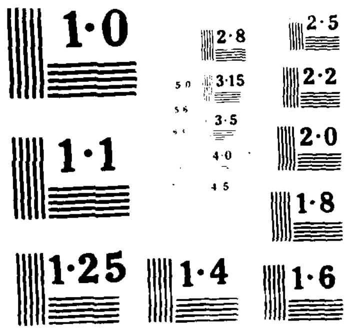
AD-A170 315 LIKELIHOOD RATIOS AND SIGNAL DETECTION FOR NONGAUSSIAN PROCESSES(U) NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS C R BAKER ET AL. MAR 86 N00014-81-K-0373 1/1

UNCLASSIFIED F/G 12/1 NL


1·0
2·8
2·5
3·15
2·2
3·5
2·0
1·1
4·0
1·8
4·5
1·25
1·4
1·6

(12)

LIKELIHOOD RATIOS AND SIGNAL DETECTION FOR NONGAUSSIAN PROCESSES

C.R. Baker
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

A.F. Gualtierotti
IDHEAP, BFSH I
University of Lausanne
Lausanne, Switzerland 1015

To appear, Stochastic Processes in Underwater Acoustics (Lecture Notes in Control and Information Science, Springer-Verlag)

Research supported by ONR Contracts N00014-84-C-0212 and N00014-86-K-0039

LISS 17
March, 1986

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) University of North Carolina Chapel Hill, North Carolina 27514	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-81-K-0373
8c. ADDRESS (City, State and ZIP Code) Statistics & Probability Program Arlington, VA 22217	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
NR	042	269	SRO 105

11. TITLE (Include Security Classification)
Likelihood Ratios & Signal Detection for . . .

12. PERSONAL AUTHOR(S)
C.R. Baker & A.F. Gualtierotti

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
TECHNICAL	FROM ______ TO ______	March, 1986	27

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	signal detection; nonGaussian stochastic processes; Shannon theory; channel capacity; likelihood ratio

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

The emphasis is on development of likelihood ratios and detection algorithms for problems involving nonGaussian data. The first problem considered is that of detecting a nonGaussian signal in Gaussian noise. This frequently arises in active sonar; it could also be important for passive sonar. General results are presented on non-singular detection and likelihood ratio. A recursive discrete-time detection algorithm is obtained and is shown to be a likelihood ratio detector when the signal-plus-noise is Gaussian.

The second major problem considered is that of detecting a signal in spherically-invariant noise (SIN). This is a model which has been proposed for some impulsive-plus-Gaussian environments, and is closely linked to detection problems encountered in some active sonar applications. General results on nonsingular detection and likelihood ratio are first obtained. For detection of a known signal, the behavior of the discrete-time likelihood ratio is analyzed as the sample size increases. Constant-false-alarm-probability detectors.

TITLE CONT.: NonGaussian Processes.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED XX SAME AS RPT XX DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
C.R. Baker	(919) 962-2189	

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE. UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

Block 19 continued:

are given, and an example based on sonar data illustrates the potential loss due to using a Gaussian model when the noise is actually nonGaussian SIN.

Abstract

The emphasis is on development of likelihood ratios and detection algorithms for problems involving nonGaussian data. The first problem considered is that of detecting a nonGaussian signal in Gaussian noise. This frequently arises in active sonar; it could also be important for passive sonar. General results are presented on nonsingular detection and likelihood ratio. A recursive discrete-time detection algorithm is obtained and is shown to be a likelihood ratio detector when the signal-plus-noise is Gaussian.

The second major problem considered is that of detecting a signal in spherically-invariant noise (SIN). This is a model which has been proposed for some impulsive-plus-Gaussian environments, and is closely linked to detection problems encountered in some active sonar applications. General results on nonsingular detection and likelihood ratio are first obtained. For detection of a known signal, the behavior of the discrete-time likelihood ratio is analyzed as the sample size increases. Constant-false-alarm-probability detectors are given, and an example based on sonar data illustrates the potential loss due to using a Gaussian model when the noise is actually nonGaussian SIN.

A-1

LIKELIHOOD RATIOS AND SIGNAL DETECTION FOR NONGAUSSIAN PROCESSES

1. INTRODUCTION

NonGaussian signal detection problems arise in several applications of underwater acoustics. NonGaussian signal processes occur for active sonar when the reflecting target (with surface undergoing random motions) has only a few dominating scatterers. The noise in such applications is frequently Gaussian, so that the detection problem is that of detecting a nonGaussian signal embedded in additive Gaussian noise.

Problems of detecting a signal in nonGaussian noise also arise; for example, for sonars operating under ice. Noise due to ice-cracking, creaking, floe-smashing, etc., contributes a component which has been found to have substantial nonGaussian behavior [6]. In addition, active sonars operating under ice near the surface may encounter a nonGaussian component due to specular reflection from the irregular under-ice surface. Another environmental situation which may produce nonGaussian noise is shallow-water reverberation.

Optimum detection algorithms require knowledge of the statistical properties of data processes. For applications involving nonGaussian noise with a strong impulsive component, a useful univariate noise model has been developed by Middleton [12]. Much remains to be done in this area. Development of optimum detection algorithms requires knowledge of multivariate statistical models for both the noise and the signal-plus-noise processes. At present, such models do not exist for some of the most important nonGaussian environments. Their development will require a mix of physics, mathematics, statistics, and extensive computational investigations. These are challenging problems whose solution must be obtained before one can obtain optimum detection algorithms.

This contribution first considers algorithms for the detection of nonGaussian signals in Gaussian noise. Results are summarized for the continuous-time problem; more attention is given to discrete-time approximations. A discrete-time recursive algorithm is given. It is shown that (under appropriate assumptions) this discrete-time algorithm is a likelihood ratio detector if the signal-plus-noise process is Gaussian. Attention is then turned to signal detection for problems involving nonGaussian spherically-invariant noise (SIN). The univariate Class A model of Middleton [12] is seen to be a special case of SIN. The likelihood ratio for detection of a signal in SIN is derived for both continuous-time and discrete-time applications. Approximations, including constant false-alarm probability (CFAP) detectors, are discussed. The effect of sample size is also considered. These results indicate that robust detection can be achieved for detection of known signals in spherically-invariant noise. In applying these results to Middleton's Class A model, it is shown that placing that model in the context of SIN provides a number of useful consequences.

All stochastic processes to be discussed are real-valued and defined on a probability space (Ω, β, P). For processes in continuous time, the parameter set is $[0,T]$, and all such processes are assumed to be mean-square continuous. All noise processes are assumed to have zero mean. $(V(t))$ will denote a stochastic process, while $V(t)$ will denote the random variable obtained by sampling the process at the time t. The argument ω in Ω will typically be suppressed: $V(t) \equiv V(t,\omega)$. $\mathcal{L}_2[0,T]$ is the linear space of all Lebesgue-square-integrable functions on $[0,T]$. $L_2[0,T]$ is the set of equivalence classes $[u]$ obtained from functions u in $\mathcal{L}_2[0,T]$. For a noise process $(N(t))$, r_N will denote the covariance function: $r_N(t,s) = E\, N(t)N(s)$. R_N will denote the covariance operator of $(N(t))$; that is, the integral operator in $L_2[0,T]$ having r_N as its kernel. R_N will be assumed strictly positive; (λ_n) is the sequence of (strictly positive) eigenvalues of R_N, with (e_n) corresponding c.o.n. eigenvectors. $\langle u,v\rangle = \int_0^T u(t)v(t)dt$ for u and v in $L_2[0,T]$. The signal-plus-noise process at time t will be $Y(t) = S(t) + N(t)$. In the continuous-time case, the likelihood ratio sought is on $L_2[0,T]$: $d\mu_Y/d\mu_N$, where μ_Y (resp., μ_N) is the probability on $L_2[0,T]$ induced by the stochastic process $(Y(t))$ (resp., $(N(t))$).

For a noise $(N(t))$ with covariance function r_N, H_N will denote the reproducing kernel Hilbert space H_N of r_N. As is well-known, there is an isometry between H_N and range$(R_N^{\frac{1}{2}})$: the element u is in H_N if and only if there exists a unique element $[u]$ in range$(R_N^{\frac{1}{2}})$ generated by u. The inner product of two elements u and v in H_N is given by $[u,v]_N = \sum_{n\geq 1} \langle [u],e_n\rangle\langle [v],e_n\rangle/\lambda_n$. Since r_N is taken to be continuous, the elements of H_N will be continuous functions. It can also be noted that range($R_N^{\frac{1}{2}}$) is a real separable Hilbert space under the above inner product (i.e., with respect to the inner product $([u],[v])_N = [u,v]_N$). For simplicity, the element $[u]$ in $L_2[0,T]$ will usually be written simply as u.

For observations $\underline{x}$ in n-dimensional Euclidean space E^n, the noise covariance matrix will be assumed strictly positive. $\underline{A}^*$ is the transpose of the matrix $\underline{A}$. For a bounded linear operator A in $L_2[0,T]$, A^* will denote the adjoint.

In discussing existence of likelihood ratios for continuous-time processes, it is necessary to introduce more mathematical structure. Thus, let $(X(t))$, t in $[0,T]$ be a m.s. continuous stochastic process. $\sigma\{X_s, s \leq t\}$ is the σ-field generated by $\{X_s, s \leq t\}$; $\underline{\sigma}_t^0(X)$ is the filtration consisting of all the σ-fields $\sigma\{X_s, s \leq t'\}$ for $0 \leq t' \leq t$. $\underline{\sigma}^0(X)$ will denote $\underline{\sigma}_T^0(X)$. $\underline{\sigma}_t(X)$ (resp., $\underline{\sigma}(X)$) will denote the filtration generated by $\underline{\sigma}_t^0(X)$ (resp., $\underline{\sigma}^0(X)$) and all sets of P-measure zero, completed with respect to the underlying probability P. If $(X(t))$ and $(V(t))$ are two such processes, then $\underline{\sigma}_t(V) \vee \underline{\sigma}_t(X)$ will denote the smallest filtration containing both $\underline{\sigma}_t(V)$ and $\underline{\sigma}_t(X)$, with $\underline{\sigma}(V) \vee \underline{\sigma}(X)$ similarly defined.

The continuous-time problems considered here will be modeled in $L_2[0,T]$. Somewhat similar results can be obtained by considering the probabilities induced on $R^{[0,T]}$, the real-valued functions on $[0,T]$. However, those results [5] are not so complete as those for $L_2[0,T]$.

For μ_Y and μ_N probabilities on the Borel sets of $L_2[0,T]$, $\mu_Y \ll \mu_N$ denotes absolute continuity of μ_Y with respect to μ_N (so that the likelihood ratio $d\mu_Y/d\mu_N$ exists). $\mu_Y \perp \mu_N$ denotes orthogonal probability measures; in detection applications, orthogonal measures imply singular (perfect) detection. $\mu_Y \sim \mu_N$ denotes mutual absolute continuity: $\mu_Y \ll \mu_N$ and $\mu_N \ll \mu_Y$. P_D will denote probability of detection = probability of correctly deciding signal present. P_{FA} denotes probability of false alarm = probability of incorrectly deciding signal present. $P_{FA} = 0$ implies $P_D = 0$ if $\mu_{S+N} \ll \mu_N$; $P_D = 1$ implies $P_{FA} = 1$ if $\mu_N \ll \mu_{S+N}$. Thus, $\mu_{S+N} \sim \mu_N$ is the situation usually assumed to hold for practical problems in signal detection. We refer to this as non-singular detection.

Treatment of the continuous-time case introduces substantial complication into the analysis. However, it clarifies structure, enables one to obtain discrete-time finite-sample algorithms as approximations, and provides valuable performance bounds.

2. DETECTION OF NONGAUSSIAN SIGNALS IN GAUSSIAN NOISE

An important active sonar application is the detection of nonGaussian signals embedded in additive Gaussian noise. For example, when the noise background is reverberation-limited, the scatterers giving rise to the reverberation can frequently be assumed to be statistically-independent in their reflecting properties. Application of the central limit theorem then gives a Gaussian process for the reverberation process. However, if the target return is primarily due to reflections from a few random scatterers (each contributing random phase and amplitude), then the composite reflection from the target will generally be nonGaussian. In this particular application, the signal and noise processes are dependent, and the noise process is nonstationary.

Other applications may also involve detection of nonGaussian signals in Gaussian noise. For example, in passive sonar the background noise can frequently be assumed to be Gaussian and stationary. However, signal sources such as ship-radiated noise need not be Gaussian.

Full solution of such problems ideally includes determining conditions for nonsingular detection, and then (when nonsingular detection holds) determining the likelihood ratio.

If the signal-plus-noise process is also Gaussian, then conditions for nonsingular detection and the form of the likelihood ratio are well known [4, 14]. If the signal is nonGaussian and independent of the noise, then sufficient conditions for nonsingular detection are given in [3]. With the noise Gaussian, the sufficient condition is that the sample paths of the signal process belong (w.p. 1) to H_N, the reproducing kernel Hilbert space of the noise. Under mild assumptions, an expression for the likelihood ratio can also be obtained from the results of [3].

If nothing is known about the signal-plus-noise process except its covariance and mean functions, then of course a likelihood ratio detector cannot be determined. However, if one limits consideration to quadratic-linear operations on the data (in forming a test statistic), then the deflection criterion can be used to determine the optimum operation. That is, let T be the class of all admissible test statistics τ. The deflection of τ is then $D_{01}(\tau) = (E_N\tau(x) - E_{S+N}\tau(x))^2/(E_N\tau^2(x) - [E_N\tau(x)]^2)$, where $E_N(\cdot)$ (resp., $E_{S+N}(\cdot)$) denotes expectation w.r.t. the noise (resp., signal-plus-noise). The problem then consists of determining $\sup_T D_{01}(\tau)$, and determining a τ achieving this supremum or a sequence (τ_n) converging to the supremum. The optimum quadratic operation for discrete-time finite-sample data is given in [1], while [2] contains the solution for the infinite-dimensional case and results linking deflection to nonsingular detection.

The results on deflection given in [1] and [2] apply to problems where the signal-plus-noise process is neither Gaussian nor consisting of the noise plus an independent signal process. For such problems, it is desirable to have general conditions for nonsingular detection and also expressions for the likelihood ratio. Currently-available data models may be inadequate to fully utilize such results, but their availability for future use is clearly desirable. Results for the special continuous-time case when the noise is the Wiener process have been known for many years [11]. However, the Wiener process has properties that are not observed in practical sonar problems: sample functions that are almost surely nondifferentiable at almost all time points, the Markov property, and the martingale property. Thus, the design of future optimum signal detection systems requires results beyond those already mentioned; such results have recently been obtained [5].

The results contained in [5] include general conditions for nonsingular detection of a possibly nonGaussian signal imbedded in additive Gaussian noise. The work is based on the spectral representation of second-order stochastic processes, particularly as developed by Hida [10]. The general problem is that of discriminating between a Gaussian noise process $(N(t))$, t in $[0,T]$, and a possibly nonGaussian process $(Y(t))$, t in $[0,T]$.

The basic assumptions made in [5] are the following:

(A.2-1) $(N(t))$ vanishes almost surely at $t = 0$;

(A.2-2) $(N(t))$ has a purely-continuous spectral representation of multiplicity $M < \infty$.

Assumption (A.2-2) is equivalent to $(N(t))$ having a representation of the form

$$N(t) = \sum_{i=1}^{M} \int_0^t F_i(t,s)dB_i(s) \qquad (2.1)$$

where $\{(B(t)): i \leq M,\ t \text{ in } [0,T]\}$ is a family of independent-increment mutually-independent path-continuous zero-mean Gaussian processes, and each F_i is a Borel-measurable function on $[0,T]\times[0,T]$ with $F_i(t,s) = 0$ for $s>t$. This representation also satisfies

$\sum_{i=1}^{M} \int_0^T \int_0^T F_i^2(t,s)d\beta_i(s)dt < \infty$, where β_i is the Borel measure on $[0,T]$ defined by the non-decreasing variance of $(B_i(t))$:

$\beta_i(a,b] = EB_i^2(b) - EB_i^2(a)$.

The representation (2-1) is taken to be the proper canonical representation for $(N(t))$ [10]. One consequence is that the completion of the σ-field $\sigma\{B_i(s): i\leq M,\ s\leq t\}$ is the same as the completion of $\sigma\{N(s): s\leq t\}$ for each t in $[0,T]$. In general, the equality (2-1) holds almost surely dP for each fixed t in $[0,T]$; by assuming that $(N(t))$ is separable w.r.t. closed sets, one obtains a.s. path equality.

The basic results on non-singular detection of a possibly nonGaussian signal embedded in additive Gaussian noise, as given in [5], entail both a set of sufficient conditions [5, Theorem 2] and a set of necessary conditions [5, Theorem 3]. The sufficient conditions for absolute continuity on $L_2[0,T]$ are given in the following result.

<u>Prop. 2.1</u> [5, Theorem 3] Let $(V(t))$ be a stochastic process independent of $(N(t))$. Suppose that $(S(t))$ is a stochastic process adapted to $\underline{\sigma}(N) \vee \underline{\sigma}(V)$ and with paths a.s. in H_N. If $Y(t) = S(t) + N(t)$ a.e. dtdP, then $\mu_Y << \mu_N$.

Both the sufficient conditions and the necessary conditions include the requirement that the signal process have a representation with almost all paths in the reproducing kernel Hilbert space of the noise covariance function r_N. With the representation (2-1), this means that almost all sample functions of the signal process have a representation of the form

$$S(t) = \sum_{i=1}^{M} \int_0^t F_i(t,s)Q_i(s)d\beta_i(s) \qquad (2\text{-}2)$$

where $(Q_i(t))$ is a stochastic process with almost all paths in $\mathcal{L}_2[\beta_i]$: $\int_0^T Q_i^2(s)d\beta_i(s) < \infty$ a.s. dP. The remaining conditions for absolute continuity embody measurability conditions on the signal process. These conditions are given in terms of the noise process $(N(t))$ and a stochastic process $(V(t))$ independent of the noise. They are essentially related to the signal process being a causal functional of the two processes $(N(t))$ and $(V(t))$. The basic idea is that the signal may be a causal functional of both the noise process (as in the case of dependent signal and noise) and an independent "message" process.

The likelihood ratio $d\mu_{S+N}/d\mu_N$ for this problem is given in [5, Theorem 7]. Define a vector stochastic process $(\underset{\sim}{Z}(t))$ by

$$Z_i(t) = \int_0^t Q_i(s)d\beta_i(s) + B_i(t) \tag{2-3}$$

where the processes $(Q_i(t))$ are those appearing in (2-2) above. Then

$$S(t) + N(t) = \sum_{i=1}^{M} \int_0^t F_i(t,s)dZ_i(s) \tag{2-4}$$

The vector processes $(\underset{\sim}{Z}(t))$ and $(\underset{\sim}{B}(t))$ define probabilities $P_{\underset{\sim}{B}}$ and $P_{\underset{\sim}{Z}}$ on the space of all M-component vector functions on [0,T] whose component functions are all continuous. Under the conditions for existence of $d\mu_{S+N}/d\mu_N$, $dP_{\underset{\sim}{Z}}/dP_{\underset{\sim}{B}}$ will also exist, and for an observation x in $L_2[0,T]$,

$$[d\mu_{S+N}/d\mu_N](x) = [dP_{\underset{\sim}{Z}}/dP_{\underset{\sim}{B}}](\underset{\sim}{m}[x]) \tag{2-5}$$

a.e. $d\mu_N(x)$. $\underset{\sim}{m}(x)$ is an M-component vector of continuous functions, defined by

$$m_i[x](t) = \sum_{n\geq 1} \langle x,e_n\rangle\langle f_t^i,e_n\rangle/\lambda_n, \tag{2-6}$$

$$f_t^i(s) = \int_0^t F_i(s,u)d\beta_i(u).$$

The likelihood ratio $dP_{\underset{\sim}{Z}}/dP_{\underset{\sim}{W}}$ of (2-5) has some explicit known representations [11], depending on the properties of $(\underset{\sim}{Z}(t))$. These representations are based on the fact that each $(B_i(t))$ is a path-continuous Gaussian martingale.

The results given above are for continuous-time observations. In sonar applications, it is desirable to have discrete-time recursive algorithms, which do not require complete recomputation of the test statistic each time a new data point is received. Moreover, it is desirable to have an algorithm with parameters that can be estimated from data, since a complete data model will not usually be available. Such an algorithm will now be derived. It will be based on the following additional assumptions:

(A.2-3) The noise process has multiplicity $M=1$, and the process $(B_1(t))$ is the standard Wiener process $(W(t))$; thus $N(t) = \int_0^t F(t,s)dW(s)$, where F is a Volterra kernel with $\int_0^T \int_0^T F^2(t,s)dsdt < \infty$;

(A.2-4) The process $(Z(t))$ defined in (2-3) is a diffusion with respect to the Wiener process and has memoryless drift function, so that $Z(t) = \int_0^t \sigma[Z(s)]ds + W(t)$. (2-7)

The assumption (A.2-3) is reasonable from several viewpoints, such as the fact that multiplicity-one processes are dense (by a mean-square distance criterion) in the space of all second-order processes, and that any Gaussian vector can be represented as the result of a lower-triangular matrix operating on white Gaussian noise. One can also show that the assumption (A.2-3) is satisfied whenever the noise process has a proper canonical representation $N(t) = \int_0^t F(t,s)dB(s)$, where the variance of $(B(t))$ is an absolutely continuous function on $[0,T]$.

The assumption (A.2-4) is less tenable; it is made primarily for computational convenience (which is in fact not very convenient, even so) when the signal-plus-noise statistics are unknown. It does permit one to consider a very large class of signal-plus-noise processes without having complete knowledge of the statistics. Of course, if a complete mathematical model is available, the assumptions (A.2-3) and (A.2-4) need not be made (if $dP_{\underset{\sim}{Z}}/dP_{\underset{\sim}{B}}$ can be determined).

For the detection problem as defined above, the general form (under a mild restriction) of the likelihood ratio is

$$[d\mu_{S+N}/d\mu_N](x) = \lim_n \exp [\Lambda^n(x)]$$

where $0 = t_0^n < t_1^n < t_2^n < \dots < t_n^n = T$ is a partition of $[0,T]$ such that $\sup_j |t_{j+1}^n - t_j^n| \to 0$,

$$\Lambda^n(x) = \sum_{i=0}^{n-1} \sigma(m[x](t_i^n))(m[x](t_{i+1}^n) - m[x](t_i^n)) - (1/2)\sum_{i=0}^{n-1} \sigma^2(m[x](t_i^n))(t_{i+1}^n - t_i^n), \quad (2\text{-}8)$$

and the limit exists in the norm of $L_1[\mu_N]$.

The representation of $(N(t))$ by $N_t = \int_0^t F(t,s)dW(s)$ yields that $R_N = FF^*$, where F is the integral operator with $F(t,s)$ as its kernel, and F^* is its adjoint. This can be used to provide an expression for the function m appearing in (2-5) and (2-6) that does not require calculation of eigenvalues and eigenvectors.

First, notice that $\langle e_j, f_t \rangle = \int_0^T \int_0^t F(s,u)du\, e_j(s)ds$ $= \int_0^t \int_0^T F(s,u)e_j(s)dsdu = [LF^*e_j](t)$, where $[Lf](t) \equiv \int_0^t f(v)dv$. Using this, the expression (2-6) for m can be rewritten as

$$m[x](t) = \lim_{k\to\infty} [LF^* \sum_1^k \langle e_j, x\rangle R_N^{-1} e_j](t)$$

$$= \lim_{k\to\infty} [LF^* R_N^{-1} P_k x](t)$$

where $P_k x$ is the projection of the function x on the subspace spanned by $\{e_1, \ldots, e_k\}$. Since $R_N^{-1} = F^{*-1}F^{-1}$, the preceding becomes $m[x](t) = \lim_{k\to\infty} [LF^{-1}P_k x](t)$.

A basic difficulty is that (with probability one [3]) the observation x will not be in the domain of the operator F^{-1}, so that $F^{-1}x$ is not defined. In fact, LF^{-1} will in general not be a bounded linear operator. However, for almost all sample functions x (either from noise or signal-plus-noise), $m[x](\cdot)$ is a continuous function of t. Thus the map m is a linear operator from $L_2[0,T]$ into $C[0,T]$ whose domain includes (with probability one) all sample functions of the noise and signal-plus-noise processes.

The difficulty in implementation of the likelihood ratio (2-8) will lie in determining the function σ and linear operator m. σ is a parameter of the signal-plus-noise process, and its estimation is a problem of considerable interest in stochastic processes (as the drift of a diffusion) and in stochastic filtering. The possibly unbounded linear operator m, mapping $L_2[0,T]$ into $C[0,T]$, depends only on the covariance function of the noise. If the noise covariance function is known, then the preceding expressions can be used to obtain a discrete-time finite-sample approximation to the likelihood ratio. Here we consider such approximations when one knows only the covariance matrix of the noise.

Let $\underline{R}_N$ denote the covariance matrix of the noise; one can write $\underline{R}_N = \underline{F}\,\underline{F}^*$, where the matrix $\underline{F}$ is lower triangular. Now, the expression for m given above is of the form

$$m[x](t) = \lim_{k\to\infty} [LF^{-1}P_k x](t),$$

where $R_N = FF^*$, L is the integration operator, and P_k is the projection of x onto the subspace spanned by $\{e_1,\dots,e_k\}$, where $\{e_n, n\geq 1\}$ are o.n. eigenvectors of R_N. Thus, a reasonable procedure is simply to replace this expression by $m[x] = \underline{L}\ \underline{F}^{-1}\underline{x}$, where $\underline{x}$ is the observed data vector, and $\underline{L}$ is the summation operator in E^k; $(\underline{L}\ \underline{x})_j = \sum_{i=1}^{j} x_i$.

There is a fundamental difference between the above approximation to m and the exact result. As previously mentioned, $F^{-1}x$ is (with probability one) not defined for the continuous-time situation; here, of course, there is no such problem for $\underline{F}^{-1}\underline{x}$.

Implementation of the discrete-time algorithm for a fixed sampling interval, Δ, will now be considered. Then, when the observation is an n-component vector, and the above approximation is used, one obtains as an approximation to the log-likelihood ratio the expression

$$\begin{aligned}\Lambda^n(\underline{x}) &= \sum_{j=0}^{n-1} (\sigma[(\underline{L}\ \underline{F}^{-1}\underline{x})_j])[(\underline{L}\ \underline{F}^{-1}\underline{x})_{j+1} - (\underline{L}\ \underline{F}^{-1}\underline{x})_j] \\ &\quad - (\Delta/2) \sum_{j=0}^{n-1} \sigma^2\ [(\underline{L}\ \underline{F}^{-1}\underline{x})_j] \qquad (2\text{-}9) \\ &= \sum_{j=0}^{n-1} (\sigma[(\underline{L}\ \underline{F}^{-1}\underline{x})_j])[(\underline{F}^{-1}\underline{x})_{j+1}] - (\Delta/2) \sum_{j=0}^{n-1} \sigma^2\ [(\underline{L}\ \underline{F}^{-1}\underline{x})_j].\end{aligned}$$

If now a new data point x_{n+1} is observed, the approximation has the recursive form

$$\Lambda^{n+1}(\underline{x}) = \Lambda^n(\underline{x}) + \sigma[(\underline{L}\ \underline{F}^{-1}\underline{x})_n](\underline{F}^{-1}\underline{x})_{n+1} - (\Delta/2)\ \sigma^2[(\underline{L}\ \underline{F}^{-1}\underline{x})_n]. \quad (2\text{-}10)$$

One notes the following:

(1) Implementation and calculation of Λ require the following operations. First, the function σ must be known and programmed. Given the value of $\Lambda^n(\underline{x}^n)$ and the observation $\underline{x}^n = (x_1,\dots,x_n)$, one stores $\Lambda^n(\underline{x}^n)$, $\underline{x}^n$, $\sigma[(\underline{L}\ \underline{F}^{-1}\underline{x})_n]$, and $(\underline{L}\ \underline{F}^{-1}\underline{x}^n)_n$. When the data point x_{n+1} is received, it is only

necessary to use $\underline{x}^{n+1}$ to calculate $(\underline{F}^{-1}\underline{x}^{n+1})_{n+1}$, which means to cross-correlate the observation $\underline{x}^{n+1}$ with the n+1 row of $\underline{F}^{-1}$. This number, say b_{n+1}, is then used to form $\Lambda^{n+1}(\underline{x}^{n+1})$,

$$\Lambda^{n+1}(\underline{x}^{n+1}) = \Lambda^{n}(\underline{x}^{n}) + \sigma[\sum_{1}^{n} b_i]b_{n+1} - (\Delta/2)\ \sigma^2\ [\sum_{1}^{n} b_i].$$

This is much simpler than a procedure whereby the function $\underline{m} = \underline{L}\ \underline{F}^{-1}$ is expressed in terms of its eigenvalues and eigenfunctions, since those quantities would have to be stored for E^n and all the sample indices $n \geq 1$, and a complete new calculation done for each new sample point observed.

(2) As already noted, the expression $\Lambda^n(\underline{x}^n)$ can only be considered as an approximation to the discrete-time likelihood ratio. This approximation becomes more valid as n increases, since it amounts to representing the noise vector $\underline{N}$ by $N_i = \sum_{j=1}^{i-1} F_{ij}(\underline{\Delta W})_j$, where $\underline{\Delta W}$ is a vector of i.i.d. $N(0,\Delta)$ random variables. As n increases, $\sum_{j=1}^{i-1} F(i\Delta,\ j\Delta)(\underline{\Delta W})_j$ will converge in mean-square to $N(i\Delta) = N(t)$, keeping $i\Delta = t$, where the function F is that appearing in the representation $N_t = \int_0^t F(t,s)dW_s$. Thus, as n increases, the representation of $\underline{N}$ by $N(t) = \sum_{j=1}^{i-1} F_{ij}(\underline{\Delta W})_j$ converges to the representation satisfying the original continuous-time models for noise and signal-plus-noise.

(3) σ can be estimated from a sample of data representative of signal-plus-noise. In discrete time, the procedure is as follows, given an observed $\underline{S}$+$\underline{N}$ vector $\underline{X}$.

a) Form $\underline{\Delta Z} = \underline{F}^{-1}\underline{X}$, where $\underline{R}_N = \underline{F}\ \underline{F}*$, F lower-triangular, $(\underline{\Delta Z})_i = Z(i\Delta) - Z([i-1]\Delta)$, $Z_0 = 0$.

b) $Z(i\Delta) = \Delta \sum_{j=1}^{i-1} \sigma(Z_{j\Delta}) + W(i\Delta)$.

Given the sample vector $\underline{Z}$ obtained from b), the function σ can be estimated. A maximum-likelihood procedure is given in [8].

Of course, the approximations (2-8) and (2-9) need not be likelihood ratios for a fixed finite set of sample points. However, it will now be shown that (2-9) is a likelihood ratio when the function σ is linear. In this case, $\underline{S}+\underline{N}$ is Gaussian, so that the likelihood ratio $d\mu_{\underline{S}+\underline{N}}/d\mu_{\underline{N}}$ can be found.

In accord with the model for $(Z(t))$, the discrete-time representation is (for σ linear)

$$Z_{k+1} = \Delta \sum_{j=1}^{k} a_j Z_j + W_{k+1}. \tag{2-11}$$

It will be shown that for an observation vector $\underline{x}$ in E^n,

$$-\underline{x}^*(\underline{R}_Z^{-1} - \underline{R}_W^{-1})\underline{x}/2 = -(\Delta/2)\sum_{i=1}^{n-1} a_i^2 x_i^2 + \sum_{i=1}^{n-1} a_i x_i (x_{i+1} - x_i) \tag{2-12}$$

The LHS of (2-12) is the log-likelihood ratio (within a constant) of $dP_{\underline{Z}}/dP_{\underline{W}}$. Given the equality (2-12), if one has that $\underline{N} = \underline{F}\ \underline{\Delta W}$, $\underline{S}+\underline{N} = F\ \underline{\Delta Z}$, then $[d\mu_{\underline{S}+\underline{N}}/d\mu_{\underline{N}}](\underline{x}) = [dP_{\underline{\Delta Z}}/dP_{\underline{\Delta W}}](\underline{F}^{-1}\underline{x})$.

Thus let

$$Z_{k+1} = \Delta \sum_{j=1}^{k} a_j Z_j + W_{k+1}, \qquad k \geq 1,$$

$$Z_1 = W_1.$$

Let $\underline{A}$ be the matrix $\mathrm{diag}[a_1, \ldots, a_n]$. The RHS of (2-9), evaluated at $\underline{y} = \underline{F}\ \underline{L}^{-1}\underline{x}$, then becomes

$$\left[\sum_{j=1}^{n-1} (Ax)_j (x_{j+1} - x_j) - \tfrac{1}{2}\Delta \sum_{j=1}^{n-1} (Ax)_j^2\right]. \tag{2-13}$$

To show that (2-9) is a likelihood ratio test statistic, it will first be shown that (2-13) is equal to $-\underline{x}^*(\underline{R}_Z^{-1} - \underline{R}_W^{-1})\underline{x}/2 = \log\,[dP_{\underline{Z}}/dP_{\underline{W}}](\underline{x})$ + constant.

The above representation for $\underline{Z}$ gives

$$(\underline{I} + \Delta\underline{A})\underline{Z} = \Delta\underline{L}\ \underline{A}\ \underline{Z} + \underline{W}$$

so
$$\underline{Z} = \underline{B}^{-1}\underline{W}$$
$$\underline{B} = I + \Delta\underline{A} - \Delta\underline{L}\ \underline{A}.$$

$\underline{Z}$ thus has covariance matrix $\underline{R}_Z = \underline{B}^{-1}\underline{R}_W\underline{B}^{*-1}$, so $\underline{R}_Z^{-1} = \underline{B}^*\underline{R}_W^{-1}\underline{B}$. Since $\underline{R}_W(i,j) = \Delta\min(i,j)$, $\underline{R}_W = \Delta LL^*$, and thus

$$\underline{R}_Z^{-1} = (I + \Delta\underline{A} - \Delta\underline{A}\ \underline{L}^*)\underline{L}^{*-1}\underline{L}^{-1}(\underline{I} + \Delta\underline{A} - \Delta\underline{L}\ \underline{A})/\Delta$$

where $(\underline{L}^{-1})_{ij} = 1$ if $i=j$
$= -1$ if $i=j+1$
$= 0$ otherwise.

This gives $\underline{R}_Z^{-1} - \underline{R}_W^{-1}$

$= [\Delta\underline{A}\underline{L}^{*-1}\underline{L}^{-1}\underline{A} + \underline{A}\underline{L}^{*-1}\underline{L}^{-1} + \underline{L}^{*-1}\underline{L}^{-1}\underline{A} - \underline{A}\underline{L}^{-1}(\underline{I}+\Delta\underline{A}) - (\underline{I}+\Delta\underline{A})\underline{L}^{*-1}\underline{A} + \Delta\underline{A}^2]$

and for a data vector $\underline{x}$,

$$\underline{x}^*(\underline{R}_Z^{-1} - \underline{R}_W^{-1})\underline{x} = \Delta(\underline{L}^{-1}\underline{A}\underline{x})^*\underline{L}^{-1}\underline{A}\underline{x} + 2(\underline{L}^{-1}\underline{x})^*\underline{L}^{-1}\underline{A}\underline{x}$$
$$- 2(\underline{L}^{-1}\underline{x})^*\underline{A}\underline{x} - 2\Delta(\underline{L}^{-1}\underline{A}\underline{x})^*\underline{A}\underline{x} + \Delta(\underline{A}\underline{x})^*\underline{A}\underline{x}.$$

The three terms containing Δ sum to $\Delta \sum_{i=1}^{n-1} a_i^2 x_i^2$, while the other two terms sum to $-2 \sum_{i=1}^{n-1} a_i x_i (x_{i+1} - x_i)$, so that

$$\underline{x}^*(\underline{R}_Z^{-1} - \underline{R}_W^{-1})\underline{x} = -2 \sum_{j=1}^{n-1} a_j x_j (x_{j+1} - x_j) + \Delta \sum_{j=1}^{n-1} a_j^2 x_j^2, \text{ as desired.}$$

This shows that (2-9), evaluated at $\underline{y} = \underline{F}\,\underline{L}^{-1}\underline{x}$, satisfies $\Lambda^n(\underline{F}\,\underline{L}^{-1}\underline{x}) = -\underline{x}^*(\underline{R}_Z^{-1} - \underline{R}_W^{-1})\underline{x}/2 = \log\,[dP_{\underline{Z}}/dP_{\underline{W}}](\underline{x}) + \text{constant}$. Moreover, $[d\mu_{\underline{S}+\underline{N}}/d\mu_{\underline{N}}](\underline{y}) = [dP_{\Delta\underline{Z}}/dP_{\Delta\underline{W}}](\underline{F}^{-1}\underline{y}) = [dP_{\underline{Z}}/dP_{\underline{W}}](\underline{L}\,\underline{F}^{-1}\underline{y})$, the last equality because $\Delta\underline{Z} = \underline{L}^{-1}\underline{Z}$, $\Delta\underline{W} = \underline{L}^{-1}\underline{W}$. With $\underline{y} = \underline{F}\,\underline{L}^{-1}\underline{x}$, $[d\mu_{\underline{S}+\underline{N}}/d\mu_{\underline{N}}](\underline{y}) = [dP_{\underline{Z}}/dP_{\underline{W}}](\underline{x})$ = (from above) $\exp[\Lambda^n(\underline{y}) + \text{constant}]$.

Thus, when the above assumptions are satisfied (including the assumption that $\underline{Z}$ is a Gaussian vector), the approximation given in (2-9) is a discrete-time finite-sample likelihood ratio.

3. SPHERICALLY-INVARIANT NOISE (SIN) MODELS

Let $(N(t))$, t in T, be a real-valued zero-mean stochastic process on a probability space (Ω, β, P). N is said to be spherically invariant if it has the representation $N(t) = AG(t)$ for each t in T, where G is a Gaussian process and A is a random variable which is independent of G and which has finite second moment. Since it can be assumed that $EA^2 = 1$, the covariance of N can be taken to be the same as that of G. Thus, the finite-dimensional distributions of N are completely determined by its covariance and by the distribution of the random variable A. SIN can thus be viewed as a first step away from Gaussian noise.

If the random variable A is discrete, then the distribution of the random vector $(N(t_1), \ldots, N(t_r))$ is given by the density function

$$f(\underline{N}) = \sum_{i=1}^{K} p_i n(0, a_i^2 R) \tag{3-1}$$

where n(a,B) is the density of a Gaussian random vector (in E^r) with mean a and covariance matrix B. In the representation (3-1), $P[A=a_i] = p_i$, and $K \leq \infty$ is the number of distinct values that A assumes with positive probability. In this paper, it will be assumed throughout that A is a discrete random variable. We also assume (WLOG) that $EA^2 = 1$ and that A is strictly positive.

The model for univariate impulsive-plus-Gaussian noise developed by Middleton [12] takes two basic forms, defined as Class A and Class B, depending on the relative bandwidth of noise and receiver. The Class A model is defined to exist when the impulsive noise pulses do not cause transients in the front end of the receiver; it is thus a model for narrowband noise. The univariate density function as developed by Middleton has the form [12]

$$f(x) = e^{-U} \sum_{m=0}^{\infty} \frac{U^m}{m!} \frac{1}{\sqrt{2\pi}\, a_m} \exp\left\{-\tfrac{1}{2} x^2/a_m^2\right\} \tag{3-2}$$

where U is the "overlap index" and (a_m^2) is a sequence of variance components. The overlap index is defined to be the average number of arrivals per second multiplied by the average length of the pulse. The variance component a_m^2 is defined by $a_m^2 = (mU^{-1} + \Gamma)/(1 + \Gamma)$, where Γ is the ratio of the intensities of the Gaussian and non-Gaussian components of the noise.

It can be seen that (3-2) is the probability distribution of a spherically-invariant random variable X = AY, where Y is a zero-mean unit-variance Gaussian r.v., and A is an independent r.v. taking the values (a_m) with

$$p_m = P[A=a_m] = U^m e^{-U}/m! \tag{3-3}$$

In fact, $E\,A^2 = \sum_{m=0}^{\infty} \frac{U^m e^{-U}}{m!} (\frac{m}{U} + \Gamma)/(1+\Gamma) = 1.$

In [16], Spaulding and Middleton analyse the problem of detecting a known signal in Class A noise by assuming independent sampling, so that the sampled noise data has joint density function (n samples)

$$p(\underline{x}) = \prod_{i=1}^{n} \sum_{m=0}^{\infty} p_m \frac{1}{\sqrt{2\pi}\, a_m} \exp\left[-\frac{x_i^2}{2a_m^2}\right]. \tag{3-4}$$

However, if the r.v. A is constant over the sampling interval, then the density of $\underline{X} = A\underline{Y}$, where the components of $\underline{Y}$ are i.i.d. N(0,1), is

$$p(\underline{x}) = \sum_{m=0}^{\infty} p_m \frac{1}{\left[2\pi\ a_m^2\right]^{n/2}} \exp\left[-\frac{\|x\|^2}{2a_m^2}\right] \tag{3-5}$$

where $\|x\|^2 \equiv \sum_{i=1}^{n} x_i^2$. When the Gaussian process Y has non-singular covariance matrix R, then the class A noise has joint density (if the r.v. A is constant over the observation interval)

$$p(\underline{x}) = \sum_{m=0}^{\infty} p_m \frac{1}{\left[2\pi\ a_m^2\right]^{n/2}(\det R)^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}\ \underline{x}^* R^{-1}\underline{x}/a_m^2\right] \tag{3-6}$$

As will be shown in the next section, for reasonably large n it is not necessary to know the values of U and Γ in order to implement this detector. This fact, as well as the joint density (3-6), illustrates some of the advantages of using a general SIN model whenever appropriate.

Of course, SIN models are not limited to the Middleton model. They cover a large family of smooth unimodal densities that are symmetric about their mean. NonGaussian examples of spherically-invariant distributions include the t and double-exponential [9].

4. DETECTION IN SPHERICALLY-INVARIANT NOISE (SIN)

In this section, (N(t)) will be SIN with representation (AG(t)). (G(t)) is a m.s. continuous zero-mean Gaussian process and A is a strictly-positive discrete random variable independent of (G(t)) and with $EA^2 = 1$. (N(t)) thus has zero mean and covariance the same as the covariance of (G(t)). A takes on the value a_i with probability $p_i > 0$.

Likelihood ratio detection of a known signal in SIN has previously been considered by Yao [18] for a very special case: the threshold on the likelihood ratio test statistic is unity. It has also been considered by Spooner [17] for a specific distribution of the mixing random variable A. A more comprehensive treatment has been given by Picinbono and Vezzosi [13]. Their work, and that of Spooner and Yao, has been for the discrete-time finite-sample-size problem. However, these authors all use or permit continuous mixing r.v.'s A. Our choice of a discrete r.v. for A permits analysis of the continuous time problem without introducing much mathematical complication. It is also sufficient to apply our results to detection in Middleton's Class

A noise.

The first topic to be addressed here is that of absolute continuity and likelihood ratio. Sufficient conditions are contained in the following result.

<u>Prop. 4.1</u>. Suppose that $(Y(t))$ is a stochastic process adapted to $\underline{\sigma}(G)$ v $\underline{\sigma}(V)$, where $(V(t))$ is any process independent of $(N(t))$. Suppose also that $Y(t) = S(t) + N(t)$ a.e. dtdP, where $(S(t))$ is a stochastic process adapted to $\underline{\sigma}(Y)$ and with almost all paths in H_N. Then $\mu_Y << \mu_N$. Moreover, $\mu_{S+aG} << \mu_{aG}$ for all $a>0$, and

$$[d\mu_{S+N}/d\mu_N](x) = \Sigma_i \; I_{C(a_i)}(x) \; [d\mu_{S+a_iG}/d\mu_{a_iG}](x) \qquad (4\text{-}1)$$

a.e. $d\mu_N(x)$. In (4-1), the sum is over all a_i such that $P[A=a_i] > 0$, I_C is the indicator function for the set C in $L_2[0,T]$, and

$$C(a_i) = \{x: \lim_n \frac{1}{n} \sum_{j=1}^{n} \langle x,e_j\rangle^2/\lambda_j = a_i^2\}.$$

Moreover, if $(S(t))$ is any process such that $P[A=a_i] > 0$ implies $\mu_{S+a_iG} << \mu_{a_iG}$, then $\mu_{S+N} << \mu_N$ and $d\mu_{S+N}/d\mu_N$ has the representation (4-1).

<u>Proof</u>. If (Y_t) is adapted to $\underline{\sigma}(G)$ v $\underline{\sigma}(V)$, then (Y_t) is adapted to $\underline{\sigma}(aG)$ v $\underline{\sigma}(V)$ for any constant a. Since $H_N = H_G = H_{aG}$, $\mu_{S+aG} << \mu_{aG}$ for any positive constant a, by Theorem 3 of [5]. Then for any Borel set B of $L_2[0,1]$,

$$\mu_{S+N}(B) = \Sigma_i \; p_i\mu_{S+a_iG}(B) = \Sigma_i \; p_i\int_B [d\mu_{S+a_iG}/d\mu_{a_iG}] \; (x) \; d\mu_{a_iG}(x).$$

Now, $\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \langle x,e_j\rangle^2/\lambda_j = a_i^2$ w.p.1 when $A = a_i$, under both μ_N and μ_{S+N}. To see this for μ_N (noise-only data), one notes that the random variable $\langle x,e_j\rangle/\lambda_j^{\frac{1}{2}}$ has the form $a_i\langle G,e_j\rangle/\lambda_j^{\frac{1}{2}}$ when only noise is present and $A = a_i$. The random variables $\{\langle G,e_j\rangle/\lambda_j^{\frac{1}{2}}: j\geq 1\}$ are i.i.d. N(0,1). Thus, by the law of large numbers, $\frac{1}{n} \sum_{1}^{n} \langle G,e_j\rangle^2/\lambda_j \to 1$ with probability one. When $A=a_i$, $\frac{1}{n} \sum_{j=1}^{n} \langle N,e_j\rangle^2/\lambda_j \to a_i^2$ w.p.1 (μ_N). If signal is present and $A=a_i$, then $\lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \langle S+a_iG,\lambda_j\rangle^2/\lambda_j$ is again equal to a_i^2 w.p.1. To see this, note that since S is in H_N w.p. 1, $\sum_1^\infty \langle S,e_j\rangle^2/\lambda_j$ is finite. This implies that both $\frac{1}{n}\sum_1^n \langle S,e_j\rangle^2/\lambda_j$ and $\frac{1}{n}\sum_1^n \langle S,e_j\rangle\langle a_iG,e_j\rangle/\lambda_j$ converge to zero w.p. 1.

The preceding shows that $\mu_{a_iG}[C(a_i)] = 1$ and that $\mu_{a_jG}[C(a_i)] = 0$ for $i\neq j$. Thus,

$$\mu_{S+N}(B) = \sum_i p_i \int_{B\cap C(a_i)} [d\mu_{S+a_iG}/d\mu_{a_iG}](x)\, d\mu_{a_iG}(x)$$

$$= \sum_i \int_{B\cap C(a_i)} [d\mu_{S+a_iG}/d\mu_{a_iG}](x)\, d\mu_N(x)$$

because $\mu_N(B\cap C(a_i)) = p_i\mu_{a_iG}(B\cap C(a_i))$. (4-1) now follows by the monotone convergence theorem.

□

The expression for the likelihood ratio given in (4-1) partitions the Borel σ-field of $L_2[0,T]$ into two major subsets. These sets are $\bigcup_i C(a_i)$ and its complement, where $C(a_i) = \{x: \lim_n \frac{1}{n}\sum_{j=1}^n \langle x,e_j\rangle^2/\lambda_j = a_i^2\}$. It is noteworthy that the likelihood ratio does not involve the probabilities $P\{A=a_i\}$. These facts show first that the important factor in determining the likelihood ratio for detection in continuous-time SIN is knowledge of the values which can be assumed by the mixing random variable A. However, no penalty is assessed if one includes too many possible values of A. That is, if b is not a possible value of A, then the set $C(b) = \{x: \lim_n \frac{1}{n}\sum_{j=1}^n \langle x,e_j\rangle^2/\lambda_j = b^2\}$ has zero μ_N-probability, and so addition of the term $I_{C(b)}(x)[d\mu_{S+bG}/d\mu_{bG}](x)$ to (4-1) will not affect (with probability one) performance of the test statistic.

A particular application of the above is the situation when the noise can be either Gaussian or spherically-invariant nonGaussian. If $P[A=1] > 0$ holds for the mixing r.v. in the nonGaussian case, then the likelihood ratio (4-1) will still be a likelihood ratio if the noise is in fact Gaussian. If $P[A=1] = 0$ for the nonGaussian SIN model, then one can add the term $I_{C(1)}(x)\dfrac{d\mu_{S+G}}{d\mu_G}(x)$ to the likelihood ratio (4-1). The resulting sum will be a likelihood ratio when either hypothesis is true.

In the remainder of this section, attention will be restricted to the problem of detecting a known signal S in additive SIN. In the case of Gaussian noise, it is well known that the likelihood ratio exists (non-singular detection) if and only if S is in range$(R_N^{\frac{1}{2}})$. The same result holds if the noise is any SIN process.

<u>Prop. 4.2</u>. If S is a fixed element in $\mathcal{L}_2[0,T]$, then either $\mu_{S+N} \perp \mu_N$ or else $\mu_{S+N} << \mu_N$ and $\mu_N << \mu_{S+N}$. Mutual absolute continuity holds if and only if S is in range$(R_N^{\frac{1}{2}})$.

<u>Proof</u>. If the $L_2[0,T]$ equivalence class generated by S is not in range$(R_N^{\frac{1}{2}})$, then $\mu_{S+a_iG} \perp \mu_{a_iG}$ for each a_i, using the known results for the Gaussian case. As shown in the proof of Prop. 4.1 (and well-known [14]), $\mu_{a_iG} \perp \mu_{a_jG}$ for $i \neq j$. Moreover, $\mu_{S+a_iG} \perp \mu_{a_jG}$ for $j \neq i$, since range$(R_N^{\frac{1}{2}})$ = range$\left[(a_i^2 R_N + a_j^2 R_N)^{\frac{1}{2}}\right]$ and S must belong to this latter range space in order to have mutual absolute continuity of μ_{S+a_iG} and μ_{a_jG} [14]. Thus $\mu_{S+a_iG} \perp \mu_{a_jG}$ for all i and j, so that $\mu_{S+a_iG} \perp \sum_{j=1}^{K} p_j \mu_{a_jG}$ for $i=1,\ldots,K$. This gives $\sum_{i=1}^{K} p_i \mu_{S+a_iG} \perp \sum_{j=1}^{K} p_j \mu_{a_jG}$, or $\mu_{S+N} \perp \mu_N$.

Conversely, if S is in range$(R_N^{\frac{1}{2}})$, $\mu_{S+a_iG} \sim \mu_{a_iG}$ for $i=1,\ldots,K$, so that

$$\sum_{i=1}^{K} p_i \mu_{S+a_iG} \sim \sum_{j=1}^{K} p_j \mu_{a_jG}. \qquad \square$$

Performance of the likelihood ratio (4-1) can be computed for the case of a known signal. When the noise is Gaussian, then it is well-known that the performance depends only on

$$d^2 = \|S\|_N^2 = \sum_{n \geq 1} \langle S, e_n \rangle^2 / \lambda_n$$

where $\{\lambda_n, n \geq 1\}$ and $\{e_n, n \geq 1\}$ are the eigenvalues and associated c.o.n. eigenvectors of the noise covariance operator R_N.

Prop. 4-3. Suppose that the signal is a known function S belonging to range($R_N^{\frac{1}{2}}$), and that $P[A=a_i] = p_i > 0$, $i \geq 1$. Then performance of a likelihood ratio test statistic is given by

$$P_{FA} = \sum_i p_i \; P[Z \geq ka_i + d/(2a_i)] \tag{4-2}$$

$$P_D = \sum_i p_i \; P[Z \geq ka_i - d/(2a_i)] \tag{4-3}$$

where Z is distributed N(0,1) and k is a constant whose value is determined by the desired value of P_{FA}.

Proof: For k > 0,

$$P_{FA} = \mu_N\{x: (d\mu_{S+N}/d\mu_N)(x) \geq e^{kd}\} = \mu_N\{x: \sum_i I_{C(a_i)}(x)\ell_i(x) \geq kd\}$$

where

$$\ell_i(x) = \log[(d\mu_{S+a_iG}/d\mu_{a_iG})(x)] = \frac{1}{a_i^2}\Big[\sum_n \langle x,e_n\rangle\langle S,e_n\rangle/\lambda_n - d^2/2\Big].$$

Thus

$$P_{FA} = \sum_i p_i \; \mu_{a_iG}\{x: \ell_i(x) \geq kd\}$$

$$= \sum_i p_i\mu_G\{x: a_i\Sigma_n \langle x,e_n\rangle\langle S,e_n\rangle/\lambda_n \geq kda_i^2 + d^2/2\}.$$

Since the random variable ℓ defined by

$$\ell(x) = \Sigma_n \langle x,e_n\rangle\langle S,e_n\rangle/\lambda_n$$

is Gaussian with respect to μ_G, and has mean zero and variance d^2,

$$P_{FA} = \Sigma_i \; p_i \; P[Z \geq ka_i + d/(2a_i)].$$

P_D is calculated in the same way.

□

As can be seen from (4-2) and (4-3), detection performance depends on d and also on the distribution of the mixing random variable A.

The likelihood ratio as given in (4-1) requires prior knowledge of the values (a_i) that can be assumed by the mixing random variable A. However, this prior knowledge is not necessary in order to implement this detector.

Prop. 4-4. For a known signal S, a likelihood ratio test is to decide signal present if and only if

$$[\ell(x)/\hat{A}^2(x) - d^2/(2\hat{A}^2(x))] \geq k \tag{4-4}$$

where k is determined from (4-2) and (4-3), and

$$\hat{A}^2(x) = \lim_n \frac{1}{n} \sum_{k=1}^{n} \langle x, e_k \rangle^2 / \lambda_k .$$

Proof. If $x \in C(a_i)$, then $\hat{A}^2(x) = a_i^2$ with probability one under μ_N or μ_{S+N}, as shown in the proof of Prop. 4-1. The result then follows directly from the expression (4-1), or by examining the proof of Prop. 4-3.

□

A likelihood ratio detector can thus be implemented without any prior knowledge of the distribution of the mixing random variable A, provided the noise is in fact SIN. However, as can be seen from the expressions (4-2) and (4-3) for P_{FA} and P_D, likelihood ratio detection performance depends on the complete distribution of A. This means that it is not possible to set a threshold for a specified P_{FA} unless one has complete knowledge of the distribution of A.

This leads one to consider the problem of CFAP (constant false alarm probability) detection, which has been treated for many years by designers of active sonar detection systems. In this traditional context, the detection problem is that of detecting a signal in Gaussian noise which is known except for a scale factor. It is desired to have the same probability of false alarm for any value of the scale factor. The scale factor has usually been treated as an unknown parameter, rather than as a random variable.

A CFAP detector can be obtained for the SIN detection problem by using the following decision procedure:

decide signal present if and only if

$$\ell(x)/\hat{A}(x) \geq kd, \tag{4-5}$$

when $\hat{A}(x) = [\hat{A}^2(x)]^{\frac{1}{2}}$. When noise only is present, $\ell(x)/\hat{A}(x) = \ell(x)/a_i$ with probability one when $x = a_i G$. Since then $\ell(x) = a_i \ell(G)$, one has that $\ell(x)/\hat{A}(x)$ is Gaussian with zero mean and variance d^2 and

$$P_{FA} = P[Z \geq k]. \tag{4-6}$$

When this detection algorithm is used, and $x = S + a_i G$, then $\ell(x)/a_i$ is Gaussian with mean d^2/a_i and variance d^2, so that

$$P_D = \Sigma\, p_i\, P[Z \geq k - d/a_i]. \tag{4-7}$$

The difference in performance between the optimum detector (4-4) and the CFAP detector (4-5) will depend on the distribution of A. Figure 1 shows an example using a distribution for A obtained from analyzing under-ice sonar data. The curves show performance for the optimum detector (4-4), the CFAP detector (4-5), and the matched filter (A=1 w.p. 1) when the noise is SIN with the given distribution for A. Also shown is the performance that one would obtain using the matched filter if the noise were truly Gaussian. The difference in performance of the matched filter and the likelihood ratio illustrates the significant performance loss that can occur if the noise is mistakenly assumed to be Gaussian. This and the similarity in performance of the CFAP detector and the likelihood ratio illustrate the wisdom of using a CFAP detector if there is a possibility that the noise is SIN with unknown distribution.

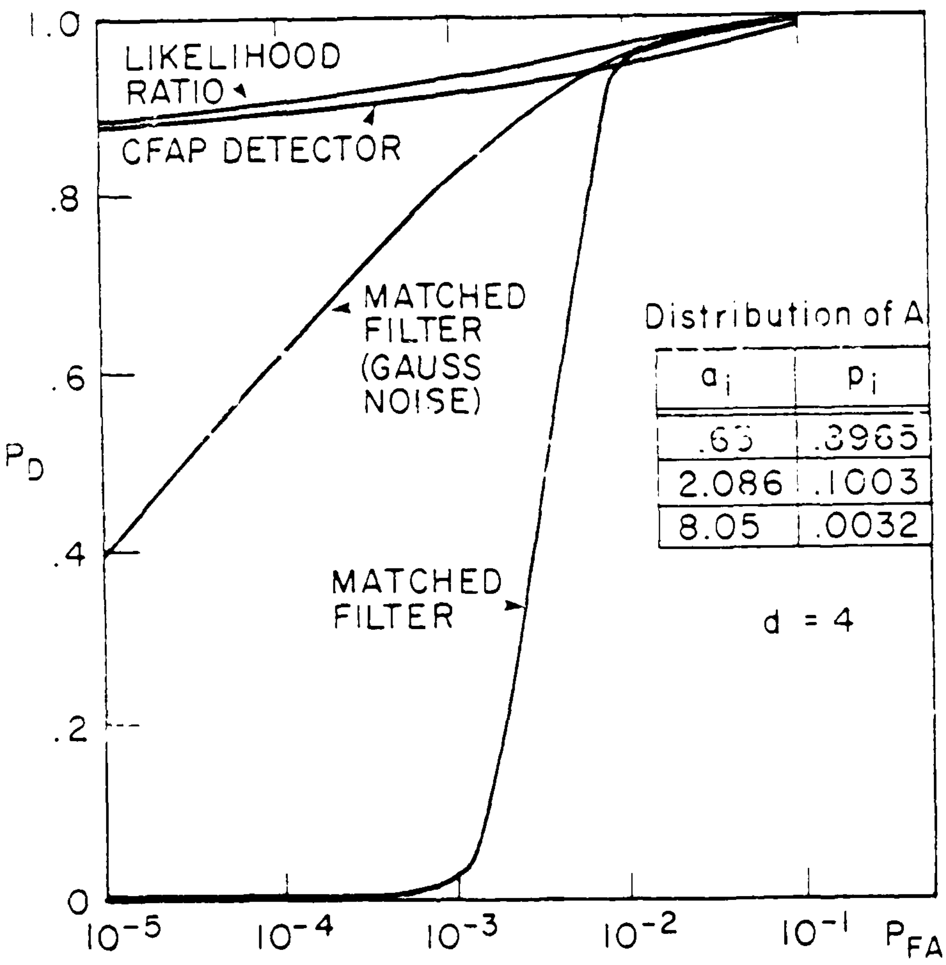


Figure 1. Detection of Known Signal in SIN

For the discrete-time finite-sample detection problem, with observation $\underline{x}$ in E^n, the likelihood ratio $dP^n_{\underline{S}+\underline{N}}/dP^n_{\underline{N}}$ is easily seen to be

$$\left[dP^n_{\underline{S}+\underline{N}}/dP^n_{\underline{N}}\right](\underline{x}) = \sum_{i=1}^{K} p_i dP^n_{\underline{S}+a_i\underline{G}}(\underline{x}) / \sum_{j=1}^{K} p_j dP^n_{a_j\underline{G}}(\underline{x}) \quad (4\text{-}8)$$

where dP^n is the multivariate density function for the probability P^n on E^n. In contrast to the continuous-time case (4-1), the probabilities $p_i = P[A=a_i]$ appear in (4-8). Moreover, the likelihood ratio (4-1) produces a non-zero value (with probability one) only if the observation involves a_iG for a_i one of the terms included in (4-1); otherwise, the value of the likelihood ratio is zero. This is not true in the discrete-time case of (4-8). In the case of a known signal S, the performance (P_{FA} and P_D) of (4-1) depends only on the distribution of A and on $d^2 = \|R_N^{-\frac{1}{2}}S\|^2$. The discrete-time detector's performance improves as the sample size n increases, with d fixed.

These differences can all be understood by examining the form of (4-8) as the sample size increases.

Suppose that noise only is present, and that the mixing r.v. A takes on the value a_i, so that the received waveform $\underline{x}$ is $a_i\underline{G}$ with $\underline{G}$ multivariate Gaussian, zero mean, non-singular covariance matrix $\underline{R}$. Let

$$Y_j^n = \log p_j \, dP^n_{a_j\underline{G}}(\underline{x}) = \log p_j - n \log a_j - a_i^2\|\underline{R}^{-\frac{1}{2}}\underline{G}\|^2/(2a_j^2) + C$$

$$= \log p_j - n \log a_j - a_i^2 \sum_{k=1}^{n} V_k^2/(2a_j^2) + C$$

where C is a constant and (V_k) is i.i.d. N(0,1). It will be shown that $Y_j^n - Y_i^n \to -\infty$ w.p. 1. Thus,

$$Y_j^n - Y_i^n = \log [p_j/p_i] - n \log [a_j/a_i] + \sum_{k=1}^{n} V_k^2\left[\tfrac{1}{2}(1 - a_i^2/a_j^2)\right]$$

$$= \gamma - n\rho u_{i,j} + \beta\chi_n^2 u_{i,j}$$

where χ_n^2 is chi-square with n degrees of freedom, $\beta = \beta(i,j) = |(1-a_i^2/a_j^2)/2|$, $\rho = \rho(i,j) = |\log (a_i/a_j)|$, $\gamma = \gamma(i,j) = \log p_j/p_i$, and $u_{i,j} = 1$ if $a_i \leq a_j$, $u_{i,j} = -1$ if $a_i > a_j$.

For $a_i < a_j$, $m > 0$, $\underline{x} = a_i\underline{G}$, one has $P[Y_j^n - Y_i^n \geq -m]$ $= P[p_j dP_{a_jG}(\underline{x}) \geq e^{-m} p_i dP^n_{a_i\underline{G}}(\underline{x})] = P[\chi_n^2 \geq -(m+\gamma)/\beta + n\rho/\beta]$.

Since $\chi_n^2/n \to 1$ w.p. 1 as $n \to \infty$, and $\rho/\beta > 1$ for $a_i < a_j$,

$P[\lim_n (\chi_n^2/n + (m+\gamma)/\beta n - \rho/\beta) \geq 0] = 0$, for any fixed $m > 0$.

If $a_i > a_j$, then $P[Y_j^n - Y_i^n \geq -m] = P[\chi_n^2 \leq (m+\gamma)/\beta + n\rho/\beta]$. In this case, $\rho/\beta < 1$, so $P[\lim_n(\chi_n^2/n - (m+\gamma)/\beta n - \rho/\beta) \leq 0] = 0$.

Using these expressions, one can obtain the value of n required for a specified approximation, once the distribution of A is known. The value of n required for $P[p_j dP_{a_j G}(\underline{x}) \geq e^{-m} p_i dP_{a_i G}(\underline{x})] \leq \alpha$ when $\underline{x} = a_i \underline{G}$ is determined from

$$
\begin{aligned}
&P[\chi_n^2 \geq -(m+\gamma)/\beta + n\rho/\beta] \leq \alpha \qquad \text{if } a_i < a_j \\
&P[\chi_n^2 \leq (m+\gamma)/\beta + n\rho/\beta] \leq \alpha \qquad \text{if } a_i > a_j.
\end{aligned}
\tag{4-9}
$$

This procedure can be repeated for the numerator of (4-8). A conservative result, which satisfies $P[p_j dP_{\underline{S}+a_j\underline{G}}(\underline{x}) \geq e^{-m} p_i dP_{\underline{S}+a_i\underline{G}}(\underline{x})] \leq \alpha$, is to require $n \geq \Delta n + n_1$, where n_1 is the value for α given by (4-9) and Δn satisfies

$$
\Delta n \rho(i,j)/\beta(i,j) \geq dZ_\alpha/a_i + d^2 u(i,j)/(2a_i^{\,2}). \tag{4-10}
$$

Applying these results to the distribution used to obtain Figure 1, with $e^{-m} = .01$ ($m = 4.6$) and $\alpha = 10^{-3}$, the required sample size for the denominator of (4-8) is $n \geq 21$, using (4-9). The value of Δn given by (4-10) is 9, so that an adequate sample size is 30. This rather small required sample size is a result of the wide separation between the three values of A, and (to a much lesser extent) the corresponding large differences in their probabilities. It can be seen from (4-9) and (4-10) that distributions for A which have more similar values will require larger sample sizes in order to achieve the above bounds, with a requirement of $n \to \infty$ as the minimum distance between A values converges to zero.

The gist of this analysis is that the likelihood ratio (4-8) converges to $dP_{\underline{S}+a_i\underline{G}}/dP_{a_i\underline{G}}$ when $\underline{x} = a_i\underline{G}$, as the sample size increases; the rate of convergence depends on the distance from a_i to the nearest value of A not equal to a_i; and the probability ratios can be ignored for large sample size. When the sample size n is sufficiently large to assume equality in this approximation, then the performance of the discrete-time detector is the same as that of the continuous-time detector so long as the value of d^2 is fixed.

For n sufficiently large, then, one can mimic the log-likelihood ratio for the continuous-time case:

$$[dP^n_{\underline{S}+\underline{N}}/dP^n_{\underline{N}}](\underline{x}) \cong \ell^n(x)/\hat{A}^2_n(\underline{x}) - d^2/2\hat{A}^2_n(\underline{x}) \tag{4-11}$$

where $\hat{A}^2_n(x) = n^{-1} \sum_{k=1}^{n} (\underline{x}*\underline{e}_k)^2/\lambda_k$,

$$\ell^n(x) = \sum_{k=1}^{n} (\underline{x}*\underline{e}_k)(\underline{S}*\underline{e}_k)/\lambda_k,$$

$\underline{R}\ \underline{e}_k = \lambda_k \underline{e}_k$, $k=1,\ldots,n$, and

$\{\underline{e}_k,\ k\geq 1\}$ is a complete orthonormal set in E^n.

The detector (4-11) has previously been given as an approximate likelihood ratio for large n by Picinbono and Vezzosi [13]. The above analysis indicates why this is so, and indicates how one can determine how large n must be in order to use the approximation.

The CFAP detector now becomes

$$\Lambda_{CFAP}(\underline{x}) = \ell^n(\underline{x})/\hat{A}_n(\underline{x}) \tag{4-12}$$

where $\hat{A}_n(x) = [\hat{A}^2_n(\underline{x})]^{\frac{1}{2}}$.

For small n, one may wish to consider the CFAP detector given by

$$\bar{\Lambda}_{CFAP}(x) = \ell^n(\underline{x})/\sigma_n(\underline{x}) \tag{4-13}$$

where $\sigma^2_n(\underline{x}) = (n-1)^{-1} \sum_{i=1}^{n} \left[\frac{(\underline{x}*\underline{e}_i)}{\lambda_i^{\frac{1}{2}}} - \frac{1}{n}\sum_{j=1}^{n} \frac{(\underline{x}*\underline{e}_j)}{\lambda_j^{\frac{1}{2}}}\right]^2$. Since the random variables $\{\underline{x}*\underline{e}_i/\sqrt{\lambda_i}:\ i=1,\ldots,n\}$ are i.i.d. $N(0,a^2)$ when A=a, and $\ell^n(\underline{x})$ is Gaussian, one may wish to assume that $\underline{S}^*_i\underline{e}_i/\sqrt{\lambda_i}$ = constant = $d/\sqrt{n}$ for $i=1,\ldots,n$. The test statistic divided by d then has a t distribution with n-1 degrees of freedom when noise only is present; this fact can be used to calculate P_{FA}. Under this same assumption, one can also obtain an expression for P_D for this detector if the distribution of A is known, using the fact that $(n-1)\sigma_n{}^2$ is chi-square distributed with n - 1 degrees of freedom. One could use these considerations to determine a worst-case value of P_D if the distribution of A is known to belong to a specified family, while maintaining a desired P_{FA}.

The problem of detecting a signal in Gaussian noise having unknown scale factor is familiar in active sonar. A detailed treatment of CFAP detection for this problem has been given by Grieve [7], who obtained CFAP optimality properties for (4-12).

5. APPLICATIONS TO DETECTION IN IMPULSIVE NOISE ENVIRONMENTS

As previously noted, Middleton's Class A univariate model is a special case of SIN. Detection in such noise has been analysed by several approaches. In [16], Spaulding and Middleton assume independent sampling and then develop bounds on likelihood ratio performance for communicating with a known signal over a channel in the presence of Class A noise. The Middleton model has also often been approximated by using only the first two or three terms: "Gauss-Gauss" or "Gauss-Gauss-Gauss" noise.

If the mixing random variable of the Middleton model remains constant over the observation interval, then the results given above can be used to provide detection results. It is not necessary to have independent sampling, but only to know the covariance matrix of the noise and the parameters U and Γ. For detection of a known signal, the preceding results can be used in several ways. They provide upper bounds on detection performance by giving the continuous-time detection performance. Secondly, they provide a method for obtaining exact detection performance for the discrete-time finite-sample-size detectors, and provide a means of calculating required sample size in order to simplify the detector structure. Thirdly, they can be used to obtain discrete-time CFAP detectors, as well as upper bounds on the performance of such detectors. Finally, the fact that the likelihood ratio detector can be implemented without knowing the distribution of the mixing r.v. A, once the sample size n is reasonably large (4-11), can provide a significant reduction of the complexity of the implementation. One need only adjust the threshold as a function of the parameters U and Γ, while the operation on the data is unchanged. Even this adjustment is not necessary if one is willing to use a CFAP detector.

The imbedding of the Middleton Class A model within the general SIN model thus provides a number of useful results. One may note the importance of the continuous-time model, which is often disregarded on the grounds that it is not relevant to practical signal detection. In the present case, the continuous-time model provides useful upper bounds on detection performance for both the likelihood ratio detector and the CFAP detectors. It also provides one with a practically-useful implementation and simplification of the apparently extremely-complicated discrete-time likelihood ratio detectors, and a rationale for making this simplification. The notion of orthogonal measures is central to these results.

6. EXTENSIONS TO THE SIN MODEL

The SIN model is not realistic for many situations, such as observation periods where the mixing r.v. A cannot be expected to take on a constant value. A more reasonable model in such situations would be generalized spherically-invariant noise, of the form $N(t) = A(t)G(t)$, where now $(A(t))$ is a stochastic process independent of the Gaussian process $(G(t))$. This reduces to a SIN model in the univariate case. Some work has previously been done for such a model [15], but general results are so far not available.

ACKNOWLEDGEMENTS

This research was supported by ONR contracts N00014-81-K-0373, N00014-84-C-0212, and N00014-86-K-0039.

The authors thank H. Cherifi for helpful discussions and assistance.

REFERENCES

1. C.R. Baker, Optimum quadratic detection of a random vector in Gaussian noise. IEEE Trans. Commun., COM-14, 802-805 (1966).

2. C.R. Baker, On the deflection of a quadratic-linear test statistic, IEEE Trans. Information Theory, IT-15, 16-21 (1969).

3. C.R. Baker, On equivalence of probability measures, Annals of Probability, 1, 690-698 (1973).

4. C.R. Baker, Absolute continuity of measures on infinite-dimensional linear spaces, Encyclopedia of Statistical Sciences, 1, 3-11, Wiley, New York (1982).

5. C.R. Baker and A.F. Gualtierotti, Discrimination with respect to a Gaussian process, Probability Theory and Related Fields, 71, 159-182 (1986).

6. R. Dwyer, A technique for improving detection and estimation of signals contaminated by underice noise, J. Acoustical Soc. America, 74, 124-130 (1983).

7. P.G. Grieve, The optimum constant false alarm probability detector for relatively coherent multichannel signals in Gaussian noise of unknown power, IEEE Trans. Information Theory, IT-23,

8. S. Geman, An application of the method of sieves: functional estimator for the drift of a diffusion, Reports in Pattern Analysis, 92, Div. Applied Math., Brown Univ. (1980).

9. J. Keilson and F.W. Steutel, Mixtures of distributions, moment inequalities, and measures of exponentiality and normality, Annals of Probability, 2, 112-130 (1974).

10. T. Hida, Canonical representations of Gaussian processes and their applications, Mem. Coll. Science, Univ. Kyoto, 33A, 109-155 (1960).

11. R.S. Liptser and A.N. Shiryayev, Statistics of Random Processes I. General Theory, Springer-Verlag, New York (1977).

12. D. Middleton, Statistical-physical models of electromagnetic interference, IEEE Trans. on Electromagn. Compat., EMC-19, 106-127 (1977).

13. B. Picinbono and G. Vezzosi, Detection d'un signal certain dans un bruit non stationnaire et non Gaussien, Annales des Telecommunications, 25, 433-439 (1970).

14. C.R. Rao and V.S. Varadarajan, Discrimination of Gaussian processes, Sankhya, 25A, 303-330 (1963).

15. A.D. Spaulding, Locally optimum and suboptimum detector performance in a non-Gaussian interference environment, IEEE Trans. Commun., COM-33, 509-517 (1985).

16. A.D. Spaulding and D. Middleton, Optimum reception in an impulsive interference environment - Part I: Coherent detection; Part II: Incoherent detection, IEEE Trans. Commun., COM-25, 910-934 (1977).

17. R.L. Spooner, On the detection of a known signal in a non-Gaussian noise process, J. Acoustical Soc. America, 44, 141-147 (1968).

18. K. Yao, A representation theorem and its application to spherically-invariant random processes, IEEE Trans. Information Theory, IT-19, 600-608 (1973).

END
DTIC
8-86