Optimum Detection of a Weak Signal with Minimal Knowledge of Dependency

GEORGE V. MOUSTAKIDES AND JOHN B. THOMAS, FELLOW, IEEE

Abstract—The optimum nonlinearity is defined for detection of a weak signal when minimal knowledge of the dependency structure of the observations is available. Specifically, it is assumed that the observations form a one-dependent strictly stationary sequence of random variables and that only a finite number of moments of the marginal density and the correlation coefficient between consecutive observations are known. It is assumed that the bivariate densities involved can be represented as diagonal series, using orthonormal polynomials. Using efficacy as a performance measure, the optimum nonlinearity is required to satisfy a saddle-point condition over this class of bivariate densities.

I. Introduction

A COMMONLY used scheme for the detection of constant signals in additive noise is to compare the sum of memoryless transformations of the observations to a fixed threshold. Optimality of this scheme can be defined in different ways. One way, for example, is in the Neyman–Pearson sense, where we try to maximize the probability of detection while keeping the false alarm probability less than a certain level. This criterion turns out to be difficult to use in cases of dependent observations, since it requires a knowledge of multiorder statistics. An alternative criterion is the maximization of the efficacy. The efficacy is an asymptotic measure of performance, and it can be used in weak signal situations. Also, this criterion is tractable when dependency is present. For the detection scheme in which we are interested, we need only second-order statistics to calculate the efficacy.

The problem of finding the memoryless transformation that maximizes the efficacy was solved in [1] for the m-dependent case and in [2] for the ϕ-mixing case. The solution requires knowledge of all bivariate densities involved. In [3], and [4] a min–max approach was used and the bivariate densities were assumed to belong to some known class of densities. In this paper we consider strictly stationary one-dependent sequences. We assume that the correlation coefficient between consecutive points is known and also that the marginal moments up to order $4m-2$ are known, where m is a positive integer. In most practical situations the dependence between consecutive points is much stronger than the dependence between points that are further apart in time. The one-dependence assumption can thus be regarded as a model of these cases and as one step away from the independent and identically distributed (i.i.d.) assumption. The assumption about the knowledge of the marginal moments is also reasonable, since there are methods for estimating them. Finally, knowledge of the correlation coefficient might be considered as the minimal knowledge we can require about dependence. It is thus interesting to see how this knowledge will change the existing results for the i.i.d. case. To derive a result we must make some assumptions about the class of bivariate densities we consider. A very common model in the literature is the expansion of the bivariate density in a diagonal series, using the set of orthonormal polynomials defined by the marginal density. This is the model we use here. What we will need is only certain properties of this expansion; further details of which can be found in [5]–[10]. In the examples we present in Section IV, the method introduced is also applied to a case where the bivariate density does not satisfy our assumptions. The resulting detection structure turns out to have performance that is always better than the linear detector, which is optimal for the i.i.d. Gaussian assumption.

Manuscript received April 5, 1984; revised March 25, 1985. This work was supported by the U.S. Office of Naval Research under Contract N00014-81-K-0146.

G. V. Moustakides was with the Department of Electrical Engineering and Computer Science, Princeton University, Princeton NJ 08544. He is now with IRISA, Avenue du General Leclerc, 35041 Rennes, France.

J. B. Thomas is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544.

IEEE Log Number 8406083.

II. Preliminaries

Let $f(x)$ be a symmetric density with unbounded support, such that all the moments exist and Fisher's information is finite. Assume that the orthonormal polynomials $\phi_n(x)$ defined by $f(x)$ form a complete orthonormal system in the $L_2(f)$ Hilbert space. To calculate the polynomials $\phi_n(x)$ it is enough to know the marginal moments $\gamma_n = E\{x^n\}$ and apply an orthonormalization procedure to the sequence $\{1, x, x^2, \cdots\}$. We are interested in bivariate densities that can be represented as

$$f(x, y) = f(x)f(y)\left\{\sum_{n=0}^{\infty} \alpha_n\phi_n(x)\phi_n(y)\right\}. \quad (1)$$

The equality in (1) is in the sense that

$$\lim_{M\to\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\frac{f(x, y)}{f(x)f(y)} - \sum_{n=0}^{M}\alpha_n\phi_n(x)\phi_n(y)\right\}^2 f(x)f(y)\,dx\,dy = 0. \quad (2)$$

A necessary condition for the expansion in (1) to be a valid density is that $\{\alpha_n\}_{n=0}^{\infty}$ has the representation

$$\alpha_n = \int_{-1}^{1} z^n h(z)\, dz, \qquad n = 0, 1, 2, \cdots \tag{3}$$

where $h(z)$ is a univariate density supported in $(-1, 1)$. For some marginals, (3) is sufficient to make the expansion in (1) a valid density, that is, nonnegative, but this is not true for every marginal [5]–[10]. Since $f(x)$ is by assumption symmetric and thus has zero mean, it is easy to see that $\phi_1(x) = x/\sqrt{\gamma_2}$. This result, using (1), yields

$$E\{xy\} = \gamma_2 E\{\phi_1(x)\phi_1(y)\} = \gamma_2 \alpha_1. \tag{4}$$

From (4) it is clear that α_1 is equal to the correlation coefficient ρ of x and y. For one-dependent, strictly stationary sequences, we know that the correlation coefficient is no greater than one-half in absolute value. We now define a class F_ρ of functions $f(x, y)$. A function $f(x, y)$ belongs to F_ρ if it satisfies (1) and (3) for some density $h(z)$ supported on $[-1, 1]$ and if

$$E\{x^n\} = \gamma_n, \qquad n = 1, 2, \cdots, 4m-2$$

$$\alpha_1 = \int_{-1}^{1} z h(z)\, dz = \rho \tag{5}$$

where ρ is the known correlation coefficient and γ_n the known marginal moments. Notice that neither the bivariate density nor the marginal density are assumed known. If we use the moments γ_n, which are known up to order $4m-2$, we can only compute the orthonormal polynomials $\phi_n(x)$ up to order $2m-1$. By allowing $h(z)$ to be supported on $[-1, 1]$, we allow degenerate functions in the class F_ρ. Clearly, F_ρ contains the bivariate densities of all one-dependent sequences that can be represented by (1).

The detection problem we would like to solve is the following. Let $\{N_n\}_{n=1}^{\infty}$ be a strictly stationary one-dependent noise sequence. Let $\{X_n\}_{n=1}^{M}$ be the observation sequence. We wish to decide between the two hypotheses

$$H_0: \quad X_n = N_n$$

$$H_1: \quad X_n = N_n + s_M \qquad n = 1, 2, \cdots, M \tag{6}$$

where s_M is a known scalar that tends to zero as $M \to \infty$. The detection scheme is the following:

$$u(T_M) = \begin{cases} 1, & \text{for } T_M > \gamma \\ p, & \text{for } T_M = \gamma \\ 0, & \text{for } T_M < \gamma \end{cases} \tag{7}$$

where

$$T_M = \frac{1}{\sqrt{M}} \sum_{n=1}^{M} \psi(X_n)$$

and where $u(T_n)$ is the probability of deciding H_1. The constants p and γ are chosen to achieve the required false alarm probability. The efficacy then [1] takes the form

$$\operatorname{eff}(\psi(x), f(x,y)) = \frac{\left[\int_{-\infty}^{\infty} \psi(x) f'(x)\, dx\right]^2}{\int_{-\infty}^{\infty} \psi^2(x) f(x)\, dx + 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y) f(x,y)\, dx\, dy}, \tag{8}$$

where $f(x, y)$ is the bivariate density of N_1 and N_2.

We assume that the nonlinearity $\psi(x)$ belongs to the class Ψ_m of all odd symmetric polynomials that have order up to $2m-1$. Since by assumption the polynomials are dense in $L_2(f)$, by letting $m \to \infty$, we have that Ψ_∞ is the class of all second-order odd symmetric nonlinearities. The restriction to odd symmetric polynomials is reasonable because we can prove that for every function in the class F_ρ the optimum nonlinearity that maximizes (8) is odd symmetric.

III. Optimum Nonlinearity

The nonlinearity that maximizes (8) is related to the actual bivariate density $f(x, y)$. Since we do not assume knowledge of this density, we define the optimum nonlinearity in a min–max way. In other words, we would like to find a pair $\psi_r(x) \in \Psi_m$ and $f_l(x, y) \in F_\rho$ such that the following saddle-point relation is satisfied:

$$\operatorname{eff}(\psi(x), f_l(x,y)) \le \operatorname{eff}(\psi_r(x), f_l(x,y)) \le \operatorname{eff}(\psi_r(x), f(x,y)) \tag{9}$$

for every $\psi(x) \in \Psi_m$ and every $f(x, y) \in F_\rho$.

First we will find the function from F_ρ that minimizes (7) for a given $\psi(x)$. This is equivalent to the following maximization:

$$\sup_{f(x,y)\in F_\rho} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y) f(x,y)\, dx\, dy. \tag{10}$$

Since $\psi(x)$ is odd symmetric, it can be expanded using only the odd symmetric orthonormal polynomials. Thus let

$$\psi(x) = \sum_{n=1}^{m} \psi_n \phi_{2n-1}(x). \tag{11}$$

Using (11), we can write

$$\begin{aligned} &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y) f(x,y)\, dx\, dy \\ &= \sum_{n=1}^{m} \psi_n^2 \alpha_{2n-1} = \sum_{n=1}^{m} \psi_n^2 \int_{-1}^{1} z^{2n-1} h(z)\, dz \\ &= \int_{-1}^{1} \left[\sum_{n=1}^{m} \psi_n^2 z^{2n-1}\right] h(z)\, dz = \int_{-1}^{1} A(z) h(z)\, dz \end{aligned} \tag{12}$$

where we define $A(z)$ as

$$A(z) = \sum_{n=1}^{m} \psi_n^2 z^{2n-1}. \tag{13}$$

These expressions hold for the case $m = \infty$ because the series is absolutely convergent for $|z| \le 1$. We can also interchange summation and integration in (12) using bounded convergence. The maximization problem now reduces to the following:

$$\sup_{h(z)} \int_{-1}^{1} A(z)h(z)\,dz \tag{14}$$

given that

$$\int_{-1}^{1} zh(z)\,dz = \rho. \tag{15}$$

Notice the following properties of $A(z)$: it is increasing and bounded in $[-1,1]$, it is analytic in $(-1,1)$, it is odd symmetric, and, for $z > 0$, it is convex. A typical form of $A(z)$ is given in Fig. 1. Now let $B(z) = \lambda z + \mu$ be the line that passes through the point $(1, A(1))$ and is tangent to

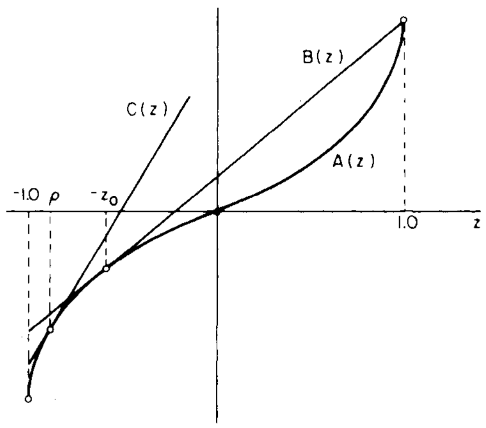

Fig. 1. Typical form of $A(z)$ and of tangent lines $B(z)$ and $C(z)$.

$A(z)$ at $-z_0$ (see Fig. 1). Then for every z, we have that

$$B(z) \geq A(z). \tag{16}$$

Notice that the point $-z_0$ can be found by solving the equation

$$\frac{A(1) - A(-z_0)}{1 + z_0} = A'(-z_0). \tag{17}$$

Proposition: The density $h(z)$ that solves the maximization problem defined by (14) and (15) is given by one of the following two cases.

Case 1: If $z_0 \geq -\rho$, the maximum is achieved by

$$h_M(z) = \frac{z_0 + \rho}{1 + z_0}\delta(z - 1) + \frac{1 - \rho}{1 + z_0}\delta(z + z_0). \tag{18}$$

Case 2: If $z_0 < -\rho$, the maximum is achieved by

$$h_M(z) = \delta(z - \rho). \tag{19}$$

Proof: We can see that in both cases, $h_M(z)$ is a valid density satisfying (15). For Case 1 maximizing (14) is equivalent to the following:

$$\sup_{h(z)}\left\{\int_{-1}^{1} A(z)h(z)\,dz - \lambda\rho - \mu\right\}; \tag{20}$$

but using (16), we have

$$\int_{-1}^{1} A(z)h(z)\,dz - \lambda\rho - \mu = \int_{-1}^{1} (A(z) - \lambda z - \mu)h(z)\,dz \leq 0. \tag{21}$$

Equality to zero is achieved when $h(z)$ is supported only on points where $A(z) = \lambda z + \mu$. For Case 2 the proof goes in a similar way. Instead of the line $B(z)$ that is tangent at $-z_0$, we use the line $C(z)$ that is tangent at ρ. The same arguments are valid because this line, as a result of the convexity, is always above $A(z)$ (see Fig. 1).

The function $f(x, y) \in F_\rho$ that corresponds to (18) is given by

$$f_M(x, y) = (1 - p)f(x)\delta(x - y) + pf(x)f(y)\left\{\sum_{n=0}^{\infty}(-z_0)^n\phi_n(x)\phi_n(y)\right\} \tag{22}$$

where

$$p = \frac{1 - \rho}{1 + z_0}. \tag{23}$$

Let us now find the optimum nonlinearity $\psi(x) \in \Psi_m$, when the function $f(x, y)$ has a form similar to the one given by (22). Since $f(x)$ has finite Fisher's information and is symmetric, we can write

$$-\frac{f'(x)}{f(x)} = \sum_{n=1}^{\infty}\beta_n\phi_{2n-1}(x). \tag{24}$$

Notice now the following:

$$\beta_n = \int -\frac{f'(x)}{f(x)}\phi_{2n-1}(x)f(x)\,dx = \int \phi'_{2n-1}(x)f(x)\,dx. \tag{25}$$

From (25), since $\phi'_n(x)$ is a polynomial, we can compute β_n using the moments γ_n. Since the polynomials $\phi_n(x)$ can be computed up to order $2m - 1$, we can compute β_n for $n = 1, 2, \cdots, m$. As it will turn out, this is all we need. The efficacy for a $\psi(x) \in \Psi_m$ takes the form

$$\operatorname{eff}(\psi(x), f_M(x, y)) = \frac{\left[\sum_{n=1}^{m}\psi_n\beta_n\right]^2}{\sum_{n=1}^{m}\psi_n^2 + 2\sum_{n=1}^{m}\psi_n^2\left[(1 - p) - pz_0^{2n-1}\right]} = \frac{\left[\sum_{n=1}^{m}\psi_n\beta_n\right]^2}{\sum_{n=1}^{m}\psi_n^2\left[3 - 2p\left(1 + z_0^{2n-1}\right)\right]}. \tag{26}$$

Equation (26) is maximized when $\psi_n = \psi_n^*$, where

$$\psi_n^* = \frac{k\beta_n}{\sqrt{3 - 2p\left(1 + z_0^{2n-1}\right)}}, \qquad n = 1, 2, \cdots, m \tag{27}$$

where k is an arbitrary constant. We assume for simplicity that $k = 1$. Thus, for a given $\psi(x)$, the function $f(x, y)$ that minimizes the efficacy is given by (22). On the other hand, if $f(x, y)$ has the form of (22), then the optimum $\psi(x)$ satisfies (27).

To find the pair we are looking for, we have to satisfy (22) and (27) simultaneously. We will assume that Case 1 of the Proposition will occur and that our $\psi_r(x)$ satisfies (27) for some $z_0 = z_r$. Thus for $\psi_r(x)$ we only have to specify z_r in some way. For every z_r we define a function $A_r(z)$ that is similar to $A(z)$ defined in (13) as follows:

$$A_r(z) = \sum_{n=1}^{m} \frac{\beta_n^2 z^{2n-1}}{3 - 2p(1 + z_r^{2n-1})} \tag{28}$$

where p is given by

$$p = \frac{1-\rho}{1+z_r}. \tag{29}$$

Since, from (9), we would like $f_l(x, y)$ to minimize the $\mathrm{eff}(\psi_r(x), f(x, y))$, the function $f_l(x, y)$ must have a form similar to (22) with $z_0 = z_r$. For this form to minimize the efficacy, z_r must be a solution of an equation similar to (17). In other words,

$$\frac{A_r(1) - A_r(-z_r)}{1 + z_r} = A_r'(-z_r). \tag{30}$$

When we substitute (28) into (30) and multiply by $(1 + z_r)$, after canceling common terms, (30) reduces to

$$\sum_{n=2}^{m} \frac{\beta_n^2(1 + z_n^{2n-1})}{3 - 2(1-\rho)\dfrac{1 + z_r^{2n-1}}{1 + z_r}} = (1 + z_r) \sum_{n=2}^{m} \frac{(2n-1)\beta_n^2 z_r^{2(n-1)}}{3 - 2(1-\rho)\dfrac{1 + z_r^{2n-1}}{1 + z_r}}. \tag{31}$$

Equation (31) has z_r as its only unknown. In the Appendix we show that a positive solution always exists and that it is no less than one-half. This means that we always have $z_r \geq \rho$ and that we do not contradict our assumption that Case 1 will occur. We will get a contradiction, though, if we assume that Case 2 will occur.

Theorem: Let z_r be a solution to (31). Define

$$f_l(x, y) = \frac{z_r + \rho}{1 + z_r} f(x)\delta(x - y) + \frac{1-\rho}{1+z_r} \cdot f(x)f(y)\left\{\sum_{n=0}^{\infty} (-z_r)^n \phi_n(x)\phi_n(y)\right\}$$

$$\psi_r(x) = \sum_{n=1}^{m} \frac{\beta_n}{\sqrt{3 - 2(1-\rho)\dfrac{1 + z_r^{2n-1}}{1 + z_r}}} \phi_{2n-1}(x). \tag{32}$$

Then $\psi_r(x)$ and $f_l(x, y)$ satisfy (9).

Proof: The proof is an immediate consequence of the way that z_r is defined. The left inequality of (9) is satisfied, because $\psi_r(x)$ satisfies (27). The right inequality is satisfied, because $f_l(x, y)$ minimizes the $\mathrm{eff}(\psi_r(x), f(x, y))$.

A few things are noteworthy. When either $m = 2$ or $\beta_n = 0$ for $n = 3, \cdots, m$, the only nonnegative solution to (31) is $z_r = 1/2$, regardless of β_2, as long as $\beta_2 \neq 0$. Also, for any $m < \infty$, when $\rho \to -1/2$ we have that $\psi_r(x) \to x$ (after it is properly normalized), so long as $\beta_1 \neq 0$. Notice that $z_r \neq 0$ even when $\rho = 0$, which means that the independence assumption is not necessarily the best candidate, when the correlation coefficient is zero.

IV. Examples

As a first example we consider the case where $f(x)$ is the standard $N(0,1)$ Gaussian density. The orthonormal polynomials are the Hermite polynomials. Also, (3) is sufficient for (1) to be a valid density [7]. Since the locally optimum nonlinearity is linear, in other words equal to $\phi_1(x)$, we have that $\beta_n = 0$ for $n \geq 2$. From (32) we conclude that $\psi_r(x)$ will be linear also. Using the formula for the expansion of the bivariate Gaussian density in a diagonal series and noting that $z_r = 1/2$, we have

$$f_l(x, y) = \frac{1 + 2\rho}{3} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)\delta(x - y) + \frac{2 - 2\rho}{3} \frac{2}{3\pi} \exp\left(-\frac{2}{3}[x^2 + y^2 + xy]\right). \tag{33}$$

As a second example, let us consider the case $m = 2$. By applying an orthonormalization procedure to $\{1, x, x^2, x^3\}$, we have

$$\phi_1(x) = \frac{x}{\sqrt{\gamma_2}}$$

$$\phi_3(x) = \frac{1}{\sqrt{\gamma}}\{\gamma_4 x - \gamma_2 x^3\} \tag{34}$$

where $\gamma = \gamma_2^2\gamma_6 - \gamma_2\gamma_4^2$. Using the Schwarz inequality, we can see that $\gamma \geq 0$. We can have $\gamma = 0$ only when $|x|$ is concentrated on a single point. From (25) we compute β_1 and β_2. We have

$$\beta_1 = \frac{1}{\sqrt{\gamma_2}}, \quad \beta_2 = \frac{1}{\sqrt{\gamma}}\{\gamma_4 - 3\gamma_2^2\}. \tag{35}$$

Since we are in the case $m = 2$, we have $z_r = 0.5$, and the nonlinearity $\psi_r(x)$ becomes after normalization

$$\psi_r(x) = \left[\frac{\gamma}{\gamma_2(\gamma_4 - 2\gamma_2^2)}\sqrt{\frac{3(1+\rho)}{2 + 4\rho}} + \gamma_4\right]x - \gamma_2 x^3. \tag{36}$$

Notice that when $f(x)$ is heavy-tailed we usually have $\gamma_4 > 3\gamma_2^2$. This will yield a nonlinearity of the form given in Fig. 2. This nonlinearity is of a similar form to the nonlinearity defined in [3]. If we now calculate the efficacy using the least-favorable function defined in (32), we have

$$\mathrm{eff}_r = \frac{\beta_1^2}{1 + 2\rho} + \frac{2\beta_2^2}{3(1+\rho)} \tag{37}$$

where β_1 and β_2 were defined in (35). Notice that this value is, in general, pessimistic, since the function defined in (32) is not always a density. Let us now see what is the

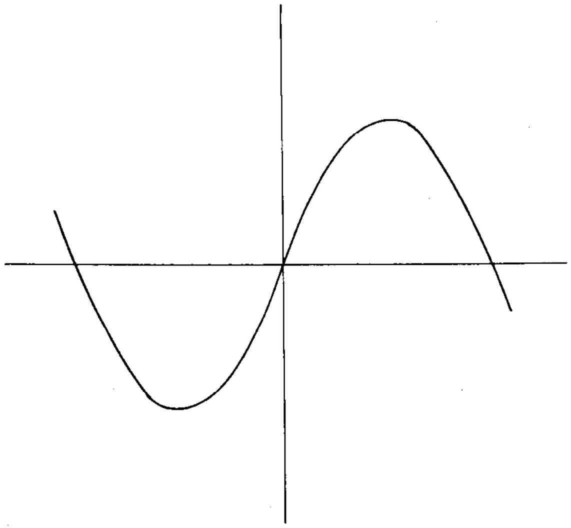

Fig. 2. Typical form of nonlinearity $\psi_r(x)$ for case $m = 2$.

optimum linear detection scheme. Let μ be the unique solution of the equation $\mu^2 - \mu/\rho + 1 = 0$ that satisfies $|\mu| < 1$. This equation always has real roots because $|\rho| \leq 0.5$ for one-dependent sequences. Let the sequence Z_n be defined as follows:

$$Z_n = -\mu Z_{n-1} + X_n, \qquad n = 1, 2, \cdots, \tag{38}$$

where X_n is the observation sequence defined by (6). Under H_0, the Z_n is zero mean and under H_1 it has a mean equal to $s/(1+\mu)$. The Z_n variables are uncorrelated. The optimum linear detector would be to compare the sum of the Z_n to a threshold, and the efficacy of this detector is

$$\text{eff}_l = \frac{1}{(1+\mu)^2 E\{Z_1^2\}}. \tag{39}$$

Calculating $E\{Z_1^2\}$ in terms of the moments of the sequence X_n and the correlation coefficient ρ gives

$$\text{eff}_l = \frac{1}{\gamma_2(1+2\rho)}. \tag{40}$$

Comparing (40) with (37) and recalling that $\beta_1 = 1/\sqrt{\gamma_2}$, we can see that $\text{eff}_r \geq \text{eff}_l$.

To see how this method behaves in a case that does not satisfy our assumptions, consider the following noise sequence:

$$N_i = Y_i + \mu Y_{i-1}, \qquad i = 1, 2, \cdots, \tag{41}$$

where the Y_i are i.i.d. and $|\mu| < 1$. If we apply the linear detector defined earlier, the random variables Z_n under H_0 are equal to Y_n (for $Z_0 = Y_0$). If $h(y)$ now is the common density of the variables Y_n, then the locally optimum detector would be to compare the sum of $l(Z_n)$ to a threshold, where $l(y) = -h'(y)/h(y)$. This gives an efficacy equal to

$$\text{eff}_0 = \frac{1}{(1+\mu)^2} E\left\{\left[\frac{h'(Y_1)}{h(Y_1)}\right]^2\right\}. \tag{42}$$

Obviously, the use of this detector presupposes the knowledge of (41) and of the density $h(y)$. To be fair in our comparisons, we will assume that we know (41) but not $h(y)$. Using (41) and the moments γ_n, we can calculate the moments of Y_n; thus we can approximate $l(y)$ with the first terms of its expansion using the orthonormal polynomials. If δ_n are the coefficients of this expansion, the efficacy of the detection scheme that uses the approximation instead of $l(y)$ is

$$\text{eff}_{s0} = \frac{1}{(1+\mu)^2} \sum_{n=1}^{m} \delta_n^2. \tag{43}$$

Consider $m = 2$; then Table I gives the asymptotic relative efficiency (ARE) of the $\psi_r(x)$ detector with respect to the linear detector, the approximate locally optimum detector, and the locally optimum detector. The density $h(y)$ was assumed to be $N(0,1)$ with probability 0.95 and $N(0, \sigma)$ with probability 0.05. Also, μ was taken equal to 0.268, thus yielding $\rho = 0.25$. We can see that $\psi_r(x)$ is always better than the linear detector and that for $\sigma \geq 5$ it is significantly better. Compared with the scheme that uses the approximation of $l(y)$, it is very slightly inferior. Compared with the locally optimum detector it behaves very badly, but as we have mentioned, the optimum scheme requires knowledge of the dependence defined by (41) and the density $h(y)$.

TABLE I
ASYMPTOTIC RELATIVE EFFICIENCY OF $\psi_r(x)$ WITH RESPECT TO DIFFERENT DETECTORS

σ	Linear	Approximate	Locally Optimum
2	1.03	0.99	0.99
3	1.13	0.99	0.92
4	1.25	0.98	0.81
5	1.36	0.97	0.70
6	1.47	0.96	0.60
7	1.57	0.96	0.52
8	1.65	0.95	0.44
9	1.73	0.95	0.38
10	1.77	0.94	0.33

Comments: We have presented a method for finding an optimum nonlinearity for signal detection when dependence is present in the additive noise. For the dependence structure, we have assumed knowledge only of the correlation coefficient between consecutive observations. Even though this method is tractable from an analytical point of view, it produces some practical problems. The generation of the orthonormal polynomials is difficult for high orders. If we consider the polynomials as an approximation to the optimum nonlinearity for the Ψ_∞ case, their convergence is slow in cases where this optimum nonlinearity is bounded. That is because we approximate a bounded function using unbounded polynomials. Also, from (32) we can see that the density $f_l(x, y)$ contains a delta-function component. This function is not a good candidate for a bivariate density of a one-dependent sequence. The reason we get this form of worst-case density is that we optimize using only necessary and not sufficient conditions. By requiring the functions $f(x, y)$ to satisfy more necessary conditions, we could probably get better results. For example, if we use the property $E\{[c_1\phi_n(x_1) + \cdots + c_k\phi_n(x_k)]^2\} \geq 0$, this yields, for one-dependent sequences satisfying (1), $c_1^2 + \cdots + c_k^2 \geq -2\alpha_n\{c_1c_2 + \cdots + c_{k-1}c_k\}$. Defining first $c_j = 1$ and letting $k \to \infty$ yields $\alpha_n \geq -0.5$. Defining then $c_j = (-1)^j$ and letting $k \to \infty$ yields $0.5 \geq \alpha_n$. We thus have that all the coefficients α_n in (1) are not greater than 0.5 in absolute value. Using this necessary condition, we can restrict further the class F_ρ of allowable functions $f(x, y)$, but a much more complicated analysis results.

Appendix

Existence of a Solution z_r

First notice that if we use (23), then

$$3 - 2p\left(1 + z_0^{2n-1}\right) \geq 3 - 2(1 - \rho) \geq 0. \quad (44)$$

This is important because in (27) we take the square root of this expression. Let us now define as $D(z_r)$ and $G(z_r)$ the left and right side of (31), respectively. Notice that each term in these two expressions is a continuous function of z_r. For $\rho > -1/2$ we can show that the two sums that define $D(z_r)$ and $G(z_r)$ are absolutely summable on [0,1] and [0,1), respectively. Using bounded convergence, we can also show that these functions are continuous. By direct calculation we have

$$D(0) \geq G(0). \quad (45)$$

We also have that

$$G(z_r) \geq (1 + z_r) \sum_{n=2}^{m} \frac{\beta_n^2 z_r^{2(n-1)}}{3 - 2(1 - \rho)\dfrac{1 + z_r^{2n-1}}{1 + z_r}}. \quad (46)$$

The right side of (46) is continuous and absolutely summable on [0,1], and thus in the limit as $z_r \to 1$ we get

$$\liminf_{z_r \to 1} G(z_r) \geq \frac{2}{3 - 2(1 - \rho)} \sum_{n=2}^{m} \beta_n^2 = D(1). \quad (47)$$

Continuity of the two expressions, combined with (45) and (47), proves existence of a solution in the interval [0,1). To show now that this solution cannot be less than 1/2, it is enough to show that every term in the difference $D(z_r) - G(z_r)$ is nonnegative for $z_r \leq 1/2$. In other words, it is enough to show that for $0 \leq z_r \leq 1/2$

$$\left(1 + z_r^{2n-1}\right) - (2n - 1)(1 + z_r) z_r^{2(n-1)} \geq 0. \quad (48)$$

The left side of (48) is decreasing with $z_r \in [0,1]$; thus it is enough to show (48) for $z_r = 1/2$ or, after some manipulation, to show

$$1 \geq \frac{3n - 2}{4^{n-1}}. \quad (49)$$

This is true for every $n \geq 1$. Thus the solution to (31) is no less than one-half. This concludes the proof.

References

[1] H. V. Poor and J. B. Thomas, "Memoryless discrete time detection of constant signals in *m*-dependent noise," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 54–61, Jan. 1979.

[2] D. R. Halverson and G. L. Wise, "Discrete time detection in ϕ-mixing noise," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 189–198, Mar. 1980.

[3] H. V. Poor, "Signal detection in the presence of weakly dependent noise; Parts I and II," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 735–752, Sep. 1982.

[4] G. V. Moustakides and J. B. Thomas, "Min–max detection of weak signals in φ-mixing noise," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 529–537, May 1984.

[5] H. O. Lancaster, "The structure of bivariate distributions," *Ann. Math. Stat.*, vol. 29, pp. 719–736, 1958.

[6] ——, "Correlations and canonical forms of bivariate distributions," *Ann. Math. Stat.*, vol. 34, pp. 532–538, 1963.

[7] O. V. Sarmanov and Z. N. Bratoeva, "Probabilistic properties of bilinear expansions of Hermite polynomials," *Theory Prob. Appl.*, vol. 12, pp. 470–481, 1967.

[8] G. L. Wise, "The bivariate structure and representation of random processes," Ph.D. dissertation, Dep. Elec. Eng., Princeton Univ., Princeton, NJ, 1974.

[9] S. G. Tyan, "The structure of bivariate distribution functions and their relation to Markov processes," Ph.D. dissertation, Dep. Elec. Eng., Princeton Univ., Princeton, NJ, 1975.

[10] J. Abrahams and J. B. Thomas, "A note on the characterization of bivariate densities with Gaussian marginal densities," *SIAM J. Appl. Math.*, vol. 37, pp. 433–435, Oct. 1979.