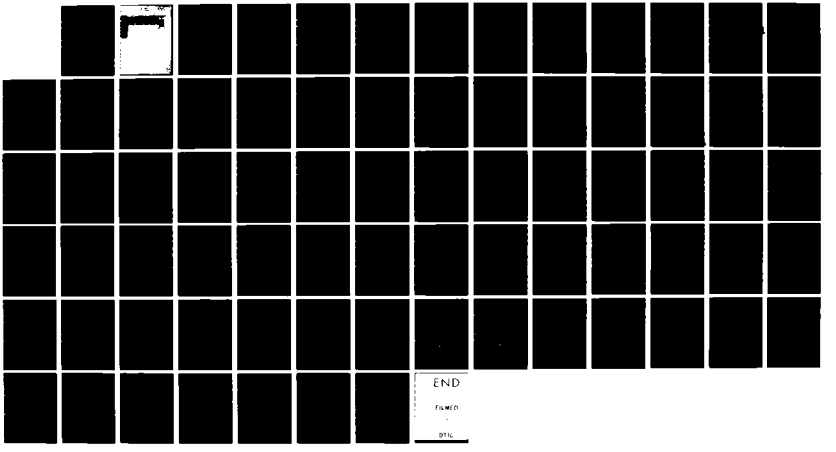
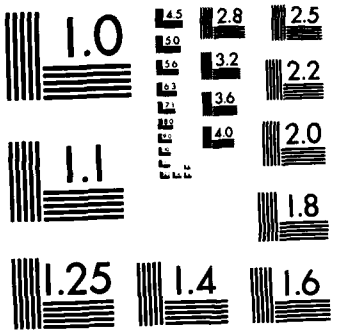
AD-A123 990 SOME DESIGN CONSIDERATIONS ON THE ADAPTIVE QUANTIZER DETECTOR IN UNKNOWN NOISE(U) ILLINOIS UNIV AT URBANA COORDINATED SCIENCE LAB W H LI APR 80 R-881 1/1

UNCLASSIFIED DAAG29-78-C-0016 F/G 9/4 NL





UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A123 990	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* SOME DESIGN CONSIDERATIONS ON THE ADAPTIVE QUANTIZER DETECTOR IN UNKNOWN NOISE		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER R-881;UILU-ENG 80-2213
7. AUTHOR(s) Wing Hong Li		8. CONTRACT OR GRANT NUMBER(s) DAAG-29-78-C-0016; N00014-79-C-0424; ENG 78-05880
9. PERFORMING ORGANIZATION NAME AND ADDRESS Coordinated Science Laboratory University of Illinois at Urbana-Champaign Urbana, Illinois 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Joint Services Electronics Program; Office of Naval Research; National Science Foundation		12. REPORT DATE April 1980
		13. NUMBER OF PAGES 71 pages
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Adaptive Signal Detection
Data Quantization
Stochastic Approximation

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

The contents of this report provide some general guidelines and insights into the design of an adaptive signal detection system based on quantized data. In particular, we focus on the adaptation of the parameters in the quantizer in a way such that the detector's probability of making errors in the presence of noise of unknown statistics will be minimized with respect to those parameters. The unknown noise is assumed throughout this report to be an independent, additive noise with a symmetric (about zero) probability density function. Within this assumption an adaptive scheme is developed, and its performance

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. (continued)

and convergence are verified via simulation.

UILU-ENG 80-2213

SOME DESIGN CONSIDERATIONS ON THE ADAPTIVE QUANTIZER DETECTOR IN UNKNOWN NOISE

by

Wing Hong Li

This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under contract DAAG-29-78-C-0016 and N00014-79-C-0424 and in part by the National Science Foundation under Grant ENG 78-05880.

SOME DESIGN CONSIDERATIONS ON THE ADAPTIVE QUANTIZER DETECTOR IN UNKNOWN NOISE

BY

WING HONG LI

B.S., University of California, Berkeley, 1977

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1980

Thesis Adviser: Professor H. V. Poor

Urbana, Illinois

ABSTRACT

The contents of this report provide some general guidelines and insights into the design of an adaptive signal detection system based on quantized data. In particular, we focus on the adaptation of the parameters in the quantizer in a way such that the detector's probability of making errors in the presence of noise of unknown statistics will be minimized with respect to those parameters. The unknown noise is assumed throughout this report to be an independent, additive noise with a symmetric (about zero) probability density function. Within this assumption an adaptive scheme is developed, and its performance and convergence are verified via simulation.

Acknowledgement

The author wishes to thank Professor H. V. Poor for his suggestions and assistance during the course of this work. Also, a special thanks is extended to Mrs. Phyllis Young for her excellent typing.

TABLE OF CONTENTS

1. INTRODUCTION

The detection of signals in the presence of independent noise can very well be accomplished if the noise characteristics were fully known; in particular, if the noise probability density function is available, an optimal detector can be formulated. However, it is often the case that the noise physical mechanism is unknown or too complex to be expressed in any simple way; moreover, the noise characteristics may not be stationary in time or space and may otherwise be impossible to represent by any fixed models.

Under these circumstances, a different approach to optimal detection is necessary which borrows the idea from adaptation. The adaptation process learns what the noise actually is at that moment and "adjusts" the detector's structure in a way to result in near-optimal detection performance.

The contents of this thesis provide some general guidelines and insights into the design of an adaptive detector based on quantized data. In particular, we focus on the adaptation of the parameters in the quantizer in a way such that the detector's probability of making errors will be minimized with respect to those parameters. The unknown noise is assumed throughout this thesis to be an independent, additive noise with a symmetric (about zero) probability density function.

2. SOME BASICS OF BINARY DETECTION WITH QUANTIZATION

2.1 Structure of an Optimal Detector Based on a Partitioned Sample Space

In binary state data communication, we transmit a positive-valued or negative-valued signal depending on whether the state is "1" or "0". After the signal has been transmitted through an additive-noise channel, the receiver determines from the observation whether it contains the positive-valued or the negative-valued signal. This problem can also be viewed as a binary hypothesis testing problem in which the hypothesis H_0 is tested versus the alternative hypothesis H_1; i.e.,

$$H_0 : Y_i = -s + N_i \qquad i = 1,2,3,\ldots,n$$

versus (2.1)

$$H_1 : Y_i = s + N_i \qquad i = 1,2,3,\ldots,n$$

where the positive and negative signals are of the same strength s and $\{N_i\}$ is a sequence of additive noise samples which are independent and identically distributed (i.i.d.) random variables with common probability density function $f_N(\cdot)$.

It is well known that the following detection scheme gives the best receiver structure (in terms of the minimum probability of error) in deciding H_0 vs. H_1:

$$\text{Decide} \begin{cases} H_0 \\ H_1 \\ H_0 \text{ or } H_1 \end{cases} \quad \text{if} \quad \frac{p(\underline{Y}|H_1)}{p(\underline{Y}|H_0)} \begin{matrix} < \\ > \\ = \end{matrix} \tau \tag{2.2}$$

where $p(\underline{Y}|H_i); i = 0,1$ is the probability density or mass function of $\underline{Y}$ ($\underline{Y} = (Y_1Y_2Y_3\ldots Y_n)$) given that H_i is the true hypothesis and τ is the

threshold to which the probability ratio is compared. The ratio $p(\underline{y}|H_1)/p(\underline{y}|H_0)$ is often called the likelihood ratio and is a function of the observation vector $\underline{y}$. If we assume the two hypotheses H_0 and H_1 are equally likely to occur, that is, they have equal a priori probability ($Pr\{H_i\} = \frac{1}{2}$; i = 0,1), and that the penalties on incorrect decision under either hypothesis are the same (no penalty on correct decision) then the best τ is 1.

Since the logarithmic function is monotonically increasing with its argument, we can write (2.2) in a different but equivalent way,

$$\text{decide} \begin{cases} H_0 \\ H_1 \\ H_0 \text{ or } H_1 \end{cases} \quad \text{if } \log \frac{p(\underline{y}|H_1)}{p(\underline{y}|H_0)} \begin{matrix} < \\ > \\ = \end{matrix} \log \tau \qquad (2.3)$$

where the function $\log(p(\underline{y}|H_1)/p(\underline{y}|H_0))$ is often called the log-likelihood ratio function of the observation $\underline{y}$.

In the problem to be considered here, we take a fixed finite number (n) of observations to determine whether H_0 of H_1 has occurred and partition the observation space (the real line in this case) into a finite number (m) of intervals. Then the detector structure of Eq. (2.2) gives an optimal m-level quantizer-detector with m preset partition intervals.

Let n_i be the number of samples from observations $\{y_i\}_{i=1}^{n}$ that fall in the i-th interval, and let p_i^0 and p_i^1 be the probabilities that the value of the observation belongs to the i-th interval under hypotheses H_0 and H_1, respectively. Hence $\sum_{i=1}^{m} n_i = n$, the total number of observations, and $\sum_{i=1}^{m} p_i^0 = 1$ and $\sum_{i=1}^{m} p_i^1 = 1$.

With the partitioning of the observation space, the probability distributions of $n = (n_1 n_2 n_3 \ldots n_m)$ given hypotheses H_0 and H_1 are multinomial, i.e.,

$$p(\underline{n}|H_0) = \frac{n!}{n_1!n_2!n_3!\ldots n_m!} (p_1^0)^{n_1}(p_2^0)^{n_2}(p_3^0)^{n_3}\ldots(p_m^0)^{n_m}$$

$$p(\underline{n}|H_1) = \frac{n!}{n_1!n_2!n_3!\ldots n_m!} (p_1^1)^{n_1}(p_2^1)^{n_2}(p_3^1)^{n_3}\ldots(p_m^1)^{n_m} \tag{2.4}$$

Now Eq. (2.2) can be written as

$$\text{decide} \begin{cases} H_0 \\ H_1 \\ H_0 \text{ or } H_1 \end{cases} \quad \text{if} \left(\frac{p_1^1}{p_1^0}\right)^{n_1}\left(\frac{p_2^1}{p_2^0}\right)^{n_2}\left(\frac{p_3^1}{p_3^0}\right)^{n_3}\cdots\left(\frac{p_m^1}{p_m^0}\right)^{n_m} \begin{matrix} < \\ > 1 \\ = \end{matrix} \tag{2.5}$$

Note that τ has been taken as 1 for the remainder of this thesis. Again since the logarithmic function is nondecreasing, (2.5) has the following equivalent form,

$$\text{decide} \begin{cases} H_0 \\ H_1 \end{cases} \quad \text{if} \sum_{i=1}^{m} n_i \log\left(\frac{p_i^1}{p_i^0}\right) \begin{matrix} < \\ \\ \geq \end{matrix} 0 \tag{2.6}$$

We arbitrarily incorporate the decision on H_1 when the test statistic $\sum_{i=1}^{m} n_i \log\left(\frac{p_i^1}{p_i^0}\right)$ equals zero; this will not affect the error-probability performance of the detector with equally likely H_0 and H_1.

2.2 The Optimal Quantizer-Detector

An m-level quantizer $Q(\cdot)$ is characterized by the two sets of numbers, $\underline{q} = (q_1 q_2 q_3 \ldots q_m)$ (the quantization levels) and $\underline{t} = (-\infty = t_0 < t_1 < \ldots < t_{m-1} < t_m = \infty)$ (the breakpoints) which partition the real line (i.e., the observation space) into m intervals. The function of the quantizer is to set $Q(y) = q_k$ if the observation y is such that $t_k \leq y < t_{k+1}$ as shown in Fig. 1.

Since $\{N_i\}_{i=1}^n$ are i.i.d. random variables and $Y_i = \pm S + N_i$ $i = 1,2,3,\ldots,n$,

$$\log \frac{p(\underline{y}|H_1)}{p(\underline{y}|H_0)} = \sum_{i=1}^{n} \log \frac{p(y_i|H_1)}{p(y_i|H_0)} \quad . \tag{2.7}$$

This equation (2.3) can be written as

$$\text{decide} \begin{cases} H_0 \\ H_1 \end{cases} \text{ if } \sum_{i=1}^{n} \log \frac{p(y_i|H_1)}{p(y_i|H_0)} \begin{matrix} < \\ \geq \end{matrix} 0 \quad . \tag{2.8}$$

Here again we have incorporated with the decision on H_1 when equality holds in the comparison. If the log-likelihood ratio function of the observation in (2.8) is quantized, we obtain a quantized version of the detector; i.e.,

$$\text{decide} \begin{cases} H_0 \\ H_1 \end{cases} \text{ if } \sum_{i=1}^{n} Q\left(\log\left(\frac{p(y_i|H_1)}{p(y_i|H_0)}\right)\right) \begin{matrix} < \\ \geq \end{matrix} 0 \tag{2.9}$$

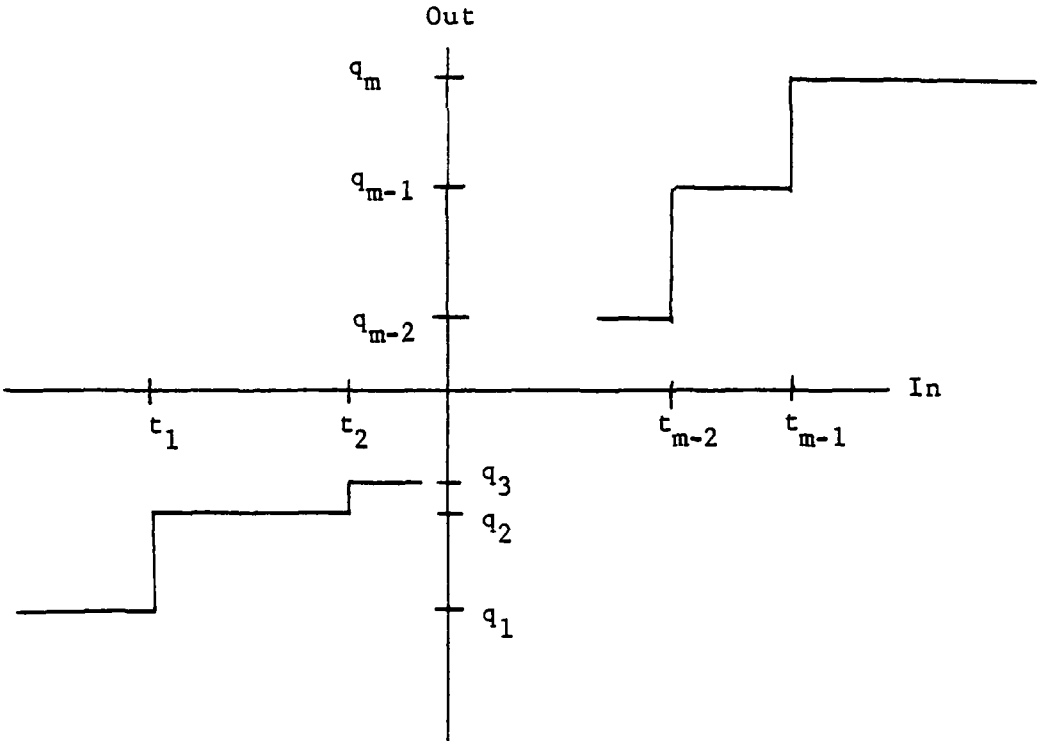


Figure 1. m-level quantizer parameters.

This can be reformed into the following by the property of the quantizer $Q(\cdot)$,

$$\text{decide} \begin{cases} H_0 & \\ & \text{if} \quad \sum_{i=1}^{m} n_k q_k \quad \begin{matrix} < \\ \\ \geq \end{matrix} \quad 0 \\ H_1 & \end{cases} \tag{2.10}$$

Now we obtain the structural form of the general quantizer detector of this problem.

Since (2.6) gives the best detector with predetermined partitions of the observation space, it can be easily recognized that if we set the q_k in (2.10) to be $\log(p_k^1/p_k^0)$, $k = 1,2,3,\ldots,m$, the resulting m-level quantizer, $Q^0(\cdot)$, with the same prechosen breakpoints, is optimal in detection performance.

The following result shows the equivalence of the test in (2.10) with $q_k = q_k^0 \triangleq \log(p_k^1/p_k^0)$ and the test in (2.3) with $\tau = 1$ as m, the number of quantization levels goes infinite.

Property: Given the following conditions on a binary hypothesis testing problem,

(1) under either H_0 or H_1; $\{Y_i\}_{i=1}^n$ are i.i.d. random variables;

(2) the cumulative distribution functions of Y_i under both H_0 and H_1, ($F_Y(y|H_0)$ or $F_Y(y|H_1)$, respectively) are continuously differentiable and strictly increasing,

then

$$\lim_{m \to \infty} \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) = \log\left(\frac{p(\underline{y}|H_1)}{p(\underline{y}|H_0)}\right) \tag{2.11}$$

where $\underline{Y} = (Y_1 Y_2 Y_3 \ldots Y_n)$, p_k^i is the probability that Y is in the k-th interval given H_i, i = 0,1 and n_k is the number of samples from $\{Y_i\}_{i=1}^n$ that fall in the k-th interval.

Proof:
$$\sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) = \sum_{i=1}^{n} \log\left(\frac{p_k^1(i)}{p_k^0(i)}\right) \tag{2.12}$$

where $p_k^j(i)$ = probability that y_i belongs to the k-th interval under H_j. Since $p_k^i(i) > 0$ and the logarithmic function is continuous

$$\lim_{m \to \infty} \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) = \sum_{i=1}^{n} \log\left\{\lim_{m \to \infty} \frac{p_k^1(i)}{p_k^0(i)}\right\} \tag{2.13}$$

By L'Hôpital's rule

$$\lim_{m \to \infty} \frac{p_k^1(i)}{p_k^0(i)} = \frac{p(y_i|H_1)}{p(y_i|H_0)} \tag{2.14}$$

Hence

$$\sum_{i=1}^{n} \log\left\{\lim_{m \to \infty} \frac{p_k^1(i)}{p_k^0(i)}\right\} = \sum_{i=1}^{n} \log\left(\frac{p(y_i|H_1)}{p(y_i|H_0)}\right) \tag{2.15}$$

and

$$\lim_{m \to \infty} \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) = \sum_{i=1}^{n} \log\left(\frac{p(y_i|H_1)}{p(y_i|H_0)}\right)$$

$$= \log \prod_{i=1}^{n} \frac{p(y_i|H_1)}{p(y_i|H_0)}$$

$$= \log\left(\frac{p(\underline{y}|H_1)}{p(\underline{y}|H_0)}\right)$$

Q.E.D.

2.3 The Locally Optimum Quantizer-Detector

In the detection of weak signals, the maximum power slope at s = 0 is an appropriate criterion. It can be shown that the locally optimal test statistic $T_{\ell o}(\underline{y})$ for the binary hypothesis testing problem of (2.1) can be obtained by differentiating the log-likelihood function in (2.3) with respect to s and setting s = 0,

$$T_{\ell o}(\underline{y}) = \sum_{i=1}^{n} 2\, g_{\ell o}(y_i) \tag{2.16}$$

where $g_{\ell o}(y) = -f_N'(y)/f_N(y)$. The generalized Neyman-Pearson lemma asserts that test with $T_{\ell o}(\cdot)$ as a test statistic mazimizes the power slope at s = 0 over all tests. The corresponding locally optimal quantized test statistics with a given set of breakpoints $\underline{t}$ is (see Kassam [2])

$$T_{Q0}(\underline{y}) = \sum_{i=1}^{n} 2\, Q_{\ell o}(y_i) \tag{2.17}$$

where $Q_{\ell o}(y)$ is the optimal quantized version of $g_{\ell o}(y)$ in the minimum-mean-squared-error sense (under H_0) with breakpoints $\underline{t}$, and it can be shown to be

$$Q_{\ell o}(y) = \frac{f_N(t_k)-f_N(t_{k-1})}{F_N(t_{k-1})-F_N(t_k)} \triangleq q_k^{\ell o} \; ; \; t_k \le y < t_{k+1} \tag{2.18}$$

where $F_N(\cdot)$ is the noise cumulative distribution function. As before, (2.17) can be reformed to give

$$T_{Q0}(\underline{y}) = \sum_{k=1}^{m} 2\, n_k q_k^{\ell o} \tag{2.19}$$

Similar results can be obtained by differentiating the optimal quantized test statistic of (2.10) with respect to s and setting s = 0,

$$\frac{d}{ds}\left[\sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right)\right]_{s=0} = \sum_{k=1}^{m} 2\, n_k \left(\frac{f_N(t_k)-f_N(t_{k-1})}{F_N(t_{k-1})-F_N(t_k)}\right) \tag{2.20}$$

Hence

$$\frac{d}{ds}\left[\sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right)\right]_{s=0}$$

is also the locally optimal quantized test statistic with the same breakpoints $\underline{t}$. Again the generalized Neyman-Pearson lemma asserts that a quantizer detector with

$$\frac{d}{ds}\left[\sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right)\right]_{s=0}$$

as its test statistic maximizes the power slope at s = 0 over all quantizer detectors of the form in (2.10).

It is known [2] that the optimal breakpoints $\underline{t}$ in the weak-signal case can be obtained by solving the following two sets of equations,

$$q_k^{\ell o} = \frac{f_N(t_k)-f_N(t_{k-1})}{F_N(t_{k-1})-F_N(t_k)} \qquad k = 1,2,3,\ldots,m \tag{2.21}$$

and

$$\frac{q_k^{\ell o} + q_{k+1}^{\ell o}}{2} = g_{\ell o}(t_k) \qquad k = 1,2,3,\ldots,m-1 \tag{2.22}$$

where $g_{\ell o}(t_k) = -f_N'(t_k)/f_N(t_k)$.

2.4 Some Properties on the Asymptotically and Locally Optimal Quantizer-Detector

Finally we want to show that, as the signal strength $s \to 0$ and the sample size $n \to \infty$ simultaneously, the set of breakpoints $\underline{t}$ which minimizes an upper bound to the probability of error in detection using the locally optimal quantization levels of (2.18) approaches the set of locally optimal breakpoints given by (2.21) and (2.22). But first we need two properties which give an upper bound to the probability of error and the symmetrical property of the optimal breakpoints minimizing the probability of error in detection of the problem depicted by (2.1) with equal a priori probabilities for H_0 and H_1 and symmetric noise density function.

Since $Pr\{H_0\} = Pr\{H_1\} = \frac{1}{2}$, the probability of error in detection using (2.6) is

$$P_e = \tfrac{1}{2} \sum_{\underline{n} \in N_1^1} P_r\{\underline{n}|H_0\} + \tfrac{1}{2} \sum_{\underline{n} \in N_2} P_r\{\underline{n}|H_1\} , \tag{2.23}$$

where

$$Pr\{\underline{n}|H_i\} = \frac{n!}{n_1!n_2!n_3!\ldots n_m!} (p_1^i)^{n_1}(p_2^i)^{n_2}(p_3^i)^{n_3}\ldots(p_m^i)^{n_m}$$

$$N_1^1 = \{\underline{n} = (n_1,n_2,n_3,\ldots,n_m) \text{ s.t. } \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) \geq 0\}$$

and

$$N_2 = \{\underline{n} = (n_1,n_2,n_3,\ldots,n_m) \text{ s.t. } \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) < 0\}$$

Notice that p_k^i's are functions of the breakpoints; thus, within smoothness conditions, the partial derivative of P_e with respect to the breakpoints yields a necessary condition on the optimal breakpoints [1],

$$\frac{\partial P_e}{\partial t_j} = \tfrac{1}{2} \sum_{\underline{n} \in N_1^1} \Pr\{\underline{n}|H_0\} \left[\frac{n_j}{p_j^0} - \frac{n_{j+1}}{p_{j+1}^0} \right] f_N(t_j+s)$$

$$+ \tfrac{1}{2} \sum_{\underline{n} \in N_2} \Pr\{\underline{n}|H_1\} \left[\frac{n_j}{p_j^1} - \frac{n_{j+1}}{p_{j+1}^1} \right] f_N(t_j-s) = 0 \qquad (2.24)$$

In general it is necessary to use a gradient search technique to solve for the optimal breakpoints with (2.24), since it is not likely that a closed form solution to this equation can be found.

The following property gives the symmetric property on $\underline{t}$; thus half of the breakpoints can be determined from their symmetric counterparts.

Property: Given the following conditions

(1) The observations $(Y_1, Y_2, Y_3, \ldots, Y_n)$ are i.i.d. random variables.

(2) $\Pr\{H_0\} = \Pr\{H_1\} = \frac{1}{2}$; i.e., H_1 and H_0 are equally likely to occur.

(3) The cumulative distribution functions of the observation y under H_0 and H_1, $F_Y(y|H_0)$ and $F_Y(y|H_1)$ are continuous and symmetrical in the sense that $F_Y(-y|H_0) = 1 - F_Y(y|H_1)$.

(4) $\underline{t} = (t_1, t_2, t_3, \ldots, t_{m-1})$ are symmetric breakpoints in the sense that $t_{m-j} = -t_j \quad j = 1,2,3,\ldots,m-1$.

(5) The j^{th} component of $\underline{t}$ is optimal so that with (2.24) $\left.\frac{\partial P_e}{\partial t_j}\right|_{\underline{t}} = 0$. Then the $(m-j)^{th}$ component of $\underline{t}$ is also optimal; that is $\left.\frac{\partial P_e}{2t_{m-j}}\right|_{\underline{t}} = 0$.

Proof: Given that $F_Y(y|H_i)$; $i = 0,1$ and $\underline{t}$ are symmetrical, the following four statements are true,

(a) $f_Y(t_j|H_0) = f_Y(t_{m-j}|H_1)$

(b) $p_j^0 = p_{m-j+1}^1$

(c) $\sum_{\underline{n} \in N_1} \Pr\{\underline{n}|H_0\} n_j = \sum_{\underline{n} \in N_2} \Pr\{\underline{n}|H_1\} n_{m-j+1}$

(d) $\sum_{\underline{n} \in N_3} \Pr\{\underline{n}|H_0\} n_j = \sum_{\underline{n} \in N_3} \Pr\{\underline{n}|H_1\} n_{m-j+1}$

where

$$N_1 = \{\underline{n} \text{ s.t. } \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) > 0\}$$

$$N_3 = \{\underline{n} \text{ s.t. } \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) = 0\}$$

$$N_2 = \{\underline{n} \text{ s.t. } \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) < 0\}$$

Hence, given t_j is optimal, $\Pr\{H_0\} = \Pr\{H_1\} = \frac{1}{2}$ and the above statements, (2.24) can be written as

$$\frac{\partial P_e}{\partial t_j} = \tfrac{1}{2} \sum_{\underline{n} \in N_2} \Pr\{\underline{n}|H_1\} \left[\frac{n_{m-j+1}}{p_{m-j+1}^1} + \frac{n_{m-j}}{p_{m-j}^1} \right] f_N(t_{m-j}-s)$$

$$+ \tfrac{1}{2} \sum_{\underline{n} \in N_3} \Pr\{\underline{n}|H_0\} \left[\frac{n_{m-j+1}}{p_{m-j+1}^1} - \frac{n_{m-j}}{p_{m-j}^1} \right] f_N(t_{m-j}+s)$$

$$+ \tfrac{1}{2} \sum_{\underline{n} \in N_1} \Pr\{\underline{n}|H_0\} \left[\frac{n_{m-j+1}}{p_{m-j+1}^1} - \frac{n_{m-j}}{p_{m-j}^1} \right] f_N(t_{m-j}+s)$$

$$= - \frac{\partial P_e}{\partial t_{m-j}}$$

Hence $\frac{\partial P_e}{\partial t_j} = 0$ implies $\frac{\partial P_e}{\partial t_{m-j}} = 0$, thus $t_{m-j} = (-t_j)$ is also optimal.

Q.E.D.

From this property we can see that the optimal quantizer is odd symmetric,

$$q_{-k}^0 = \log\left(\frac{p_{-k}^1}{p_{-k}^0}\right) = \log\left(\frac{p_k^0}{p_k^1}\right) = -q_k^0 \tag{2.25}$$

The next result gives the upper bound to the probability of error in detection of (2.6).

Property: Given the test statistic $T = \sum_{k=1}^{m} n_k \log \frac{p_k^1}{p_k^0}$, with

$E\{T|H_0\} < 0$ and $E\{T|H_1\} \geq 0$, and $Pr\{H_0\} = Pr\{H_1\} = \frac{1}{2}$

the probability of error of the test (2.6) is upper bounded by

$$P_e \leq \tfrac{1}{2} \inf_{\omega}\left\{\left[\sum_{k=1}^{m}\left(\frac{p_k^1}{p_k^0}\right)^{\omega} p_k^0\right]^n\right\} + \tfrac{1}{2} \inf_{\omega}\left\{\left[\sum_{k=1}^{m}\left(\frac{p_k^1}{p_k^0}\right)^{\omega} p_k^1\right]^n\right\} \tag{2.26}$$

where n is the sample size and ω is real valued.

Proof: If $E\{T|H_0\} < 0$ and $E\{T|H_1\} \geq 0$, then by applying the Chernoff Bound to P_e,

$$P_e \leq \tfrac{1}{2} \inf_{\omega}[G_T(\omega|H_0)] + \tfrac{1}{2} \inf_{\omega}[G_T(\omega|H_1)] \quad , \tag{2.27}$$

where

$$P_e = \sum_{\underline{n} \in N_1^1} \tfrac{1}{2} Pr\{\underline{n}|H_0\} + \sum_{\underline{n} \in N_2} \tfrac{1}{2} Pr\{\underline{n}|H_1\}$$

$N_1^1 = \{\underline{n} \text{ s.t. } T \geq 0\}$ and $N_2 = \{\underline{n} \text{ s.t. } T < 0\}$, and

$G_T(\omega|H_j)$ is the moment generating function of the random variable T given hypothesis H_j; $j = 0,1$ which is given as

$$G_T(\omega|H_j) = \left[\sum_{k=1}^{m}\left(\frac{p_k^1}{p_k^0}\right)^{\omega} p_k^j\right]^n \qquad j = 0,1 \tag{2.28}$$

Substituting (2.28) into (2.27) gives (2.26).

Q.E.D.

By the property on symmetry of $\underline{t}$ and the fact that $q_{-k}^{0} = -q_{k}^{0}$, we can consider only the positive half of the entire partition and rewrite (2.26) as

$$P_e \le 1/2 \inf_{\omega} \left\{ \left[\sum_{k=1}^{m^*} \left(\left(\frac{p_k^1}{p_k^0} \right)^{\omega} p_k^0 + \left(\frac{p_k^1}{p_k^0} \right)^{-\omega} p_k^1 \right) \right]^n \right\}$$

$$+ \inf_{\omega} \left\{ \left[\sum_{k=1}^{m^*} \left(\left(\frac{p_k^1}{p_k^0} \right)^{\omega} p_k^1 + \left(\frac{p_k^1}{p_k^0} \right)^{-\omega} p_k^0 \right) \right]^n \right\} \qquad (2.29)$$

where $m^* = m/2$. The minimizing values of ω can be found for the first and second terms in (2.29) by differentiating them with respect to ω and setting them to be zero separately, thus for $k = 1,2,3,\ldots,m^*$

$$p_k^0 \left(\frac{p_k^1}{p_k^0} \right)^{\omega} \log\left(\frac{p_k^1}{p_k^0} \right) - p_k^1 \left(\frac{p_k^1}{p_k^0} \right)^{-\omega} \log\left(\frac{p_k^1}{p_k^0} \right) = 0 \qquad (2.30)$$

$$p_k^1 \left(\frac{p_k^1}{p_k^0} \right)^{\omega} \log\left(\frac{p_k^1}{p_k^0} \right) - p_k^0 \left(\frac{p_k^1}{p_k^0} \right)^{-\omega} \log\left(\frac{p_k^1}{p_k^0} \right) = 0 \qquad (2.31)$$

(2.30) and (2.31) give $\omega = \frac{1}{2}$ and $\omega = -\frac{1}{2}$ as the minimizing values for the first and second terms in (2.29), respectively; thus

$$P_e \le \tfrac{1}{2} \left[\sum_{k=1}^{m^*} \left\{ \left(\frac{p_k^1}{p_k^0} \right)^{\frac{1}{2}} p_k^0 + \left(\frac{p_k^1}{p_k^0} \right)^{-\frac{1}{2}} p_k^1 \right\} \right]^n$$

$$+ \tfrac{1}{2} \left[\sum_{k=1}^{m^*} \left\{ \left(\frac{p_k^1}{p_k^0} \right)^{-\frac{1}{2}} p_k^1 + \left(\frac{p_k^1}{p_k^0} \right)^{\frac{1}{2}} p_k^0 \right\} \right]^n$$

$$\le \left[\sum_{k=1}^{m^*} 2 (p_k^1 p_k^0)^{\frac{1}{2}} \right]^n \qquad (2.32)$$

The quantity on the right is Hellinger's integral for the partitioned data sequence. The design of quantizers in terms of this latter quantity has been considered by Poor and Thomas in [6].

To study the performance of a fixed sample size detector for weak signal or equivalently for large sample size n, it is usual to assume that the signal strength s is of the order of $1/\sqrt{n}$ since n is a parameter under control. So let $s = K/\sqrt{n}$, where K is any positive constant, so that both $s \to 0$ and $n \to \infty$ at the same time. However, with $s = K/\sqrt{n}$, the bound in (2.32) approaches 1^{∞}, an indeterminate form as $n \to \infty$. Applying the L'Hôpital Rule twice to the bound in (2.32) we can obtain an "asymptotic" upper bound to P_e as $s \to 0$ and $n \to \infty$, which is

$$P_e\Big|_{\substack{s \to 0 \\ n \to \infty}} \leq \exp\left\{ - K^2 \sum_{k=1}^{m^*} \left[\left(f_N'(t_{k-1}) - f_N'(t_k) \right) + \frac{\left(f_N(t_k) - f_N(t_{k-1}) \right)^2}{F_N(t_k) - F_N(t_{k-1})} \right] \right\} \qquad (2.33)$$

A detailed derivation of (2.33) is in the Appendix.

We now try to obtain a set of breakpoints which will give the smallest possible upper bound to P_e in the case of weak signal and large sample size by taking the derivative of the bound in (2.33) with respect to t_k. It turns out that t_k has to satisfy the equation

$$\frac{d}{dt_k} \left[\left(f_N'(t_{k-1}) - f_N'(t_k) \right) + \frac{\left(f_N(t_k) - f_N(t_{k-1}) \right)^2}{F_N(t_k) - F_N(t_{k-1})} + \left(f_N'(t_k) - f_N'(t_{k+1}) \right) + \frac{\left(f_N(t_{k+1}) - f_N(t_k) \right)^2}{F_N(t_{k+1}) - F_N(t_k)} \right] = 0$$

This becomes

$$-f_N''(t_k) + \frac{2\left(f_N(t_k)-f_N(t_{k-1})\right)f_N'(t_k)}{F_N(t_k)-F_N(t_{k-1})} - \frac{\left(f_N(t_k)-f_N(t_{k-1})\right)^2 f_N(t_k)}{\left(F_N(t_k)-F_N(t_{k-1})\right)^2}$$

$$+f_N''(t_k) + \frac{2\left(f_N(t_{k+1})-f_N(t_k)\right)}{F_N(t_{k+1})-F_N(t_k)}\left(-f_N'(t_k)\right) - \frac{\left(f_N(t_{k+1})-f_N(t_k)\right)^2}{\left(F_N(t_{k+1})-F_N(t_k)\right)^2}\left(f_N(t_k)\right) = 0 \tag{2.34}$$

From (2.34) and with

$$q_k = \frac{f_N(t_{k-1})-f_N(t_k)}{F_N(t_k)-F_N(t_{k-1})} \qquad k = 1,2,3,\ldots,m-1 \tag{2.35}$$

the breakpoints $\underline{t}$ can be determined by the following set of equations

$$\frac{q_{k+1} + q_k}{2} = -\frac{f_N'(t_k)}{f_N(t_k)} \qquad k = 1,2,3,\ldots,m-1 \tag{2.36}$$

Surprisingly, this is the same set of equations, (2.21) and (2.22), necessary for $\underline{t}$ to be locally optimal. Hence we can conclude that, with respect to the breakpoints $\underline{t}$, the upper bound to the probability of error in detection of an optimal quantized test given in the above theorem, minimizes simultaneously, as $s \to 0$ and $n \to \infty$, with the inverse of the test's power slope at $s \to 0$ which is, as noted earlier, an appropriate criterion for detection of small signal instead of the test's power. This result is a special case of that obtained by Poor and Thomas in [6].

3. THE PROBABILITY OF ERROR OF THE 4-LEVEL QUANTIZER-DETECTOR

3.1 Structure of the 4-Level Quantizer Detector

From earlier discussion, we conclude that an optimal m-level quantizer detector for arbitrary signal strength s is given by

$$\text{decide} \begin{cases} H_0 & \text{if } \sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) \lessgtr 0 \\ H_1 & \end{cases} \tag{3.1}$$

where $p_k^i = [F_N(t_k - (-1)^i s) - F_N(t_{k-1} - (-1)^i s)]$ for $i = 0,1$, and t_k, $k = 1,2,3,\ldots,m-1$ have to satisfy the set of equations

$$\frac{\partial P_e}{\partial t_k} = \tfrac{1}{2} \sum_{\underline{n} \in N_1^1} \Pr\{\underline{n}/H_0\} \left[\frac{n_k}{p_k^0} - \frac{n_{k-1}}{p_{k-1}^0}\right] f_N(t_k+s)$$

$$+ \tfrac{1}{2} \sum_{\underline{n} \in N_2^1} \Pr\{\underline{n}/H_1\} \left[\frac{n_k}{p_k^1} - \frac{n_{k-1}}{p_{k-1}^1}\right] f_N(t_k-s) = 0 \tag{3.2}$$

where $N_1^1 = \{\underline{n} = (n_1,n_2,n_3,\ldots,n_m)$ s.t. $\sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) \geq 0\}$

and $N_2 = \{\underline{n} = n_1,n_2,n_3,\ldots,n_m)$ s.t. $\sum_{k=1}^{m} n_k \log\left(\frac{p_k^1}{p_k^0}\right) < 0\}$

But (3.2), as mentioned before, generally does not have a closed form solution to t_k and it can only be solved by some root-searching technique. However, from the property on symmetry of $\underline{t}$, we can easily see that, for m even, $t_{\frac{m}{2}} = 0$ and $t_j = t_{m-j}$. So there are only $(\frac{m}{2} - 1)$ different t_k's left to determine as the others can be set according to the property.

Only when m = 4 are the root searching techniques one-dimensional, and we will confine ourselves only to this case for the rest of this thesis, although it is conceptually as simple to use some higher-dimensional searching techniques to locate all the $(\frac{m}{2} - 1)$ t_k's in (3.2) for $m \geq 6$ (Note: m is taken to be even).

Once $\underline{t}$ is set from some searching methods, its corresponding optimal quantization levels q_k and, consequently, the optimal quantized test statistic, are determined. With symmetric noise density function $f_N(\cdot)$, the odd symmetric property of q_k implies that (3.1) can be rewritten as

$$\text{decide} \begin{cases} H_0 & \text{if } \sum_{k=1}^{m/2} (n_{m-k+1} - n_k) q_{m-k+1} < 0 \\ H_1 & \text{if } \sum_{k=1}^{m/2} (n_{m-k+1} - n_k) q_{m-k+1} \geq 0 \end{cases} \tag{3.3}$$

with m = 4, (3.3) becomes

$$\text{decide} \begin{cases} H_0 & \text{if } \sum_{k=1}^{2} (n_{5-k} - n_k) q_{5-k} < 0 \\ H_1 & \text{if } \sum_{k=1}^{2} (n_{5-k} - n_k) q_{5-k} \geq 0 \end{cases} \tag{3.4}$$

We can see that normalizing the test statistics in (3.4) with q_3 has no effect on the quantizer detector performance, hence

$$\text{decide} \begin{cases} H_0 & \text{if } (n_4 - n_1) q_r + (n_3 - n_2) < 0 \\ H_1 & \text{if } (n_4 - n_1) q_r + (n_3 - n_2) \geq 0 \end{cases} \tag{3.5}$$

where $q_r = q_4/q_3$. Also with m = 4, $t_1 = -t_3$ and $t_2 = 0$, (3.2) can be written as

$$\frac{dP_e}{dt_3} = \tfrac{1}{2} \sum_{\underline{n} \in N_1^1} \Pr\{\underline{n}|H_0\} \left[\begin{matrix} n_3 \\ 0 \\ p_3 \end{matrix} - \begin{matrix} n_2 \\ 0 \\ p_2 \end{matrix} \right] f_N(t_3+s)$$

$$+ \tfrac{1}{2} \sum_{\underline{n} \in N_2} \Pr\{\underline{n}|H_1\} \left[\begin{matrix} n_3 \\ 1 \\ p_3 \end{matrix} - \begin{matrix} n_2 \\ 1 \\ p_2 \end{matrix} \right] f_N(t_3-s) = 0 \qquad (3.6)$$

with $N_1^1 = \{\underline{n} = (n_1,n_2,n_3,n_4)$ s.t. $(n_4-n_1)q_r + (n_3-n_2) \geq 0\}$

and $N_2 = \{\underline{n} = (n_1,n_2,n_3,n_4)$ s.t. $(n_4-n_1)q_r + (n_3-n_2) < 0\}$

$p_1^i = F_N(-t_3 \pm s)$, $p_2^i = F_N(\pm s) - F_N(-t_3 \pm s)$, $p_3^i = F_N(t_3 \pm s) - F_N(\pm s)$,

$p_4^i = 1 - F_N(t_3 \pm s)$ with $\pm$ for $i = \begin{matrix} 0 \\ 1 \end{matrix}$.

(3.6) is a function of t_3 only and many one-dimensional root-seeking methods can give a solution to t_3.

For the reason given in the following chapter, we prefer and will use instead of (3.6), the probability of error itself in solving t_3, i.e.,

$$P_e(t_3) = \tfrac{1}{2} \sum_{\underline{n} \in N_1^1} \Pr\{\underline{n}|H_0\} + \tfrac{1}{2} \sum_{\underline{n} \in N_2} \Pr\{\underline{n}|H_1\} . \qquad (3.7)$$

Again using any one-dimensional "peak" seeking methods on (3.7), t_3 can be located as well.

3.2 Formation of the $P_e^{so}(t_3)$ and $P_e^{o}(t_3)$ Curves

$P_e(t_3)$ can be plotted versus t_3 in two different ways. P_e^{so} is the curve plotted with q_r, which determines N_1^1 and N_2 in (3.7), held fixed; this implies the same N_1^1 and N_2 are used in calculating $P_e(t_3)$ for all t_3. Minimizing P_e^{so} with respect to t_3 corresponds to the minimization of the

probability of error by moving t_3 around until the minimum probability of error is attained with quantization level ratio q_r fixed all through the process. Obviously, the t_3 so obtained is optimal only for the class of quantizer detectors using that particular level ratio.

Several $P_e^{so}(t_3)$ curves are shown in Fig. 2 for different values of q_r (with Gaussian and Cauchy noise). It can be seen that for any t_3 there is a corresponding q_r which gives the minimum value to the probability of error at that particular t_3. Since $q_k = \log(p_k^1/p_k^0)$ gives an optimal quantizer with breakpoints $\underline{t}$, therefore the optimal q_r corresponds to each t_3 is

$$q_r = \log \left\{ \frac{1-F_N(t_3-s)}{1-F_N(t_3+s)} \right\} \Big/ \log \left\{ \frac{F_N(t_3-s) - F_N(-s)}{F_N(t_3+s) - F_N(s)} \right\} \tag{3.8}$$

$P_e^o(t_3)$ is then the curve which picks off the minimum of all the probabilities of error over all possible q_r at each t_3; it is shown in Fig. 3 and in Fig. 2 along with the $P_e^{so}(t_3)$ curves. Obviously $P_e^o(t_3)$ is the greatest lower bound to all possible $P_e^{so}(t_3)$ at each t_3 and consequently the curve $P_e^o(t_3)$ always stays on or below all $P_e^{so}(t_3)$ curves. On the other hand, $P_e^o(t_3)$ can also be obtained analytically, at each t_3, by evaluating (3.7) with q_r from (3.8) for every t_3; hence the sets N_1^1 and N_2 are different for every different t_3. The t_3 so obtained by minimizing $P_e^o(t_3)$ will yield a truly optimal quantizer detector. However, as the number of possible elements $\underline{n}$ in N_1^1 and N_2 gets large for large sample size n, the necessary search for $\underline{n} = (n_1,n_2,n_3,n_4)$ in N_1^1 and N_2 for every t_3 will become time-consuming.

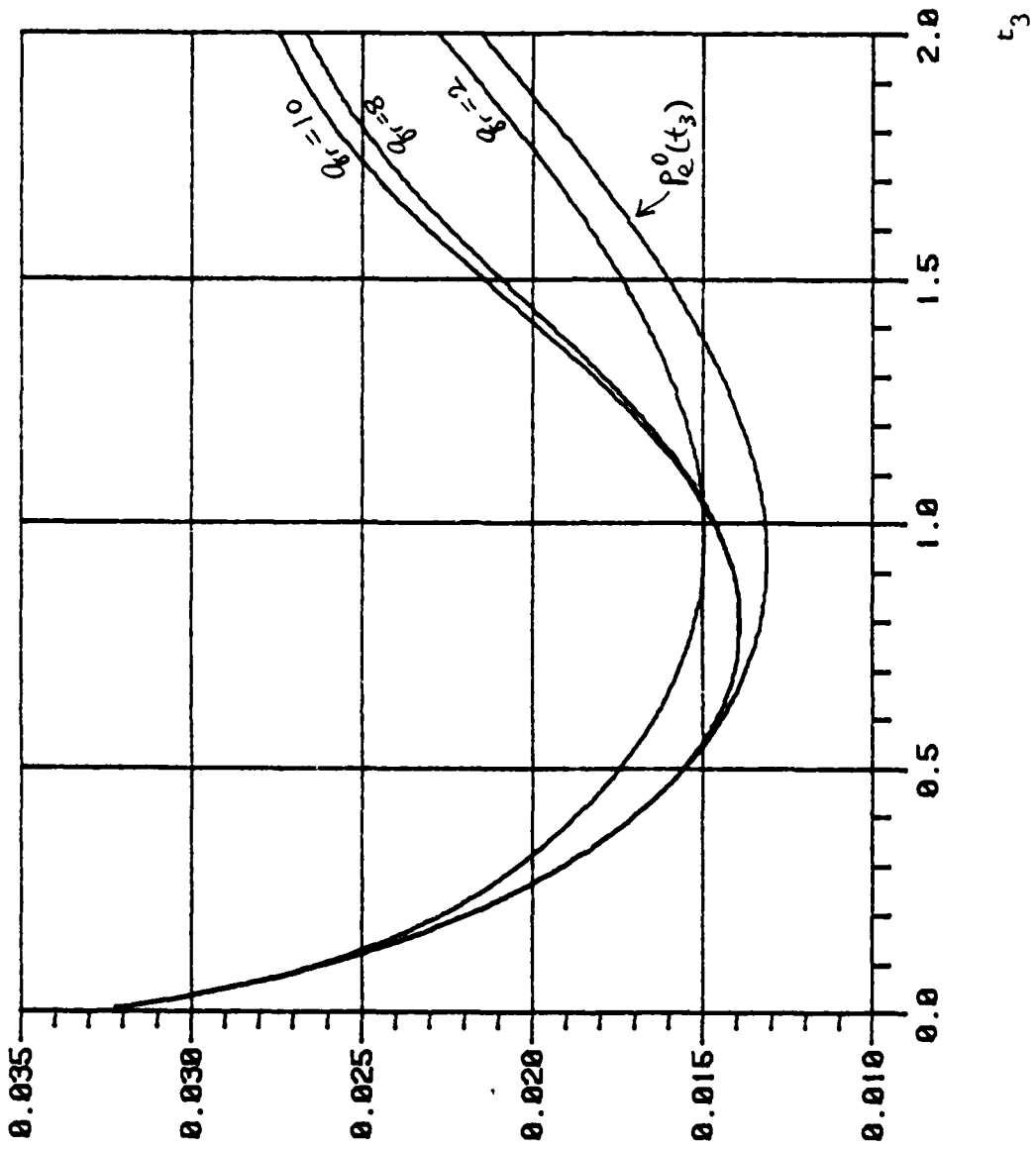


Figure 2a. $P_e^{so}(t_3)$ curves for Gaussian noise with SNR = 0.75 and n = 10.

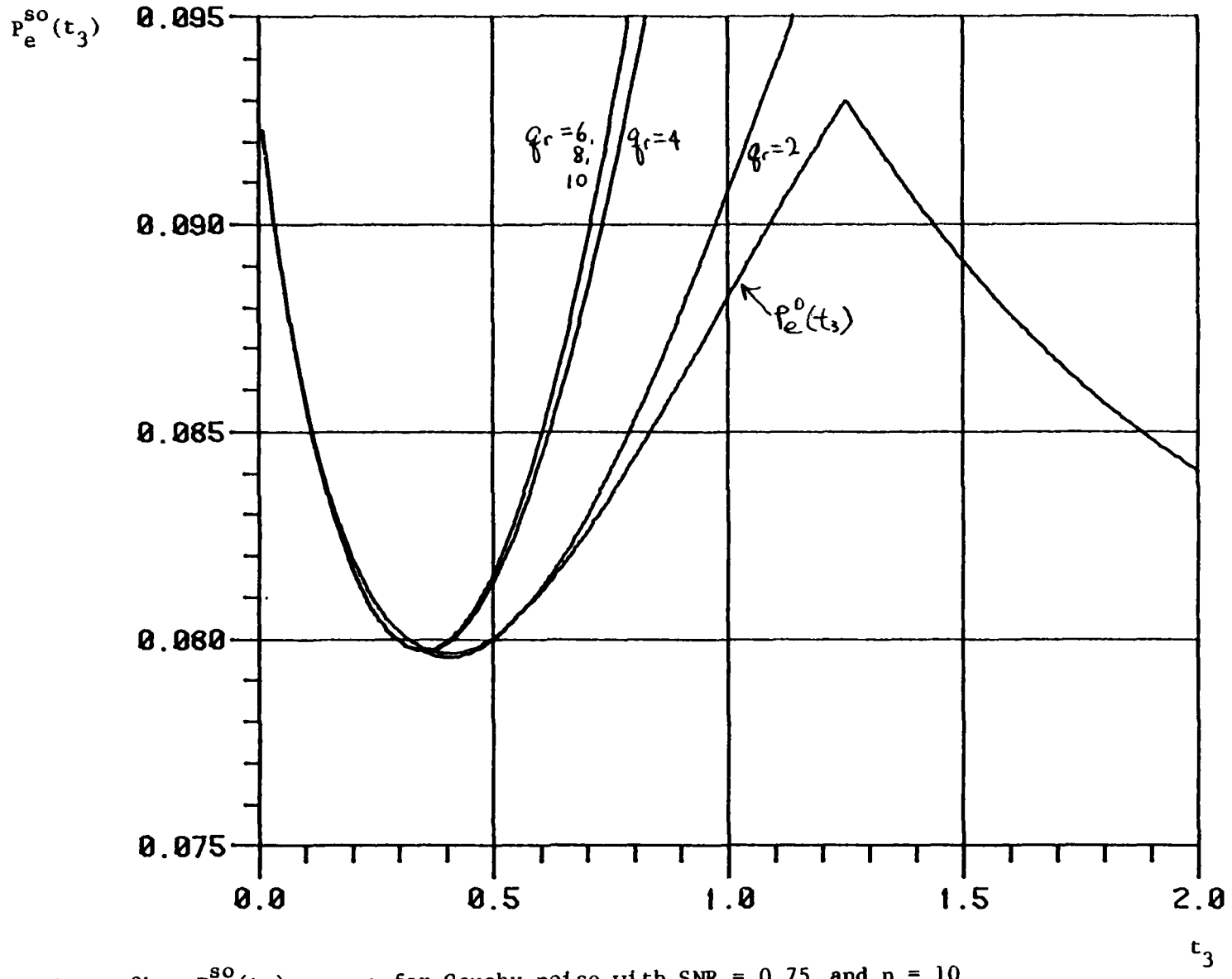


Figure 2b. $P_e^{so}(t_3)$ curves for Cauchy noise with SNR = 0.75 and n = 10.

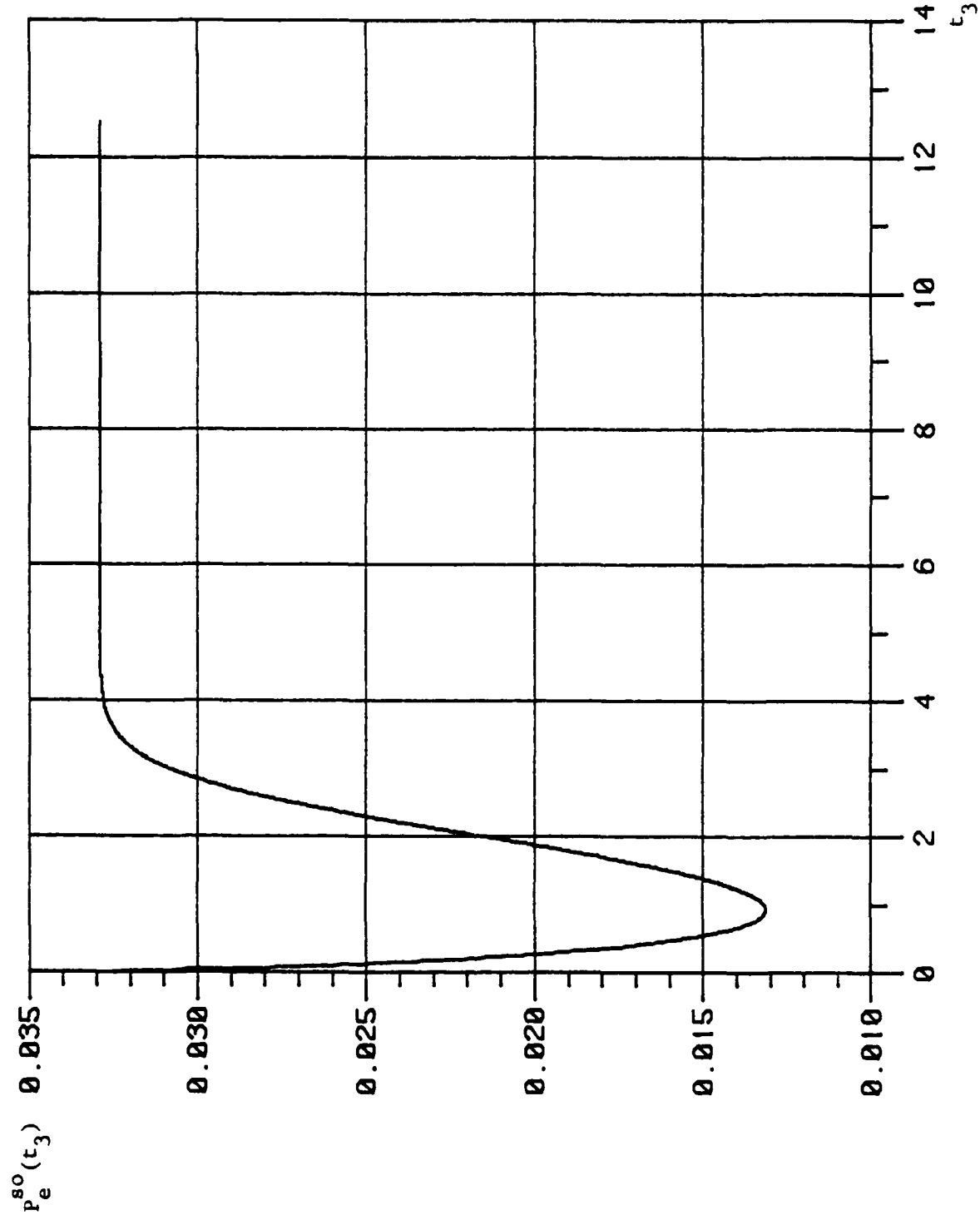


Figure 3a. $P_e^{O}(t_3)$ curve for Gaussian noise with SNR = 0.75 and n = 10.

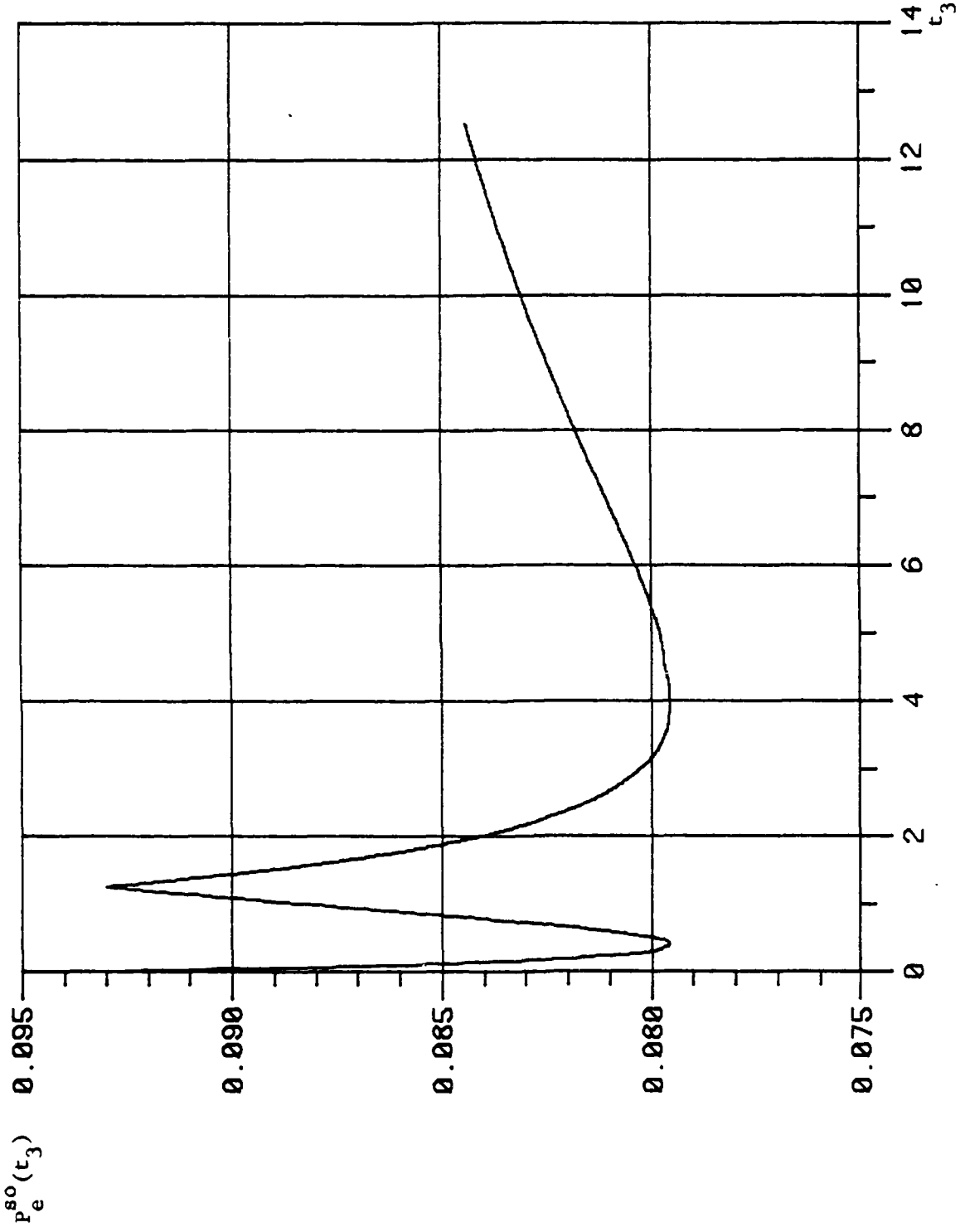


Figure 3b. $P_e^{o}(t_3)$ curve for Cauchy noise with SNR = 0.75 and n = 10.

3.3 Some Characteristics of the $P_e^{so}(t_3$ and $P_e^{o}(t_3)$ Curves

After the formulation of $P_e^{so}(t_3)$ and $P_e^{o}(t_3)$ have been considered, we now turn to discuss their characteristics from the observations of their curves in Fig. 2. First we notice that for some t_3 there is a range of q_r values that give rise to the same probability of error; that is, for some t_3 in which the probability of error is insensitive to certain range of q_r values. This can be better illustrated by plotting the probability of error versus the level ratio q_r given t_3. As shown in Fig. 4-Fig. 7, each of these curves is actually a series of steps and the width of each step corresponds to the range of q_r which gives equivalent probability of error at that t_3. From these figures (Fig. 4-Fig. 7) it is clear that for every t_3, the probability of error depends only on the ranges of q_r (i.e., it is a function of the ranges of q_r only) and not on the actual q_r values. This is because for a given t_3, $P_e^{so}(t_3)$ depends on q_r through the sets N_1^1 and N_2 and with the sample size n finite, there may be a range of values of q_r which gives rise to the same sets of N_1^1 and N_2. Although the value $[(n_4-n_1)q_r + (n_3-n_2)]$ itself changes for every different q_r, the two sets of $\underline{n} = (n_1,n_2,n_3,n_4)$ that give N_1^1 and N_2 such that $[(n_4-n_1)qr + (n_3-n_2)] \gtrless 0$, respectively may be invariant under different q_r. We expect the sets N_1^1 and N_2 to be more distinguishable for different q_r and the staircase-like curves in Fig. 4- Fig. 7 to smooth out as n gets large.

Next we notice that for large enough q_r and fixed n, the probability of error is independent of q_r for every t_3. This can be seen from Fig. 2 or better from Fig. 4-Fig. 7 where the last step extends all the way from $q_r = 10$ given any t_3, this is due to the fact that the sample size is

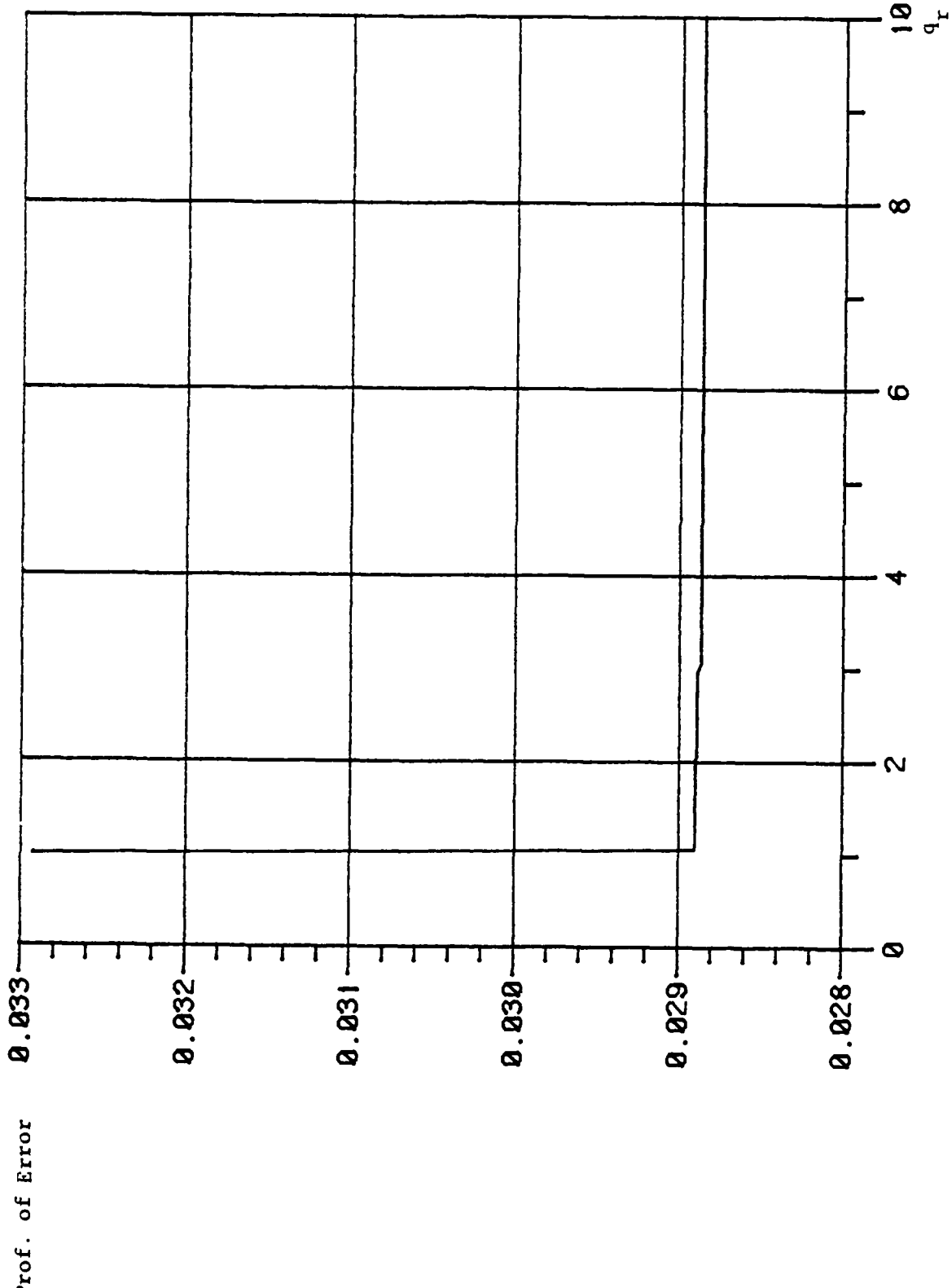


Fig. 4a. Probability of error vs q_r for Gaussian noise SNR = 0.75, t_3 = 0.05, n = 10.

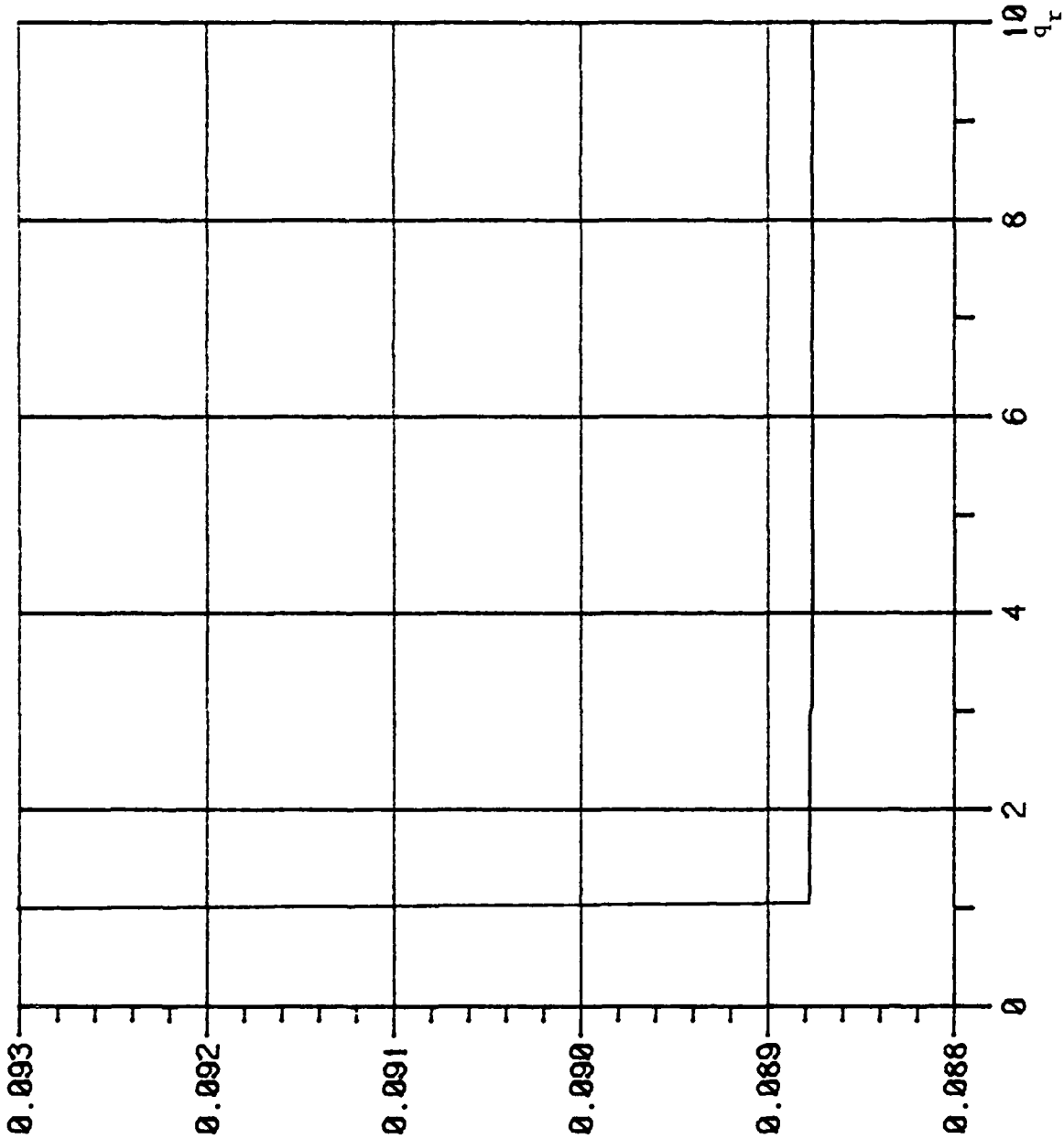


Figure 4b. Probability of error vs q_r for Cauchy noise SNR = 0.75, t_3 = 0.05, n = 10.

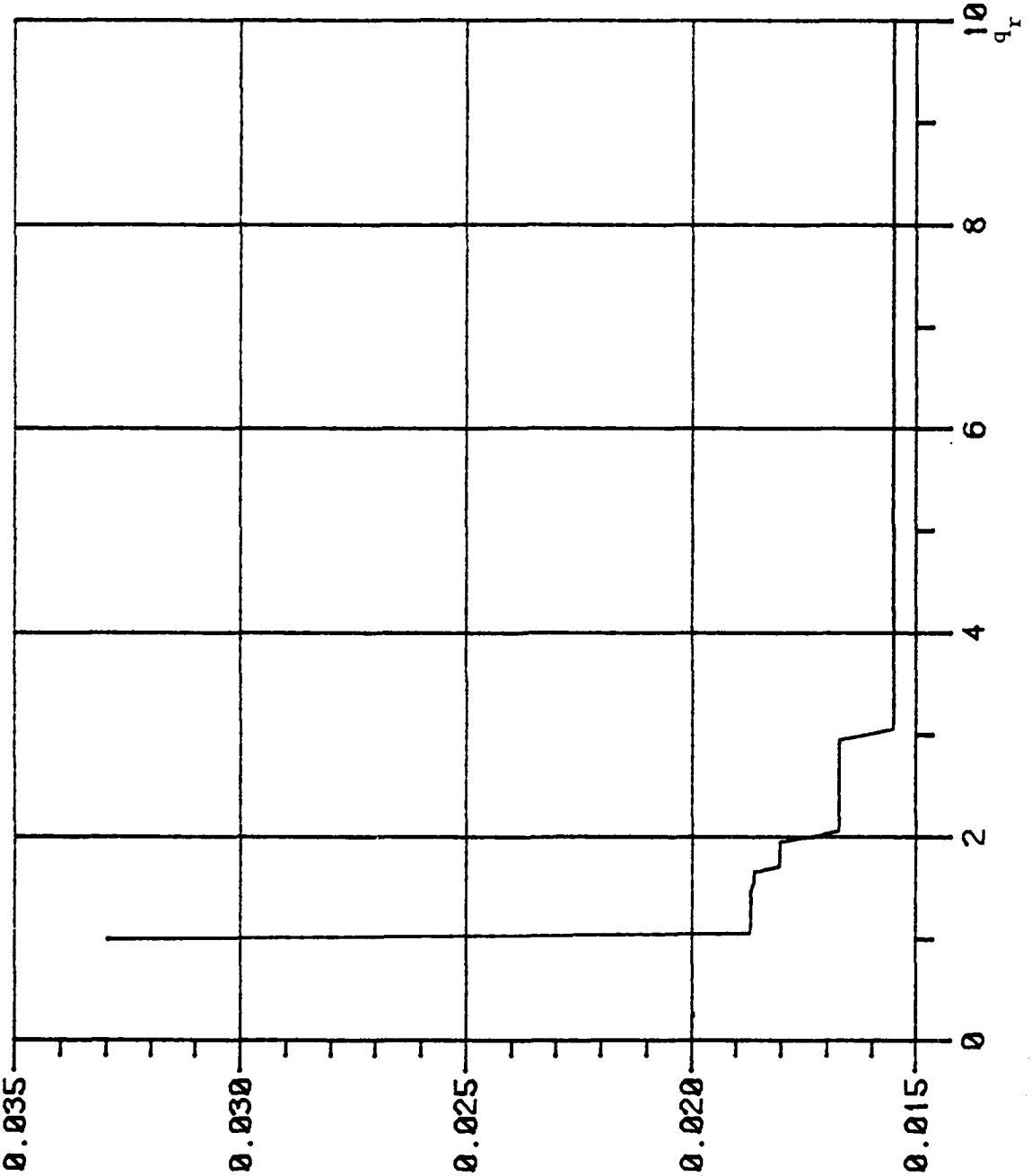


Figure 5a. Probability of error vs q_r for Gaussian noise SNR = 0.75, t_3 = 0.5, n = 10.

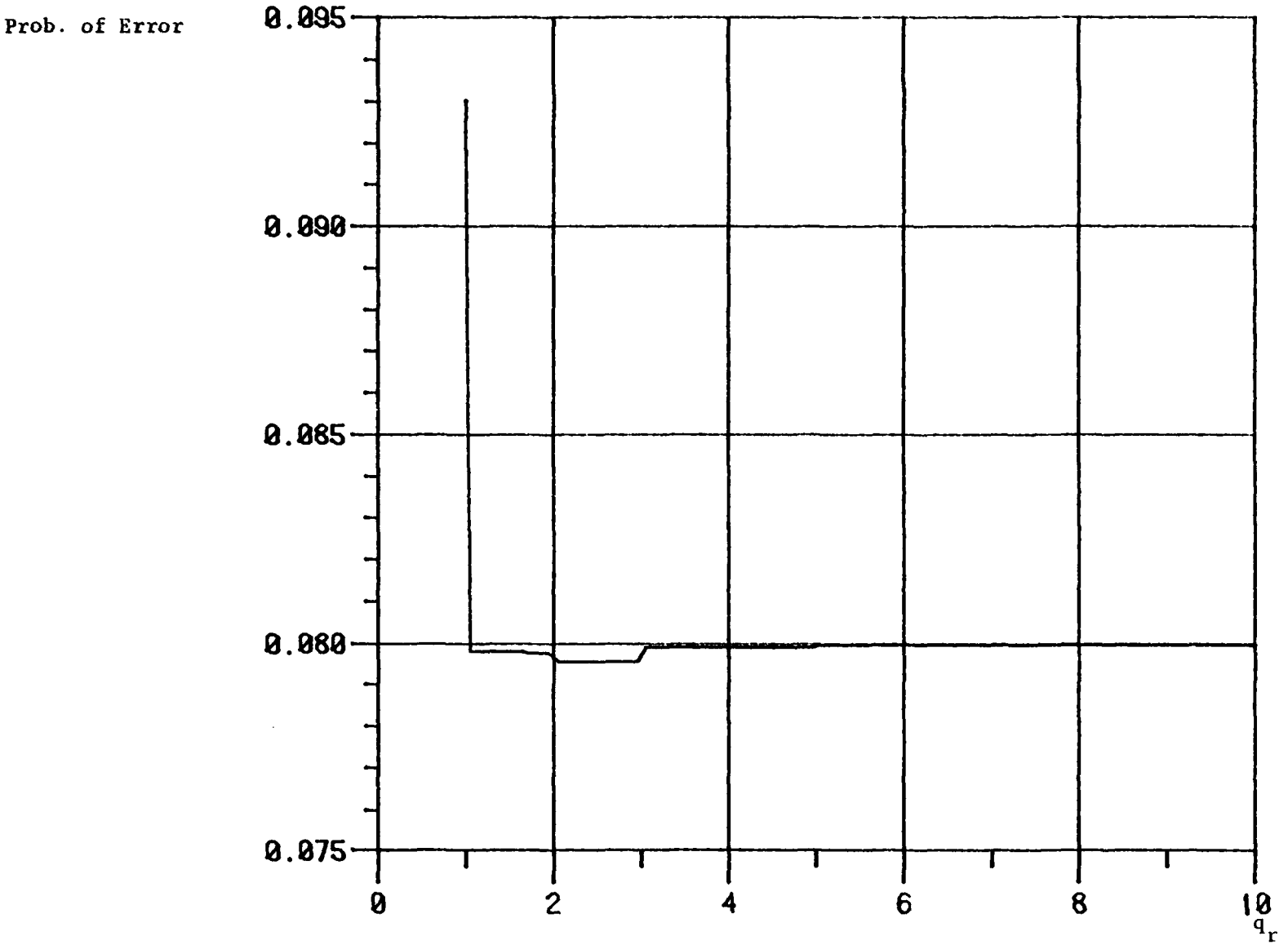


Figure 5b. Probability of error vs q_r for Cauchy noise SNR = 0.75, t_3 = 0.4 , n = 10.

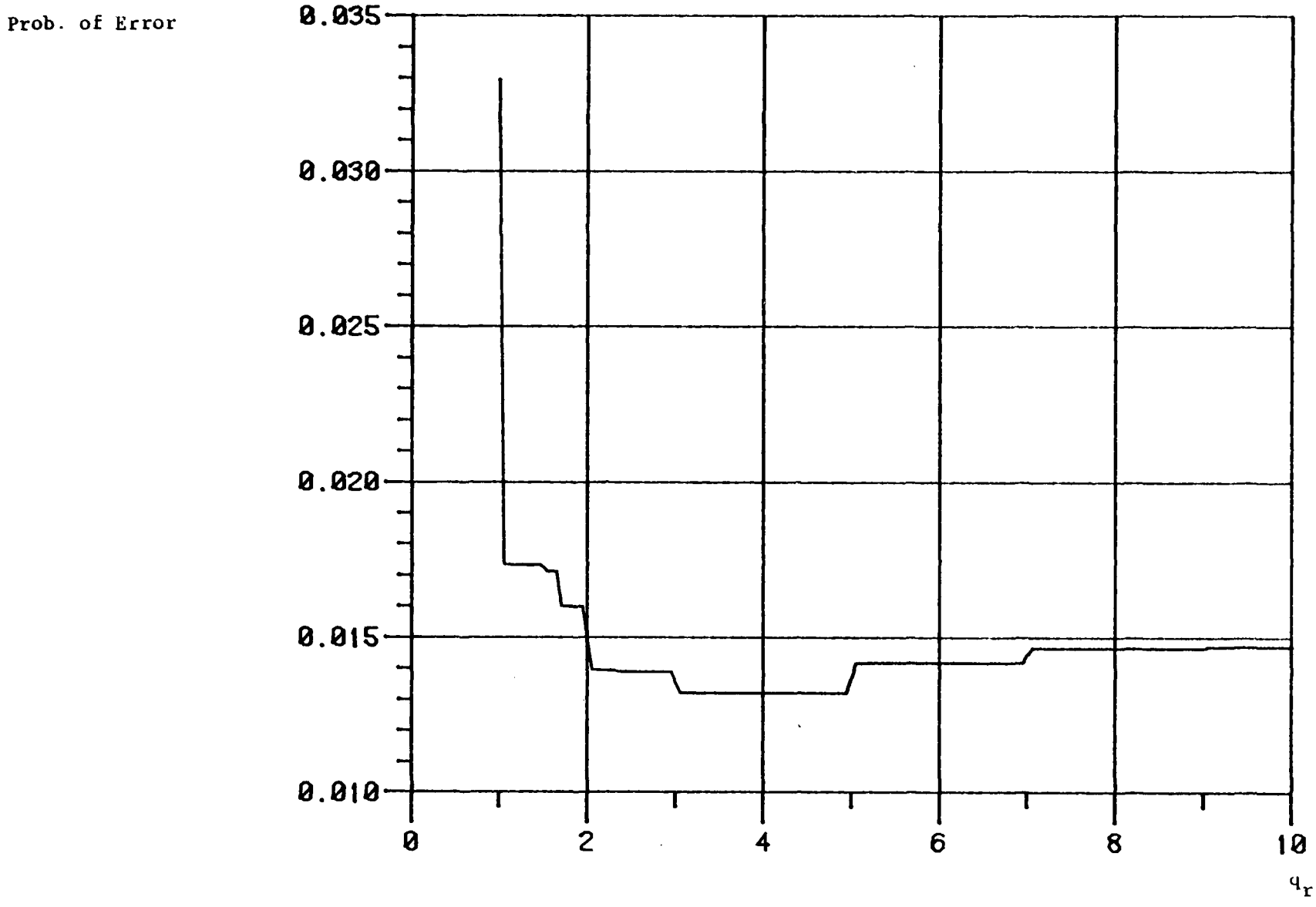


Figure 6a. Probability of error vs q_r for Gaussian noise SNR = 0.75, t_3 = 1.0, n = 10.

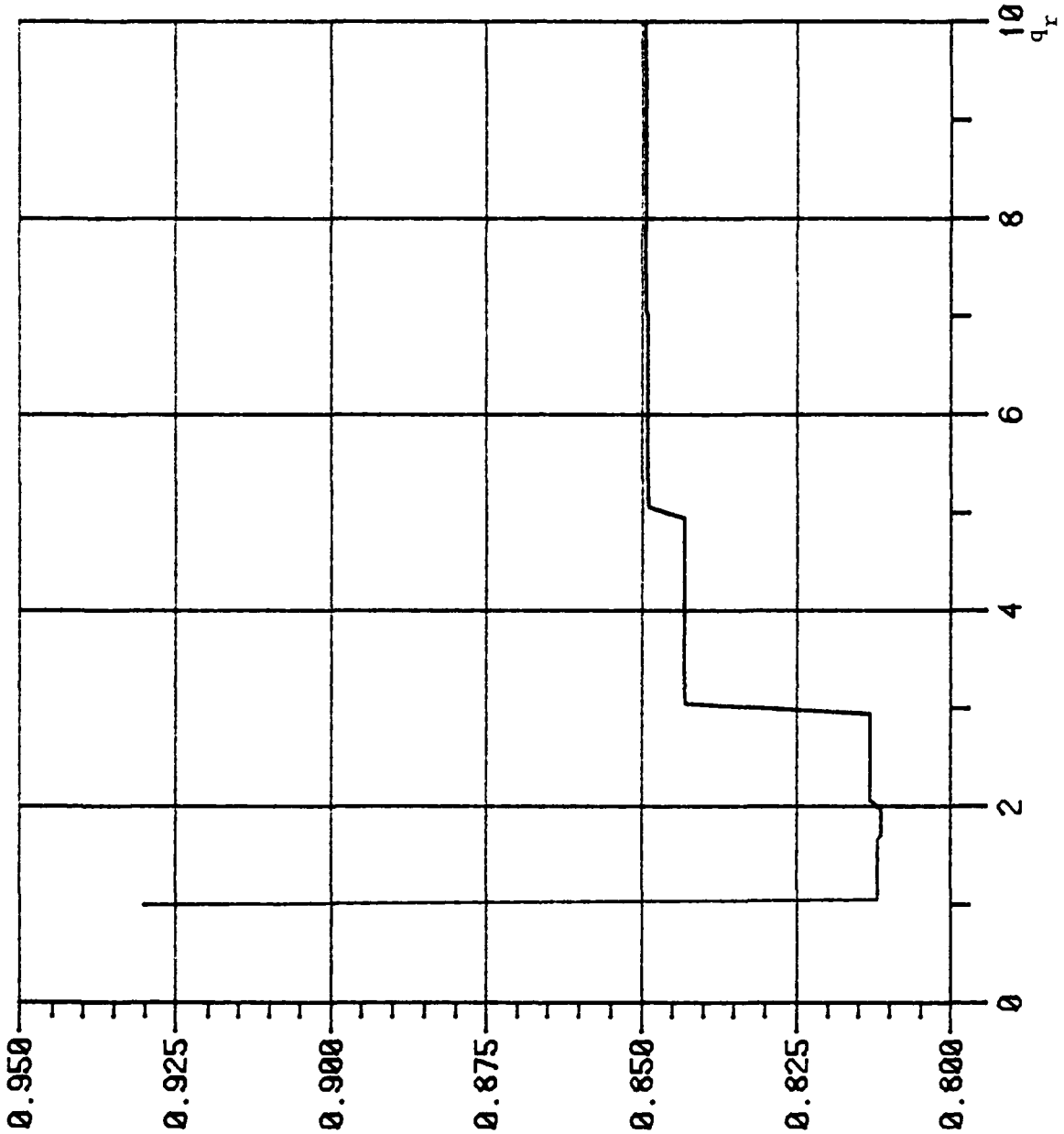


Figure 6b. Probability of error vs q_r for Cauchy noise SNR = 0.75, t_3 = 0.6, n = 10.

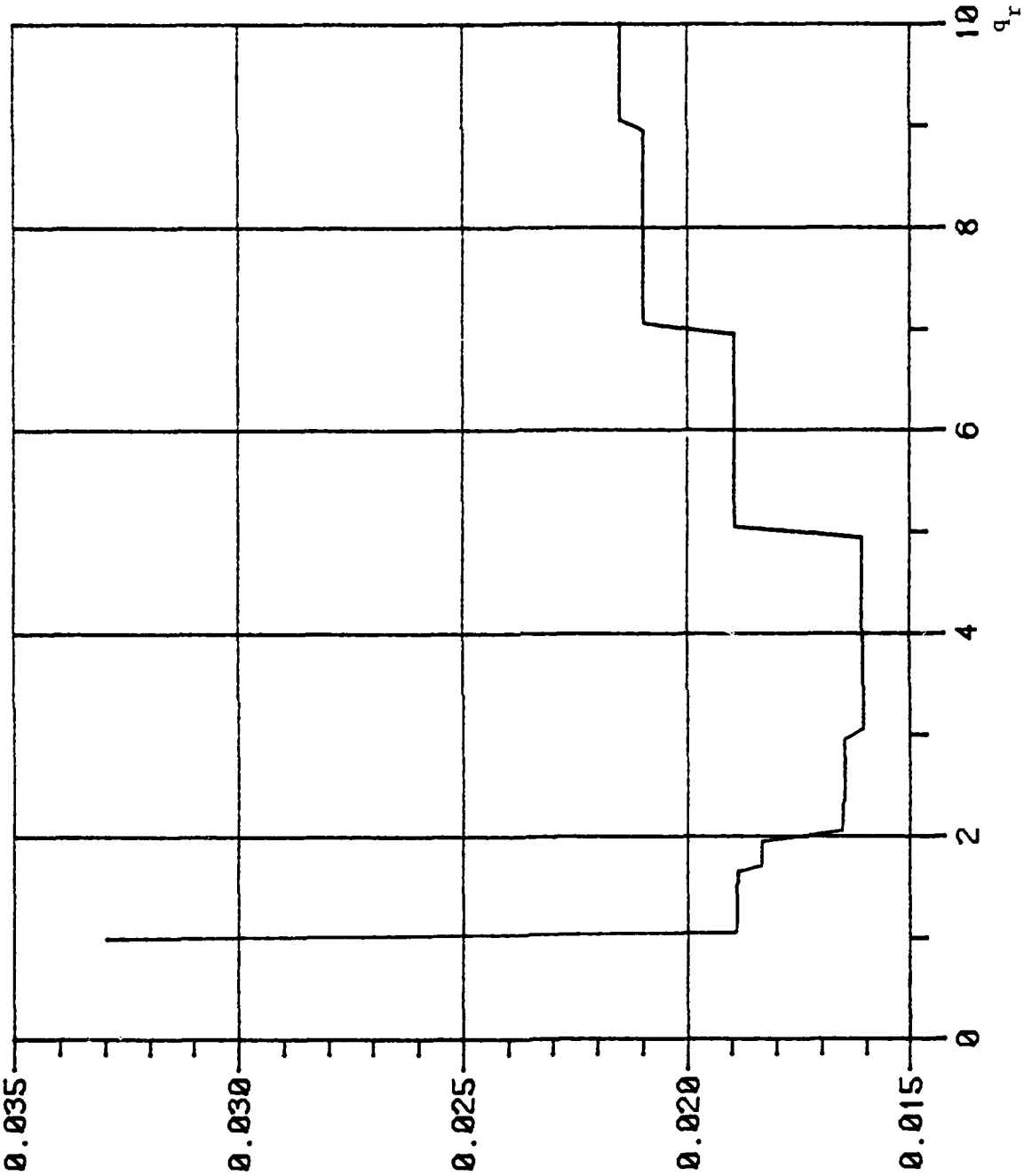


Figure 7a. Probability of error vs q_r for Gaussian noise SNR = 0.75, t_3 = 1.5, n = 10.

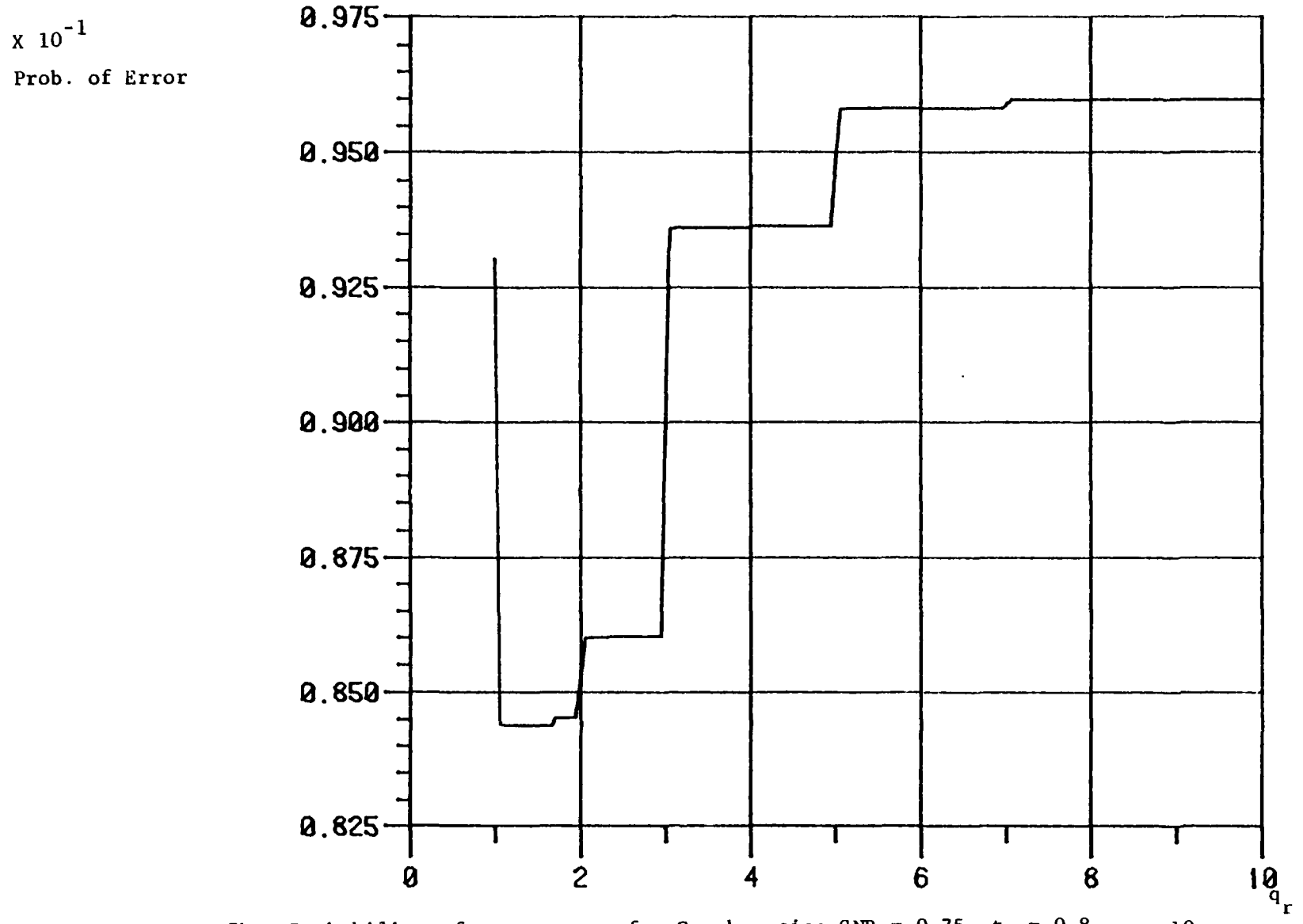


Figure 7b. Probability of error vs q_r for Cauchy noise SNR = 0.75, t_3 = 0.8, n = 10.

finite causing N_1^1 and N_2 to be invariant under large q_r. Again it will occur at larger q_r when n gets larger.

The insensitivity of $P_e^{so}(t_3)$ to q_r becomes significant for small t_3. As shown in the staircase-like curves for small t_3, the probability of error is roughly constant for all q_r values greater than one. Thus the $P_e^{o}(t_3)$ curve almost coincides perfectly with all the $P_e^{so}(t_3)$ curves for small t_3 values (see Fig. 2) and hence the minimum points of the $P_e^{o}(t_3)$ and $P_e^{so}(t_3)$ curves are found located close to each other. These characteristics of the curves have a very important implication on adaptation discussed later.

Since the probability of error given t_3 depends only on the range of q_r for finite n, there must be a range of q_r values that gives the same minimum probability of error. Let us denote this range as an optimal range of q_r values; it is quite obvious that the value $\log\left(\frac{p_4^1}{p_4^0}\right) \Big/ \log\left(\frac{p_3^1}{p_3^0}\right)$ must fall in the optimal range. For example, from Fig. 2, the optimal breakpoint t_3 is shown to be about 0.4 for Cauchy noise when s = 0.75, the $q_r = \log\left(\frac{p_4^1}{p_4^0}\right) \Big/ \log\left(\frac{p_3^1}{p_3^0}\right)$ with $t_3 = 0.4$ is about 2.65 which is within the optimal range as seen in Fig. 4. As mentioned earlier, the staircase-like curves smooth out as n gets large and eventually ($n \to \infty$) the optimal range for q_r will collapse to a single point of value $\log\left(\frac{p_4^1}{p_4^0}\right) \Big/ \log\left(\frac{p_3^1}{p_3^0}\right)$.

Before leaving this section, two main characteristics of $P_e^{so}(t_3)$ are worth pointing out again; i.e., given the signal strength s, the probability of error is insensitive to q_r for any t_3 below a certain value; and the probability of error depends only on t_3 for any q_r above a critical value which was mentioned earlier. This critical q_r is related to the sample size n. If the sample size n is finite, for quantizers with 4 levels, it is easily seen from the definitions of N_1^1 and N_2,

$$N_1^1 = \{(n_1,n_2,n_3,n_4) \text{ s.t. } (n_4-n_1)q_r + (n_3-n_2) \geq 0\}$$

$$N_2 = \{(n_1,n_2,n_3,n_4) \text{ s.t. } (n_4-n_1)q_r + (n_3-n_2) < 0\}$$

that any q_r greater than or equal to n will definitely give the same sets of N_1^1 and N_2; hence $P_e^{so}(t_3)$ is identical for all $q_r \geq n$. Note that Figs. 2-7 are created with sample size n = 10.

3.4 Some Considerations on the Design of the Adaptive Quantizer Detector

From previous discussion, we point out that the minimum points of the $P_e^o(t_3)$ and $P_e^{so}(t_3)$ curves are situated closely to each other; hence if we are willing to suffer a little more probability of error near the minima, we may just as well consider $P_e^{so}(t_3)$ instead of $P_e^o(t_3)$ since $P_e^o(t_3)$ will (as noted above) take much more processing time than $P_e^{so}(t_3)$.

Now it becomes necessary to decide which $P_e^{so}(t_3)$ curve (correspondingly, which q_r) to work on. However, there is a rule of thumb in picking q_r condensed from the previous descriptions on the general characteristics of the curves, which gives a guaranteed performance for the adaptive quantizer detector.

If q_r is chosen such that it is strictly less than n and greater than 1, we are guaranteed that the performance of the adaptive quantizer detector is better than the worst possible performance with that particular noise. This is because, for small t_3, $P_e^{so}(t_3)$ is roughly the same for almost every q_r, and for large t_3, $P_e^{so}(t_3)$ is increasing with q_r but upper bounded by that value of $P_e^{so}(t_3)$ with $q_r \geq n$. So, using any $P_e^{so}(t_3)$ with $1 < q_r < n$ will have performance always better than the lower bound performance.

It appears from the curves that the smaller the q_r used, the better the adaptive quantizer detector's performance will be; however, we note from Fig. 2 that $q_r = 2$ gives the largest minimum probability of error over all q_r for the Gaussian noise case though the performance during the adaptive process is almost the best we can get. Since the noise is unknown to the detector, we really do not have any good guess on the initial t_3 to start our iterative process for adaptation. If we start with "small" initial t_3, then it does not really matter which q_r we use since performance of the adaptive process is insensitive to q_r in the range of small t_3. But if our initial choice on t_3 turns out to be "large", we have a tradeoff between better performance with smaller q_r and faster convergence to the final optimal operating point with larger q_r, which is due to its relatively steeper slope.

One might arrive at the conclusion that, if we can start with arbitrarily small t_3, we can then forget about choosing q_r and still have both fast convergence and an almost uniform performance over all q_r. But we simply cannot start with arbitrarily small t_3 because, as will be

seen in the simulation, small t_3 gives a very bad estimation on the probability of error, especially when n is small; besides, the smallest initial t_3 that can be used is also dictated by the particular iterative scheme being used.

Finally we observe from Fig. 8 that the slope of the $P_e^{so}(t_3)$ curve may be steeper for smaller s. Hence, the absolute amount of errors saved from adapting the quantizer detector to its optimal operating point may be larger for smaller s; however, the percentage of improvement in the probability of detection is less as compared with larger s in adaptation. Thus it depends on the particular design objective whether or not the adaptation process to the optimal quantization parameters is worth doing for large or small signal strength.

So far only Gaussian and Cauchy distributions are considered; but since they represent two extremes, we may consider all these trends to be typical.

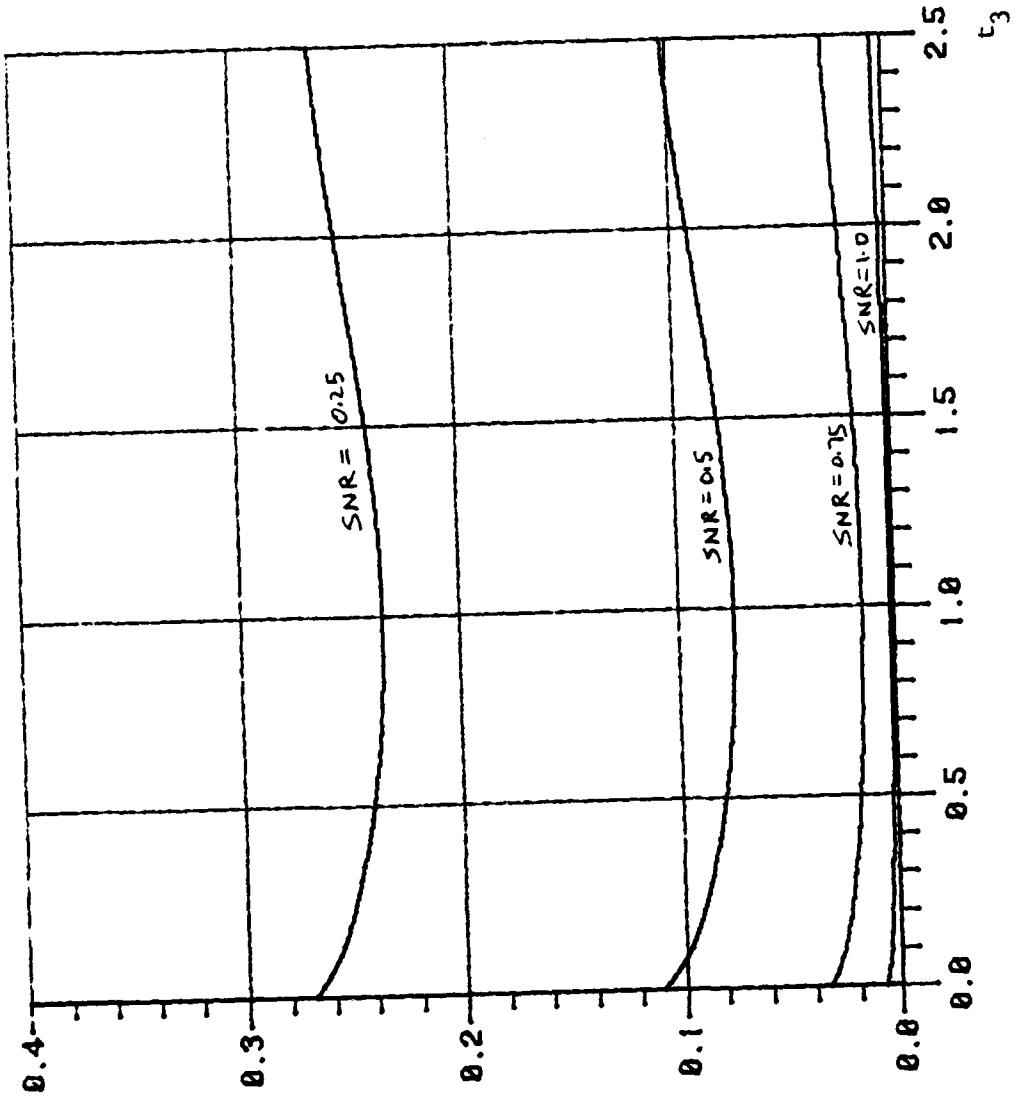


Figure 8a. $P_e^{so}(t_3)$ curves for Gaussian noise with $q_r = 2.0$, $n = 10$.

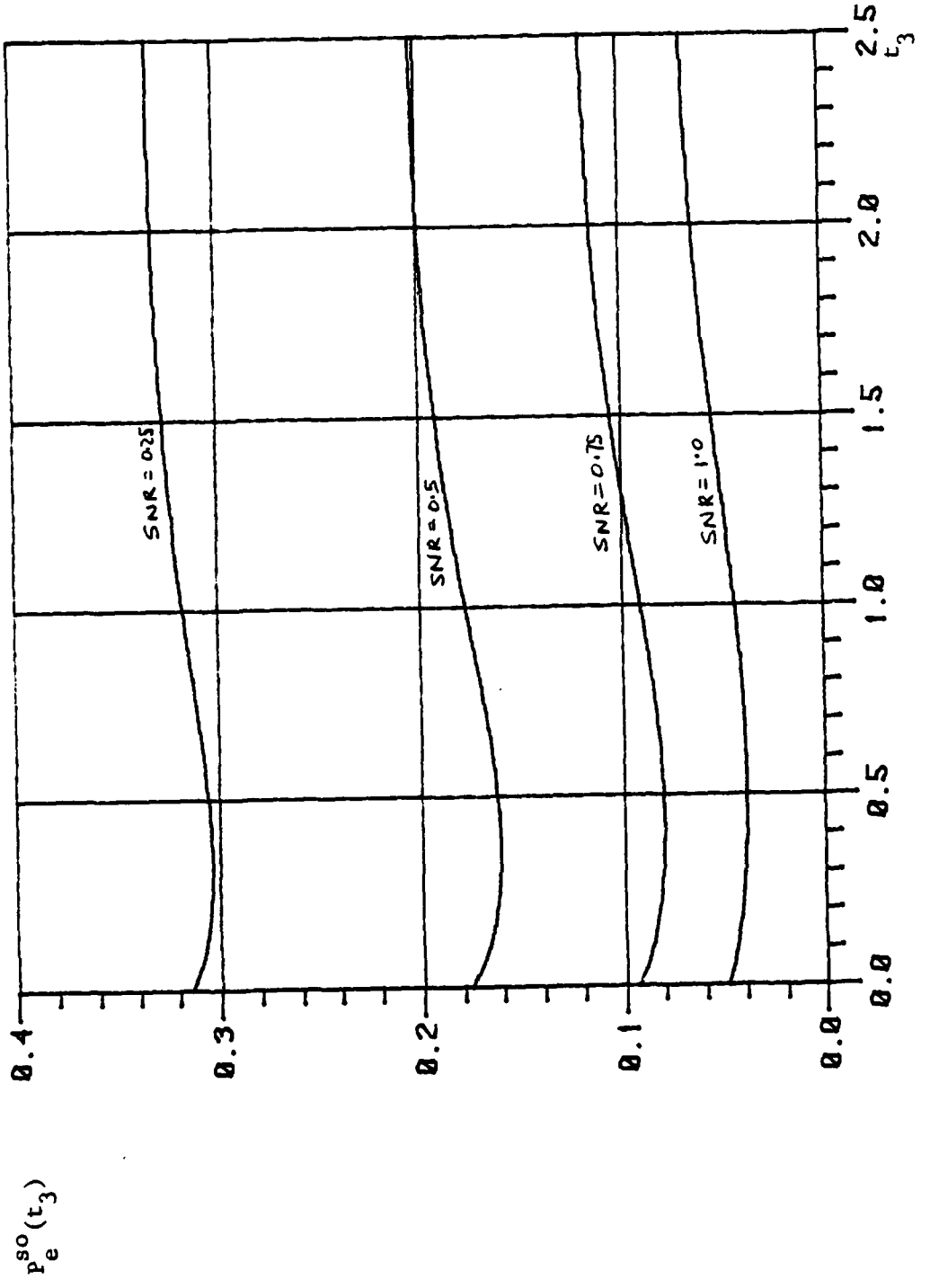


Figure 8b. $P_e^{so}(t_3)$ curves for Cauchy noise with $q_r = 2.0$, $n = 10$.

4. THE ADAPTIVE DETECTION SYSTEM

4.1 Two Methods of Adaptation

There are several possible ways to adapt the quantizer to the unknown noise. For example we may use a training sequence of signals which is known to the detector before transmission; and the method is to adjust the parameters in the quantizer (i.e., breakpoints and levels) until a maximum number of samples from the sequence are correctly "detected." This method requires a certain idling period for training before any actual transmission and detection of real data. This may not be acceptable in some cases. Furthermore if the background noise is time-varying, though it may be changing very slowly, the training process may be necessary quite often.

One of the other ways is to use the method of unsupervised decision directed adaptation, in which the detector runs with real data while the adaptation of the quantizer is taking place. This way the detector can operate on a full time basis and can keep up with any change in the noise up to a certain time lag due to the transient response of the particular adaptation scheme being used in the system. In this method, every decision made on the real data is assumed correct and is used as a training sequence for the optimal quantizer parameter values.

The potential disaster of this method is the possibility of system runaway if enough decisions made were actually incorrect and the modification on the quantizer values based on these incorrect decisions drives the quantizer away from its optimal state. This results in more errors in decisions. This happens most likely in the case when the initial probability of error of the detector is large.

4.2 Structure of the Adaptive Quantizer-Detector System

The structure of the whole detection system in our simulation is shown in Fig. 9. The detection scheme follows the structure in (3.5) where n_i is the number of samples, from an observation size of 10 samples, that fall in the i-th interval which is characterized by the value of the breakpoint t_3 (note: i runs from 1 to 4 for a 4-level quantizer detector). Then with level ratio q_r, the quantity $(n_4-n_1)q_r + (n_3-n_2)$ is compared to a threshold (which is zero in our case) to make a decision on which hypothesis (H_0 or H_1) those 10 samples are from, depending on whether the quantity is below or above the threshold.

As mentioned in a previous chapter, additional complexity goes into the system when $P_e^o(t_3)$ is used instead of $P_e^{so}(t_3)$. $P_e^{so}(t_3)$ is the probability of error as a function of t_3 for a fixed q_r and hence the sets N_1^1 and N_2 are fixed at all times; while $P_e^o(t_3)$ requires new sets of N_1^1 and N_2 which correspond to the new $q_r = \log\left(\frac{P_4^1}{P_4^0}\right) \Big/ \log\left(\frac{P_3^1}{P_3^0}\right)$ with every newly iterated t_3. Unless it is necessary to go to the true optimal point of the detector by using $P_e^o(t_3)$, we will consider the $P_e^o(t_3)$ case only.

In iterative procedures, there are two ways to locate the optimal t_3 which gives minimum P_e^{so}. One is to find the zero of the derivative function of $P_e^{so}(t_3)$ with respect to t_3. The other is to locate the minimum of the function $P_e^{so}(t_3)$. In the 4-level quantizer-detector,

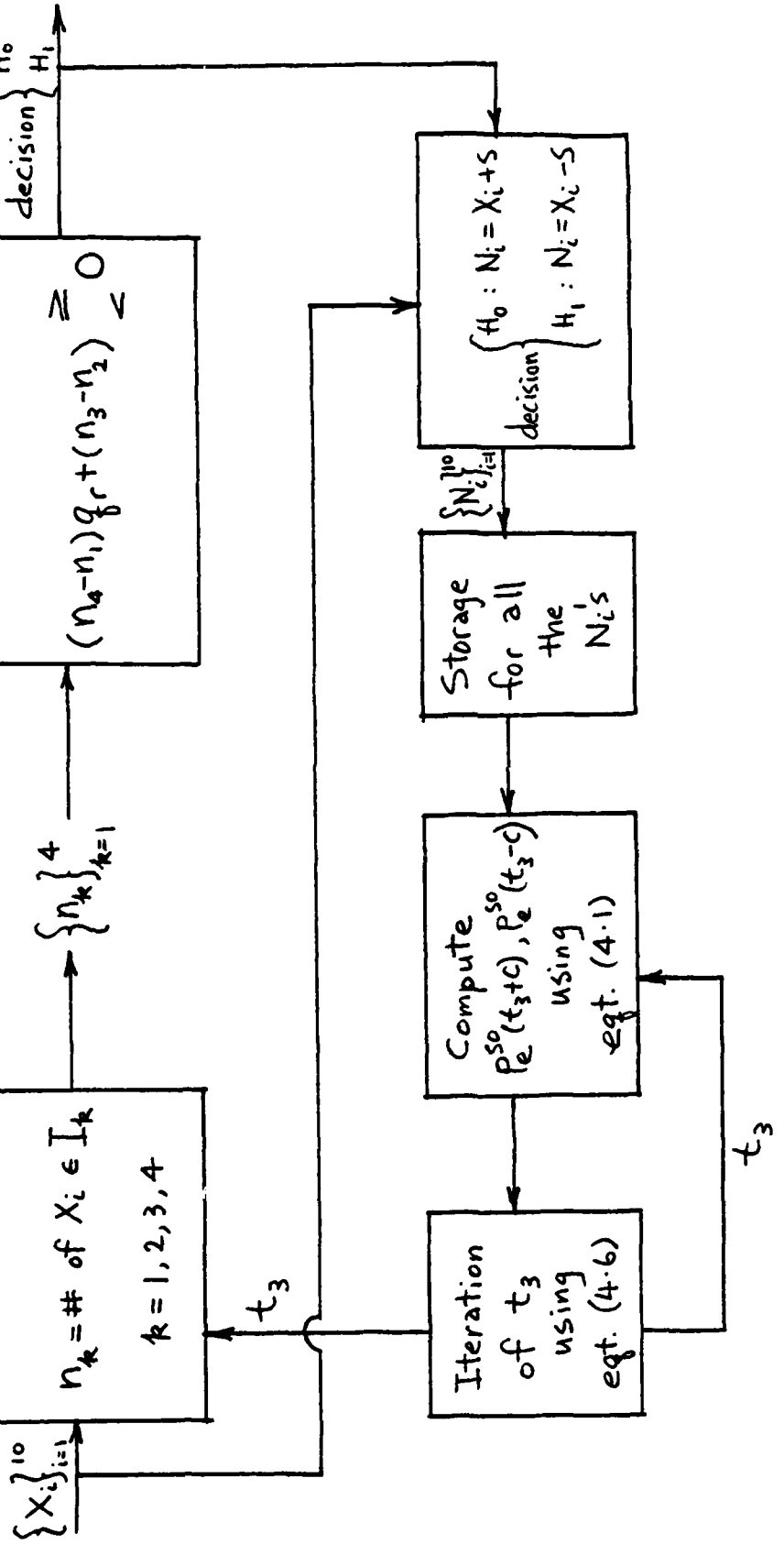


Figure 9. The decision directed adaptive quantizer detector structure with constant level ratio q_r and SNR = S.

the $P_e^{so}(t_3)$ and $\frac{\partial P_e^{so}(t_3)}{\partial t_3}$ are given as

$$P_e^{so}(t_3) = \tfrac{1}{2} \sum_{\underline{n} \in N_1^1} Pr\{\underline{n}|H_0\} + \tfrac{1}{2} \sum_{\underline{n} \in N_2} Pr\{\underline{n}|H_1\} \tag{4.1}$$

and

$$\frac{\partial P_e^{so}(t_3)}{\partial t_3} = \tfrac{1}{2} \sum_{\underline{n} \in N_1^1} Pr\{\underline{n}|H_0\} \left[\frac{n_3}{p_3^0} - \frac{n_2}{p_2^0} \right] f_N(t_3+s)$$

$$+ \tfrac{1}{2} \sum_{\underline{n} \in N_2} Pr\{\underline{n}|H_1\} \left[\frac{n_3}{p_3^1} - \frac{n_2}{p_2^1} \right] f_N(t_3-s) \tag{4.2}$$

where

$$Pr\{\underline{n}|H_i\} = \frac{n!}{n_1!n_2!n_3!n_4!} (p_1^i)^{n_1} (p_2^i)^{n_2} (p_3^i)^{n_3} (p_4^i)^{n_4}$$

$$N_1^1 = \{\underline{n} = (n_1,n_2,n_3,n_4) \text{ s.t. } (n_4-n_1)q_r + (n_3-n_2) \geq 0\}$$

$$N_2 = \{\underline{n} = (n_1,n_2,n_3,n_4) \text{ s.t. } (n_4-n_1)q_r + (n_3-n_2) < 0\}$$

We can see from these equations that finding the zero of $\frac{\partial P_e^{so}(t_3)}{\partial t_3}$ involves an additional estimation of the noise density function $f_N(\cdot)$; it was found that it may not be well approximated by any simple means. Besides, the computation required to obtain $\frac{\partial P_e^{so}(t_3)}{\partial t_3}$ is more involved and time consuming. So, in working with $P_e^{so}(t_3)$ directly and using some "peak seeking" methods to locate its minimum point, all we need is to have a good approximation of the p_i^j $i = 1,2,3,4$; $j = 0,1$, which are easily estimated.

The adaptation scheme used in our simulation as shown in Fig. 9 is decision directed. Every decision made from a block of 10 samples is assumed correct and, based on this decision (H_0 or H_1), the 10 samples are then modified so that their noise and signal content are revealed. Explicitly, the modification is as follows:

$$\text{if decision is } \begin{cases} H_0 & N_i = Y_i + s \\ H_1 & N_i = Y_i - s \end{cases} \qquad i = 1,2,3,\ldots, \quad n = 10$$

Hence, if all the decisions ever made were correct, all the noise data N_i so obtained will distribute according to the true noise present in the environment. With these noise data, we can approximate the p_i^0 and p_i^1 $i = 1,2,3,4$, necessary for the computation of P_e^{so} which is going to be minimized with respect to t_3. The approximation is done in the usual way; that is,

$$p_i^0 = \frac{\text{number of noise data from memory s.t. the value (noise data-s) is in the i-th interval}}{\text{total number of noise data stored in memory}} \tag{4.4}$$

$$p_i^1 = \frac{\text{number of noise data from memory s.t. the value (noise data+s) is in the i-th interval}}{\text{total number of noise data store in memory}} \tag{4.5}$$

Notice that the location of the i-th interval is determined by the current iterated breakpoint t_3, so p_i^0, p_i^1 and hence P_e^{so} change accordingly with t_3 in each iteration. The only way we can update the values of p_i^0 and p_i^1 is to check through the entire storage of the noise data and perform the above approximation for p_i^0 and p_i^1 in each iteration. In fact, this is the most troublesome thing to do in the whole algorithm in the simulation.

Once $P_e^{so}(t_3^{(\ell)})$ is found in the ℓ-th iteration using $t_3^{(\ell)}$, the iterated value of the breakpoint in the ℓ-th iteration, the $(\ell+1)$-st breakpoint value can be obtained by the following iterative process,

$$t_3^{(\ell+1)} = t_3^{(\ell)} - a_\ell \left\{ \frac{P_e^{so}(t_3^{(\ell)}+C_\ell) - P_e^{so}(t_3^{(\ell)}-C_\ell)}{2C_\ell} \right\} \quad . \tag{4.6}$$

This is the well-known Kiefer-Wolfowitz method in stochastic approximation. With this, the $t_3^{(\ell)}$ approaches, as ℓ (the number of iterations) goes to infinity, the limit t_3^0 which gives minimum value to the function $P_e^{so}(t_3)$. However, it is necessary for the two sequences a_ℓ (the stepping sequence) and C_ℓ satisfy the following conditions for convergence,

(1) $\lim_{\ell \to \infty} a_\ell = 0$

(2) $\lim_{\ell \to \infty} C_\ell = 0$

(3) $\sum_{\ell=1}^{\infty} a_\ell = \infty$

(4) $\sum_{\ell=1}^{\infty} \left(\frac{a_\ell}{C_\ell}\right)^2 < \infty$

In our simulation, a_ℓ and C_ℓ are chosen as $1/\ell$ and $1/(4\ell^{\frac{1}{4}})$, respectively and it can be shown that this choice of a_ℓ and C_ℓ does satisfy (1)-(4). With this, Equation (4.6) becomes

$$t_3^{(\ell+1)} = t_3^{(\ell)} - 2 \left\{ \frac{P_e^{so}(t_3^{(\ell)} + \frac{1}{4}\ell^{-\frac{1}{4}}) - P_e^{so}(t_3^{(\ell)} - \frac{1}{4}\ell^{-\frac{1}{4}})}{\ell^{3/4}} \right\} \quad . \tag{4.7}$$

As we have mentioned in a previous chapter, the smallest initial t_3 that can be used is governed by the largest element of the C_ℓ sequence, which is the first element ($C_1 = \frac{1}{4}$) in the case where $C_\ell = \frac{1}{4}\,\ell^{-\frac{1}{4}}$ (note: this sequence is monotone decreasing). The reason is simply because if the initial t_3 was smaller than $\frac{1}{4}$, $\left(t_3^{(1)} - \frac{1}{4}\right)$ would be negative and P_ℓ^{so} will be undefined. For a different choice of C_ℓ sequence and hence a different type of convergence behavior, the initial t_3 can be made as small as desired.

Computer simulations of the system in Fig. 9 are done with Gaussian and Cauchy noises, their density functions are $f_N(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ and $f_N(x) = \frac{1}{\pi(1+x^2)}$, respectively. The iterative scheme of (4.7) is used to iterate the optimal t_3 with signal-to-noise ratio S/N = 0.75 and various initial breakpoint values.

Figures 10 and 11 show how the iterative process has brought t_3 toward its optimal values (in cases where $t_3^{(1)} = 0.25$ and 2.0 with q_r fixed at 2.0 for Gaussian noise). However, the algorithm is far from converging even after 2500 iterations. Also, we see from Figs. 12 and 13 that the probability of making errors of the system approximates the theoretical values after a large number of iterations for Gaussian noise given in Fig. 2. Similar curves for Cauchy noise are given in Figs. 14-17. Notice that the curves in Figs. 12, 13, 16, and 17 are generated according to the following definition

$$\text{Probability of error at } \ell\text{-th stage} = \frac{\text{total incorrect decisions made by system up to the } \ell\text{-th stage}}{\text{total decisions made by system up to the } \ell\text{-th stage}}$$

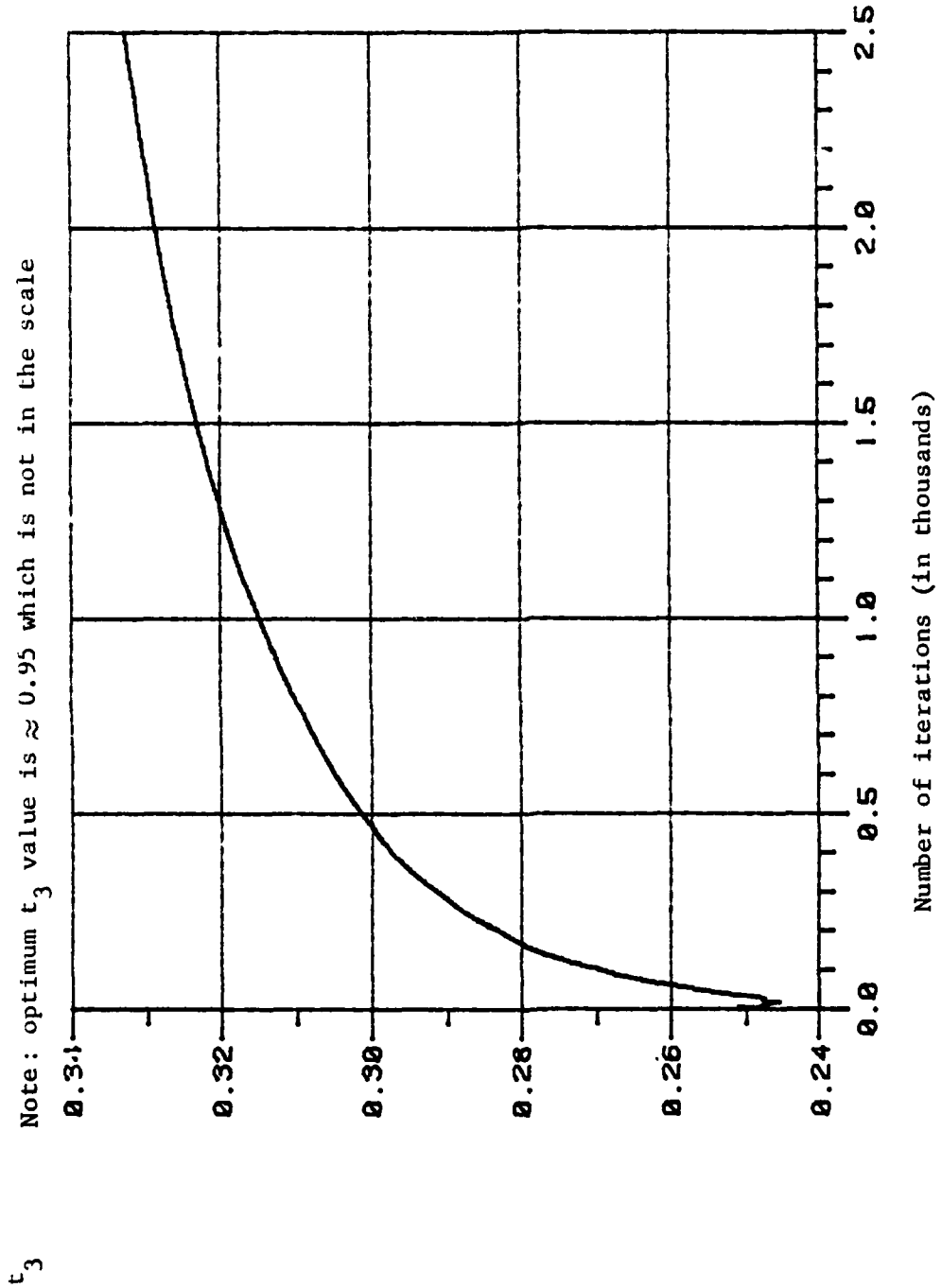


Figure 10. Iterated t_3 vs number of iterations for Gaussian noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 0.25.

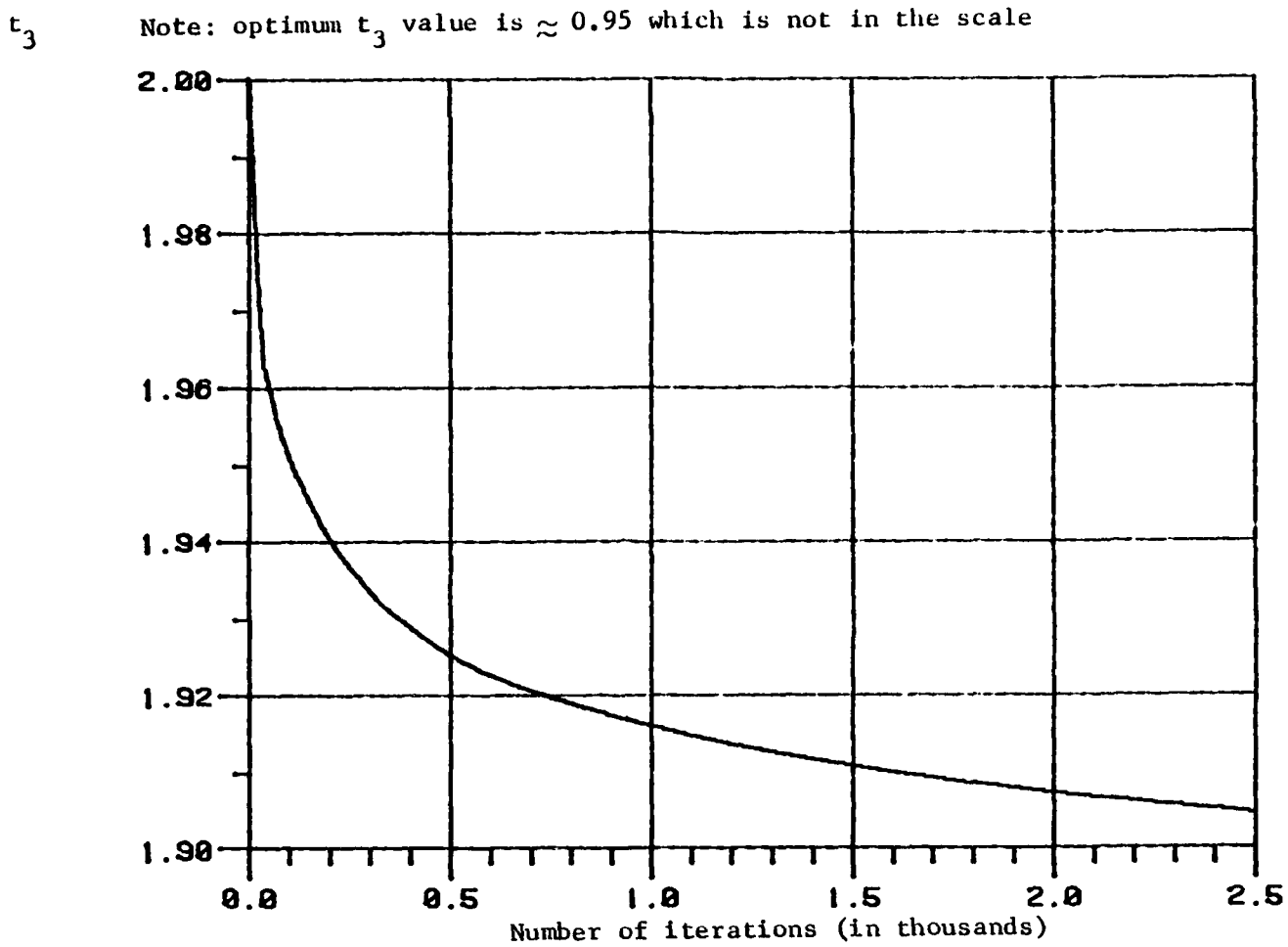


Figure 11. Iterated t_3 vs number of iterations for Gaussian noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 2.0 .

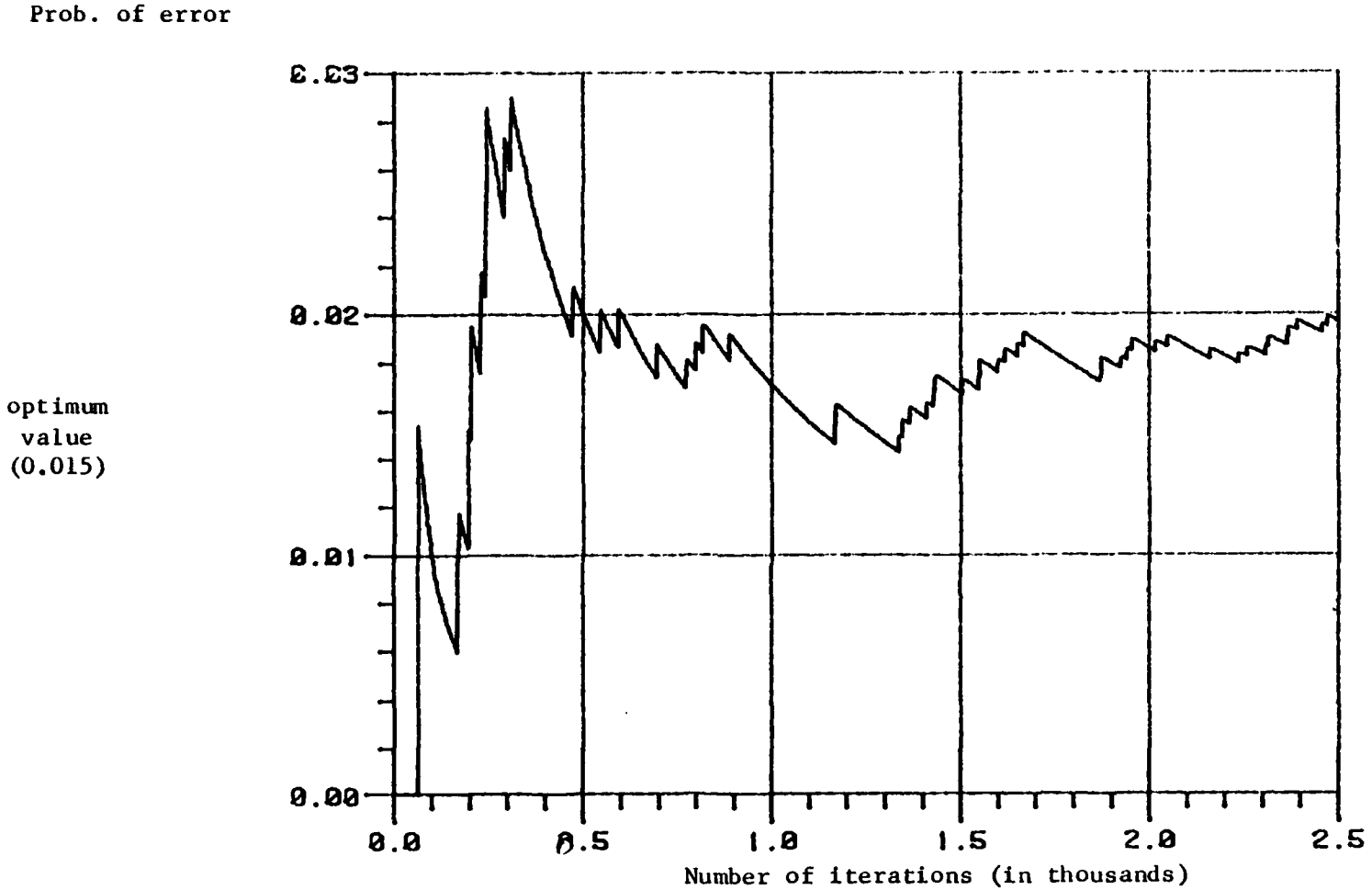


Figure 12. Iterated probability of error vs number of iterations for Gaussian noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 0.25.

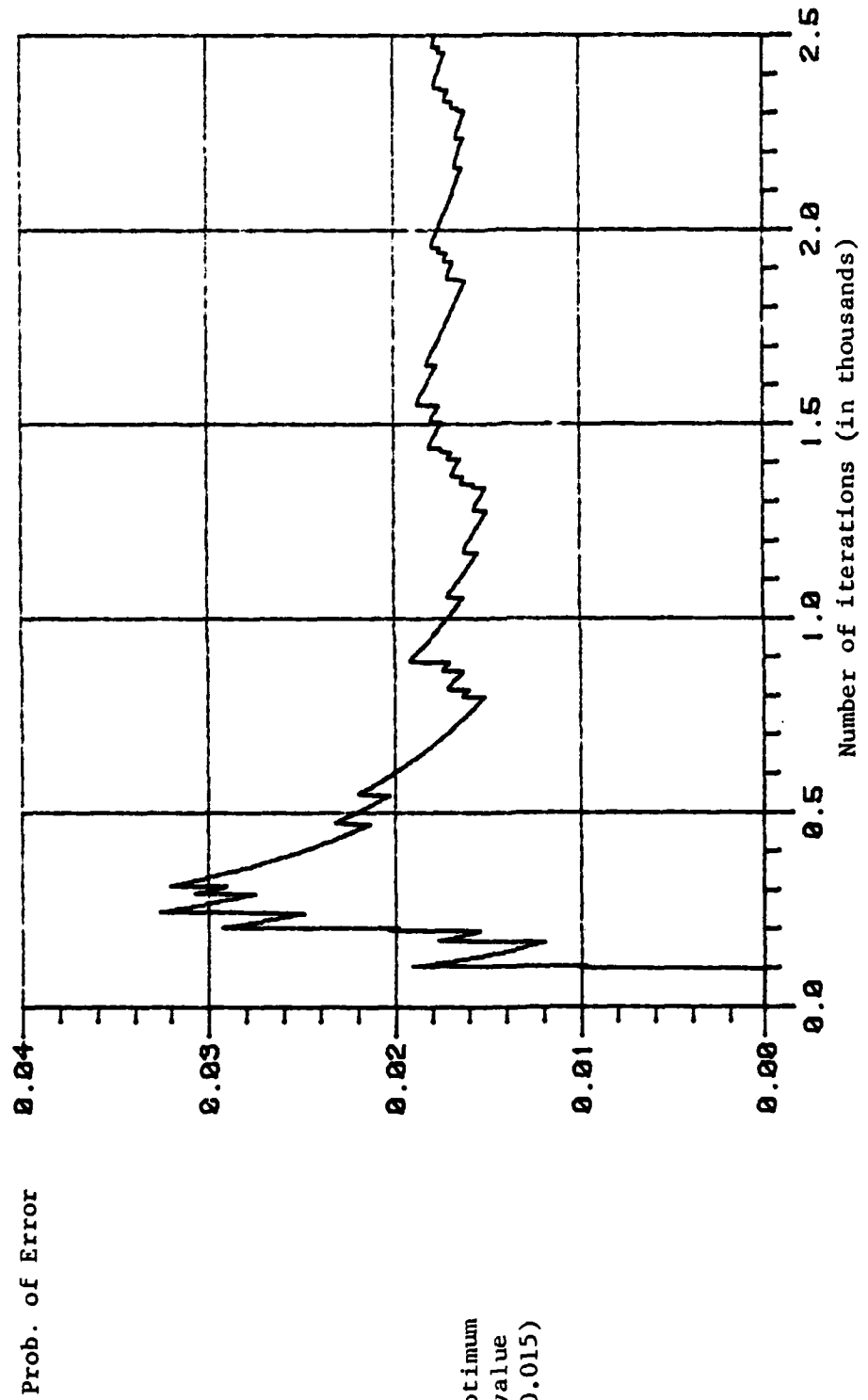


Figure 13. Iterated probability of error vs number of iterations for Gaussian noise with SNR = 0.75, $q_r = 2.0$, n=10 and initial $t_3 = 2.0$.

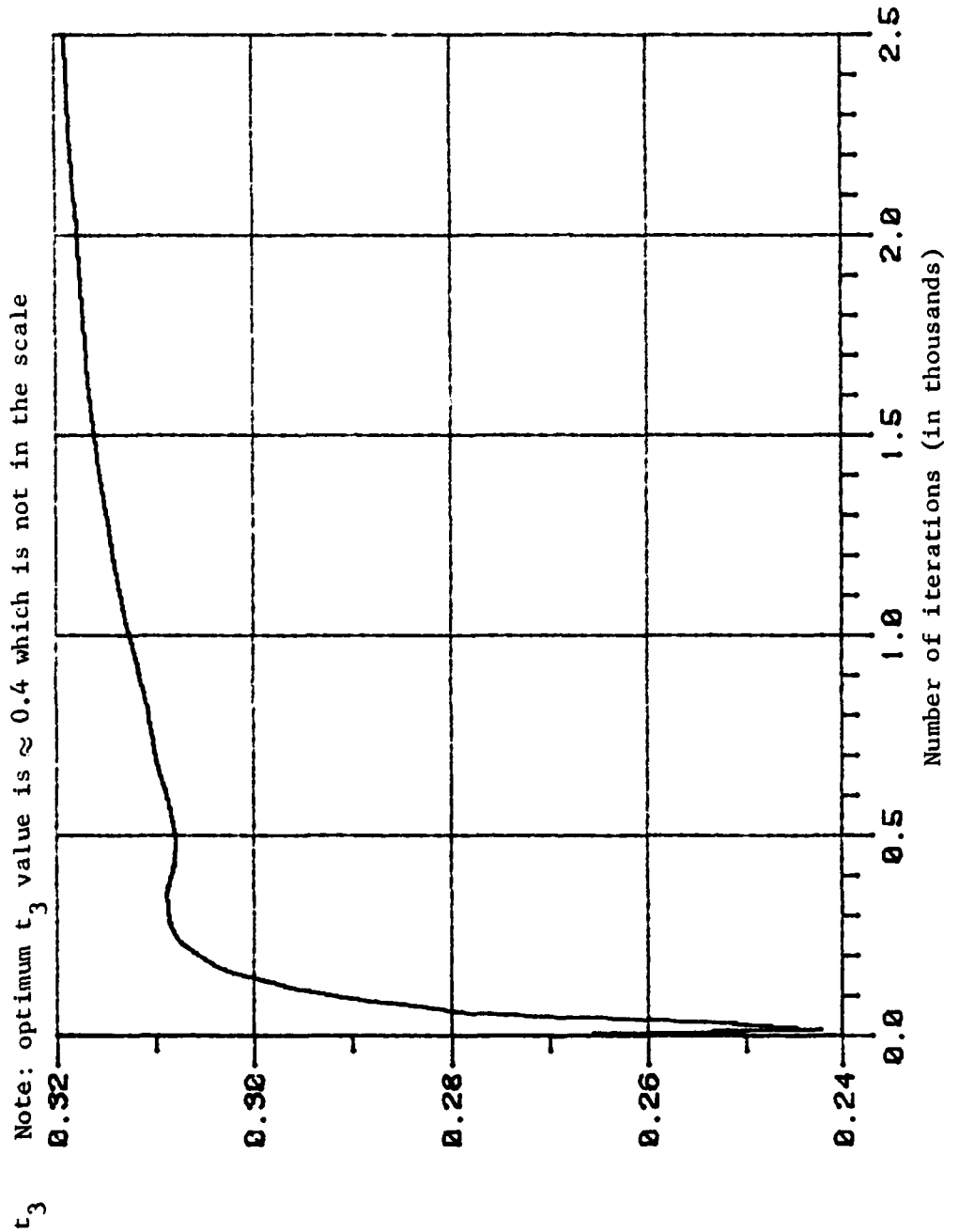


Figure 14. Iterated t_3 vs number of iterations for Cauchy noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 0.25 .

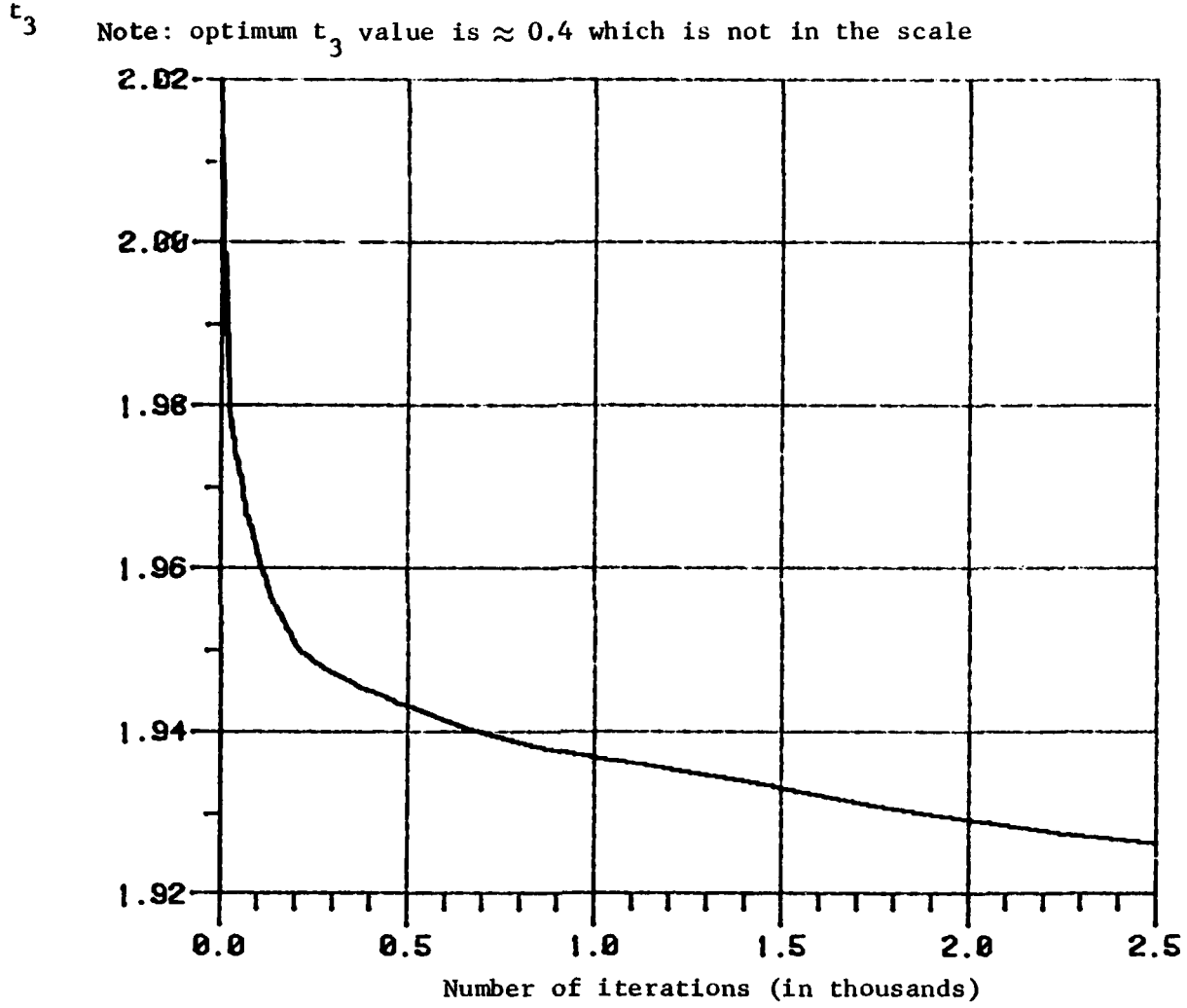


Figure 15. Iterated t_3 vs number of iterations for Cauchy noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 2.0 .

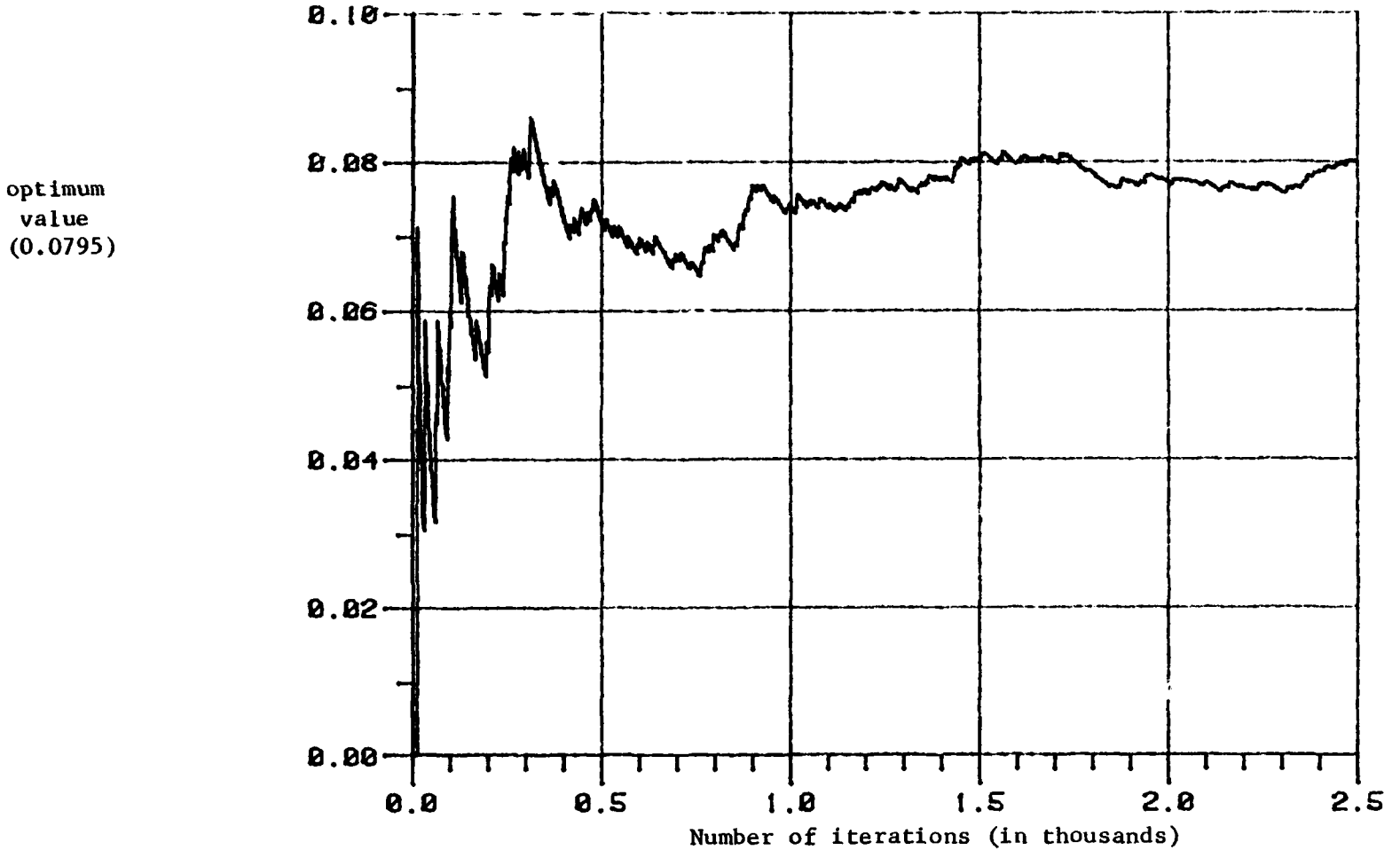


Figure 16. Iterated probability of error vs number of iterations for Cauchy noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 0.25.

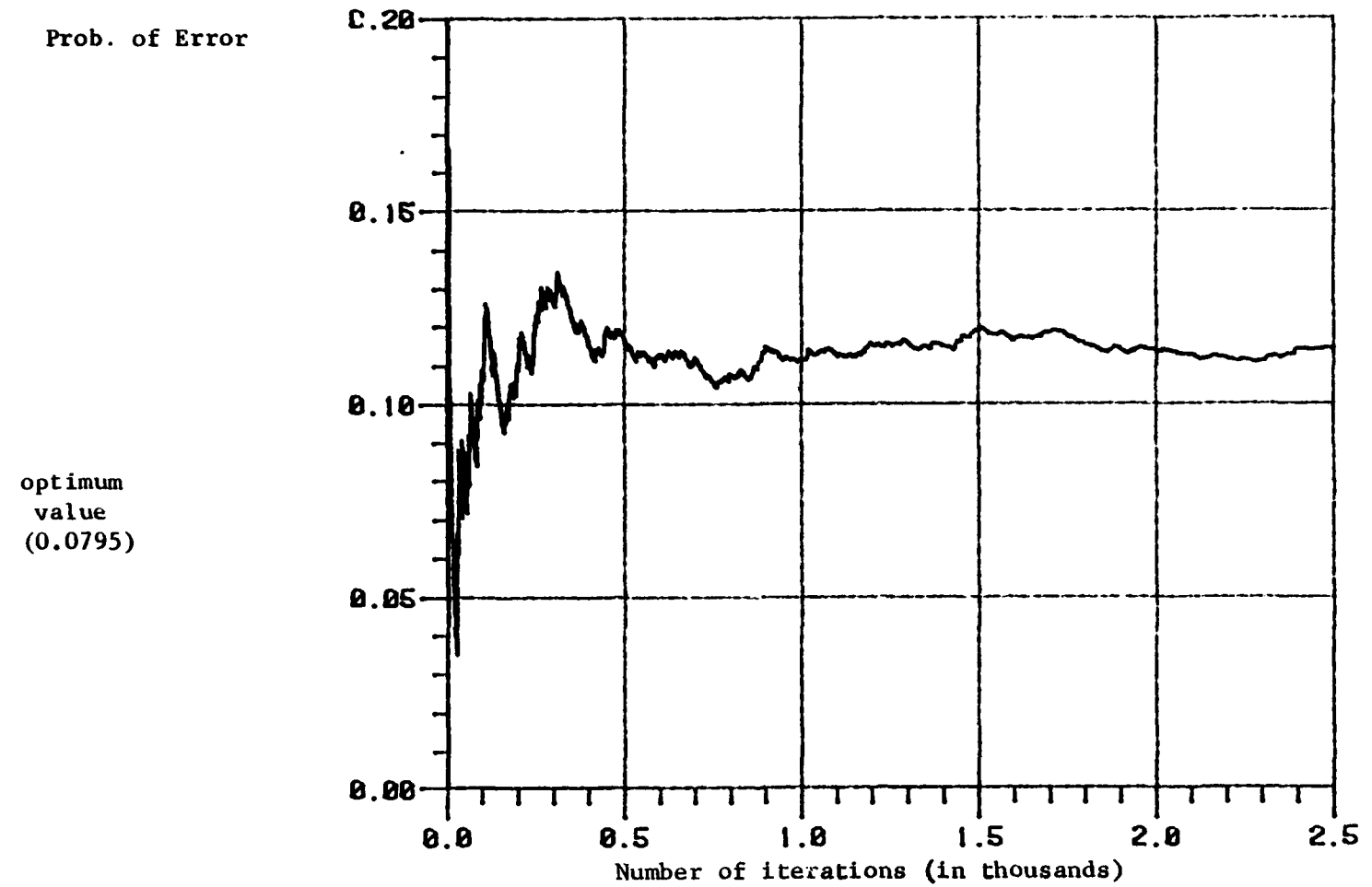


Figure 17. Iterated probability of error vs number of iterations for Cauchy noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 2.0.

This way of generating the probability of error of the detection system at each stage can only show that the system's probability of error does approach its minimal value but in no way indicates the system's "current potential" of making errors. This "current potential" of making errors by the system is actually the P_e^{so} evaluated at the current iterated breakpoint $t_3^{(\ell)}$.

4.3 The Modified Iterative Procedure

In order to speed up the convergence by a considerable amount, the above iterative scheme (4.6) is modified in the following way. If the sign of the quantity $(P_e^{so}(t_3^{(\ell)}+C_\ell) - P_e^{so}(t_3^{(\ell)}-C_\ell))$ is different from that of the previous quantity, a_ℓ will take on the next value (following the one used by $a_{\ell-1}$) in the stepping sequence; otherwise, the same value used by $a_{\ell-1}$ is used.

It is necessary that C_ℓ be constant valued and the stepping sequence be monotone decreasing (in addition to $\lim_{\ell \to \infty} a_\ell = 0$ and $\sum_{\ell=1}^{\infty} a_\ell = \infty$) for the modified iterative procedure be convergent. In our simulation with the modified scheme, $C_\ell = 0.125$ and the stepping sequence is again $1/\ell$ (harmonic sequence is monotone decreasing). Now, the smallest initial t_3 that can be used is 0.125 in this modified version of the adaptive system. We use Table 1 to help illustrate this modification.

Table 1. Illustration for modified iterative procedure.

ℓ	1	2	3	4	5	6	7	8
Sign of $P_e^{so}(t_3^{(\ell)}+C) - P_e^{so}(t_3^{(\ell)}-C)$	+	+	+	-	-	+	-	+
a_ℓ used	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$

Intuitively, this scheme gives faster convergence because we put large modification on t_3 when the direction of the search for the optimum does not change from that of the previous search and reduce the size of modification only when the search direction changes which indicates an overshoot of the iterated t_3 about its optimum and then that we need a finer search.

Figures 18-23 show t_3 converges to its optimal value in a much faster rate as compared with those in Figs. 10, 11, 14, and 15 for the same S/N (= 0.75), with q_r = 2.0 and initial t_3 = 0.13, 1.7 and 2.2 for Gaussian and Cauchy noises.

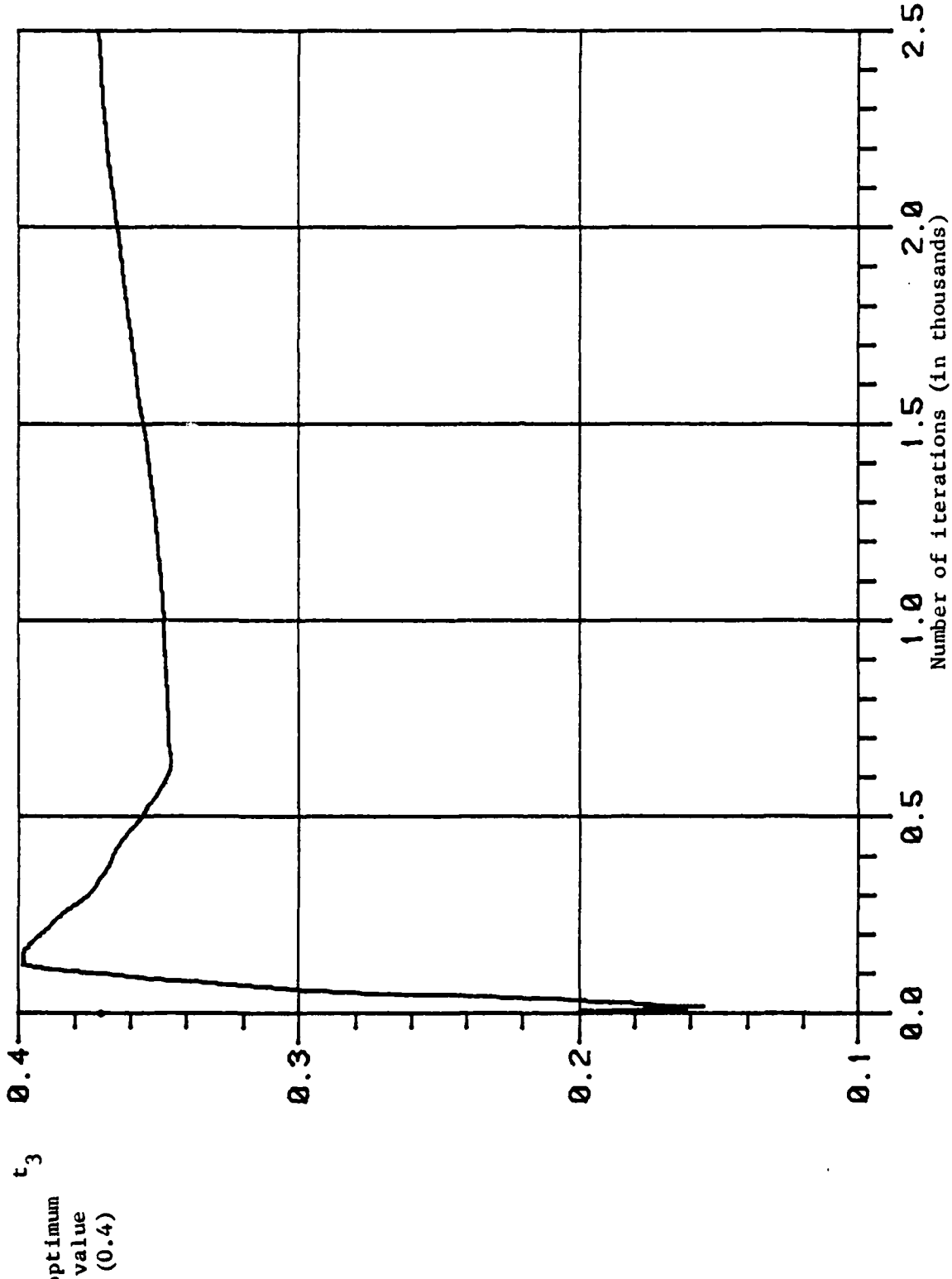


Figure 18. Iterated t_3 (with modified stepping sequence) for Cauchy noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 0.13.

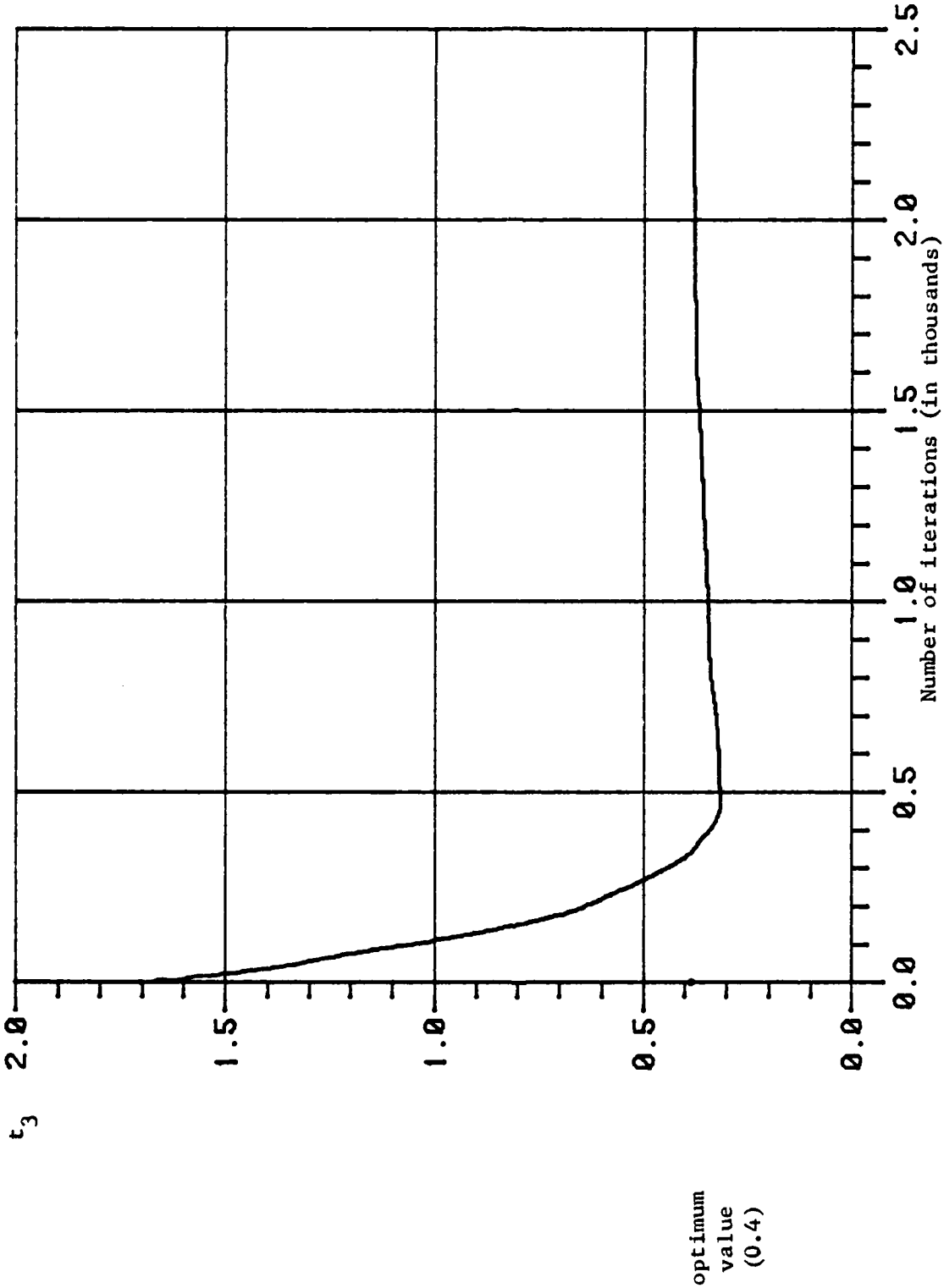


Figure 19. Iterated t_3 (with modified stepping sequence) for Cauchy noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 1.7.

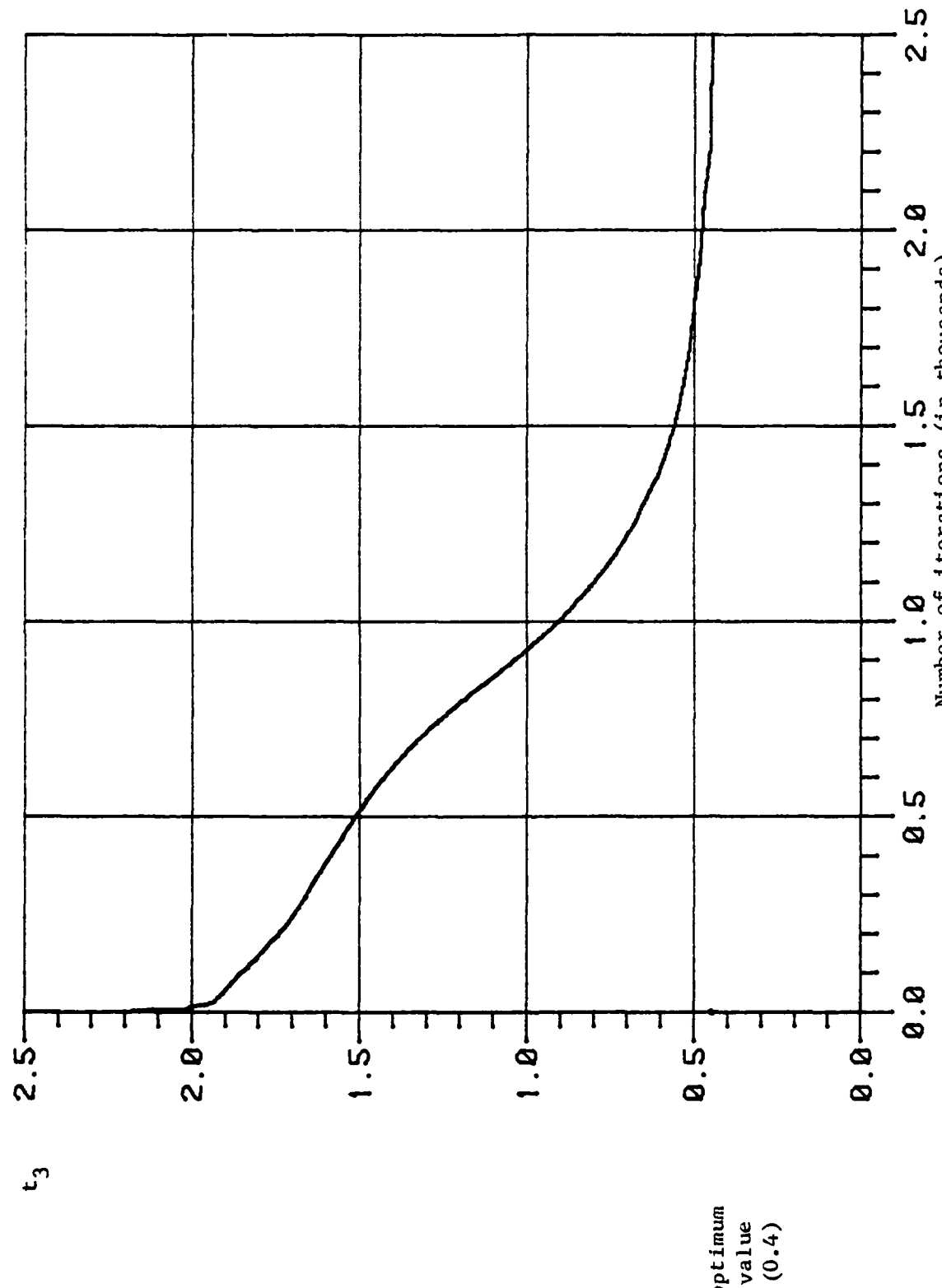


Figure 20. Iterated t_3 (with modified stepping sequence) for Cauchy noise with SNR = 0.75, $q_r = 2.0$, $n=10$ and initial $t_3 = 2.2$.

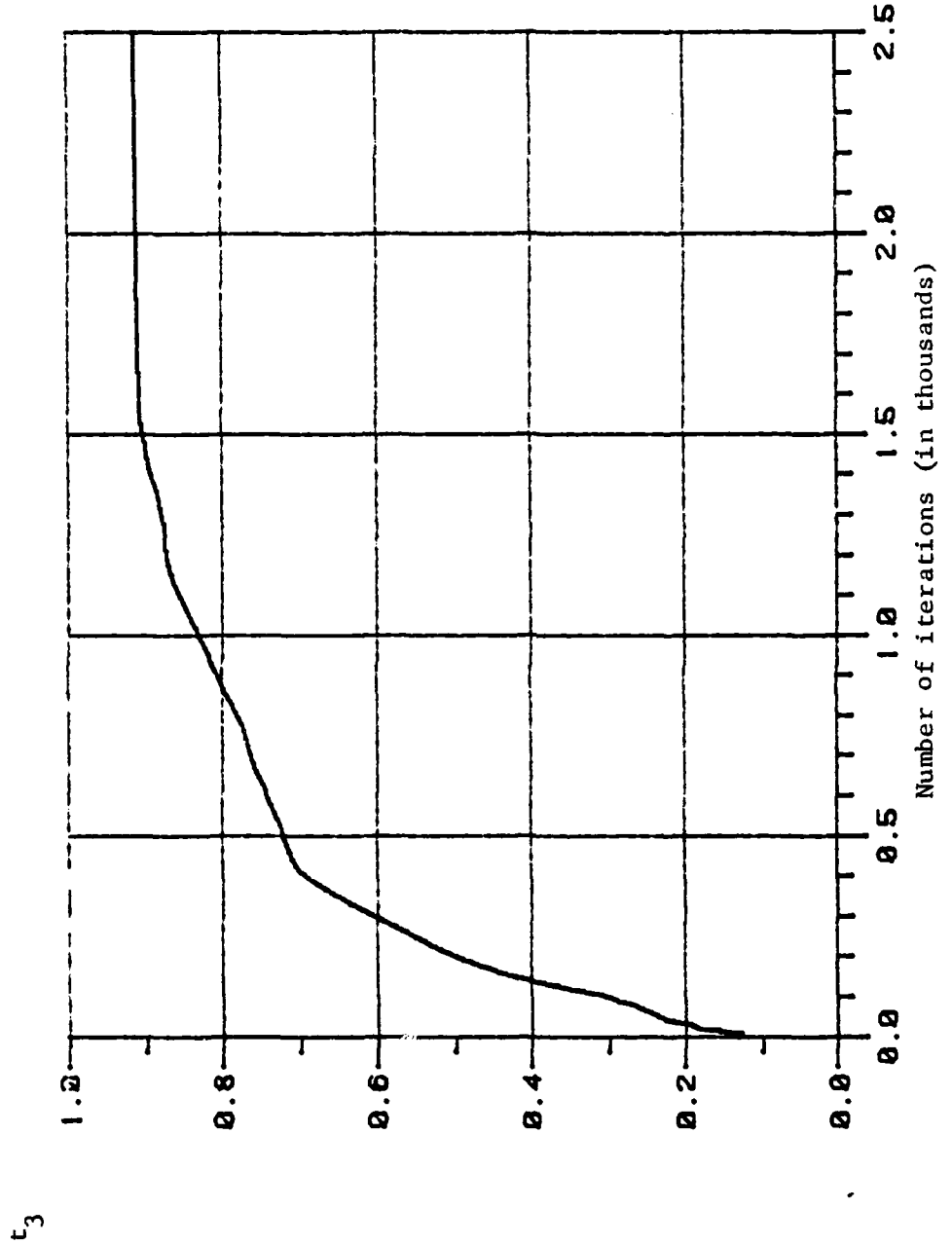


Figure 21. Iterated t_3 (with modified stepping sequence) for Gaussian noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 0.13.

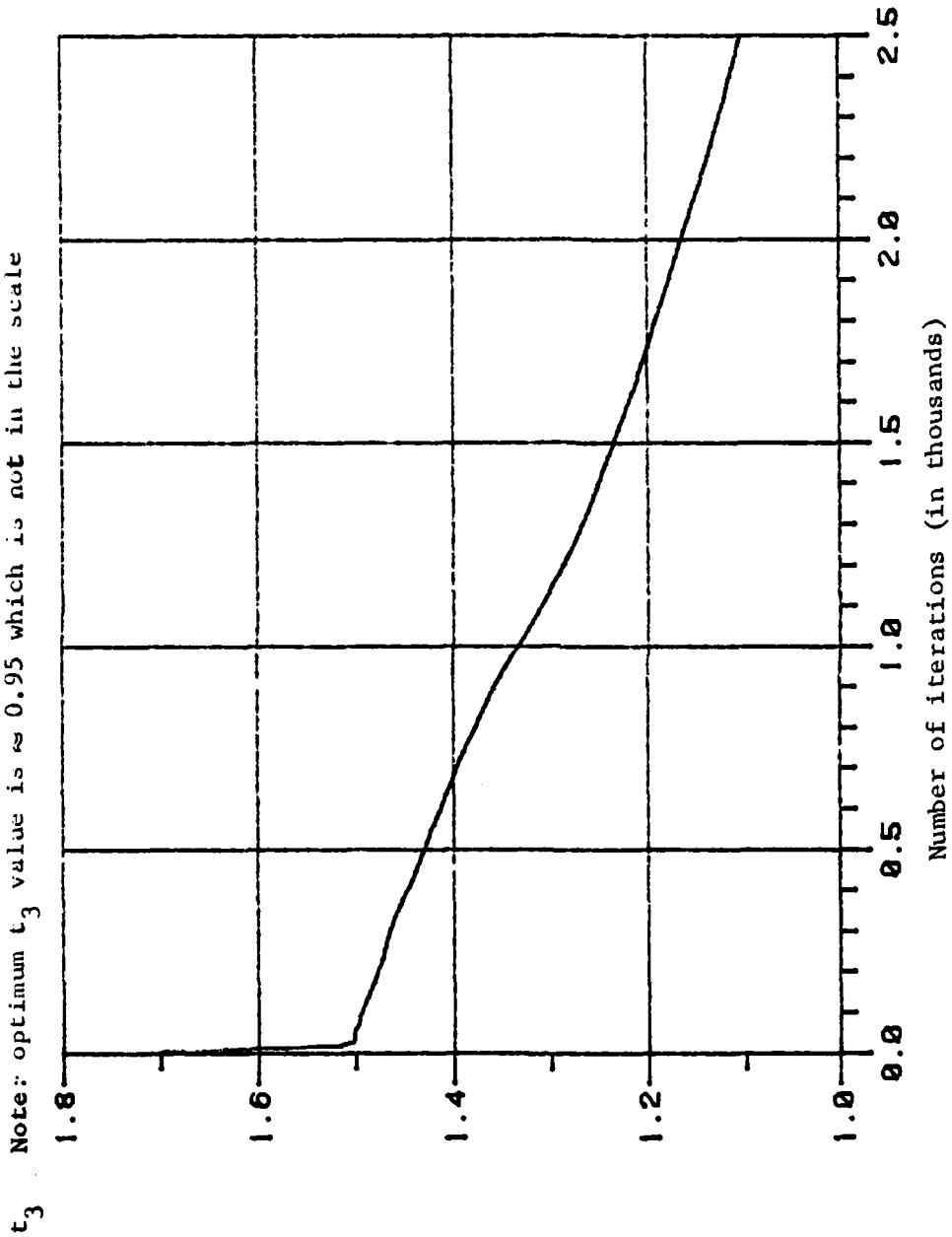


Figure 22. Iterated t_3 (with modified stepping sequence) for Gaussian noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 1.7.

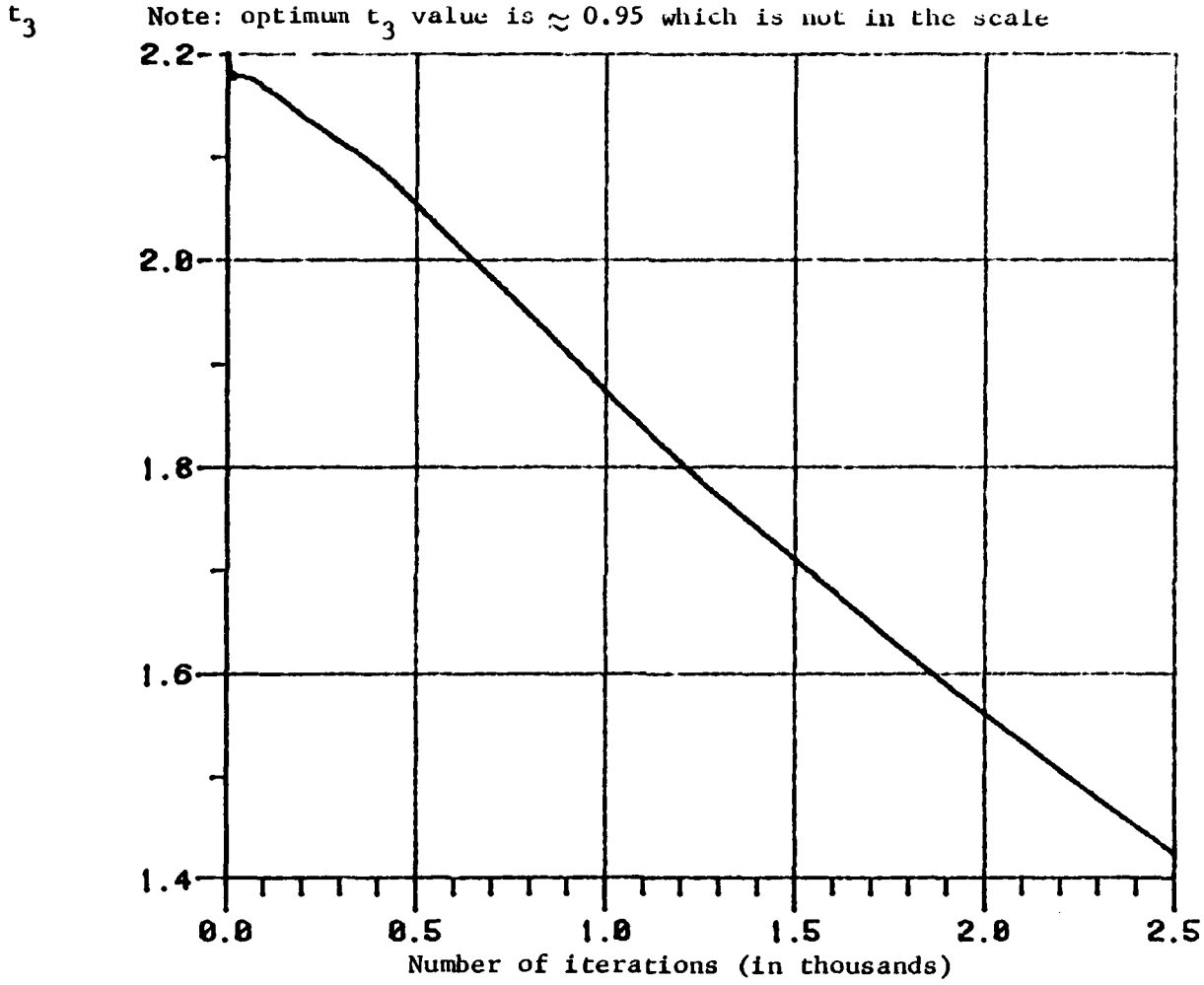


Figure 23. Iterated t_3 (with modified stepping sequence) for Gaussian noise with SNR = 0.75, q_r = 2.0, n=10 and initial t_3 = 2.2.

5. CONCLUSION

The simulation in the previous chapter shows that the adaptive detection system does not run away but eventually operates in its optimal state, under the conditions that the signal-to-noise ratio S/N = 0.75 and sample size n = 10, with Gaussian and Cauchy noises. It is expected that the system will work just as well with smaller signal-to-noise ratio level and is left to those who are interested to try with some other signal-to-noise ratio levels.

Comments on the size of the memory required to store the noise data is necessary. In the simulation, we store all the noise data available, which amounts to (sample size times the number of decisions made) 25000 storage locations in the final stage. However, the actual amount of noise data needing to be stored can be determined from the simulated curves in the previous chapter. The general guidelines in deciding the storage size are the size of the memory available in the system, the time allowed in processing the data during each iteration and the accuracy of the estimations of the P_i^0 and P_i^1 necessary to achieve the desired detector performance.

Finally, we note that other simulations were conducted in which the levels of the quantizer were adapted. The results of this analysis indicate that, although the levels do adapt to the noise, the performance gained in doing this is negligible when the initial t_3 is "small", that is, the performance of the adaptive quantizer-detector using $P_e^o(t_3)$ in the iteration process is the same as that using any $P_e^{so}(t_3)$ curves with $1 < q_r < n$. This coincides exactly to the observations discussed in Sections 3.3 and 3.4.

APPENDIX

DERIVATION OF THE UPPER BOUND TO P_e IN (2.33)

From (2.32), the upper bound to P_e is

$$P_e \le \left[\sum_{k=1}^{m'} 2\left(p_k^1 p_k^0\right)^{\frac{1}{2}} \right]^n \tag{A1}$$

With $s = K/\sqrt{n}$, $p_k^i = F_N\left(t_k \pm \frac{K}{\sqrt{n}}\right) - F_N\left(t_{k-1} \pm \frac{K}{\sqrt{n}}\right)$ "-" when $i = 1$ and "+" when $i = 0$, the bound in (A1) can be written as

$$\left\{ \sum_{k=1}^{m'} 2 \left[F_N\left(t_k - \frac{K}{\sqrt{n}}\right) - F_N\left(t_{k-1} - \frac{K}{\sqrt{n}}\right) \right]^{\frac{1}{2}} \left[F_N\left(t_k + \frac{K}{\sqrt{n}}\right) - F_N\left(t_{k-1} + \frac{K}{\sqrt{n}}\right) \right]^{\frac{1}{2}} \right\}^n \tag{A2}$$

(A2) approaches 1^∞, as $n \to \infty$, which is an indeterminate form. To apply L'Hospital Rule we first reform (A2) into

$$\exp\left\{ \frac{\log\left\{ \sum_{k=1}^{m'} 2 \left[F_N\left(t_k - \frac{K}{\sqrt{n}}\right) - F_N\left(t_{k-1} - \frac{K}{\sqrt{n}}\right) \right]^{\frac{1}{2}} \left[F_N\left(t_k + \frac{K}{\sqrt{n}}\right) - F_N\left(t_{k-1} + \frac{K}{\sqrt{n}}\right) \right]^{\frac{1}{2}} \right\}}{n^{-1}} \right\} \tag{A3}$$

Now the fraction inside the outermost bracket in (A3) is of the $\frac{0}{0}$ indeterminate form as $n \to \infty$.

Apply the L'Hospital Rule on this fraction yields $\frac{A^{-1}B}{n^{-\frac{1}{2}}}$, which goes to $\frac{0}{0}$ also as $n \to \infty$, with

$$A = \sum_{k=1}^{m'} 2 \left(p_k^0 p_k^1\right)^{\frac{1}{2}}$$

and

$$B = \sum_{k=1}^{m'} \frac{K}{2}\left[\left(\frac{p_k^1}{p_k^0}\right)^{\frac{1}{2}}\left(f_N\left(t_k + \frac{K}{\sqrt{n}}\right) - f_N\left(t_{k-1} + \frac{K}{\sqrt{n}}\right)\right)\right.$$

$$\left.\left(\frac{p_k^1}{p_k^0}\right)^{-\frac{1}{2}}\left(f_N\left(t_k - \frac{K}{\sqrt{n}}\right) - f_N\left(t_{k-1} - \frac{K}{\sqrt{n}}\right)\right)\right]$$

Using the L'Hospital Rule on $\frac{A^{-1}B}{n^{-\frac{1}{2}}}$ yields $\left[\frac{A^{-1}\frac{dB}{dn}}{-\frac{1}{2}n^{-3/2}} - 2\left(\frac{B}{A}\right)^2\right]$.

Since as $n \to \infty$, $B \to 0$, $A \to 1$ and $\left[\frac{dB}{dn}\Big/\left(-\frac{1}{2}n^{-3/2}\right)\right]$ approaches

$$K^2 \sum_{k=1}^{m'}\left\{\left(f_N'(t_k) - f_N'(t_{k-1})\right) - \frac{\left(f_N(t_k) - f_N(t_{k-1})\right)^2}{F_N(t_k) - F_N(t_{k-1})}\right\},$$

we obtain the upper bound to P_e in equation (2.33).

REFERENCES

1. W. J. Bushnell and L. Kurz, "The optimization and performance of detectors based on partition tests," Proc. 12th Annual Allerton Conference, Circuit and System Theory, pp. 1016-1023, Oct. 1974.

2. S. A. Kassam, "Optimum quantization for signal detection," IEEE Transactions on Communications, vol. COM-25, pp. 479-484, May 1977.

3. T. S. Ferguson, Mathematical Statistics - A Decision Theoretic Approach, Academic Press, New York 1967.

4. D. J. Sakrison, "Stochastic approximation: A recursive method for solving regression problems," in Advances in Communications Systems, vol. 2, pp. 51-106, Academic Press, New York 1966.

5. D. J. Wilde, Optimal Seeking Methods, Prentice Hall, Englewood Cliffs, N. J. 1964.

6. H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey Distance Measures in the Design of Generalized Quantizers for Binary Decision Systems," IEEE Transactions on Communications, vol. COM-25, pp. 893-900, Sept. 1977.