(NASA-CR-143203) ON OPTIMAL SOFT-DECISION DEMODULATION (Notre Dame Univ.) 24 p HC $3.25 CSCL 09C

N75-28315

Unclas
G3/33 29840

Department of

ELECTRICAL ENGINEERING

UNIVERSITY OF NOTRE DAME, NOTRE DAME, INDIANA

ON OPTIMAL SOFT-DECISION DEMODULATION*

Lin-nan Lee
Department of Electrical Engineering
University of Notre Dame
Notre Dame, Indiana 46556

Technical Report No. EE-753

July 1, 1975

Abstract: Wozencraft and Kennedy have suggested that the appropriate demodulator criterion of goodness is the cut-off rate of the discrete memoryless channel created by the modulation system; the criterion of goodness adopted in this note is the "symmetric" cut-off rate which differs from the former criterion only in that the signals are assumed equally likely. Massey's necessary condition for optimal demodulation of binary signals is generalized to M-ary signals. It is shown that the optimal demodulator decision regions in likelihood space are bounded by hyperplanes. An iterative method is formulated for finding these optimal decision regions from an initial "good guess." For additive white Gaussian noise, the corresponding optimal decision regions in signal space are bounded by hypersurfaces with hyperplane asymptotes; these asymptotes themselves bound the decision regions of a demodulator which, in several examples, is shown to be virtually optimal. In many cases, the necessary condition for demodulator optimality is also sufficient, but a counterexample to its general sufficiency is given.

* This research was supported by the National Aeronautics and Space Administration under NASA Grant NSG 5025 at the University of Notre Dame in liaison with the Goddard Space Flight Center. A portion of this paper was presented orally at the IEEE International Symposium on Information Theory, Notre Dame, Indiana, October 27-31, 1974.

I. INTRODUCTION

The block diagram of a one-way, coded, digital communications system is given in Figure 1. From this figure, it is apparent that modulation and coding are both aspects of the "signal design" problem, whereas demodulation and decoding are both aspects of the "signal detection" problem. The natural question then is how to coordinate the design of the modulation system and the coding system so as to produce an efficient and effective communications system.

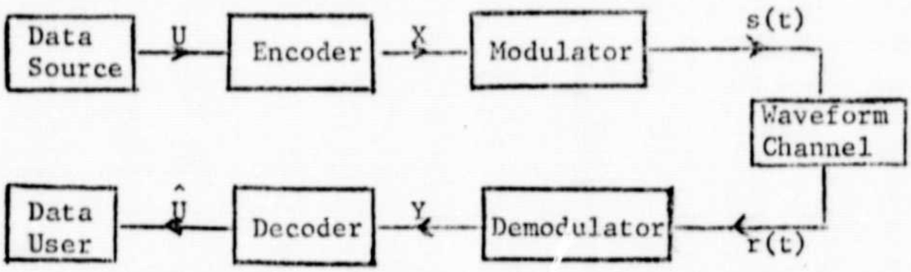


Figure 1. A one-way, coded, digital communications system.

Suppose that the modulator is M-ary; then, without loss of generality, we may consider the modulator input alphabet to be the set {0, 1, 2, ..., M-1}. Suppose the demodulator is restricted to J different decisions, then we may take its output alphabet to be {0, 1, 2, ..., J-1}. We say that the demodulator makes "hard-decisions" or "soft-decisions" according as to whether $J = M$ or $J > M$ respectively. Clearly, the "classical" modulation system design criterion of "error probability" is applicable only for hard-decisions. Unfortunately for classicists, the use of a hard-decision demodulator generally reduces substantially the effectiveness of the coding system.

Wozencraft and Kennedy [1] were the first to suggest that the proper modulation system design criterion is the "cut-off rate," R_o, of the M-input, J-output discrete memoryless channel (DMC) presented by the modulation system to the coding system. This DMC is completely described by the transition probabilities, $P(j|m)$, that the demodulator decision is j given that the modulator input was m, $0 \leq j < J$, $0 \leq m < M$. Mathematically, the cut-off rate is given by

$$R_o = -\log_2 \{\min_Q \sum_{j=0}^{J-1} [\sum_{m=0}^{M-1} Q(m)\sqrt{P(j|m)}]^2\} \quad (1)$$

where Q is a probability distribution for the channel input letters, i.e., $Q(m) \geq 0$ for all m and $Q(0) + Q(1) + \ldots + Q(M-1) = 1$.

Wozencraft and Kennedy were led to the choice of R_o (or, as it was then usually denoted, "R_{comp}") because R_o is the upper limit of code rates for which the average decoding computation per digit is finite when sequential decoding is used. More recently, Massey [2] has pointed out a more persuasive reason for choosing R_o as a design criterion. Viterbi [3] has shown that, when convolutional coding is used with maximum likelihood decoding on the DMC, then the decoding error probability is upper bounded by

$$P_e < c_R L 2^{-NR_o}, \text{ if } R < R_o, \quad (2)$$

where N is the code constraint length, R is the code rate (number of data bits per decoded letter), L is the number of bits encoded, and c_R is an unimportant constant independent of N and L. Hence, as Massey observed, the single number R_o determines both a range of rates over which reliable operation is possible as well as a measure of the necessary coding complexity to obtain a specified error probability. R_o is thus even more informative than the channel capacity of the DMC which, although it determines the entire region of rates over which reliable communications is possible, says nothing about the coding complexity needed for a specified decoding error probability at any given code rate.

In the same paper [2], Massey established a number of fundamental results about modulation systems under the R_o criterion. He gave a general expression for R_o for unquantized demodulation ($J=\infty$), and proved that, for any given M, the M-ary simplex signal maximizes the unquantized R_o for the additive white

Gaussian noise (AWGN) channel. For binary modulation (M=2) and any given J, Massey also gave a necessary condition for the demodulator decision regions to be optimal, and showed how to use this condition as the basis of an iterative computational technique for finding the optimal decision regions.

In this note, we extend Massey's necessary condition for optimal demodulation to the non-binary (M>2) case, and we give an example which shows that the condition is not sufficient even in the binary case. We show that, in likelihood space, the optimal demodulator decision regions are always bounded by hyperplanes, and we give some examples to illustrate the nature of these regions.

II. THE SYMMETRIC CUT-OFF RATE

We define the symmetric cut-off rate, $\tilde{R}_o$, to be the value of the righthand side of (1) when Q is the uniform distribution [$Q(m) = 1/M$ for $0 \leq m < M$] rather than the minimizing distribution. Thus,

$$\tilde{R}_o = \log_2 M - \log_2 \{\frac{1}{M} \sum_{j=0}^{J-1} [\sum_{m=0}^{M-1} \sqrt{P(j|m)}]^2\}. \quad (3)$$

Evidently, $\tilde{R}_o \leq R_o$. Moreover, $\tilde{R}_o = R_o$ in the binary case (M=2) for which the uniform distribution is always the minimizing distribution, and also $\tilde{R}_o = R_o$ in most other cases of practical interest where the modulation signal set and the demodulator decision regions are reasonably "symmetric." Furthermore, the bound of (1) becomes

$$P_e < c_R L 2^{-N\tilde{R}_o}, \text{ if } R < \tilde{R}_o, \quad (4)$$

when the code is such that each letter in the code alphabet appears in the same fraction of codewords, a situation that always occurs in the conventional convolutional codes that would be used in practice. Thus, both to reflect this practical situation and to obviate the awkward minimization over Q in (1), we henceforth take $\tilde{R}_o$ of the resultant DMC as the measure of quality for the modulation system.

III. A NECESSARY CONDITION FOR OPTIMAL DEMODULATION

Henceforth, we assume that we have made the standard transformation [4] from waveforms to signal space so that s(t) and the "relevant" component of r(t) in Figure 1 may be replaced by the corresponding vectors $\underline{s}$ and $\underline{r}$ in n-dimensional Euclidean space. We let $\underline{s}_m$ denote the transmitted signal when the modulation input is m. Any demodulator then may be viewed as a partition $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{J-1}$ of m-space, where the "decision region" $\mathcal{D}_j$ is the set of all received vectors $\underline{r}$ that cause the demodulator to emit the decision j. We now derive a necessary condition for the decision regions to be optimal for a given signal set and channel.

Let $p(\underline{r}|m)$ be the probability density function, which we assume to be everywhere continuous, for the received vector $\underline{r}$ given that signal $\underline{s}_m$ is transmitted over the channel. The transition probabilities of the resultant DMC, for a given demodulator, are then given by

$$P(j|m) = \int_{\mathcal{D}_j} p(\underline{r}|m)\,d\underline{r}. \tag{5}$$

Let a and b, $a \neq b$, be two output letters of the demodulator such that $\mathcal{D}_a$ and $\mathcal{D}_b$ are adjacent regions, i.e. the boundary between $\mathcal{D}_a$ and $\mathcal{D}_b$ is a hypersurface in n-space, and let $\underline{\rho}$ be any point on this boundary. Next, consider transferring from $\mathcal{D}_a$ to $\mathcal{D}_b$ a small region, $\mathcal{D}_{ab}$, which includes the point $\underline{\rho}$. [We show this situation in Figure 2 for the case n=2]. The resulting variation in the transition probabilities is then seen from (5) to be

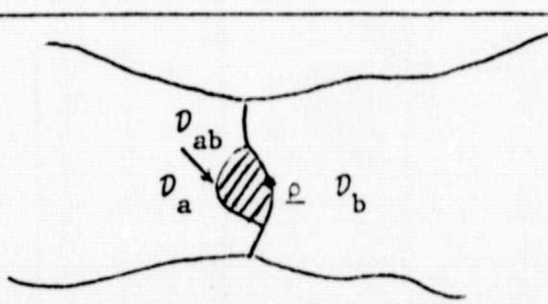


Figure 2. The variation of the decision regions $\mathcal{D}_a$ and $\mathcal{D}_b$ by transfer of the small region $\mathcal{D}_{ab}$.

$$\delta P(j|m) = \begin{cases} +\ p(\underline{\rho}|m)\delta V & j = b \\ -\ p(\underline{\rho}|m)\delta V & j = a \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where we have now assumed that each $p(\underline{r}|m)$, $0 \leq m < M$, is continuous at $\underline{r} = \underline{\rho}$, and where

$$\delta V = \int_{\mathcal{D}_{ab}} d\underline{r}$$

is the volume of the small region $\mathcal{D}_{ab}$. If the decision regions are optimal, the resulting variation, $\delta\tilde{R}_o$, of the symmetric cut-off rate, $\tilde{R}_o$, must be 0. As we see from (3), the condition $\delta\tilde{R}_o = 0$ is equivalent to the condition $\delta S = 0$ where

$$S = \sum_{j=0}^{J-1} [\sum_{m=0}^{M-1} \sqrt{P(j|m)}]^2. \tag{7}$$

We then begin with

$$\delta S = \sum_{j=0}^{J-1} \sum_{m=0}^{M-1} \frac{\delta S}{\delta P(j|m)} \delta P(j|m)$$

which, with the aid of (6), becomes

$$\delta S = \sum_{m=0}^{M-1} [\frac{\delta S}{\delta P(b|m)} - \frac{\delta S}{\delta P(a|m)}]\ p(\underline{\rho}|m)\delta V. \tag{8}$$

We next note that direct differentiation in (7) gives

$$\frac{\delta S}{\delta P(j|m)} = [\sum_{i=0}^{M-1} \sqrt{P(j|i)}] \frac{1}{\sqrt{P(j|m)}}, \tag{9}$$

provided that $P(j|m) \neq 0$. Then, by using (9) in (8), we obtain

$$\delta S = \sum_{m=0}^{M-1} [\frac{1}{\sqrt{P(b|m)}} \sum_{i=0}^{M-1} \sqrt{P(b|i)} - \frac{1}{\sqrt{P(a|m)}} \sum_{i=1}^{M-1} \sqrt{P(a|i)}]\quad p(\underline{\rho}|m)\delta V.$$

Thus, the condition that $\delta S = 0$ for an arbitrary δV becomes

$$\sum_{m=0}^{M-1} [\frac{1}{\sqrt{P(b|m)}} \sum_{i=0}^{M-1} \sqrt{P(b|i)} - \frac{1}{\sqrt{P(a|m)}} \sum_{i=0}^{M-1} \sqrt{P(a|i)}] p(\underline{\rho}|m) = 0. \tag{10}$$

We have thus proved:

Theorem: The demodulator decision regions $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{J-1}$ in signal space can maximize R_o only if, for every a and b, $a \neq b$, such that $P(b|m) \neq 0$ and $P(a|m) \neq 0$ for $0 \leq m < M$, and such that $\mathcal{D}_a$ and $\mathcal{D}_b$ share a hypersurface boundary, it is the case that (10) holds at every point $\underline{r} = \underline{\rho}$ on this boundary which is a point of continuity of $P(\underline{r}|m)$ for $0 \leq m < M$.

In the next section we shall give a more illuminating form of condition (10).

IV. DECISION REGIONS IN LIKELIHOOD SPACE

For the received vector $\underline{r}$, we define the waveform channel likelihood-ratio vector, $\underline{\Lambda}(\underline{r}) = [\Lambda_1(\underline{r}), \Lambda_2(\underline{r}), \ldots, \Lambda_{M-1}(\underline{r})]$, by

$$\underline{\Lambda}(\underline{r}) = [\frac{p(\underline{r}|1)}{p(\underline{r}|0)}, \frac{p(\underline{r}|2)}{p(\underline{r}|0)}, \ldots, \frac{p(\underline{r}|M-1)}{p(\underline{r}|0)}]. \tag{11}$$

We note that, as pointed out by Massey [2], the demodulator can always, at its "front end", map $\underline{r}$ to $\underline{\Lambda}(\underline{r})$ with no loss of optimality. Thus, it becomes of interest to determine the form of the decision regions $D_0, D_1, \ldots, D_{M-1}$ in likelihood space which correspond to the optimal decision regions $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{M-1}$ in signal space. But, seeing from (11) that (10) may, after division by $p(\underline{\rho}|0)$ (which we now assume to be non-zero), be rewritten as the linear equation

$$\sum_{m=1}^{M-1} [\frac{1}{\sqrt{P(b|m)}} \sum_{i=0}^{M-1} \sqrt{P(b|i)} - \frac{1}{\sqrt{P(a|m)}} \sum_{i=0}^{M-1} \sqrt{P(a|i)}]\Lambda_m(\underline{\rho}) +$$

$$+ \frac{1}{\sqrt{P(b|0)}} \sum_{i=0}^{M-1} \sqrt{P(b|i)} - \frac{1}{\sqrt{P(a|0)}} \sum_{i=0}^{M-1} \sqrt{P(a|i)} = 0, \tag{12}$$

we have immediately our main result

Corollary 1: The demodulator decision regions $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_{J-1}$ in signal space can maximize $\tilde{R}_o$ only if, for every a and b, $a \neq b$, such that $P(b|m) \neq 0$ and $P(a|m) \neq 0$ for $0 \leq m < M$, and such that $\mathcal{D}_a$ and $\mathcal{D}_b$ share a hypersurface boundary, it is the case that every point $\underline{r} = \underline{\rho}$ on this boundary, which is a point of continuity of $\underline{\Lambda}(\underline{r})$ lies on the hyperplane defined by (12).

In other words, the optimal decision regions in likelihood space are always bounded by hyperplanes. This fact has considerable practical significance as it is difficult to implement circuitry which determines to what decision region some vector $\underline{\Lambda}(\underline{r})$ belongs except in the case when the decision regions are bounded by hyperplanes.

Condition (12) can be placed in an even more transparent form. We note that when (12) is satisfied, then

$$T_m = - \frac{\frac{1}{\sqrt{P(b|0)}} \sum_{i=0}^{M-1} \sqrt{P(b|i)} - \frac{1}{\sqrt{P(a|0)}} \sum_{i=0}^{M-1} \sqrt{P(a|i)}}{\frac{1}{\sqrt{P(b|m)}} \sum_{i=0}^{M-1} \sqrt{P(b|i)} - \frac{1}{\sqrt{P(a|m)}} \sum_{i=0}^{M-1} \sqrt{P(a|i)}} , \quad (13)$$

(provided the denominator is non-zero) is just the intercept on the m-th axis in likelihood space of the boundary hyperplane. [See Figure 3 for a graphical interpretation of T_i.] Thus, we have

<u>Corollary 2</u>: The demodulator decision regions $D_0, D_1, \ldots, D_{J-1}$ in likelihood space can maximize $\tilde{R}_o$ only if, for every a and b, $a \neq b$, such that $P(a|m) \neq 0$ and $P(b|m) \neq 0$ for $0 \leq m < M$, and such that D_a and D_b share a hypersurface boundary, it is the case that this boundary is a hyperplane whose intercepts with the coordinate axes are given by equation (13).

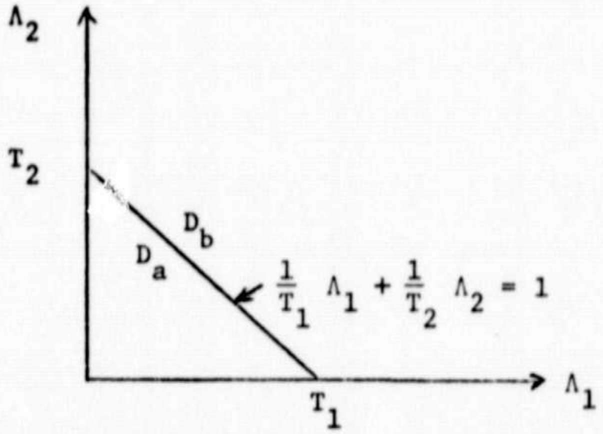


Figure 3: A hyperplane boundary in likelihood space separating the decision regions D_a and D_b for the case $M = 3$.

Analogous to our definition of $\underline{\Lambda}(\underline{r})$, we now define the likelihood vector, $\underline{\lambda}(j) = [\lambda_1(j), \lambda_2(j), \ldots, \lambda_{M-1}(j)]$, of the DMC, which is created by the modulation system, by

$$\lambda_m(j) = \frac{P(j|m)}{P(j|0)}, \tag{14}$$

where we assume $P(j|0) \neq 0$ for $0 \leq j < J$. From (12), (13), and (14), it follows after some tedious algebraic manipulation that, when the decision regions are optimal.

$$\sum_{m=1}^{M-1} \frac{\sqrt{\lambda_m(b)\lambda_m(a)}}{T_m} = 1. \tag{15}$$

For the case of binary modulation ($M = 2$), we note that (15) reduces to the necessary condition for optimality,

$$T = \sqrt{\lambda(b)\lambda(a)} \tag{16}$$

that was given by Massey [2].

V. EXAMPLES AND AN ITERATIVE OPTIMIZATION TECHNIQUE

In this section, we give some examples to illustrate the use of the necessary condition for demodulator optimality given by Theorem 1 and its corollaries. We also formulate a systematic method for finding the optimal demodulator decision regions by iteration from an initial guess.

Example 1: Hard-decision demodulation (i.e., $J = M$) for M-ary phase modulation in additive white Gaussian noise (AWGN). In this case, the signal space is 2-dimensional and the signal vector, $\underline{s}_m$ for $0 \leq m < M$, may be taken as the point on the circle of radius $\sqrt{E}$ (where E is the signal energy) at an angle of $(2\pi/M)m$. The ternary (i.e., $M = 3$) case is shown in Figure 4. The heavy lines in this figure are the boundaries of the decision regions for a

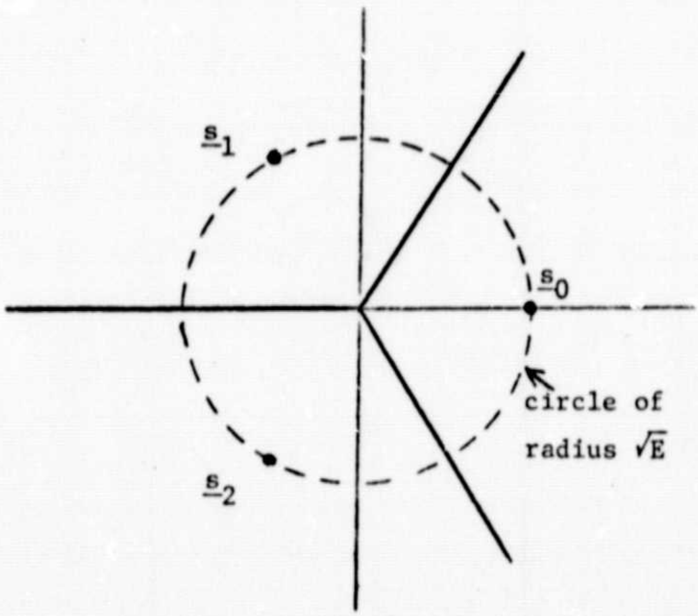


Fig. 4: Maximum-likelihood demodulation of ternary phase-modulated signals.

maximum likelihood (ML) demodulator which, of course, is the hard-decision demodulator that minimizes error probability when the signals are equally likely.

We now show that the ML demodulator for phase modulation is also the hard-decision demodulator which maximizes $\tilde{R}_o$. Let $\underline{s}_a$ and $\underline{s}_b$ be any two adjacent signals, i.e., their phase difference is $2\pi/M$. By the symmetry of the signal set and by the spherical symmetry of the additive white Gaussian noise, it follows that the ML demodulator causes the probabilities $P(b|0)$, $P(b|1)$, ..., $P(b|M-1)$ to be a permutation of $P(a|0)$, $P(a|1)$, ..., $P(a|M-1)$, and also that for each m such that $P(b|m) \neq P(a|m)$ there is a corresponding m' such that $P(b|m) = P(a|m')$, $P(a|m) = P(b|m')$ and $p(\underline{\rho}|m) = p(\underline{\rho}|m')$ for $\underline{\rho}$ on the boundary between $\mathcal{D}_a$ and $\mathcal{D}_b$. Thus, the terms in the summation on the lefthand side of (10) either vanish singly [when $P(b|m) = P(a|m)$] or cancel in pairs. Thus, the ML decision regions satisfy the necessary condition for maximizing $\tilde{R}_o$ given by Theorem 1. Symmetry considerations indicate this is the only local maximum of $\tilde{R}_o$ and hence is the global maximum.

As a specific numerical example, we take the M = [illegible] case of Figure 3 where $E/N_o = 1$, N_o being the one-sided noise power spectral density so that the variance of the noise in each dimension of signal space is $N_o/2$ The value of $\tilde{R}_o$ yielded by the optimal hard-decision demodulator is 0.3971. The unquantized $\tilde{R}_o$ for this case can be found from Massey's results [2] to be 0.6254 so that the penalty for hard-decisions is 1.97 dB.

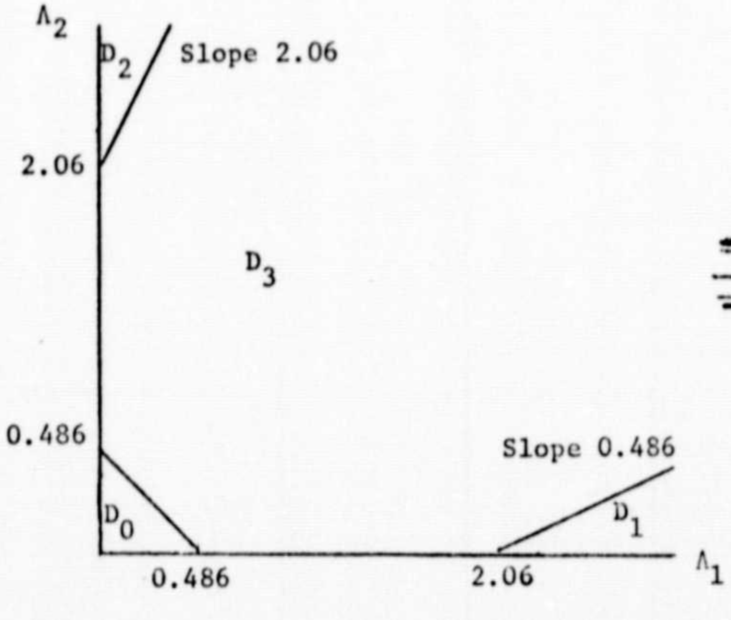
Example 2: Quaternary demodulation ($J = 4$) for ternary phase modulation ($M = 3$) in AWGN. Symmetry considerations suggest that the optimal decision regions will consist of three regions, $\mathcal{D}_0$, $\mathcal{D}_1$ and $\mathcal{D}_2$ having 120° rotational symmetry and containing the signals $\underline{s}_0$, $\underline{s}_1$ and $\underline{s}_2$ respectively, together with an "erasure" region $\mathcal{D}_3$ containing the origin in signal space. In this case, there will be probabilities p and q such that $P(3|m) = q$ for all m, $P(j|m) = p$ for $j \neq m$ and $j \neq 3$, and $P(j|m) = 1 - 2p - q$ for $j = m$. Substituting these parameters into the necessary condition for optimality (13), we find that the resulting optimal intercepts correspond to the straight lines

$$\Lambda_1 + \Lambda_2 = c$$
$$c\Lambda_1 - \Lambda_2 = 1$$
$$-\Lambda_1 + c\Lambda_2 = 1$$

where

$$c = 2\sqrt{p}/\sqrt{1-2p-q}.$$

Thus, the optimal decision regions in likelihood space are known up to the parameter c. By trying various choices of c for the specific case $E/N_o = 1$ and calculating R_o for the DMC resulting from the demodulator corresponding to these decision regions, we find that, for the optimal decision regions, c = .486 and the attained value of R_o is 0.4402 which is a 0.45 dB improvement over hard decisions. In Fig. 5(a), we show the optimal decision regions in likelihood space, while in Fig. 5(b) we show the corresponding regions in signal space.



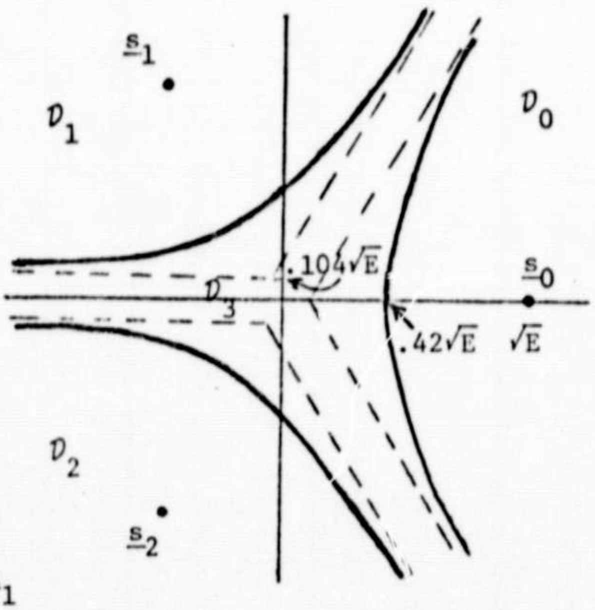


Fig. 5: Optimal decision regions for quaternary demodulation of ternary phase modulation in AWGN with $E/N_o = 1$ shown (a) in likelihood space, and (b) in signal space.

We now describe a general iterative procedure that may be used to find the optimal decision regions for J-ary demodulation given a particular M-ary signal set and a given channel. The basic idea is quite simple. Given decision regions $D_0, D_1, \ldots, D_{J-1}$ bounded by hyperplanes, we note that, for some $a \neq b$ such that D_a and D_b are adjacent regions, the intercepts of the bounding hyperplane with the coordinate axes in likelihood space will satisfy (13) if the decision regions are optimal. If they are not optimal, we can use the numbers determined by (13) as the intercepts of a hyperplane which will be a better approximation to the optimal bounding hyperplane between D_a and D_b. Our procedure may be stated as:

<u>Iterative Demodulator Optimization</u>:

<u>Step 0</u>: Make an initial guess, $D_0^{(1)}, D_1^{(1)}, \ldots, D_{J-1}^{(1)}$, for the optimal hyperplane-bounded decision regions in likelihood space. Set $k = 1$.

<u>Step 1</u>: Calculate $P(j|m)$, $0 \leq j < J$ and $0 \leq m < M$, for the DMC created by the decision regions $D_0^{(k)}, D_1^{(k)}, \ldots, D_{J-1}^{(k)}$.

<u>Step 2</u>: Choose an a and b, $a \neq b$, such that $D_a^{(k)}$ and $D_b^{(k)}$ are adjacent and calculate $T_1^{(k+1)}, T_2^{(k+1)}, \ldots, T_{M-1}^{(k+1)}$ from equation (13). [Note: If the decision regions are optimal, then these T's will be the intercepts of the boundary between $D_a^{(k)}$ and $D_b^{(k)}$ with the coordinate axes.]

<u>Step 3</u>: Take the boundary between $D_a^{(k+1)}$ and $D_b^{(k+1)}$ as the hyperplane whose intercepts with the coordinate axes are $T_1^{(k+1)}, T_2^{(k+1)}, \ldots, T_{J-1}^{(k+1)}$.

<u>Step 4</u>: Repeat steps (2) and (3) until all such pairs a and b have been considered.

<u>Step 5</u>: If $D_0^{(k+1)}, D_1^{(k+1)}, \ldots, D_{J-1}^{(k+1)}$ are "sufficiently close" to $D_0^{(k)}, D_1^{(k)}, \ldots, D_{J-1}^{(k)}$, stop and take the former decision regions as the result of this optimization method. Otherwise, increase k by 1 and return to step 1.

Two remarks about the above iterative method are in order. First, the most time-consuming part of the procedure is the calculation of the transition probabilities $P(j|m)$ in step 1. This would ordinarily be done by mapping from the decision regions $D_0^{(k)}, D_1^{(k)}, \ldots, D_{J-1}^{(k)}$ in likelihood space to the corresponding decision regions $\mathcal{D}_0^{(k)}, \mathcal{D}_1^{(k)}, \ldots, \mathcal{D}_{J-1}^{(k)}$ in signal space, then evaluating the integral in (5) either analytically or numerically. Secondly, we note that when $M = 2$, the above iterative procedure requires more calculation than the iterative method given by Massey [2] which is based on equation (16). Unfortunately, Massey's method does not generalize to cases where $M > 2$.

We now give two examples to illustrate the use of the above iterative optimization method. The first of these is a "hard-decision" case which serves also to illustrate the fact that, when the signal set is not sufficiently symmetric, even in this case the decision regions which minimize demodulator error probability may not maximize $\tilde{R}_o$.

Example 3: Hard-decision demodulation for the two-dimensional signal set $\underline{s}_0 = [0,0]$, $\underline{s}_1 = [2,0]$ and $\underline{s}_2 = [0,2]$ in AWGN with variance $N_o/2 = 1/2$ in each component. [The average signal-energy-to-noise-power-spectral-density-ratio, E/N_o, is 2.67 (or 4.26 dB).] In this case, the received vector $\underline{r} = [x,y]$ has a likelihood ratio vector

$$\underline{\Lambda}\ (\underline{r}) = [\Lambda_1, \Lambda_2] = [e^{4x-4}, e^{4y-4}].$$

Thus, the boundary straight-line (or hyperplane in two-dimensional space)

$$\frac{1}{T_1}\Lambda_1 + \frac{1}{T_2}\Lambda_2 = 1$$

corresponds to the curve

$$\frac{1}{T_1} e^{4x-4} + \frac{1}{T_2} e^{4y-4} = 1 \tag{17}$$

in signal space.

As the boundaries between decision regions in the initialization step 0 of the algorithm, we choose those which minimize demodulator error probability, viz., the straight lines $\Lambda_1 = 1$, $\Lambda_2 = 1$, and $\Lambda_1 = \Lambda_2$ shown in Fig. 6(a). Note that the boundary between D_o and D_2 is a reflection around the line $\Lambda_1 = \Lambda_2$ of that between D_o and D_1. This symmetry is preserved at successive iterations of the algorithm because of the corresponding symmetry of the signal set about the line x = y and the symmetry of the AWGN. Thus, it suffices to determine the boundary between D_o and D_1. Letting (T_1, T_2) be the intercepts of this boundary with the Λ_1 and Λ_2 axes, respectively, and taking $(1, -10^6)$ as an approximation to the initial $(1, -\infty)$, we find, from application of the interative procedure in which we use (17) to determine the region over which the integral in (5) is to be evaluated numerically, the following succession of intercepts:

$(1, -10^6)$	initialization
(0.8307, -0.5732)	1st iteration
(0.8145, -0.4030)	2nd iteration
(0.8252, -0.3501)	3rd iteration
⋮	⋮
(0.8646, -0.3100)	15th iteration

up to the point where there is no further change in the 4th significant digit on further iteration. In Figures 6(b) and 6(c), we show the optimal decision regions for this example, as found by the iterative optimization procedure, in likelihood space and in signal space, respectively.

The values of $\tilde{R}_o$ obtained as successive iterations were as follows:

0.6388	initialization
0.6462	1st iteration
0.6476	2nd iteration
0.6480	3rd iteration
⋮	⋮
0.6482	15th iteration.

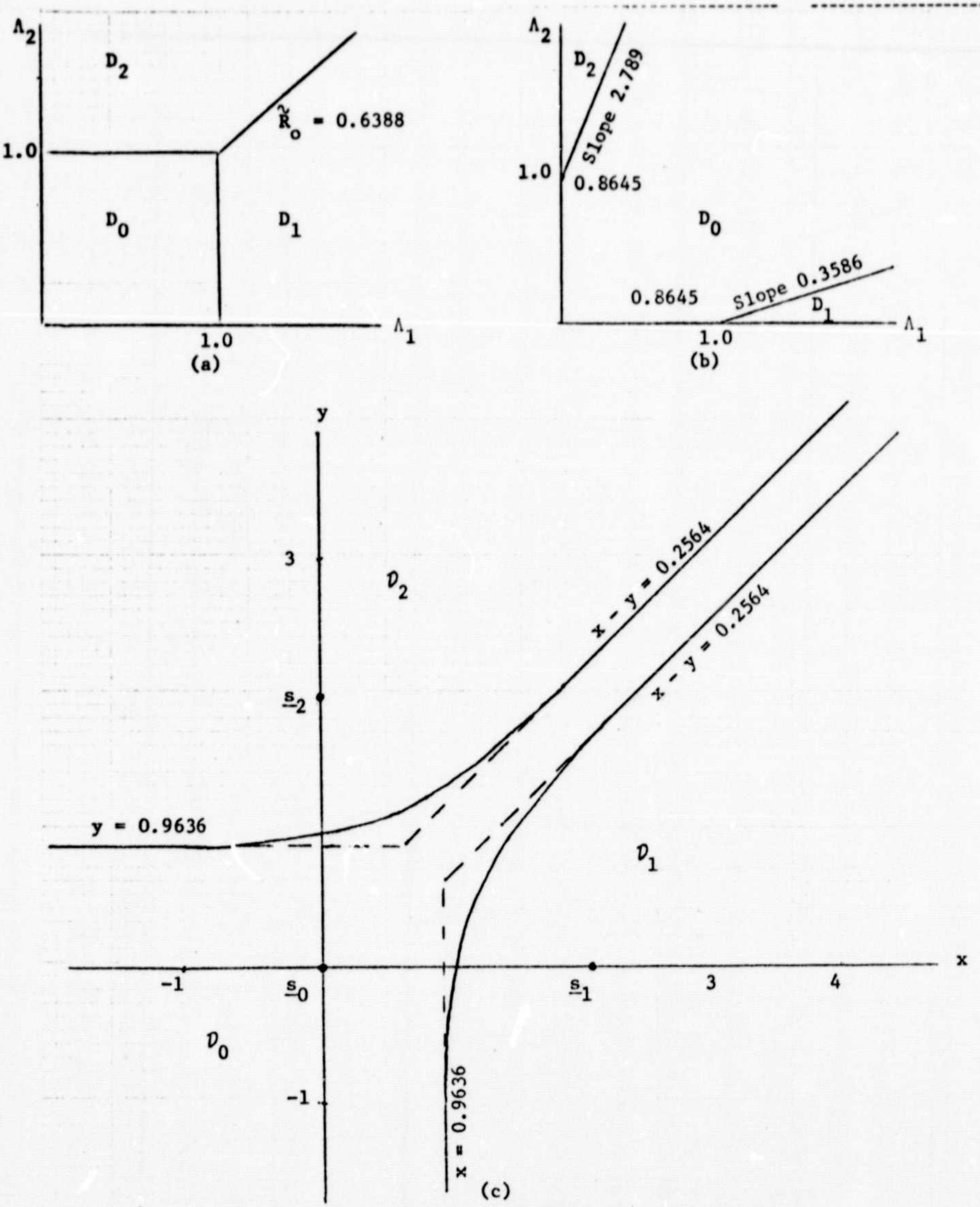


Figure 6. Hard decision demodulation of ternary signals:
(a) The initial decision regions in likelihood space,
(b) the optimal decision regions in likelihood space,
(c) the optimal decision regions in signal space.

The optimal value of $\tilde{R}_o$ is about 1.5% (or .06 dB) above that obtained for the hard-decision demodulator which minimizes error probability. We see also that, for this example, the iterative procedure converged rapidly to the optimal demodulator--after only one iteration the resulting demodulator was effectively optimal.

As can be seen from Fig. 6(c), the optimal decision boundaries have asymptotes which are straight lines (hyperplanes in two-dimensional space.) These asymptotes are shown by the dashed lines in the figure. If one uses these asymptotes as the boundaries of the decision regions for a sub-optimal demodulator, one finds the resultant $\tilde{R}_o$ to be 0.6459 which is only .015 dB inferior to the optimal hard-decision demodulator. We shall later discuss the practical significance of the near-optimality of these asymptotic linear boundaries in signal space.

<u>Example 4</u>: Quaternary demodulation ($J = 4$) for the same ternary signal set ($M = 3$) and noise as in example 3.

In Figure 7(a), we show the $J = 4$ decision regions used to initiate the iterative optimization procedure. The optimal decision regions in likelihood space and in signal space are shown in Figs. 7(b) and 7(c), respectively. Convergence to four significant digits of accuracy in the intercepts of the boundary lines with the coordinate axes required 25 iterations. The values of $\tilde{R}_o$ at successive steps were as follows:

0.6536	initialization
0.6986	1st iteration
0.7045	2nd iteration
0.7057	3rd iteration
⋮	⋮
0.7062	25th iteration;

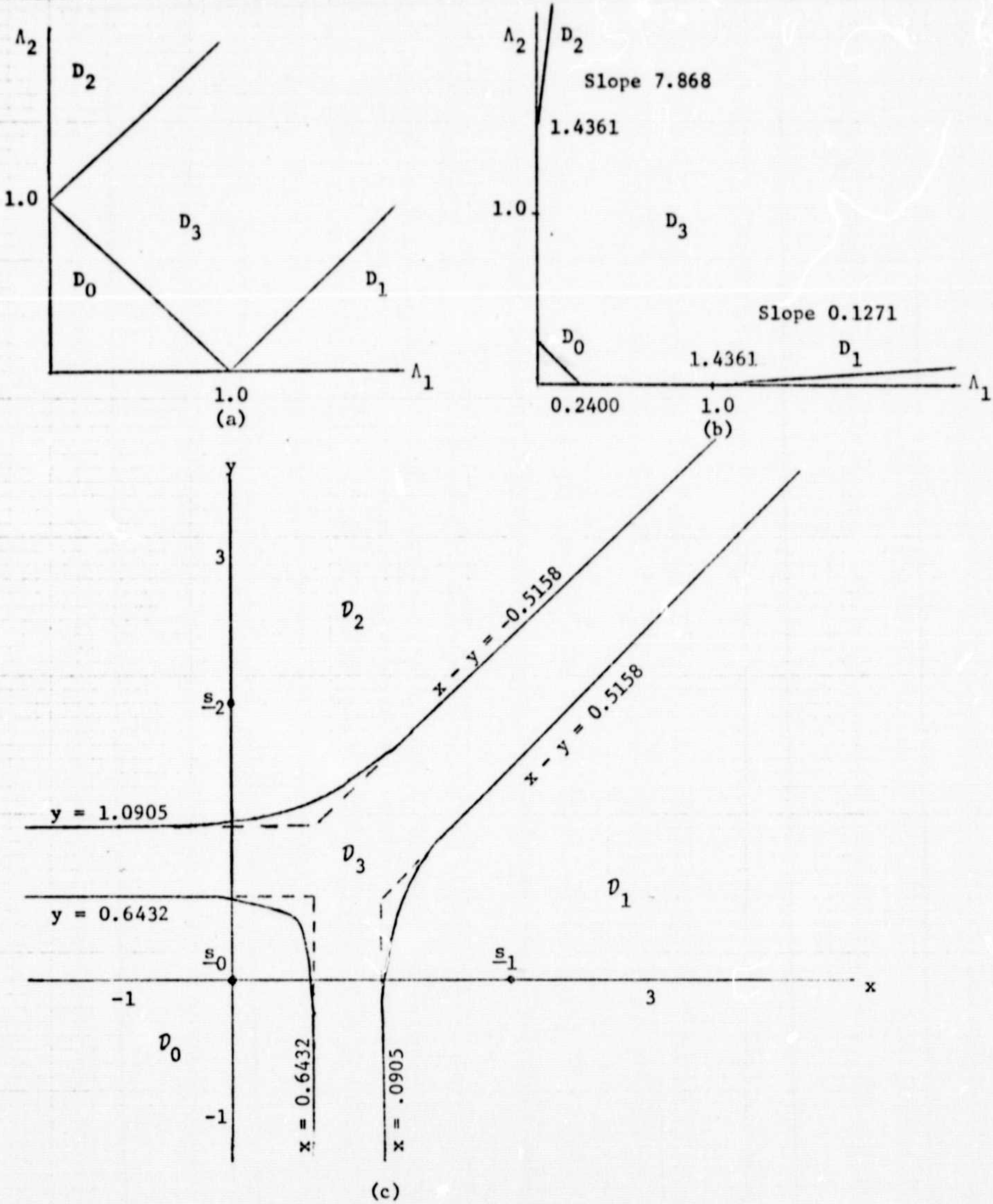


Figure 7. Quarternary demodulation of ternary signals:
(a) the initial decision regions likelihood space,
(b) the optimal decision regions in likelihood space,
(c) the optimal decision regions in signal space.

we see that the decision regions after only two iterations were effectively optimal. The optimal value of $\tilde{R}_o$ for quaternary demodulation is about 8% (or 0.34 dB) above that for the initial decision regions of Fig. 7(a), and is about 9% (or 0.37 dB) above that for optimal hard-decision demodulation.

In Fig. 7(c), the dashed lines (or hyperplanes in two-dimensional space) again are the asymptotes to the optimal decision boundaries in signal space. If these asymptotes are used as the actual boundaries between the decision regions in signal space, we find $\tilde{R}_o$ of the resultant demodulator to be .7031 which is only .02 dB below optimal.

In both Examples 3 and 4, we have seen that the linear (hyperplane) asymptotes to the optimal decision regions in signal space themselves bound the decision regions for a demodulator which is virtually optimal. The practical significance of this fact is that the resulting sub-optimal decision rule can be as easily implemented directly in signal space as can the optimal decision rule in likelihood space. There is no need for the conversion from signal space to likelihood space in order to obtain conveniently-implemented decision regions with linear (hyperplane) boundaries.

In fact, it can be shown generally, for AWGN in an n-dimensional signal space, that the optimal demodulation regions in signal space are such that each bounding hypersurface has (at most) 2^{n-1} hyperplane asymptotes. We conjecture that these hyperplane asymptotes form the boundaries of decision regions for a demodulator that is virtually optimum, and hence that the mapping from signal space to likelihood space is not necessary to obtain virtually optimal demodulation together with an easily-implemented decision rule.

VI. A COUNTEREXAMPLE TO THE SUFFICIENCY OF OPTIMALITY CONDITION (10)

As we have pointed out above, condition (10) is actually the condition for an extremum of $\tilde{R}_o$, and hence not in general a sufficient condition for optimal demodulation. In the examples which we have studied wherein the noise was additive with a "smooth" density function, there has generally been only one set of decision regions satisfying (10) so that the extremum was necessarily the global maximum of $\tilde{R}_o$. We now give an example to show, however, that it is possible for (10) to be satisfied for decision regions that define only a local maximum, or even a local minimum, of $\tilde{R}_o$.

Example 5: Hard-decision demodulation for binary signals such that the conditional density functions for the likelihood ratio Λ are

$$p_0(\Lambda) = \begin{cases} 0.25 & 0 < \Lambda \leq 0.9 \\ 2.75 & 0.9 < \Lambda \leq 1.1 \\ 0.25 & 1.1 < \Lambda \leq 2.0 \\ 0 & 2.0 < \Lambda \end{cases}$$

and $p_1(\Lambda) = \Lambda p_0(\Lambda)$ when the signals $\underline{s}_0$ and $\underline{s}_1$ are transmitted, respectively. [These are valid choices for these conditional density functions as they satisfy the constraints

$$\int_0^\infty p_0(\Lambda)d\Lambda = \int_0^\infty p_1(\Lambda)d\Lambda$$

and

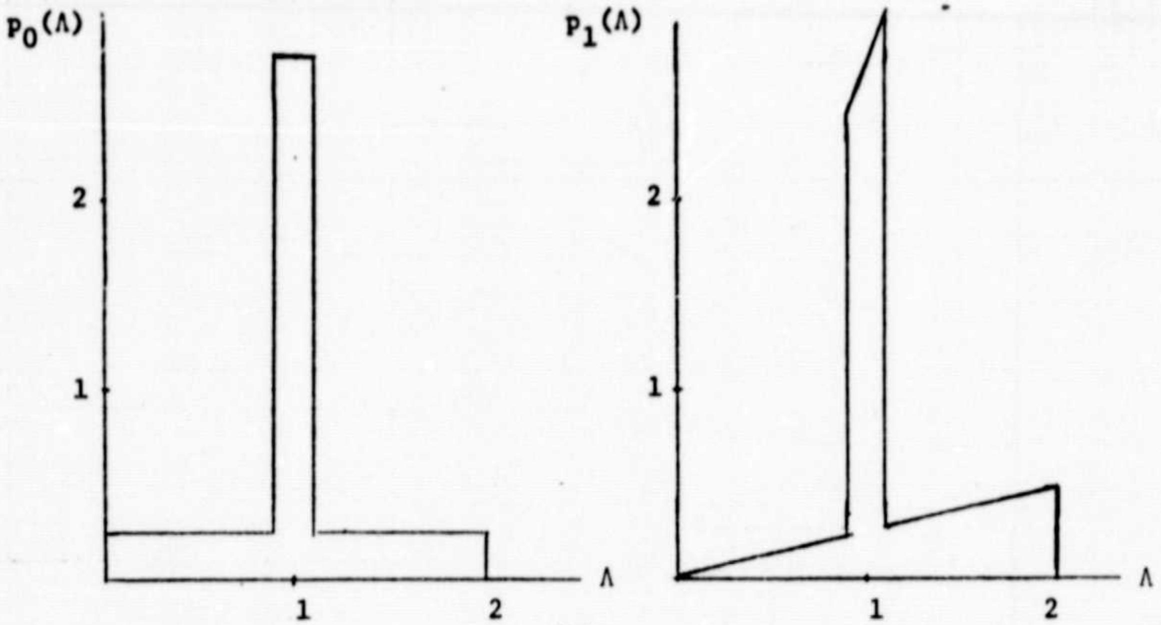
$$\Lambda = p_1(\Lambda)/p_0(\Lambda)$$

that are the only ones that must be observed in the binary case.]

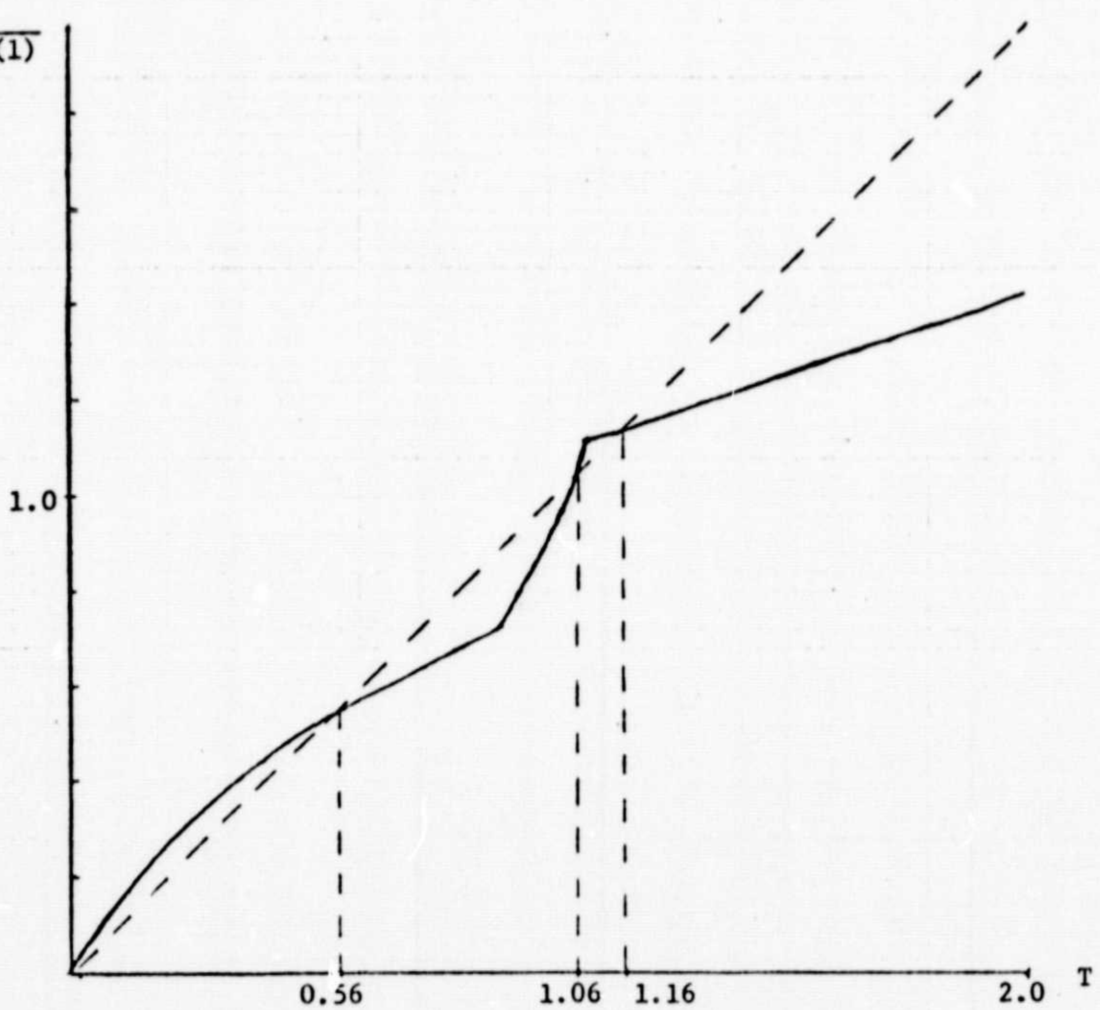
For hard-decision demodulation with binary signalling, condition (10) reduces to Massey's condition (16), namely,

$$T = \sqrt{\lambda(1)\lambda(0)} \tag{18}$$

where T is the threshold between decision regions in likelihood space.



(a)



(b)

Figure 8. The binary, hard-decision, demodulation situation used to demonstrate the insufficiency of optimality condition (10).

In Fig. 8(a), we show the conditional density functions for the likelihood ratio Λ, and in Fig 8(b) we show $\sqrt{\lambda(1)\lambda(0)}$ as a function of the threshold T between the decision regions in likelihood space. We see that condition (18), the necessary condition for optimal demodulation, is satisfied at three places, viz., 0.56, 1.06, and 1.16; the corresponding values of $\tilde{R}_o$ are 0.0123, 0.0066, and 0.0069, respectively. The third of these corresponds to a local, but not global, maximum of $\tilde{R}_o$. The second corresponds to a local minimum of $\tilde{R}_o$, while the first corresponds to the desired global maximum of $\tilde{R}_o$. That is, T = 0.56 is the threshold between decision regions in likelihood space for the optimal demodulator.

VII. SUMMARY

In this note, we have derived a necessary condition for optimal J-ary demodulation of M-ary signals, where optimality is taken to mean maximality of the symmetric cut-off rate, $\tilde{R}_o$, of the resulting discrete memoryless channel. By means of a counterexample, we have shown that this condition is not in general sufficient for optimality. We have also used this necessary condition for optimality as the foundation for an iterative optimization method to find the optimal demodulator decision regions from an initial "good guess."

In general, the optimal demodulator decision regions are bounded by hyperplanes in likelihood space. For the important case of additive white Gaussian noise, the corresponding optimal decision regions in signal space have hyperplane asymptotes. In some examples, we have shown that the regions in signal space bounded by these asymptotic hyperplanes define demodulator decision regions which are virtually optimal, and we conjectured that this happy state of affairs (which permits near optimal performance with a decision rule that can be simply-implemented directly in signal space) holds in general.

ACKNOWLEDGMENT

I would like to thank Professor James L. Massey for his guidance and suggestions during the course of this research. He provided the motivation to generalize his necessary condition for optimal demodulation of binary signals, as well as shaped this note into its present form.

REFERENCES

[1] J. M. Wozencraft and R. S. Kennedy, "Modulation and Demodulation for Probabilistic Coding," IEEE Trans. on Information Theory, Vol. IT-12, pp. 291-297, July 1966.

[2] J. L. Massey, "Coding and Modulation in Digital Communications," Proceedings, International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 12-15, 1974, pp. E2(1)-E2(4).

[3] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," IEEE Trans. on Information Theory, IT-13, pp. 260-269, April 1967.

[4] J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering. New York: John Wiley and Sons Inc., 1965.

July 31, 1975

<u>Errata</u> for:

"Semi-Annual Status Report, Coordinated Design of Coding and Modulation Systems, NASA Grant NSG 5025," Dept. of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, June 25, 1975.

Page 2, K = 9 row in Table I, present entry of "7" should be "8"

Page 4, K = 9 row under Systematic Codes in Table II, present entry of "7C*" should be "7C"

July 31, 1975

Errata for:

J. L. Massey, "A Recommended R = 1/2, K = 32, Quick-Look-In Convolutional Code for NASA Use," Tech. Rpt. No. EE-751, Dept. of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, April 28, 1975.

Page 5, line 16: "all 10,000 frames" should be "all but 23 of the 10,000 frames"

Page 10, second and third columns of Table III: last two entries, namely "69" and "0", should be "46" and "23" respectively in both columns.

July 31, 1975

<u>Errata</u> for:

J. L. Massey, "A Recommended R = 1/2, K = 32, Quick-Look-In Convolutional Code for NASA Use," Tech. Rpt. No. EE-751, Dept. of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, April 28, 1975.

Page 5, line 16: "all 10,000 frames" should be "all but 23 of the 10,000 frames"

Page 10, second and third columns of Table III: last two entries, namely "69" and "0", should be "46" and "23" respectively in both columns.

July 31, 1975

Errata for:

"Semi-Annual Status Report, Coordinated Design of Coding and Modulation Systems, NASA Grant NSG 5025," Dept. of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana, June 25, 1975.

Page 2, K = 9 row in Table I, present entry of "7" should be "8"

Page 4, K = 9 row under Systematic Codes in Table II, present entry of "7C*" should be "7C"