DETECTION IN INCOMPLETELY CHARACTERIZED COLORED NON-GAUSSIAN NOISE VIA PA. (U) RHODE ISLAND UNIV KINGSTON DEPT OF ELECTRICAL ENGINEERING S KAY ET AL.

UNCLASSIFIED AUG 86 TR-5 N00014-84-K-0527

F/G 12/1
Detection in Incompletely Characterized Colored Non-Gaussian Noise via Parametric Modeling

STEVEN KAY AND DEBASIS SENGUPTA
Department of Electrical Engineering
University of Rhode Island
Kingston, Rhode Island 02881

August 1986

Prepared for

OFFICE OF NAVAL RESEARCH
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract N00014-84-K-0527
S.M. Kay, Principal Investigator

Approved for public release; distribution unlimited
# REPORT DOCUMENTATION PAGE

1. REPORT NUMBER 5

4. TITLE and Subtitle: Detection in Incompletely Characterized Colored Non-Gaussian Noise via Parametric Modeling

7. AUTHOR(S): Steven Kay and Debasis Sengupta

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Electrical Engineering Department, University of Rhode Island, Kingston, Rhode Island 02881

11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, Virginia 22217

12. REPORT DATE: August 1986

13. NUMBER OF PAGES: 33

14. MONITORING AGENCY NAME AND ADDRESS (IF DIFFERENT FROM CONTROLLING OFFICE): Office of Naval Research, Department of the Navy, Arlington, Virginia 22217

15. SECURITY CLASS: Unclassified


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

18. SUPPLEMENTARY NOTES:


21. ABSTRACT: The problem of detecting a signal known except for amplitude in incompletely characterized colored non-Gaussian noise is addressed. The problem is formulated as a testing of hypotheses using parametric models for the statistical behavior of the noise. A generalized likelihood ratio test is employed. It is shown that for a symmetric noise probability density function the detection performance is asymptotically equivalent to that obtained for a detector designed with a priori knowledge of the noise.
Abstract

The problem of detecting a signal known except for amplitude in incompletely characterized colored non-Gaussian noise is addressed. The problem is formulated as a testing of composite hypotheses using parametric models for the statistical behavior of the noise. A generalized likelihood ratio test is employed. It is shown that for a symmetric noise probability density function the detection performance is asymptotically equivalent to that obtained for a detector designed with a priori knowledge of the noise parameters. Non-Gaussian distributions of the noise are found to be more favorable for the purpose of detection as compared to the Gaussian distribution.
I. Introduction

The theory of detection of a known signal in presence of Gaussian noise having a known covariance matrix is well developed [Van Trees 1968]. In many applications, however, the covariance matrix is not known a priori. This difficulty can be alleviated by characterizing the correlation pattern of the noise by a simple model and using estimates of the model parameters to design a detector [Whalen 1971], [Bowyer et al 1979], [Kay 1983]. The difficulty increases when full information regarding the noise probability density function (PDF), usually assumed to be Gaussian, is unavailable due to insufficient knowledge about the noise source [Knight et al 1981]. There is no uniformly most powerful (UMP) test in this case because the use of a Neyman-Pearson criterion leads to a detector which depends on the unknown parameters. The Bayesian method of assigning priors to the unknown parameters of the noise PDF produces an ‘optimal’ detector [Lee et al 1977], but requires a multidimensional integration. Its performance is critically dependent on the accuracy of the choice of priors. A robust detector [Kassam and Poor 1985], on the other hand, does not use any partial knowledge about the noise PDF and therefore is not expected to perform well. Locally optimal (LO) detectors for this problem have been studied extensively by Czarnecki, Martinez, Thomas and others [Czarnecki and Thomas 1984], [Martinez and Thomas 1982]. Their results however rely on a known covariance matrix and marginal PDF of the noise. A third dimension is added to the problem if the amplitude of the signal is not known [Kay 1985]. A locally optimal detector can not be used since it depends on the polarity or sign of the amplitude, which is usually unknown.

This paper addresses the problem of detecting a deterministic signal known except for amplitude in the presence of incompletely characterized non-white non-Gaussian noise. The approach chosen here is to use the theory of the generalized likelihood ratio test (GLRT) for composite hypothesis testing [Kendall and Stuart 1979]. The work presented here is an extension of the work of Kay [1985] in which the noise is assumed
to be non-Gaussian but white. In this case the covariance matrix is assumed to be known except for a few parameters. Maximum likelihood estimates (MLE) for these parameters are then used in the GLRT. The asymptotic performance of the GLRT detector is shown to be equivalent to the asymptotic performance of the clairvoyant GLRT detector (one which uses perfect knowledge of the unknown parameters) for a symmetric noise PDF. Therefore the GLRT asymptotically achieves an upper bound in performance and is *optimal* in this sense.

The paper is organized as follows. Section II gives the theory of the GLRT which will be used extensively in the subsequent sections. Section III formulates the detection problem and derives the GLRT for it. The case of autoregressive (AR) noise is considered separately. Section IV discusses the performance of the GLRT detector and compares it to that of the clairvoyant GLRT detector. Section V draws some general conclusions about the performance of the GLRT. Section VI summarizes the results and discusses the implementation aspect of the problem.

II. Review of Generalized Likelihood Ratio Test

Consider the problem of testing the value of the parameter $\Theta = [\Theta_r^T \Theta_s^T]^T$ based on the data set $y = [y_1 y_2 \cdots y_N]$. $\Theta_r$ and $\Theta_s$ are assumed to be vectors of dimension $r$ and $s$, respectively. A common hypothesis test is

$$
\mathcal{H}_0 : \Theta^T = [0^T \Theta_s^T] \quad (1)
$$

$$
\mathcal{H}_1 : \Theta^T = [\Theta_r^T \Theta_s^T] \quad \Theta_r \neq 0
$$

$\Theta_s$, referred to as the vector of nuisance parameters, is of no concern and may assume any value. Assuming the observed data $y$ has a joint probability function $f(y; \Theta_r, \Theta_s)$, a *generalized likelihood ratio test* for testing (1) is to decide $\mathcal{H}_1$ if

$$
\tau_G = \frac{f(y; \hat{\Theta}_r, \hat{\Theta}_s)}{f(y; 0, \hat{\Theta}_s)} > \gamma \quad (2)
$$

for some threshold $\gamma$. $0$ is an $r$-dimensional vector of zeroes. $\hat{\Theta}_s$ is the MLE of $\Theta_s$ assuming $\mathcal{H}_0$ is true while $\hat{\Theta}_r$ and $\hat{\Theta}_s$ are *joint* MLE's of $\Theta_r$ and $\Theta_s$ assuming $\mathcal{H}_1$ is
true. \( \hat{\Theta}_s \) is found by maximizing \( f(y; 0, \Theta_s) \) over \( \Theta_s \). Similarly, \( \hat{\Theta}_r, \hat{\Theta}_s \) are obtained by maximizing \( f(y; \Theta_r, \Theta_s) \) over \( \Theta_r \) and \( \Theta_s \).

The statistics of \( \ell_G \) are difficult to obtain in general. For large data records (asymptotically) it may be shown that \( 2 \ln \ell_G \) is distributed in the following manner [Kendall and Stuart 1979].

\[
2 \ln \ell_G \sim \chi^2_r \quad \text{under } \kappa_0 \quad (3a)
\]
\[
2 \ln \ell_G \sim \chi'^2(r, \lambda) \quad \text{under } \kappa_1 \quad (3b)
\]

Here \( \chi^2_r \) represents a chi-square distribution with \( r \) degrees of freedom and \( \chi'^2(r, \lambda) \) represents a noncentral chi-square distribution with \( r \) degrees of freedom and noncentrality parameter \( \lambda \). Note that \( \chi'^2(r, 0) = \chi^2_r \) or the distribution under \( \kappa_0 \) is a special case of the distribution under \( \kappa_1 \) and occurs when \( \lambda = 0 \). The noncentrality parameter \( \lambda \), which is a measure of the discrimination between the two hypotheses, is given by

\[
\lambda = \Theta_r^T [I_{\Theta, \Theta}(0, \Theta_s) - I_{\Theta, \Theta}(0, \Theta_s)I_{\Theta, \Theta}^{-1}(0, \Theta_s)I_{\Theta, \Theta}^T] \Theta_r \quad (4)
\]

where \( \Theta_r, \Theta_s \) are the true values. The terms in the brackets of (4) are found by partitioning the Fisher information matrix for \( \Theta \) as

\[
I(\Theta) = \begin{pmatrix}
I_{\Theta, \Theta}(\Theta_r, \Theta_s) & I_{\Theta, \Theta}(\Theta_r, \Theta_s) \\
I_{\Theta, \Theta}(\Theta_r, \Theta_s) & I_{\Theta, \Theta}(\Theta_r, \Theta_s)
\end{pmatrix}
\quad (5)
\]

and the partitions are defined as

\[
I_{\Theta, \Theta}(\Theta_r, \Theta_s) = E \left[ \left( \frac{\partial \ln f}{\partial \Theta_r} \right) \left( \frac{\partial \ln f}{\partial \Theta_s} \right)^T \right] \quad r \times r
\]
\[
I_{\Theta, \Theta}(\Theta_r, \Theta_s) = E \left[ \left( \frac{\partial \ln f}{\partial \Theta_r} \right) \left( \frac{\partial \ln f}{\partial \Theta_s} \right)^T \right] \quad r \times s
\]
\[
I_{\Theta, \Theta}(\Theta_r, \Theta_s) = I_{\Theta, \Theta}(\Theta_r, \Theta_s) \quad s \times r
\]
\[
I_{\Theta, \Theta}(\Theta_r, \Theta_s) = E \left[ \left( \frac{\partial \ln f}{\partial \Theta_s} \right) \left( \frac{\partial \ln f}{\partial \Theta_s} \right)^T \right] \quad s \times s \quad (6)
\]
All the partitions of the Fisher information matrix are evaluated at \( \Theta_r = 0 \) and the true values of \( \Theta_s \) for use in (4).

The motivation for using a GLRT is that for large data records it exhibits certain optimality properties. A uniformly most powerful (UMP) test does not exist in many situations. However, of all the tests which are invariant to a natural set of transformations the GLRT exhibits the largest probability of detection. The GLRT is said to be the asymptotically uniformly most powerful invariant (UMPI) test [Lehmann 1959]. It is also a consistent test in the sense that the probability of deciding \( \kappa_0 \) when \( \kappa_1 \) is actually true approaches 0 for large data records. Asymptotically the GLRT is unbiased, i.e., the probability of detection when \( \kappa_1 \) is true is larger than the probability of false alarm. (This result follows from (3) and properties of the chi-square distribution.) Finally, although the GLRT does not usually exhibit a constant false alarm rate (CFAR) it does so for large data records. It is difficult to find the conditions under which the asymptotic results apply to finite length data records. The following heuristic conditions follow from [Cox and Hinkley 1974].

1) The asymptotic statistics of the MLE’s used in the likelihood ratio should be applicable, i.e., they should be Gaussian with mean equal to the true parameter value and covariance matrix equal to the inverse of the Fisher information matrix.

2) The two hypotheses should be reasonably close and only slight departures of \( \Theta_r \) from zero should be tested.

III. Formulation of the Problem and GLRT Solution

The detection problem considered here is the following.

\[
\kappa_0 : y = \mathbf{W}u \\
\kappa_1 : y = \mathbf{W}u + \mu \mathbf{s}
\]

where \( \mathbf{s} = [s_1 \ s_2 \ \cdots \ s_N]^T \) is a vector of known signal amplitudes, \( \mathbf{u} = [u_1 \ u_2 \ \cdots \ u_N]^T \)
is a vector of independent and identically distributed (i.i.d.) noise with a symmetric PDF, $\mu$ is an unknown scalar (either positive or negative) and $W$ is an invertible $(N \times N)$ matrix whose elements are functions of a set of unknown parameters $\Psi = [\psi_1 \psi_2 \cdots \psi_M]$.

$$[W]_{ij} = w_{ij}(\Psi)$$

Since $u_n$, $n = 1, 2, \cdots N$ are i.i.d., the PDF of $u$ can be expressed as

$$f(u; \Phi) = \prod_{n=1}^{N} f(u_n; \Phi)$$

(8)

where $f(u_n; \Phi)$ is the marginal PDF of each $u_n$ dependent on the unknown parameter vector $\Phi$. $f$ is assumed to be symmetric, i.e., $f(-u) = f(u)$. Note that the covariance matrix of the noise is $\sigma^2 W W^T$ where $\sigma^2$ is the variance of $u_n$.

(7) represents a general set of problems. The unknown matrix $W$ allows for a large class of spectral characteristics or correlation patterns of the background noise. For large data records autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) processes can be represented by the above formulation of the underlying random process if $W$ is the impulse response matrix of the corresponding filter. Secondly, the PDF of $u_n$ can be chosen to characterize specific problems in a realistic way. The parameter vector $\Phi$ is left unknown in order to add flexibility to the noise PDF model. Thirdly, by allowing $\mu$ to be positive or negative the detector will be able to accommodate a change of polarity in the signal.

The problem of (7) can be recast as

$$\chi_0 : \Theta^T = [\Theta^T \Theta_r^T]$$

(9a)

$$\chi_1 : \Theta^T = [\Theta_r^T \Theta_s^T] \quad \Theta_r \neq 0$$

(9b)

where

$$\Theta_r = \mu \quad \text{a scalar}$$

$$\Theta_s = [\Psi^T \Phi^T]^T \quad \text{(vector of nuisance parameters)}$$

(10)
Since (9) is equivalent to (1), the GLRT for testing $\lambda_1$ vs. $\lambda_0$ is given by (2). In order to evaluate the MLE's it is necessary to find the joint PDF of $y$ under either hypothesis which can be found from the joint PDF of $u$ in the following way. From (7) it follows that

$$u = W^{-1}y \quad \text{under } \lambda_0 \quad (11a)$$

$$u = W^{-1}(y - \mu_s) \quad \text{under } \lambda_1 \quad (11b)$$

$W^{-1}$ exists because $W$ is assumed to be invertible. The elements of $W^{-1}$ are also known functions of $\Psi$.

$$[W^{-1}]_{ij} = \omega_{ij}(\Psi) \quad (12)$$

(11) being an affine transformation, the joint PDF of $y$ can be written as

$$f(y; \Psi, \Phi) = \frac{1}{|\det(W)|} f(u; \Phi) \bigg|_{u = W^{-1}y} \quad \text{under } \lambda_0$$

$$f(y; \Psi, \Phi) = \frac{1}{|\det(W)|} f(u; \Phi) \bigg|_{u = W^{-1}(y - \mu_s)} \quad \text{under } \lambda_1$$

which in view of (8) and (11) reduces to

$$f(y; \Psi, \Phi) = \frac{1}{|\det(W)|} \prod_{n=1}^{N} \left( f(u_n; \Phi) \bigg|_{u_n = \sum_{j=1}^{N} \omega_{nj}(\Psi)y_j} \right) \quad \text{under } \lambda_0 \quad (13a)$$

$$f(y; \Psi, \Phi) = \frac{1}{|\det(W)|} \prod_{n=1}^{N} \left( f(u_n; \Phi) \bigg|_{u_n = \sum_{j=1}^{N} \omega_{nj}(\Psi)(y_j - \mu_{sj})} \right) \quad \text{under } \lambda_1 \quad (13b)$$

Therefore the GLRT for this problem is to decide $\lambda_1$ if

$$\mathcal{L}_G = \frac{\prod_{n=1}^{N} f\left( \sum_{j=1}^{N} \omega_{nj}(\hat{\Psi})(y_j - \hat{\mu}_{sj}); \hat{\Phi} \right)}{\prod_{n=1}^{N} f\left( \sum_{j=1}^{N} \omega_{nj}(\hat{\Psi})y_j; \hat{\Phi} \right)} > \gamma \quad (14)$$

where hat's denote MLE under $\lambda_0$ and double hat's denote MLE under $\lambda_1$. It is assumed that the values of $|\det(W)|$ under $H_0$ and $\lambda_1$ are nearly the same. This assumption
simplifies the problem considerably. The threshold $\gamma$ is adjusted to achieve a given probability of false alarm, as will be discussed in the next section.

Note that if $\Psi$ is known so that $W^{-1}y$ can be computed, then (7) reduces to

$$
\mathcal{H}_0 : W^{-1}y = u \\
\mathcal{H}_1 : W^{-1}y = u + \mu W^{-1}s
$$

which is simply the problem of detecting the transformed signal $W^{-1}s$ of unknown amplitude $\mu$ in i.i.d. noise from the transformed observation vector $W^{-1}y$. The likelihood ratio corresponding to the GLRT for this problem is

$$
\mathcal{L}_G = \frac{\prod_{n=1}^{N} f \left( \sum_{j=1}^{N} \omega_{nj}(\Psi)(y_j - \hat{\mu}s_j); \hat{\phi} \right)}{\prod_{n=1}^{N} f \left( \sum_{j=1}^{N} \omega_{nj}(\Psi)y_j; \hat{\phi} \right)}
$$

The same statistic is used for the case of unknown $\Psi$ by replacing it with its MLE under the respective hypotheses for numerator and denominator as per (14).

Another special case of (7) arises when the noise is white, i.e., $W = I$, where $I$ is the identity matrix. (14) then reduces to [Kay 1985]

$$
\mathcal{L}_G = \frac{\prod_{n=1}^{N} f((y_n - \hat{\mu}s_n); \hat{\phi})}{\prod_{n=1}^{N} f(y_n; \hat{\phi})} > \gamma
$$

It was indicated earlier in this section that the linear model (7) is capable of representing the case of AR noise for large data records. The advantage of AR modeling of the noise as opposed to an ARMA or MA model is that it is easier to estimate the unknown parameters as required by the GLRT. This case is now examined in detail. The detection problem for AR noise is

$$
\mathcal{H}_0 : y = x \\
\mathcal{H}_1 : y = x + \mu s
$$

(15)
with

\[ x = [x_1\ x_2\ \cdots\ x_N]^T \]

It is assumed that the sequence \( \{x_1, x_2, \cdots, x_N\} \) is the output of a \( p \)th order all-pole filter excited by white driving noise or

\[ x_n = -\sum_{j=1}^{p} a_j x_{n-j} + u_n, \quad n = 1, 2, \cdots N \]

alternately,

\[ u_n = x_n + \sum_{j=1}^{p} a_j x_{n-j} = \sum_{j=0}^{p} a_j x_{n-j}, \quad n = 1, 2, \cdots N \]

assuming \( a_0 = 1 \). \( u_n \) can also be written as a function of \( y \) under either hypothesis

\[ u_n = \sum_{j=0}^{p} a_j y_{n-j}, \quad n = 1, 2, \cdots N \quad \text{under } \mathcal{H}_0 \quad (16a) \]

\[ = \sum_{j=0}^{p} a_j (y_{n-j} - \mu s_{n-j}), \quad n = 1, 2, \cdots N \quad \text{under } \mathcal{H}_1 \quad (16b) \]

Note that \( u_1, u_2, \cdots u_p \) involves samples prior to \( y_1 \) which are outside the observation interval. These are assumed to be zero for simplicity. For large data records this assumption will not change the character of the GLRT. In the matrix form

\[
\begin{pmatrix}
  u_1 \\
  \vdots \\
  u_p \\
  u_{p+1} \\
  u_N
\end{pmatrix} =
\begin{pmatrix}
  1 \\
  \vdots \\
  a_p \\
  0 \\
  \vdots
\end{pmatrix}
\begin{pmatrix}
  a_1 \\
  \vdots \\
  \vdots \\
  a_p \\
  0
\end{pmatrix}
\begin{pmatrix}
  y_1 - \mu s_1 \\
  \vdots \\
  y_p - \mu s_p \\
  y_{p+1} - \mu s_{p+1} \\
  y_N - \mu s_N
\end{pmatrix}
\]

under \( \mathcal{H}_1 \). The equation is the same under \( \mathcal{H}_0 \), except that \( \mu = 0 \). (17) is a special case of (11) with \( \Psi = a = [a_1\ a_2\ \cdots\ a_p] \) and \( M = p \). \( W^{-1} \) is a lower triangular Toeplitz matrix given by

\[ \omega_{ij}(a) = \begin{cases} 
0, & \text{if } i < j, \\
 a_{i-j}, & \text{if } j \leq i \leq j + p, \\
0, & \text{if } j + p < i.
\end{cases} \]
To avoid having to assume that \( \{y_{-(p-1)}, y_{-(p-2)}, \ldots, y_0\} \) are zero one can proceed as follows. Considering only the last \((N-p)\) equations of (16) which expressed in the matrix form are

\[
\begin{pmatrix}
  u_{p+1} \\
  \vdots \\
  u_{2p} \\
  u_{2p+1} \\
  \vdots \\
  u_N
\end{pmatrix} =
\begin{pmatrix}
  1 & \cdots & \cdots & \cdots & \cdots \\
  \vdots & \ddots & \cdots & \cdots & \cdots \\
  0 & a_p & \cdots & 1 \\
  \vdots & \ddots & \ddots & \cdots & \cdots \\
  0 & \cdots & 0 & a_p & 1
\end{pmatrix}
\begin{pmatrix}
  y_{p+1} - \mu s_{p+1} \\
  \vdots \\
  y_{2p} - \mu s_{2p} \\
  y_{2p+1} - \mu s_{2p+1} \\
  \vdots \\
  y_N - \mu s_N
\end{pmatrix}
\]

under \( \mathcal{H}_1 \). Substitution of \( \mu = 0 \) in (18) gives the corresponding equation for \( \mathcal{H}_0 \). (18) is also a special case of (11b) except that only the last \((N-p)\) of the \( N \) scalar equations implied by (11b) are used. Since the added vector causes a departure from the general model, the previous results can not be used. To determine the GLRT first consider the conditional likelihood function. In this case the conditional likelihood of \( y_{p+1}, y_{p+2}, \ldots, y_N \) given \( y_1, y_2, \ldots, y_p \) is

\[
f(y_{p+1}, y_{p+2}, \ldots, y_N|y_1, y_2, \ldots, y_p)
\]

\[
= \prod_{n=p+1}^{N} f\left( \sum_{j=0}^{p} a_j y_{n-j}; \Phi \right) \quad \text{under } \mathcal{H}_0 \quad (19a)
\]

\[
= \prod_{n=p+1}^{N} f\left( \sum_{j=0}^{p} a_j (y_{n-j} - \mu s_{n-j}); \Phi \right) \quad \text{under } \mathcal{H}_1 \quad (19b)
\]
The likelihood ratio is given by
\[
\ell_G = \frac{f(y_{p+1}, y_{p+2}, \ldots, y_N | y_1, y_2, \ldots, y_p; \hat{\Theta}_r, \hat{\Theta}_s) f(y_1, y_2, \ldots, y_p; \hat{\Theta}_r, \hat{\Theta}_s)}{f(y_{p+1}, y_{p+2}, \ldots, y_N | y_1, y_2, \ldots, y_p; 0, \hat{\Theta}_s) f(y_1, y_2, \ldots, y_p; 0, \hat{\Theta}_s)}
\]
\[
= \left[ \prod_{n=p+1}^{N} \frac{f \left( \sum_{j=0}^{p} \hat{\alpha}_j (y_{n-j} - \hat{\mu} s_{n-j}); \hat{\Phi} \right)}{f(y_1, y_2, \ldots, y_p; \hat{\mu}, \hat{\alpha}, \hat{\Phi})} \right] \left[ \prod_{n=p+1}^{N} \frac{f \left( \sum_{j=0}^{p} \hat{\alpha}_j y_{n-j}; \hat{\Phi} \right)}{f(y_1, y_2, \ldots, y_p; 0, \hat{\alpha}, \hat{\Phi})} \right]
\]
where \( \hat{\alpha}_0 \) and \( \hat{\alpha}_0 \) are defined to be unity. The second term is dropped for ease of computation. A heuristic justification for ignoring the second term is that when \( N \) is large, its contribution to \( \ell_G \) will be negligible. The closer the poles of the AR model to the unit circle, larger is the requirement for \( N \) [Box and Jenkins 1970], [Kay 1981].

With this simplification, the GLRT decides \( \mathcal{H}_1 \) if
\[
\ell_G = \left[ \prod_{n=p+1}^{N} \frac{f \left( \sum_{j=0}^{p} \hat{\alpha}_j (y_{n-j} - \hat{\mu} s_{n-j}); \hat{\Phi} \right)}{f(y_1, y_2, \ldots, y_p; \hat{\mu}, \hat{\alpha}, \hat{\Phi})} \right] > \gamma
\]  

(20)

A comparison of (14) and (20) shows that the latter uses fewer terms in the product. However both formulations are clearly asymptotically equivalent. Figure 1 is a block diagram to generate \( 2 \ln \ell_G \) from the data. The reason for computing \( 2 \ln \ell_G \) instead of \( \ell_G \) will be clear from the discussion in the next section. The block diagram is very much similar to that obtained by [Kay 1983] for the detection of a completely known signal in unknown colored Gaussian noise. In the Gaussian case \( \ln f \) is a simple quadratic, while for the general non-Gaussian case it will be highly non-linear. Figure 1 also uses an estimator for \( \mu \) which was assumed to be known in [Kay 1983].

IV. Asymptotic Performance of the GLRT Detector

Asymptotic distributions of \( 2 \ln \ell_G \) under \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \) are given by (3a) and (3b) respectively. In this case \( \Theta_r = \mu, \Theta_s = [\Psi^T \Phi^T]^T \) for the general linear model and
\[ \Theta_s = [a^T \Phi^T]^T \] for the AR case. Hence the noncentrality parameter is

\[ \lambda = \mu^2 [I_{\mu\mu}(0, \Theta_s) - I_{\mu\Theta_s}(0, \Theta_s)I_{\Theta_s\Theta_s}^{-1}(0, \Theta_s)I_{\mu\Theta_s}(0, \Theta_s)] \quad (21) \]

The probability of false alarm is

\[ P_{FA} = P\{2 \ln \ell_G > \gamma'|\lambda_0\} \quad (22) \]

where \( \gamma' = 2 \ln \gamma \). The probability of detection is

\[ P_D = P\{2 \ln \ell_G > \gamma'|\lambda_1\} \quad (23) \]

Both the probabilities can be calculated from the tables of noncentral and central chi-squared distributions, respectively. In practice, \( \gamma' \) can be set to produce a given false alarm rate and \( P_D \) can be calculated from (23) accordingly.

As indicated before, there is no UMP test for the detection problem considered in this paper. Therefore there is no upper bound to which the performance of any detector may be compared. However the performance of the GLRT is better appreciated when compared to that of a clairvoyant GLRT. A clairvoyant GLRT is one which uses perfect knowledge of the nuisance parameters \( \Theta_s \). The likelihood ratio in this case is

\[ \ell_{GC} = \frac{f(y; \hat{\Theta}_r, \Theta_s)}{f(y; 0, \Theta_s)} \]

which in view of (13) is

\[ \ell_{GC} = \prod_{n=1}^N f \left( \sum_{j=1}^N \omega_{nj}(\Psi)(y_j - \hat{\mu}_s); \Phi \right) \]

\[ \prod_{n=1}^N f \left( \sum_{j=1}^N \omega_{nj}(\Psi)y_j; \Phi \right) \]

where \( \Psi \) and \( \Phi \) are assumed to be known. Asymptotically, \( 2 \ln \ell_{GC} \) is distributed as

\[ 2 \ln \ell_{GC} \sim \chi^2 \quad \text{under } \lambda_0 \quad (24a) \]

\[ 2 \ln \ell_{GC} \sim \chi^2(r, \lambda_c) \quad \text{under } \lambda_1 \quad (24b) \]
where
\[ \lambda_c = \Theta_r^T I_{\Theta, \theta_r}(0, \Theta_s) \Theta_r \]

For the problem considered here, \( r = 1 \) and \( \Theta_r = \mu \). Hence
\[ \lambda_c = \mu^2 I_{\mu\mu}(0, \Theta_s) \quad (25) \]

Comparing \( \lambda \) and \( \lambda_c \) as given by (21) and (25), respectively, it is apparent that \( \lambda \) is equal to \( \lambda_c \) less an additional term. Assuming that \( I_{\Theta, \theta_r}^{-1} \) is positive semidefinite,
\[ \lambda_c \geq \lambda \]

From the theory of noncentral chi-square distribution it can be shown that \( P_D \) as given by (23) is a monotonic function of the noncentrality parameter, which implies that \( P_D \) for the clairvoyant GLRT is greater than or equal to that for the GLRT [Sengupta 1986]. Therefore the clairvoyant GLRT detector, although impractical in this case, provides an upper bound on the performance of the GLRT detector. In order that the upper bound be achieved, \( \lambda \) should be equal to \( \lambda_c \). This will occur if
\[ I_{\mu\theta_r}(0, \Theta_s) = 0 \quad (26) \]

Appendices A and B show that this is indeed the case for the general linear model of (7) and the AR noise model of (15), respectively, if as assumed \( f \) is a symmetric PDF. Therefore the asymptotic performance of the GLRT is equivalent to the performance of the clairvoyant GLRT for detection in presence on non-Gaussian noise modeled as in (7) or (15). This implies that one can do as well in detecting a signal of unknown amplitude as if the unknown noise parameters were known.

V. General Conclusions about the Performance of the GLRT

A key to the asymptotic performance of the GLRT detector is the noncentrality parameter \( \lambda \), which is found to be equivalent to \( \lambda_c \). It would be interesting to examine
how \( \lambda \) depends on the statistical properties of the noise. First consider

\[
I_{\mu \nu}(\mu, \Theta_s) = E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right)^2 \right] = E \left[ \frac{\partial \ln \left( \prod_{n=1}^{N} f \left( \sum_{j=1}^{N} \omega_{nj} (y_j - \mu s_j); \Phi \right) \right)}{\partial \mu} \right]^2
\]

\[
= E \left[ \frac{\partial \ln \left( \prod_{n=1}^{N} f(u_n; \Phi) \right)}{\partial \mu} \right]^2
\]

using (11b) and (12)

\[
= \sum_{n=1}^{N} E \left[ \left( \frac{\partial}{\partial \mu} \ln f(u_n; \Phi) \right)^2 \right] \quad (27)
\]

The last step follows from the facts that \( u_n \)'s are i.i.d. and that the cross-terms are zero, since

\[
E \left[ \frac{\partial}{\partial \mu} \ln f(u_n; \Phi) \right] = \frac{\partial}{\partial \mu} E[\ln f(u_n; \Phi)] = 0
\]

under certain regularity assumptions on \( f \) [Bickel and Doksum 1977]. Writing (11b) explicitly as

\[
u_n = \sum_{j=1}^{N} \omega_{nj} (y_j - \mu s_j) \quad (28)
\]

it follows that

\[
\frac{\partial u_n}{\partial \mu} = -\sum_{j=1}^{N} \omega_{nj} s_j \quad (29)
\]

Hence (27) can be rewritten as

\[
I_{\mu \nu}(\mu, \Theta_s) = \sum_{n=1}^{N} E \left[ \left( \frac{\partial}{\partial u_n} \ln f(u_n; \Phi) \right) \left( \frac{\partial u_n}{\partial \mu} \right)^2 \right]
\]

\[
= \sum_{n=1}^{N} E \left[ \left( -\sum_{j=1}^{N} \omega_{nj} s_j \right)^2 \left( \frac{\partial \ln f}{\partial u_n} \right)^2 \right]
\]
\[
\begin{align*}
&= \sum_{n=1}^{N} \left( -\sum_{j=1}^{N} \omega_{nj} s_j \right)^2 E \left[ \left( \frac{\partial \ln f}{\partial u_n} \right)^2 \right] \\
&= \sum_{n=1}^{N} \left( -\sum_{j=1}^{N} \omega_{nj} s_j \right)^2 I_f(\Phi) \\
&= \left[ \frac{1}{\sigma^2} \sum_{n=1}^{N} \left( -\sum_{j=1}^{N} \omega_{nj} s_j \right)^2 \right] \sigma^2 I_f(\Phi) \\
\end{align*}
\]

where \( \sigma^2 \) is the variance of \( u_n \) and

\[
I_f(\Phi) = E \left[ \left( \frac{\partial \ln f}{\partial u_n} \right)^2 \right]
\]

does not depend on \( \mu \) or \( u_n \). The expectation is with respect to \( u_n \) only since \( u_n \)'s are i.i.d. It follows from (25) that

\[
\lambda = \lambda_c = \left[ \frac{\mu^2}{\sigma^2} \sum_{n=1}^{N} \left( -\sum_{j=1}^{N} \omega_{nj} s_j \right)^2 \right] \sigma^2 I_f(\Phi)
\]

\[
= \frac{\mu^2}{\sigma^2} (W^{-1}s)^T (W^{-1}s) \sigma^2 I_f(\Phi)
\]

\[
= \frac{\mu^2}{\sigma^2} s^T (W^{-1}W^{-1}) s \sigma^2 I_f(\Phi)
\]

\[
= (\mu s)^T (\sigma^2 WW^T)^{-1} (\mu s) \sigma^2 I_f(\Phi)
\]

\[
= s_0^T R^{-1}s_0 \sigma^2 I_f(\Phi)
\]

where \( R = \sigma^2 WW^T \) is the \( N \times N \) autocorrelation matrix of the colored noise and \( s_0 = \mu s \) is the signal vector including the amplitude. \( s_0^T R^{-1}s_0 \) is the signal to noise ratio (SNR) at the output of a prewhitener followed by a matched filter (or correlator), both built with perfect knowledge of the filter parameters (\( \Psi \)). To be more precise, if the data is passed through an ideal whitener (a filter which will completely whiten the noise) and correlated (multiplied term-by-term and summed) with the output of a similar filter through which only the known signal is passed then \( s_0^T R^{-1}s_0 \) is the
squared ratio of the contributions from the signal and noise parts of the data. In the case of AR noise, a similar derivation using (16b) (instead of (28)) gives

\[ I_{\mu\mu}(\mu, \Theta_s) = \left[ \frac{1}{\sigma^2} \sum_{n=p+1}^{N} \left( -\sum_{j=0}^{p} a_j s_{n-j} \right) \right]^2 \sigma^2 I_f(\Phi) \]  

(32)

and

\[ \lambda = \left[ \frac{\mu^2}{\sigma^2} \sum_{n=p+1}^{N} \left( -\sum_{j=0}^{p} a_j s_{n-j} \right) \right]^2 \sigma^2 I_f(\Phi) \]

(33)

\[ \lambda = \frac{\mu^2}{\sigma^2} (A^H A)^{-1} \sigma^2 I_f(\Phi) \]

where \( A \) is the \((N-p) \times N\) Toeplitz matrix

\[ A = \begin{pmatrix}
    a_p & \ldots & a_1 & 1 & \ldots & 0 \\
    0 & a_p & \ldots & a_1 & 1 & \ldots \\
    \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
    0 & \ldots & 0 & a_p & \ldots & a_1 & 1
\end{pmatrix} \]

Therefore

\[ \lambda = s_0^T R^{-1} s_0 [\sigma^2 I_f(\Phi)] \]  

(34)

\( s_0 = \mu s \) as before and \( R = \sigma^2 (A^T A)^{-1} \) is approximately the covariance matrix of the noise. Clearly, \( \lambda \) is proportional to the SNR at the output of a prewhitener-correlator (using true value of \( a \)) in the AR case also.

Having established similar results in the cases of AR noise and the general linear model, an attempt is now made to examine them. The AR noise model is chosen for this purpose because of its intuitive frequency-domain interpretation. Figure 2 is a block diagram representing (33). It shows that \( \lambda \) can be obtained by inverse filtering the signal and summing the squares of the output of the filter. If the signal has most of its power at the frequency where the inverse filter has a zero, the output power and hence \( \lambda \) will be small leading to a small probability of detection. In other words, it
is difficult to detect the signal if the peaks of the signal spectrum coincides with the peaks of the noise PSD. This makes perfect intuitive sense. On the other hand it is possible to maximize $\lambda$ by choosing a suitable signal $s$ for a given noise background. This can be done by constraining the signal energy to be constant and maximizing the SNR at the output of a prewhitener-correlator over all possible signal shapes. Writing the Toeplitz matrix $R$ in terms of its orthonormal eigenvectors $\{v_1, v_2, \ldots, v_N\}$ and eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_N\}$

$$R = \sum_{j=1}^{N} \lambda_j v_j v_j^T$$

it follows that

$$R^{-1} = \sum_{j=1}^{N} \frac{1}{\lambda_j} v_j v_j^T$$

Hence

$$s_0^T R^{-1} s_0 = \sum_{j=1}^{N} \frac{1}{\lambda_j} (s_0^T v_j)^2$$

$$= \sum_{j=1}^{N} \frac{1}{\lambda_j} \xi_j^2$$

where $\xi_j = s_0^T v_j$ is the component of the signal $s_0$ along the eigenvector $v_j$. The condition of constant signal energy can be written as

$$s_0^T s_0 = P_s$$

Since the eigenvectors are orthonormal

$$\sum_{j=1}^{N} \xi_j^2 = \sum_{j=1}^{N} s_0^T v_j v_j^T s_0 = s_0^T \left[ \sum_{j=1}^{N} v_j v_j^T \right] s_0 = s_0^T s_0 = P_s$$

The SNR given by the weighted sum (36) has to be maximized subject to constraint that the unweighted sum of the squares is fixed at $P_s$ as in (38). In general, $R$ will be positive definite and all the eigenvalues will be positive. If there exists a minimum
eigenvalue \( \lambda_k \), then the SNR is maximized by choosing \( \xi_k = \sqrt{P_s} \) and \( \xi_j = 0 \) for \( j \neq k \), i.e., by choosing \( s_0 \) to be proportional to \( v_k \). Since the probability of detection given by (23) is a monotonic function of \( \lambda \) which is proportional to the SNR at the output of a prewhitener-correlator, the above choice of the signal shape for a given signal energy also maximizes the probability of detection. If one of the eigenvalues \( \lambda_k \) is zero, then it is possible to choose the signal in such a way that there is no component of noise along the signal vector and therefore the SNR is infinite giving rise to singular detection. Therefore the probability of detection is maximized by choosing the signal in the direction of the smallest noise component. This is the discrete time equivalent of a well-known result for the continuous case [Van Trees 1968]. An interesting special case occurs when \( N \to \infty \) such that the eigenvectors become

\[
\nu_j \to \frac{1}{\sqrt{N}} \left( 1 e^{j2\pi f_j} \ldots e^{j2\pi(N-1)f_j} \right)
\]

Hence the optimum signal is a sinusoid in the direction of the eigenvector associated with the minimum eigenvalue. For very large data records the eigenvalue \( \lambda_j \) corresponding to the eigenvector \( v_j \) approaches the value of PSD at the frequency \( f_j \). Hence the signal easiest to detect would be a sinusoid at the frequency at which the noise PSD has a minimum. This is also apparent from the frequency domain equivalent of (36) (using Perseval's theorem)

\[
s_0^T R^{-1} s_0 = \int_{-0.5}^{0.5} \frac{|S_0(f)|^2}{P_{uu}(f)} df
\]

where \( S_0(f) \) is the signal spectrum (Fourier transform of \( s_0 \)) and \( P_{uu} \) is the noise PSD. With the constraint that

\[
\int_{-0.5}^{0.5} |S_0(f)|^2 df = 1
\]

which is equivalent to (37), the integral is maximized if the numerator is large only where the denominator is small or zero. This result has a nice intuitive justification.
However if the filter parameters are completely unknown the above result can not be used to select a suitable signal.

The next issue of interest is the effect of the noise PDF on the detection performance. A reasonable basis of comparison should be formed for this purpose. Therefore the Gaussian and non-Gaussian noise processes are assumed to have the same PSD, i.e., the same spectral shape and power and detection of the same signal is considered. Consequently, the comparison is done on the basis of a fixed signal to noise ratio (or $s_0^T R^{-1} s_0$). Under these assumptions, the probability of detection is larger for that noise PDF which has a larger value of $\sigma^2 I_f$. In other words, given two noise backgrounds with the same PSD but different underlying noise PDF's, in order to achieve the same probability of detection, more SNR is required for that background for which $\sigma^2 I_f$ is smaller. It is known that among all symmetric and integrable PDF's, the Gaussian PDF is the only one for which $\sigma^2 I_f$ attains its minimum value of unity [Sengupta and Kay 1986]. Therefore for a given noise PSD, it is easier to detect a signal known except for amplitude in non-Gaussian noise than in Gaussian noise. From (34) it follows that in order to have the same noncentrality parameters in the non-Gaussian and Gaussian cases

$$SNR_{NG}(\sigma^2 I_f) = SNR_G$$

where $SNR_{NG}$ and $SNR_G$ are the SNR's required in non-Gaussian and Gaussian noises, respectively, in order to achieve a given probability of detection (i.e., a given $\lambda$). The above equation can also be written as

$$10 \log_{10} \frac{SNR_G}{SNR_{NG}} = 10 \log_{10}(\sigma^2 I_f)$$

(39)

Therefore $10 \log_{10}(\sigma^2 I_f)$ is a measure of the SNR bonus in dB for a non-Gaussian distribution. The result also holds for the special case of white noise when (33) becomes

$$\lambda = \left[ \frac{\mu^2}{\sigma^2} \sum_{n=1}^{N} s_n^2 \right] \sigma^2 I_f(\Phi)$$

(40)
The quantity $\sigma^2 I_f$ is now shown to be independent of scaling. If the random variable $u$ has a PDF $f(u)$ then the normalized random variable $\hat{u} = u/\sigma$ has the PDF

$$g(\hat{u}) = \sigma f(u)$$

hence

$$\sigma^2 I_f = \sigma^2 \int_{-\infty}^{\infty} \frac{|f'(u)|^2}{f(u)} du$$

$$= \sigma^2 \int_{-\infty}^{\infty} \left[ g\left(\frac{u}{\sigma}\right) \frac{1}{\sigma} \right]^2 du$$

$$= \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sigma^4} \left( \frac{g'(\hat{u})}{g(\hat{u})} \right)^2 \sigma d\hat{u}$$

$$= \int_{-\infty}^{\infty} \left( \frac{g'(\hat{u})}{g(\hat{u})} \right)^2 d\hat{u}$$

Hence $\sigma^2 I_f$ depends only on the shape of the PDF and is unaffected by scaling. Therefore the SNR bonus quantified by (39) is the same for any value of the noise power as long as the powers of the non-Gaussian and Gaussian processes are the same.

It is interesting to note that the same quantity represents the amount of departure from Gaussianity for the problem of estimating AR filter parameters of a non-Gaussian AR process. The CR bounds for these parameters are found to be less in the case of a non-Gaussian PDF than the corresponding bounds in the Gaussian case by a factor of $\sigma^2 I_f$ [Sengupta and Kay 1986].

As an illustration of the improvement made by the proposed detector over the Gaussian detector, consider the mixed-Gaussian noise PDF

$$f(u) = \frac{1 - \epsilon}{\sqrt{2\pi\sigma_B^2}} e^{-\frac{u^2}{2\sigma_B^2}} + \frac{\epsilon}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{u^2}{2\sigma_I^2}}$$

The first term on the right hand side is referred to as the background component with variance $\sigma_B^2$ and the second term is called the interference component with variance $\sigma_I^2$. $\epsilon$ is called the mixture parameter and is regarded as a measure of the degree of
contamination of the background Gaussian process by the interference process. The model is useful in representing a nominally Gaussian noise background characterized by the presence of sharp spikes or impulses [Sengupta and Kay 1986]. Assuming \( \sigma_B^2 = 1 \) and \( \sigma_I^2 = 1000 \), Figure 3 plots the SNR bonus given by (39) vs. \( \epsilon \) (in this case \( \Phi = \epsilon \)). It shows how much improvement can be expected over the Gaussian case in terms of SNR while detecting a signal known except for amplitude in colored noise. The comparison is made, as indicated before on the basis of the same PSD in the Gaussian and mixed-Gaussian cases. It should be mentioned, however, that introduction of impulses in an otherwise Gaussian environment does not improve the probability of detection, which is expected intuitively. This is because of the fact that introduction of impulses also increases the noise power by a considerable amount. This increase in noise power is alleviated by employing a non-Gaussian detector. As an example, for \( \epsilon = 0.1 \), the overall noise variance is approximately \( 100\sigma_B^2 \) (as compared to \( \sigma_B^2 \) before the introduction of impulses), i.e., the noise power increases by 20 dB. It can be observed from Figure 3 that the SNR bonus is also approximately 20 dB for \( \epsilon = 0.1 \). Therefore the mixed-Gaussian detector does not suffer from a loss of performance unlike the Gaussian detector whose threshold of detection is expected to go down considerably with the introduction of impulses.

VI. Summary

The GLRT for the detection of a signal known except for amplitude in unknown colored non-Gaussian noise was derived in section III through parametric modeling of the noise PDF and covariance matrix. The popular time series models such as AR, MA and ARMA for the noise are asymptotically special cases of the proposed linear model for large data records. The GLRT was found to achieve the performance of a clairvoyant GLRT asymptotically, i.e., knowledge of the nuisance parameters is not required to attain an upper bound in performance. The effects of the signal spectrum
and the noise PSD on the detection performance was discussed. It was observed that it is difficult to detect a signal whose spectrum matches the noise PSD. If, however, most of the signal is along a direction of low noise component, it is very easy to detect. The asymptotic performance of the GLRT for Gaussian and non-Gaussian noise models were compared. It was concluded that detection in non-Gaussian noise is easier than detection in Gaussian noise for the same noise PSD. The improvement in performance of the GLRT in a non-Gaussian noise background over the Gaussian case is easily quantified in terms of the SNR 'bonus' as a function of the noise PDF parameters.

In order to implement the GLRT described in section III one needs to find the MLE’s of the unknown parameters under each hypothesis. Some work along this line has been done for the case of AR noise [Sengupta and Kay 1986, 2], using reasonable approximations to reduce computation. This work is appropriate for estimation under the null hypothesis ($\mathcal{H}_1$) and extension to the case of alternative hypothesis ($\mathcal{H}_1$) is not straightforward except for the special case of a d.c. signal ($s_j = 1, \quad j = 1, 2, \ldots, N$). Evaluating the joint MLE of the mean or the location parameter ($\mu$) and the AR filter parameters may be particularly difficult for most non-Gaussian processes. Computationally efficient approximations to the GLRT, such as the Rao test and the Wald test [Rao 1973] can be used for this purpose. Estimation of the mean and the other parameters under $\mathcal{H}_1$ can thus be avoided for small signal amplitudes [Sengupta 1986]. This problem will be addressed in a future paper.

References


APPENDIX A

Asymptotic Optimality of the GLRT
for a General Linear Model of the Noise

Assuming that $f$ is an even distribution and $\Theta_s$ is as given in (10), it will now be shown that (26) holds for the detection problem defined in (7). It suffices to prove that

$$I_{\mu\psi}(\mu, \psi, \Phi) = E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \psi} \right)^T \right] = 0 \quad (A.1)$$

and

$$I_{\mu\Phi}(\mu, \psi, \Phi) = E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \Phi} \right)^T \right] = 0 \quad (A.2)$$
To prove (A.1) it is observed that

$$E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \psi_j} \right) \right] = E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \sum_{i=1}^{N} \sum_{k=1}^{N} \left( \frac{\partial \ln f}{\partial \omega_{ik}} \right) \left( \frac{\partial \omega_{ik}}{\partial \psi_j} \right) \right]$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{N} \left( \frac{\partial \omega_{ik}}{\partial \psi_j} \right) E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \omega_{ik}} \right) \right]$$

(A.3)

$\omega_{ik}$ is written without its argument (see (12)) to make the notation easier.

$$E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \omega_{ik}} \right) \right]$$

$$= E \left[ \frac{\partial \ln \left( \prod_{n=1}^{N} f \left( \sum_{j=1}^{N} \omega_{nj} (y_j - \mu s_j); \Phi \right) \right)}{\partial \mu} \frac{\partial \ln \left( \prod_{n=1}^{N} f \left( \sum_{j=1}^{N} \omega_{nj} (y_j - \mu s_j); \Phi \right) \right)}{\partial \omega_{ik}} \right]$$

$$u_n$$ can be used as the argument of $f$ (see (28)) to simplify the equation.

$$E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \omega_{ik}} \right) \right] = E \left[ \frac{\partial \ln \left( \prod_{n=1}^{N} f(u_n; \Phi) \right)}{\partial \mu} \frac{\partial \ln \left( \prod_{n=1}^{N} f(u_n; \Phi) \right)}{\partial \omega_{ik}} \right]$$

$$= \sum_{n=1}^{N} \sum_{m=1}^{N} E \left[ \left( \frac{\partial}{\partial \mu} \ln f(u_n; \Phi) \right) \left( \frac{\partial}{\partial \omega_{ik}} \ln f(u_m; \Phi) \right) \right]$$

$$= \sum_{n=1}^{N} E \left[ \left( \frac{\partial}{\partial \mu} \ln f(u_n; \Phi) \right) \left( \frac{\partial}{\partial \omega_{ik}} \ln f(u_n; \Phi) \right) \right]$$

(A.4)

All the cross-terms are zero because $u_n$'s are i.i.d. and

$$E \left[ \frac{\partial}{\partial \mu} \ln f(u_n; \Phi) \right] = \frac{\partial}{\partial \mu} E[\ln f(u_n; \Phi)] = 0$$

(A.5)

under certain regularity assumptions on $f$ [Bickel and Doksum 1977]. The derivatives w.r.t. $\mu$ and $\omega_{ik}$ can be written in terms of the derivative w.r.t. $u_n$. Note from (28) that

$$\frac{\partial u_n}{\partial \omega_{ik}} = \begin{cases} y_k - \mu s_k & \text{if } n = i \\ 0 & \text{if } n \neq i \end{cases}$$

(A.6)
From (A.6) and (29) it follows that (A.4) can be rewritten as

\[
E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \omega_{ik}} \right) \right] = \sum_{n=1}^{N} E \left[ \left( \frac{\partial \ln f}{\partial u_n} \right) \left( \frac{\partial \ln f}{\partial u_n} \right) \left( \frac{\partial u_n}{\partial \omega_{ik}} \right) \right]
\]

\[
= \left( -\sum_{j=1}^{N} \omega_{ij} s_j \right) E \left[ \left( \frac{\partial}{\partial u_i} \ln f(u_i; \Phi) \right)^2 \left( \sum_{j=1}^{N} \omega_{kj} u_j \right) \right]
\]

\[
= \left( -\sum_{j=1}^{N} \omega_{ij} s_j \right) \sum_{j=1}^{N} \omega_{kj} E \left[ \left( \frac{\partial}{\partial u_i} \ln f(u_i; \Phi) \right)^2 u_j \right]
\]

\[
= \left( -\sum_{j=1}^{N} \omega_{ij} s_j \right) \sum_{j=1}^{N} \omega_{kj} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial u_i} \ln f(u_i; \Phi) \right)^2 u_j \left( \prod_{m=1}^{N} f(u_m; \Phi) \right)
\]

\[
= \left( -\sum_{j=1}^{N} \omega_{ij} s_j \right) \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial u_i} \ln f(u_i; \Phi) \right)^2 f(u_i; \Phi) du_i \sum_{j=1}^{N} \omega_{kj} \int_{-\infty}^{\infty} u_j f(u_j; \Phi) du_j
\]

\[
+ \omega_{ki} \int_{-\infty}^{\infty} \left( \frac{\partial}{\partial u_i} \ln f(u_i; \Phi) \right)^2 u_i f(u_i; \Phi) du_i
\]

\[
= 0
\]

This is true for each \( i \) and \( k \), so that

\[
E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \omega_{ik}} \right) \right] = 0 \quad i, k = 1, 2, \ldots, N
\]
(A.1) follows directly from (A.3) and (A.7).

(A.2) can be proved in a similar way. Consider

\[ E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \phi_i} \right) \right] \]

\[ = E \left[ \frac{\partial \ln \left( \prod_{n=1}^{N} f \left( \sum_{j=1}^{N} \omega_{nj} (y_j - \mu s_j); \Phi \right) \right)}{\partial \mu} \frac{\partial \ln \left( \prod_{n=1}^{N} f \left( \sum_{j=1}^{N} \omega_{nj} (y_j - \mu s_j); \Phi \right) \right)}{\partial \phi_i} \right] \]

\[ = \sum_{n=1}^{N} \sum_{m=1}^{N} E \left[ \left( \frac{\partial \ln f(u_n; \Phi)}{\partial \mu} \right) \left( \frac{\partial \ln f(u_m; \Phi)}{\partial \phi_i} \right) \right] \]

\[ = \sum_{n=1}^{N} E \left[ \left( \frac{\partial \ln f(u_n; \Phi)}{\partial \mu} \right) \left( \frac{\partial \ln f(u_n; \Phi)}{\partial \phi_i} \right) \right] \quad \text{(as } u_n \text{'s are i.i.d.)} \]

Using (A.6) this becomes

\[ E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \phi_i} \right) \right] \]

\[ = \sum_{n=1}^{N} \left( -\sum_{j=1}^{N} \omega_{nj} s_j \right) E \left[ \left( \frac{\partial \ln f(u_n; \Phi)}{\partial u_n} \right) \left( \frac{\partial \ln f(u_n; \Phi)}{\partial \phi_i} \right) \right] \quad \text{(odd)} \]

Under the assumption that \( f \) is even, \( \ln f \) is even, derivative of \( \ln f \) w.r.t. \( u_n \) is odd and the derivative of \( \ln f \) w.r.t. \( \phi_i \) is even [Kay 1985]. Therefore the expectation is taken on an odd function and should be equal to zero as explained while proving (A.7). This being true for each \( \phi_i \) one can conclude that (A.2) holds. (26) is a direct implication of (A.1) and (A.2).
APPENDIX B

Asymptotic Optimality of the GLRT for an AR Model of the Noise

It is now shown that (26) also holds in the case of the GLRT given by (20) for detection in AR noise. Note that the conditional likelihood function (see (19)) is used and the vector of nuisance parameters is

$$ \Theta_* = [a \Phi] $$

with the notations used before. Proving (26) in this case is equivalent to proving that

$$ I_{\mu a}(\mu, a, \Phi) = E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial a} \right)^T \right] = 0 \quad (B.1) $$

and

$$ I_{\mu \Phi}(\mu, a, \Phi) = E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \Phi} \right)^T \right] = 0 \quad (B.2) $$

To prove (B.1) it is observed that

$$ E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial a_i} \right) \right] = E \left[ \frac{\partial \ln \left( \prod_{n=p+1}^N f \left( \sum_{j=0}^p a_j (y_j - \mu s_j); \Phi \right) \right)}{\partial \mu} \right] \frac{\partial \ln \left( \prod_{n=p+1}^N f \left( \sum_{j=0}^p a_j (y_j - \mu s_j); \Phi \right) \right)}{\partial a_i} $$

(16) can be used to simplify the argument of $f$,

$$ E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial a_i} \right) \right] = E \left[ \frac{\partial \ln \left( \prod_{n=p+1}^N f(u_n; \Phi) \right)}{\partial \mu} \frac{\partial \ln \left( \prod_{n=p+1}^N f(u_n; \Phi) \right)}{\partial a_i} \right] $$

$$ = \sum_{n=p+1}^N \sum_{m=p+1}^N E \left[ \left( \frac{\partial \ln f(u_n; \Phi)}{\partial \mu} \right) \left( \frac{\partial \ln f(u_m; \Phi)}{\partial a_i} \right) \right] $$

$$ = \sum_{n=p+1}^N E \left[ \left( \frac{\partial \ln f(u_n; \Phi)}{\partial \mu} \right) \left( \frac{\partial \ln f(u_n; \Phi)}{\partial a_i} \right) \right] \quad (B.3) $$

28
The last step follows from (A.5) and the fact that $u_n$'s are iid. The derivatives w.r.t. $\mu$ and $a_i$ can be written in terms of the derivative w.r.t. $u_n$. From (16b) it follows that

$$\frac{\partial u_n}{\partial \mu} = -\sum_{j=0}^{p} a_j s_{n-j} \quad (B.4)$$

and

$$\frac{\partial u_n}{\partial a_i} = (y_{n-i} - \mu s_{n-j}) \quad (B.5)$$

Using these results (B.3) can be rewritten as

$$E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial a_i} \right) \right]$$

$$= \sum_{n=p+1}^{N} E \left[ \left( \frac{\partial \ln f}{\partial u_n} \right) \left( \frac{\partial u_n}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial u_n} \right) \left( \frac{\partial u_n}{\partial a_i} \right) \right]$$

$$= \sum_{n=p+1}^{N} \left( -\sum_{j=0}^{p} a_j s_{n-j} \right) E \left[ \frac{\partial}{\partial u_n} \ln f(u_n; \Phi) \right]^{2} \frac{y_{n-i} - \mu s_{n-i}}{even}$$

$$= \sum_{n=p+1}^{N} \left( -\sum_{j=0}^{p} a_j s_{n-j} \right) E \left[ \frac{\partial}{\partial u_n} \ln f(u_n; \Phi) \right]^{2} \frac{y_{n-i} - \mu s_{n-i}}{odd}$$

$(y_{n-i} - \mu s_{n-i})$ is a linear function of $\{u_{n-i}, u_{n-i-1}, \ldots, u_1\}$ and hence is an odd function of each $u_n$. Therefore the expectation is taken on an odd function which should be equal to zero since the PDF $f$ itself is even. This being true for each $a_i$, it can be concluded that (B.1) holds.

Proof of (B.2) is similar. Consider
\[ E \left[ \left( \frac{\partial \ln f}{\partial \mu} \right) \left( \frac{\partial \ln f}{\partial \phi_i} \right) \right] \]

\[ = E \left[ \frac{\partial \ln \left( \prod_{n=p+1}^{N} f(u_n; \Phi) \right)}{\partial \mu} \frac{\partial \ln \left( \prod_{n=p+1}^{N} f(u_n; \Phi) \right)}{\partial \phi_i} \right] \]

\[ = \sum_{n=p+1}^{N} \sum_{m=p+1}^{N} E \left[ \left( \frac{\partial}{\partial \mu} \ln f(u_n; \Phi) \right) \left( \frac{\partial}{\partial \phi_i} \ln f(u_m; \Phi) \right) \right] \]

\[ = \sum_{n=p+1}^{N} E \left[ \left( \frac{\partial}{\partial \mu} \ln f(u_n; \Phi) \right) \left( \frac{\partial}{\partial \phi_i} \ln f(u_n; \Phi) \right) \right] \quad \text{(as } u_n \text{'s are iid)} \]

\[ = \sum_{n=p+1}^{N} \left( -\sum_{j=0}^{p} a_{j} s_{n-j} \right) E \left[ \left( \frac{\partial}{\partial u_n} \ln f(u_n; \Phi) \right) \left( \frac{\partial}{\partial \phi_i} \ln f(u_n; \Phi) \right) \right] \]

\[ \text{(using } (B.4)) \]

Since \( f \) is even, derivative of \( \ln f \) w.r.t. \( u_n \) is odd and that w.r.t. \( \phi_i \) is even. The expectation is therefore taken on and odd function and must equal zero. Since this is true for each \( \phi_i \), it can be concluded that \( (B.2) \) holds. Consequently, \( (26) \) holds for the case of AR noise when the GLRT is computed on the basis of conditional likelihood function as in \( (20) \).
Figure 1 Block diagram of GLRT
Figure 2
Composition of $\lambda$

$A(2) = \sum_{j=0}^{\infty} a_j z^{-j}$, $a_0 = 1$

MA filter

$s_l, \ldots, s_N$

square

$\Sigma$

$\mu z^{-1}(\theta)$

$\lambda$
Figure 3  SNR bonus vs. $\epsilon$
for a Mixed-Gaussian process
OFFICE OF NAVAL RESEARCH  
STATISTICS AND PROBABILITY PROGRAM  

BASIC DISTRIBUTION LIST  
FOR  
UNCLASSIFIED TECHNICAL REPORTS  

FEBRUARY 1982  

<table>
<thead>
<tr>
<th>Copies</th>
<th>Copies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Statistics and Probability Program (Code 411(SP))  
Office of Naval Research  
Arlington, VA 22217 |        |
|        | Navy Library  
National Space Technology Laboratory  
Attn: Navy Librarian  
Bay St. Louis, MS 39522 | 3 |
|        | U. S. Army Research Office  
P.O. Box 12211  
Attn: Dr. J. Chandra  
Research Triangle Park, NC 27706 | 1 |
|        |        |
| Defense Technical Information Center  
Cameron Station  
Alexandria, VA 22314 | 12 |
|        | Director  
National Security Agency  
Attn: RS1, Dr. Maar  
Fort Meade, MD 20755 | 1 |
|        |        |
| Commanding Officer  
Office of Naval Research  
Eastern/Central Regional Office  
Attn: Director for Science  
Barnes Building  
495 Summer Street  
Boston, MA 02210 | 1 |
|        | ATAA-SL, Library  
U.S. Army TRADOC Systems  
Analysis Activity  
Department of the Army  
White Sands Missile Range, NM 88002 | 1 |
|        |        |
| Commanding Officer  
Office of Naval Research  
Western Regional Office  
Attn: Dr. Richard Lau  
1030 East Green Street  
Pasadena, CA 91101 | 1 |
|        | ARI Field Unit-USAREUR  
Attn: Library  
c/o ODCSPER  
HQ USAEREUR & 7th Army  
APO New York 09403 | 1 |
|        |        |
| U. S. ONR Liaison Office - Far East  
Attn: Scientific Director  
APO San Francisco 96503 | 1 |
|        | Library, Code 1424  
Naval Postgraduate School  
Monterey, CA 93940 | 1 |
|        |        |
| Applied Mathematics Laboratory  
David Taylor Naval Ship Research and Development Center  
Attn: Mr. G. H. Gleissner  
Bethesda, Maryland 20084 | 1 |
|        | Technical Information Division  
Naval Research Laboratory  
Washington, DC 20375 | 1 |
|        |        |
| Commandant of the Marine Corps  
(Code AX)  
Attn: Dr. A. L. Slafkosky  
Scientific Advisor  
Washington, DC 20380 | 1 |
|        | OASD (I&L), Pentagon  
Attn: Mr. Charles S. Smith  
Washington, DC 20301 | 1 |
<table>
<thead>
<tr>
<th>Copies</th>
<th>Copies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Director</strong>&lt;br&gt;AMSAA&lt;br&gt;Attn: DRXSY-MP, H. Cohen&lt;br&gt;Aberdeen Proving Ground, MD 21005</td>
<td><strong>Reliability Analysis Center (RAC)</strong>&lt;br&gt;RADC/RRAC&lt;br&gt;Attn: I. L. Krulac&lt;br&gt;Data Coordinator/ Government Programs&lt;br&gt;Griffiss AFB, New York 13441 1</td>
</tr>
<tr>
<td>Dr. Gerhard Heiche&lt;br&gt;Naval Air Systems Command&lt;br&gt;(NAIR 03)&lt;br&gt;Jefferson Plaza No. 1&lt;br&gt;Arlington, VA 20360 1</td>
<td>Technical Library&lt;br&gt;Naval Ordnance Station&lt;br&gt;Indian Head, MD 20640 1</td>
</tr>
<tr>
<td>Dr. Barbara Bailar&lt;br&gt;Associate Director, Statistical Standards&lt;br&gt;Bureau of Census&lt;br&gt;Washington, DC 20233 1</td>
<td>Library&lt;br&gt;Naval Ocean Systems Center&lt;br&gt;San Diego, CA 92152 1</td>
</tr>
<tr>
<td>Leon Slavin&lt;br&gt;Naval Sea Systems Command&lt;br&gt;(NSEA 05H)&lt;br&gt;Crystal Mall #4, Rm. 129&lt;br&gt;Washington, DC 20036 1</td>
<td>Technical Library&lt;br&gt;Bureau of Naval Personnel&lt;br&gt;Department of the Navy&lt;br&gt;Washington, DC 20370 1</td>
</tr>
<tr>
<td>B. E. Clark&lt;br&gt;RR #2, Box 647-B&lt;br&gt;Graham, NC 27253 1</td>
<td>Mr. Dan Leonard&lt;br&gt;Code 8105&lt;br&gt;Naval Ocean Systems Center&lt;br&gt;San Diego, CA 92152 1</td>
</tr>
<tr>
<td>Naval Underwater Systems Center&lt;br&gt;Attn: Dr. Derrill J. Bordelon&lt;br&gt;Code 601&lt;br&gt;Newport, Rhode Island 02840 1</td>
<td>Dr. Alan F. Petty&lt;br&gt;Code 7930&lt;br&gt;Naval Research Laboratory&lt;br&gt;Washington, DC 20375 1</td>
</tr>
<tr>
<td>Naval Coastal Systems Center&lt;br&gt;Code 741&lt;br&gt;Attn: Mr. C. M. Bennett&lt;br&gt;Panama City, FL 32401 1</td>
<td>Dr. M. J. Fischer&lt;br&gt;Defense Communications Agency&lt;br&gt;Defense Communications Engineering Center&lt;br&gt;1860 Wiehle Avenue&lt;br&gt;Reston, VA 22090 1</td>
</tr>
<tr>
<td>Naval Electronic Systems Command&lt;br&gt;(NELEX 612)&lt;br&gt;Attn: John Schuster&lt;br&gt;National Center No. 1&lt;br&gt;Arlington, VA 20360 1</td>
<td>Mr. Jim Gates&lt;br&gt;Code 9211&lt;br&gt;Fleet Material Support Office&lt;br&gt;U. S. Navy Supply Center&lt;br&gt;Mechanicsburg, PA 17055 1</td>
</tr>
<tr>
<td>Defense Logistics Studies Information Exchange&lt;br&gt;Army Logistics Management Center&lt;br&gt;Attn: Mr. J. Dowling&lt;br&gt;Fort Lee, VA 23801 1</td>
<td>Mr. Ted Tupper&lt;br&gt;Code M-311C&lt;br&gt;Military Sealift Command&lt;br&gt;Department of the Navy&lt;br&gt;Washington, DC 20390 1</td>
</tr>
</tbody>
</table>
Mr. F. R. Del Priori  
Code 224  
Operational Test and Evaluation  
Force (OPTEVFOR)  
Norfolk, VA 23511  

1