I. INTRODUCTION

The open-loop Gram-Schmidt (GS) technique for adaptive cancellation [1-6] has been shown to yield superior performance simultaneously in arithmetic efficiency, stability, and convergence times over other adaptive algorithms. Arithmetic efficiency results from using systolic processing architectures that take advantage of the GS structure. In addition, the stability of the GS algorithm is enhanced because it does not require the direct calculation of an inverse covariance matrix as does the sampled matrix inversion (SMI) algorithm [7]. Also, the GS canceler algorithm is very suitable for a nonstationary noise environment because the adaptive weights can be updated in a numerically efficient manner, using "sliding window" techniques on the input data instead of "batch" or "block" processing. Two types of batch processing techniques are concurrent processing and nonconcurrent processing. For concurrent processing, the adaptive weights are calculated from an input data set and reapplied to the same input data set. For nonconcurrent processing, the weights are applied to a different data set.

The optimal weights associated with an adaptive canceler are generally not known a priori and thus must be estimated by using finite averaging. Because of the use of estimated weights, suboptimal canceler performance results. Reed, Mallet, and Brennan [7, 8] quantified this performance for the SMI algorithm in the transient state under certain input conditions, one of these being that the input noise must be Gaussian. They mathematically demonstrated that the SMI canceler has relatively fast convergence characteristics and also that the convergence is independent of the input covariance matrix.

In this paper, we show that the GS canceler algorithm is numerically identical with the SMI algorithm in the transient state if infinite numerical accuracy is assumed (Section IV). By transient state, we mean that a finite number of time-coincident samples from channel to channel are used to obtain an estimate of the optimal weights. Thus the convergence rate or any other measures of effectiveness of the two algorithms in the transient state are identical. In addition, we reproduce many of the results of [7 and 8] by using the GS canceler structures as an analysis tool (Sections V to VIII). Also, new results are generated for the case when the input noises are not necessarily Gaussian (Section IX).

Note that the analysis presented here pertains to the adaptive processor in the "sidelobe canceler" (SLC) configuration, where the desired signal is assumed to be present only in the main channel; and auxiliary channels are used to cancel correlated noises in the main channel. However as is shown in [7], any nonconstrained, linear adaptive array processor can be transformed into an SLC configuration without...
changing the convergence properties. Hence, the results presented here apply to any nonconstrained linear adaptive array processor.

II. GS CANCELER

Consider the general N-input GS canceler structure as seen in Fig. 1(a). Let \( x_M(t), x_1(t), \ldots, x_{N-1}(t) \) represent the complex data in the 0th, 1st, \ldots, \( N-1 \)th channels, respectively. We call the leftmost input \( x_M(t) \) the main channel, and we call the remaining \( N-1 \) inputs the auxiliary channels. The signal of the main channel consists of a desired signal plus additive noise. The noise consists of internal noise plus external noise. Cancellation of the signals from external interfering sources relies on the correlation of simultaneously received signals in the main and auxiliary channels. The internal noises in each channel are assumed uncorrelated between channels. The canceler operates so as to decorrelate the auxiliary inputs one at a time from the other inputs by use of the basic two-input GS processor as is shown in Fig. 1(b). For example, Fig. 1(a) shows that \( x_{N-1}(t) \) is decorrelated with \( x_M(t), x_1(t), \ldots, x_{N-2}(t) \) in the first level of decomposition. Next, the output channel that results from decorrelating \( x_{N-1}(t) \) with \( x_{N-2}(t) \) is decorrelated with the other outputs of the first-level GSs. The decomposition proceeds until a final output channel is generated. If the decorrelation weights in each of the two-input GSs are computed from an infinite number of input samples, this output channel is totally decorrelated with the input: \( x_1(t), x_2(t), \ldots, x_{N-1}(t) \).

If there is not an infinite number of input samples, the decorrelation weights associated with each two-input GS canceler are estimated by using finite averaging. In this section we discuss two methods of processing data through the GS canceler. The first is called concurrent processing whereby the weights are estimated from a block of input data and applied back onto the same input data set. The second method is called nonconcurrent processing whereby the weights are estimated from a block of input data and applied to subsequent or previous input data. Inherent in both techniques is the "block processing" of data. Reference [8] shows that the average output noise power residue can vary quite differently, depending on whether concurrent or nonconcurrent processing is used.

We now briefly describe the concurrent and nonconcurrent GS canceler. For the concurrent canceler, let \( x_{(m)}^{(m)} \) represent the time-coincident outputs of the two-input GSs on the \( (m-1) \)th level. Then outputs of the two-input GSs at the \( m \)th level are given by

\[
x_n^{(m+1)} = x_n^{(m)} - w_n^{(m)} x_{N-m-1}^{(m)},
\]

\( n = 0, 1, \ldots, N - m - 1, \quad m = 1, 2, \ldots, N - 1. \) (1)

Note that \( x_0^{(1)} = x_0 \) and \( x_1^{(1)} = x_1, \) \( n = 1, 2, \ldots, N - 1. \)

The weight \( w_n^{(m)} \), seen in (1), is computed so as to decorrelate \( x_n^{(m)} \) with \( x_{N-m-1}^{(m)} \). For \( K \) input samples per channel, this weight is estimated as

\[
w_n^{(m)} = \frac{\sum_{k=1}^{K} x_{N-m-k}^{(m)} x_{N-m-k}^{*(m)}(k)}{\sum_{k=1}^{K} |x_{N-m-k}^{(m)}(k)|^2}
\]

where * denotes the complex conjugate and \( | \cdot | \) is the magnitude. Here \( k \) indexes the time-coincident sampled data.

For the nonconcurrent canceler, let \( X_n^{(m)} \) represent the outputs of the two-input GSs on the \( (m-1) \)th level. Then the outputs of the two-input GSs at the
mth level are given by

\[ X'_n^{(m+1)} = X'_n^{(m)} - w_n^{(m)} X'_n^{(m)}, \]

\[ n = 0, 1, \ldots, N - m - 1, \quad m = 1, 2, \ldots, N - 1 \]  \hspace{1cm} (3)

where \( X'_n^{(1)} = X_m, \quad X'_n^{(m+1)} = X_n, \quad n = 1, 2, \ldots, N - 1 \), and \( w_n^{(m)} \) is calculated by the use of (2); i.e., these weights are computed from a block of data that does not include \( X_n \).

Let \( x_0 \) and \( X_0 \) represent the additive noises in the main channel for concurrent and nonconcurrent processing, respectively. For this development unless otherwise noted we make the following assumptions.

1) The \( X_0, X_1, \ldots, X_{N-1} \) and \( x_0, X_1, \ldots, X_{N-1} \) are identically distributed Gaussian complex random variables (RVs).

2) These same RVs are samples from stationary processes with zero mean and equal variance.

3) For \( k_1 \neq k_2 \), \( x_n(k_1) \) is independent of \( x_n(k_2) \), and \( X_n(k_1) \) is independent of \( X_n(k_2) \).

4) For all \( k_1, k_2, n_1, n_2, x_n(k_1) \) is independent of \( X_n(k_2) \).

5) The desired signals are not present during weight computation for nonconcurrent processing.

6) The desired signals are not present in the auxiliary channels.

Note that in Section IX, we remove the assumption that the RVs are Gaussian.

The following definition is used often in the upcoming development. A normalized \( L \)-length multivariate complex circular Gaussian vector has \( L \) elements, each of which has real and imaginary parts that are independent Gaussian RVs with zero mean and variance equal to \( 1/2 \) (note the magnitude variance is one). In addition, the \( L \) elements are independent of one another.

III. OUTPUT MEASURES

The \( N \)-input GS canceler structures for concurrent and nonconcurrent processing are simplified by the representations as seen in Fig. 2(a). There are \( 0.5N \) \( (N - 1) \) weights computed in the GS structure. We call these weights the GS interior weights. The notation \( \text{GS}_{K,N} \) indicates that an \( N \)-input GS structure uses \( K \) samples from each channel to compute the GS interior weights in the GS structure. Note that for the nonconcurrent structure the weights are computed from the \( x_0, X_1, \ldots, X_{N-1} \) data block and applied to \( X_0, X_1, \ldots, X_{N-1} \). The 0th channel (or the far left channel in Fig. 2(a)) is always designated as the main channel, and the others are called auxiliary channels (or AUXes). The output of the concurrent (weighting) processor is denoted by \( z_{cw} \), and the output of the nonconcurrent processor is denoted by \( Z_{cw} \). We also represent the GS structure as shown in Fig. 2(b), where the \( N \) orthogonal outputs are displayed.

As previously mentioned, the input signal in the main channel consists of a desired signal \( s \) plus noise \( x_0 \). For nonconcurrent processing, it is assumed that the desired signal passes from input-to-output unperturbed and its power equals one. However, for concurrent processing the presence of the signal in the main channel causes signal cancellation through the GS canceler as reported in [8]. Because of linearity, the \( \text{GS}_{K,N} \) canceler can be decomposed as shown in Fig. 3. The left-hand \( \text{GS}_{K,N} \) canceler seen in this figure has only the desired signal in the main channel, and the right-hand \( \text{GS}_{K,N} \) has only noise \( x_0 \) in the main channel. Note that the interior weights of each \( \text{GS}_{K,N} \) are not identical because of the different main channel input in each (actually only the weights along the main channel path of each differ). We will use this decomposition when defining the output noise-to-signal power ratio (NSR).

It can be shown that for any set of GS interior weights that are estimated and applied to the auxiliary input channels there is an equivalent linear weighting of the input auxiliary channels. We denote this equivalent linear weighting by the \((N-1)\)-length
vector \( \hat{w}_a \), where

\[ \hat{w}_a = (\hat{w}_1, \ldots, \hat{w}_{N-1})^T. \] (4)

Thus, the outputs of the GS processed main channel are identical to the outputs of a main channel derived by subtracting the linear weighted auxiliary channels from the main channel input. With respect to the decomposition configuration seen in Fig. 3, we see that \( \hat{w}_a \) is actually the sum of two \((N - 1)\)-length weighting vectors, \( \hat{w}_{a,s} \) and \( \hat{w}_{a,n} \) where \( \hat{w}_{a,s} \) is the auxiliary linear weighting vector associated with only the desired signal in the main channel and \( \hat{w}_{a,n} \) is the auxiliary linear weighting vector associated with only noise \( x_0 \) in the main channel. Note that as \( K \to \infty, \hat{w}_{a,s} \to 0; \) also for nonconcurrent processing, \( \hat{w}_a = \hat{w}_{a,n} \). For the GS canceler the weighting on the main channel is constrained to be one.

Let \( \sigma_{\text{min}}^2 \) be the steady state \((K \to \infty)\) output noise power residue, and let \( \text{SNR}_{\text{opt}} \) be the steady state output signal-to-noise power ratio. Note that \( \sigma_{\text{min}}^2 \) is identical for both concurrent and nonconcurrent processing as is the steady state signal-to-noise power ratio, \( \text{SNR}_{\text{opt}} \). We define

\[ R_a = \text{Steady state} (N - 1) \times (N - 1) \text{ input noise covariance matrix of the auxiliary channels.} \]

\[ \hat{R}_a = \text{Estimated auxiliary input covariance matrix using } x_1, x_2, \ldots, x_{N-1} \text{ data (} K \text{ samples per input channel).} \]

\[ \hat{R}_X = \text{Estimated auxiliary input noise covariance matrix using } X_1, X_2, \ldots, X_{N-1}, \text{ (note, no desired signal assumed in this calculation).} \]

\[ \delta_{\text{aw}}^2 = \text{Transient output noise power associated with nonconcurrent weighting normalized by dividing by } \sigma_{\text{min}}^2. \]

\[ \text{SNR}_{\text{aw}} = \text{Transient output SNR associated with nonconcurrent weighting normalized by dividing by } \text{SNR}_{\text{opt}} \]

\[ |s'|^2 = \text{Transient output signal power associated with concurrent weighting normalized by dividing by signal power } |s|^2. \]

\[ \delta_{\text{cw}}^2 = \text{Transient output noise power associated with concurrent weighting normalized by SNR}_{\text{aw}}. \]

\[ \text{NSR}_{\text{cw}} = \text{Transient output noise-to-signal power ratio associated with concurrent weighting normalized by dividing by } \text{SNR}_{\text{opt}}. \]

Note that the last eight quantities defined are RVs. \( \text{NSR}_{\text{cw}} \) is defined as a noise-to-signal ratio because it is easier to obtain an analytical result pertaining to this quantity as opposed to the SNR.

By using the above definitions, it can be shown that

\[ \delta_{\text{aw}}^2 = \frac{|Z_{aw}|^2}{\sigma_{\text{min}}^2} = \frac{|x_0|^2 - \hat{w}_s^* \hat{R}_X \hat{w}_a}{\sigma_{\text{min}}^2} \] (5a)

and

\[ \delta_{\text{cw}}^2 = E_X \{ \delta_{\text{aw}}^2 \} = \frac{E\{|x_0|^2\} - \hat{w}_a^* \hat{R}_a \hat{w}_a}{\sigma_{\text{min}}^2} \] (5b)

where \( t \) denotes conjugate transpose, \( E\{\cdot\} \) denotes the expected value, and \( E_X \{\cdot\} \) denotes that the expectation is taken over the RVs \( X_0, X_1, \ldots, X_{N-1} \).

Furthermore,

\[ \text{SNR}_{\text{aw}} = \frac{(1/\delta_{\text{aw}}^2)}{\text{SNR}_{\text{opt}}} \] (6)

\[ \delta_{\text{cw}}^2 = \frac{|Z_{cw}|^2}{\sigma_{\text{min}}^2} = \frac{|x_0|^2 - \hat{w}_{a,n}^* \hat{R}_a \hat{w}_{a,n}}{\sigma_{\text{min}}^2} \] (7)

\[ \text{NSR}_{\text{cw}} = \frac{\delta_{\text{cw}}^2}{|s'|^2} \] (8)

We define

\[ \sigma_{\text{aw}}^2(K,N) = E\{\delta_{\text{aw}}^2\} = E\{\delta_{\text{aw}}^2\} \] (9)

\[ \text{SNR}_{\text{aw}}(K,N) = E\{\text{SNR}_{\text{aw}}\} \] (10)

\[ \sigma_{\text{cw}}^2(K,N) = E\{\delta_{\text{cw}}^2\} \] (11)

\[ s_{\text{cw}}(K,N) = E\{|s'|^2\} \] (12)

and

\[ \text{NSR}_{\text{cw}}(K,N) = E\{\text{NSR}_{\text{cw}}\}. \] (13)

Equations (9) to (13) are the first moments or average transient values of the previously defined output measures of the GS cancelers. These output measures...
are commonly used to grade the convergence performance of the SMI canceler. In addition, the \( i \)th moment of SNR is defined as
\[
\text{SNR}_{\text{ii}}(k, N) = E\{(|\text{SNR}|)^i\} \quad (14)
\]
and the \( i \)th moment of \( |s|^2 \) is defined as
\[
\text{sg}(K, N) = E\{(|s|^2)^i\} \quad (15)
\]
where \( s(N, K) = s^*(N, K) \).

In the succeeding sections, expressions for \( u_{\text{ii}}(K, N) \), \( \text{SNR}_{\text{lwl}}(K, N) \), \( u_{\text{sw}}(K, N) \), \( \text{NSR}_{\text{sw}}(K, N) \), \( \text{SNR}_{\text{ti}}(K, N) \), and \( s(N, K) \) are derived by using the GS canceler as an analysis tool.

IV. SMI AND GS CANCELER EQUIVALENCE

In this section the SMI canceler and the GS canceler are shown to be equivalent in the sense that the estimated linear weighting vector of the SMI is identical to the equivalent estimated vector of the GS. Hence, if either concurrent or nonconcurrent processing is used, the output of the SMI and GS canceler is identical in the transient state (finite averaging). For this equivalence to be true, infinite computational accuracy and the nonsingularity of the estimated input covariance is assumed.

We briefly describe the SMI algorithm for the SLC
\[
\hat{w}_0 = R_{\text{ii}}^{-1}r_{\text{am}}. \quad (16)
\]
For the SMI algorithm, \( R_{\text{ii}} \) and \( r_{\text{am}} \) are estimated, the auxiliary weighting vector is calculated by using (16), this vector is applied to the auxiliary channels, and the resultant is subtracted from the main channel.

Define the input data vector
\[
x_a(k) = [x_1(k), x_2(k), \ldots, x_{N-1}(k)]^T, \quad k = 1, 2, \ldots, K \quad (17)
\]
where \( a \) refers to the auxiliary channels and \( T \) denotes the vector (or matrix) transpose operation. The estimates of \( R_{\text{ii}} \) and \( r_{\text{am}} \), denoted by \( \hat{R}_{\text{ii}} \) and \( \hat{r}_{\text{am}} \), are given by the expressions
\[
\hat{R}_{\text{ii}} = \frac{1}{K} \sum_{k=1}^{K} x_a(k)x_a^*(k) \quad (18)
\]
and
\[
\hat{r}_{\text{am}} = \frac{1}{K} \sum_{k=1}^{K} x_a(k)x_0(k) \quad (19)
\]
where the estimated linear weighting vector can be found by using the equation
\[
\hat{R}_{\text{ii}}\hat{w} = \hat{r}_{\text{am}}. \quad (20)
\]
We define the following \( K \)-length input data vectors
\[
x_n = [x_n(1), x_n(2), \ldots, x_n(K)]^T, \quad n = 0, 1, \ldots, N - 1. \quad (21)
\]
and a \( K \times (N - 1) \) auxiliary input data matrix \( A \) where
\[
A = [x_1, x_2, \ldots, x_{N-1}]. \quad (22)
\]
It is straightforward to show that
\[
\hat{R}_a = A^tA \quad (23)
\]
and
\[
\hat{r}_{\text{am}} = A^tx_0. \quad (24)
\]
Thus using (20),
\[
\hat{w} = \hat{R}_{\text{ii}}^{-1}\hat{r}_{\text{am}} = (A^tA)^{-1}A^tx_0. \quad (25)
\]
Note that a necessary condition that \( \hat{R}_{\text{ii}} \) be nonsingular (and hence a unique solution for \( \hat{w} \) exists) is that \( K \geq N - 1 \). To show this, assume that \( K < N - 1 \) and define an \( (N - 1) \times (N - 1) \) augmented matrix \( A_{\text{aug}} \) as
\[
A_{\text{aug}} = \begin{bmatrix} A \\ 0 \end{bmatrix} \quad (26)
\]
where the last \( N - K - 1 \) rows are zero filled. Now
\[
\hat{R}_{\text{ii}} = A_{\text{aug}}^{-1}A_{\text{aug}}. \quad (27)
\]
However the determinant of \( \hat{R}_{\text{ii}} \), denoted by \( \det(\hat{R}_{\text{ii}}) \), equals \( \det(A_{\text{aug}}) \det(A_{\text{aug}}) \). Since \( \det(A_{\text{aug}}) = 0 \), it follows that \( \det(\hat{R}_{\text{ii}}) = 0 \) so that \( \hat{R}_{\text{ii}} \) is singular if \( K < N - 1 \).

We now show that if \( K = N - 1 \), the output noise residue is zero for the concurrent processor implementation. Set \( \hat{w} = (\hat{w}_1, \hat{w}_2, \ldots, \hat{w}_{N-1})^T \) and let \( z \) be the \( K \)-length output residue vector of the concurrent canceler. As a result
\[
z = x_0 - \sum_{n=1}^{N-1} \hat{w}_n x_n. \quad (28)
\]
Using (25), it can be shown that
\[
z = x_0 - A(A^tA)^{-1}A^tx_0 = (I_k - A(A^tA)^{-1}A^t)x_0 \quad (29)
\]
where \( I_k \) denotes the \( K \times K \) identity matrix. For \( K = N - 1 \), \( A(A^tA)^{-1}A^t = I_k \), so that \( z = 0 \).

As a result of the preceding discussion, in the following development for nonconcurrent processing we restrict \( K \geq N - 1 \) and for concurrent processing, \( K \geq N \).
It can be shown that (20) reduces to solving the following system of linear equations:

\[
\begin{align*}
\sum_{n=1}^{N-1} x_n^1 x_n \hat{w}_n &= x_0^1 x_0 \\
\sum_{n=1}^{N-1} x_n^2 x_n \hat{w}_n &= x_0^2 x_0 \\
\vdots \\
\sum_{n=1}^{N-1} x_{N-1}^n x_n \hat{w}_n &= x_{N-1}^1 x_0.
\end{align*}
\]

(30)

We show that the solution for the SMI weights, \( \hat{w}_1, \hat{w}_2, \ldots, \hat{w}_{N-1} \), which orthogonalizes the auxiliary input data vectors \( x_n, n = 1, 2, \ldots, N - 1 \) to the output residue vector \( z \) is given by solving (30); i.e., the condition

\[ x_n^i z = 0, \quad n = 1, 2, \ldots, N - 1 \]

(31)

results in a system of equations identical to (30). Note that (28) and (31) imply concurrent processing. However the weight calculation is valid for either concurrent or nonconcurrent processing.

If the concurrent GS canceler orthogonalizes the output data vector with respect to the auxiliary input data vectors \( x_n, n = 1, 2, \ldots, N - 1 \), it follows that the equivalent linear weighting vector associated with the GS structure is identical to that computed for the SMI algorithm. We show that a concurrent GS canceler orthogonalizes the output data vector \( z \) with respect to the auxiliary input data vector by using mathematical induction. This is obviously true for \( N = 2 \); we assume that it is true for all integers less than or equal to \( N - 1 \) and demonstrate that it is true when the number of channels equals \( N \).

From Fig. 2 and our assumptions:

\[ x_n^{(2)} z = 0, \quad n = 1, 2, \ldots, N - 2 \]

(32)

and

\[ x_{N-1}^{(1)} x_0^{(2)} = 0, \quad n = 0, 1, 2, \ldots, N - 2 \]

(33)

where \( x_n^{(1)}, x_n^{(2)} \) are the \( K \)-length data vectors associated with \( x_n^{(1)}, x_n^{(2)} \), respectively. Furthermore, the output vector can be written as

\[ z = x_0^{(2)} - \sum_{n=1}^{N-2} W_n^{(2)} x_n^{(2)} \]

(34)

where \( W_1^{(2)}, W_2^{(2)}, \ldots, W_{N-2}^{(2)} \) is representative of the equivalent linear weighting of the input vectors from level 2 through \( N - 1 \) of the GS structure (see Fig. 2).

Using (1) in (32) results in

\[
\begin{align*}
x_n^{(2)} z &= (x_n^{(2)} - W_n^{(2)} x_{N-1}^{(1)}) z \\
&= x_n^{(2)} z - W_n^{(2)} x_{N-1}^{(1)} z = 0, \quad n = 1, 2, \ldots, N - 2.
\end{align*}
\]

(35)

From (33) and (34) it can be shown that

\[ x_{N-1}^{(1)} z = x_{N-1}^{(1)} x_0^{(2)} - \sum_{n=1}^{N-2} W_n^{(2)} x_n^{(2)} x_{N-1}^{(1)} z = 0. \]

(36)

Thus, from (35) and (36), it follows that

\[ x_n^{(1)} z = 0, \quad n = 1, 2, \ldots, N - 1. \]

(37)

Since \( x_n = x_n^{(1)}, n = 1, 2, \ldots, N - 1 \), we have shown that the auxiliary input data vectors are orthogonal to the output vector.

V. INVARIANT TRANSFORMS AND GS THEOREMS

In this section, we discuss two types of matrix transforms on the input data that significantly simplify the forthcoming analysis. Let \( C \) be any \( N \times N \) nonsingular matrix. It is well known [7] that transforming the input channels \( x_1, x_2, \ldots, x_{N-1} \) by this transform does not change the transient or steady state performance of the SMI (or GS) canceler. GS cancellation is equivalent to a specific matrix transformation of the input channels. For the GS canceler, the transform matrix \( C \) has the upper triangular matrix form. An equivalent configuration of a GS canceler in the transient state is illustrated in Fig. 4. Here \( C \) is implemented by passing the input channels through a \( GS_{aux,N} \) structure followed by a power equalizer on the output auxiliary channels. The output powers of the AUX channels after power equalization are equal to \( \sigma^2_{aux} \). Note that each input channel into the \( GS_{aux,N} \) structure is orthogonal in the steady state to the other channels and that all input channels have the same power level, \( \sigma^2_{aux} \). Also, without loss of generality we can define \( \sigma^2_{aux} = 1 \).

The structure shown in Fig. 4 illustrates that any GS canceler structure can be divided into two parts: a deterministic steady state front-end processor whereby the main channel is decorrelated from the auxiliary channels and a stochastic back-end processor that is driven by uncorrelated equal powered noise in each channel. The back-end processor is independent of the input covariance matrix, and the auxiliary weights associated with the back-end processor go to zero as \( K \to \infty \). Hence the convergence properties of the GS canceler can be studied by analyzing the convergence properties of the back-end processor. Thus from this point on the input channels are assumed to be orthogonal and of equal power.

A second matrix transform that significantly simplifies the forthcoming analysis is now discussed.

GERLACH & KRETSCHMER: CONVERGENCE PROPERTIES OF ADAPTIVE ALGORITHMS

49
Let \( \Phi \) be any \( K \times K \) unitary matrix, i.e., \( \Phi^* \Phi = I_K \).

Let us transform each input channel noise vector \( x_n \), \( n = 0, 1, 2, \ldots, N - 1 \) by \( \Phi \) such that
\[
    x'_n = \Phi x_n, \quad n = 0, 1, \ldots, N - 1
\]  

(38)

where \( x'_n, n = 0, 1, \ldots, N - 1 \) is the resultant output noise vector. If we input this noise vector into a \( GS_{K,N} \) canceler, we can show that the estimated weights using the \( x'_n \) inputs are identical to those using the \( x_n \) inputs. This is easily proved by substituting \( x'_n \) as given by (38) into (30), which is the system of equations that solves for the auxiliary weights. Because
\[
    x'_m' x'_n = (\Phi x_n)' (\Phi x_m) = x'_m x'_n, \quad \text{for any } n, m
\]  

(39)

an identical system of linear equations results. Thus, the estimated weights are identical.

One simplification that readily presents itself because of the above invariant transform pertains to the signal representation for concurrent processing. Let the input signal be represented by the \( K \)-length vector \( s \) where
\[
    s = (s_1, s_2, \ldots, s_K)^T.
\]  

(40)

It is known that a unitary matrix transform \( \Phi \) exists that transforms \( s \) into a \( K \)-length vector with a nonzero first element and all other elements equal to zero. In fact
\[
    \tilde{s} = \Phi s = \left( \sqrt{s_1}, 0, 0, \ldots, 0 \right)^T.
\]  

(41)

Thus we need only consider an input signal vector which is proportional to the form \((1, 0, \ldots, 0)^T\) and \( \sqrt{s_1} \).

VI. TWO-INPUT GS CANCELER

The basis for understanding the convergence properties of a GS canceler begins with studying the two-input GS canceler illustrated in Fig. 5 where from the discussion in Section V, \( x_0 \) and \( x_1 \) are RVs that are assumed to be equal powered and uncorrelated. Let
\[
    z_{cw} = (z_{cw}(1), z_{cw}(2), \ldots, z_{cw}(K))^T
\]

be the concurrent output noise residue vector and \( z_{nw} \) be the nonconcurrent output noise residue. For a two-input GS canceler
\[
    z_{cw} = x_0 - \hat{w} x_1
\]

(42)

\[
    z_{nw} = X_0 - \hat{w} X_1
\]

(43)

where
\[
    \hat{w} = \frac{x_1' x_0}{x_1' x_1}.
\]

(44)

Furthermore, the transient output noise powers are given by
\[
    \frac{\sigma^2_{cw}}{\sigma^2_{nw}} = \frac{1}{K} \frac{x_1' x_0}{x_1' x_1}
\]

(45)

and
\[
    \frac{\sigma^2_{nw}}{\sigma^2_{cw}} = 1 + \left| \hat{w} \right|^2 = 1 + \frac{|x_1' x_0|^2}{(x_1' x_1)^2}.
\]

(46)

In Appendix A (and [8]), it is shown under assumptions 1 to 6 given in Section II that \( \frac{\sigma^2_{nw}}{\sigma^2_{cw}} = \eta \) has the following probability density function (pdf).
\[
    p(\eta) = \frac{K}{\eta^K + 1}, \quad \eta \geq 1.
\]

(47)

Now the transient SNR is equal to the reciprocal of \( \frac{\sigma^2_{nw}}{\sigma^2_{cw}} \). Thus if SNR \( = 1/\frac{\sigma^2_{nw}}{\sigma^2_{cw}} = \rho \), then by the use of elementary probability theory,
\[
    p(\rho) = K \rho^{K-1}, \quad 0 \leq \rho \leq 1.
\]

(48)
By using (47) and (48), it can be shown that
\[ \sigma_{ow}^2(K,2) = 1 + \frac{1}{K-1} \]  
(49)
\[ \text{SNR}_{ow}(K,2) = \frac{K}{K+1} \]  
(50)
and
\[ \text{SNR}_{ow}^{(i)}(K,2) = \frac{K}{K+i}. \]  
(51)

From (45), the expected value of \( \sigma_{ow}^2 \) conditioned on \( x_1 \) is given by
\[ E(\sigma_{ow}^2 | x_1) = \frac{1}{K} \left[ E(x_0^2) - \frac{x_1^2}{x_1^2} E(x_0 x_1^*) \right]. \]  
(52)
By assumption \( E(x_0 x_1^*) = I_K \), where \( I_K \) is the \( K \times K \) identity matrix and \( E(x_0^2) = K \). Thus (52) reduces to
\[ E(\sigma_{ow}^2 | x_1) = 1 - \frac{1}{K}. \]  
(53)
Note that this expression is independent of \( x_1 \) so that
\[ E(\sigma_{ow}^2) = 1 - \frac{1}{K}. \]  
(54)
or
\[ \sigma_{ow}^2(K,2) = 1 - \frac{1}{K}. \]  
(55)
Also note the assumption that the inputs are Gaussian RVs was not used in this derivation. In fact each data point in either channel can have any pdf so long as it has a zero mean and identical variances.

The unnormalized output signal vector through the two-input GS canceler is given by the expression
\[ s'' = \left( I_K - \frac{x_1 x_1^*}{x_1^2} \right) s. \]  
(56)
Hence the sample average output signal power is given by
\[ |s''|^2 = \frac{1}{K} s'' s'' = \frac{1}{K} \left( s s - \frac{|x_1|^2}{x_1 x_1^*} \right). \]  
(57)
We showed in Section V that \( s \) and \( x_1 \) can be transformed by a unitary matrix without affecting the resultant output measures. If we transform \( s \) and \( x_1 \) by the unitary \( \Phi_1 \) matrix defined in Section V, then (57) reduces to
\[ |s''|^2 = \frac{1}{K} s s \left[ 1 - \frac{|x_1|^2}{\sum_{k=1}^{K} |x_1(k)|^2} \right]. \]  
(58)
Now the normalized signal power is given by
\[ |s'|^2 = \frac{|s''|^2}{s s}. \]  
(59)
so that
\[ |s'|^2 = 1 - \frac{|x_1(1)|^2}{\sum_{k=1}^{K} |x_1(k)|^2}. \]  
(60)
It is straightforward to show that if \( x_1 \) is a normalized \( K \)-length multivariate complex circular Gaussian vector, then \( \xi = |s'|^2 \) has the following pdf
\[ p(\xi) = (K-1)\xi^{K-2}, \quad \xi \geq 0. \]  
(61)
Furthermore, the moments of \( |s'|^2 \) can be found by using the above pdf and are given by
\[ s_{ow}(K,2) = \frac{K-1}{K-1+i}. \]  
(62)
From (62) it follows that
\[ s_{ow}(K,2) = 1 - \frac{1}{K}. \]
Finally, we can show that
\[ \text{NSR}_{ow} = \frac{\sigma_{ow}^2}{|s'|^2} = \frac{1}{K} \left( \frac{x_0^2 - |x_1|^2}{x_1 x_1^*} \right) + \frac{1}{K} \left( \frac{|x_1(1)|^2}{\sum_{k=1}^{K} |x_1(k)|^2} \right). \]  
(63)
If \( \text{NSR}_{ow} \) is averaged over \( x_0 \), then
\[ E(\text{NSR}_{ow} | x_1) = \frac{K-1}{K} \left( 1 + \frac{|x_1(1)|^2}{\sum_{k=2}^{K} |x_1(k)|^2} \right). \]  
(64)
Now since \( x_1 \) is a normalized \( K \)-length multivariate complex circular Gaussian vector,
\[ E(|x_1(1)|^2) = 1 \quad \text{and} \quad \text{and} \]
\[ E \left( \frac{1}{\sum_{k=2}^{K} |x_1(k)|^2} \right) = \frac{1}{K-2}. \]  
(65)
Thus
\[ \text{NSR}_{ow}(K,2) = E(\text{NSR}_{ow}) = \frac{K-1}{K} \left( 1 + \frac{1}{K-2} \right) = \frac{(K-1)^2}{K(K-2)}. \]  
(66)
VII. GENERAL MOMENT THEOREM FOR GS CANCELERS

Let $\hat{A}_{K,N}$ denote a transient unnormalized moment of an output measure (output signal or noise power residue, NSR, or SNR) associated with a concurrent or nonconcurrent GS canceler with $N$ input channels and $K$ independent samples per channel. Define the normalized average transient moment as

$$A(K,N) = \frac{E\{\hat{A}_{K,N}\}}{K^{N-1}}$$

(67)

where $A(K,1) = 1$ and $K > N$. In this section we prove the following theorem.

**GENERAL MOMENT THEOREM FOR GS CANCELERS.** If assumptions 1 to 6 hold, then

$$A(K,N) = A(K,2)A(K-1, N-1)$$

(68)

or equivalently

$$A(K,N) = \prod_{k=K-N+2}^{K} A(k,2).$$

(69)

**PROOF.** We prove this by mathematical induction. First, the Theorem is obviously true for $N = 2$. Thus, we can assume that the Theorem is true for all integers less than or equal to some upper bound, $N - 1$. We can then show that it is true for $N$, which implies that it is true for any $N \geq 2$.

Again assume that all input channels are of equal power and uncorrelated. It is shown in [7] and discussed in Section V that this assumption does not change the output measures. A GS$_{K,N}$ structure can be decomposed as shown in Fig. 6 into a first-level processor followed by a GS$_{K-1,N-1}$ structure. The output data $K$-length vectors of the first-level processor can be written as

$$y_n = x_n - \hat{y}_n x_{N-1}, \quad \hat{y}_n = \frac{x_{N-1}^{H} x_n}{x_{N-1}^{H} x_{N-1}},$$

(70)

or

$$y_n = x_n - \frac{x_{N-1}^{H} x_n}{x_{N-1}^{H} x_{N-1}} x_{N-1}$$

or

$$y_n = (I_K - \frac{x_{N-1}^{H} x_{N-1}}{x_{N-1}^{H} x_{N-1}}) x_n,$$

(71)

$$n = 0, 1, 2, \ldots, N-1.$$  

It can be shown that

$$I_K - \frac{x_{N-1}^{H} x_{N-1}}{x_{N-1}^{H} x_{N-1}} = \Phi^{H} \Lambda \Phi$$

(72)

where $\Phi$ is a $K \times K$ unitary matrix and $\Lambda$ is a diagonal matrix where the first element is 0 and all other diagonal elements are equal to 1. Thus

$$y_n = \Phi^{H} \Lambda \Phi x_n, \quad n = 0, 1, \ldots, N - 2.$$  

(73)

As shown in Section V, the output data set $y_n, n = 0, 1, \ldots, N - 2$ can be transformed by a unitary matrix $\Phi$ and not change the equivalent transient weighting vector of the GS$_{K-1,N-1}$ structure. Thus

$$u_n = \Phi y_n = \Lambda \Phi x_n, \quad n = 0, 1, \ldots, N - 2.$$  

(74)

Now set $v_n = \Phi x_n$. Because $x_n$ is a normalized $K$-length multivariate complex circular Gaussian vector, then $v$ is the same. As a result, using the form of $\Lambda$ and setting $u_n = (u_{n1}, u_{n2}, \ldots, u_{nk})^{T}$, $v_n = (v_{n1}, v_{n2}, \ldots, v_{nk})^{T}$, it follows from (74) that

$$u_{n1} = 0$$

(75)

and

$$u_{nk} = v_{nk}, \quad k = 2, 3, \ldots, K.$$  

Hence, the input RVs to the GS$_{K-1,N-1}$ structure are identically distributed to the input RVs to the GS$_{K,N}$ structure except that their number has been reduced by one as illustrated in Fig. 7.

Consider the implications of this new structure by using concurrent processing. Let $\hat{B}_{K-1,N-1}$ denote a transient unnormalized moment of any output measure (output noise residue or SNR) associated with the GS$_{K-1,N-1}$ structure. Note that $\hat{A}_{K,N} = \hat{B}_{K-1,N-1}$. Then according to the General Moment Theorem,

$$\lim_{K \to \infty} \frac{E\{\hat{B}_{K-1,N-1} | x_N\}}{K^{N-1}} = A(K-1,N-1).$$

(76)

Note in the limit taken above that $K$ goes to infinity only in the GS$_{K-1,N-1}$ structure and not in the first-level processor that precedes the GS$_{K-1,N-1}$
VIII. CONVERGENCE RESULTS

In this section, the General Moment Theorem for GS cancelers (69) is used to derive a number of results. Most of these results are demonstrated in [7 and 8].

Employing the General Moment Theorem under assumptions 1 to 6 and the expressions for $\sigma_{\text{sn}}(K,2)$, $s_{\text{snw}}(K,2)$, $s_{\text{snw}}(K,2)$, $\text{SNR}_{\text{snw}}(K,2)$, $\sigma_{\text{snw}}(K,2)$, $\text{SNR}_{\text{snw}}(K,2)$ and $\text{SNR}_{\text{snw}}(K,2)$ given in Section IV, we can show that

$$\sigma_{\text{snw}}(K, N) = 1 - \frac{N - 1}{K} \cdot (82)$$

$$s_{\text{snw}}(K, N) = 1 - \frac{N - 1}{K} \cdot (83)$$

$$s_{\text{snw}}(K, N) = (K - 1)(K + i - N)! \cdot (84)$$

$$\text{SNR}_{\text{snw}}(K, N) = \frac{(K - 1)(K - N + 1)}{K(K - N)} \cdot (85)$$

$$\sigma_{\text{snw}}(K, N) = \frac{K}{K - N + 1} \cdot (86)$$

$$\text{SNR}_{\text{snw}}(K, N) = \frac{K - N + 2}{K + 1} \cdot (87)$$

$$\text{SNR}_{\text{snw}}(K, N) = \frac{K!(K + i - N + 1)!}{(K - N + 1)!((K + i)!} \cdot (88)$$

Equations (82) to (84) and (86) to (88) are given in [7 and 8].

Equation (88) can be used in ad hoc fashion to find the pdf of $\text{SNR}_{\text{snw}}$ for any $K$ and $N$. If $\rho = \text{SNR}_{\text{snw}}$, then the pdf that yields moments as given by (88) is

$$p(\rho) = \frac{K!}{(N - 2)!(K - N + 1)!} \frac{(1 - \rho)^{N - 2} \rho^{K - N + 1}}{0 \leq \rho \leq 1} \cdot (89)$$

From (89), the pdf of $\sigma_{\text{snw}}^2$ can be obtained. Let $\eta = \sigma_{\text{snw}}^2 = 1/\rho$. It is straightforward to show that

$$p(\eta) = \frac{K!}{(N - 2)!(K - N + 1)!} \frac{(\eta - 1)^{N - 2} \eta^{K + 1}}{1 \leq \eta \leq \infty}, \cdot (90)$$
Note that (89) and (90) were first derived in [7 and 8], respectively.

IX. LOWER BOUND

In this section, we derive a lower bound associated with convergence of a nonconcurrent GS canceler when the input data are not necessarily Gaussian. In the analysis, assumptions 2 to 6 of Section II hold. An element of the input data vectors \( x_0, x_1, \ldots, x_{N-1} \) can have any pdf.

A result for the two-input GS canceler is first established. For this case, it was previously shown that

\[
\hat{\sigma}_w^2 = 1 + |w|^2 = 1 + \frac{x_1^2 x_0^2}{(x_1^2)^2}.
\]

(91)

We can show that because \( E\{x_0 x_0^*\} = I_K \),

\[
E\{\hat{\sigma}_w^2 | x_1\} = 1 + \frac{1}{x_1^2}.
\]

(92)

Thus

\[
\sigma_w^2(K,2) = E\{\hat{\sigma}_w^2\} = 1 + E\left\{\frac{1}{x_1^2}\right\}.
\]

(93)

Appendix B shows that if \( z \) is any RV with a nonzero mean, then

\[
E\left\{\frac{1}{z}\right\} \geq \frac{1}{E\{z\}}.
\]

(94)

Applying this inequality to (93) results in the inequality

\[
\sigma_w^2(K,2) \geq 1 + \frac{1}{K}.
\]

(95)

We now state the following theorem.

**NONCONCURRENT PROCESSING THEOREM.** Let an input data element in the vectors \( x_0, x_1, \ldots, x_{N-1} \) have any pdf. Then under assumptions 2 to 6, Section II,

\[
\sigma_w^2(K,N) \geq \frac{K + 1}{K - N + 2}.
\]

(96)

**PROOF.** The Proof of this Theorem is by induction and again closely follows the Proof of the General Moment Theorem. We have shown that it is true for \( N = 2 \) (95). Thus the Theorem is true for all integers less than or equal to some upper bound, \( N - 1 \). We show this is true for any \( N \), which implies that it is true for any \( N \geq 2 \).

Again the GS\(_{K,N}\) processor is decomposed as shown in Fig. 6 and further reduced as shown in Fig. 7. Neither this decomposition nor reduction depend on the pdfs of the input data. Also the concurrent data entering the GS\(_{K-1,N-1}\) processor satisfy assumptions 2 to 6. In addition, the nonconcurrent data conditioned on \( x_{N-1} \) and \( X_{N-1} \) satisfy assumptions 2 to 6.

Thus if \( \hat{\sigma}_{w-1,N-1}^2 \) is the transient unnormalized output noise power residue using nonconcurrent processing of the GS\(_{K-1,N-1}\) canceler where \( \hat{\sigma}_{w-1,N-1}^2 = \hat{\sigma}_{w-1,N-1}^2 \) and if the Nonconcurrent Processing Theorem is true for \( N = 1 \), then

\[
\lim_{K \to \infty} E\{\hat{\sigma}_{w-1,N-1}^2 | x_{N-1}, X_{N-1}\} \geq \frac{K}{K - N + 2}.
\]

(97)

Again note that in the limit taken above, \( K \) goes to infinity only in the GS\(_{K-1,N-1}\) structure and not in the first-level processor that precedes the GS\(_{K-1,N-1}\) structure.

From (97), it follows that

\[
\lim_{K \to \infty} E\{\hat{\sigma}_{w-1,N-1}^2\} \geq \frac{K}{K - N + 2} \lim_{K \to \infty} E\{\hat{\sigma}_{w-1,N-1}^2\}.
\]

(98)

We know that

\[
\lim_{K \to \infty} E\{\hat{\sigma}_{w-1,N-1}^2\} = \sigma_w^2(K,2) \geq 1 + \frac{1}{K}.
\]

(99)

Substituting (99) into (98) results in the Theorem being proved.

Note that if the input noises are Gaussian, then the lower bound given by (96) is almost achieved. Hence, this assumption results in almost the "best case" performance.

X. OVERMATCHING DEGREES OF FREEDOM

If we examine the expression for SNR\(_{nw}(K,N)\) given by (87) for the nonconcurrent GS canceler, we find that for a fixed number of input samples \( K \), the SNR\(_{nw}(K,N)\) decreases monotonically as the order of the GS structure \( N \). Hence, for a given input noise scenario, increasing the order of the GS canceler may lead to a noisier output.

To illustrate this problem, let us say that for a variety of input noise scenarios a GS\(_{K,N}\) canceler yields good cancellation performance and is therefore specified in the design. However, suppose for a specific noise scenario that only an \( L \)th-order GS canceler is needed for good performance where \( L < N \). Hence, for this specific scenario there is a loss of cancellation performance by using a GS\(_{K,N}\) canceler instead of a GS\(_{K,L}\) canceler. For this case, what occurs is that at the \( (L-1) \)th level of cancellation in the GS\(_{K,N}\) canceler, the input noises in the main channel are essentially cancelled. Thereafter, in each succeeding level the noise residue will only increase. This phenomenon is called "overmatching the degrees of freedom (DOF)." For optimality, we should have stopped the cancellation process after the \( (L-1) \)th level. Note that the number of DOFs for a GS\(_{K,N}\) canceler is \( N - 1 \).

By using the expression given for SNR\(_{nw}(K,N)\) in (87), we can quantify this loss for nonconcurrent processing. This loss is given by

\[
\text{LOSS}_w = \frac{\text{SNR}_{nw}(K,L)}{\text{SNR}_{nw}(K,N)} = \frac{K - L + 2}{K - N + 2}.
\]

(100)
One method suggested by Lewis and Kretschmer [9] for overcoming the effects of overmatching the DOFs is to monitor the nonconcurrent noise powers at each level of the main channel in the GS structure. The main channel is terminated in the GS structure where the noise power is a minimum. This point in the GS structure varies as the noise environment changes or equivalently as more or fewer DOFs are needed.

\[ @ = \text{a } K \times K \text{ unitary matrix and } A \text{ is the diagonal matrix of eigenvalues where the first diagonal element equals 1 and all others equal 0. Thus} \]
\[ v = x_0^t \Phi^t \Lambda \Phi x_0 = (\Phi x_0)^t \Lambda \Phi x_0. \] (A6)

Set a new K-length vector \( y = \Phi x_0 \). It is easy to show that if \( x_0 \) is a normalized K-length multivariate complex circular Gaussian vector, then \( y \) is the same. If \( y_1 \) is the first element of \( y \), then from (A6)
\[ v = |y_1|^2. \] (A7)

XI. SUMMARY

The open-loop GS canceler was shown to be numerically identical with the SMI algorithm in the transient state if infinite numerical accuracy is assumed. Two forms of the GS canceler were discussed and analyzed: concurrent and nonconcurrent. Previous and new results for concurrent and nonconcurrent SMI cancelers that assume Gaussian inputs have been produced by using the GS structures as an analysis tool. In addition, new results were obtained when the input noises are not Gaussian. The deleterious effect of overmatching the DOFs was discussed.

APPENDIX A. PROBABILITY DENSITY FUNCTION ASSOCIATED WITH TWO-INPUT GS CANCELER

In this Appendix, we outline a derivation for obtaining the pdf of the transient noise power residue and the transient SNR of a two-input GS. For this analysis the estimated weight is applied to a data set that is independent of the data set that calculated the weight (nonconcurrent processing).

Let \( x_0 \) and \( x_1 \) be the K-length input vectors associated with the main and auxiliary channels, respectively. The estimated weight of the GS canceler is then given by
\[ \hat{w} = \frac{x_1^t x_0}{x_1^t x_1}. \] (A1)
For nonconcurrent processing the transient noise power residue is given by
\[ \sigma^2 = 1 + |\hat{w}|^2. \] (A2)
We derive the pdf of \(|\hat{w}|^2\), from which the pdf of \( \sigma^2 \) is easily attainable.

Now let \( z = |\hat{w}|^2 \) or
\[ z = \frac{|x_1^t x_0|^2}{(x_1^t x_1)^2} = \frac{x_0^t x_1 x_1^t x_0}{x_1^t x_1 - x_1^t x_1}. \] (A3)
Let
\[ \nu = x_0^t x_1 x_1^t x_0. \] (A4)
We derive \( p(\nu \mid x_1) \). The matrix \( x_1^t x_1^t x_1 x_1 \) can be written as
\[ x_1^t x_1^t = \Phi^t \Lambda \Phi \] (A5)
where \( \Phi \) is a \( K \times K \) unitary matrix and \( \Lambda \) is the diagonal matrix of eigenvalues where the first diagonal element equals 1 and all others equal 0. Thus
\[ v = x_0^t \Phi^t \Lambda \Phi x_0 = (\Phi x_0)^t \Lambda \Phi x_0. \] (A6)

The pdf of \(|y_1|^2\) is the well-known second-order chi-square and thus it can be shown that
\[ p(\nu \mid x_1) = e^{-\nu}, \quad \nu \geq 0. \] (A8)
Note that this pdf is independent of \( x_1 \). As a result,
\[ p(\nu) = e^{-\nu}, \quad \nu \geq 0. \] (A9)
From (A3) and elementary probability theory we can show that
\[ p(z \mid x_1) = x_0^t x_1 e^{-z x_1^t x_1}, \quad z \geq 0 = u e^{-uz}, \quad u \geq 0 \] (A10)
where \( u = x_0^t x_1 \). It is known that the pdf of \( u \) is given by
\[ p(u) = \frac{1}{(K-1)!} u^{K-1} e^{-u}, \quad u \geq 0. \] (A11)
Hence
\[ p(z) = \int_0^\infty p(x \mid u) p(u) du \] (A12)
\[ = \int_0^\infty \frac{1}{(K-1)!} u^{K} e^{-u + 1} du. \] (A13)
It is elementary to show that the above integral reduces to
\[ p(z) = \frac{K}{(z+1)^{K+1}}, \quad z \geq 0. \] (A14)
If we set \( \eta = \sigma^2 \), then it follows from (A14) and (A2) that
\[ p(\eta) = \frac{K}{\eta^{K+1}}, \quad \eta \geq 1. \] (A15)

APPENDIX B. FIRST MOMENT BOUND

If \( z \) is an RV with \( z \geq 0, 0 < E(z) < \infty \), and \( E(1/z) < \infty \), then
\[ E \left[ \frac{1}{z} \right] \geq \frac{1}{E(z)}. \] (B1)
We use the Cauchy-Schwarz inequality to show this. Let \( p(z) \) be the pdf of \( z \). Now
\[ \int_0^\infty p(z) dz = \int_0^\infty \sqrt{\frac{p(z)}{z}} \sqrt{2 p(z)} dz = 1 \] (B2)
where the square root function shown above is the positive square root function. Using the Cauchy-Schwarz inequality,

$$\int_0^\infty \frac{p(z)}{z} \, dz \cdot \int_0^\infty z p(z) \, dz \geq \int_0^\infty \sqrt{\frac{p(z)}{z}} \cdot \sqrt{(z p(z))} \, dz. \tag{B3}$$

Thus

$$\int_0^\infty \frac{p(z)}{z} \, dz \int_0^\infty z p(z) \, dz \geq 1. \tag{B4}$$

Equation (B1) follows from (B4).

REFERENCES


Karl Gerlach (M'81) was born in Chicago, IL. He received his B.S. in 1972 from the University of Illinois, Urbana, and his M.S. and D.Sc. from George Washington University, Washington, DC, in 1975 and 1981, respectively, in electrical engineering.

Since 1972, he has been employed by the Naval Research Laboratory in Washington, DC. From 1972 to 1976, he worked on experimental submarine communications systems and from 1976 to the present he has been with the Radar Division where his research interests include adaptive signal processing and space-based radar.

Dr. Gerlach was the 1986 recipient of the Radar Systems Panel award.

Frank F. Kretschmer, Jr. (M'64—SM'77—F'88) was born in Philadelphia, PA. He received the B.S. degree in 1957 from Pennsylvania State University, University Park, the M.S. degree in 1961 from Drexel Institute of Technology, Philadelphia, PA, and the Ph.D. degree in 1970 from The Johns Hopkins University, Baltimore, MD, all in electrical engineering.

He was employed by the Burroughs Corporation, Paoli, PA, from 1957 to 1958. From 1958 to 1964 he was employed by Bendix Radio in Towson, MD, where he was engaged in radar systems work. From 1964 to 1970 he was a Research Associate at The Johns Hopkins University. Since 1970 he has been with the Naval Research Laboratory, Washington, DC, where his research interests include adaptive signal processing, clutter processing, and waveform coding.

Dr. Kretschmer is a member of Sigma Xi.