline{F}$ a real symmetric matrix and $\underline{G}$ a positive semidefinite matrix that:

$$\mathrm{Tr}[\underline{FG}] \leq \max\{\mathrm{eig}\{\underline{F}\}\}\mathrm{Tr}[\underline{G}], \tag{C-3}$$

$$\mathrm{Tr}[\underline{FG}] \geq \min\{\mathrm{eig}\{\underline{F}\}\}\mathrm{Tr}[\underline{G}]. \tag{C-4}$$

By repeated application of (C-3) on (C-2) the trace of $\underline{C}_{WW}(j+1)$ may be bounded:

$$\begin{aligned}\mathrm{Tr}[\underline{C}_{WW}(j+1)] \leq {} & \beta_{max}^2\mathrm{Tr}[\underline{C}_{WW}(j)] \\ & + 4\mu^2\{\gamma_{max}^2\mathrm{Tr}[\underline{C}_{WW}(j)] + \gamma_{max}\mathrm{Tr}[\underline{R}]\mathrm{Tr}[\underline{C}_{WW}(j)] \\ & + \mathrm{Tr}[\underline{R}]E\{e^2(j)|_{W=\overline{W}(j)}\} \\ & + [P-\underline{R}\overline{W}(j)]^T[P-\underline{R}\overline{W}(j)]\} .\end{aligned} \tag{C-5}$$

$$\mathrm{Tr}[\underline{C}_{WW}(j+1)] \leq (\beta_{max}^2 + 4\mu^2\gamma_{max}^2 + \gamma_{max}\mathrm{Tr}[\underline{R}])\mathrm{Tr}[\underline{C}_{WW}(j)]$$
$$+ 4\mu^2\{\mathrm{Tr}[\underline{R}]E\{e^2(j)|_{W=\overline{W}(j)}\}$$
$$+ [P-\underline{R}\overline{W}(j)]^T[P-\underline{R}\overline{W}(j)]\} \qquad (C\text{-}6)$$

where

$$\beta_{max} = \max\{|\mathrm{eig}\{\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})\}|\} \qquad (C\text{-}7)$$

$$\gamma_{max} = \max\{\mathrm{eig}\{\underline{R}\}\} \quad . \qquad (C\text{-}8)$$

This inequality fits the form:

$$\mathrm{Tr}[\underline{C}_{WW}(j+1)] \leq a\mathrm{Tr}[\underline{C}_{WW}(j)] + c(j) \qquad (C\text{-}9)$$

where a is a positive constant, and c(j) is bounded in value. From linear system theory it is known that $\mathrm{Tr}[\underline{C}_{WW}(j)]$ will remain bounded when $a < 1$. Therefore if

$$\beta_{max}^2 + 4\mu^2\gamma_{max}^2 + 4\mu^2\gamma_{max}\mathrm{Tr}[\underline{R}] < 1 \qquad (C\text{-}10)$$

then $\mathrm{Tr}[\underline{C}_{WW}(j)]$ will remain bounded.

To evaluate the inequality of (C-10) requires knowledge of β_{max}. β_{max} is the absolute value of the eigenvalue of $[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]$ which has the greatest magnitude. Now, the eigenvalues of $[\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})]$ are $1-2\mu\lambda_i$, where the λ_i are the eigenvalues of $\underline{R}+\underline{A}^T\underline{B}\underline{A}$. The maximum eigenvalue is $1-2\mu\lambda_{min}$, and minimum eigenvalue is $1-2\mu\lambda_{max}$, because $\lambda_i > 0$. Thus β_{max} is the absolute value of one of these expressions:

$$\beta_{max} = \max\{|1-2\mu\lambda_{min}|, |1-2\mu\lambda_{max}|\} \quad . \qquad (C\text{-}11)$$

Now, since $\lambda_{max} > \lambda_{min}$, $1-2\mu\lambda_{max} < 1-2\mu\lambda_{min}$. The only way that $|1-2\mu\lambda_{max}|$ might be greater than $|1-2\mu\lambda_{min}|$ is when $1-2\mu\lambda_{max}$ is negative but has a greater magnitude than $1-2\mu\lambda_{min}$. Thus the requirement is $1-2\mu\lambda_{min} > 2\mu\lambda_{max}-1$ for $\beta_{max} = 1-2\mu\lambda_{min}$. Solving this inequality for μ finally yields the result that

$$\beta_{max} = \begin{cases} 1-2\mu\lambda_{min}, & \text{if } \mu \leq \dfrac{1}{\lambda_{max}+\lambda_{min}} \\ 2\mu\lambda_{max}-1, & \text{if } \mu \geq \dfrac{1}{\lambda_{max}+\lambda_{min}} \end{cases} \quad . \tag{C-12}$$

These expressions for β_{max} are now used to determine where (C-10) holds true.

Case 1: when $\mu \leq \dfrac{1}{\lambda_{max}+\lambda_{min}}$, substitution for β_{max} in (C-10) results in:

$$(1-2\mu\lambda_{min})^2 + 4\mu^2\gamma_{max}^2 + 4\mu\gamma_{max}\mathrm{Tr}[\underline{R}] < 1 \tag{C-13}$$

which can be solved to yield the additional condition on μ:

$$\mu < \frac{\lambda_{min}}{\lambda_{min}^2+\gamma_{max}^2+\gamma_{max}\mathrm{Tr}[\underline{R}]} \quad . \tag{C-14}$$

Thus if

$$0 \leq \mu < \min\left(\frac{1}{\lambda_{max}+\lambda_{min}}, \frac{\lambda_{min}}{\lambda_{min}^2+\gamma_{max}^2+\gamma_{max}\mathrm{Tr}[\underline{R}]}\right) \tag{C-15}$$

then $a < 1$, and $\mathrm{Tr}[\underline{C}_{WW}(j)]$ is guaranteed to remain bounded.

Case 2: when $\mu \geq \frac{1}{\lambda_{max}+\lambda_{min}}$, substituting β_{max} into (C-10) results in:

$$(2\mu\lambda_{max}-1)^2 + 4\mu^2\gamma_{max}^2 + 4\mu^2\gamma_{max}Tr[\underline{R}] < 1 \qquad (C-16)$$

which yields the conditions on μ of:

$$\mu < \frac{\lambda_{max}}{\lambda_{max}^2+\gamma_{max}^2+\gamma_{max}Tr[\underline{R}]} \quad . \qquad (C-17)$$

Thus if

$$\frac{1}{\lambda_{max}+\lambda_{min}} \leq \mu < \frac{\lambda_{max}}{\lambda_{max}^2+\gamma_{max}^2+\gamma_{max}Tr[\underline{R}]} \qquad (C-18)$$

then $a < 1$, and $Tr[\underline{C}_{WW}(j)]$ will remain bounded.

The conditions on μ derived in the two cases may be manipulated to yield a more acceptable form. Note that if:

$$\gamma_{max}^2+\gamma_{max}Tr[\underline{R}] \leq \lambda_{max}\lambda_{min} \qquad (C-19)$$

then the condition of the first case is satisfied when μ is in the range

$$0 \leq \mu < \frac{\lambda_{min}}{\lambda_{min}^2+\gamma_{max}^2+\gamma_{max}Tr[\underline{R}]} \qquad (C-20)$$

and the conditions for the second case are never satisfied. This proves the first half of the theorem's conclusion.

Now consider the situation when

$$\gamma_{max}^2+\gamma_{max}Tr[\underline{R}] \geq \lambda_{max}\lambda_{min} \quad . \qquad (C-21)$$

Then the conditions of the first case are satisfied when

$$0 \leq \mu < \frac{1}{\lambda_{max}+\lambda_{min}} \tag{C-22}$$

and the conditions of the second case are satisfied when

$$\frac{1}{\lambda_{max}+\lambda_{min}} \leq \mu < \frac{\lambda_{max}}{\lambda_{max}^2+\gamma_{max}^2+\gamma_{max}Tr[\underline{R}]} \quad . \tag{C-23}$$

These two ranges for μ may be combined to yield the single range of

$$0 \leq \mu < \frac{\lambda_{max}}{\lambda_{max}^2+\gamma_{max}^2+\gamma_{max}Tr[\underline{R}]} \quad . \tag{C-24}$$

This is the proof of the second half of the theorem's conclusion.

APPENDIX D

PROOF OF THEOREM 4: NECESSARY CONDITIONS FOR BOUNDEDNESS OF THE TRACE OF THE WEIGHT VECTOR COVARIANCE MATRIX

The statement of Theorem 4 is:

Theorem 4: Necessary conditions for boundedness of the trace of the weight vector covariance matrix

For $Tr[\underline{C}_{WW}(j)]$ to be a bounded sequence, it is necessary that

1) Theorem 2 hold, and also that

2a) When $\gamma_{min}^2 + \gamma_{min} Tr[\underline{R}] \geq \lambda_{max}^2$,

that $0 \leq \mu < \dfrac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{min}^2 + \gamma_{min} Tr[\underline{R}]}$;

b) When $\lambda_{min}^2 \leq \gamma_{min}^2 + \gamma_{min} Tr[\underline{R}] \leq \lambda_{max}^2$,

that $0 \leq \mu < \dfrac{1}{\sqrt{\gamma_{min}^2 + \gamma_{min} Tr[\underline{R}]}}$

c) When $\gamma_{min}^2 + \gamma_{min} Tr[\underline{R}] \leq \lambda_{min}^2$,

that $0 \leq \mu < \dfrac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{min}^2 + \gamma_{min} Tr[\underline{R}]}$.

To find a necessary condition for $Tr[\underline{C}_{WW}(j)]$ to remain bounded, consider finding a sequence $\delta(j)$ which is a lower bound for the sequence $Tr[\underline{C}_{WW}(j)]$. Then the sequence $\delta(j)$ must remain bounded for $Tr[\underline{C}_{WW}(j)]$, to be bounded, since if $\delta(j)$ grows without bound, and $\delta(j) \leq Tr[\underline{C}_{WW}(j)]$, then $Tr[\underline{C}_{WW}(j)]$ will also grow without bound. Applying (C-1) and then (C-4) to Eq. (9-7) in a similar manner as in Appendix C yields:

$$\begin{aligned} Tr[\underline{C}_{WW}(j+1)] \geq\ & \beta_{min}^2 Tr[\underline{C}_{WW}(j)] + 4\mu^2\gamma_{min}^2 Tr[\underline{C}_{WW}(j)] \\ & + 4\mu^2\gamma_{min} Tr[\underline{R}]Tr[\underline{C}_{WW}(j)] \\ & + 4\mu^2\Big\{Tr[\underline{R}]E\{e^2(j)|_{W=\overline{W}(j)}\} \\ & \quad + [P-\underline{R}\overline{W}(j)]^T [P-\underline{R}\overline{W}(j)]\Big\} \end{aligned} \tag{D-1}$$

where

$$\beta_{min} = \min\{|eig\{\underline{I}-2\mu(\underline{R}+\underline{A}^T\underline{B}\underline{A})\}|\} \tag{D-2}$$

$$\gamma_{min} = \min\{eig\{\underline{R}\}\} \quad . \tag{D-3}$$

Define the recursion for $\delta(j)$ as

$$\begin{aligned} \delta(j+1) =\ & (\beta_{min}^2+4\mu^2\gamma_{min}^2+4\mu^2\gamma_{min}Tr[\underline{R}])\delta(j) \\ & + 4\mu^2[Tr[\underline{C}_{WW}(j)]E\{e^2(j)|_{W=\overline{W}(j)}\} \\ & \quad + (P-\underline{R}\overline{W}(j))^T(P-\underline{R}\overline{W}(j))] \end{aligned} \tag{D-4}$$

and

$$\delta(0) = Tr[\underline{C}_{WW}(0)] \quad . \tag{D-5}$$

The equation for $\delta(j)$ is of the form:

$$\delta(j+1) = a\delta(j) + c(j) \qquad (D-6)$$

where a is a positive constant, and $c(j)$ is a positive bounded sequence. Again from linear system theory, $\delta(j)$ will be bounded if and only if $a<1$. Evaluation of a requires knowledge of β_{min}. Recall that $1-2\mu\lambda_{min} > 1-2\mu\lambda_{max}$. Therefore, if $1-2\mu\lambda_{max} > 0$, $\beta_{min}=1-2\mu\lambda_{max}$. However, if $1-2\mu\lambda_{min} < 0$, then $\beta_{min} = 2\mu\lambda_{min}-1$. Otherwise, it is possible that for some i, $1-2\mu\lambda_i = 0$, so $\beta_{min}=0$ in this case. Thuse

$$\beta_{min} = \begin{cases} 1-2\mu\lambda_{max} & \text{if } \mu \leq \dfrac{1}{2\lambda_{max}} \\ 2\mu\lambda_{min}-1 & \text{if } \mu \geq \dfrac{1}{2\lambda_{min}} \\ 0 & \text{otherwise} \end{cases} . \qquad (D-7)$$

Case 1: Consider the value of a when $\mu \leq \dfrac{1}{2\lambda_{max}}$. Then $\beta_{min}=1-2\mu\lambda_{max}$, resulting in

$$a = (1-2\mu\lambda_{max})^2 + 4\mu^2\gamma_{min}^2 + 4\mu^2\gamma_{min}Tr[\underline{R}] < 1 \quad . \qquad (D-8)$$

Solving for μ yields

$$\mu < \frac{\lambda_{max}}{\lambda_{max}^2+\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]} \quad . \qquad (D-9)$$

Combining the two conditions on μ yields the result that if

$$\mu < \min\left(\frac{\lambda_{max}}{\lambda_{max}^2+\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]}\ ,\ \frac{1}{2\lambda_{max}}\right) \qquad (D\text{-}10)$$

then $a < 1$.

Case 2: Now consider the case when $\frac{1}{2\lambda_{max}} \leq \mu \leq \frac{1}{2\lambda_{min}}$. Then $\beta_{min}=0$, and the condition on a is:

$$a = 4\mu^2\gamma_{min}^2 + 4\mu^2\gamma_{min}Tr[\underline{R}] < 1 \quad . \qquad (D\text{-}11)$$

Solving for μ yields

$$\mu < \frac{1}{2\sqrt{\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]}} \quad . \qquad (D\text{-}12)$$

The conclusion is that if

$$\frac{1}{2\lambda_{max}} \leq \mu < \min\left(\frac{1}{2\sqrt{\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]}}\ ,\ \frac{1}{2\lambda_{min}}\right)$$

then $a < 1$. (Note: this requires that

$$\frac{1}{2\lambda_{max}} < \frac{1}{2\sqrt{\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]}} \qquad (D\text{-}13)$$

or else no range exists.)

Case 3: Finally, consider the remaining case when $\mu \geq \frac{1}{2\lambda_{min}}$. Then $\beta_{min} = 2\mu\lambda_{min}-1$, and the condition on a is written:

$$a = (2\mu\lambda_{min}-1)^2 + 4\mu^2\gamma_{min}^2 + 4\mu\gamma_{min}Tr[\underline{R}] < 1 \quad . \tag{D-14}$$

Solving for μ yields

$$\mu < \frac{\lambda_{min}}{\lambda_{min}^2+\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]} \quad . \tag{D-15}$$

Thus, if

$$\frac{1}{2\lambda_{min}} \leq \mu < \frac{\lambda_{min}}{\lambda_{min}^2+\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]} \tag{D-16}$$

then $a < 1$. Note that if

$$\frac{\lambda_{min}}{\lambda_{min}^2+\gamma_{min}^2+\gamma_{min}Tr[\underline{R}]} < \frac{1}{2\lambda_{min}} \tag{D-17}$$

then $a \geq 1$ whenever $\mu \geq \frac{1}{2\lambda_{min}}$.

Grouping these results for the three cases of μ together in a manner similar to that used in the proof of Theorem 3 yields the result:

$a < 1$, if:

1) When $\gamma_{min}^2 + \gamma_{min}Tr[\underline{R}] \geq \lambda_{max}^2$,

$$0 \leq \mu < \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{min}^2 + \gamma_{min}Tr[\underline{R}]} ;$$

2) When $\lambda_{min}^2 \leq \gamma_{min}^2 + \gamma_{min}Tr[\underline{R}] \leq \lambda_{max}^2$,

$$0 \leq \mu < \frac{1}{2\sqrt{\gamma_{min}^2 + \gamma_{min}Tr[\underline{R}]}} ;$$

AD-A081 388 SYRACUSE UNIV NY F/G 9/5
RESEARCH ON LARGE ADAPTIVE ARRAYS.(U)
DEC 79 B WIDROW, R CHESTEK, H MESIWALA F30602-78-C-0083
UNCLASSIFIED RADC-TR-79-308 NL

2 OF 2
AD A081-6

END
DATE FILMED
4-80
DTIC

3) When $\gamma_{min}^2 + \gamma_{min} Tr[\underline{R}] \leq \lambda_{min}^2$

$$0 \leq \mu \leq \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{min}^2 + \gamma_{min} Tr[\underline{R}]} \quad .$$

If none of these conditions are satisfied, then $a \geq 1$, and $\delta(j)$ will be an unbounded sequence, which implies that $Tr[\underline{C}_{WW}(j)]$ is an unbounded sequence. Thus satisfaction of the above conditions is a necessary (but not sufficient) condition for $Tr[\underline{C}_{WW}(j)]$ to be a bounded sequence.

APPENDIX E

DERIVATION OF A CALCULABLE BOUND ON μ WHICH SATISFIES THE CONDITIONS FOR CONVERGENCE OF THE MEAN WEIGHT VECTOR AND GUARANTEED BOUNDEDNESS OF THE WEIGHT VECTOR COVARIANCE MATRIX

Section 9 proposes a bound on μ which is calculable and satisfies the conditions on μ in Theorems 1 and 3. This appendix demonstrates that the bound satisfies these criteria.

The bound (9-8) proposed is:

$$0 < \mu < \frac{1}{3\mathrm{Tr}[\underline{R}] + \mathrm{Tr}[\underline{A}^T\underline{B}\underline{A}]} \quad . \qquad (E$$

Satisfying Theorem 1.

This bound satisfies the conditions for convergence of the mean weight vector (Theorem 1) since:

$$\frac{1}{\lambda_{max}} \geq \frac{1}{\mathrm{Tr}[\underline{R}] + \mathrm{Tr}[\underline{A}^T\underline{B}\underline{A}]} \geq \frac{1}{3\mathrm{Tr}[\underline{R}] + \mathrm{Tr}[\underline{A}^T\underline{B}\underline{A}]} \quad . \qquad (E$$

Thus if

$$0 < \mu < \frac{1}{3\mathrm{Tr}[\underline{R}] + \mathrm{Tr}[\underline{A}^T\underline{B}\underline{A}]} \quad , \qquad (E$$

then

$$0 < \mu < \frac{1}{\lambda_{max}} \quad , \qquad (E$$

satisfying the condition on μ of Theorem 1.

Satisfying Theorem 3.

This bound (E-1) also satisfies the conditions on μ which guarantees boundedness of the weight vector covariance matrix (Theorem 3):

Case 1 of Theorem 3: using the condition of case 1, the following inequality may be written:

$$\frac{\lambda_{min}}{\lambda_{min}^2 + \lambda_{max}\lambda_{min}} \leq \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]} .$$

Then

$$\frac{1}{\lambda_{min} + \lambda_{max}} \leq \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]}$$

$$\frac{1}{Tr[\underline{R}] + Tr[\underline{A}^T\underline{B}\underline{A}]} \leq$$

$$\frac{1}{3Tr[\underline{R}] + Tr[\underline{A}^T\underline{B}\underline{A}]} \leq \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]} .$$

Thus if

$$0 < \mu < \frac{1}{3Tr[\underline{R}] + Tr[\underline{A}^T\underline{B}\underline{A}]} , \qquad (E-7)$$

then

$$0 < \mu < \frac{\lambda_{min}}{\lambda_{min}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]} , \qquad (E-8)$$

satisfying the condition on μ of Case 1 of Theorem 3.

Case 2 of Theorem 3: Begin with the equality:

$$\frac{1}{\lambda_{max} + \gamma_{max}\frac{\gamma_{max}}{\lambda_{max}} + \frac{\gamma_{max}}{\lambda_{max}}Tr[\underline{R}]} = \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]}$$

Now $\lambda_{max} \geq \gamma_{max}$, so

$$\frac{1}{\lambda_{max} + \gamma_{max} + Tr[\underline{R}]} \leq \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]}$$

$$\frac{1}{Tr[\underline{R}+\underline{A}^T\underline{B}\underline{A}] + Tr[\underline{R}] + Tr[\underline{R}]} \leq$$

$$\frac{1}{Tr[\underline{A}^T\underline{B}\underline{A}] + 3Tr[\underline{R}]} \leq \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]}$$

Thus if

$$0 < \mu < \frac{1}{3Tr[\underline{R}] + Tr[\underline{A}^T\underline{B}\underline{A}]} \quad . \tag{E-7}$$

Then

$$0 < \mu < \frac{\lambda_{max}}{\lambda_{max}^2 + \gamma_{max}^2 + \gamma_{max}Tr[\underline{R}]} \quad , \tag{E-8}$$

satisfying the condition on μ of case 2 of Theorem 3.

Since both cases of Theorem 3 are satisfied, the proposed bound (E-1) on μ will guarantee boundedness of the weight vector covariance matrix.

Calculability

$Tr[\underline{A}^T\underline{B}\underline{A}]$ is certainly calculable, since the matrices are all specified by the designer. $Tr[\underline{R}]$ is also computable as discussed in section IX; it is n times the input power to the filter.

APPENDIX F

SIMULTANEOUS DIAGONALIZATION OF TWO HERMITIAN MATRICES

This appendix shows that a transformation exists which simultaneously diagonalizes two hermitian matrices. This is a previously solved problem, included here for completeness [25,26]. This appendix follows Noble's [25] development. Gantmacher [26] arrives at the same results by a different path.

The theorem is stated as follows: Given two hermitian matrices $\underline{A}$ and $\underline{B}$ with $\underline{B}$ nonsingular, a matrix $\underline{S}$ exists such that $\underline{S}^{+}\underline{B}\underline{S} = \underline{I}$, the identity matrix (where $\underline{S}^{+}$ denotes the complex conjugate transpose of the matrix $\underline{S}$), and $\underline{S}^{+}\underline{A}\underline{S} = \underline{\Psi}$, where $\underline{\Psi}$ is a diagonal matrix.

The proof begins by noting that for any hermitian matrix there exists a unitary transformation which will diagonalize that hermitian matrix. Thus, for the hermitian matrix $\underline{B}$ there exists a unitary matrix $\underline{P}$ such that $\underline{P}^{+}\underline{B}\underline{P} = \underline{\Psi}_B$, where $\underline{\Psi}_B$ is the diagonal matrix with diagonal elements which are the eigenvalues of matrix $\underline{B}$.

Denote the diagonal matrix which has as elements the square roots of the eigenvalues of $\underline{B}$ as $\underline{\Psi}_B^{1/2}$, yielding the relation $(\underline{\Psi}_B^{1/2})^{+}\underline{\Psi}_B^{1/2} = \underline{\Psi}_B$. Under the assumption that $\underline{\Psi}_B$ (and thus $\underline{B}$) is nonsingular, $\underline{\Psi}_B^{1/2}$ is also nonsingular. Using the inverse of $\underline{\Psi}_B^{1/2}$ results in the relation:

$$[(\underline{\Psi}_B^{1/2})^{+}]^{-1}\underline{P}^{+}\underline{B}\underline{P}(\underline{\Psi}_B^{1/2})^{-1} = \underline{I} \quad . \tag{F-1}$$

Thus the matrix $\underline{P}(\underline{\Psi}_B^{1/2})^{-1}$ transforms $\underline{B}$ to an identity matrix.

Now consider applying this transformation to the hermitian matrix $\underline{A}$. This would result in the matrix:

$$[(\underline{\Psi}_B^{1/2})^+]^{-1}\underline{P}^+\underline{AP}(\underline{\Psi}_B^{1/2})^{-1} \quad . \tag{F-2}$$

This matrix is also hermitian. Therefore, there exists a unitary matrix $\underline{Q}$ such that the resulting unitary transformation diagonalizes the matrix of expression (F-2), yielding:

$$\underline{Q}^+[(\underline{\Psi}_B^{1/2})^+]^{-1}\underline{P}^+\underline{AP}(\underline{\Psi}_B^{1/2})^{-1}\underline{Q} = \underline{\Psi} \quad , \tag{F-3}$$

where $\underline{\Psi}$ is a diagonal matrix. Note that $\underline{P}(\underline{\Psi}_B^{1/2})^{-1}\underline{Q}$ is a transformation which will diagonalize the matrix A. Denote this transformation by $\underline{S}$:

$$\underline{S} \triangleq \underline{P}(\underline{\Psi}_B^{1/2})^{-1}\underline{Q} \quad . \tag{F-4}$$

Now apply this transformation to the matrix $\underline{B}$.

Since $\underline{P}$ was originally chosen so that $\underline{P}^+\underline{BP} = \underline{\Psi}_B$,

$$\underline{S}^+\underline{BS} = \underline{Q}^+[(\underline{\Psi}_B^{1/2})^+]^{-1}\underline{P}^+\underline{BP}(\underline{\Psi}_B^{1/2})^{-1}\underline{Q} \quad . \tag{F-5}$$

By application of (F-1)

$$\begin{aligned} \underline{S}^+\underline{BS} &= \underline{Q}^+\underline{Q} \\ &= \underline{I} \quad , \end{aligned} \tag{F-6}$$

since $\underline{Q}$ is a unitary transformation. Therefore the matrix $\underline{S}$ satisfies the two relationships

$$\underline{S}^+\underline{BS} = \underline{I} \quad , \tag{F-7}$$

$$\underline{S}^+\underline{AS} = \underline{\Psi} \quad . \tag{F-8}$$

where $\underline{\Psi}$ is a diagonal matrix and $\underline{S} = \underline{P}\underline{\Psi}_B^{1/2}\underline{Q}$, where $\underline{P}$ is a unitary matrix which diagonalizes $\underline{B}$, $\underline{\Psi}_B^{1/2}$ is a diagonal composed of the square roots of the eigenvalues of $\underline{B}$, and $\underline{Q}$ is a unitary matrix which diagonalizes $[(\underline{\Psi}_B^{1/2})^+]^{-1}\underline{P}^+\underline{A}\underline{P}(\underline{\Psi}_B^{1/2})^{-1}$.

Although deeper knowledge of $\underline{S}$ and $\underline{\Psi}$ is not required, the following properties can be shown:

1) The columns of $\underline{S}$ are the eigenvectors of the matrix $\underline{B}^{-1}\underline{A}$

2) The values of the diagonal elements of $\underline{\Psi}$ are the eigenvalues of the matrix $\underline{B}^{-1}\underline{A}$.

These properties can be shown as follows: Begin with the relation:

$$\underline{S}^+\underline{A}\underline{S} = \underline{\Psi} \quad . \tag{F-9}$$

Then

$$\underline{A}\underline{S} = (\underline{S}^+)^{-1}\underline{\Psi} \quad . \tag{F-10}$$

Now, premultiply by $\underline{B}^{-1}$:

$$\underline{B}^{-1}\underline{A}\underline{S} = \underline{B}^{-1}(\underline{S}^+)^{-1}\underline{\Psi} \quad . \tag{F-11}$$

An expression for $\underline{B}^{-1}(\underline{S}^+)^{-1}$ can be found by considering the relation:

$$\underline{S}^+\underline{B}\underline{S} = \underline{I} \quad . \tag{F-12}$$

Inverting both sides yields:

$$(\underline{S}^+\underline{B}\underline{S})^{-1} = \underline{S}^{-1}\underline{B}^{-1}(\underline{S}^+)^{-1} = \underline{I} \quad . \tag{F-13}$$

Then premultiplying by $\underline{S}$ yields:

$$\underline{B}^{-1}(\underline{S}^+)^{-1} = \underline{S} \quad . \tag{F-14}$$

Substituting this relation in (F-11) yields:

$$\underline{B}^{-1}\underline{A}\underline{S} = \underline{S}\underline{\Psi} \qquad \text{(F-15)}$$

which shows that the columns of $\underline{S}$ are the eigenvectors of $\underline{B}^{-1}\underline{A}$ and the diagonal elements of $\underline{\Psi}$ are the corresponding eigenvalues.

APPENDIX G

PROOF OF THEOREM 5: OPTIMUM WEIGHT VECTOR FOR SOFT-CONSTRAINT LMS ALGORITHM GOES TO OPTIMUM WEIGHT VECTOR FOR FROST'S HARD CONSTRAINT LMS ALGORITHM

The theorem statement is:

Theorem 5: Soft Constraint Solution Goes to Hard Constraint Solution

If

1) The weighting matrix $\underline{B}$ in the soft constraint algorithm is replaced by $\gamma\underline{B}$, so that the weighting on the constraints may be increased simultaneously by increasing the scalar γ, and if

2) The data vector autocorrelation matrix $\underline{R}$ is nonsingular and if

3) The weighting matrix $\underline{B}$ is nonsingular, and the matrix $\underline{A}$ is full rank, then

$$\lim_{\gamma \to \infty} W_{opt} = W_{hc} \quad . \tag{G-1}$$

Proof: The proof begins by writing W_{opt} in terms of $\gamma\underline{B}$ from (6-3):

$$W_{opt} = (\underline{R} + \gamma \underline{A}^T \underline{B} \underline{A})^{-1} (P + \gamma \underline{A}^T \underline{B} H) \quad .$$

Now apply the matrix inversion lemma (section 5.7 of [25]) to $\underline{R} + \gamma \underline{A}^T \underline{B} \underline{A}$:

$$W_{opt} = [\underline{R}^{-1} - \underline{R}^{-1}\underline{A}^T(\tfrac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{A}\underline{R}^{-1}]\underline{P}$$

$$+ [\underline{R}^{-1} - \underline{R}^{-1}\underline{A}^T(\tfrac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{A}\underline{R}^{-1}]\gamma\underline{A}^T\underline{B}\underline{H}$$

The conditions of the lemma are guaranteed to be satisfied in this use and in the next, due to the assumptions made on $\underline{R}$, $\underline{B}$, and on $\underline{A}$. This assumption is also required by Frost to yield a unique optimum weight vector for his algorithm. Now apply the matrix inversion lemma to $\frac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T$ in the second term:

$$W_{opt} = [\underline{R}^{-1} - \underline{R}^{-1}\underline{A}^T(\tfrac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{A}\underline{R}^{-1}]\underline{P}$$

$$+ \left\{\underline{R}^{-1} - \underline{R}^{-1}\underline{A}^T\left[(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\right.\right.$$

$$\left.\left. -(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}[\gamma\underline{B}+(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}]^{-1}(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\right]\underline{A}\underline{R}^{-1}\right\}\gamma\underline{A}^T\underline{B}\underline{H}. \quad \text{(G-3)}$$

Now factoring out $\underline{R}^{-1}$ as a postmultiplier in the first term and as a premultiplier in the second term, and multiplying the $\underline{A}^T$ factor of $\underline{A}^T\underline{B}\underline{H}$ in the second term through yields:

$$W_{opt} = [\underline{I} - \underline{R}^{-1}\underline{A}^T(\tfrac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{A}]\underline{R}^{-1}\underline{P}$$

$$+ \underline{R}^{-1}\left\{\underline{A}^T - \underline{A}^T\left[(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\right.\right.$$

$$\left.\left. -(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}[\gamma\underline{B}+(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}]^{-1}(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\right]\underline{A}\underline{R}^{-1}\underline{A}^T\right\}\gamma\underline{B}\underline{H}. \quad \text{(G-4)}$$

Since $(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}(\underline{A}\underline{R}^{-1}\underline{A}^T) = \underline{I}$, the second term reduces, yielding:

$$W_{opt} = [\underline{I} - \underline{R}^{-1}\underline{A}^T(\tfrac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{A}]\underline{R}^{-1}\underline{P}$$

$$+ \underline{R}^{-1}\left\{\underline{A}^T - \underline{A}^T\left[\underline{I} - (\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}[\gamma\underline{B}+(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}]^{-1}\right]\right\}\gamma\underline{B}\underline{H}$$

$$= [\underline{I} - \underline{R}^{-1}\underline{A}^T(\tfrac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{A}]\underline{R}^{-}$$

$$+ \underline{R}^{-1}\{\underline{A}^T - \underline{A}^T + \underline{A}^1(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}[\gamma\underline{B} + (\underline{A}$$

Now,

$$\lim_{\gamma\to\infty}(\tfrac{1}{\gamma}\underline{B}^{-1} + \underline{A}\underline{R}^{-1}\underline{A}^T)^{-1} = (\underline{A}\underline{R}^{-}$$

and

$$\lim_{\gamma\to\infty}[\gamma\underline{B} + (\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}]^{-1}\gamma\underline{B} = \lim_{\gamma\to\infty}(\quad$$

So

$$\lim_{\gamma\to\infty} W_{opt} = [\underline{I} - \underline{R}^{-1}\underline{A}^T(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}$$

$$+ \underline{R}^{-1}\underline{A}^T(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{H} \qquad (G\text{-}8)$$

and since from (2.8) of Frost [9] (when rewritten in the [illegible]

of this paper):

$$W_{hc} = [\underline{I} - \underline{R}^{-1}\underline{A}^T(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{A}\underline{R}^{-1}$$

$$+ \underline{R}^{-1}\underline{A}^T(\underline{A}\underline{R}^{-1}\underline{A}^T)^{-1}\underline{H}\ , \qquad (G\text{-}9)$$

the conclusion is

$$\lim_{\gamma\to\infty} W_{opt} = W_{hc}\ , \qquad (G\text{-}10)$$

and the theorem is proved.

REFERENCES

1. B. Widrow, "Adaptive Filters," in Aspects of Network and System Theory, R. E. Kalman and N. DeClaris, Eds. New York: Holt, Reinhart, and Winston, Inc., pp. 563-587, 1970.

2. B. Widrow and J. M. McCool, "A Comparison of Adaptive Algorithms Based on the Methods of Steepest Descent and Random Search," IEEE Trans. on Antennas and Propagation, Vol. AP-24, No. 5, pp. 615-637, Sept. 1976.

3. B. Widrow, et al, "Adaptive Noise Cancelling: Principles and Applications," Proceedings of the IEEE, Vol. 63, No. 12, pp. 1692-1716, December 1975.

4. J. R. Treichler, "Transient and Convergent Behavior of the Adaptive Line Enhancer," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-27, No. 1, pp. 53-62, Feb. 1979.

5. D. R. Morgan and S. E. Craig, "Real Time Adaptive Linear Prediction using the Least Mean Square Gradient Algorithm," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-24, No. 6, pp. 494-507, Dec. 1976.

6. R. L. Riegler and R. T. Compton, Jr., "An Adaptive Array for Interference Rejection," Proceedings of the IEEE, Vol. 61, No. 6, pp. 748-758, June 1973.

7. R. D. Gitlin, J. E. Mazo, and M. G. Taylor, "On the Design of Gradient Algorithms for Digitally Implemented Adaptive Filters," IEEE Trans. on Circuit Theory, Vol. CT-20, No. 2, pp. 125-136, March 1973.

8. L. J. Griffiths, "A Simple Adaptive Algorithm for Real-Time Processing in Antenna Arrays," Proceedings of the IEEE, Vol. 57, No. 10, pp. 1696-1704, October 1969.

9. O. L. Frost, III, "An Algorithm for Linearly Constrained Adaptive Array Processing," Proceedings of the IEEE, Vol. 60, No. 8, pp. 926-935, August 1972.

10. C. S. Williams, "Adaptive Channel Modeling for Digital Communication in the Presence of Multipath and Doppler," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, 1977.

11. J. R. Treichler, "The Spectral Line Enhancer - the Concept, an Implementation, and an Application," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University, 1977.

12. N. Ahmed, G. R. Elliot, and S. D. Stearns, "Long Term Instability Problems in Adaptive Noise Cancellers," Sandia Report SAND78-1032, Sandia Laboratories, Albuquerque, N. M. 87185, August 1978.

13. W. D. White, "Artificial Noise in Adaptive Arrays," IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-14, No. 2, pp. 380-384, March 1978.

14. C. L. Zahm, "Application of Adaptive Arrays to Suppress Strong Jammers in the Presence of Weak Signals," IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-9, No. 2, pp. 260-271, March 1973.

15. H. L. Van Trees, Detection, Estimation, and Modulation Theory, Part I, New York: John Wiley and Sons, 1968.

16. A. P. Sage and J. L. Melsa, Estimation Theory with Applications to Communication and Control, New York: McGraw-Hill, 1971.

17. T. Kailath, Lectures on Linear Least Squares Estimation, New York: Springer Verlag, 1976.

18. D. S. Wilde, Optimum Seeking Methods, Englewood Cliffs, N. J.: Prentice-Hall, 1964.

19. M. T. Wasan, Stochastic Approximation, New York: Cambridge University Press, 1969.

20. T. P. Daniell, "Adaptive Estimation with Mutually Correlated Training Sequences," IEEE Trans. on Systems Science and Cybernetics, Vol. SSC-6, No. 1, pp. 12-19, Jan. 1970.

21. L. D. Davisson, "Steady State Error in Adaptive Mean Square Minimization," IEEE Trans. on Information Theory, Vol. IT-16, No. 4, pp. 382-385, July 1970.

22. J. K. Kim and L. D. Davisson, "Adaptive Linear Estimation for Stationary M-Dependent Processes," IEEE Trans. on Information Theory, Vol. IT-21, No. 1, pp. 23-31, Jan. 1975.

23. J. L. Melsa and A. P. Sage, An Introduction to Probability and Stochastic Processes, Englewood Cliffs, N. J.: Prentice-Hall, 1973.

24. J. L. Moschner, "Adaptive Filter with Clipped Input Data," Stanford Electronics Laboratories Tech. Report No. 6796-1, Stanford University, Stanford, CA 94305, June 1970.

25. B. Noble, Applied Linear Algebra, Englewood Cliffs, N. J.: Prentice-Hall, 1969.

26. F. R. Gantmacher, The Theory of Matrices, Vol. 1, New York: Chelsea Publishing Co., 1977.