3. BACKGROUND THEORY AND RESULTS

In this section we will summarize briefly the pertinent concepts from Statistical-Decision Theory which we require in our search for optimum detection algorithms and in the determination of the performance of these algorithms. The approach to be used is, naturally, the standard Bayes Theory approach, and we want to specify the techniques developed in the last few years which we will use for the real-world, impulsive, non-Gaussian channel.

One of the main assumptions which is made is that the waveform of the received interference can be represented by a vector of independent samples. We will discuss the ramifications of this assumption in section 3.2 for our class A and class B interference (sec. 2.3).

3.1 Elements of Statistical-Decision Theory

In this section we summarize the basic principles of Bayes Theory, use these to obtain Hall's (1966) optimum receiver, to specify the Schwartz (1969) series algorithm, and to present the ideas of locally optimum Bayes detection (LOBD).

We start with some basic definitions:

(1) Γ is the observation space which has elements $[x]$. The element x is our received data (waveform) which is composed of our desired signal plus interference (or interference above).

(2) A is the decision space with elements $[a]$, called <u>decisions</u> or <u>actions</u>.

(3) Ω is the signal space having elements [s], our desired signals. For binary problems Ω has two elements, s_1 and s_2. The elements [s] have an a priori distribution q(s).

(4) A decision function $\delta(a|x)$ is the conditional probability of deciding $a \in A$ given a point $x \in \Gamma$.

(5) A cost function, denoted by C(s,a), is the cost (or penalty) that is assessed for choosing a when s was transmitted.

(6) The average risk is the expected value of the cost function. The average risk depends on the decision function $\delta(a|x)$ and the a priori signal distribution q(s). Thus,

$$R(q,\delta) = E[C(s,a)] \quad .$$

(7) The evaluation of a system consists of determining $R(q,\delta)$ for a given q(s) and $\delta(a|x)$.

(8) The decision function $\delta_B(a|x)$ which minimizes the average risk is termed the Bayes decision rule. That is

$$R(q,\delta_B) \leq R(q,\delta) \text{ for all } \delta \; ,$$

and $R(q,\delta_B)$ is the Bayes risk.

(9) The optimum receiver is that receiver which performs the operations indicated by the Bayes decision rule.

In general,

$$R(q,\delta) = \int_{\Omega} ds\ q(s) \int_{A} da\ C(s,a) \int_{\Gamma} dx\ \delta(a|x)\ p(x|s) \ , \tag{3.1}$$

where $p(x|s)$ is the probability of observing x, given that s was transmitted.

Specializing to a signal space Ω of two elements for $A = (a_1,a_2)$ and $\Omega = (s_1,s_2)$ we define a cost matrix

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \ , \tag{3.2}$$

where C_{ij} is the cost of deciding a_i given s_j was transmitted. By definition of "correct" and "incorrect" we must assume that the cost of making a correct decision to be less than the cost of making an incorrect decision, so that $C_{ii} < C_{ij}$ for all $i \neq j$. Under these conditions it is easy to show (Weber, 1968), using (3.1), that the decision function that minimizes $R(q,\delta)$ is given by

$$\delta_B(a_2|x) = \begin{cases} 1 \text{ when } \dfrac{q_2C_{21}p(x|s_2) + q_1C_{11}p(x|s_1)}{q_1C_{12}p(x|s_1) + q_2C_{22}p(x|s_2)} \geq 1 \\ 0 \qquad \text{otherwise} \ , \end{cases} \tag{3.3}$$

and

$$\delta_B(a_1|x) = 1 - \delta_B(a_2|x) \quad . \tag{3.4}$$

In (3.3), we have assumed that the signals, s_i, are completely known so that the probability densities are completely given. This special case falls into the classification of simple hypothesis testing. In general, the signals may be random or contain random parameters. In this case, the pdf's $p(x|s_i)$ in (3.3) must be replaced by $\langle p(x|s_i)\rangle_{s_i}$, where $\langle\cdot\rangle_{s_i}$ denotes required statistical averages over the random parameters of the signal s_i (or in some cases, over s_i itself).

If we now define a likelihood ratio $\Lambda(x)$,

$$\Lambda(x) = \frac{\langle p(x|s_2)\rangle_{s_2}}{\langle p(x|s_1)\rangle_{s_1}} \quad , \tag{3.5}$$

then (3.3) becomes

$$\delta_B(a_2|x) = \begin{cases} 1 \text{ if } \Lambda(x) \geq \dfrac{q_1(C_{12}-C_{11})}{q_2(C_{21}-C_{22})} \equiv K \\ \\ 0 \text{ if } \Lambda(x) < K \quad . \end{cases} \tag{3.6}$$

Thus the optimal decision function is a threshold operator. The optimal receiver (decision function) simply forms the likelihood ratio* and compares it to a threshold K and decides a_2 occurs if $\Lambda(x) \geq K$ and a_1 otherwise. The average risk is

---

*An earlier and alternative definition of likelihood ratio is the generalized likelihood ratio $\Lambda_{gen} = q_2/q_1\Lambda$ [Middleton, 1960, chapter 19].

$$R(q,\delta_B) = q_1C_{12}P_{12} + q_2C_{22}P_{22} + q_1C_{11}P_{11} + q_2C_{21}P_{21} \quad , \tag{3.7}$$

where P_{ij} = Prob[deciding $a_j | s_i$ was transmitted].

In some situations the a priori probabilities q(s) are unknown and cannot be estimated. In this situation the minimax criterion is often an acceptable alternative for determining the best receiver operation. Since the q(s) is unknown, we use a conditional risk $r(s,\delta)$, where

$$r(s,\delta) = \int_A da \int_\Gamma C(s,a)\ \delta(a|x)\ p(x|s) \quad . \tag{3.8}$$

A decision function δ_M is said to be minimax if

$$\max_s r(s,\delta_M) = \min_\delta \max_s r(s,\delta_M) \le \max_s r(s,\delta) \text{ for all } \delta.$$

This minimizes the set of worst possible cases and does not require knowledge of q(s). It can be shown that the minimax decision function corresponds to the Bayes decision function for the a priori probabilities which makes the Bayes risk a maximum [Middleton, 1960, sec. 18]. So the minimax receiver is also a threshold device and can be obtained from the likelihood ratio (3.5) with the threshold adjusted (via q_1 and q_2) to maximize the Bayes risk (3.7).

In other situations, such as radar detection, the a priori probabilities are unknown, and the individual costs are not usually assigned. In this case the Neyman-Pearson criterion is appropriate. The concept (for signal spaces with two elements, s_1 and s_2) is to fix the error P_{12} at an acceptable low level (termed the false alarm probability α) and then minimize the error P_{21} so that the detection probability P_{22} is maximized. This Neyman-Pearson receiver is also a threshold device and can be determined from the likelihood ratio, with the threshold K determined by

$$P_{12} = \text{Prob}[\Lambda(x) \geq K | s_1 \text{ was sent}] = \alpha \quad . \qquad (3.9)$$

[Of course, fixing P_{12} automatically assigns a cost ratio and a q_2/q_1 ratio at the same time.]

Thus we see that the Bayes receiver, the minimax receiver, and the Neyman-Pearson receiver are all determined by a Bayes decision function; i.e., a likelihood ratio, with an appropriate threshold.

In communications problems it is customary to use the cost matrix (called the "Ideal Observer")

$$C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} , \qquad (3.10)$$

so that the corresponding average Bayes risk (3.7) becomes $R(q,\delta_B) = q_2P_{21} + q_1P_{12}$. The average probability of error is $P_e = q_2P_{21} + q_1P_{12}$ so that the Bayes risk and the average probability of error are synonymous in this case. The Bayes decision rule minimizes the average probability of error.

If we receive our desired signals corrupted by additive interference, so that our received waveform is composed of the desired signal plus the interfering waveform $Z(t)$, then, corresponding to the above, we have two hypothesis states

$$H_1: \quad x(t) = s_1(t) + Z(t), \quad 0 < t \le T$$

$$H_2: \quad x(t) = s_2(t) + Z(t), \quad 0 < t \le T \quad . \qquad (3.11)$$

The optimum receiver forms the likelihood ratio $\Lambda(x)$ and decides $H_1(s_1(t)$ sent) if $\Lambda(x) \le K$, and $H_2(s_2(t)$ sent) otherwise. The average probability of error is given by

$$P_e = q_2 \; \text{Prob}[\Lambda(x) \ge K] + q_1 \; \text{Prob}[\Lambda(x) < K] \quad , \qquad (3.12)$$

so that the pdf of $\Lambda(x)$ is required in order to evaluate performance.

The method most often used in analyzing communication systems with time varying random waveforms is to replace all waveforms by vectors of samples from the waveforms. This is done since many communications receivers physically operate with discrete data and also there are as yet, in general, no useful mathematical techniques available for directly handling the continuous data problem (the white Gaussian noise channel being the exception). The problem (3.11) becomes

$$H_1: \quad \underline{X} = \underline{S_1} + \underline{Z}$$

$$H_2: \quad \underline{X} = \underline{S_2} + \underline{Z} \quad , \qquad (3.13)$$

where $\underline{X} = \{x(t_1), x(t_2), \ldots, x(t_N)\}$, etc. When the interference Z is white Gaussian, the vector formulation is completely equivalent to the continuous formulation; i.e., the resulting discrete operations carry over directly to the corresponding continuous operations. [See Hancock and Wintz, 1966, appendix B, for the connection between processing finite dimensional random vectors and processing continuous random waveforms.] If Z is non-Gaussian, we operate so as to generate (essentially) independent samples, z_i, so that only first order pdf's are required in our analysis. This is a necessary simplification, which leads to upper bounds on optimality. We discuss this assumption of independence and its practical implications in section 3.2 for our present interference processes (Middleton's class A and class B interference). For non-Gaussian situations, the required pdf of $\Lambda(\underline{X})$, containing the various effects of intersample correlations, can almost never be determined generally, so that it is then impossible to obtain performance estimates precisely. One usually must be satisfied with bounds (upper and lower) on performance, which are almost always quite adequate for the practical purposes of assessing performance and making system comparisons.

3.1.1 Hall's Optimum Receiver

In chapter 2 we reviewed the various models which have been proposed for the real world non-Gaussian interference and noted that the only model that has been used to date to determine an optimum receiver is the Hall (1966) model. We now determine Hall's receiver, to illustrate the concepts of detection theory outlined above and to get some idea of how non-Gaussian statistics for the interference influence detector structure.

The pdf of the instantaneous amplitude of the received interference is given by (2.30) as

$$p_Z(z) = \frac{\Gamma\left(\frac{\theta}{2}\right)}{\Gamma\left(\frac{\theta-1}{2}\right)} \frac{\gamma^{\theta-1}}{\sqrt{\pi}} \frac{1}{[z^2+\gamma^2]^{\theta/2}} \quad . \qquad (3.14)$$

The likelihood ratio for the Hall statistics (3.14), for completely known signals, the cost matrix (3.10), and $q_1 = q_2 = 1/2$ [so that K = 1], is

$$\Lambda(\underline{X}) = \frac{P(\underline{X}|\underline{S_2})}{P(\underline{X}|\underline{S_1})} = \frac{P_Z(\underline{X}-\underline{S_2})}{P_Z(\underline{X}-\underline{S_1})} \underset{S_2}{\overset{S_1}{\lessgtr}} 1 \quad , \qquad (3.15)$$

where $\underline{X}$ is our vector of N samples $\{x(t_1), \ldots, x(t_n)\}$, etc. We assume independent interference samples, $z_i = Z(t_i)$, so that the required Nth order pdf in (3.15) is given by the Nth product of the first order pdf (3.14) [see remarks in sec. 3.2].

Thus we obtain

$$\Lambda(\underline{X}) = \frac{\prod_{i=1}^{N} C(\theta)\, \gamma^{\theta-1}[\gamma^2+(x_i+s_{2i})^2]^{-\theta/2}}{\prod_{i=1}^{N} C(\theta)\, \gamma^{\theta-1}[\gamma^2+(x_i+s_{1i})^2]^{-\theta/2}} \underset{S_2}{\overset{S_1}{\lessgtr}} 1 \quad . (3.16)$$

Taking logarithms, we find that

$$\sum_{i=1}^{N} \left\{-\frac{\theta}{2} \ln[\gamma^2+(x_i-s_{2i})^2]\right\} - \sum_{i=1}^{N} \left\{-\frac{\theta}{2} \ln[\gamma^2+(x_i-s_{1i})^2]\right\} \underset{S_2}{\overset{S_1}{\lessgtr}} 0 , \tag{3.17}$$

or

$$\sum_{i=1}^{N} \ln[\gamma^2+(x_i-s_{1i})^2] - \sum_{i=1}^{N} \ln[\gamma^2+(x_i-s_{2i})^2] \underset{S_2}{\overset{S_1}{\lessgtr}} 0 . \tag{3.18}$$

We therefore obtain the well-known receiver (3.18), which Hall has termed a "log-correlator" receiver, illustrated in figure 3.1.

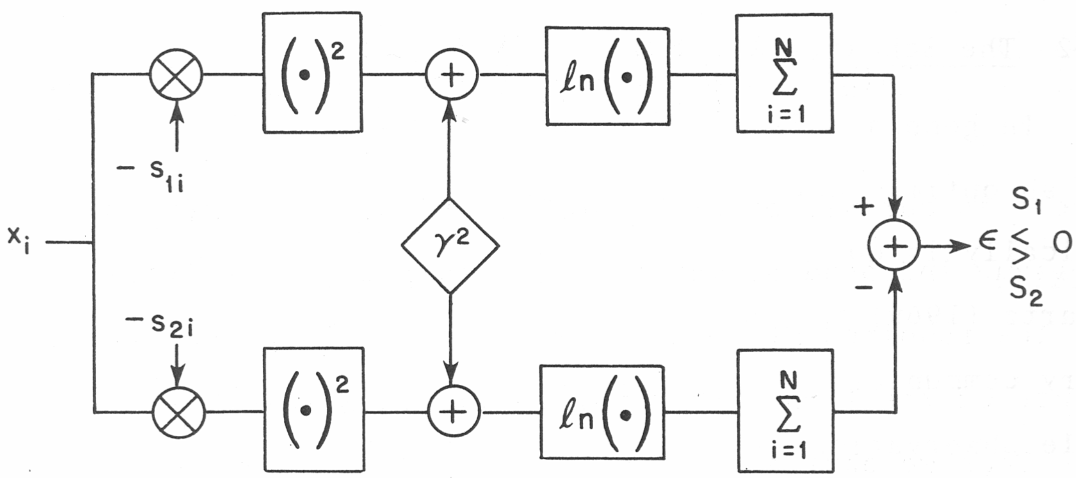

Figure 3.1. The Hall receiver for a completely known set of binary signals, $s_1(t)$ and $s_2(t)$.

We note that the receiver (3.18) must be adaptive, in that it must be able to determine the bias γ^2, which changes with changing interference conditions. Nirenberg

(1974) has developed an algorithm for the estimation of the parameter γ. Hall (1966) analyzed the performance of his receiver (3.18) for ON-OFF digital signaling and Spaulding et al. (1969) performed a more extensive analysis of performance for some special cases of interest. A version of Hall's receiver has been physically constructed (Morozumi and McCormick, 1968) and its performance measured.

Hall (1966) also demonstrated that in order to gain any advantage over normal receivers (i.e., those optimum for Gaussian interference), the number of independent samples, N, must be large. That is, a large time-bandwidth produce is required. (We shall comment on this fact in sections 4 and 5.)

3.1.2 The Schwartz Algorithm

In general, direct use of the likelihood ratio to obtain an optimum detection algorithm that can be implemented physically is quite difficult (for non-Gaussian interference). Schwartz (1969) presented an algorithm that is optimum for binary communication problems where decisions are based on a single observation (N=1). Schwartz's idea is as follows: Instead of using the likelihood ratio

$$\Lambda(x) = \frac{q_2 p_2(x)}{q_1 p_1(x)} \underset{s_2}{\overset{s_1}{\lessgtr}} 1 \quad , \tag{3.19}$$

where $p_2(x) = \langle p_Z(x-s_2)\rangle_{s_2}$ and $p_1(x) = \langle p_Z(x-s_1)\rangle_{s_1}$, consider the completely equivalent test statistic

$$T(x) = q_2 p_2(x) - q_1 p_1(x) \underset{s_2}{\overset{s_1}{\lessgtr}} 0 \quad . \tag{3.20}$$

Suppose the densities $p_i(x)$, $i = 1, 2$, can be represented by pointwise convergent series

$$\left| p_i(x) - \sum_{j=0}^{q} a_{ij} \, \psi_j(x) \right| < B_i(q,x), \; i = 1,2 \; , \tag{3.21}$$

where, for every x,

$$\lim_{q \to \infty} B_i(q,x) = 0 \quad .$$

Then (3.20) is given by

$$T(x) = \sum_{j=0}^{\infty} (q_2 \, a_{2j} - q_1 \, a_{1j}) \, \psi_j(x) \quad . \tag{3.22}$$

Define a truncated test statistic

$$T_q(x) = \sum_{j=0}^{q} (q_2 a_{2j} - q_1 a_{1j}) \, \psi_j(x) \quad . \tag{3.23}$$

It follows that $T_q(x)$ differs from $T(x)$ by

$$|T(x) - T_q(x)| < q_1 B_1(q,x) + q_2 B_2(q,x) \equiv B(q,x). \tag{3.24}$$

With a knowledge of this bound, the following procedure, which uses the truncated series $T_q(x)$, is optimum. The observation is used to evaluate $T_q(x)$ for some initial q. If, for the given values of x and q, the magnitude of $T_q(x)$ is greater than $B(q,x)$, the decision using $T_q(x)$ is the same as when

$T(x)$ is used; $T(x)$ will lie somewhere between $T_q(x) \pm B(q,x)$, and this quantity will be on the same side of the threshold as $T_q(x)$. Hence if $|T_q(x)| > B(q,x)$, H_2 is announced if $T_q(x)$ is positive and H_1 if $T_q(x)$ is negative. If the inequality is not satisfied, another term of the series and the bound $B(q+1,x)$ are computed and the procedure repeated until the inequality is meet for some q.

The above procedure, while optimum, cannot be used for performance determination. Also, the required pointwise convergent series representation of the pdf's and the bound $B(x,q)$ may be very difficult to obtain. In our case, however, the pdf's for class A and class B interference are already given as pointwise convergent infinite series (e.g., 2.56 and 2.61). Correspondingly, we will discuss the above series algorithm in detail in section 4.1 for our case of class A interference.

3.1.3 Threshold Receivers

Another popular way of overcoming the inherent difficulties in obtaining physically realizable detection structure from the likelihood ratio for non-Gaussian interference is to investigate optimum detection as the desired signal becomes vanishing small. The receivers that result from such investigations are termed threshold or locally optimum Bayes detectors (LOBD).

This problem of optimum threshold detection was apparantly first attacked by Middleton (1954, 1960). Later Rudnick (1961) examined the threshold problem for the case

$$H_1: \quad \underline{X} = \underline{Z}$$
$$H_2: \quad \underline{X} = \underline{Z} + \underline{S} \quad , \qquad (3.25)$$

using the likelihood ratio

$$\Lambda(\underline{X}) = \frac{p_2(\underline{X})}{p_1(\underline{X})} \mathop{\gtrless}_{H_2}^{H_1} 1 \quad , \qquad (3.26)$$

where $p_1(\underline{X}) = p_Z(\underline{X})$ and where the desired signal may be random so that

$$p_2(\underline{X}) = \int p_Z(\underline{X}-\underline{S}) \; d \; F(\underline{S}) \quad , \qquad (3.27)$$

where $F(\underline{S})$ is the cumulative distribution function of the signal. Rudnick assumes that $p_Z(\underline{X}-\underline{S})$ has a Taylor expansion about $\underline{X}$ so that

$$\Lambda(\underline{X}) - 1 = \frac{1}{p_1(\underline{X})} \left[- \sum_{i=1}^{N} \frac{\partial p_Z(\underline{X})}{\partial \; x_i} \langle s_i \rangle + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 p_Z(\underline{X})}{\partial x_i \; \partial x_j} \langle s_i s_j \rangle + \ldots \right]$$

$$= \sum_{i=1}^{N} y_i \langle s_i \rangle + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i y_j - v_{ij}) \langle s_i s_j \rangle + \ldots \quad , \qquad (3.29)$$

with

$$y_i = - \frac{\partial}{\partial x_i} \ln p_Z(\underline{X}) \quad ,$$

and

$$v_{ij} = \frac{\partial^2}{\partial x_i \, \partial x_j} \ln p_Z(\underline{X}) \ .$$

The threshold receiver is then obtained by assuming that the signal is small enough so that the higher order terms in s can be neglected. The result (3.29) is canonical in that it does not depend on the particular form of $p_Z(z)$.

Middleton (1966) presented a detailed treatment for both coherent and incoherent signaling situations, obtaining a generalized matrix form of Rudnick's development. Middleton (1966) also showed that in some situations the LOBD must contain a bias term that depends on the signal-to-noise ratio. This bias term is required to insure that $P_e \to 0$ as $S \to \infty$ and that $P_e \to 0$, given S, as $N \to \infty$ (consistency). We will see an example of this in section 5.1.2 where we treat threshold reception of ON-OFF incoherent signals. Middleton (1966) also gives sufficient, and necessary and sufficient, conditions on the bias term and on the threshold test statistic itself to insure that the LOBD is at least asymptotically as efficient, or equivalent to, the corresponding Bayes detector in the limit $(S \to 0)$. [See section 4.4.1 for discussion and examples of asymptotic efficiency.]

Antonov (1967) treated the coherent ON-OFF signalling situation, using the same approach as Middleton (1966), however, assuming independent samples. Antonov's analysis proceeds as follows:

The likelihood ratio (3.26), for N independent samples, is

$$\Lambda(\underline{X}) = \exp\left[\sum_{i=1}^{N} \ln p_Z(x_i - s_i) - \sum_{i=1}^{N} \ln p_Z(x_i)\right] \overset{H_1}{\underset{H_2}{\lessgtr}} 1 . \tag{3.30}$$

Expanding $\ln p_Z(x_i - s_i)$ about x_i, collecting terms, and interchanging summations, one gets the result

$$\ln \Lambda(\underline{X}) = \sum_{k=1}^{\infty} \sum_{i=1}^{N} \frac{(-1)^k}{k!} s_i^{\,k} \frac{d^k}{dx_i^{\,k}} \ln p_Z(x_i) . \tag{3.31}$$

Then ignoring terms k > 1, one finds the threshold receiver to be

$$\ln \Lambda(\underline{X}) = \left[- \sum_{i=1}^{N} s_i \frac{d}{dx_i} \ln p_Z(x_i)\right] \overset{H_1}{\underset{H_2}{\lessgtr}} 0 . \tag{3.32}$$

Antonov (1967b) extended this result to the corresponding incoherent signaling case.

Kushner and Levin (1968) and Ribin (1972) showed that these canonical threshold receivers are asymptotically optimum.

In sections 4 and 5 we will obtain the canonical threshold receivers for a variety of binary coherent and incoherent signaling situations and determine their performance in class A interference.