Philippine Engineering Journal (1990) XI(2): 1-12

MOST POWERFUL AND LOCALLY OPTIMUM DETECTION IN ADDITIVE NOISE

Efren F. Abaya, Ph.D.
Associate Professor
Department of Electrical Engineering
University of the Philippines
Diliman, Quezon City

ABSTRACT

A fundamental problem in many types of digital communication receivers is that of discriminating between the presence or absence of a signal of known form mixed with additive noise. This paper considers the structure of two detectors designed for different optimality criteria respectively known as the most powerful (or Neyman-Pearson) detector and the locally optimum detector. It is shown that an affine relation between the nonlinearities of the two detectors is a necessary and sufficient condition for the two detectors to reduce to identical forms. This property is used to search for noise densities which have common detector structures.

INTRODUCTION

A fundamental problem in many types of digital communication receivers is that of discriminating between the presence or absence of a signal of known form mixed with additive noise. Oftentimes, the decision process is mathematically modelled as a statistical test between two hypotheses. Then the design of an optimum receiver or detector becomes, to a first approximation, equivalent to the design of an optimum statistical test.

This paper considers the structure of two detectors designed for different optimality criteria respectively known as the most powerful (or Neyman-Pearson) detector and the locally optimum detector. Conditions are derived under which the two detectors reduce to the same detector structure. Stated differently, the conditions derived imply that the underlying statistical tests are identical.

A most powerful detector can be employed in situations where the signal shapes and amplitudes are completely known (so-called "sure" signal). On the other hand, a locally optimum detector has been proposed for detecting small signals of unknown (possibly variable) amplitude. Generally, these two detectors take very different mathematical forms or physical structures. However, if the same detector structure should happen to be optimum for both criteria, then the designer has a very favorable situation in the sense that desirable detection properties are obtained with both large and small signals.

It is shown that an affine relation between the nonlinearities of the two detectors is a necessary and sufficient condition for the two detectors to reduce to identical forms. It follows from this that a first order linear difference-differential equation with retarded argument given in Eq. (16) is a necessary and sufficient condition for a detector to be simultaneously optimum in both senses for the problem of detecting a signal in additive noise with statistically independent samples.

For the case of a constant signal in independent, identically distributed noise samples the difference-differential equation gives two noise probability density functions--the normal or Gaussian pdf and the extreme value distributions, including the Gumbel distribution--as the pdf's for which both detectors become identical. Except for these noise pdf's therefore, it seems that the designer must choose between the two design criteria.

This paper is organized as follows. The next section reviews the relevant theories of most powerful and locally optimum detectors and defines the notation and terminology used in the paper. Section III defines the concept of a bioptimum detector and derives the principal existence results stated as a difference-differential equation. Section IV solves the difference-differential equation and derives the noise pdf's mentioned above.

REVIEW OF DETECTION THEORY

The statistical approach to the problem of detecting a known time-varying signal in additive noise is briefly described below to introduce the terminology and notation used in this paper.

The Detection Problem. A received waveform at the receiver input is sampled at discrete intervals to produce a series of measurements described by the sample vector

$$\tilde{X} = (x_1, x_2, \ldots x_m)^T.$$

In one case, the received signal consists only of noise, so that the following null hypothesis H_0 holds:

$$H_0: \; x_i = n_i \qquad i = 1, 2, \ldots m \tag{1}$$

where the n_i are statistically independent random variables with known probability density functions (usually assumed to have zero means)

$$f_0(\tilde{X}) = \prod_{i=1}^{m} f_{0i}(x_i) \tag{2}$$

In the other case, the received signal consists of a known signal plus the additive noise, so that the following alternate hypothesis H_1 holds:

$$H_1: \; x_i = \theta s_i + n_i \qquad i = 1, 2, \ldots m \tag{3}$$

where θ is a positive amplitude parameter for the transmitted signal sequence $S = (s_1, s_2, \ldots s_m)^T$. Then the pdf of the samples can be written as:

$$\tilde{f}_1(X;\theta) = f_0(\tilde{X} - \theta\tilde{S})$$

$$= \sum_{i=1}^{m} f_{0i}(x_i - \theta s_i) \qquad (4)$$

in which the dependence of f_1 on the parameter θ is explicitly shown for emphasis.

The detection problem of the communication receiver is to decide on the basis of the sample vector $\tilde{X}$ whether the desired signal $\tilde{S}$ is absent (H_0) or present (H_1).

The detection problem described above models the reception of unipolar digital pulses. By a simple modification, it may also be used for other common signalling methods such as Manchester coding or frequency shift keying (FSK).

A detector (henceforth used synonymously with hypothesis test) separates the m-dimensional space of samples X into two disjoint and exhaustive regions--the acceptance region R_0 leading to a decision to accept the null hypothesis, and the rejection or critical region R_1 leading to a decision to accept the alternate hypothesis. A decision may result in two kinds of errors. If the null hypothesis is falsely rejected, a Type I (false alarm) error is committed, which occurs with false alarm probability (size of the test)

$$\alpha = \int_{R_1} f_0(\tilde{X})\, d\tilde{X} \qquad (5)$$

If the null hypothesis is falsely accepted, a Type II (miss) error occurs with probability

$$1 - \beta(\theta) = \int_{R_0} f_1(\tilde{X};\theta)\, d\tilde{X} \qquad (6)$$

The power function $\beta(\theta)$ measures the performance of the detector with different signal parameter values.

Most Powerful Detectors. The so-called Neyman-Pearson criterion [8,11] maximizes the power function for a given amplitude parameter θ (i.e., minimizes the probability of a Type II error) while maintaining the false alarm probability not greater than a specified value. This detector is used when the signal shapes and amplitudes are completely known.

The resulting most powerful (or Neyman-Pearson) detector has a critical region composed of all sample vectors satisfying

$$L = \log \frac{f_1(\tilde{X};\theta)}{f_0(\tilde{X})} > t_{NP} \tag{7}$$

Using Eqs. (2-3), the log likelihood ratio L may also be expressed as a sum of nonlinear functions (or nonlinearities) in the following manner:

$$L = \sum_{i=1}^{m} g_{NPi}(x_i;\theta) > t_{NP}$$

$$g_{NPi}(x_i;\theta) = \log \frac{f_{0i}(X_i - \theta_{si})}{f_{0i}(X_i)} \tag{8}$$

The detection threshold t_{NP} is chosen so that the size of the test as given by Eq. (5) is α for the fixed and known parameter θ.

If the same detector (7) is most powerful for all (positive) values of θ, then it is said to be "uniformly most powerful".

The detector just described is an example of a more general class of likelihood ratio tests that compare the likelihood ratio L against a threshold as in Eq. (7). This class includes maximum a posteriori (MAP) or minimum probability of error detectors, Bayes detectors, and minimax detectors [10]. Although designed to optimize different criteria, they differ only in the values used for the threshold of Eqs. (7-8).

__Locally Optimum Detectors.__ A different situation occurs when θ is small but has unknown magnitude, e.g., a weak signal in a strongly fading channel. A robust type of detector for this problem maximizes the slope of the power function $\beta(\theta)$ at the origin while maintaining a specified false alarm probability [4,8]. Subject to some simple regularity conditions [2,4], this so-called locally optimum detector has the following critical region

$$L_0 = \left. \frac{\frac{\partial}{\partial\theta} f_1(\tilde{X};\theta)}{f_0(\tilde{X})} \right|_{\theta=0} > t_{LO} \tag{9}$$

$$L_0 = \left. \frac{\partial}{\partial\theta} \log f_1(\tilde{X};\theta) \right|_{\theta=0}$$

This function may also be expressed in sum form as

$$L_0 = \sum_{i=1}^{m} g_{LOi}(x_i) > t_{LO}$$

$$g_{LOi}(x_i) = \frac{\partial}{\partial \theta} \log f_{0i}(x_i - \theta s_i) \Big|_{\theta=0}$$

$$= -s_i f'_{0i}(x_i)/f_{0i}(x_i) \qquad (10)$$

The detection threshold t_{LO} is chosen so that the size of the test is α.

Note that the two detectors described by Eqs. (8, 10) both have the nonlinearity-summer-comparator structure commonly found in the literature [7], differing only in the nonlinearities g_{NPi} and g_{LOi} and perhaps the thresholds tNP and t_{LO}.

DERIVATION OF DIFFERENCE-DIFFERENTIAL EQUATION

This paper characterizes the conditions under which the most powerful and locally optimum detectors reduce to the same structure or test. That is, both have the same nonlinearities and critical regions, and therefore the same detection performance.

Define a detector with nonlinearity-summer-comparator structure to be **bioptimum** for testing

$$H_0: \theta = 0$$

$$H_1: \theta = \theta' > 0 \qquad (11)$$

if for all θ' under H_1, the detector is simultaneously most powerful and locally optimum.

Note that a detector is bioptimum if and only if it is uniformly most powerful for discriminating between the two hypotheses.

The following result relates the most powerful and locally optimum nonlinearities for a bioptimum detector.

Theorem 1. In independent noise with at least two samples, a detector is bioptimum if and only if for each i = 1,2,...m

$$g_{NPi}(x_i;\theta) = a(\theta)\, g_{LOi}(x_i) + b_i(\theta) \qquad (12)$$

where $a(\theta)$ is a positive function of θ only.

Proof of Sufficiency. Suppose Eq. (12) is true. Then the critical region of the most powerful detector can be expressed in the following form by using Eq. (8)

$$L_{NP} = \sum_{i=1}^{m} g_{LOi}(x_i) > \frac{t_{NP} - \Sigma\, b_i(\theta)}{a(\theta)}$$

$$= t'_{NP} \qquad (13)$$

Notice that $L_{NP} = L_O$. Now for any $\theta > 0$ the thresholds t_{LO} and t'_{NP} are determined using Eq. (5). Since this integral is independent of θ, it follows that the thresholds are equal and that the given detector is bioptimum.

Proof of Necessity. To show the only if part, assume that a detector is bioptimum. Then Eqs. (8) and (10) describe the same critical region and it should be possible to algebraically transform one inequality to the other. That is, there should exist a transformation T such that

$$T\left(\sum_{i=1}^{m} g_{NPi}(x_i;\theta)\right) = \sum_{i=1}^{m} g_{LOi}(x_i) \qquad (14)$$

This functional equation was studied in [1] where the result paraphrased below was proven. For consistency with the present discussion, the notation of this paper has been adopted.

> Assume that the ranges of each of the functions g_{NPi} contain a nondegenerate interval about the origin. If the transformation T has a property that excludes the density of its graph in the real plane, then the general form of T is given by $T(z) = cz + b$ for some constants c and b. Furthermore, there exist constants b_i such that the g_{NPi} and g_{LOi} are linearly related as in Eq. (12).

There are many easily obtained conditions on T that would exclude the density of its graph in the real plane, including

- continuity at a single point; or
- monotonicity in an arbitrarily small interval; or
- measurability on an arbitrarily small interval [9].

Therefore, the postulated property of T is easily attained.

The conditions on the ranges of g_{NPi} are likewise easy to verify. For the pdf's in the detection problem must be differentiable (almost everywhere) in order for a locally optimum detector (10) to exist. This means that the nonlinearities in (8) are all continuous. The point $g_{NPi} = 0$ corresponds to an intersection of the graphs of the pdf's under the null and alternate hypotheses, which occurs at or near the threshold values of most detection problems.

Since both conditions of the paraphrased result are satisfied, the desired equation (12) holds. QED

The condition posed in Thm. 1 significantly restricts the possible forms of noise pdf's for the additive noise detection problem, as the following result shows.

Theorem 2. A bioptimum detector for the detection problem (11) exists only if the pdf's $f_{0i}(x_i - \theta s_i)$ under the alternate hypothesis are all members of the exponential family of pdf's described by [6]

$$f(x;\theta) = \exp(Q(\theta)R(x) + P(x) + S(\theta)) \tag{15}$$

Proof. Assume that a bioptimum detector exists. By Eqs. (8,12)

$$f_{0i} (x_i - \theta s_i) = \exp(a(\theta) g_{LOi}(x_i) + b_i(\theta) + \ln f_{0i}(x_i))$$

showing that the pdf is from the exponential class. QED

The main result on the forms of nonlinearity functions for bioptimum detectors (which implicitly determines the forms of pdf's also) is given by the following.

Theorem 3. In additive noise with independent samples, a necessary and sufficient condition for a bioptimum detector for (11) to exist is that the detector nonlinearities satisfy the difference-differential equations

$$a\, s_i g'_{Loi}(x) = g_{LOi}(x) - g_{LOi}(x - \theta s_i) \tag{16}$$

for any x and constants $a > 0$, $\theta > 0$ and s_i. (The notation g' denotes the first derivative.)

Proof. The noise density for which g_{LOi} is locally optimum can be obtained by solving Eq. (10) as a first-order differential equation. When this is done we get

$$f_{0i}(x_i) = K \exp(- \int_{-\infty}^{x_i} g_{LOi}(z)\, dz / s_i) \tag{17}$$

where K is a normalizing constant. Substituting this into Eq. (8) gives a relation between the most powerful and locally optimum nonlinearities at the i-th sampling time:

$$g_{NPi}(x;\theta) = \ln f_{0i}(x - \theta s_i) - \ln f_{0i}(x)$$

$$= \int_{x-\theta s_i}^{x} g_{LOi}(z)\, dz / s_i \tag{18}$$

This integral equation can be differentiated to get an equivalent relation:

$$g_{NPi}(x;\theta) = [\, g_{LOi}(x) - g_{LOi}(x - \theta s_i) \,]/s_i \qquad (19)$$

The above equation is true for any detection problem. For bioptimum detectors, Thm. 1 gives a second necessary and sufficient relation that must be satisfied. Combining Eqs. (12) and (19) and letting $a = a(\theta)$ yields Eq. (16) as a necessary and sufficient condition for a bioptimum detector to exist in independent noise problems.
QED

SOLUTION OF DIFFERENCE-DIFFERENTIAL EQUATIONS

Eq. (16) belongs to the well-studied class of first order homogeneous linear difference-differential equations (LDDE) with retarded argument, also known as delay LDDE [3]. For the succeeding development, we assume a constant signal case, so that signal samples may be normalized to $s_i = 1$. At the end of this section, we will consider the more general case of time-varying signals. Then we have the delay LDDE

$$a\, g'_{LO}(x) = g_{LO}(x) - g_{LO}(x - \theta)$$

$$a,\ \theta > 0 \qquad (20)$$

subject to the condition that the solutions correspond to valid pdf's. That is, Eq. (17) should be integrable as follows

$$\int_{-\infty}^{\infty} \exp - \int_{-\infty}^{x} g_{LO}(z)\, dz\, dx < \infty \qquad (21)$$

Derivation of Series Solution. It is known that the delay LDDE has a continuous solution in the range $x > \theta$ that has a continuous first derivative for $x > \theta$ and a continuous second derivative for $x > 2\theta$. Moreover, a linear combination of solutions is also a valid solution to the same delay LDDE [3].

In fact, the function exp(sx) is a solution if and only if s is a complex root of the transcendental characteristic equation

$$a \cdot s \cdot \exp(sx) = \exp(sx) - \exp(s(x-\theta))$$

or

$$C(S) = as - 1 + \exp(-\theta s) = 0 \qquad (22)$$

Among the important properties of the so-called characteristic roots of the above equation are the following [3]:

1. The complex roots occur in complex conjugate pairs which are always simple. This is easily seen by solving simultaneously $C(s) = C'(s) = 0$ which gives $s = 1/a - 1/\theta$, a real number.

2. There are at most two real characteristic roots. In fact C(s) in Eq. (22) has exactly two real roots, one if which is s = 0, which may be seen by graphing the equation. If a θ then s = 0 is a double root. If a < θ the other real root is positive. If a > θ the other root is negative.

3. There is at most one multiple root of multiplicity two given by $s = 1/a - 1/\theta$.

4. All the roots lie entirely to the left of some vertical line in the complex s-plane. In fact, by pplying the Argument Principle [5], every complex root of Eq. (20) may be shown to have a negative real part [2].

5. There are roots with arbitrarily small (i.e., negatively large) real part.

6. The characteristic roots can be arranged (i.e., indexed) in the order of increasing absolute value, or of decreasing real part, or of increasing imaginary part.

According to a theorem for delay LDDE's [3] all continuous solutions of Eq. (20) are given by:

When $a = \theta$

$$g_{L0}(x) = \sum_{i=2}^{\infty} c_i \exp(r_i x) + c_0 + c_1 x \quad (23)$$

When $a \neq \theta$

$$g_{L0}(x) = \sum_{i=2}^{\infty} c_i \exp(r_i x) + c_0 + c_1 \exp(\sigma_1 x) \quad (24)$$

The r_i are complex characteristic roots and σ_1 is the nonzero real root, if there is one. The constants c_i can be arbitrarily chosen as long as the series converges to a real number for all x.

Eqs. (23-24) can be used to search for nonlinearity forms that yield pdf-like forms when subjected to integration as in Eq. (17). It still remains to be verified that the corresponding f_0 is indeed a pdf, i.e., that Eq. (21) is satisfied.

Normal Distribution. Taking the last three terms in Eq. (23), which is equivalent to taking the double root zero, and integrating according to Eq. (17) gives the pdf form

$$f_0(x) = \exp(c'_0 + c'_1 x + b'x^2) \quad (25)$$

This is easily recognized as a gaussian or normal pdf. Thus

$$f_0(x) = (2\pi)^{-1/2} \sigma \exp(-(x-\mu)^2/2 \sigma^2) \quad (26)$$

for arbitrary mean μ (although zero mean is common) and variance σ^2. In the constant signal independent identically distributed noise sample case, the detector nonlinearities are

$$g_{NP}(x;\theta) = (x-\mu)\theta/\sigma^2 - \theta^2/2\sigma^2$$

$$g_{LO}(x) = (x-\mu)/\sigma^2 \qquad (27)$$

It may be seen that Eq. (12) is satisfied. Because of the linear relationship of nonlinearities, the critical regions of both the most powerful and locally optimum detectors may be expressed as

$$\sum_{i=1}^{m} x_i > t \qquad (28)$$

which is seen to involve the sufficient statistic for the mean of the normal distribution.

<u>Extreme Value Distribution</u>. Taking the last three terms in Eq. (24), which is equivalent to choosing the zero root and a nonzero real root λ, and integrating according to Eq. (17) gives the pdf form

$$f_0(x) = K\exp(\, c'_1 \exp(\lambda x) + b'x\,) \qquad (29)$$

Trying various combinations for the arbitrary constants c'_1 and b', we find that the integrable solution is the extreme value distribution (also called the double exponential distribution) [6]

$$f_0(x) = \frac{\lambda C_2^{C_1/\lambda}}{\Gamma(c_1/\lambda)} \exp(\, c_1 x - c_2\exp(\lambda x)\,) \qquad (30)$$

where Γ is the gamma function, and the constants satisfy $c_2 > 0$ and $c_1\lambda > 0$.

This family of pdf's includes the Gumbel distribution [6] or "first double exponential distribution"

$$f_0(x) = K\exp(\, -\exp(-x) - x\,)$$

as well as Gumbel's second exponential distribution

$$f_0(x) = K\exp(\, -\exp(x) + x\,)$$

The detector nonlinearities for the case of a constant signal in independent identically distributed noise samples are

$$g_{NP}(x;\theta) = c_2(1 - \exp(-\lambda\,\theta))\exp(\lambda\, x) - c_1\theta$$

$$g_{LO}(x) = c_2 \lambda \exp(\lambda x) - c_1 \quad (31)$$

The critical regions of the most powerful and locally optimum detectors can be expressed in terms of a sufficient statistic as

$$\sum_{i=1}^{m} \exp(\lambda x_i) > t$$

<u>Time Varying Signals</u>. We now turn to the detection of a time varying signal in additive noise with independent identically distributed samples. Here the samples under the alternate hypothesis are independent but not identically distributed. In general, the most powerful and the locally optimum detector nonlinearities will also be time-varying. Thm. 1 is still the necessary and sufficient condition for a bioptimum detector to exist. Therefore, pdf's satisfying Eq. (12) under signal-varying conditions must also satisfy it under constant-signal conditions.

It may be verified by computing the most powerful and locally optimum nonlinearities that a bioptimum detector exists for a time varying signal in additive gaussian noise. In fact, the two tests have a common critical region

$$\sum_{i=1}^{m} s_i x_i > t$$

This detector performs a discrete-time correlation of the received samples with the known signal samples.

It may be similarly verified that a bioptimum detector does not exist for the extreme value distribution unless the signal is constant. Thus the gaussian pdf is unique in this respect.

CONCLUSION

This paper has studied the structure of most powerful and locally optimum detectors for discriminating between the presence or absence of a known signal in additive noise. It has been shown that the two detectors reduce to the same structure whenever the detector nonlinearities have the affine relationship expressed in Thm 1. This further implies that the nonlinearities must satisfy a difference-differential equation given in Thm 3. Two integrable solutions of this equation yield the gaussian pdf and the extreme value pdf as the noise pdf's for which most powerful and locally optimum detectors coincide. Unfortunately, it seems that there are no other pdf's within this model, so that in most realistic detection problems, the designer cannot have both types of detectors.

REFERENCES

1. ABAYA, EFREN F., *"A Functional Equation in Detection Theory,"* Philippine Engineering Journal, Vol. VII, pp. 88-93, Dec. 1986.

2. ABAYA, EFREN F., A Relation Between Optimum Detectors and Locally Optimum Detectors, M.S. Thesis, The University of Texas at Austin, 1980.

3. BELLMAN, RICHARD and KENNETH L. COOKE, Differential-Difference Equations N.Y.: Academic Press, 1963.

4. CAPON, JACK, *"On the Asymptotic Relative Efficiency of Locally Optimum Detectors,"* IRE Trans. Inform. Theory, vol. IT-7, pp. 67-71, April 1961.

5. CHURCHILL, RUEL V., Complex Variables and Applications, 2nd ed.; N.Y.: McGraw-Hill Book Company, 1960.

6. JOHNSON, NORMAN L. and SAMUEL KOTZ, Continuous Univariate Distributions, Vols 1,2, Boston: Houghton Mifflin Company, 1970.

7. MILLER, JAMES H. and JOHN B. THOMAS, Detectors for Discrete-Time Signals in Non-Gaussian Noise, IEEE Trans. Inform. Theory, vol. IT-18, pp. 241-50, March 1972.

8. RAO, C.R., Linear Statistical Inference and Its Applications, N.Y.: John Wiley & Sons, Inc., 1963.

9. SAATY, THOMAS L., Modern Nonlinear Equations, N.Y.: McGraw-Hill Book Company, 1967.

10. SHANMUGAN, K. SAM and ARTHUR M. BREIPOHL, Random Signals: Detection, Estimation and Data Analysis, N.Y.: John Wiley & Sons, Inc., 1988.

11. VAN TREES, HARRY L. DETECTTION, Estimation, and Modulation Theory, Part 1, N.Y.: John Wiley & Sons, 1968.