exponential family. The term estimator–correlator originated as an interpretation of the detector for the problem of discriminating a Gaussian signal in Gaussian noise. Since then it has been extended to a variety of situations both in discrete- and continuous-time problems. (See [6] for the genesis of the term and further discussion.) Equation (17) demonstrates that this interpretation is valid for any discrete signal (prior pdf) in Gaussian noise, and, in fact, for any detection problem where the noise has exponential statistics.

A similar test results when H_0 is a noise alone hypothesis with known statistics. The likelihood ratio is the average of the parameter-known likelihood and is sometimes called the average likelihood ratio [3, p. 151]

$$l(t) = \int_\Theta l(t \mid \theta)\pi_1(\theta)\, d\theta = \int_\Theta \frac{f_1(t \mid \theta)\pi_1(\theta)\, d\theta}{f_0(t \mid \theta_0)}. \quad (18)$$

Equations (16) and (18) are also related to a result of Esposito [4], [5] who showed that, under mild regularity conditions, the average likelihood ratio could be written in terms of the conditional likelihood ratio and a pseudoestimate $\tilde{\theta}(t)$, i.e., $l(t) = l(t \mid \tilde{\theta}(t))$. The quantity $\tilde{\theta}(t)$ is called a pseudoestimate because it may not be a good estimate for estimation purposes; it is the estimate required for optimum detection. Esposito then showed that, for $f(t \mid \theta)$ Gaussian, this pseudoestimate is equal to the CME, $\tilde{\theta}(t) = \hat{\theta}(t)$. Equation (16) shows that this equality of estimates holds true for all members of the exponential family.

Detection criteria are sometimes used which lead to an "ideal observer" detector, i.e., the ratio of the posterior pdf's [3, p. 84]. For purposes of notational simplicity, assume the conditional pdf's are the same under either hypothesis and that under H_0, θ_0 is known. From (9) and (15), we have

$$\log (f_1(\theta \mid t)/f_0(t \mid \theta_0)) = \int^t (\theta_1 - \hat{\theta}_1(\tau))\, d\tau - \theta_0 t + \log \pi_1(\theta_1) - k(b(\theta_1) - b(\theta_0)) - \log c_1. \quad (19)$$

As is easily seen, the detector factors into two terms. The first is the integral of the error in the CME and obviously depends on the data while the second term is solely determined by prior information. In practice, prior information is usually incomplete. Properties of the CME could then be used to evaluate the relative contributions of the two terms. For example, the CME is an efficient estimate, independent of the prior. With partial knowledge of the prior, one could determine bounds on the two terms and on how many samples would be required before decisions can be made (utilizing an approximation to the CME) with a high degree of confidence. An investigation of this type could justify approximate decision procedures such as the often used maximum likelihood test even for a moderate number of samples.

References

[1] T. S. Ferguson, *Mathematical Statistics, A Decision Theoretic Approach*. New York: Academic, 1967.
[2] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Wiley, 1959.
[3] C. W. Helstrom, *Statistical Theory of Signal Detection*, 2nd ed. Elmsford, N.Y.: Pergamon, 1968.
[4] R. Esposito, "A class of estimators for optimum adaptive detection," *Inform. Contr.*, vol. 10, pp. 137–148, Feb. 1967.
[5] ——, "On a relation between detection and estimation in decision theory," *Inform. Contr.*, vol. 12, pp. 116–120, Feb. 1968.
[6] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 350–361, May 1969.
[7] S. Schwartz, "Simultaneous detection and estimation for the exponential family," Information Sciences and Systems Lab., Dep. Elec. Eng., Princeton Univ., Princeton, N.J., Tech. Rep. 36, Nov. 1974.

An Algorithm for a Selective Nearest Neighbor Decision Rule

G. L. RITTER, H. B. WOODRUFF, S. R. LOWRY, AND T. L. ISENHOUR

***Abstract*—A procedure is introduced to approximate nearest neighbor (NN) decision boundaries. The algorithm produces a selective subset of the original data so that 1) the subset is consistent, 2) the distance between any sample and its nearest selective neighbor is less than the distance from the sample to any sample of the other class, and 3) the subset is the smallest possible.**

The nearest neighbor (NN) decision rule [1]–[3] is a conceptually simple, yet powerful, decision method. An unclassified sample is assigned to the same class as the nearest or most similar of n known samples. Cover and Hart [4], [5] have shown that under many reasonable (mild) conditions, for a many sample problem the risk in making an NN decision is less than twice the minimum or Bayes risk. Since, however, an NN rule is a search procedure on the known data file, all samples must be examined to classify each unknown. Thus when the data file becomes large, storage and computational requirements may make application prohibitively expensive.

If, in a preprocessing step, a subset of the n samples can be found such that all decision boundaries are unchanged, then the risk remains the same. Any sample not among the boundary-maintaining subset may then be neglected in an NN classification. This implies that, given any unknown sample x, the class of its NN in the n sample set will be the same as the class of its NN in the subset. To be practical, fewer calculations must be involved in obtaining this subset than in determining all of the NN decision surfaces. One alternative is to edit the data before applying an NN procedure. The edited k-NN rule eliminates samples that are incorrectly classified by the k-NN rule and the remainder of the data [6], [7]. Another alternate preprocessing step selects samples near the decision boundaries such that the class of the NN in the n sample set is the same as the class of the NN in the subset with a probability nearly equal to one.

Since the difficulty in finding a boundary-maintaining subset has been appreciated, techniques have been developed to choose points near the decision boundaries. The condensed NN decision rule [8] iteratively produces a consistent subset—a reference set that correctly classifies all samples in the known set. When the deletion of a sample in the condensed subset produces no change in the classification of any member of the complete set, the deleted sample may be excluded from the condensed subset. This is the basis of the reduced NN modification [9] of the condensed NN algorithm. Both authors indicate that the optimization of the consistent subset (finding the smallest or minimal consistent subset) is a problem without an obvious solution.

This correspondence presents an alternate method for approximating NN decision surfaces. Three criteria serve as the basis for the set of selective NN's:

1) the subset must be a consistent subset;
2) all samples must be nearer (more similar) to a selective neighbor of the same class than to *any* sample in the other class;
3) there must be no subset that satisfies criteria 1) and 2) and contains fewer members than the selective subset.

Manuscript received August 29, 1974; revised March 5, 1975. This work was supported by the National Science Foundation under Grant GP-43720X.

The authors are with the Department of Chemistry, University of North Carolina, Chapel Hill, N.C. 27514.

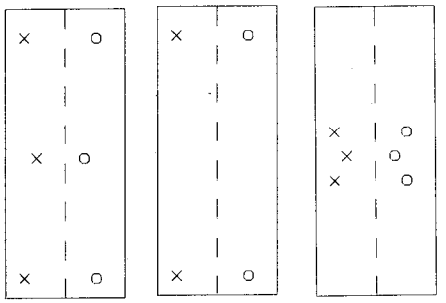

Fig. 1. (a) Illustrates two three-sample classes separated by NN decision boundary shown. (b) Gives possible condensed subset of (a). (c) Shows altered picture of (a) such that only two center points must be included in selective subset.

The second criterion is the principal difference between the condensed and the newly calculated selective subset. Criterion 2) for the condensed NN technique would read as follows:

> all samples must be nearer to a condensed neighbor of the same class than to any *condensed* neighbor of the other class.

This is simply a restatement of criterion 1). Furthermore, the second criterion for the selective NN subset allows a smallest subset to be calculated without requiring that every possible permutation of the samples be tested. However, it will be seen that the subset need not be unique (for an example, see the Appendix).

Since the criteria for the selective subset are more specific than those for the condensed subset, the selective set will not necessarily be minimally consistent. Likewise, in general, it will not be a reduced subset of the condensed NN set.

The difference between the selective and condensed sets may be illustrated by a simple figure. In Fig. 1(a), the decision boundary is the vertical line shown in the figure. The condensed NN algorithm may give the consistent subset shown in Fig. 1(b). This subset does not satisfy the selective NN second criterion—the deleted "X" is classified correctly by the condensed subset, but it is closer to the deleted "0" than to any point in the condensed set. For this example, the selective set will contain all the points in the set. If the points are altered as shown in Fig. 1(c), the selective set consists of only the two center points. Thus when the samples are closer together, the selective NN criteria appear to favor the samples near the decision surface.

Finding the Selective NN Subset

Let the set of n samples be represented by $\{(x_i,\theta_i)\}$, where θ_i denotes the class of the sample represented by x_i. Define x_j as a related neighbor to x_i if 1) x_j is in the same class as x_i (i.e., $\theta_j = \theta_i$) and 2) x_i is nearer to x_j than to any sample in the other class. Then if Y_i is the set of all related neighbors to sample x_i,

$$Y_i = \{x_j \mid \theta_j = \theta_i \text{ and } d(x_i,x_j) < \min_{\text{all } k} d(x_i,x_k) \text{ where } \theta_k \neq \theta_i\}.$$

The selective subset will be the smallest subset of $\{(x_i,\theta_i)\}$ which contains at least one member of Y_i for each sample x_i.

The algorithm needed to calculate the selective subset requires a mathematical representation of the related neighbors to each x_i. These related neighbors may be concisely described by an $n \times n$ binary matrix A, where

$$A_{ji} = \begin{cases} 1, & \text{if } x_j \in Y_i \\ 0, & \text{if } x_j \notin Y_i. \end{cases}$$

That is, a *one* appears in the matrix element (j,i), if x_j is a related neighbor to x_i. *Ones* along a vertical (column i) indicate the presence of related neighbors to x_i, and *ones* along the horizontal (row j) indicate that x_j is a related neighbor to other samples. The selective NN problem may now be restated in terms of the binary representation. A solution implies finding the fewest number of samples so that each has a related neighbor in the set, or finding the minimum number of rows such that at least one *one* appears in each column.

Given the matrix A, the following algorithm may be used to calculate the selective subset.

0) Set up an array to store members of the selective subset.

1) Sample x_j is placed into the selective subset if it is the only (remaining) related neighbor to some sample x_i. In the matrix notation, this implies saving x_j if the only (remaining) *one* in column i appears in row j. Delete all columns k such that x_j is a related neighbor to x_k, i.e., $A_{jk} = 1$.

In the first pass through 1), any sample point that has no related neighbor other than itself is included in the selective set. In subsequent passes through this step, sample x_j is placed in the selective set if it is the only remaining related neighbor to some sample x_m. Since the ultimate goal is to include in the selective set a related neighbor for each sample, all samples for which x_j is a related neighbor need no longer be considered. Therefore, all columns k that represent samples that have x_j as a related neighbor may be deleted. (The deleting process may be thought of as physically crossing out or removing the columns in the matrix representation.)

2) Row j is deleted if for all (remaining) columns i and for some (remaining) row k,

$$A_{ji} \leq A_{ki}.$$

Row j is deleted if, whenever row j contains a *one*, row k also contains a *one*, or if, whenever x_j is a related neighbor to some sample, x_k is also a related neighbor to that sample. Thus there is no need to consider x_j for the selective set, since all the same related neighbor information is included with sample x_k.

3) Delete column i if for all (remaining) rows j and some (remaining) column k

$$A_{ji} \geq A_{jk}.$$

At least one sample that is a related neighbor to x_k must be chosen for the selective NN subset. However, whenever a sample is a related neighbor to x_k, it is also a related neighbor to x_i. Thus when a sample that is a related neighbor to x_k is placed into the selective subset, by necessity it will also be a related neighbor to x_i. Then there is no need to consider the related neighbors of x_i or in the matrix delete column i. The columns that remain in the matrix then represent the samples that do not have a related neighbor in the subset. The remaining rows represent samples that may still be included in the selective set.

4) If no changes have been made in the matrix in 1)–3) and undeleted columns remain, go to 5). If no undeleted columns remain, the selective subset is complete. Otherwise, return to 1).

If it is necessary to go to 5), the matrix can be reduced no further by the simple deletion of rows and columns. Then it is necessary to search for the next sample that should be placed into the selective subset. Rather than requiring an exhaustive trial of each combination of remaining rows, a particular ordering scheme is implemented. Fortunately, in our experience this step is often unnecessary, and rarely necessary when many undeleted columns remain.

5) Employ a branching procedure on the remaining rows and columns.

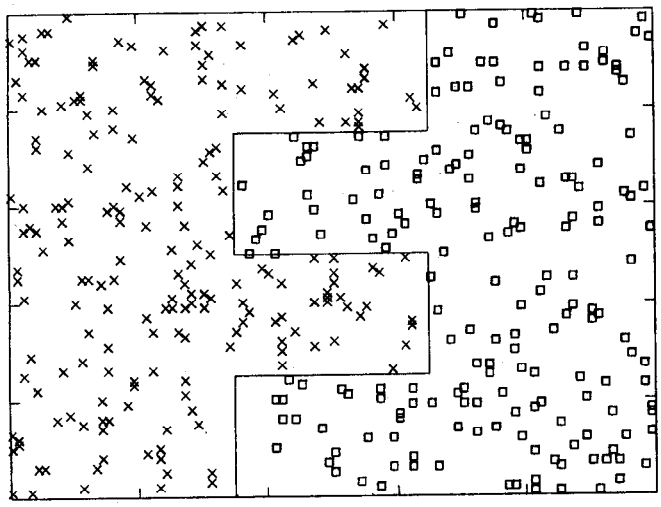

Fig. 2. Random points generated according to decision boundaries shown.

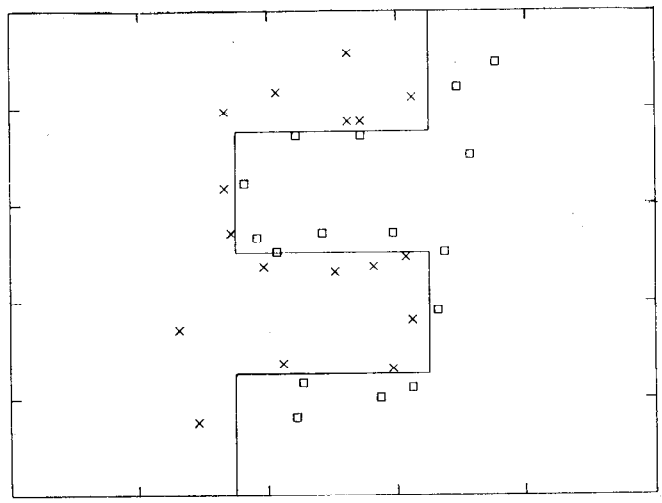

Fig. 3. Samples selected by selective NN decision rule.

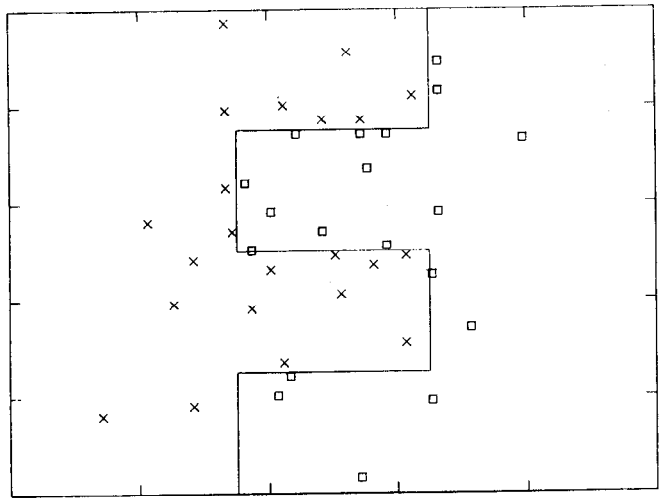

Fig. 4. Samples selected by condensed NN decision rule.

a) In order, assume that each remaining row will be in the selective subset. Subject to that assumption, calculate the minimum number of rows that *may* be needed to complete the subset. Or, given that a row is in the subset, determine the fewest number of rows such that the sum of the entries in the binary matrix for the rows and the remaining columns is at least equal to the number of remaining columns. The absolute minimum is the smallest value in the set of calculated minima.

b) For each of the rows that *may* complete the subset in the minimum number of samples, take that row as being in the subset. Thus place that row tentatively into the subset and apply 1)–4) to find out how many samples *are* required to complete the selective subset. If at any point this number of samples matches the absolute minimum, the selective subset is complete. If not, save the smallest number of samples that has been needed to complete the subset.

c) Since the selective subset cannot be described in the absolute minimum number of samples, increment the number. If the smallest value from b) matches the incremented absolute minimum, the selective subset is completed. Otherwise, return to b).

These steps may be summarized as follows: find the row that when assumed in the selective subset requires the fewest other rows to complete the subset.

One possible simple modification to the algorithm may be made.

2a) If $A_{ij} = A_{kj}$, for all (remaining) columns j, delete row m ($m = i$ or j) such that

$$\sum_l (A_{ml} - A_{lm}) = \min_{n=i \text{ or } j} \left\{ \sum_l (A_{nl} - A_{ln}) \right\}.$$

This step is necessary when the remaining elements of two rows are identical. Intuitively, it would be better to delete the sample that is farther from the decision surface. However, if the samples are about the same distance from the boundary, the one that is a related neighbor to fewer of the total samples should be deleted. These considerations are balanced in 2a) so that the deleted row represents a sample that is both far from the decision boundary (i.e., has many related neighbors) and that is a related neighbor to fewer of the samples. The algorithm is illustrated by the example problem shown in the Appendix.

Testing the Algorithm

The selective NN algorithm has been applied to real and artificial problems. Class boundaries as described by Hart [6] have been reproduced. Two two-dimensional uniform distributions are generated according to the decision surface and range shown in Fig. 2. 400 points, approximately 200 in each class, are chosen from the distributions, and the selective NN algorithm is applied. The algorithm terminates without using 5), and the subset contains 33 samples. These results are illustrated in Fig. 3. Fig. 4 shows the results for the condensed NN algorithm where 41 samples are necessary. (Interestingly, using the condensed NN subset, selective NN criterion 2) is disobeyed for nine of the 400 samples.)

Similarly, a collection of 630 low resolution mass spectra has been used to test the algorithm [10]. Each spectrum is coded in a 119-dimensional space and 16 chemically significant two-state questions are imposed. In each case, the relative prior class probability of the larger class to the smaller class is less than 2.1:1. For the 16 questions, the mean in the number of samples in the selective subset is 149. By comparison, the condensed NN algorithm selects an average of 180 samples, and the reduced NN algorithm requires 149 samples. Thus it appears that the selective NN algorithm gives sample reduction comparable to the reduced algorithm. These results are summarized in columns 2–4 in Table I.

The data set has also been divided into a 430 sample training set and 200 sample classification set. Condensed, reduced, and selective subsets are computed from the training set. The prediction of the NN, condensed NN, reduced NN, and selective NN

TABLE I
RESULTS USING CONDENSED, REDUCED, AND SELECTIVE NN DECISION RULES

Question	CNN	RNN	SNN	MAX. LIKELIHOOD *	% CORRECT NN	CNN	RNN	SNN
more than 12 hydrogens	164	132	131	50.3	79.5	78.0	78.0	81.0
branching carbon	175	148	139	51.0	85.0	85.5	85.0	86.0
more than 13 hydrogens	163	132	127	51.1	80.5	79.0	79.0	82.0
presence of ethyl group(s)	290	227	237	54.4	71.0	69.0	65.5	64.0
carbon without a hydrogen	248	205	210	54.8	66.5	62.5	62.5	66.0
more than 6 carbons	160	133	143	55.7	83.0	82.0	80.5	81.5
more than 7 carbons	163	144	145	56.7	53.5	59.0	58.0	57.5
presence of double bond(s)	188	156	157	61.8	82.5	80.0	81.0	84.0
presence of a ring	162	137	130	62.2	87.0	82.0	80.0	83.0
more than 11 hydrogens	167	138	136	62.2	80.0	78.5	77.0	75.5
more than 10 hydrogens	176	146	145	64.3	79.5	78.0	78.0	78.0
more than 2 methyl groups	195	163	158	65.1	81.5	76.0	74.5	78.0
presence of a 4 or more member ring	140	116	116	65.4	85.0	81.0	82.0	83.5
more than 1 methyl group	201	160	165	66.7	86.5	81.5	82.0	82.5
more than 14 hydrogens	148	121	120	67.5	71.5	70.5	70.5	72.0
more than 15 hydrogens	146	120	118	67.9	71.0	69.0	69.0	71.5
AVERAGE	180	149	149	59.8	77.7	75.7	75.2	76.6
STD. DEV.	39	30	33	6.3	8.9	7.5	7.8	8.3

* Max. Likelihood indicates relative size of classes.

algorithm on the classification set is shown in Table I (column 6-9). Although the results vary for the individual questions, the average result for prediction is that the error rate of the selective subset falls between the error rates for NN and condensed NN sets. Therefore, the NN decision boundaries are more precisely reproduced by the selective subset.

Although the work thus far has been encouraging, theoretical and practical problems remain. The increase in the risk caused by using a selective NN algorithm is yet to be characterized other than empirically. Practically, all training set distances must still be calculated as part of the preprocessing step. This serves as a practical limit in application when only a small number of unknowns are to be classified.

APPENDIX

Given that the binary matrix representation of eight data points is shown in the following matrix, the selective subset will be found (ignoring 2a))

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	1	0	0	0	0	1	0	0
(2)	0	1	0	0	0	0	1	0
(3)	0	0	1	0	0	0	0	0
(4)	0	1	0	1	1	0	0	0
(5)	0	0	0	0	1	0	1	0
(6)	0	0	1	0	0	1	0	0
(7)	0	0	0	1	0	0	1	0
(8)	0	0	0	1	0	0	0	1.

Once again, the *ones* in row 1 imply that sample 1 is a related neighbor to both sample 1 and sample 6. Similarly the *ones* in column 4 imply that samples 4, 7, and 8 are related neighbors to sample 4. The selective subset is a smallest subset such that each sample has a related neighbor in the set, or that at least one *one* appears in each column.

Continuing with the algorithm we have the following.

Step 1: Columns 1 and 8 have only one entry. The sample represented by row 1 must be in the selective subset as the only related neighbor to sample 1. (The *one* in column 1 allows that column to be deleted.) Taking row 1 to represent the first member of the selective set also forces a *one* in column 6. Thus column 6 is deleted from the matrix since the sample represented by that column has a related neighbor in the selective subset. Similarly the sample represented by row 8 must be in the selective subset, and columns 4 and 8 may be deleted; column 4 since that sample is in the subset and column 8 since it now has a related neighbor in the subset. Rows 1 and 8 will be deleted as representing members of the selective subset. The remaining matrix is shown

	(2)	(3)	(5)	(7)
(2)	1	0	0	1
(3)	0	1	0	0
(4)	1	0	1	0
(5)	0	0	1	1
(6)	0	1	0	0
(7)	0	0	0	1.

Step 2: All elements of row 7 are less than or equal to the corresponding elements in row 2. Thus row 7 may be deleted. Furthermore, $A_{3l} \leq A_{6l}$, for all remaining i, (since $A_{3l} = A_{6l}$, for all i, this is a case where 2a) could be used) therefore, row 3 may be deleted. This completes step 2, and leaves the following matrix

	(2)	(3)	(5)	(7)
(2)	1	0	0	1
(4)	1	0	1	0
(5)	0	0	1	1
(6)	0	1	0	0.

Step 3: There are no columns for which $A_{jl} \geq A_{jk}$, for all j, and thus no columns may be deleted at this step. Changes have been made during this iteration and thus return to step 1.

Step 4: There is only one entry in column 3, therefore, sample 6 must be included in the selective set. Deleting column 3 and the selective subset row 6 leaves the matrix shown

	(2)	(5)	(7)
(2)	1	0	1
(4)	1	1	0
(5)	0	1	1.

The next iteration through steps 2, 3, 1, 2, 3 allows no changes to be made, and thus step 5 is employed.

Step 5a): The sum of the entries for the rows is

row 2 gives sum = 2

row 4 gives sum = 2

row 5 gives sum = 2.

Since three columns remain, for each of the rows the minimum number of rows to complete the set will be two.

Step 5b): Select sample 2 as tentatively in the selective set. From step 1, columns 2 and 7 (and row 2) may be deleted. This leaves

	(5)
(4)	1
(5)	1.

$A_{45} \leq A_{55}$, and thus row 4 may be deleted. Thus assuming sample 2 is in the subset implies that sample 5 is in the subset (by one further application of step 1). The selective subset has been completed in two samples. This is the minimum value predicted, and thus the algorithm is complete.

A selective subset consists of samples 1, 2, 5, 6, and 8. This is a minimal subset, but it is not unique. For example, in the last step, sample 4 may have been selected instead of sample 5, and an equally valid subset consists of samples 1, 2, 4, 6, 8.

References

[1] E. Fix and J. L. Hodges, Jr., "Discriminatory analysis, non-parametric discrimination, consistency properties," USAF Sch. Aviation Medicine, Randolph Field, Tex., Proj. 21-49-004, Rep., Feb. 1951.

[2] ——, "Discriminatory analysis: small sample performance," USAF Sch. Aviation Medicine, Randolph Field, Tex., Proj. 21-49-004, Rep. 11, Aug. 1952.

[3] D. O. Loftsgaarden and C. P. Quesenberry, "A nonparametric density function," *Ann. Math. Statist.*, vol. 36, pp. 1049–1051, 1965.

[4] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 21–27, Jan. 1967.

[5] T. M. Cover, "Estimation by the nearest neighbor rule," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 50–55, May 1968.

[6] D. L. Wilson, "Asymptotic properties of nearest neighbor rules using edited data," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-2, pp. 408–421, July 1972.

[7] T. J. Wagner, "Convergence of the edited nearest neighbor," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-19, pp. 696–697, Sept. 1973.

[8] P. E. Hart, "The condensed nearest neighbor rule," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-14, pp. 515–516, May 1968.

[9] G. W. Gates, "The reduced nearest neighbor rule," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-18, pp. 431–433, May 1972.

[10] T. L. Isenhour and P. C. Jurs, "Some chemical applications of machine intelligence," *Anal. Chem.*, vol. 43, pp. 20A–35A, 1971.

The Relationship Between an Adaptive Quantizer and a Variance Estimator

DAVID L. COHN, MEMBER, IEEE, AND
JAMES L. MELSA, MEMBER, IEEE

***Abstract*—In this correspondence, it is shown that an adaptive quantizer with a one word memory can be viewed as one that estimates the variance of its input and normalizes the input by the square root of the estimate. It is shown that, even though the estimate is an exponential average, the effect of transmission errors does not die out. Finally, a method of combating the effect of such errors is described.**

I. Introduction

Several authors have proposed schemes for using adaptive quantizers when digitizing signals with variable dynamic ranges [1]–[9]. The technique has been applied primarily to digitizers for speech, and several interesting systems have resulted [10]–[12]. The basic principle of the adaptation is that in each time interval the range scale of the quantizer is multiplied by an expansion–contraction factor that is determined by the prior quantizer output. Thus the quantizer range tends to track the dynamic range of the input. In this correspondence, it will be shown that this adaptive quantizer can be made equivalent to a quantizer that forms a maximum likelihood estimate of the variance of its input stream, normalizes the input by the square root of the estimate, and quantizes the resulting ratio with a fixed quantizer. The equivalence requires a fixed relationship between the expansion–contraction factors and the output levels of the quantizer. It has been found [9] that this relationship leads to the minimum-mean-square error in one application.

The estimation procedure in the equivalent variance estimator is an exponential average. Thus it would seem that the effect of transmission errors would die out over time. It is shown that this is not the case, and a new method of limiting the effect of such errors is described.

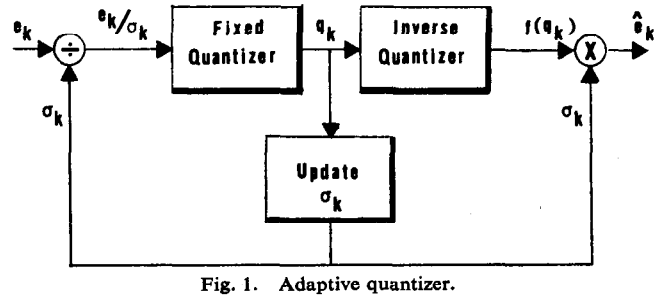

Fig. 1. Adaptive quantizer.

II. Adaptive Quantizer

An adaptive quantizer can be viewed either as one with a variable range or as one that normalizes the input with a variable and uses a fixed range. The latter model will be used, and the class of adaptive quantizers to be considered is illustrated in Fig. 1. The input to the quantizer e_k is normalized by the quantizer state variable σ_k. The resulting ratio is the input to a fixed quantizer. The output of the quantizer q_k denotes the level into which the ratio fell. The quantization level q_k is used to determine the quantizer output and also to update the quantizer state

Manuscript received December 27, 1974; revised April 2, 1975. This work was supported in part by the Defense Communications Agency under Contract DCA-100-74-C-0037.

The authors are with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, Ind. 46556.