TECHNICAL
RESEARCH
REPORT

SYSTEMS
RESEARCH
CENTER

*Supported by the
National Science Foundation
Engineering Research Center
Program (NSFD CD 8803012),
Industry and the University*

One-Step Memory Nonlinearities for Signal Detection and Discrimination from Correlated Observations

by D. Sauder and E. Geraniotis

TR 92-41

ONE-STEP MEMORY NONLINEARITIES FOR SIGNAL DETECTION AND DISCRIMINATION FROM CORRELATED OBSERVATIONS

D. Sauder and E. Geraniotis

ONE-STEP MEMORY NONLINEARITIES FOR SIGNAL DETECTION AND DISCRIMINATION FROM CORRELATED OBSERVATIONS

D. Sauder and E. Geraniotis

ABSTRACT

New detectors based on one-step memory nonlinearities and employing the test statistic $\sum_{j=0}^{n-1} g(X_j, X_{j+1})$ are introduced. Problems of discrimination between two arbitrary stationary m-dependent or mixing sequences of observations and problems of detecting a weak signal in additive stationary m-dependent or mixing noise are considered in this context. For each problem, the nonlinearity g is optimized for performance criteria, such as the generalized signal-to-noise ratio and the efficacy and is obtained as the solution to an appropriate linear integral equation. Moreover, the schemes considered can be robustified to statistical uncertainties determined by 2-alternating capacity classes, for the second-order joint pdfs of the observations, and by bounds on the correlation coefficients of time-shifts of the observation sequence, for the third- and fourth-order joint pdfs. Evaluation of the performance of the new schemes via simulation reveals significant gains over that of detectors employing memoryless nonlinearities or the i.i.d. nonlinearity.

D. Sauder is with Locus, Inc., Alexandria, VA, and the Department of Electrical Engineering, University of Maryland at College Park, MD 20742.
E. Geraniotis is with the Department of Electrical Engineering and the Systems Research Systems, University of Maryland at College Park, MD 20742.
This research was supported in part by the Systems Research Center at the University of Maryland, College Park, through the National Science Foundation's Engineering Research Centers Program, NSF CDR 8803012.

I. Introduction and Problem Formulation

Over the last decade, a considerable amount of attention has been devoted to the study of detectors and discriminators based on memoryless nonlinearities. These tests use a test statistic of the form

$$T_n(\mathbf{x}) = \sum_{i=1}^{n} g(x_i). \tag{1}$$

A decision is usually made by comparing T_n to a threshold and declaring the hypothesis H_1 to be true if and only if T_n exceeds the threshold. Most of the research in this area has focused on optimal memoryless nonlinearities ([1], [2]), and robust memoryless nonlinearities ([3]) for weak-signal detection. More recently, Sadowsky and Bucklew [4] and Sauder and Geraniotis [5] have considered the more general problem of signal discrimination, where the test must decide which of two possible random signals is observed. In this paper, we advance further in discriminator/detector complexity by considering nonlinearities possessing one-step memory.

The signal discrimination problem involves deciding between the two hypotheses

$$\begin{aligned} H_0 : &\quad \{X_k\} \text{ has the distribution } F_0 \\ H_1 : &\quad \{X_k\} \text{ has the distribution } F_1 \end{aligned} \tag{2}$$

where we use F_i to denote symbolically all the known or assumed information about the entire process $\{X_k\}$ under H_i. Problems of this type arise in areas such as radar target discrimination, where one must decide which of two possible targets is present on the basis of returned radar pulses. In some instances, the distribution of a process is known in principle, as the underlying mechanism which generates the process is known, but a closed form expression for the complete multivariate density is not available (e.g. the Rice process which is obtained as the envelope of a complex Gaussian process with nonzero mean). For the derivation of the optimal one-step memory nonlinearities in this paper, all of the marginal probability densities of dimensions one, two, three, and four are required. Therefore, for this paper F_i must denote at least the complete set of marginal densities up to four dimensions. By contrast, only univariate and bivariate densities are required for the derivation of optimal memoryless nonlinearities [5].

A problem that is related to the discrimination problem is the weak-signal detection problem,

$$\begin{aligned} H_0 : &\quad X_k = Y_k \\ H_1 : &\quad X_k = Y_k + \theta, \end{aligned} \tag{3}$$

where $\theta > 0$ represents the constant signal and $\{Y_k\}$ the additive noise. The statistical distribution F_0 of the noise is assumed to be known, in the sense described in the preceding paragraph. To obtain the optimal one-step memory nonlinearity derived here, the marginal densities corresponding to F_0 up to four dimensions must be available. So far, research in the weak-signal detection area has yielded memoryless nonlinearities which are optimal under the asymptotic relative efficiency (ARE) performance measure [1], [2] and robust [3]. These results required knowledge of the univariate and bivariate densities only.

Memoryless discriminators possess a form that is simple to implement, and this makes them attractive from a practical standpoint. However, since correlated observations possess memory, we are inclined to think a priori that a discriminator with memory will perform substantially better. It is interesting to note that the situations in which one obtains the best improvement over the i.i.d. discriminator by using an optimal memoryless discriminator also appear to be the most favorable for further improvement by using an optimal one-step memory discriminator. In some earlier simulation results [5], we compared the error probabilities for the memoryless discriminator that is optimal for independent and identically distributed (i.i.d.) processes, i.e. $g(x) = \log[f_1(x)/f_0(x)]$, with that of the memoryless discriminator that is optimal for correlated processes under a generalized signal-to-noise ratio (SNR) performance measure. It was observed that the i.i.d. discriminator tends to have the lower error probabilities when one of two conditions prevails: (1) when the correlation is very weak and (2) when the marginal densities f_1 and f_0 are significantly different. A plausible explanation is obvious. In the first situation, the processes are close enough to being i.i.d. that the i.i.d. discriminator, which is actually the likelihood ratio test (LRT) for i.i.d. processes, yields nearly optimal error probabilities. In the second case, the signals under the two hypotheses are so different that any SNR-type performance measure is not likely to be useful in predicting of how discriminators perform under a criterion involving the error probabilities (e.g. Bayes risk). This conclusion was also obtained by Sadowsky and Bucklew in [4]. Thus the optimal memoryless discriminators derived in [5] seem to offer the greatest improvement over the i.i.d. discriminator when the univariate densities are not too dissimilar and the correlation is moderate to strong. Intuition suggests, however, that in this case most of the discrimination potential lies in the correlation, especially if the strength of the correlation is significantly different under the two hypotheses. It is primarily this line of thinking which motivates us to investigate one-step memory nonlinearities for discrimination and detection.

The form of the test statistic we employ is

$$T_n(\mathbf{x}) = \sum_{j=1}^{n-1} g(x_j, x_{j+1}) \tag{4}$$

where $g(\cdot,\cdot)$ is a one-step memory nonlinearity that characterizes the test. Since we are concerned primarily about asymptotic performance, we ignore the possibility of an initial term $g_0(x_1)$. A decision is made by comparing T_n to a threshold, as in the case of the memoryless discriminator. The test statistic (4) maintains some of the advantages of the memoryless test statistic (1), while introducing a means to discriminate on the basis of correlation. In particular, with the proper condition of asymptotic independence of the observations, T_n obeys a central limit theorem. One must use caution, though, not to infer too much from the asymptotic normality of the test statistic. For example, if the error probabilities vanish asymptotically, as they do for a consistent test, one cannot use the central limit theorem to accurately approximate them. See [4] for a more complete discussion of issues related to nonlocal detection and an approach based on large deviations. The asymptotic normality of the test statistic is important, however, in the context of weak-signal detection when the efficacy performance measure (defined below) is used.

Dependency conditions that are sufficient for a central limit theorem (CLT) to hold are well known. For a good survey paper on CLTs for dependent data, see [6]. The two conditions we consider most useful are m-dependence and ρ-mixing. To define these conditions, let $\mathcal{F}_a^b$ denote the σ-field generated by the random variables $\{X_a, \ldots, X_b\}$. The process $\{X_k\}$ is said to be m-dependent under a distribution F if $\mathcal{F}_{-\infty}^k$ and $\mathcal{F}_{k+m}^\infty$ are independent for any k. If U and V are random variables which are measurable with respect to $\mathcal{F}_{-\infty}^k$ and $\mathcal{F}_{k+n}^\infty$, respectively, and square integrable under the distribution F, then ρ_n is defined as $\sup_{U,V} \mathrm{Cov}(U,V)/\sqrt{\mathrm{Var}U\,\mathrm{Var}V}$ where the supremum is taken over all such U and V. The process $\{X_k\}$ is then defined to be ρ-mixing under the distribution F if the sequence $\{\rho_n\}$ converges to zero as $n \to \infty$. If the observation sequence $\{X_k\}$ is m-dependent, then the sequence $\{g(X_k, X_{k+1})\}$ is $(m+1)$-dependent for any square-integrable nonlinearity g. Also, if $\{X_k\}$ is ρ-mixing, so is $\{g(X_k, X_{k+1})\}$. Thus a central limit theorem applies as in the memoryless case. Here we assume that all processes are stationary and either m-dependent or ρ-mixing.

The asymptotic performance of the test statistic T_n may be measured by its asymptotic mean and variance, which have the expressions

$$\mu_i(g) = \lim_{n\to\infty} \frac{1}{n} \mathrm{E}_i\, T_n = \mathrm{E}_i\, g(X_1, X_2) \tag{5}$$

and

$$\begin{aligned} \sigma_i^2(g) &= \lim_{n\to\infty} \frac{1}{n} \mathrm{Var}_i\, T_n \\ &= \mathrm{Var}_i\, g(X_1, X_2) + 2\sum_{j=1}^{\infty} \mathrm{Cov}_i\left[g(X_1, X_2), g(X_{j+1}, X_{j+2})\right] \end{aligned} \tag{6}$$

where E_i, Var_i, and Cov_i respectively denote the expectation, variance, and covariance operators under H_i. When the process is m-dependent, then the upper limit on the sum in (6) is m.

In our work [5] on memoryless nonlinearities for discrimination, we used for a performance measure a generalized signal-to-noise ratio

$$S_\zeta(g) = \frac{|\mu_1(g) - \mu_0(g)|^2}{(1-\zeta)\sigma_0^2(g) + \zeta\sigma_1^2(g)} \tag{7}$$

where $0 \leq \zeta \leq 1$. In this paper, we take the optimal one-step memory nonlinearity to be the one which maximizes the performance measure S_ζ. The parameter ζ determines the weighting of the asymptotic variances. Although a value of ζ near $\frac{1}{2}$ yields a balanced mixture, we also obtained good simulation results in [5] for $\zeta = 1$, especially in the case of strong correlation under H_1 and weak correlation under H_0. It is also possible to take ζ to be a function of g. In particular, if ζ is taken to be $\zeta(g) = (2\sigma_0\sigma_1 + \sigma_1^2)/(\sigma_1^2 - \sigma_0^2)$, then the performance measure becomes

$$S(g) = \frac{|\mu_1 - \mu_0|^2}{|\sigma_0 + \sigma_1|^2} \tag{8}$$

which is the performance measure obtained in [4]. In [5], it was explained how the optimal nonlinearity for (8) may be obtained by solving iteratively a series of linear integral equations. The idea is to choose an initial value for ζ, optimize the performance measure S_ζ, use the nonlinearity thus obtained to compute σ_0 and σ_1, and then compute a new value for ζ. The procedure is then repeated until the value of ζ does not change significantly.

When the problem under consideration is weak-signal detection (3), the appropriate performance measure is the efficacy functional [7]

$$\eta(g) = \frac{|\mu_0'(g)|^2}{\sigma_0^2(g)} \tag{9}$$

where $\mu_0'(g) = (d/d\theta)\mathrm{E}_\theta\, g(X_1, X_2)\big|_{\theta=0}$. We denote by E_θ the expectation taken under H_1 when the signal strength is θ. The use of the efficacy functional requires that

$$\lim_{\theta \to 0} \frac{\sigma_\theta^2(g)}{\sigma_0^2(g)} = 1 \tag{10}$$

where $\sigma_\theta^2(g)$ is the asymptotic variance of T_n when the signal strength is θ. A few other mild regularity conditions on the distribution are also required [7], including the uniform weak convergence of the distribution to normality for θ in an interval around zero. The relation of (9) to (7) is evident.

We show in this paper that one-step memory nonlinearities maximizing the generalized SNR (Section II) and efficacy (Section III) performance measures may be obtained through the solution of certain linear integral equations. For m-dependent observations, the integral equations yield precisely the optimal nonlinearities, while for ρ-mixing observations the optimal nonlinearities may be approximated as closely as desired by the integral equation solutions. Because the integral equations generally lack closed form solutions, numerical approximations are required, and we show how the approximations may be obtained by solving a large but manageable system of linear equations. We have also obtained discriminators that have a guaranteed level of performance over uncertainty classes for the distributions (Section IV). This result is related to minimax robustness, the difference being that we have not been able to prove that there actually exists a pair of distributions in the uncertainty classes for which the lower bound on the performance is actually obtained. In our numerical results in Section V, we compare the performance of several discriminators (i.i.d. LRT, optimal memoryless, and optimal one-step memory) for the problem of discriminating between Rayleigh and lognormal processes. We also show the ARE of several detectors relative to the i.i.d (locally optimal) detector for the problem of signal detection in Cauchy noise. The results show the significant advantage in the discriminator or weak-signal detector performance that becomes possible through the use of one-step memory nonlinearities.

II. Derivation of Optimal Nonlinearity for Signal Discrimination

The one-step memory discriminator is characterized by the one-step memory nonlinearity $g(\cdot,\cdot)$ used in the test statistic T_n, as defined by (4), and the threshold γ. Once a nonlinearity g is given, the distribution of T_n can be estimated through simulation and the threshold determined on the basis of the estimated distribution and the desired error probabilities. Thus the important design step is the selection of the nonlinearity. Our task is to find the nonlinearity g which maximizes $S_\zeta(g)$.

To avoid problems with the interchanging of limits, we must begin with the optimization of $S_\zeta(g)$ for the m-dependent model. Results for the ρ-mixing model are obtained via approximation by m-dependent models with sufficiently large m. We first note that $S_\zeta(g)$ remains invariant under scaling of g; therefore we may maximize the numerator while constraining the denominator to be constant. Introducing a Lagrange multiplier λ, our maximization problem becomes equivalent to maximizing the functional

$$J(g) = \mu_1(g) - \mu_0(g) - \lambda\big[(1-\zeta)\sigma_0^2(g) + \zeta\sigma_1^2(g)\big]. \tag{11}$$

This is the approach used in [1], [2], [4], [5]. A necessary condition for g to maximize J is that $\frac{d}{d\epsilon}J(g+\epsilon h)\big|_{\epsilon=0} = 0$, for an arbitrary nonlinearity h satisfying $\mathrm{E}_i \left|h(X_1,X_2)\right|^2 < \infty$ for $i = 0, 1$. First we compute

$$\frac{d}{d\epsilon}\mu_i(g+\epsilon h)\big|_{\epsilon=0} = \mathrm{E}_i\, h(X_1,X_2) \tag{12}$$

and

$$\begin{aligned}\frac{d}{d\epsilon}\sigma_i^2(g+\epsilon h)\big|_{\epsilon=0} &= 2\mathrm{E}_i\, g(X_1,X_2)h(X_1,X_2)\\ &\quad + 2\sum_{j=1}^{m}\left[\mathrm{E}_i\, g(X_1,X_2)h(X_{j+1},X_{j+2}) + \mathrm{E}_i\, g(X_{j+1},X_{j+2})h(X_1,X_2)\right]\\ &\quad - 2(2m+1)\mathrm{E}_i\, g(X_1,X_2)\mathrm{E}_i\, h(X_1,X_2).\end{aligned} \tag{13}$$

We now introduce the notation $f_i^{(j,k)}$, $f_i^{(j,k,l)}$, and $f_i^{(j,k,l,p)}$ denoting the densities of (X_j,X_k), (X_j,X_k,X_l), and(X_j,X_k,X_l,X_p), respectively, under H_i. Using (12) and (13), we obtain

$$\begin{aligned}\frac{d}{d\epsilon}J(g+\epsilon h)\big|_{\epsilon=0} &= \iint h(x,y)\big[f_1^{(1,2)}(x,y) - f_0^{(1,2)}(x,y)\big]\,dx\,dy\\ &\quad - 2\lambda\big[(1-\zeta)W_0(g,h) + \zeta W_1(g,h)\big]\end{aligned} \tag{14}$$

where the operator W_i is defined by

$$
\begin{aligned}
W_i(g,h) = &\iint g(x,y)h(x,y)f_i^{(1,2)}(x,y)\,dx\,dy \\
&+ \iiint \big[g(x,y)h(y,u)+g(y,u)h(x,y)\big] f_i^{(1,2,3)}(x,y,u)\,dx\,dy\,du \\
&+ \sum_{j=2}^{m} \iiiint \big[g(x,y)h(u,v)+g(u,v)h(x,y)\big] f_i^{(1,2,j+1,j+2)}(x,y,u,v)\,dx\,dy\,du\,dv \\
&- (2m+1)\iint g(x,y)f_i^{(1,2)}(x,y)\,dx\,dy \iint h(x,y)f_i^{(1,2)}(x,y)\,dx\,dy.
\end{aligned}
\tag{15}
$$

The right-hand side of (14) becomes zero for an arbitrary h if and only if the part of the integrand which multiplies h is identically 0. Thus we obtain the integral equation

$$
\begin{aligned}
f_1^{(1,2)}(x,y) - f_0^{(1,2)}(x,y) = &\, 2\lambda g(x,y)\big[(1-\zeta)f_0^{(1,2)}(x,y)+\zeta f_1^{(1,2)}(x,y)\big] \\
&+ 2\lambda\int g(u,x)K_1(u,x,y)\,du + 2\lambda\int g(y,u)K_1(x,y,u)\,du \\
&+ 2\lambda\iint g(u,v)K_2(x,y,u,v)\,du\,dv
\end{aligned}
\tag{16}
$$

where

$$
K_1(u,x,y) = (1-\zeta)f_0^{(1,2,3)}(u,x,y)+\zeta f_1^{(1,2,3)}(u,x,y) \tag{17}
$$

$$
\begin{aligned}
K_2(x,y,u,v) = &\sum_{j=2}^{m}\big[(1-\zeta)f_0^{(1,2,j+1,j+2)}(u,v,x,y)+\zeta f_1^{(1,2,j+1,j+2)}(u,v,x,y) \\
&+ (1-\zeta)f_0^{(1,2,j+1,j+2)}(x,y,u,v)+\zeta f_1^{(1,2,j+1,j+2)}(x,y,u,v)\big] \\
&- (2m+1)\big[(1-\zeta)f_0^{(1,2)}(u,v)f_0^{(1,2)}(x,y)+\zeta f_1^{(1,2)}(u,v)f_1^{(1,2)}(x,y)\big].
\end{aligned}
\tag{18}
$$

It is evident that λ determines the scaling of g, which is irrelevant to the performance, provided that the threshold is properly scaled. It seems most convenient to set $\lambda = \frac{1}{2}$, which yields the integral equation

$$
\begin{aligned}
f_1^{(1,2)}(x,y) - f_0^{(1,2)}(x,y) = &\, g(x,y)\big[(1-\zeta)f_0^{(1,2)}(x,y)+\zeta f_1^{(1,2)}(x,y)\big] \\
&+ \int g(u,x)K_1(u,x,y)\,du + \int g(y,u)K_1(x,y,u)\,du \\
&+ \iint g(u,v)K_2(x,y,u,v)\,du\,dv.
\end{aligned}
\tag{19}
$$

It is interesting to note that the generalized SNR (7) has a property similar to that of the SNR and the matched filter. If we multiply both sides of (19) by $g(x, y)$ and integrate, we find that, if g^* solves the integral equation, then (19) becomes

$$\mu_1(g^*) - \mu_0(g^*) = (1 - \zeta)\sigma_0^2(g^*) + \zeta\sigma_1^2(g^*).$$

This gives us the value of the performance measure

$$S_\zeta(g^*) = \mu_1(g^*) - \mu_0(g^*) \tag{20}$$

evaluated at the optimal nonlinearity.

Under mismatch conditions, which is the case when the nonlinearity g is not the optimal nonlinearity for the densities $f_i^{(1,2)}$, $f_i^{(1,2,3)}$, and $f_i^{(1,2,j+1,j+2)}$ involved in (17)-(19), the performance measure $S_\zeta(g)$ can be computed from (7) where the numerator is given by

$$\mu_1(g) - \mu_0(g) = \iint g(x, y)\big[f_1^{(1,2)}(x, y) - f_0^{(1,2)}(x, y)\big]\, dx\, dy \tag{21}$$

and the denominator by

$$\begin{aligned}(1 - \zeta)\sigma_0^2(g) + \zeta\sigma_1^2(g) = &\iint g^2(x, y)\big[(1 - \zeta)f_0^{(1,2)}(x, y) + \zeta f_1^{(1,2)}(x, y)\big]\, dx\, dy \\ &+ \iiint g(x, y)\big[g(u. x)K_1(u, x, y) + g(y, u)K_1(x, y, u)\big]\, du\, dx\, dy \\ &+ \iiiint g(x, y)g(u, v)K_2(x, y, u, v)\, dx\, dy\, du\, dv.\end{aligned} \tag{22}$$

The integral equation (19) generally cannot be solved for a closed form solution; numerical methods must be used. We recommend a normalization of (19) in order to facilitate a more accurate numerical solution by introducing the substitution $h(x, y) = g(x, y)\sqrt{D(x, y)}$, with $D(x, y) = (1 - \zeta)f_0^{(1,2)}(x, y) + \zeta f_1^{(1,2)}(x, y)$. The new integral equation

$$\begin{aligned}\frac{f_1^{(1,2)}(x, y) - f_0^{(1,2)}(x, y)}{\sqrt{D(x, y)}} = h(x, y) &+ \int \frac{h(u, x)K_1(u, x, y)}{\sqrt{D(u, x)D(x, y)}} du + \int \frac{h(y, u)K_1(x, y, u)}{\sqrt{D(y, u)D(x, y)}} du \\ &+ \iint \frac{h(u, v)K_2(x, y, u, v)}{\sqrt{D(u, v)D(x, y)}}\, du\, dv\end{aligned} \tag{23}$$

is more "balanced" and results in a more stable solution. In particular, the kernel $K_2(x,y,u,v)/\sqrt{D(u,v)D(x,y)}$ of the double integral is symmetric in the pairs (x,y) and (u,v).

To solve the integral equation numerically we introduce the nodes $\{x_0,\ldots,x_{N-1}\}$ and weights $\{w_0,\ldots,w_{N-1}\}$ from some numerical integration method (such as Simpson's rule). The numerical form of the integral equation (23) is then

$$\begin{aligned}\frac{f_1^{(1,2)}(x_i,x_j)-f_0^{(1,2)}(x_i,x_j)}{\sqrt{D(x_i,x_j)}} &= u_{ij}+\sum_{k=0}^{N-1}\frac{u_{ki}K_1(x_k,x_i,x_j)w_k}{\sqrt{D(x_k,x_i)D(x_i,x_j)}}+\sum_{k=0}^{N-1}\frac{u_{jk}K_1(x_i,x_j,x_k)w_k}{\sqrt{D(x_i,x_j)D(x_j,x_k)}}\\ &\quad+\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\frac{u_{kl}K_2(x_k,x_l,x_i,x_j)\,w_k\,w_l}{\sqrt{D(x_k,x_l)D(x_i,x_j)}}\end{aligned}\tag{24}$$

with $u_{i,j}\approx h(x_i,x_j)=g(x_i,x_j)\sqrt{D(x_i,x_j)}$. Eq. (24) can be solved as a system of N^2 linear equations in the N^2 variables u_{ij}, $i=0,\ldots,N-1$, $j=0,\ldots,N-1$. Note that, without the presence of the terms involving K_1, the matrix involved is symmetric.

The results of this section may be applied to ρ-mixing processes. The method involves approximating the ρ-mixing process by an m-dependent process; that is, one assumes that observations are independent if they are farther than m steps apart in time. The approximation is sufficient if m is sufficiently large. More precisely, if $g^{(m)}$ is the the nonlinearity obtained by solving the integral equation (19), then $S_\zeta(g^{(m)})\to\sup_g S_\zeta(g)$, as $m\to\infty$. A proof of this result was given in [5] for memoryless nonlinearities; the extension to one-step memory nonlinearities is straightforward.

III. Derivation of the Optimal Nonlinearity for Weak-Signal Detection

For weak-signal detection, the hypotheses are given by (3), where $\{Y_i\}$ an m-dependent noise process and $\theta>0$ is a constant signal. The n-variate densities are thus $f^{(n)}(\mathbf{x})$ and $f^{(n)}(\mathbf{x}-\underline{\theta})$, where $\underline{\theta}=(\theta,\theta,\cdots,\theta)$, under H_0 and H_1, respectively. The performance measure here is the efficacy functional (9). The method for maximizing $\eta(g)$ is similar to that for maximizing $S_\zeta(g)$.

Assuming that

$$\mu_0'(g) = \iint g(x,y) \frac{d}{d\theta} f^{(1,2)}(x-\theta, y-\theta)\big|_{\theta=0}\, dx\, dy \tag{25}$$

as well as some other regularity conditions [1], [2], we can obtain the integral equation

$$\begin{aligned} -\left[\frac{\partial}{\partial x} f^{(1,2)}(x,y) + \frac{\partial}{\partial y} f^{(1,2)}(x,y)\right] = g(x,y) f^{(1,2)}(x,y) \\ + \int \left[g(u,x) f^{(1,2,3)}(u,x,y) + g(y,u) f^{(1,2,3)}(x,y,u)\right] du \\ + \iint g(u,v) K(x,y,u,v)\, du\, dv \end{aligned} \tag{26}$$

where

$$\begin{aligned} K(x,y,u,v) = \sum_{j=2}^{m} \left[f^{(1,2,j+1,j+2)}(u,v,x,y) + f^{(1,2,j+1,j+2)}(x,y,u,v)\right] \\ - (2m+1) f^{(1,2)}(u,v) f^{(1,2)}(x,y) \end{aligned} \tag{27}$$

and $f^{(j,k)}$, $f^{(j,k,l)}$, and $f^{(j,k,l,p)}$ denote the densities of (Y_j, Y_k), (Y_j, Y_k, Y_l), and (Y_j, Y_k, Y_l, Y_p), respectively. This integral equation is also given in [2].

The efficacy functional of (25) is closely related to the generalized SNR functional of (7). Indeed, we can write

$$\eta(g) = \lim_{\theta \to 0} \frac{S_\zeta(g)}{\theta^2}$$

where $S_\zeta(g)$ is evaluated from (7) by using $f_1^{(1,2)}(x,y) = f^{(1,2)}(x-\theta, y-\theta)$, $f_0^{(1,2)}(x,y) = f^{(1,2)}(x,y)$, and similar relationships for the third- and fourth-order joint densities.

The efficacy has also a property similar to that of the SNR with the matched filter. If we multiply both sides of (26) by $g(x,y)$ and integrate, we find that, if g^* solves the integral equation, then (26) implies that

$$\mu_0'(g^*) = \sigma_0^2(g^*).$$

This gives the value

$$\eta(g^*) = \mu_0'(g^*) = -\iint g^*(x,y) \left[\frac{\partial}{\partial x} f^{(1,2)}(x,y) + \frac{\partial}{\partial y} f^{(1,2)}(x,y)\right] dx\, dy \tag{28}$$

of the efficacy evaluated at the optimal nonlinearity.

Under mismatch conditions, that is, when $g(x,y) \neq g^*(x,y)$, the efficacy $\eta(g)$ can be obtained from (9), where the numerator by

$$\mu_0'(g) = -\iint g(x,y)\left[\frac{\partial}{\partial x} f^{(1,2)}(x,y) + \frac{\partial}{\partial y} f^{(1,2)}(x,y)\right] dx\, dy \tag{29}$$

and the denominator is given by

$$\begin{aligned}\sigma_0^2(g) = &\iint g^2(x,y) f^{(1,2)}(x,y)\, dx\, dy \\ &+ \iiint g(x,y)\left[g(u,x) f^{(1,2,3)}(u,x,y) + g(y,u) f^{(1,2,3)}(x,y,u)\right] du\, dx\, dy \\ &+ \iiiint g(x,y) g(u,v) K(x,y,u,v)\, dx\, dy\, du\, dv.\end{aligned} \tag{30}$$

IV. A Lower Bound for Worst-Case Performance

In our work on memoryless discriminators [5], we were able to obtain a nonlinearity which is robust in a minimax sense. We have not been able to extend these results fully to the one-step memory case. What we have obtained is a lower bound on performance, in terms of the SNR performance measure, which holds for every distribution in the uncertainty classes. This is not a minimax robustness result because we have not been able to demonstrate the existence of least favorable distributions whose optimal nonlinearity achieves the lower bound. However, this information is useful, since the lower bound may be able to tell us that significant performance degradation from the nominal performance can be avoided by using the robust test. The treatment here is brief, since most of the work here is an extension of the results in [5].

The notation in this section is cumbersome. F_0 and F_1 denote the complete distributions under H_0 and H_1 (i.e. the complete set of finite dimensional distributions), while f_0 and f_1 denote the bivariate distributions of (X_1, X_2) under H_0 and H_1, respectively. We use $S_\zeta(g; F_0, F_1)$ to denote the performance measure (7) evaluated under the indicated distributions. A related performance measure $S'_\zeta(g; f_0, f_1)$, which involves only the bivariate distributions, is defined by

$$S'_\zeta(g; f_0, f_1) = \frac{|\mathrm{E}_1\, g(X_1, X_2) - \mathrm{E}_0\, g(X_1, X_2)|^2}{(1-\zeta)(1+2R_0)\mathrm{Var}_0\, g(X_1, X_2) + \zeta(1+2R_1)\mathrm{Var}_1\, g(X_1, X_2)}, \tag{31}$$

where $R_i = \sum_{j=2}^{m} r_{ij}$ are known constants. The parameters r_{ij} are defined below.

Uncertainty classes for the distributions are defined in the following way. The bivariate densities are assumed to belong to classes defined by ϵ-mixtures. (The results hold for other uncertainty classes, as well, such as the 2-alternating capacity classes [8]). Thus every bivariate density under H_i is assumed to have the form

$$f_i(x, y) = (1 - \epsilon_i) f_i^0(x, y) + \epsilon_i h_i(x, y) \tag{32}$$

where f_i^0 is a nominal density and h_i is unknown. The parameter $\epsilon_i \in (0, 1)$ determines the degree of uncertainty in the bivariate distribution. A further condition on the uncertainty classes, given below, completely defines them. This condition requires that the bound

$$\sup_g \frac{\mathrm{Cov}_i[g(X_1, X_2), g(X_{j+1}, X_{j+2})]}{\sqrt{\mathrm{Var}_i\, g(X_1, X_2)\mathrm{Var}_i\, g(X_{j+1}, X_{j+2})}} \leq r_{ij} \tag{33}$$

be satisfied for a known sequence r_{ij}, for $i = 0, 1$ and $j = 2, 3, \ldots, m$. The sequence of r parameters determines the degree of uncertainty in the higher-order distributions. Note that for ρ-mixing processes, we have $r_i, j \leq \rho_j$ for all j.

For the performance measure $S_\zeta(g; f_0, f_1)$, which involves only the bivariate distributions, we may obtain a complete robustness result. One can show [5] that the least-favorable bivariate densities $\hat{f}_0$, $\hat{f}_1$ are given by Huber [9]. These are

$$\begin{aligned}
\hat{f}_0(x, y) &= \begin{cases} (1 - \epsilon_0) f_0^0(x, y) & \text{if} \quad f_1^0(x, y)/f_0^0(x, y) < c'' \\ (1/c'')(1 - \epsilon_0) f_1^0(x, y) & \text{if} \quad f_1^0(x, y)/f_0^0(x, y) \geq c'' \end{cases} \\
\hat{f}_1(x, y) &= \begin{cases} (1 - \epsilon_1) f_1^0(x, y) & \text{if} \quad f_1^0(x, y)/f_0^0(x, y) > c' \\ c'(1 - \epsilon_1) f_0^0(x, y) & \text{if} \quad f_1^0(x, y)/f_0^0(x, y) \leq c' \end{cases}
\end{aligned} \tag{34}$$

where the constants c' and c'' are chosen so that the functions integrate to one. The nonlinearity $\hat{g}$ that is matched to these densities is minimax robust under the performance measure S'_ζ, that is, the inequalities

$$S'_\zeta(\hat{g}; f_0, f_1) \geq S'_\zeta(\hat{g}; \hat{f}_0, \hat{f}_1) \geq S'_\zeta(g; \hat{f}_0, \hat{f}_1) \tag{35}$$

are satisfied for any other densities f_0 and f_1 in the uncertainty class and for any other nonlinearity g. This nonlinearity is given by the solution to the integral equation

$$\hat{g}(x,y) = \frac{\hat{f}_1(x,y) - \hat{f}_0(x,y)}{A\hat{f}_1(x,y) + \hat{f}_0(x,y)} + \iint \frac{A\hat{f}_1(x,y)\hat{f}_1(u,v) + \hat{f}_0(x,y)\hat{f}_0(u,v)}{A\hat{f}_1(x,y) + \hat{f}_0(x,y)}\hat{g}(u,v)\, du\, dv \tag{36}$$

where

$$A = \frac{\zeta(1+2R_1)}{(1-\zeta)(1+2R_0)}.$$

The integral equation (36) may be obtained directly by the method used to obtain (19). The solution to this integral equation was obtained in [5] and is given by

$$\hat{g}(x,y) = \left[\iint \frac{\hat{f}_0(x,y)\hat{f}_1(x,y)}{A\hat{f}_1(x,y) + \hat{f}_0(x,y)}\, dx\, dy\right]^{-1} \frac{\hat{f}_1(x,y)}{A\hat{f}_1(x,y) + \hat{f}_0(x,y)} \tag{37}$$

We now obtain the lower bound for the complete uncertainty classes. The lower bound is given by

$$S_\zeta(\hat{g}; F_0, F_1) \geq S'_\zeta(\hat{g}; f_0, f_1) \geq S'_\zeta(\hat{g}; \hat{f}_0, \hat{f}_1). \tag{38}$$

The first inequality in (38) is obtained from the bounds (33) and the expression (6). The second inequality holds because of the minimax robustness of $\hat{g}$ over the ϵ-contamination classes for the bivariate densities, the proof of which is an extension of the proof given in [5] for univariate densities.

In order to demonstrate true minimax robustness for the complete uncertainty classes, one must construct least favorable multivariate distributions $\hat{F}_0$ and $\hat{F}_1$ that satisfy the bounds

$$S_\zeta(\hat{g}; F_0, F_1) \geq S_\zeta(\hat{g}; \hat{F}_0, \hat{F}_1) \geq S_\zeta(g; \hat{F}_0, \hat{F}_1) \tag{39}$$

for arbitrary distributions F_0, F_1, and an arbitrary nonlinearity g. The way to do this is to extend the least favorable bivariate densities $\hat{f}_0$ and $\hat{f}_1$ to multivariate distributions such that equality holds in (33) for each j. If this this were the case, then we would have the equality

$$S'_\zeta(\hat{g}; \hat{f}_0, \hat{f}_1) = S_\zeta(\hat{g}; \hat{F}_0, \hat{F}_1). \tag{40}$$

Then, combining (36) and (38) would yield the left inequality in (35). We could also obtain the optimal nonlinearity $\hat{g}$ by the method of Section II, and thus we would also have the right inequality. For memoryless nonlinearities, it is possible to construct such multivariate distributions [3] and thus prove minimax robustness [5]. It should be noted, however, that the distributions in the memoryless case possess a very peculiar property which is necessary for showing robustness inequalities. This property seems to be lost when dimensionality is extended.

V. Numerical Results

A. *Discrimination Between Rayleigh and Lognormal Correlated Time-Series*

In Section II, we provided a method for obtaining an approximate solution of the integral equation (19) by solving a system of linear equations (24). We used this method, with $N = 33$ nodes, to obtain (approximately) the optimal one-step memory nonlinearity for discrimination between a Rayleigh and lognormal process. The performance of the one-step memory discriminator was then evaluated through computer simulation. The results are presented here.

The discrimination problem we consider assumes a Rayleigh distribution under H_0 and a lognormal distribution under H_1. The Rayleigh process $\{X_k\}$ involves two underlying Gaussian processes $\{Y_k\}$ and $\{Z_k\}$ through the relation $X_k = \sqrt{Y_k^2 + Z_k^2}$. We assume that the distributions of Y_k and Z_k have mean $\mu_Y = \mu_Z = 0$, variance $\sigma_Y^2 = \sigma_Z^2 = 4$, and covariances $\mathrm{Cov}[Y_k, Y_{k+j}] = \mathrm{Cov}[Z_k, Z_{k+j}] = e^{-|j|/\tau_0}$, with $\tau_0 = 13$. The lognormal process $\{X_k\}$ is related to a Gaussian process $\{U_k\}$ through $X_k = \exp(U_k)$. We assume that $\mu_U = 0.8$, $\sigma_U^2 = 0.25$, and $\mathrm{Cov}[U_k, U_{k+j}] = e^{-|j|/\tau_1}$, with $r_1 = 130$. With these parameters, the Rayleigh and lognormal distributions have equal first and second moments. The tails are different, however, with the lognormal distribution having a heavier tail to the right (toward $+\infty$) and the Rayleigh distribution having a heavier tail to the left (toward zero). The correlation is considerably different, being very strong under H_1 (lognormal) and moderate under H_0 (Rayleigh). We assume that an m-dependent model with $m = 300$ is sufficient approximation for either hypothesis. The forms of the Rayleigh and lognormal second-, third-, and fourth-order densities are given in Appendix A.

Figure 1 shows the optimal one-step memory nonlinearity, which is symmetric in x_i and x_{i+1}. Figures 2 and 3 show plots of P_1 vs. P_0, where P_i is the probability of error when H_i is true, for three different discriminators. These discriminators are denoted $T_n^{(0)}$ for the i.i.d. discriminator (Eq. (1) with $g(x) = \log[f_1(x)/f_0(x)]$); $T_n^{(1)}$ for the memoryless discriminator (1) that is optimal under the performance measure (7) with $\zeta = \frac{1}{2}$; and $T_n^{(2)}$ for the optimal one-step memory discriminator obtained from the integral equation (19) with $\zeta = \frac{1}{2}$. The memoryless discriminator $T_n^{(1)}$ is obtained through the solution of an integral equation similar to (19), as shown in [5]. Figure 2 was generated from 10^6 simulations with $n = 25$ samples for the test statistic. The superiority of $T_{25}^{(2)}$ is evident. This superiority is even more evident in Figure 3, also generated from 10^6 simulations but with $n = 50$ samples.

Generally, 10^6 trials is not sufficient to accurately approximate the error probabilities unless importance sampling is used. Since we did not use importance sampling, the curve for $T_n^{(2)}$ in Figure 3 should not be regarded as an accurate estimation of the true error probability curve. Nevertheless, the plot does indicate a substantial improvement in performance obtained by the one-step memory discriminator over the memoryless discriminator.

B. Detection of a Weak-Signal in Correlated Cauchy Noise

The univariate density for Cauchy noise is $f(x) = [\pi(1 + x^2)]^{-1}$. This distribution has extremely heavy tails and is a good model for impulsive noise. Correlated Cauchy noise may be obtained from a memoryless transformation of a Gaussian process. For our model, $\{Z_k\}$ is assumed Gaussian with zero mean, unit variance, and covariance $\mathrm{Cov}[Z_k, Z_{k+j}] = e^{-|j|/\tau}$, and the Cauchy noise process is obtained by $Y_k = \tan\left[\pi \operatorname{erf}(Z_k/\sqrt{2})/2\right]$. The forms of the Cauchy second-, third-, and fourth-order densities are given in Appendix B.

Figures 4 and 5 show graphs of the nonlinearities for $\tau = 1$ and $\tau = 15$. Table 1 shows the value of the efficacy for several detectors and values of τ. The values are normalized in Table 2 to show the ARE to the locally optimal i.i.d. (memoryless) detector with $g(x) = -f'(x)/f(x)$. The Markov detector is a one-step memory detector with

$$\begin{aligned} g(x,y) &= \frac{d}{d\theta} \log \left[\frac{f(x-\theta, y-\theta) f(x)}{f(x,y) f(x-\theta)} \right]_{\theta=0} \\ &= -\frac{\frac{\partial}{\partial x} f(x,y) + \frac{\partial}{\partial y} f(x,y)}{f(x,y)} + \frac{f'(x)}{f(x)}, \end{aligned}$$

which is the locally optimal detector if the noise process is Markov. It is interesting to note that the efficacy of the memoryless detectors decreases with increasing τ, while the efficacy of the one-step memory detectors increases. Since the Markov detector behaves like a memoryless detector for relatively weak correlation and a one-step memory detector for stronger correlation, its efficacy at first decreases and then increases, as τ increases.

VI. Conclusions

The contribution of our research has been to extend results obtained for memoryless discriminators to discriminators involving one-step memory. In Section II, we showed that the optimal one-step memory nonlinearity can be obtained as the solution of the linear integral equation (19). We also showed how an approximate solution of the integral equation can be obtained by solving a system of linear equations. This numerical approximation actually involves a quantization of the nonlinearity, which we do not claim to be optimal. However, our simulation results demonstrate a significant improvement over memoryless discriminators even for a crude solution with $N = 33$ nodes. Evidently, any loss of accuracy in the approximation was not significant enough to override the discrimination power contributed by the memory. This numerical approximation to the one-step memory nonlinearity involved the solution of a 1089×1089 system of linear equations, which is well within the computation capability of many computer systems. In our preliminary results, we used only 17 nodes for the integration and still obtained a significant improvement over the optimal memoryless discriminator. We actually used interpolation to generate values of g, though, and not quantization, in our simulations. Results on optimal quantization may be possible, although they are more complex than results for quantization of the memoryless nonlinearities.

We have also been able to apply our results to the problem of weak-signal detection. We showed that the optimal one-step memory nonlinearity for detection under the ARE criterion is given by the solution of the linear integral equation (26). The performance of the optimal one-step memory detector as compared to the optimal memoryless detector under the ARE performance measure is significantly better and improves with increasing correlation.

Results for robustness for memoryless discriminators do not completely extend to the case of memory, as we showed in Section IV. What we obtained there is a one-step memory nonlinearity $\hat{g}$ and a lower bound on its performance when the distributions are known to belong to given uncertainty classes.

VII. References

[1] H. V. Poor and J. B. Thomas, "Memoryless Discrete-Time Detection of a Constant Signal in m-Dependent Noise," *IEEE Trans. Inform. Theory,* vol. IT-25, pp. 54–61, January 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, March 1980.

[3] J. S. Sadowsky, "A Maximum Variance Model for Robust Detection and Estimation with

Dependent Data," *IEEE Trans. Inform. Theory,* vol. IT-32, pp. 220–226, March 1986.

[4] J. S. Sadowsky and J. A. Bucklew, "A Nonlocal Approach for Asymptotic Memoryless Detection Theory," *IEEE Trans. Inform. Theory,* vol. IT-32, pp. 115–120, January 1986.

[5] D. Sauder and E. Geraniotis, "Optimal and Robust Memoryless Discrimination From Dependent Observations," *IEEE Trans. Inform. Theory,* vol. IT-37, pp. 73–91, January 1991.

[6] M. Peligrad, "Recent Advances in the Central Limit Theorem and Its Weak Invariance Principle for Mixing Sequences of Random Variables," in *Dependence in Probability and Statistics*, Birkhauser, 1985.

[7] H. V. Poor, *An Introduction to Signal Detection and Estimation*, New York: Springer-Verlag, 1988.

[8] P. J. Huber and V. Strassen, " Minimax Tests and the Neyman-Pearson Lemma for Capacities," *Ann. Statistics*, vol. 1, pp. 251–263, 1973.

[9] P. J. Huber " A Robust Version of the Probability Ratio Test," *Ann. Math. Stat.*, vol. 36, pp. 1753–1758, 1965.

Appendix A

Second, Third, and Fourth Order Joint pdfs of the Rayleigh and Lognormal Observations Used in Section V.A

For the **Rayleigh** case (under hypotheses H_0) the bivariate joint pdf is

$$f_0(z,w) = \frac{zw}{(\sigma_0^2)^2(1-\rho_0^2)} \exp\left\{-\frac{z^2+u^2}{2(1-\rho_0^2)\sigma_0^2}\right\} I_0\left[\frac{\rho_0 zw}{(1-\rho_0^2)\sigma_0^2}\right]$$

Furthermore, the third order joint pdf

$$f_0(z,w,v) = \frac{zwv}{(\sigma_0^2)^3(1-\rho_0^2)^2} \exp\left\{-\frac{1}{2(1-\rho_0^2)\sigma_0^2}[z^2+(1+\rho_0^2)w^2+v^2]\right\} I_0\left[\frac{\rho_0 zw}{(1-\rho_0^{2l})\sigma_0^2}\right] I_0\left[\frac{\rho_0 wv}{(1-\rho_0^2)\sigma_0^2}\right]$$

Finally, the fourth-order joint pdf of $(X_1, X_2, X_{l+1}, X_{l+2})$ is

$$f_0(z,w,v,u;l) = \frac{zwvu}{(\sigma_0^2)^4(1-\rho_0^2)^2(1-\rho_0^{2(l-1)})} \exp\left\{-\frac{1}{2(1-\rho_0^2)\sigma_0^2}\left[z^2+u^2+\frac{1-\rho_0^{2l}}{1-\rho_0^{2(l-1)}}(w^2+v^2)\right]\right\}$$
$$\cdot I_0\left[\frac{\rho_0 zw}{(1-\rho_0^2)\sigma_0^2}\right] I_0\left[\frac{\rho_0^{l-1} wv}{(1-\rho_0^{2(l-1)})\sigma_0^2}\right] I_0\left[\frac{\rho_0 vu}{(1-\rho_0^2)\sigma_0^2}\right]$$

Similarly, for the **lognormal** case (under hypothesis H_1) the bivariate joint pdf is

$$f_1(z,w) = \frac{1}{2\pi\sigma_1^2\sqrt{1-\rho_1^2}} \frac{1}{zw} \exp\left\{-\frac{1}{2(1-\rho_1^2)\sigma_1^2}[(\ln z-\mu_1)^2+(\ln w-\mu_1)^2-2\rho_1(\ln z-\mu_1)(\ln w-\mu_1)]\right\}$$

and the third-order joint pdf is

$$f_1(z,w,v) = \frac{1}{(2\pi\sigma_1^2)^{3/2}(1-\rho_1^2)} \frac{1}{zwv} \exp\left\{-\frac{1}{2(1-\rho_1^2)\sigma_1^2}\left[(\ln z-\mu_1)^2+(1+\rho_{j,1}^2)(\ln w-\mu_1)^2+(\ln v-\mu_1)^2\right.\right.$$
$$\left.\left. -2\rho_1[(\ln z-\mu_1)(\ln w-\mu_1)+(\ln w-\mu_1)(\ln v-\mu_1)]\right]\right\}$$

Finally, the fourht-order joint pdf of $(X_1, X_2, X_{l+1}, X_{l+2})$ is given by

$$
\begin{aligned}
f_1(z,w,v,u;l) = {} & \frac{1}{(2\pi\sigma_1^2)^2(1-\rho_1^2)\sqrt{1-\rho_1^{2(l-1)}}}\,\frac{1}{zwvu}\exp\Bigg\{-\frac{1}{2(1-\rho_1^2)\sigma_1^2}\Bigg[(\ln z-\mu_1)^2+(\ln u-\mu_1)^2 \\
& + \frac{1-\rho_1^{2l}}{1-\rho_1^{2(l-1)}}[(\ln w-\mu_1)^2+(\ln v-\mu_1)^2] \\
& - 2\rho_1[(\ln z-\mu_1)(\ln w-\mu_1)+(\ln v-\mu_1)(\ln u--\mu_1 \\
& - 2\frac{\rho_1^{l-1}(1-\rho_1^2)}{1-\rho_1^{2(l-1)}}(\ln w-\mu_1)(\ln v-\mu_1)\Bigg]\Bigg\}
\end{aligned}
$$

Appendix B

Second, Third, and Fourth Order Joint Pdfs of the Cauchy Noise Observations Used in Section V.B

We start with the Gaussian bivariate, third-order, and fourth-order joint pdfs, namely

$$f_G(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[x^2 - 2\rho xy + y^2]\right\}$$

$$f_G(u,x,y) = \frac{1}{(2\pi)^{3/2}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}[u^2 + (1+\rho^2)x^2 + y^2 - 2\rho ux - 2\rho xy]\right\}$$

$$\begin{aligned} f_G^{(1,2,k+1,k+2)}(u,v,x,y) &= \frac{1}{(2\pi)^2(1-\rho^2)\sqrt{1-\rho^{2k-2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}(uvxy)^T \Lambda (uvxy)\right\} \\ &= \frac{1}{(2\pi)^2(1-\rho^2)\sqrt{1-\rho^{2k-2}}} \exp\left\{-\frac{u^2 + sv^2 + sx^2 + y^2 - 2\rho uv - 2rvx - 2\rho xy}{2(1-\rho^2)}\right\} \end{aligned}$$

where $\Lambda = \begin{Bmatrix} 1 & -\rho & 0 & 0 \\ -\rho & s & -r & 0 \\ 0 & -r & s & -\rho \\ 0 & 0 & -\rho & 1 \end{Bmatrix}$ with $s = \frac{1-\rho^{2k}}{1-\rho^{2k-2}}$ and $r = \frac{\rho^{k-1}(1-\rho^2)}{1-\rho^{2k-2}}$

Suppose X is Gaussian distributed and $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$, then the transformation $X = \psi(Y)$ where

$$\psi(x) = \Phi^{-1}\left[\frac{1}{\pi}\tan^{-1} x + \frac{1}{2}\right],$$

implies that Y is Cauchy distributed; the inverse transformation $Y = \psi^{-1}(X)$ is given by

$$\psi^{-1}(x) = \tan\left(\left[\Phi(x) - \frac{1}{2}\right]\pi\right).$$

The first and second order derivatives of $\psi(x)$ are given by

$$\psi'(x) = \frac{f(x)}{f_G[\psi(x)]} = \frac{[\pi(1+x^2)]^{-1}}{\frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}\psi^2(x)]} = \sqrt{\frac{2}{\pi}}\frac{\exp[\frac{1}{2}\psi(x)^2]}{1+x^2}$$

and

$$\psi''(x) = \sqrt{\frac{2}{\pi}} \frac{\exp[\frac{1}{2}\psi(x)^2]\psi(x)\psi'(x)(1+x^2) - 2x\exp[\frac{1}{2}\psi(x)^2]}{(1+x^2)^2}$$

$$= \psi(x)[\psi'(x)]^2 - \frac{2x}{1+x^2}\psi'(x)$$

Furthermore, the bivariate, third-order, and fourth-order joint pdfs are given by

$$f(x,y) = f_G[\psi(x),\psi(y)]\psi'(x)\psi'(y),$$

$$f(u,x,y) = f_G[\psi(u),\psi(x),\psi(y)]\psi'(u)\psi'(x)\psi'(y),$$

and

$$f^{(1,2,k+1,k+2)}(u,v,x,y) = f_G[\psi(u),\psi(v),\psi(x),\psi(y)]\psi'(u)\psi'(v)\psi'(x)\psi'(y),$$

respectively.

Moreover, the partial derivative $\frac{\partial}{\partial x}f(x,y)$ can be computed as

$$\frac{\partial}{\partial x}f(x,y) = f_{G,1}[\psi(x),\psi(y)]\psi'(x)^2\psi'(y) + f_G[\psi(x),\psi(y)]\psi''(x)\psi'(y)$$

where $\psi'(x)$ and $\psi''(x)$ have been evaluated above and

$$f_{G,1}(u,v) = \frac{d}{du}f_G(u,v) = f_G(u,v)\left(\frac{\rho v - u}{1-\rho^2}\right).$$

Consequently, we can deduce that

$$\frac{\partial}{\partial x}f(x,y) = f_G[\psi(x),\psi(y)]M(x,y)\psi'(x)\psi'(y) = f(x,y)M(x,y)$$

where

$$M(x,y) = \frac{\rho\psi(y) - \psi(x)}{1-\rho^2}\psi'(x) + \psi(x)\psi'(x) - \frac{2x}{1+x^2}$$

Similarly,

$$\frac{\partial}{\partial y}f(x,y) = f(x,y)M(y,x)$$

and thus finally

$$\frac{\partial}{\partial x}f(x,y) + \frac{\partial}{\partial y}f(x,y) = [M(x,y) + M(y,x)]f(x,y)$$

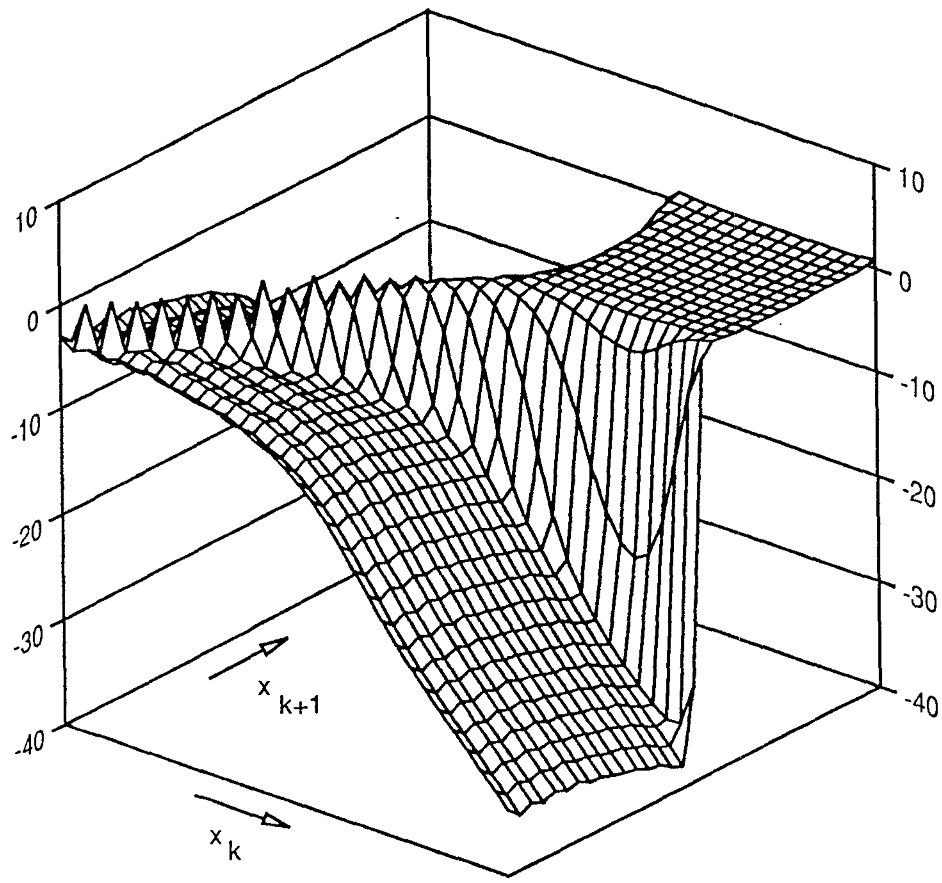

Figure 1. Graph over the region [0,16]x[0,16] of the optimal one-step memory nonlinearity for discrimination of Rayleigh vs. lognormal processes.

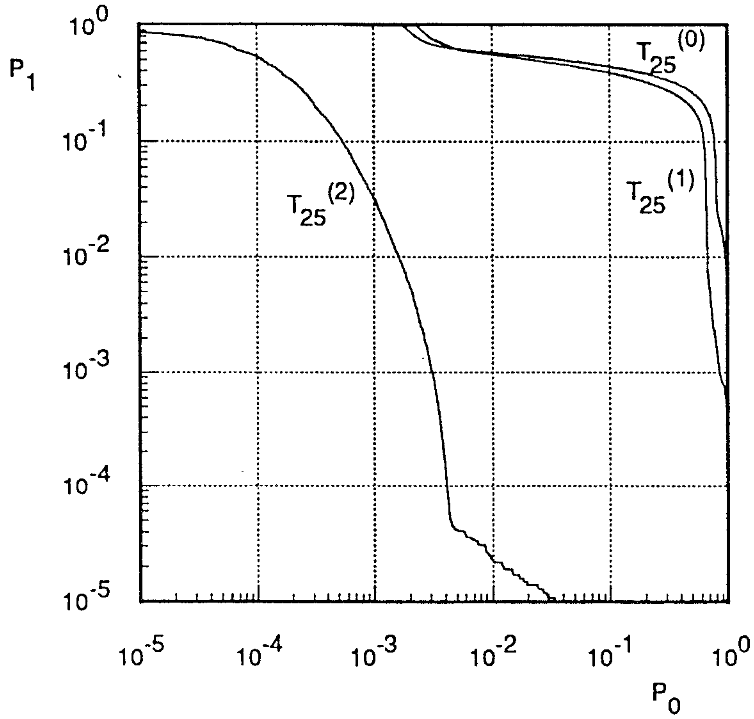

Figure 2. Plots of the error probabilities for the i.i.d. discriminator $T^{(0)}$, the optimal memoryless discriminator $T^{(1)}$, and the optimal one-step memory discriminator $T^{(2)}$ for a sample size of n = 25.

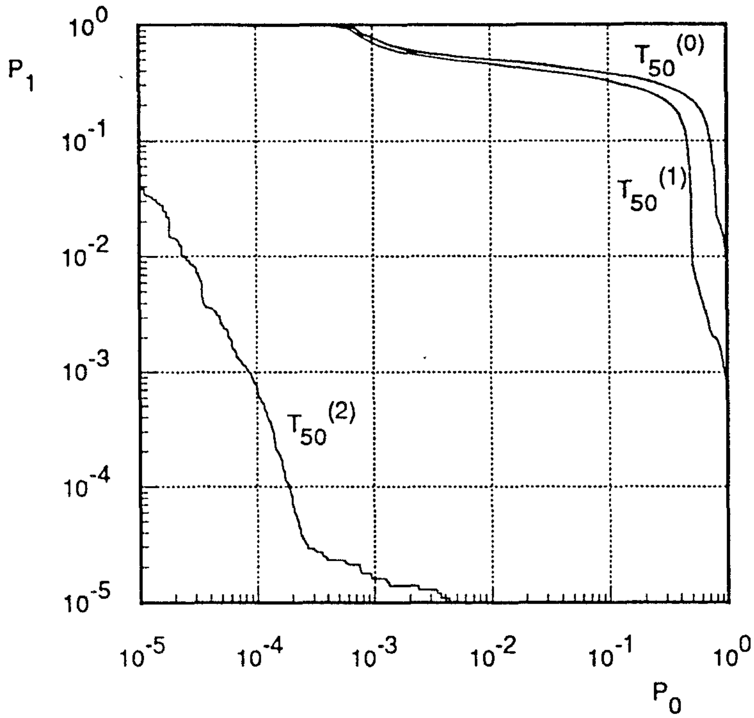

Figure 3. Plots of the error probabilities for the i.i.d. discriminator $T^{(0)}$, the optimal memoryless discriminator $T^{(1)}$, and the optimal one-step memory discriminator $T^{(2)}$ for a sample size of $n = 50$.

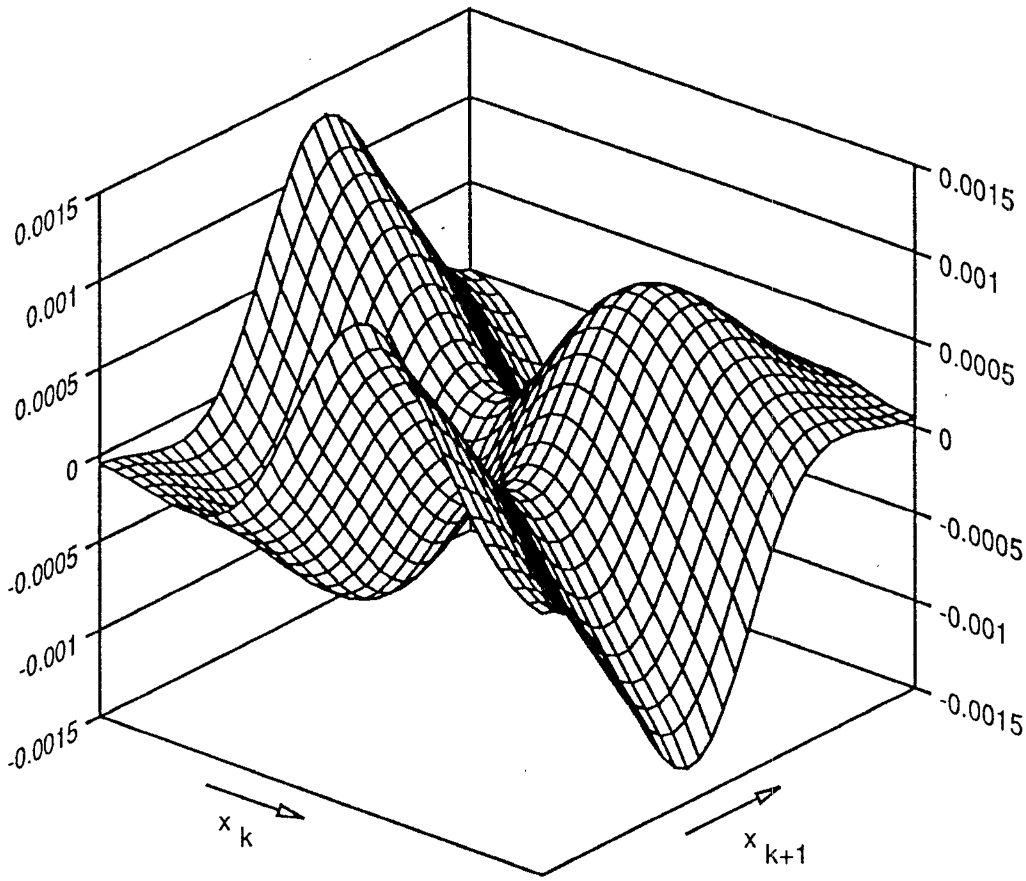

Figure 4. Graph over the region [-233,233]x[-233,233] of the optimal one-step memory nonlinearity for signal detection in Cauchy noise with $\tau = 1$.

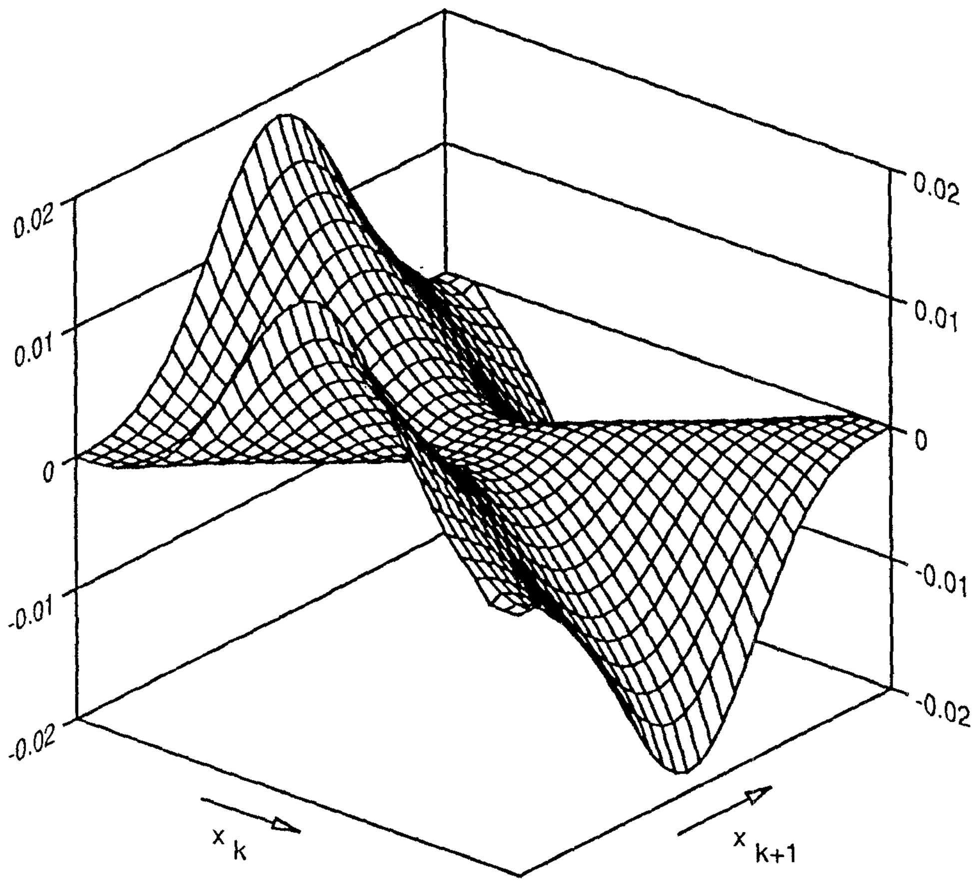

Figure 5. Graph over the region [-233,233]x[-233,233] of the optimal one-step memory nonlinearity for signal detection in Cauchy noise with $\tau = 15$.

Table 1. Efficacy

τ	i.i.d. L.O.	optimal memoryless	Markov L.O.	optimal one-step
1	0.3320	0.3733	0.4003	0.4242
5	0.0922	0.1512	0.3921	0.4253
15	0.0316	0.0567	0.4138	0.4260
30	0.0158	0.0288	0.4225	0.4270

Table 2. ARE

τ	i.i.d. L.O.	optimal memoryless	Markov L.O.	optimal one-step
1	1.000	1.124	1.206	1.278
5	1.000	1.640	4.253	4.614
15	1.000	1.794	13.10	13.48
30	1.000	1.817	26.67	26.95