Some Results on NonGaussian Signal Detection

C.R. Baker*, M.R. Frey*, and A.F. Gualtierotti**

*Department of Statistics
University of North Carolina
Chapel Hill, North Carolina, 27599, USA

**IDHEAP
University of Lausanne
CH 1015 Lausanne, Switzerland

Summary

Recent results on detection of random nonGaussian signals in additive Gaussian noise are discussed. A discrete-time approximation to a likelihood ratio detection algorithm is discussed, and preliminary results are presented for a computational evaluation of the performance for this approximation.

Résumé

D'abord nous présentons un résumé des résultats récents sur la détection des signaux aléatories non-gaussien dans le bruit gaussien additif. Ensuite, après une discussion sur une approximation à temps discret à une algorithme de détection par le rapport des vraisemblances, nous présentons les résultats préliminaires d'une évaluation numérique de l'efficacité de cette approximation.

Introduction

Detection of nonGaussian signals is a problem that arises frequently in underwater acoustics. See, for example, the discussion in [1]. The noise in such problems is typically additive and frequently Gaussian. In some applications, it is nonGaussian, but with properties similar in several respects to those of a Gaussian process.

Detection of Gaussian signals in Gaussian noise is a subject on which there is a voluminous literature, and likelihood ratios for such problems are by now well-understood [2]. However, detection of non-Gaussian signals in Gaussian noise is a quite different matter. Until very recently, the results on nonsingular detection and likelihood ratios for such problems were limited to the case where the noise is a Wiener process [2], [3].

For detection of nonGaussian signals in Gaussian noise when only moment information is used, one can assume a convenient form for the detector, and choose the optimum detector from this class using a signal-to-noise-ratio criterion. The most commonly used class of detection algorithms based on moment criteria is the quadratic-plus-linear detector with optimization done according to the deflection criterion [4], [5], [6]. Thus, the test statistic for an observation x is $\Lambda(x) = \langle x, Wx\rangle + \langle x, h\rangle$, where $\langle \cdot,\cdot \rangle$ denotes the inner product in (as appropriate) $L_2[0,1]$ or n-dimensional Euclidean space E^n. The operator W and vector h are selected to maximize $D(W,h) \equiv \dfrac{[E_{S+N}\Lambda(x) - E_N\Lambda(x)]^2}{E_N\Lambda^2(x) - (E_N\Lambda(x))^2}$.

$E_N(\cdot)$ denotes expectation w.r.t. the noise, $E_{S+N}(\cdot)$ w.r.t. signal-plus-noise. The optimum (W,h) is then given by [5], [6]: $W = R_{S+N}^{-1}R_N R_{S+N}^{-1}$, $h = R_N^{-1}m$, where R_{S+N} and R_N denote the covariance operators of the signal-plus-noise and noise processes, respectively, and m is the mean of the signal process.

Since the information required to implement the deflection criterion quadratic-linear detector is relatively easy to obtain, and the algorithm is well-known and widely-used, it is an appropriate benchmark for evaluation of the performance of any algorithm for detection of nonGaussian signals in Gaussian noise.

Of course, the desired detection algorithm for any detection problem (using various criteria) is well-known to be a monotone function of the likelihood ratio. The difficulty is that such algorithms require more information than is usually available. In particular, in sonar applications one will typically not know the family of finite-dimensional distributions for the signal-plus-noise process.

However, there are some models of this detection problem for which the implementation of the likelihood ratio does not require advance knowledge of the statistics of the signal-plus-noise process. Such a detector is described in [7], based on theoretical results obtained in [8]. It assumes that the signal-plus-noise is a filtered diffusion process. This algorithm will be briefly discussed in this paper, and some preliminary computational results on performance will be presented, including comparison with performance of the optimum quadratic-linear detector based on the deflection criterion.

The theoretical results are based on the spectral representation of purely nondeterministic second-order processes, as developed by Cramér [9] and Hida [10]. In particular, if (N_t) is any m.s. continuous purely nondeterministic stochastic process on [0,1], then (N_t) has the proper canonical representation [10]

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

where the multiplicity M can be infinite, $\{B_i,\ i \le M\}$ is a family of mutually orthogonal orthogonal-increment processes with (non-decreasing) variances $\{\beta_i,\ i \le M\}$, and the non-random functions $\{F_i,\ i \le M\}$ satisfy $\sum_{i=1}^{M}\int_0^1\int_0^t F_i^2(t,s)d\beta_i(s)dt < \infty$.

In the case where (N_t) is Gaussian, the processes $\{B_i,\ i \le M\}$ are mutually-independent independent-increment Gaussian processes, which can be assumed path-continuous.

The equality in the above expression for (N_t) is an equality in the mean-square sense for each t. However, by assuming separability, one obtains path equality in the almost sure sense.

Nonsingular Detection and Likelihood Ratios

A basic question to be considered for any detection problem is that of model validity. Since physical detection problems are typically nonsingular (non-zero probability of error), nonsingularity is an important consideration, especially in the case of continuous-time processes. In fact, since one is typically seeking to determine a likelihood ratio, the model should first by analyzed to determine whether or not a likelihood ratio exists; this is one aspect of nonsingular detection. The version of nonsingular detection required in order for the likelihood ratio to exist is of the following form: for any detection algorithm giving zero probability of false alarm, one must also have zero probability of detection. This is absolute continuity of the signal-plus-noise probability measure μ_{S+N} with respect to the noise measure μ_N ($\mu_{S+N} << \mu_N$). These probabilities are induced on the space of sample functions by the stochastic processes (S_t+N_t) and (N_t); they can be taken on either $\mathbb{R}^{[0,1]}$, the space of all real-valued functions on $[0,1]$, or on $L_2[0,1]$. Sufficient conditions and necessary conditions have been obtained for each case [8], along with expressions for the likelihood ratio. These conditions all involve a signal-plus-noise representation of the form

$$S_t + N_t = \sum_{i=1}^{M} \int_0^t F_i(t,s)dX_i(s)$$

where $X_i(t) = \int_0^t Q_i(s)d\beta_i(s) + B_i(t)$, and the vector stochastic process $(\underline{Q}(t))$ satisfies certain measurability conditions and, moreover, $\Sigma_{i=1}^{M}\int_0^t Q_i^2(s)d\beta_i(s) < \infty$ with probability one. The last condition is equivalent to the signal process having almost all paths in the reproducing kernel Hilbert space of the noise.

A Discrete-Time Detection Algorithm

The development to be described here is given in more detail in [7]. First, it is assumed that the original continuous-time noise process has multiplicity one and that (X_t) is a diffusion with respect to a Wiener process (W_t):

$$X(t) = \int_0^t \sigma[X_s]ds + W(t),$$

See [7] for a discussion of these assumptions. Of course, this does not reduce the class of noise processes, since any discrete-parameter zero-mean Gaussian noise process can be obtained as a causal linear operation on white Gaussian noise. Thus, one can write the noise vector $\underline{N}$ as $\underline{N} = \underline{F}\ \underline{\delta W}$, where $\underline{\delta W}$ consists of normalized increments of the Wiener process. The discretized form of the signal-plus-noise process under these same assumptions is then $\underline{S} + \underline{N} = \underline{F}\ \underline{\delta X}$, where for a sampling interval Δ,

$$X(i\Delta) = \Delta \sum_{j=1}^{i-1} \sigma[X(j\Delta)] + W(i\Delta),$$

$$(\underline{\delta X})_i = X(i\Delta) - X([i-1]\Delta), \quad X(0) = 0.$$

Moreover, $R_N = FF^*$, where F is a lower-triangular matrix. Let L be the summation matrix, $(LX)_i = \Sigma_{j=1}^{i} X_j$ ($L_{ij} = 1$ for $i \leq j$; $L_{ij} = 0$ for $i > j$). The discrete-time detection algorithm then can be expressed as follows. For an observation vector in $\mathbb{R}^n$, an approximation to the log-likelihood ratio is given by the test statistic

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

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

$$\begin{aligned} \tau^{n+1}(\underline{x}) = \tau^n(\underline{x}) &+ \sigma[(\underline{L}\ \underline{F}^{-1}\underline{x})_n])[(\underline{F}^{-1}\underline{x})_{n+1} \\ &- (\Delta/2)\ \sigma^2[(\underline{L}\ \underline{F}^{-1}\underline{x})_n]. \end{aligned}$$

Implementation and calculation of τ require the following operations. First, the function σ must be known and programmed. Given the value of $\tau^n(\underline{x}^n)$ and the observation $\underline{x}^n = (x_1,\ldots,x_n)$, one stores $\tau^n(\underline{x}^n)$, $\underline{x}^n$, $\sigma[(\underline{L}\ \underline{F}^{-1}\underline{x})_n]$, 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 $\tau^{n+1}(\underline{x}^{n+1})$, $\tau^{n+1}(\underline{x}^{n+1}) = \tau^n(\underline{x}^n) + \sigma[\Sigma_1^n b_i]b_{n+1} - (\Delta/2)\sigma^2[\Sigma_1^n b_i]$. Other approximations can be given, including one [1] based on the eigenvalues and eigenvectors (for an observation in E^n) of the n×n matrix with elements $\{R_N(i,j): i,j \leq n\}$. The form given here has the advantage of computational efficiency, including recursive properties.

Of course even under the above assumptions, one cannot assert that the discretized version of the continuous-time likelihood ratio is the actual discrete-time likelihood ratio. For the Gaussian case (σ linear), optimality can be shown [7], but for a general σ one can only assert asymptotic optimality. The designation of this algorithm as a "likelihood ratio" should be understood to mean that it is an approximation to the log-likelihood ratio.

In many applications, one will know R_N (thus the matrix F), while σ will be unknown. One procedure for implementing this detector is then as follows. Given an observed discrete-time sample function, $\underline{y}$, obtain (an assumed) $\underline{\delta X}$ by $\underline{\delta X} = F^{-1}\underline{y}$. Using $X(0) = 0$, form $\underline{X}$ and use this sample vector to estimate σ. Substitute $\hat{\sigma}$ into the likelihood ratio and then evaluate the likelihood ratio using the original observation $\underline{y}$. Compare the likelihood ratio with a threshhold, the threshold being determined by the desired value of the probability of false alarm, using the estimated value of σ and the known R_N.

Computational development of such an algorithm, based on single-sample-path information, is now underway. Some preliminary results are given in the next section. Since determination of the matrix F is not expected to be a major problem, the results were obtained for F = I. An algorithm for estimating σ was developed, based upon assuming that the drift function σ can be approximated by a low-order polynomial. The simulation is described in the next section.

Simulation Results

A series of simulations were conducted. In each case an ensemble of sampled noise paths and an ensemble of sampled signal-plus-noise paths were generated and processed by the likelihood ratio detector. The same noise and signal-plus-noise ensembles were also passed through the deflection criterion detector to provide a relative measure of performance. It should be noted that in obtaining these preliminary results, the diffusion drift function was estimated from an observed diffusion path. In practice the diffusion drift will be estimated from the observation vector, which may be only noise.

The Wiener noise paths and diffusion signal paths considered here are indexed by an interval of length T and sampled at a rate R. Each path is then represented by RT samples. The size of the signal path ensemble and the size of the noise path ensemble are each N. Thus, the diffusion signal-plus-noise sampled path ensemble $\{X_{n,t},\ n = 1,\ldots,RT\}$ and the Wiener noise sampled path ensemble $\{W_{n,t},\ n = 1,\ldots,N,\ t = 1,\ldots,RT\}$ are each represented by NRT sample points. $W_{n,0} = 0$ and $X_{n,0} = 0$ for each n. For each simulation, noise path samples were generated recursively using $W_{n,t} = W_{n,t-1} + (\rho/\sqrt{R})N(0,1)$, where N(0,1) represents independent realizations of a standard normal random variable and ρ is the variance parameter of the Wiener process. Samples representing the diffusion path ensemble were generated using $X_{n,t} = X_{n,t=1} + (1/R)\sigma(X_{n,t-1}) + (\rho/\sqrt{R})N(0,1)$, where σ is the drift of the diffusion. The realizations of the standard normal random variable used in the formation of the diffusion samples were generated independently of the set of samples in the noise path ensemble.

To calculate the deflection detector output, the known inverse covariance matrix for the Wiener process was used. For the diffusion signal-plus-noise process, the covariance matrix was estimated from the ensemble of sampled paths:

$$R_X = \frac{1}{N}\sum_{n=1}^{N} \underline{X}_n \underline{X}_n^T - \overline{X}\,\overline{X}^T$$

where $\underline{X}_n$ is the column vector of samples of the n^{th} path in the signal-plus-noise ensemble and $\overline{X}$ is the sample mean $\overline{X} = \frac{1}{N}\sum_{n=1}^{N}\underline{X}_n$.

Polynomial regression provides one approach to the estimation of the drift function of the diffusion. The diffusion (X_t), $t \in [0,T]$ is given by the equation:

$$X_t(\omega) = \int_0^t \sigma(X_s(\omega))ds + W_t(\omega)$$

where (W_t) is a Wiener process and σ is the unknown drift function. An increment of the diffusion can be expressed as $X_{t+\Delta} - X_t = \int_t^{t+\Delta}\sigma(X_s)ds + (W_{t+\Delta} - W_t)$. Define $\delta X_t = X_{t+\Delta} - X_t$, $\delta W_t = W_{t+\Delta} - W_t$, and choose Δ small enough that $\int_t^{t+\Delta}\sigma(X_s)ds \doteq \sigma(X_t)\Delta$. Suppose also that Δ divides T; then $\delta X_t \doteq \sigma(X_t)\Delta + \delta W_t$. Suppose σ is a polynomial, $\sigma(x) = \sigma_0 + \sigma_1 x + \ldots + \sigma_p x^p$ and let $\underline{\sigma}$ be the p+1-dimensional vector of coefficients of $\sigma(x)$. Then one can write $\underline{Y} \doteq X\underline{\sigma} + \underline{N}$ where

$$Y_k = \frac{1}{\delta}\,\delta X_{k\Delta},\quad N_k = \frac{1}{\Delta}\,\delta W_{k\Delta},\quad X_{ki} = X_{k\Delta}^{i-1}$$

where $1 \le k \le T/\Delta - 1$ and $1 < i \le p+1$. $\underline{N}$ is a vector of i.i.d. Gaussian r.v.'s with zero mean and standard deviation of $\rho/\sqrt{\Delta}$. The coefficients of σ can be estimated by solving the regression $\underline{Y} = X\underline{\sigma} + \underline{N}$.

The order p of the polynomial σ is unknown. It is estimated concurrently with $\underline{\sigma}$ by solving the regression problem $\underline{Y} = X\underline{\sigma} + \underline{N}$ first for p = 0 and then for successively higher values of p until increases in the coefficient of determination

$$R^2 = \sum_{k=1}^{n}(\hat{Y}_k - \overline{Y})^2 / \sum_{k=1}^{n}(Y_k - \overline{Y})^2$$

are no longer significant. This same procedure can also be applied in cases in which the drift function is not a polynomial, although in many such cases it may fail to give reasonable estimates and could be improved upon in a variety of ways. However, this estimation procedure does serve to demonstrate the potential of the likelihood ratio detector and is the procedure used in obtaining the simulation results shown in Figures 1, 2, and 3. The results shown in these figures were obtained using N = 5000, R = 100, and T = 1. The variance parameter ρ of (W_t) is unity. The false alarm probabilities are 90% confidence-bounded.

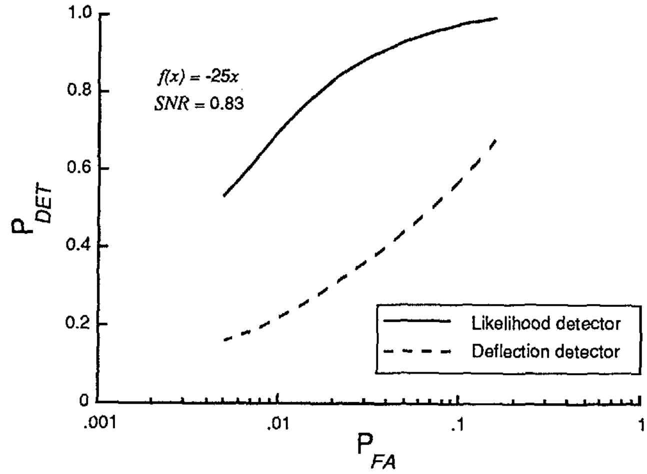

Figure 1. Detector Performance, Linear Drift.

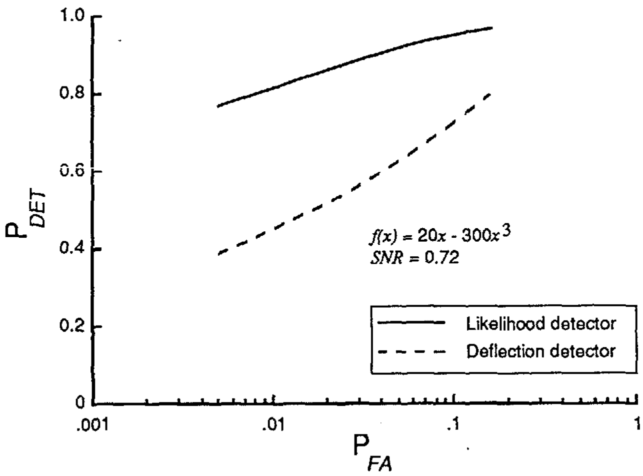

Figure 2. Detector Performance, Cubic Drift.

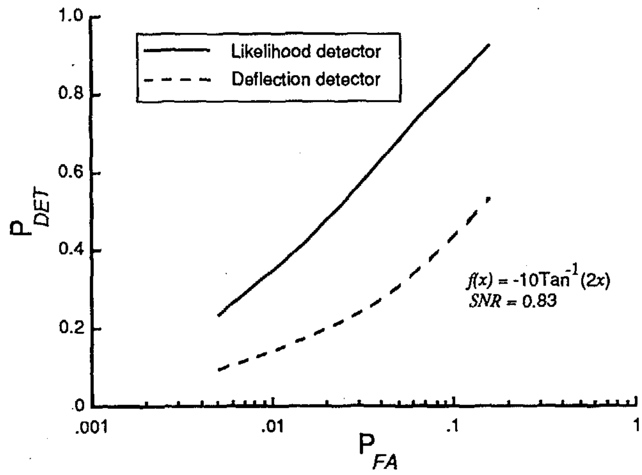

Figure 3. Detector Performance, Arctangent Drift.

Acknowledgement

This paper is based on research supported by ONR Contract N00014-86-K-0039, ONR Grant N00014-89-J-1175, and ARO Grant DAAL03-86-G-0022.

References

[1] C.R. Baker and A.F. Gualtierotti, Detection of nonGaussian signals in Gaussian noise, *Statistical Signal Processing*, E.J. Wegman and J.G. Smith, editors, 107-116, Marcel Dekker, Inc., New York (1984).

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

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

[4] J.L. Lawson and G.E. Uhlenbeck, *Threshold Signals*, McGraw-Hill, New York (1950).

[5] C.R. Baker, Optimum quadratic detection of a random vector in Gaussian noise, *IEEE Trans. on Comm. Tech.*, COM-14, 802-805 (1966).

[6] __________, On the deflection of a quadratic-linear test statistic, *IEEE Trans. Inform. Th.*, IT-15, 16-21 (1969).

[7] C.R. Baker and A.F. Gualtierotti, Likelihood ratios and signal detection for nonGaussian processes, *Lectures Notes in Control and Information Science (Stochastic Processes in Underwater Acoustics)*, 85, 154-180 (1986).

[8] ______________________________, Discrimination with respect to a Gaussian process, *Probab. Th. Rel. Fields*, 71, 159-182 (1986).

[9] H. Cramér, On some classes of nonstationary stochastic processes, *Proc. Fourth Berkeley Symp. on Math. Stat. and Probab.*, 2, 57-77 (1961).

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