An Adaptive Detection Algorithm

E.J. KELLY
Lincoln Laboratory

A general problem of signal detection in a background of unknown Gaussian noise is addressed, using the techniques of statistical hypothesis testing. Signal presence is sought in one data vector, and another independent set of signal-free data vectors is available which share the unknown covariance matrix of the noise in the former vector. A likelihood ratio decision rule is derived and its performance evaluated in both the noise-only and signal-plus-noise cases.

Manuscript received February 18, 1985; revised June 26, 1985.

This work was supported by the United States Air Force under Contract F19628-85-C-0002.

The views expressed are those of the author and do not reflect the official policy or position of the U.S. Government.

Author's address: Lincoln Laboratory, Massachusetts Institute of Technology, P.O. Box 73, Lexington, MA 02173.

0018-9251/86/0300-0115 $1.00 © 1986 IEEE

I. INTRODUCTION

In a well-known paper [1], Reed, Mallett, and Brennan (RMB) discuss an adaptive procedure for the detection of a signal of known form in the presence of noise (interference) which is assumed to be Gaussian, but whose covariance matrix is totally unknown. Two sets of input data are used, which for convenience will be called the primary and secondary inputs. The possibility of signal presence is accepted for the primary data, while the secondary inputs are assumed to contain only noise, independent of and statistically identical to the noise components of the primary data.

In the RMB procedure, the secondary inputs are used to form an estimate of the noise covariance, from which a weight vector for the detection of the known signal is determined. This weight vector is then applied to the primary data in the form of a standard colored noise matched filter. The implication is that the output of this filter is compared with a threshold for signal detection, but no rule is given for the determination of this threshold, whose value controls the probability of false alarm (PFA). In fact, no predetermined threshold can be assigned to achieve a given PFA, since the detector is supposed to operate in an interference environment of unknown form and intensity.

Instead, the RMB paper provides an analysis of the signal-to-noise ratio (SNR) of the filter output, for given values of the secondary data. This SNR is a function of the secondary data and is therefore a random variable. The probability density function (PDF) of this SNR is deduced, and this PDF has the remarkable property of being independent of the actual noise covariance matrix; it is a function only of the dimensional parameters of the problem. In the intended application, the secondary data would (hopefully) be sufficient in quantity to support a good estimate of the noise covariance, and a threshold could presumably be determined from this estimate. The resulting final decision statistic is then more complicated than the matched filter output itself, and being a nonlinear function of the secondary inputs, the detection and false alarm probabilities are not functions of the SNR alone, leaving the actual performance of the procedure undetermined.

In this study the original problem (with a slight generalization of the signal model) is reconsidered as an exercise in hypothesis testing, and the ad hoc RMB procedure is replaced by a likelihood ratio test. No optimality properties are claimed for this test, involving as it does the maximization of two likelihood functions over a set of unknown parameters. The form of the test is, however, reasonable, and the RMB matched filter output appears as a portion of the likelihood ratio detection statistic. This test exhibits the desirable property that its PFA is independent of the covariance matrix (level and structure) of the actual noise encountered. This is a generalization of the familiar constant false alarm rate (CFAR) behavior of detectors using scalar input data, in which only the level of the noise is unknown. In addition, it is shown that the effect of signal presence depends only on the dimensional parameters of the problem and a parameter which is the same as the SNR of a conventional colored noise matched filter.

The PDF of this detection statistic itself is derived here, for both the noise-alone case and the signal-plus-noise case. In the former, the PDF proves to be very simple, and the dependence of the PFA on threshold is exactly the same as that of a simple scalar CFAR problem, in which detection is based on one complex sample and the threshold is proportional to the sum of the squares of a number of independent noise samples. The corresponding PDF in the presence of a signal component exhibits the effect of an SNR loss factor, which obeys the same beta distribution as the SNR loss factor studied in the RMB paper. In fact, the behavior of the detector is identical to that of a simple scalar CFAR system with a fluctuating target, where the latter fluctuation is governed by the beta distribution instead of one of the more usual radar models.

The final detection probability of the likelihood ratio algorithm is obtained in the form of a finite series. The method used here leans heavily on the techniques of the RMB paper (which in turn leans directly on the analysis of Capon and Goodman [2]) with the difference that the matrix transformations required here are carried out directly on the variables of the problem, so that much less reliance is placed on known properties of the Wishart distribution.

It should be emphasized that the output of this likelihood ratio algorithm, like that of the RMB procedure, is a decision on signal presence, and not a sequence of processed data samples from which the interference component has been reduced (nulled) and in which actual signal detection remains to be accomplished. For this reason the direct application of the algorithm to a real radar problem would require the storage of data from an array of inputs (such as adaptive array elements), perhaps also sampled to form range-gated outputs for each pulse and collected for a sequence of pulses (such as those forming a coherent processing interval). The practicality of such a processor, which involves the inversion of a correspondingly large matrix, or the possibility of its simplification, are topics not addressed in this paper.

II. FORMULATION OF THE PROBLEM

The mathematical setting for the formulation of this detection problem is actually quite general, but it is introduced here first in a relatively specific way, in order to lend concreteness by way of example. Suppose that the antenna system of a radar provides a number, say N_a, of RF signals. These may be the outputs of array elements, subarrays, beamformers, or any mix of the above. The radar waveform is supposed to be a simple burst of identical pulses, say N_p in number, and target detection is to be based upon the returns from this burst.

In effect, the radar front end carries out amplification, filtering, and reduction to base band, at which point the quadrature signals are subjected to pulse compression, the final stage of filtering. The order in which these things are carried out is immaterial to the present model, since it is not addressed to the problems of realization and channel matching, although these are of great importance in practice. The I and Q output pairs are next sampled to form range-gate samples for each pulse, say, N_g range gates/pulse. This results in a total of $N_a N_p N_g$ complex samples for the burst. Signal presence is sought in one range gate at a time, hence the primary data consists of the $N_a N_p$ samples from a single, unnamed range gate. These samples are arranged in a column vector z of dimension $N = N_a N_p$. The secondary data consist of the outputs of K range gates, forming a subset of the $N_g - 1$ remaining ones, and these are described by the set of vectors $z(k)$, where $k = 1 \ldots K$. The decision rule will be formulated in terms of the totality of input data, without the a priori assignment of different functions to the primary and secondary inputs.

The secondary data are assumed to be free of signal components, at least in the design of the algorithm, and any selection rules applied to make this assumption more plausible are ignored. The primary data may contain a signal vector, written in the form bs, where b is an unknown complex scalar amplitude, and s is a column vector of N components describing the signal which is sought. The modeled variation of signal amplitude and phase among the array inputs is included in s, as well as pulse-to-pulse variations, such as those relating to a particular target Doppler velocity. The problem of unknown Doppler, or other unknown signal parameters, is mentioned briefly below. It should be noted that the signal vector s can be normalized in any convenient way, since an unknown amplitude factor is already included, and we retain the freedom to assign a norm to s at a later point, where it will be most advantageous to make a specific choice.

The total noise components of the data vectors, representing all sources of internal and external noise and interference, are modeled as zero-mean complex Gaussian random vectors. The noise component of the primary vector z is characterized by the unknown covariance matrix M. Each of the $z(k)$ is assumed to share this $N \times N$ covariance matrix, and the vectors z and the $z(k)$ are all mutually independent. All Gaussian vectors are assumed to have the circular property usually associated with I and Q pairs.

The key features of this model are the Gaussian assumption, the independence of the range-gated output vectors, and the assumption of a common covariance matrix. The structure of the N vectors, in particular the doubly indexed model used to account for multiple pulses and multiple array outputs, is not exploited at all in the following, and a notation for the components of these vectors is not required.

III. THE LIKELIHOOD RATIO TEST

Consider a single input vector from the secondary data set, say $z(k)$. If the covariance matrix of this vector is M,

$$M = E\{z(k)\, z(k)^{\dagger}\}$$

then the N-dimensional Gaussian PDF of this complex random vector is

$$f[z(k)] = \frac{1}{\pi^N \|M\|}\, e^{-z(k)^{\dagger} M^{-1} z(k)}.$$

In the notation used here the double bars signify the determinant of a matrix, and the superscript dagger symbolizes the conjugate transpose of a vector. Each of the secondary data vectors has this same PDF, and under the noise-alone hypothesis, the primary vector does so as well, hence the joint PDF of all the input data is the product:

$$f_0[z, z(1), \ldots, z(K)] = f[z] \prod_{k=1}^{K} f[z(k)].$$

If v is any N vector, we can write the following inner product in the form of a matrix trace:

$$v^{\dagger} M^{-1} v = \mathrm{tr}(M^{-1} V)$$

where V is the open product matrix

$$V = v v^{\dagger}.$$

When this equivalence is applied to all the factors of the joint PDF, it is seen that the latter may be written in the convenient form

$$f_0[z, z(1), \ldots, z(K)] = \left\{\frac{1}{\pi^N \|M\|} \exp[-\mathrm{tr}(M^{-1} T_0)]\right\}^{K+1}$$

where

$$T_0 = \frac{1}{K+1}\left(z\, z^{\dagger} + \sum_{k=1}^{K} z(k)\, z(k)^{\dagger}\right).$$

Under the signal-plus-noise hypothesis, the $z(k)$ have the same PDF as before, and the PDF of the primary vector is obtained by replacing z by

$$z - E\{z\} = z - bs.$$

The resulting joint PDF of the inputs is then

$$f_1[z, z(1), \ldots, z(K)] = \left\{\frac{1}{\pi^N \|M\|} \exp[-\mathrm{tr}(M^{-1} T_1)]\right\}^{K+1}$$

where now

$$T_1 = \frac{1}{K+1}\left((z - bs)(z - bs)^{\dagger} + \sum_{k=1}^{K} z(k)\, z(k)^{\dagger}\right).$$

In the likelihood ratio testing procedure, the PDF of the inputs is maximized over all unknown parameters, separately for each of the two hypotheses. The ratio of these maxima is the detection statistic, and the hypothesis whose PDF is in the numerator is accepted as true if it exceeds some preassigned threshold. The maximizing parameter values are, by definition, the maximum likelihood (ML) estimators of these parameters, hence the maximized PDFs are obtained by replacing the unknown parameters by their ML estimators.

We begin with the noise-alone hypothesis, maximizing over the unknown covariance matrix M. Of all positive definite M matrices, the one which maximizes the expression inside the curly brackets of this PDF is simply T_0. This is equivalent to the statement that the ML estimator of a covariance matrix is equal to the sample covariance matrix, which is well known [3]. When this estimator is substituted in the PDF, the trace which appears there becomes the trace of the $N \times N$ unit matrix, which is just N, and we find

$$\max_M f_0 = \left(\frac{1}{(e\pi)^N \|T_0\|}\right)^{K+1}$$

The same procedure applied to the signal-plus-noise hypothesis yields the formula

$$\max_M f_1 = \left(\frac{1}{(e\pi)^N \|T_1\|}\right)^{K+1}$$

and it remains to maximize this expression over the complex unknown signal amplitude b. Since b appears only in this PDF, we can form a likelihood ratio $L(b)$ at this point and subsequently maximize it over b. It is more convenient to work with the $(K+1)$st root of this ratio, and we put

$$L(b) \equiv \{\ell(b)\}^{K+1}.$$

Obviously,

$$\ell(b) = \frac{\|T_0\|}{\|T_1\|}$$

and the final likelihood ratio test takes the form

$$\max_b \ell(b) = \frac{\|T_0\|}{\min_b \|T_1\|} > \ell_0.$$

The threshold parameter on the right will evidently be greater than unity, since the denominator on the left equals the numerator for the choice $b = 0$, and we are maximizing over b.

To proceed, we define the matrix

$$S = \sum_{k=1}^{K} z(k)\, z(k)^{\dagger}$$

which involves only the secondary data. This matrix is K times the sample covariance matrix of these data, and it satisfies the well-known Wishart distribution. The only property of this distribution that we need here is the fact that for $K > N$, a condition we now impose, the matrix S is nonsingular with probability one. S is, of course, positive definite, and hence Hermitian. We use an easily proved Lemma to evaluate the determinants of both sides of the equation

$$(K+1)\, T_0 = S + z\, z^\dagger$$

with the result

$$(K+1)^N \|T_0\| = \|S\|\, (1 + z^\dagger S^{-1} z).$$

Similarly, we have

$$(K+1)^N \|T_1\| = \|S\|\, (1 + (z - bs)^\dagger S^{-1} (z - bs)).$$

Now is a good time to minimize this quantity over b, and we do this by completing the square:

$$\begin{aligned}(z-bs)^\dagger S^{-1}(z-bs) &= (z^\dagger S^{-1} z) + |b|^2 (s^\dagger S^{-1} s) \\ &\quad - 2\,\mathrm{Re}\,\{b(z^\dagger S^{-1} s)\} \\ &= (z^\dagger S^{-1} z) + (s^\dagger S^{-1} s) \\ &\quad \times \left| b - \frac{(s^\dagger S^{-1} z)}{(s^\dagger S^{-1} s)} \right|^2 \\ &\quad - \frac{|(s^\dagger S^{-1} z)|^2}{(s^\dagger S^{-1} s)}.\end{aligned}$$

The minimum is clearly attained when the positive factor containing b is made to vanish, and the resulting likelihood ratio is given by

$$\ell = \max_b \ell(b) = \frac{1 + (z^\dagger S^{-1} z)}{1 + (z^\dagger S^{-1} z) - \dfrac{|(s^\dagger S^{-1} z)|^2}{(s^\dagger S^{-1} s)}}.$$

It is convenient to introduce the quantity η, defined by

$$\eta = \frac{|(s^\dagger S^{-1} z)|^2}{(s^\dagger S^{-1} s)\, [1 + (z^\dagger S^{-1} z)]}$$

so that

$$\ell = \frac{1}{1 - \eta}.$$

Then the test

$$\ell > \ell_0$$

is equivalent to the test

$$\eta > \eta_0 = \frac{\ell_0 - 1}{\ell_0}.$$

We note that η_0 lies between the values zero and one.

If the target model is generalized, so that the signal vector still contains one or more unknown parameters (such as target Doppler), the likelihood ratio obtained above must next be maximized over these parameters. It is clear that this is equivalent to maximizing η itself over the remaining target parameters. This maximization generally cannot be carried out explicitly, and the standard technique is to approximate it by evaluating the test statistic, in this case η, for a discrete set of target parameters, forming a filter bank, and declaring target presence if any filter output exceeds the threshold. Our purpose in discussing this here is only to show how our test can be generalized in this straightforward way, but from now on we ignore any additional target parameters, which is equivalent to concentrating on the performance of a single member of the filter bank.

For comparison with the RMB procedure, we introduce $\hat{M}$, the ML estimator of the noise covariance, based on the secondary data alone. We have already noted that this estimator is equal to

$$\hat{M} = \frac{1}{K} S.$$

The likelihood ratio test can then be written in the form

$$\frac{|(s^\dagger \hat{M}^{-1} z)|^2}{(s^\dagger \hat{M}^{-1} s)\, [1 + \dfrac{1}{K} (z^\dagger \hat{M}^{-1} z)]} > K\eta_0.$$

We note that the secondary inputs enter this test only through the sample covariance matrix $\hat{M}$ and also that

$$(s^\dagger \hat{M}^{-1} z) = (\hat{w}^\dagger z)$$

where $\hat{w}$ is the RMB weight vector

$$\hat{w} \equiv \hat{M}^{-1} s.$$

The RMB test itself is just

$$|(\hat{w}\, z)|^2 > \text{threshold}$$

which has the form of the colored noise matched filter test, with $\hat{M}$ replacing the usual known covariance matrix of the noise.

The presence of the signal-dependent factor in the denominator of the expression for η causes this detection statistic to be unchanged if the signal vector is altered by a scalar factor. Since the normalization of this vector has been left arbitrary, this invariant is highly desirable. In effect, this factor in the denominator is normalizing s for us, in terms of the estimated noise covariance. The entire detection statistic is also invariant to a common change of scale of all the input data vectors, a minimal CFAR requirement. Further properties of η are developed in the following section.

In the limit of very large K, one expects the estimator $\hat{M}$ to converge to the true covariance matrix M at least in probability. Moreover, it can be shown that the quantity

$$(z^\dagger M^{-1} z)$$

an inner product utilizing the actual covariance matrix instead of its estimator, obeys the chi-squared distribution, with $2N$ degrees of freedom, and hence this term, when divided by K, converges to zero in probability, as K grows without bound. In this sense the likelihood ratio test passes over into the conventional colored noise matched filter test, as the number of sample vectors in the secondary data set becomes very large.

IV. PROPERTIES OF THE LIKELIHOOD RATIO TEST

The likelihood ratio test is discussed in terms of the random variable η, the decision statistic eventually obtained in the preceding section. The definition of η, as well as that of the matrix S on which it depends, are reproduced here for convenience:

$$\eta = \frac{|(s^\dagger S^{-1} z)|^2}{(s^\dagger S^{-1} s)\,[1 + (z^\dagger S^{-1} z)]}$$

$$S = \sum_{k=1}^{K} z(k)\, z(k)^\dagger .$$

The random variable η is, of course, a function of both the primary and secondary data, and as a preliminary to discussing its actual PDF, some useful properties are first derived. We begin with the noise-alone case, and assume that the actual noise covariance matrix is M.

The matrix M is positive definite, and hence a positive definite square root matrix can be defined. Since M can be diagonalized by a unitary transformation, it can be represented in the form

$$M = U \Lambda U^\dagger$$

where the columns of the unitary matrix U are eigenvectors of M, and Λ is diagonal. The diagonal elements of Λ, say $\lambda(n)$, $(n = 1 \ldots N)$, are the real, positive eigenvalues of M. In case of degeneracy of an eigenvalue, the corresponding eigenvectors are assumed to have been orthogonalized. The square root may be defined by the representation

$$M^{1/2} = U \Lambda^{1/2} U^\dagger$$

where $\Lambda^{1/2}$ is diagonal, with diagonal elements $[\lambda(n)]^{1/2}$. The matrix $M^{-1/2}$ is similarly defined in terms of $\Lambda^{-1/2}$, and it is easily seen to be the inverse of $M^{1/2}$. Uniqueness of the square roots is not necessary for our purpose, only their existence and positive definite (hence also Hermitian) character.

Now consider the vector

$$\mathsf{z} \equiv M^{-1/2} z$$

and the similarly transformed secondaries

$$\mathsf{z}(k) \equiv M^{-1/2} z(k).$$

The new vectors are zero-mean Gaussian variables, but with covariance matrix equal to I_N, the $N \times N$ identity matrix. This follows directly from the definitions:

$$E\{\mathsf{z}\, \mathsf{z}^\dagger\} = M^{-1/2} E\{z\, z^\dagger\} M^{-1/2} = M^{-1/2} M M^{-1/2} = I_N$$

with identical reasoning for the transformed secondaries. The linear transformation introduced here is, of course, a whitening transformation.

We note that the scalar η depends on the data and signal vectors only through inner products. By inverting the whitening transformation we may evaluate, for example, the product

$$(z^\dagger S^{-1} z) = (\mathsf{z}^\dagger M^{1/2} S^{-1} M^{1/2} \mathsf{z})$$
$$= (\mathsf{z}^\dagger (M^{-1/2} S M^{-1/2})^{-1} \mathsf{z}).$$

We define the new matrix

$$\mathcal{S} \equiv M^{-1/2} S M^{-1/2}$$

and substitute for S, finding

$$\mathcal{S} = \sum_{k=1}^{K} M^{-1/2} z(k)\, z(k)^\dagger M^{-1/2} = \sum_{k=1}^{K} \mathsf{z}(k) \mathsf{z}(k)^\dagger .$$

Therefore, the new S matrix is K times the sample covariance matrix of the whitened secondaries, and the random variable

$$\Sigma \equiv (z^\dagger S^{-1} z) = (\mathsf{z}^\dagger \mathcal{S}^{-1} \mathsf{z})$$

is seen to be independent of M, being expressible as a function of $K+1$ independent Gaussian vectors, each of dimension N, and each sharing the covariance matrix, I_N. The PDF of Σ, like that of the RMB signal-to-noise ratio, is therefore a universal function of the dimensional parameters, N and K, alone.

The other inner products in the decision statistic are handled in an analogous manner; thus

$$(s^\dagger S^{-1} z) = (s^\dagger M^{-1/2} \mathcal{S}^{-1} \mathsf{z}) = (\mathsf{t}^\dagger \mathcal{S}^{-1} \mathsf{z})$$

where t stands for the whitened signal vector

$$\mathsf{t} \equiv M^{-1/2} s.$$

At this point we make the deferred definition of signal normalization, by taking t to be a unit vector:

$$(\mathsf{t}^\dagger \mathsf{t}) = (s^\dagger M^{-1} s) \equiv 1.$$

This choice gives specific meaning to the signal amplitude parameter b, whose square is now a proper signal-to-noise ratio, for which we introduce the symbol a:

$$a \equiv |b|^2 = (E\{z\}^\dagger M^{-1} E\{z\}).$$

When the obvious substitutions are made in the final inner product, we obtain

$$\eta = \frac{|(\mathsf{t}^\dagger \mathcal{S}^{-1} \mathsf{z})|^2}{(\mathsf{t}^\dagger \mathcal{S}^{-1} \mathsf{t})\,[1 + (\mathsf{z}^\dagger \mathcal{S}^{-1} \mathsf{z})]} .$$

The dependence on M is now confined to t, and it is shown below that even this dependence on the true covariance matrix is illusory. When a signal is present, z is replaced by

$$z = bs + n$$

where n has all the properties attributed to z in the noise-alone case. In this situation, the whitened data vector is

$$\mathsf{z} = M^{-1/2} z = b\mathsf{t} + \nu$$

where

$$\nu \equiv M^{-1/2} n$$

which is statistically identical to the whitened data vector in the noise-alone case. We have therefore found that when a signal is present, the PDF of η depends on M only through b and t, and the dependence on the unit vector t is again only apparent, as we now show.

Suppose the whitening transformation is followed by a unitary one, in which the whitened vectors are expressed as the products of a unitary matrix and a new set of random vectors. These new random vectors are statistically indistinguishable from their predecessors, and it would only be confusing to introduce a new notation for them. Tracing this transformation through the inner products, we find that only the normalized signal vector is changed: t is replaced by

$$t_1 \equiv U_1 t$$

where U_1 is the unitary matrix characterizing this last transformation. Any unit vector in the complex N space can be realized as t_1 by such a transformation. In particular, we can cause t_1 to be a coordinate vector, for which a single element is unity, the remaining $(N-1)$ elements vanishing. It is for this reason that the PDF of η depends on M only through the meaning of the signal amplitude parameter, b. In fact, this PDF can depend only on b, N, and K, and hence the false alarm probability of the likelihood ratio detector, namely

$$\mathrm{PFA} = \Pr\{\eta > \eta_0\}$$

is independent of M, and this is the generalized CFAR property claimed in Section I.

V. PROBABILITY DISTRIBUTION OF THE TEST STATISTIC

We now take advantage of our freedom to make a unitary transformation, and choose for t_1 a vector whose first element is unity, all others being zero. This can be accomplished by choosing for U_1 a matrix whose first row is the conjugate transpose of t, and whose other rows are the conjugates of unit vectors orthogonal to t. Understanding that this choice has been made, we drop the subscript on t, so that η is still given by the formula in Section IV.

This form for t makes it expedient to decompose all vectors into two components, an A component consisting of the first element only, and a B component consisting of the rest of the vector. Thus we write

$$z = \begin{bmatrix} z_A \\ z_B \end{bmatrix}$$

where the A component is a scalar and the B component is an $(N-1)$ vector. In this notation, the signal vector is just

$$t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

the zero being $(N-1)$ dimensional. Matrices are decomposed in analogous fashion, and we write

$$\mathcal{S} = \begin{bmatrix} \mathcal{S}_{AA} & \mathcal{S}_{AB} \\ \mathcal{S}_{BA} & \mathcal{S}_{BB} \end{bmatrix}.$$

Note that the AA element is a scalar, the BA element is an $(N-1)$ dimensional column vector, and so on. We also give a name to the inverse of this matrix, decomposing it as well:

$$\mathcal{S}^{-1} \equiv \mathcal{P} = \begin{bmatrix} \mathcal{P}_{AA} & \mathcal{P}_{AB} \\ \mathcal{P}_{BA} & \mathcal{P}_{BB} \end{bmatrix}.$$

With this notation we have, simply,

$$(t^\dagger \mathcal{S}^{-1} t) = (t^\dagger \mathcal{P} t) = \mathcal{P}_{AA}$$

while

$$(t^\dagger \mathcal{S}^{-1} z) = [1\ 0] \begin{bmatrix} \mathcal{P}_{AA} & \mathcal{P}_{AB} \\ \mathcal{P}_{BA} & \mathcal{P}_{BB} \end{bmatrix} \begin{bmatrix} z_A \\ z_B \end{bmatrix}$$

$$= \mathcal{P}_{AA} z_A + \mathcal{P}_{AB} z_B.$$

It is important to keep in mind that we now have a fourfold decomposition of the total input data set into primary and secondary vectors, each of which is divided into A and B components.

According to the Frobenius relations for partitioned matrices,

$$\mathcal{P}_{AA} = (\mathcal{S}_{AA} - \mathcal{S}_{AB} \mathcal{S}_{BB}{}^{-1} \mathcal{S}_{BA})^{-1}$$

which is a scalar, and also

$$\mathcal{P}_{BA} = -\mathcal{S}_{BB}{}^{-1} \mathcal{S}_{BA} \mathcal{P}_{AA}.$$

Since the $N \times N$ sample covariance matrix and its inverse are Hermitian, we obtain

$$\mathcal{P}_{AB} = \mathcal{P}_{BA}{}^\dagger = -\mathcal{P}_{AA} \mathcal{S}_{AB} \mathcal{S}_{BB}{}^{-1}$$

and therefore

$$(t^\dagger \mathcal{S}^{-1} z) = \mathcal{P}_{AA} (z_A - \mathcal{S}_{AB} \mathcal{S}_{BB}{}^{-1} z_B).$$

The final inner product is expanded as follows:

$$(z^\dagger \mathcal{S}^{-1} z) = \mathcal{P}_{AA} |z_A|^2 + 2\,\mathrm{Re}\,\{z_A^* \mathcal{P}_{AB} z_B\} + (z_B^\dagger \mathcal{P}_{BB} z_B)$$

where we have applied the identity

$$(u^\dagger v) = (v^\dagger u)^*$$

to the $(N-1)$ dimensional inner product

$$z_B^\dagger \mathcal{P}_{BA}.$$

Next we complete the square in this last expression, writing

$$(z^\dagger \mathcal{S}^{-1} z) = \mathcal{P}_{AA} |z_A + \mathcal{P}_{AA}^{-1} \mathcal{P}_{AB} z_B|^2$$
$$+ z_B^\dagger (\mathcal{P}_{BB} - \mathcal{P}_{AA}^{-1} \mathcal{P}_{BA} \mathcal{P}_{AB}) z_B$$

and by using a Frobenius relation in reverse we see that

$$\mathcal{P}_{BB} - \mathcal{P}_{AA}^{-1} \mathcal{P}_{BA} \mathcal{P}_{AB} = \mathcal{S}_{BB}^{-1}.$$

Finally, combining these results, we find

$$(z^\dagger \mathcal{S}^{-1} z) = \mathcal{P}_{AA} |z_A - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B|^2 + (z_B^\dagger \mathcal{S}_{BB}^{-1} z_B).$$

When these evaluations are substituted into our expression for η, the result can be expressed in the apparently simple form:

$$\eta = \frac{X}{1 + X + \Sigma_B}.$$

We have introduced here the notation

$$\Sigma_B = (z_B^\dagger \mathcal{S}_{BB}^{-1} z_B)$$

which is retained, and the temporary notation

$$X = \mathcal{P}_{AA} | z_A - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B |^2.$$

Note that Σ_B is just like the quantity Σ defined earlier, except that the dimensionality of the vectors involved is now $N-1$. The last form of the decision test, namely

$$\eta > \eta_0$$

is evidently equivalent to

$$X > \frac{\eta_0}{1 - \eta_0} (1 + \Sigma_B).$$

We leave the test in this form for a time, while we examine the statistical properties of the quantities which enter into it.

A previous evaluation for the leading factor in X can be used to obtain the following form:

$$X = \frac{| z_A - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B |^2}{\mathcal{S}_{AA} - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} \mathcal{S}_{BA}}.$$

We make use of the definitions to express the denominator as a sum:

$$\mathcal{S}_{AA} - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} \mathcal{S}_{BA} = \sum_{k=1}^{K} (z_A(k) - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B(k)) z_A(k)^*.$$

This is the same as the sum of squares

$$\sum_{k=1}^{K} | z_A(k) - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B(k) |^2$$

because the terms supplied to complete this square add up to zero:

$$\sum_{k=1}^{K} \left(z_A(k) - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B(k) \right) \left(\mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B(k) \right)^*$$

$$= \sum_{k=1}^{K} z_A(k) z_B(k)^\dagger \mathcal{S}_{BB}^{-1} \mathcal{S}_{BA} - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1}$$

$$\times \sum_{k=1}^{K} z_B(k) z_B(k)^\dagger \mathcal{S}_{BB}^{-1} \mathcal{S}_{BA}$$

$$= \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} \mathcal{S}_{BA} - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} \mathcal{S}_{BB} \mathcal{S}_{BB}^{-1} \mathcal{S}_{BA} = 0.$$

The evaluation of the sums here follows from the definitions of the partitioned matrix elements. We introduce the notation

$$y(k) \equiv z_A(k) - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B(k)$$

for the terms of the sum, and the analogous notation

$$y \equiv z_A - \mathcal{S}_{AB} \mathcal{S}_{BB}^{-1} z_B$$

for the quantity appearing in the numerator of X, so that the likelihood ratio test can be written in the more explicit form

$$X = \frac{|y|^2}{\sum_{k=1}^{K} |y(k)|^2} > \frac{\eta_0}{1 - \eta_0} (1 + \Sigma_B).$$

We proceed by fixing the B vectors temporarily, and consider the probability densities of all quantities entering into the decision statistic to be conditioned on these values. The conditional probabilities of detection and false alarm are evaluated first, and the condition is then removed by taking expectation values over the joint PDF of the B vectors. With the B vectors fixed, only the $K+1$ scalar A components are random, and we show now, under this condition, that y and the $y(k)$ are Gaussian variables, that y is uncorrelated with the $y(k)$, and that the latter have a covariance matrix with simple properties.

Using the definitions of the ys and of the AB matrix element which enters there, we can express these quantities in the form

$$y = z_A - \sum_{k=1}^{K} z_A(k) z_B(k)^\dagger \mathcal{S}_{BB}^{-1} z_B$$

and

$$y(k) = z_A(k) - \sum_{i=1}^{K} z_A(i) z_B(i)^\dagger \mathcal{S}_{BB}^{-1} z_B(k).$$

This represents the ys as linear combinations of the A components, and hence proves their conditional Gaussian character. Moreover, the $y(k)$ have zero mean in all cases, while the conditional mean, written E_B, of y in the general case is

$$E_B\, y = E z_A = b$$

as a result of our choice of signal vector.

The linear dependence of the ys on the A components is best expressed in terms of the quantities

$$q(k) \equiv z_B(k)^\dagger \mathcal{S}_{BB}^{-1} z_B$$

and

$$Q(i,k) \equiv z_B(i)^\dagger \mathcal{S}_{BB}^{-1} z_B(k)$$

which are constants under the conditioning. Obviously,

$$y = z_A - \sum_{k=1}^{K} z_A(k) q(k)$$

and

$$y(k) = z_A(k) - \sum_{k=1}^{K} z_A(i) Q(i,k).$$

The $q(k)$ may be considered as the components of a K vector q, and the $Q(i,k)$ as the elements of a $K \times K$ matrix Q. The desired properties of the ys flow from the following facts about this new vector and matrix:

$$Qq = q$$

and

$$Q^2 = Q.$$

To prove the first of these, we write it out in component form:

$$\sum_{i=1}^{K} Q(k,i)q(i) = \sum_{i=1}^{K} z_B(k)^\dagger \mathcal{S}_{BB}^{-1} z_B(i) z_B(i)^\dagger \mathcal{S}_{BB}^{-1} z_B.$$

The sum over i regenerates the BB matrix element:

$$\sum_{i=1}^{K} z_B(i) z_B(i)^\dagger = \mathcal{S}_{BB}$$

(as happened when the denominator of X was expressed as a sum of squares), and the result follows immediately. The idempotent character of Q is proved in the same way. We also note that Q is Hermitian, and that its trace is $N-1$:

$$\begin{aligned}\mathrm{Tr}(Q) &= \sum_{k=1}^{K} z_B(k)^\dagger \mathcal{S}_{BB}^{-1} z_B(k) \\ &= \mathrm{Tr}\Big(\mathcal{S}_{BB}^{-1} \sum_{k=1}^{K} z_B(k) z_B(k)^\dagger\Big) \\ &= \mathrm{Tr}(I_{N-1}) = N - 1.\end{aligned}$$

Note that we are dealing with the trace of a $K \times K$ matrix on the left side here, and of $(N-1) \times (N-1)$ matrices on the right. All of these results will be required in the following.

The fact that y and the $y(k)$ are conditionally uncorrelated now follows easily from the independence of the A components themselves:

$$\begin{aligned}E_B y y(k)^* &= -E_B\left(\sum_{i=1}^{K} q(i) z_A(i) z_A(k)^*\right) \\ &\quad + E_B\left(\sum_{i=1}^{K} q(i) z_A(i) \sum_{n=1}^{K} z_A(n)^* Q(n,k)^*\right) \\ &= -q(k) + \sum_{i=1}^{K} q(i) Q(k,i) = 0.\end{aligned}$$

Next, consider the conditional variance of y:

$$E_B|y-b|^2 = 1 + \sum_{k=1}^{K} |q(k)|^2.$$

Substituting for the $q(k)$, we have

$$\begin{aligned}\sum_{k=1}^{K} |q(k)|^2 &= \sum_{k=1}^{K} z_B^\dagger \mathcal{S}_{BB}^{-1} z_B(k) z_B(k)^\dagger \mathcal{S}_{BB}^{-1} z_B \\ &= z_B^\dagger \mathcal{S}_{BB}^{-1} z_B = \Sigma_B\end{aligned}$$

and hence

$$E_B|y-b|^2 = 1 + \Sigma_B.$$

This last result is responsible for a significant simplification of the statistics of the likelihood ratio test.

Finally, we compute the conditional covariance of the $y(k)$. We use the notation $\delta(i,k)$ for the elements of the unit matrix, so that $y(k)$ can be written

$$y(k) = \sum_{i=1}^{K} z_A(i)[\delta(i,k) - Q(i,k)].$$

Using the independence of the A variables again, we obtain

$$\begin{aligned}Ey(k)y(n)^* &= \sum_{i=1}^{K} [\delta(i,k) - Q(i,k)][\delta(i,n) - Q(i,n)]^* \\ &= \delta(k,n) - Q(n,k) - Q(k,n)^* \\ &\quad + \sum_{i=1}^{K} Q(i,k)Q(i,n)^*.\end{aligned}$$

Since Q is Hermitian and idempotent, we find the simple result

$$Ey(k)y(n)^* = \delta(n,k) - Q(n,k).$$

The likelihood ratio test is now rearranged slightly to read

$$\frac{|y|^2}{1+\Sigma_B} > \frac{\eta_0}{1-\eta_0} \sum_{k=1}^{K} |y(k)|^2.$$

In view of fact that the conditional variance of y equals the denominator on the left, it makes sense to define a normalized variable

$$w \equiv \frac{y}{(1+\Sigma_B)^{1/2}}.$$

Conditioned on the B vectors, w is Gaussian and independent of the $y(k)$. It has a conditional variance of unity, and a conditional mean value:

$$E_B w = \frac{b}{(1+\Sigma_B)^{1/2}}.$$

In the noise-alone case, the conditioning has no effect on the PDF of w. The sum over k is also given a name:

$$T \equiv \sum_{k=1}^{K} |y(k)|^2$$

and the test is now written

$$|w|^2 > (\ell_0 - 1)T.$$

where the original threshold constant has been reintroduced. In fact, it is easily verified that our original likelihood ratio is given by

$$\ell = 1 + \frac{|w|^2}{T}.$$

We now turn to the properties of T. Given the B vectors, the joint PDF of the $y(k)$ is zero-mean Gaussian with covariance matrix J:

$$J(i,k) \equiv \delta(i,k) - Q(k,i).$$

The conditional characteristic function of T is therefore

$$\Phi_B(\lambda) \equiv E_B\{e^{i\lambda T}\} = \|I - i\lambda J\|^{-1}.$$

Since Q is idempotent, its eigenvalues are either zero or one, and from the value of its trace we see that Q must have exactly $N-1$ unit eigenvectors. It follows that J has $K+1-N$ unit eigenvectors, the others being zero, and thus

$$\Phi_B(\lambda) = (1-i\lambda)^{-(K+1-N)}.$$

This is the characteristic function of a chi-squared random variable, and the PDF of T is simply

$$f_B(T) = \frac{T^{K-N}}{(K-N)!}e^{-T}.$$

It is remarkable that the statistical properties of T are independent of the actual values of the conditioning B-vectors, and we can consequently drop the subscript on its PDF. Moreover, T is statistically equivalent to the sum of the squares of $K+1-N$ independent, complex Gaussian variables, each of which has zero mean and unit variance. If we let $w(k)$, $(k=1 \ldots K+1-N)$, be such a set, then T is statistically indistinguishable from the sum

$$\sum_{k=1}^{K+1-N} |w(k)|^2.$$

The properties of the likelihood ratio test are therefore identical to the properties of the simple test

$$|w|^2 > (\ell_0 - 1)\sum_{k=1}^{K+1-N} |w(k)|^2$$

where the $w(k)$ are now also taken to be independent of w. The probability of the truth of this inequality is still conditioned on the B vectors, but this conditioning appears only through the quantity Σ_B, which is contained in the conditional mean of w.

This equivalent decision rule represents the behavior of a simple scalar CFAR test, in which the power in one complex sample (a single radar hit), being tested for signal presence, is compared to a threshold proportional to the sum of the powers of $K+1-N$ samples of noise. This problem is quite familiar, and the test just described is also a likelihood ratio test in the corresponding situation. The performance of the scalar CFAR detector is very simple, and in particular, its PFA is just

$$\left(\frac{1}{\ell_0}\right)^{K+1-N}$$

In this case, when the signal amplitude is zero, the conditioning B vectors do not appear at all, and hence this simple formula gives the PFA for our original likelihood ratio test.

The probability of detection (PD) of the scalar CFAR test is also well known, and in our case it depends on the conditional SNR, which is the squared magnitude of the conditional mean of w. In terms of the colored noise matched filter SNR a defined earlier, and the quantity

$$r = \frac{1}{1+\Sigma_B}$$

this conditional SNR is just ra. The factor r represents a loss factor, applied to the SNR, and caused by the necessity of estimating the noise covariance matrix. The PD of the CFAR detector can be expressed in a particularly convenient way as a finite sum [4]:

$$P_D = 1 - \frac{1}{\ell_0^L}\sum_{k=1}^{L}\binom{L}{k}(\ell_0-1)^k G_k\left(\frac{ra}{\ell_0}\right)$$

where $L = K+1-N$. The function G which enters here is itself a finite sum:

$$G_k(y) = e^{-y}\sum_{n=0}^{k-1}\frac{y^n}{n!}.$$

In order to complete our computation of the PD of the likelihood ratio test, we must take the expectation value of this conditional PD over the joint PDF of the B vectors. These, however, enter the final result only through the loss factor r which acts as a fluctuation model for the signal. Unlike more familiar fluctuation models, this one is characterized by a factor lying in the range zero to one. The present situation is similar to that discussed in the RMB paper, except that besides an SNR loss, our test will suffer a CFAR loss as well, when compared with a colored noise matched filter test in which everything is known concerning the noise or interference.

Although the loss factor found here depends on the primary data, through its B component, while the RMB loss factor is a function only of the secondary data, it turns out that the two factors have exactly the same PDF. The proof of this interesting result is deferred to the Appendix, in which the RMB loss factor is also expressed in our notation, and the evaluation of the PDFs of both these quantities is carried out in parallel.

The PDF shared by these loss functions is the beta distribution:

$$f(r) = \frac{(N+L-1)!}{L!(N-2)!}(1-r)^{N-2}r^L$$

and the final expression for the PD of our test can be written

$$P_D = 1 - \frac{1}{\ell_0^L}\sum_{k=1}^{L}\binom{L}{k}(\ell_0-1)^k H_k\left(\frac{a}{\ell_0}\right).$$

In this formula, the H functions are the expected values of the Gs:

$$H_k(y) = \int_0^1 G_k(ry)f(r)\,dr.$$

These integrals are elementary, although not simple, and their detailed evaluation is not given here. The results of Section VI are based on these formulas.

VI. NUMERICAL RESULTS AND DISCUSSION

The performance of the likelihood ratio test depends only on the dimensional integers N and K and the SNR parameter a. The latter is a function of the true signal strength and the intensity and character of the actual noise and interference. Our analysis deals with a very general problem, and nothing can be said about the anticipated values of a. The ability of a system to function effectively in interference depends principally on the arrangements which have been made in its design to achieve a good colored noise matched filter SNR in its intended environment. These arrangements will usually take the form of diversity of RF inputs in one form or another. An additional requirement is the need to have inputs available from which the actual noise characteristics can be estimated, and this is the aspect of the problem which has been addressed here. In particular, for given values of PD and PFA, we can determine what SNR is actually required to achieve those values using the likelihood ratio detector, and compare that number with the SNR which would be adequate to achieve identical performance if the noise covariance matrix were known in advance. The difference is the penalty for having to estimate the noise covariance, and we expect that penalty to vary sharply with the number K of available secondary input vectors.

This penalty has two components, one due to the CFAR character of the decision rule and another due to the effective SNR loss factor. The latter is expected to behave much as the results of the RMB analysis would predict, based on the statistical properties of the loss factor alone. The CFAR loss will decrease as the value of K increases, and it may be expected to depend largely on that parameter, while the SNR loss effect depends roughly on the ratio of K to N.

These expectations are borne out by the numerical consequences of our analysis, as shown in the accompanying figures. In Figs. 1–4, PD is shown as a function of a, the SNR, for three detectors (PFA is fixed at 10^{-6} for these curves). The detector performing best is a matched filter with known noise covariance, and the worst is the likelihood ratio detector which, of course, is estimating the noise covariance. The middle curve (dashed) in these plots shows the performance of a simple, scalar CFAR detector using $L=K+1-N$ noise samples, and it differs from the behavior of the likelihood ratio detector only in that the SNR loss factor has been ignored. This detector is included in the comparison in order to show how much of the degradation imposed by noise estimation is due to each of the two contributing effects.

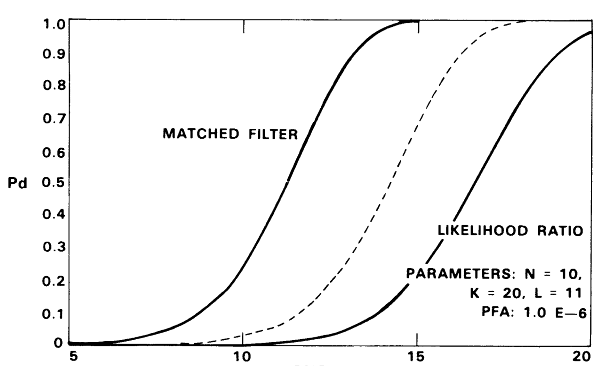


Fig. 1.

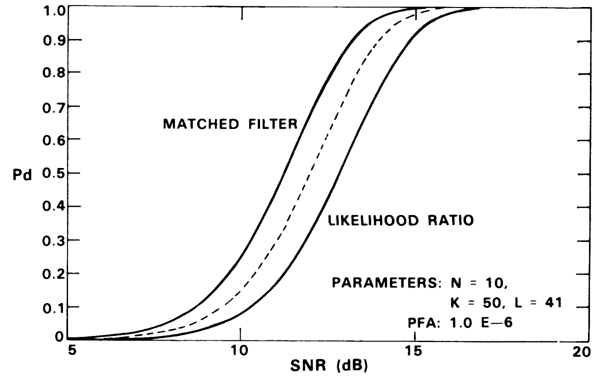


Fig. 2.

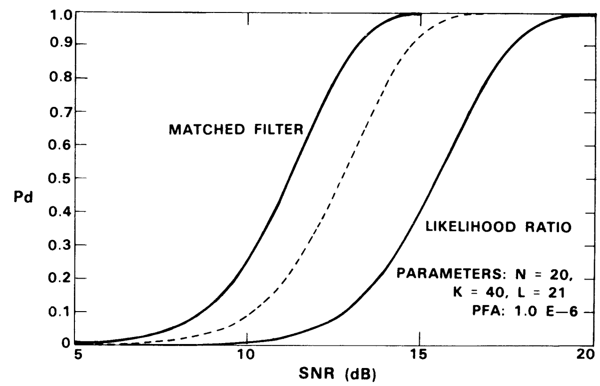


Fig. 3.

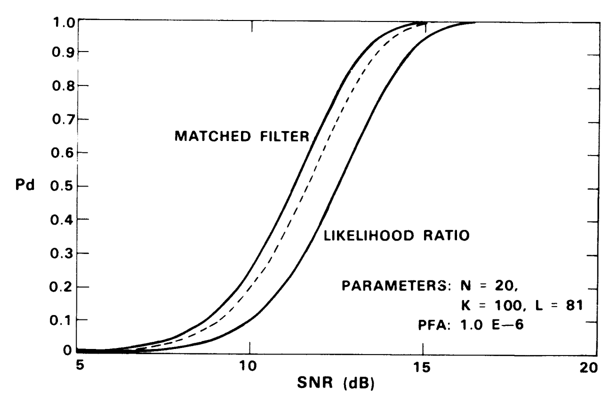


Fig. 4.

We note that doubling both K and N has little effect on the portion of the degradation due to SNR loss, while the CFAR part is reduced, simply because K is being increased. The curves also show the significant improvement which results from increasing the ratio of K to N. When this ratio is equal to 5, the SNR loss contribution to the performance degradation is about 0.9 dB, in agreement with the mean value of the SNR loss as obtained from the beta distribution. Likewise, the CFAR contributions are directly comparable with the ordinary CFAR loss for a detector of nonfluctuating targets with no noncoherent integration (i.e., a single radar hit).

The detector performance is characterized in a different way in Figs. 5–8, which show the additional SNR required, when estimating the noise covariance, to achieve the same PD and PFA as a matched filter for known noise. In all these figures, the PD is specified at 0.9, but the results will not depend strongly on the chosen PD level, since the curves of PD versus SNR are nearly parallel for the two detectors. Three PFA values are represented on each plot. The independent variable for these curves is the number of secondary vectors, and this variable always covers the range $2N$ through $5N$, for 4 different N values. The SNR loss is not strictly a function of the ratio K/N, but generally decreases with increasing N, with this ratio held constant. The loss shown is the total loss, due to the CFAR effect and the SNR loss factor itself.

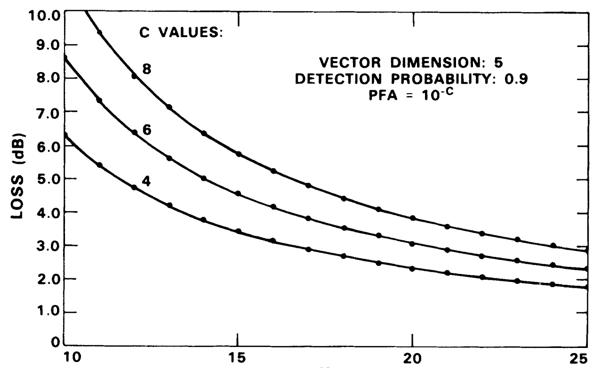


Fig. 5.

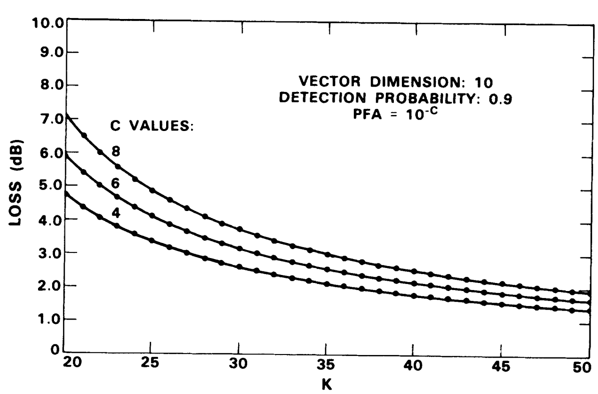


Fig. 6.

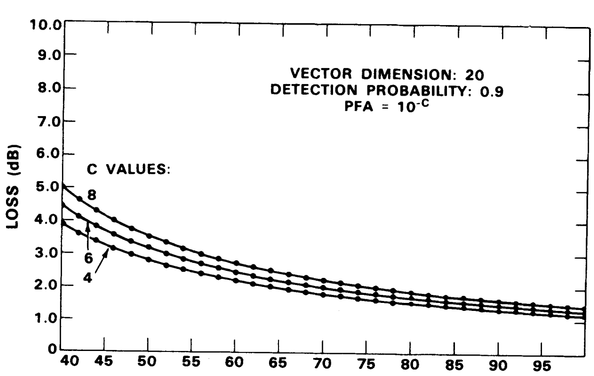


Fig. 7.

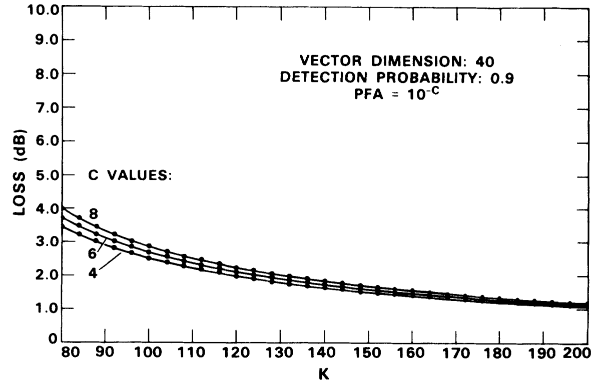


Fig. 8.

It was noted earlier that K, the number of secondary vectors, must exceed N, the dimension of each of the data vectors, in order to have a nonsingular sample covariance matrix. It is clear from the results just discussed that K must exceed N by a significant factor if noise estimation is not to cause a serious loss in performance. Since N is the dimension of the total vector of data used for detection, the requirements on the number of secondaries can be extremely large. In the radar example mentioned in Section II, N was the product of the number of RF channels N_a and the number of pulses N_p in a coherent processing interval. To supply enough secondaries from range-gate outputs, other than the one being tested for signal presence, the bandwidth of the radar would have to exceed some minimum value.

The chief reason for the requirement of many secondaries is the generality of our formulation, in which any interference covariance matrix is allowed. In the radar example, this includes the possibility of arbitrary correlation between interference inputs from pulse returns widely spaced in time, although it might well be realistic to assume independence (but not statistical identity) of the interference inputs accompanying distinct pulse returns. This is a more difficult problem, for which only an approximate solution has been found [5].

APPENDIX. THE PROBABILITY DENSITY FUNCTION OF THE LOSS FACTOR

The SNR loss factor derived in the text was expressed in the form

$$r = \frac{1}{1 + \Sigma_B}$$

where

$$\Sigma_B = z_B^\dagger \mathcal{S}_{BB}^{-1} z_B .$$

Before discussing the PDF of Σ_B, from which the PDF of r follows easily, we express the RMB loss factor ρ in our notation. From the RMB paper,

$$\rho = \frac{|(\hat{w}^\dagger s)|^2}{(s^\dagger M^{-1} s)(\hat{w}^\dagger M \hat{w})}$$

where M is the actual noise covariance matrix, s is the signal vector, and $\hat{w}$ is a weight vector:

$$\hat{w} = k\hat{M}^{-1}s.$$

In this last formula, $\hat{M}$ is the sample covariance matrix of the secondary data and k is an arbitrary constant. The loss factor itself is the ratio of the conditional SNR of the output of a filter which uses $\hat{w}$ as a weight, relative to the SNR of the colored noise matched filter for known M. The conditioning in this case corresponds to given values of the secondary data.

Choosing $k = 1/K$, we obtain

$$\hat{w} = S^{-1}s$$

in terms of the S matrix used in the text, and then

$$\rho = \frac{(s^\dagger S^{-1}s)^2}{(s^\dagger M^{-1}s)(s^\dagger S^{-1}MS^{-1}s)}.$$

Note that ρ is unaffected if s is changed by a constant factor.

We now carry out the whitening transformation, as in the text, and normalize the signal vector as before. The result is

$$\rho = \frac{(t^\dagger \mathcal{S}^{-1}t)^2}{(t^\dagger \mathcal{S}^{-2}t)}$$

$$= \frac{(t^\dagger \mathcal{P} t)^2}{(t^\dagger \mathcal{P}^2 t)}.$$

A unitary transformation is now applied to convert the signal vector to the final one used in the text, and the matrices are decomposed in the same way. This gives the simple expression

$$\rho = \frac{(\mathcal{P}_{AA})^2}{(\mathcal{P}^2)_{AA}}.$$

It is clear at this point that the PDF of ρ will be independent of the actual covariance matrix M.

Using the Frobenius relations, we obtain

$$(\mathcal{P}^2)_{AA} = (\mathcal{P}_{AA})^2 + \mathcal{P}_{AB}\mathcal{P}_{BA}$$

$$= (\mathcal{P}_{AA})^2 \left(1 + \mathcal{S}_{AB}\mathcal{S}_{BB}^{-2}\mathcal{S}_{BA}\right)$$

and therefore

$$\rho = \frac{1}{1 + \Sigma_\rho}$$

where

$$\Sigma_\rho \equiv \mathcal{S}_{AB}\mathcal{S}_{BB}^{-2}\mathcal{S}_{BA}.$$

Note that the RMB loss factor depends on the secondary data only, both A and B components, while r depends on the B components of both primary and secondary data.

We proceed to analyze the two loss factors together, and begin by conditioning on the B components of the secondary data vectors, on which both loss factors depend. Then

$$\mathcal{S}_{BB} = \sum_{k=1}^{K} z_B(k) z_B(k)^\dagger$$

is a constant matrix, positive definite, and nonsingular for all sets of conditioning vectors (except for a set of probability zero). We can therefore introduce the square root of this matrix and define the vectors

$$\xi_B \equiv \mathcal{S}_{BB}^{-1/2} z_B$$

and

$$\xi_\rho \equiv \mathcal{S}_{BB}^{-1}\mathcal{S}_{BA}.$$

With the conditioning, these quantities are zero-mean Gaussian vectors; the former is a linear function of the elements of the B component of the whitened primary vector, and the latter is expressible in terms of the secondary A components:

$$\xi_\rho = \mathcal{S}_{BB}^{-1} \sum_{k=1}^{K} z_B(k) z_A(k)^\dagger.$$

We use the subscript C to denote the present conditioning, and compute the conditional covariance matrices of these vectors:

$$E_C \xi_B \xi_B^\dagger = \mathcal{S}_{BB}^{-1/2} E_C z_B z_B^\dagger \mathcal{S}_{BB}^{-1/2} = \mathcal{S}_{BB}^{-1}$$

and

$$E_C \xi_\rho \xi_\rho^\dagger = \mathcal{S}_{BB}^{-1} W \mathcal{S}_{BB}^{-1}$$

where

$$W = E_C \sum_{k=1}^{K} z_B(k) z_A(k)^\dagger \sum_{i=1}^{K} z_A(i) z_B(i)^\dagger$$

$$= \sum_{k=1}^{K} z_B(k) z_B(k)^\dagger = \mathcal{S}_{BB}.$$

Therefore

$$E_C \xi_\rho \xi_\rho^\dagger = \mathcal{S}_{BB}^{-1}$$

and the two ξ vectors are statistically equivalent under the conditioning. They therefore share the same final PDF when the conditioning is removed by averaging over the secondary B vectors. Since

$$\Sigma_B = (\xi_B^\dagger \xi_B)$$

and

$$\Sigma_\rho = (\xi_\rho^\dagger \xi_\rho)$$

this proves that the loss factors themselves are statistically identical, and we continue with the loss factor r.

Since ξ_B is a Gaussian $(N-1)$ vector, the conditional joint PDF of its components is

$$f_c(\xi_B) = \frac{\|\mathcal{S}_{BB}\|}{\pi^{N-1}} e^{-(\xi_B^\dagger \mathcal{S}_{BB} \xi_B)}.$$

The S matrix which enters here is itself subject to the Wishart PDF, which in the present case (in which the sample vectors are of dimension $N-1$ and the covariance matrix is equal to the identity) takes the form

$$f_w(A) = \frac{\|A\|^{K+1-N}}{C(N-1,K)} e^{-\mathrm{tr}(A)}.$$

In this formula

$$C(N,K) = \pi^{\frac{N(N-1)}{2}} \prod_{n=1}^{N} (K-n)!$$

is the Wishart normalization factor. The volume element for this PDF will be written $d(A)$. It is $(N-1)^2$ dimensional, ranging over the diagonal elements and the real and imaginary parts of the upper off-diagonals of all positive definite matrices, A. For our purpose, only the normalization integral of the Wishart PDF is required, hence we need not dwell on the detailed properties of this fascinating distribution.

Since ξ_B depends on the conditioning data only through the S matrix, its unconditioned PDF can be written

$$f(\xi_B) = \frac{1}{\pi^{N-1}} \int \cdots \int \|A\| \, e^{-\mathrm{tr}(A\mathcal{C})} f_w(A) \, d(A).$$

As in the text, we have replaced the exponential part of the Gaussian PDF by a trace, this time involving the open product matrix

$$\mathcal{C} \equiv \xi_B \xi_B^{\dagger}.$$

When we substitute for the Wishart PDF in the expression above, we encounter the integral

$$\int \cdots \int \|A\|^{K+2-N} e^{-\mathrm{tr}[A(I_{N-1} + \mathcal{C})]} \, d(A).$$

This is the same as the normalization integral for another Wishart PDF, of dimensions $N-1$ and $K+1$, and for which the underlying sample vectors share the covariance matrix

$$\mathcal{M} \equiv (I_{N-1} + \mathcal{C})^{-1}.$$

The normalization factor for this slightly more general Wishart PDF is just

$$C(N-1,K+1)\|\mathcal{M}\|^{K+1} = \frac{C(N-1,K+1)}{\|I_{N-1} + \mathcal{C}\|^{K+1}}$$

$$= \frac{C(N-1,K+1)}{[1 + (\xi_B^{\dagger}\xi_B)]^{K+1}}$$

(the evaluation of the determinant uses the same Lemma utilized by Section III. Combining these facts, we obtain the simple result

$$f(\xi_B) = \frac{1}{\pi^{N-1}} \frac{C(N-1,K+1)}{C(N-1,K)} [1 + (\xi_B^{\dagger}\xi_B)]^{-(K+1)}$$

$$= \frac{K!}{\pi^{N-1}(K+1-N)!} (1+\Sigma_B)^{-(K+1)}.$$

The remainder of the derivation is identical to the final few steps given in [1, Appendix]. The norm of the vector ξ_B is interpreted as the square of the radial coordinate in a $(2N-2)$-dimensional Cartesian space, a change to polar coordinates is made, and the angular coordinates integrated out. This process yields the PDF of Σ_B, and then a simple change of variable provides the desired PDF of r:

$$f(r) = \frac{K!}{(K+1-N)!(N-2)!} (1-r)^{N-2} r^{K+1-N}.$$

REFERENCES

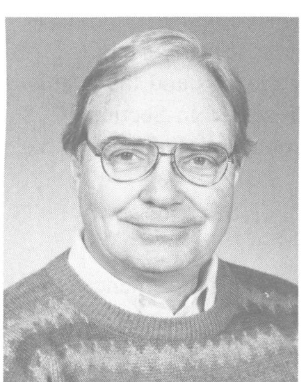
[1] Reed, I.S., Mallett, J.D., and Brennan, L.E. (1974)
Rapid convergence rate in adaptive arrays.
IEEE Transactions on Aerospace and Electronic Systems, AES-10 (Nov. 1974), 853–863.

[2] Capon, J., and Goodman, N.R. (1970)
Probability distributions for estimators of the frequency-wavenumber spectrum.
Proceedings of the IEEE, 58 (Oct. 1970), 1785–1786.

[3] Goodman, N.R. (1963)
Statistical analysis based on a certain multivariate complex Gaussian distribution.
Annals of Mathematical Statistics, 34 (Mar. 1963), 152–177.

[4] Kelly, E.J. (1981)
Finite-sum expressions for signal detection probabilities.
Technical Report 566, Lincoln Laboratory, M.I.T., May 20, 1981.

[5] Kelly, E.J. (1985)
Adaptive detection in non-stationary interference.
Technical Report 724, Lincoln Laboratory, M.I.T., June 25, 1985.



Edward J. Kelly was born in Philadelphia PA, in 1924. He received the S.B. degree in 1945, and the Ph.D. degree in 1950, both from the Massachusetts Institute of Technology, Cambridge, MA, in theoretical physics.

He was a staff member at Brookhaven National Laboratory, Upton, NY, from 1950 to 1952, and joined M.I.T. Lincoln Laboratory in 1952, where he is now a member of the Senior Staff.

His work at Lincoln Laboratory has been in the areas of radar detection and estimation theory, applied seismology, air traffic control, spread spectrum communications and, currently, space-based radar systems analysis.