IS-T-836

PROPORTION ESTIMATION AND CLASSIFICATION OF MIXED PIXELS IN MULTISPECTRAL DATA

Kenneth Ray Crouse

Based on a Ph.D. thesis submitted to Iowa State University

Ames Laboratory, USDOE
Iowa State University
Ames, Iowa 50011

Date Transmitted: January 1979

PREPARED FOR THE U. S. DEPARTMENT OF ENERGY
UNDER CONTRACT NO. W-7405-eng-82

DISCLAIMER

NOTICE

Available from: National Technical Information Service
U. S. Department of Commerce
P.O. Box 1553
Springfield, VA 22161

Price: Microfiche $3.00

TABLE OF CONTENTS

I. INTRODUCTION

A. The LANDSAT System

In response to an increasing awareness of our need for information about the earth's resources, the science of remote sensing has progressed rapidly in the past few years. Generally speaking, remote sensing involves observing objects without coming into physical contact with them. Information may be transmitted to the observer via magnetic, gravitational, or electromagnetic fields. Some types of sensors that are used in remote sensing include photographic cameras, television cameras, radar systems, and multispectral scanners.

Aerial photography has traditionally been the most widely used method of remote sensing, but the launching of the earth resources satellites, LANDSAT-1 in 1972 and LANDSAT-2 in 1975, has introduced a new dimension to remote sensing. The satellites offer a repetitive, synoptic view of the earth that has never before been possible. Examples of some of the areas in which LANDSAT data has been utilized include:

crop surveys

mineral and petroleum exploration

forest inventories

water resources monitoring

land use mapping

marine studies.

Other applications are discussed in Reeves (1975).

LANDSAT is in a sun-synchronous, near-polar orbit at an altitude of some 560 miles. The sun-synchronous nature of its orbit means that it passes over all locations at the same local sun time (9:42 a.m.). It circles the globe 14 times a day and covers the entire earth in 18 days. Since virtually every spot on the earth is covered once each 18 days, data can potentially be collected 20 times a year for any given location.

The sensor systems on LANDSAT include a return beam vidicon (RBV) and a multispectral scanner (MSS). The RBV suffered a malfunction shortly after the launch of LANDSAT-1, and consequently, the bulk of the data collection task has been handled by the MSS. The MSS functions basically as follows.

Reflected and emitted electromagnetic radiation (EMR) from a point on the ground is transmitted through the atmosphere and strikes an oscillating mirror in the lower part of the scanner. The mirror deflects the radiation to a set of optics that separate it into four distinct spectral bands. Radiation in each spectral band strikes a different electro-optical detector which transforms the EMR into an electrical signal that is recorded on magnetic tape and later telemetered to ground receiving stations. The spectral bands, which include two visible and two near-infrared bands, are:

MSS 4 0.5-0.6 micrometers

MSS 5 0.6-0.7 micrometers

MSS 6 0.7-0.8 micrometers

MSS 7 0.8-1.1 micrometers.

Because the responses in all four spectral bands are detected simultaneously, the recorded data is four-dimensional in nature and is often referred to as multispectral data.

As the satellite passes over the surface of the earth, the scanner mirror traces out a scan path perpendicular to the motion of the spacecraft. This is depicted in Figure 1. Six lines are scanned and recorded simultaneously. At a given instant of time the scanner views an element in each line which has the nominal dimensions of 80 meters by 80 meters on the earth's surface. There are approximately 3240 such elements per line. One complete image contains 2340 scan lines and covers a ground area of 185 kilometers on a side. Both photographic products and digital tapes are produced from the image data.

B. Analysis of LANDSAT Data

One approach to the analysis of LANDSAT data has been to simply apply the standard techniques of photo interpretation to the images in photographic form. This is and will continue to be an important method of analysis, but it suffers the following limitations:

1. It is very difficult for a human interpreter to simultaneously deal with data in four dimensions, such as is generated by LANDSAT;

2. The results of photo interpretation tend to be highly subjective and nonrepeatable since they depend on the skill and experience of the analyst;

3. The data throughput rates required to perform large-scale

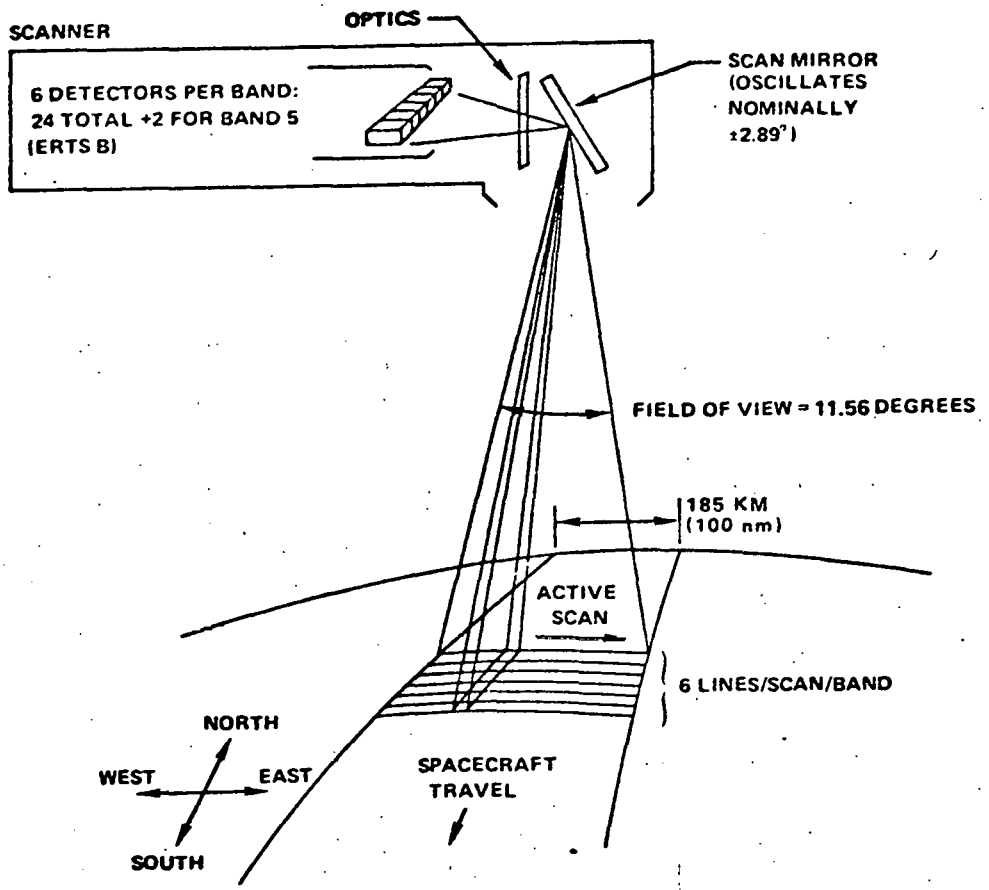


Figure 1. Schematic of LANDSAT MSS scan (after NASA, 1972)

surveys in a timely manner make manual interpretation of the data impractical for these applications.

These reasons, as well as others, have given rise to the development of techniques to automatically analyze LANDSAT data.

Central to the automatic analysis of multispectral data is the concept of a spectral signature. The spectral signature of a material is its relative response in terms of reflected plus emitted EMR as a function of wavelength. An example of hypothetical spectral signatures for three materials is given in Figure 2. Here it is evident that the responses for all three materials are fairly distinct at wavelengths

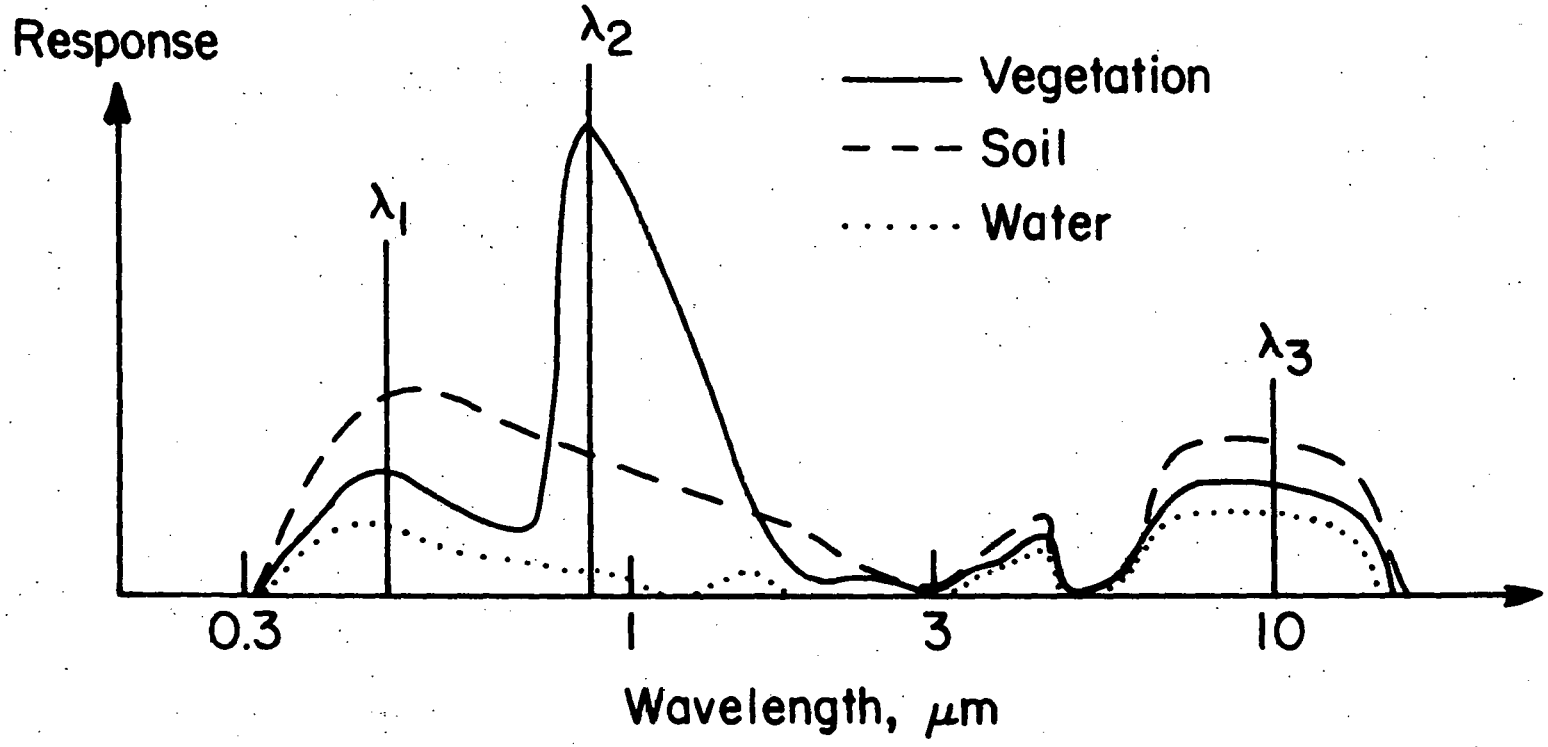


Figure 2. Spectral signatures of three common materials (from Landgrebe, 1971)

λ_1 and λ_2. Plotting the responses in two dimensions as in Figure 3 shows that the materials are easily distinguishable if one observes their responses at λ_1 and λ_2.

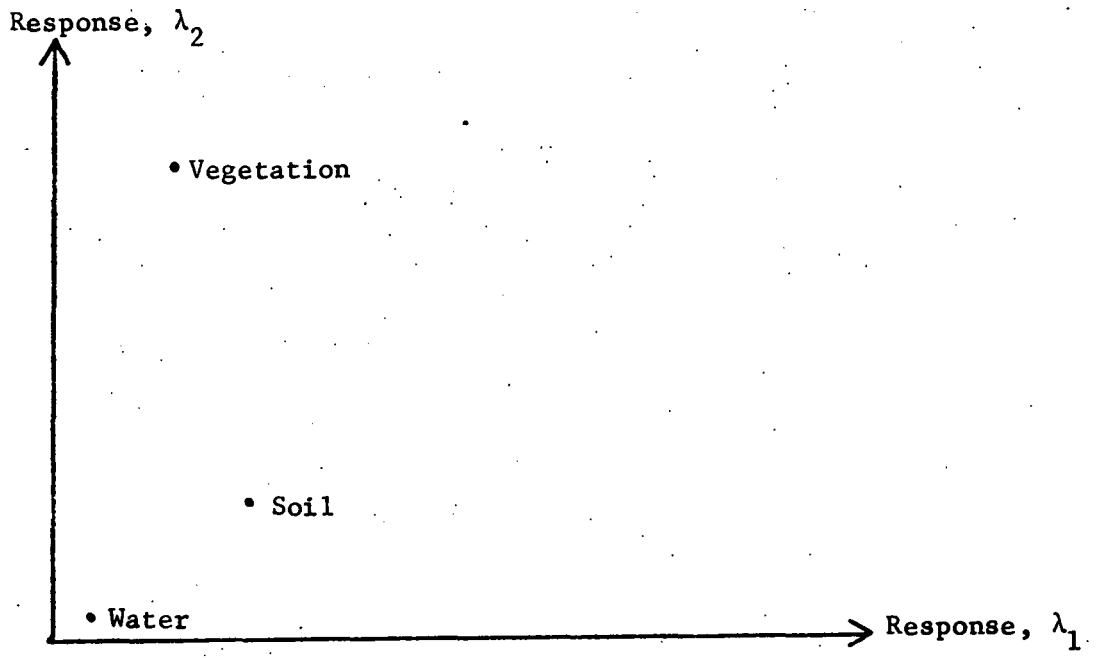


Figure 3. Spectral responses in the $\lambda_1\lambda_2$ plane (after Landgrebe, 1971)

In the real world things are not so neat and simple. A given material, corn, for example, exhibits variation in the amount of EMR it reflects and emits depending on many factors: maturity, moisture content, vigor, underlying soil type, variety, and others. Therefore, when a scanner makes multiple observations of a given material, the recorded responses can be expected to vary about some mean value as shown in Figure 4.

Now suppose in addition to the responses displayed in Figure 4 there is another observation u whose true identity is unknown. In order to classify the unknown point, one would like to divide the

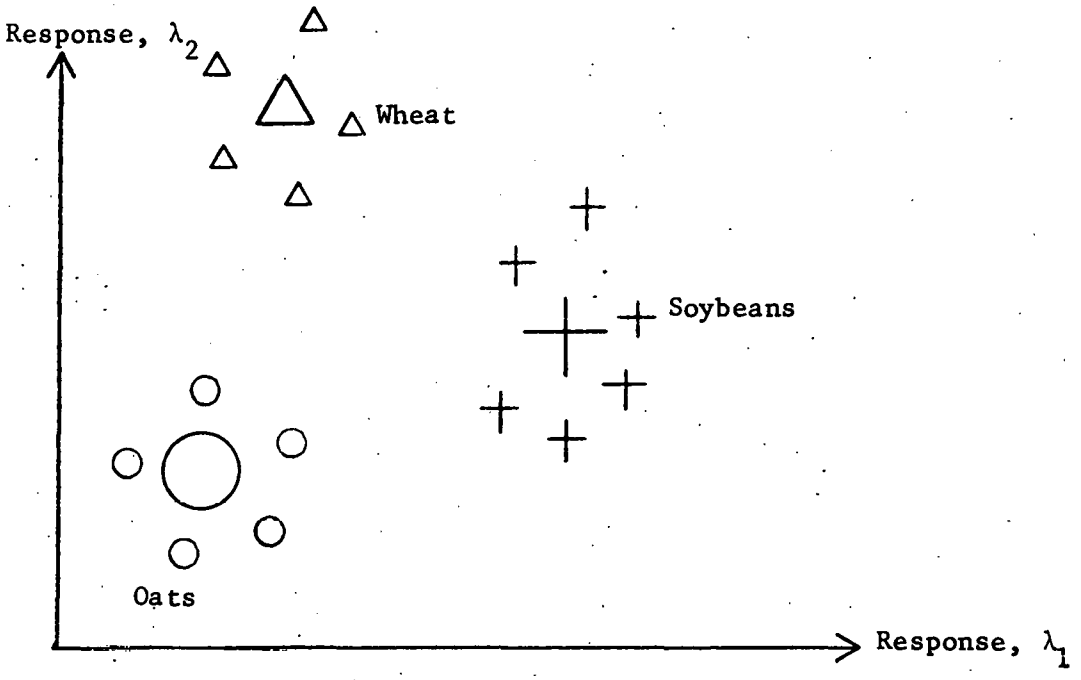


Figure 4. Spectral responses showing variation about the mean (after Landgrebe, 1971)

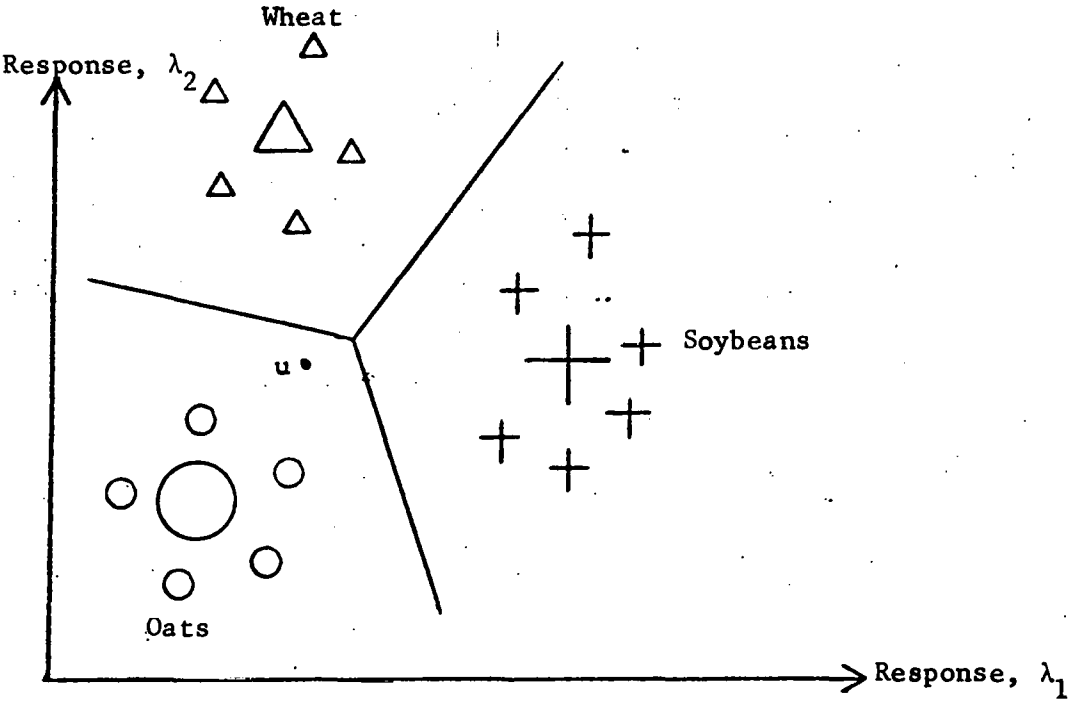


Figure 5. Classification by minimum distance to means (after Landgrebe, 1971)

response space into mutually exclusive and exhaustive regions and assign the unknown point to exactly one class corresponding to the region into which it falls. In this simple example the space may be partitioned by drawing the perpendicular bisectors of the lines joining the means as shown in Figure 5. This effects a classification based on minimum distance to means. The point u would be assigned to the Oats class under this classification rule.

Many procedures exist for carrying out the task of classification. It is not intended to attempt a comprehensive review of such methods here, but one method, known as the maximum likelihood, or Bayes, classifier, merits special attention because of its widespread use in remote sensing data analysis and its relationship to the techniques appearing in later chapters.

Let $\pi_1, \pi_2, \ldots, \pi_m$ denote distinct classes of material with a priori probabilities $q_1, q_2, \ldots, q_m$. Let x be an n-dimensional random observation, and let the class-conditional density functions be denoted

$$P(x|\pi_i), \qquad i = 1, \ldots, m.$$

If $C(i|j)$ is the cost of misclassifying an observation from class j into class i, assume that

$$C(i|j) > 0, \qquad i \neq j, \; i, j = 1, \ldots, m$$

$$C(i|i) = 0, \qquad i = 1, \ldots, m.$$

A Bayes rule R is one which partitions the observation space into mutually exclusive and exhaustive regions $R_1, R_2, \ldots, R_m$ such that the expected cost of misclassification is minimized. Given an

observation x, the expected cost of misclassifying x into π_i is

$$L_x(i) = \sum_{j=1}^{m} C(i|j)p(\pi_j|x).$$

Applying Bayes' formula, this may be written as

$$L_x(i) = \sum_{j=1}^{m} C(i|j)p(x|\pi_j)q_j/p(x),$$

where p(x) is the unconditional probability of observing x. It is not hard to show that minimizing $L_x(i)$ with respect to i is equivalent to choosing i = k such that

$$\sum_{\substack{j=1 \\ j\neq k}}^{m} C(k|j)p(x|\pi_j)q_j \leq \sum_{\substack{j=1 \\ j\neq \ell}}^{m} C(\ell|j)p(x|\pi_j)q_j, \quad \ell = 1, \ldots, m. \tag{1.1}$$

Anderson (1958, page 143) has shown that a procedure that assigns observation x to region R_k whenever (1.1) holds is a Bayes procedure.

In the special case of equal costs of misclassification,

$$C(i|j) = C, \qquad i \neq j,$$

and equal prior probabilities, condition (1.1) reduces to

$$\sum_{\substack{j=1 \\ j\neq k}}^{m} p(x|\pi_j) \leq \sum_{\substack{j=1 \\ j\neq \ell}}^{m} p(x|\pi_j), \qquad \ell = 1, \ldots, m,$$

which is equivalent to

$$p(x|\pi_k) = \max_i \; p(x|\pi_i) \; , \tag{1.2}$$

the criterion for the maximum likelihood solution.

In remote sensing the assumption is usually made that the spectral response for class i follows a multivariate normal distribution with

mean μ_i and covariance matrix Σ_i. Let n be the dimensionality. If the prior probabilities are taken to be

$$q_i = 1/m, \qquad i = 1, \ldots, m,$$

and the cost structure is

$$C(i|j) = 1, \qquad i \neq j$$

$$C(i|i) = 0, \qquad i = 1, \ldots, m,$$

then the Bayes rule R reduces to assigning x to class k if (1.2) is true. Thus, the Bayes rule and the method of maximum likelihood are equivalent in this case.

With the means and covariance matrices as given above, one may write out the class-conditional probabilities as

$$p(x|\pi_i) = (2\pi)^{-n/2}|\Sigma_i|^{-1/2}e^{-1/2(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)}, \qquad i = 1, \ldots, m. \tag{1.3}$$

Typically, the class means and covariance matrices are unknown and must be estimated from samples by $\hat{\mu}_i$ and $\hat{\Sigma}_i$, respectively. If one substitutes these estimates for the true parameters in (1.3) and expands (1.2) in terms of the expressions given in (1.3), one has, after a little manipulation, that

$$-\ln|\hat{\Sigma}_k| - (x-\hat{\mu}_k)'\hat{\Sigma}_k^{-1}(x-\hat{\mu}_k) \geq -\ln|\hat{\Sigma}_i| - (x-\hat{\mu}_i)'\hat{\Sigma}_i^{-1}(x-\hat{\mu}_i), \quad i = 1, \ldots, m. \tag{1.4}$$

The rule that assigns x to class π_k whenever (1.4) is true is called the maximum likelihood classifier. It is usually applied to each data point in an image on a point-by-point basis. It has been used extensively

in remote sensing applications because it is relatively simple to apply, and it gives excellent accuracy in many instances.

C. The Mixtures Problem

One question that an analyst frequently would like to answer is, "How much of a certain material is present in a specified region on the ground?" For example, he may want to know how many acres are planted to each of several crops in a county. If each resolution element viewed by a scanner were to contain exactly one type of material, the acreage estimation task would be straightforward: count the number of elements assigned to each crop and multiply by the size of a resolution element.

Earlier in the chapter it was stated that the size of a resolution element viewed by LANDSAT is about 80 meters square, or 1.1 acres. Obviously, objects smaller than 80 meters by 80 meters will not completely fill the field of view and will be seen as a mixture of the object and its background. Even for objects larger than 1.1 acres, a resolution element that overlaps the boundary between two large objects will be viewed as a mixture of the radiation emanating from each object. In this case the spectral response recorded by the sensor will not be characteristic of either object.

Suppose corn and bare soil have the spectral signatures shown in Figure 6. Then mixtures of 20% bare soil – 80% corn and 50% bare soil – 50% corn would have signatures as shown in Figure 7. A classifier trained to recognize corn and bare soil based on the signatures of

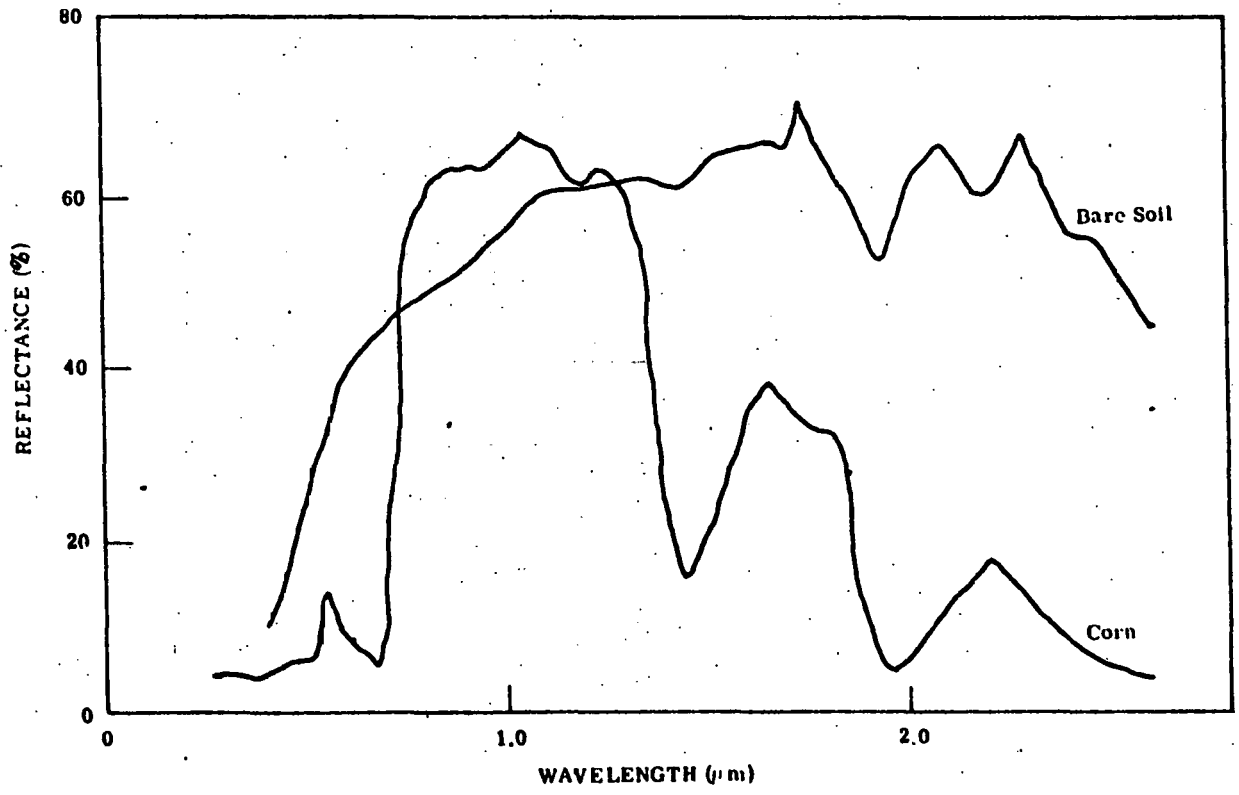


Figure 6. Spectral signatures of bare soil and corn (from Nalepka and Hyde, 1972)

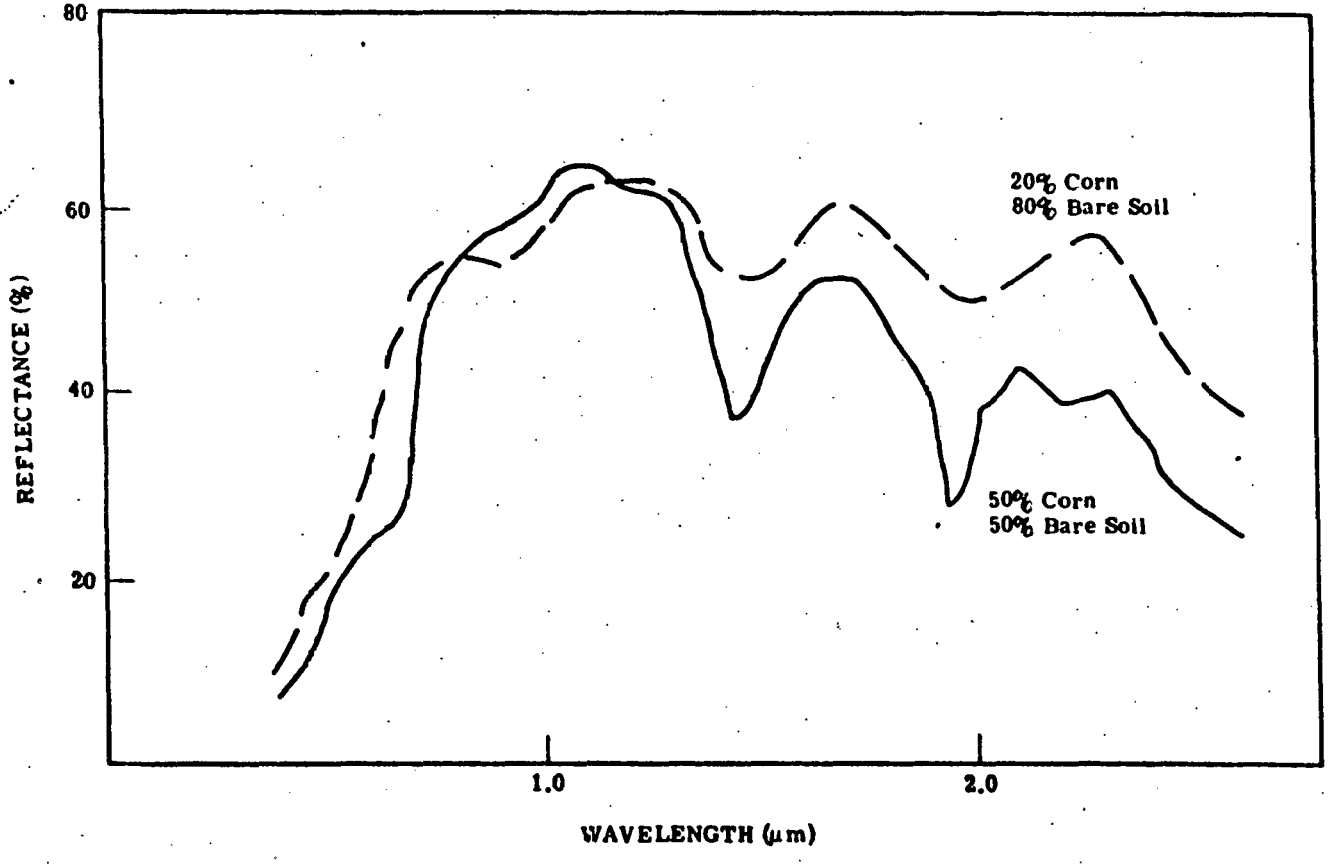


Figure 7. Spectral signatures of mixtures of bare soil and corn (from Nalepka and Hyde, 1972)

Figure 6 would be apt to classify the mixtures as being from some class other than corn or bare soil.

To see how resolution elements containing a mixture of two or more materials can affect classification accuracy in an agricultural application, consider Figure 8. The solid lines delineate "fields" and the dashed lines delineate the resolution elements seen by a scanner. Assuming the size of a resolution element to be 1.1 acres, the fields are approximately 10 acres in size. If the center field contains a different crop than its neighboring fields, it is evident that mixtures of crops will be an important factor in classifying the center field. Only four resolution elements fall entirely within the center field, while twelve elements overlap other fields. Misclassification of the overlapping elements would result in a 55% underestimation of the crop acreage for the center field.

Nalepka and Hyde (1972) have estimated the percentage of square fields that would be seen as a mixture for various field sizes. They took the size of a resolution element to be 300 feet square and assumed that the direction of scanning was parallel to field boundaries.

For small fields of 20 acres or so, they determined the mixture percentage to be around 40%. Even for fields of between 60 and 100 acres, which are large in many areas with mixed agriculture, the probable mixture percentage exceeded 20%. To be able to accurately determine the amount of a crop present in typical agricultural fields, it becomes necessary to have some means of dealing with mixtures. One must be able to estimate the proportion of each crop contained in a mixture element.

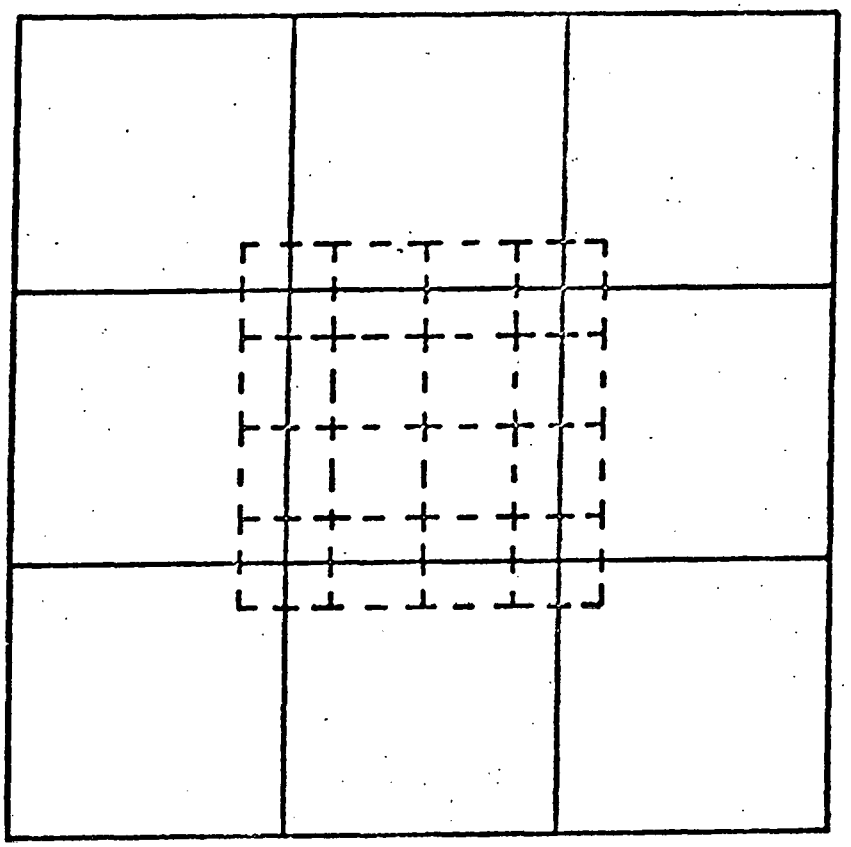


Figure 8. Effect of mixed pixels on classification accuracy (from Nalepka and Hyde, 1972)

Several approaches have been proposed to account for mixtures in classifying remotely sensed data. One model, the ERIM[1] model, will be described here since it forms the basis for the estimation procedures of the next chapter. Other approaches will be mentioned briefly for comparison.

The ERIM model as presented by Horwitz et al. (1971) assumes the spectral responses of the materials of interest follow normal distributions. If there are m classes (materials) and n spectral channels,

[1]Environmental Research Institute of Michigan, Ann Arbor, Michigan.

the ith class is distributed as n-dimensional normal with mean A_i and covariance matrix M_i. The proportion of class i in a given resolution element is denoted by λ_i.

Assume a resolution element consists of N small cells of equal size and let N_i be the number of cells containing the ith material. With the jth of these N_i cells, associate a random variable X_{ij} representing the spectral response of material i from that particular cell. The situation is shown in Figure 9, where the cells associated with class i are taken for convenience to form a contiguous block.

Let X_{ij}, $j = 1, \ldots, N_i$, have mean A_i^* and covariance matrix M_i^* for $i = 1, \ldots, m$. Let Y represent the total response for the resolution element. Then

$$Y = \sum_{i=1}^{m} \sum_{j=1}^{N_i} X_{ij}, \text{ where } \sum_{i=1}^{m} N_i = N.$$

If the entire resolution element were to consist of material i, the mean of Y would be

$$E(Y) = NA_i^* = A_i,$$

and its covariance matrix would be

$$V(Y) = NM_i^* = M_i,$$

assuming independence between the cells of class i. Since there are actually N_i cells of material i, the mean of Y is

$$A_\lambda = E(Y) = \sum_{i=1}^{m} N_i A_i^* = \sum_{i=1}^{m} \lambda_i N A_i^* = \sum_{i=1}^{m} \lambda_i A_i. \qquad (1.5)$$

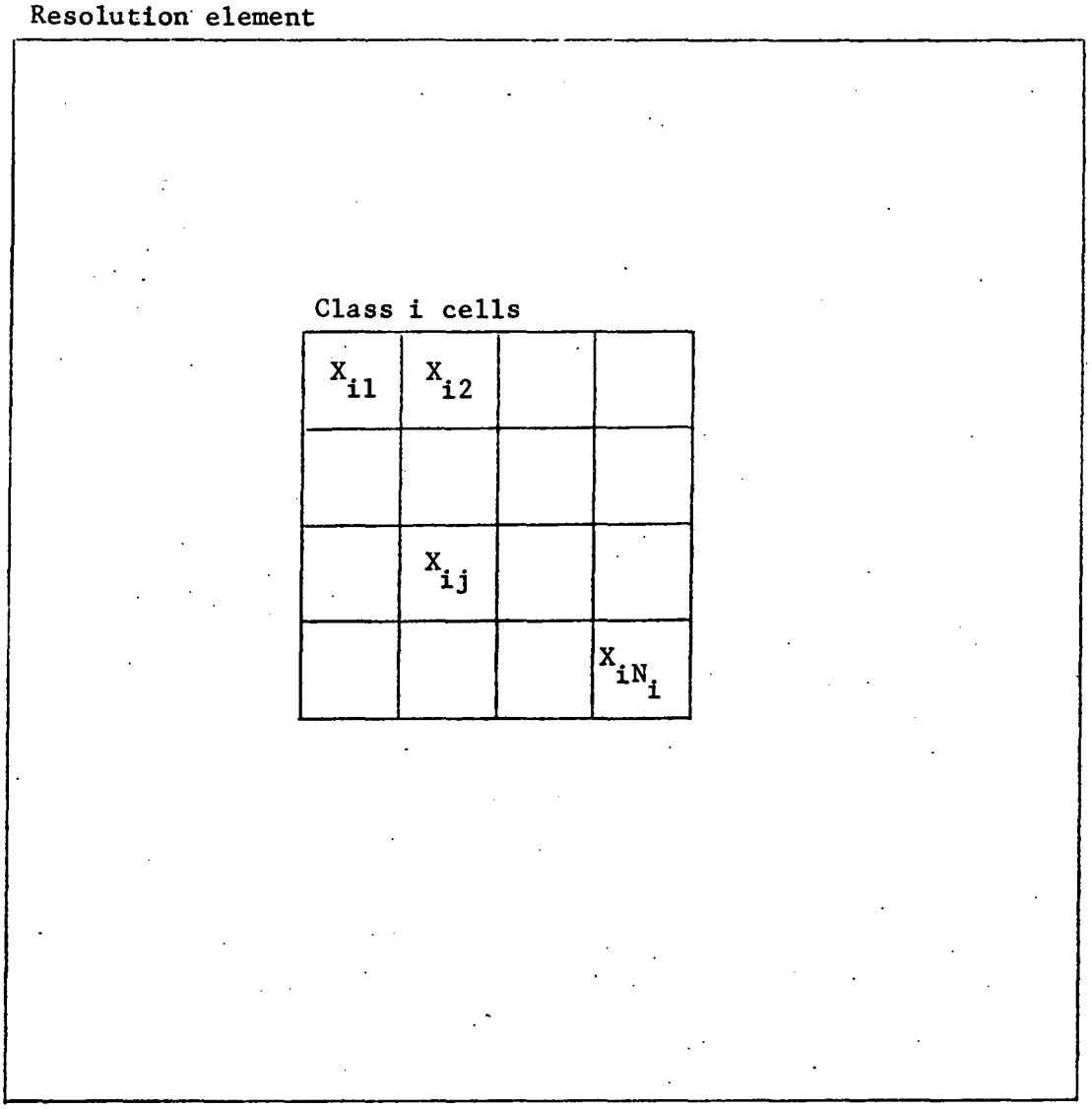


Figure 9. Random variables associated with cells of the ith material in a resolution element

If the random variables associated with cells from different classes are assumed to be independent, the covariance matrix is

$$M_\lambda = V(Y) = \sum_{i=1}^{m} N_i M_i^* = \sum_{i=1}^{m} \lambda_i N M_i^* = \sum_{i=1}^{m} \lambda_i M_i . \qquad (1.6)$$

Thus, the distribution of a given mixture of classes with the associated proportions λ_i is n-dimensional normal with mean A_λ and covariance matrix M_λ.

Given an observation vector y from a mixture distribution, one would like to estimate the true class proportions λ_i. Two methods for obtaining such estimates are presented in the next chapter.

Various researchers have suggested approaches to the mixtures problem that deviate to different degrees from the ERIM approach. Detchmendy and Pace (1972) developed a model for mixtures that is based on different fundamental assumptions from those of the ERIM model. In their formulation the spectral signatures of all pure materials are considered to exhibit no variation; rather, the variations in observed responses are due to variations in the proportions of materials and their backgrounds within a resolution element. Thus, the proportions rather than the class signature vectors are taken to be random variables. Salvato (1973) gives the conditions under which the model of Detchmendy and Pace is mathematically equivalent to the ERIM model.

Smedes et al. (1975) used the ERIM model to generate mixture signatures from the known signatures of each pure material. The proportions associated with the mixtures were specified beforehand in terms of fixed increments. The mixtures corresponding to the generated signatures were treated as additional classes besides the pure classes,

and a standard maximum likelihood classification was carried out.

Odell and Basu (1975) have developed several proportion estimators based on the theory of mixing distributions. Their methods can be utilized to obtain an overall estimate of the proportion of a region covered by a given crop. However, they do not produce estimates for each individual resolution element, and thus do not specifically deal with boundary elements.

D. Outline of Approach

In Chapter II two types of proportion estimators are defined, the standard estimator and the simplified estimator. Examples are presented to illustrate the operation of the two estimation methods and to show how they differ. The computational procedures involved in implementing each type of estimator are discussed, and flowcharts of the programs are provided. An alternative method of computing the simplified estimator based upon a closed-form solution of the least squares problem with interval constraints is given. The two simplified estimation methods are demonstrated to be algebraically equivalent. It is also shown that the closed-form solution is computationally faster than the usual simplified estimation procedure, but its use is subject to more stringent requirements.

The approach used in testing the accuracy of the standard and simplified estimators is given in Section C of Chapter II. Two methods of generating test data are discussed. With one method the proportions of classes in a simulated mixture are fixed in advance.

With the other method the proportions are randomly selected for each data point. The mean square error criterion used to measure the performance of the estimators is defined, and some known theoretical results are given. The results from two tests are presented and discussed, one test involving a specially structured 2-dimensional data set and the other involving LANDSAT-type data with random proportions. The times required to compute the standard and simplified estimates are recorded and tabulated for various numbers of classes and spectral bands.

In Chapter III the concept of data averaging is introduced as a technique for obtaining an estimate of the proportions over an entire region in less time than with the point-by-point estimation methods of Chapter II. The test involving LANDSAT-type data is repeated on the standard and simplified estimators using data averaging, and the results are compared to those of Chapter II.

Chapter IV examines the assumption of equal class covariance matrices and its relationship to proportion estimation. The importance of making the assumption is briefly discussed, and a likelihood-ratio test for equal covariances is applied to data covariance matrices extracted from actual LANDSAT data.

In the third section of Chapter IV, two tests are reported that investigate the effect of assuming equal covariances when they are in fact unequal. The first test uses a specially constructed 2-dimensional data set designed to illustrate the effect of transferring some of the variation from the covariance matrix of one

class to that of another class and of introducing positive and negative correlations. The second test involves two data sets of simulated LANDSAT data. One is constructed using the covariance matrices extracted from LANDSAT data as discussed in Chapter IV, Section B, and the other is constructed using the average of these matrices as the common class covariance matrix. Several proportion estimation procedures are applied to each data set, and the results are compared.

The assumption that the data is distributed as multivariate normal is examined in Chapter V. Some evidence indicating nonnormality of the data is discussed, and the use of normed exponential densities is considered as an alternative to the normal model. A general r-normed exponential density is defined, and a model based on the L_1 norm is presented in detail. The salient properties of the L_1 norm are discussed and illustrated, including an example of how the L_1 norm can out perform L_2 due to the relative insensitivity of the L_1 norm to outliers in the data.

The implementation of a classifier based on the L_1 norm is discussed, and it is shown that the L_1 classifier is computationally more efficient than the corresponding L_2 classifier and leads to an exact evaluation of the probabilities of misclassification, which the L_2 classifier does not. Several sets of simulated data are constructed to test the L_1 classifier. Some of the data sets contain normally distributed data which has been contaminated by a Cauchy or Laplace distribution or the introduction of extreme points. The

results obtained using the L_1 classifier are compared to the results obtained with the L_2 classifier, and the computation times of both are measured.

The final section of Chapter V considers how one might go about applying the L_1 model in dealing with the mixtures problem.

II. METHODS OF PROPORTION ESTIMATION

A. Definition of Methods

1. Standard estimator

In this section the standard proportion estimator is presented as first given by Horwitz et al. (1971). The model used is the ERIM model described in Chapter I.

Let n be the number of bands or channels of the scanner and let m be the number of classes of material. We assume class i follows an n-dimensional multivariate normal distribution with mean A_i and covariance matrix M_i, i = 1, ..., m. Let λ_i be the proportion of class i contained in a mixture of materials and define the proportion vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$.

The mixture associated with λ is then distributed as multivariate normal with mean

$$A_\lambda = \sum_{i=1}^{m} \lambda_i A_i$$

and covariance matrix

$$M_\lambda = \sum_{i=1}^{m} \lambda_i M_i \ .$$

The mixture density function may be written as

$$f(y) = (2\pi)^{-n/2} |M_\lambda|^{-1/2} e^{-1/2(y-A\lambda)' M_\lambda^{-1} (y-A\lambda)} .$$

The log likelihood is then

$$L(\lambda) = \ln(f(y)) = - n/2 \ln(2\pi) - 1/2 \ln|M_\lambda| - 1/2(y - A_\lambda)' M^{-1} (y - A_\lambda).$$

Thus, the constrained maximum likelihood estimate of λ will minimize

$$F(\lambda) = \ln|M_\lambda| + (y - A_\lambda)'M_\lambda^{-1}(y - A_\lambda) \tag{2.1}$$

subject to the constraints

$$\sum_{i=1}^{m} \lambda_i = 1, \quad \lambda_i \geq 0, i = 1, \ldots, m. \tag{2.2}$$

At this point the simplifying assumption is made that all the covariance matrices M_i are equal to a common covariance matrix, say M. This reduces the minimization problem to

$$\underset{\lambda}{\text{minimize}} \ (y - A_\lambda)'M^{-1}(y - A_\lambda)$$

subject to constraints (2.2).

With M positive definite we can perform the Cholesky decomposition

$$M = LL',$$

where L is lower triangular. Taking

$$\begin{aligned} z &= L^{-1}y \\ B_i &= L^{-1}A_i \qquad i = 1, \ldots, m \\ B_\lambda &= L^{-1}A_\lambda \end{aligned} \tag{2.3}$$

the problem becomes one of finding λ that minimizes

$$G(\lambda) = \|z - B\lambda\|^2$$

subject to the constraints (2.2). Under the equal covariance assumption this is completely equivalent to minimizing (2.1) with respect to λ since

$$\|z - B_\lambda\|^2 = (z - B_\lambda)'(z - B_\lambda) = (L^{-1}y - L^{-1}A_\lambda)'(L^{-1}y - L^{-1}A_\lambda)$$

$$= (y - A_\lambda)'L^{-1'}L^{-1}(y - A_\lambda) = (y - A_\lambda)'M^{-1}(y - A_\lambda),$$

and since $\ln|M|$ does not depend on λ, it may be dropped.

Let the λ that minimizes $G(\lambda)$ be denoted $\hat{\lambda}$. This will subsequently be referred to as the standard estimator.

A simple geometric interpretation can be given to the minimization of $G(\lambda)$. Let A be the matrix whose columns are the class mean vectors. Thus,

$$A = (A_1 A_2 \ \cdots \ A_m)$$

and let λ satisfy (2.2). Then the set of all points $A\lambda(= A_\lambda)$ is the convex hull of the A_i and is called the signature simplex. Similarly, the set of points $B\lambda(= B_\lambda)$ is called the transformed signature simplex.

Finding the λ that minimizes $G(\lambda)$ is equivalent to finding the point B_λ on the transformed signature simplex that is closest to the transformed data point z. This is depicted in Figure 10 where z is projected onto the plane determined by the B_i at Pz and B_λ is the orthogonal projection of Pz onto the transformed signature simplex.

An important restriction in order that the optimal λ be uniquely determined in the above formulation is that the number of classes, m, be less than or equal to n + 1, the number of bands plus one. This is implicit in the requirement that the signature simplex have positive (m - 1)-dimensional volume.

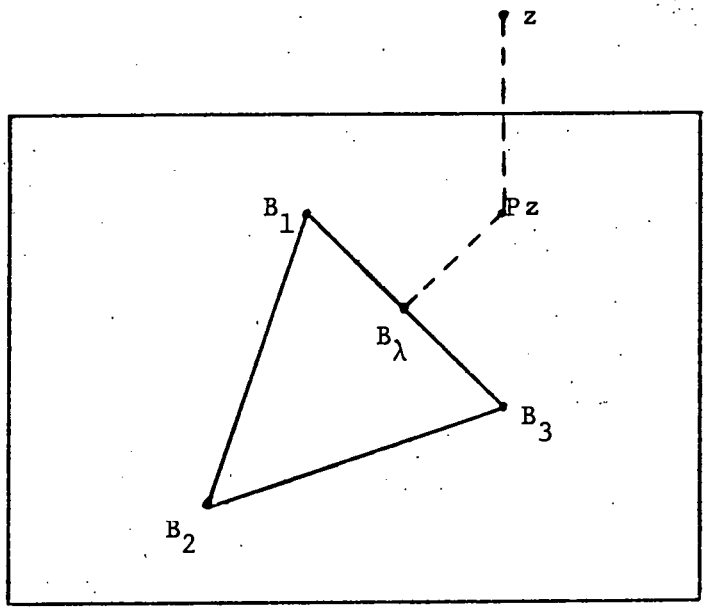


Figure 10. Geometric interpretation of standard estimator

2. Simplified estimator

In order to reduce the amount of computation required to obtain a proportion estimate, a modification of the standard estimator was proposed by Horwitz et al. (1974). The problem is to minimize

$$G(\lambda) = \|z - B_\lambda\|^2$$

subject to

$$\Sigma\lambda_i = 1 \tag{2.4}$$

but not requiring

$$\lambda_i \geq 0 \qquad i = 1, \ldots, m. \tag{2.5}$$

Minimizing $G(\lambda)$ subject only to constraint (2.4) is equivalent to projecting the transformed data vector z onto the hyperplane determined

by the transformed means B_i, i = 1, ..., m. If the projection falls within the transformed signature simplex, then constraint (2.5) will be satisfied, and the estimate will be the same as in the standard case. If the projection falls outside the simplex, however, some of the λ_i will be negative. In this case an estimate is obtained by setting the negative λ_i to zero and normalizing the remaining components.

To precisely specify the estimation procedure, some definitions are needed. Since the covariance-removing transformation given in (2.3) of the previous section is not essential to a basic definition of the estimator, the untransformed means A_i and data vector y will be used in the subsequent discussion.

A <u>proportion vector</u> is a vector λ satisfying (2.4) and (2.5). Let S_A be the signature simplex associated with the mean vectors A_i. Then S_A is the set of all vectors $A\lambda$ where λ is a proportion vector. Let L_A be the set of all vectors $A\eta$ where η satisfies (2.4) but may or may not satisfy (2.5). Clearly S_A is a proper subset of L_A.

Let η^+ be the vector obtained by setting all negative components of η to zero, and let w be the sum of the positive components of η. Then the normalized vector given by

$$\tilde{\eta} = \frac{1}{w}\,\eta^+$$

is a proportion vector.

If Py is the orthogonal projection of y onto L_A, then

$$Py = A\eta \tag{2.6}$$

for some vector η satisfying (2.4). Thus $A\eta$ is the point in L_A closest to y. As an estimator of λ, take

$$\tilde{\lambda} = \tilde{\eta}.$$

The estimator $\tilde{\lambda}$ is called the simplified estimator.

It is evident from the definition of $\tilde{\eta}$ that $A\tilde{\lambda}$ is in S_A, but it may not necessarily correspond to the point $A\hat{\lambda}$ determined by the standard estimator. In the special case where Py falls in S_A, the vector η in (2.6) will itself be a proportion vector, and

$$\tilde{\lambda} = \eta = \hat{\lambda}.$$

To see how the standard and simplified estimators may in general differ, consider the following example in two dimensions where there are three channels and three classes (n = 3 and m = 3).

Let the mean vectors be given in terms of (x_1, x_2) coordinates in the L_A plane by

$$A_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad A_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad A_3 = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

as shown in Figure 11. Let the projection of y onto L_A be

$$Py = A\eta = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Solving for η yields

$$\eta = \begin{pmatrix} 1 \\ -2/3 \\ 2/3 \end{pmatrix}.$$

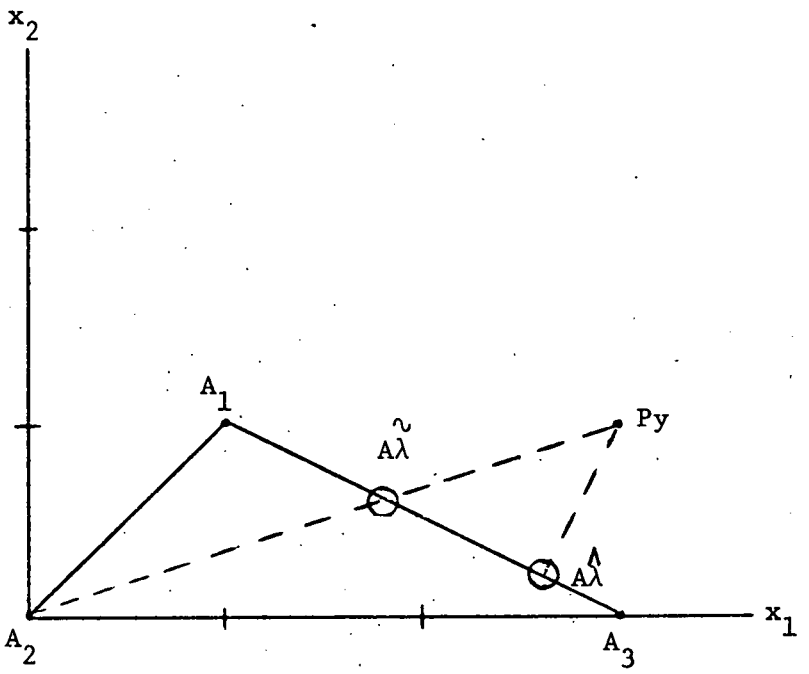


Figure 11. Illustration of difference in standard and simplified estimators

The orthogonal projection of Py onto the line A_1A_3 is

$$\hat{A\lambda} = \begin{pmatrix} 13/5 \\ 1/5 \end{pmatrix}$$

which gives

$$\hat{\lambda} = \begin{pmatrix} 1/5 \\ 0 \\ 4/5 \end{pmatrix}$$

as the value of the standard estimator.

For the simplified estimator one finds the intersection of the lines A_2Py and A_1A_3 to be

$$A\tilde{\lambda} = \begin{pmatrix} 9/5 \\ 3/5 \end{pmatrix}$$

which gives

$$\tilde{\lambda} = \begin{pmatrix} 3/5 \\ 0 \\ 2/5 \end{pmatrix}.$$

Using the definition of the simplified estimator, one has

$$\eta^{+} = \begin{pmatrix} 1 \\ 0 \\ 2/3 \end{pmatrix} \text{and } w = 5/3,$$

so that

$$\tilde{\lambda} = \tilde{\eta} = \frac{1}{w}\eta^{+} = \begin{pmatrix} 3/5 \\ 0 \\ 2/5 \end{pmatrix}.$$

It is not hard to see that in some cases the two estimators will differ considerably in the results they give. In a later section the performance of these estimators will be compared.

B. Computational Procedures

1. Standard estimator

It was shown in Section A of this chapter that under the equal covariance assumption estimating the proportion vector by the standard estimator involves finding the λ that will

$$\underset{\lambda}{\text{minimize}} \; \|z - B_\lambda\|^2$$

subject to

$$\sum_{1}^{m} \lambda_i = 1; \; \lambda_i \geq 0, \qquad i = 1, \ldots, m,$$

where z and B_λ are the transformed observation vector and matrix of means, respectively. With a little manipulation the problem can be re-expressed as a quadratic programming problem. Since

$$\|z - B_\lambda\|^2 = (z - B_\lambda)'(z - B_\lambda) = z'z - 2z'B\lambda + \lambda'B'B\lambda,$$

the expression to be minimized can be reduced to

$$\underset{\lambda}{\text{minimize}} \; [-2z'B\lambda + \lambda'B'B\lambda].$$

In the notation of quadratic programming, the problem becomes

$$\underset{\lambda}{\text{minimize}} \; [p\lambda + \lambda'Q\lambda]$$

such that

$$J'\lambda = 1$$

$$\lambda_i \geq 0, \; i = 1, \ldots, m,$$

where

$$p = -2z'B$$

$$Q = B'B$$

$$J' = (1, 1, \ldots, 1).$$

The program STDEST was written to implement the standard estimator. It employs the quadratic programming package QP360 in solving for the estimate of λ. QP360 was developed by the Rand Corporation and is based on Wolfe's Algorithm, which takes a simplex approach to quadratic programming by utilizing the Kuhn-Tucker conditions. A special feature of QP360 is a parametric programming option which allows one to vary the linear part of the objective function. QP360 is supported at Iowa State by the Numerical Analysis Section of the Statistical Laboratory.

The flow diagram of Figure 12 shows the structure of STDEST. The essential functions performed by the program are:

1. read data vectors from the region of interest and estimate the proportions of materials associated with each vector;
2. save the estimated proportion for mapping the results;
3. print the overall percentage of the region covered by each material.

A few aspects of the program require additional explanation.

The user may supply either a single covariance matrix common to all classes or a separate covariance matrix for each class. If the latter option is chosen, the program will take the average of the matrices as the mixture covariance matrix. The rationale for averaging the covariance matrices is discussed by Horwitz et al. (1971).

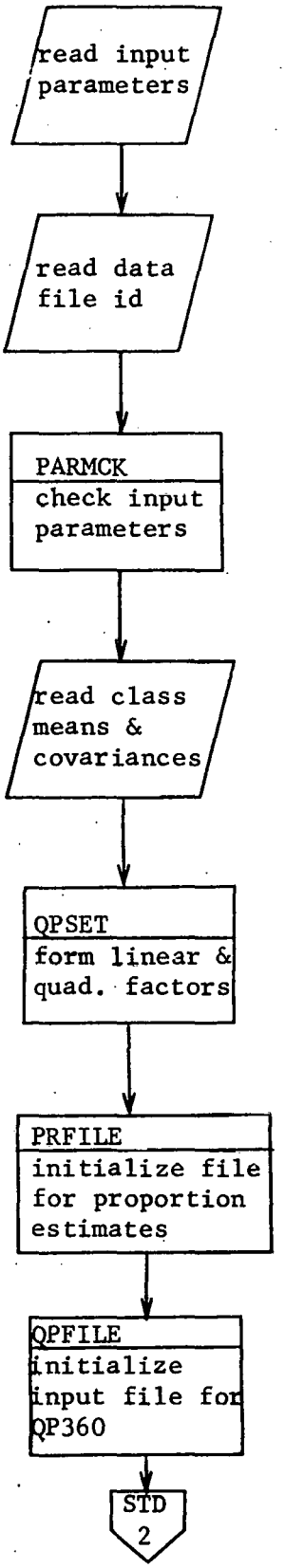
Since the original version of QP360 expected the input to be on cards, a minor modification was necessary to take the input from disk. An example of the input file constructed on disk for QP360 is given in Figure 13. The statements through the MATRIX command are read from

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

Figure 12. Flowchart for STDEST



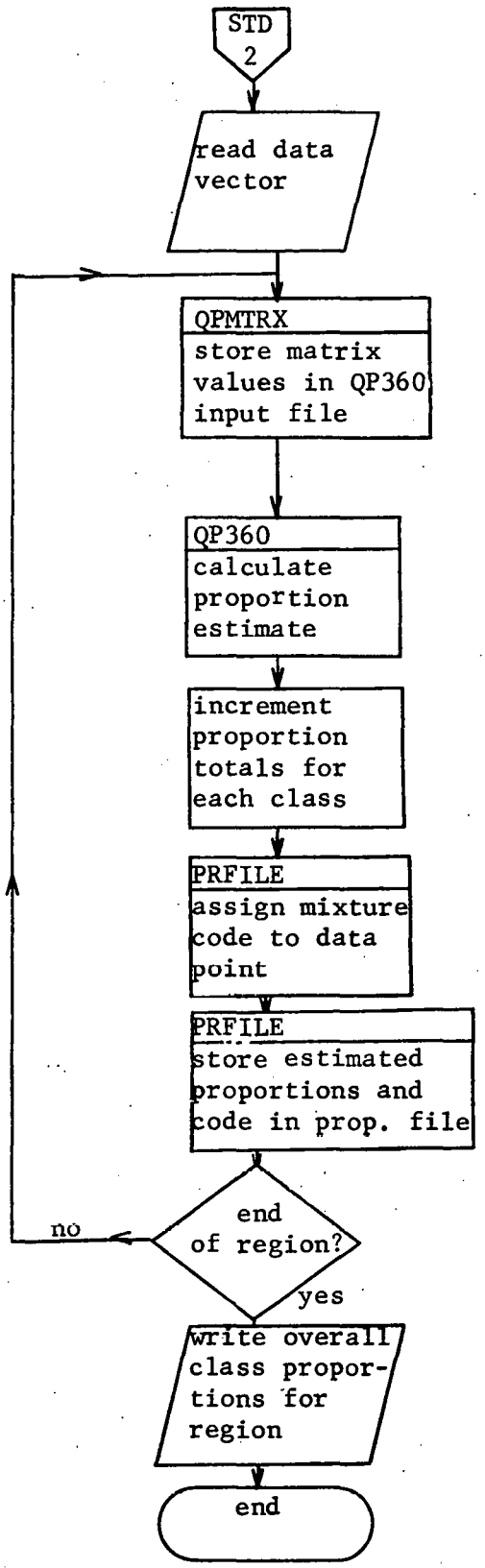
number of classes, bands requested, region coordinates

identification record on input data file

if separate class covariance matrices are supplied, the average is computed; a single matrix is taken to be the common covariance

do Cholesky decomposition of covariance matrix to get linear and quadratic parts of objective function

read partial QP360 input from cards and put on disk (see Figure 13)



place values for linear and quadratic factors after MATRIX statement

read input file from disk and solve QP problem to get estimate

used for estimate of overall proportion of each class in region

based on value of estimate and thresholds read from cards

Figure 12. Continued

```
BEGIN
PRMODE
0022222234
ROWS

                EQUAL*
                & LINEAR

END
RHS

                EQUAL*          1.0

END
MATRIX
        X1      EQUAL*          1.0
        X1      LINEAR          P1 a
        X1      X1              Q1,1 b
        X1      X2              2Q2,1
        X2      EQUAL*          1.0
        X2      LINEAR          P2
        X2      X2              Q2,2

END
SETINV
SOLVE
EXIT
```

[a]P is the vector of coefficients for the linear part of the objective function.

[b]Q is the matrix of coefficients for the quadratic part of the objective function.

Figure 13. Sample input to QP360

cards during initialization and can be varied from run to run if necessary. Since the data in the MATRIX section depend on the value of a particular observation, the input file must be rewritten each time through the loop.

In order to invoke QP360 from STDEST, the main routine in QP360 was made into a subroutine callable from STDEST. When the optimal values for the proportions are calculated in QP360, they are passed back to STDEST via COMMON storage.

STDEST converts the estimated proportions to the nearest percent and stores them in a file in the same format as is used for LANDSAT data. Thus, a vector of percentages, one percentage for each class, is stored corresponding to each data point in the region of interest. This enables the same programs that are used to produce gray-scale maps of LANDSAT data to be used in producing maps of mixture processing results.

At the time the proportions corresponding to a certain data point are stored, a mixture code is calculated and stored to indicate the particular mixture represented by the data point. To determine the code, the proportions are sorted into descending order and compared against a set of cumulative thresholds and a set of minimum thresholds. Thus, if there were m classes, there would be m cumulative and m minimum thresholds.

Proceeding from combinations of one class to combinations of m classes, a data point is taken to be a mixture of k classes ($1 \leq k \leq m$) if the sum of the k largest proportions exceeds the kth cumulative threshold, and each of the k largest proportions exceeds the kth minimum

threshold. The smallest value of k for which both thresholds are satisfied is considered to be the number of classes in the mixture, and a unique code is assigned corresponding to the particular k classes in the mixture.

The cumultative and minimum thresholds are specified as input parameters to the program, and as such reflect a subjective judgment on the part of the user. The code assignment procedure is, after all, not a statistical classification method, but merely a convenient means of labeling the proportion estimation results.

2. Simplified estimator

The heart of the simplified estimator calculation consists of projecting the transformed data vector z onto the hyperplane determined by the transformed means. The problem may be expressed as

$$\underset{\lambda}{\text{minimize}} \ \frac{1}{2} (z - B\lambda)'(z - B\lambda)$$

such that

$$\sum_{i=1}^{m} \lambda_i = 1,$$

where $B = (B_1, B_2, \ldots, B_m)$ is the matrix of transformed means. By introducing the Lagrange multiplier Δ, the objective function may be written

$$\phi(\lambda, \Delta) = \frac{1}{2} (z - B\lambda)'(z - B\lambda) + \Delta(\Sigma\lambda_i - 1).$$

Applying the Kuhn-Tucker conditions for optimality gives

$$\frac{\partial \phi}{\partial \lambda} = B'B\lambda - B'z + \Delta J = 0$$

and

$$\frac{\partial \phi}{\partial \Delta} = \Sigma\lambda_i - 1 = 0,$$

where J is an m x 1 vector of ones. These equations may be rewritten as

$$B'B\lambda + \Delta J = B'z$$

$$\Sigma\lambda_i = 1,$$

which in matrix notation becomes

$$\left[\begin{array}{c|c} B'B & J \\ \hline J' & 0 \end{array}\right] \left[\begin{array}{c} \lambda \\ \hline \Delta \end{array}\right] = \left[\begin{array}{c} B'z \\ \hline 1 \end{array}\right]. \tag{2.7}$$

If the augmented B'B matrix in (2.7) is nonsingular, the solution for λ may be found from

$$\left[\begin{array}{c} \lambda \\ \hline \Delta \end{array}\right] = \left[\begin{array}{c|c} B'B & J \\ \hline J' & 0 \end{array}\right]^{-1} \left[\begin{array}{c} B'z \\ \hline 1 \end{array}\right]. \tag{2.8}$$

The projection z* of z onto the hyperplane of the B_i is thus given by

$$z^* = B\lambda.$$

If z* falls outside the transformed signature simplex, some of the λ_i will be negative. In this case the negative λ_i are set to zero, and the resulting vector is normalized so that its components sum to one.

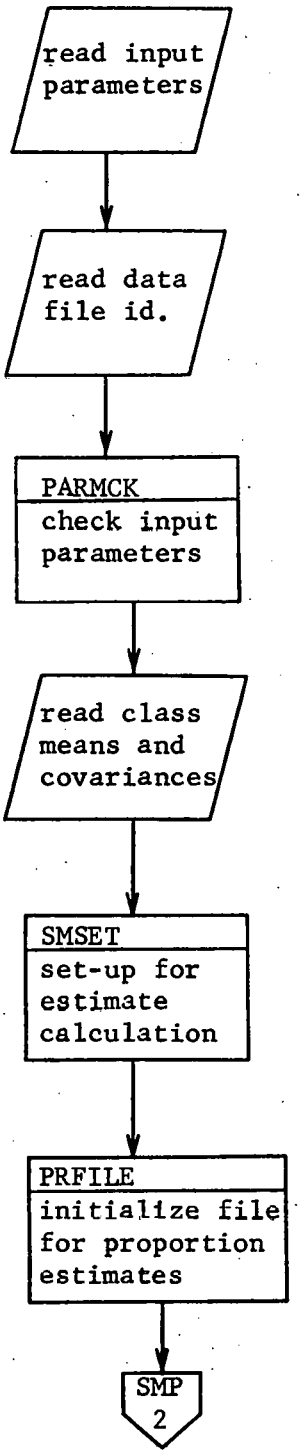
The program SMPEST performs the computations indicated above in calculating the simplified estimate. The augmented B'B matrix is formed, and its inverse is computed and saved. For each data vector in the region of interest, the multiplication indicated in (2.8) is carried out to obtain λ. Each component of λ is tested, any negative component is set to zero, and a flag is turned on to indicate the presence of one or more negative proportions. Subsequently, the flag is examined, and if it is on, λ is normalized by dividing by the sum of the positive components. The estimated proportions are converted to the nearest percent and stored in a file along with a mixture code as described in the discussion of STDEST in the previous section. A flow chart for SMPEST is presented in Figure 14.

It was stated earlier that the signature simplex must be nondegenerate. This assumption is related to nonsingularity of the augmented B'B matrix. Nondegeneracy of the (transformed) signature simplex means that it has positive (m - 1)-dimensional volume, or, equivalently, that the m vectors $\begin{pmatrix} B_i \\ \overline{1} \end{pmatrix}$ are linearly independent.

Augmenting the B_i vectors with ones causes dependency to be considered in mixture space. Vector B_j, $j = 1, 2, \ldots, m$, is in the mixture space defined by the vectors $B_1, \ldots, B_{j-1}, B_{j+1}, \ldots, B_m$ if

$$\sum_{\substack{i=1 \\ i \neq j}}^{m} \lambda_i B_i = B_j \quad \text{, and} \quad \sum_{\substack{i=1 \\ i \neq j}}^{m} \lambda_i = 1.$$

The situation for three classes and two channels is illustrated in Figure 15. Here the signature simplex formed by B_1, B_2 and B_3 is degenerate since it does not have positive 2-dimensional volume. Since



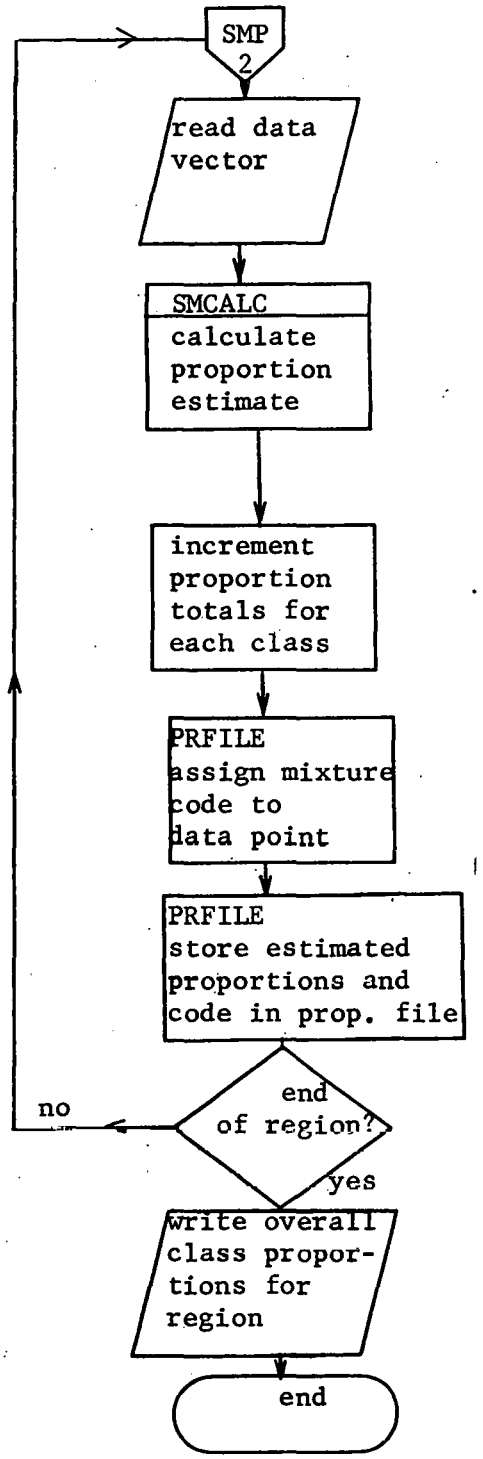
number of classes, bands requested, region coordinates

identification record on input data file

if separate class covariance matrices are supplied, the average is computed; a single matrix is taken to be the common covariance

perform Cholesky decomposition of covariance matrix and transform means

Figure 14. Flowchart for SMPEST



multiply transformed data vector by inverse of transformed means matrix to solve for estimate

used for estimate of overall proportion of each class in region

based on value of estimate and thresholds read from cards

Figure 14. Continued

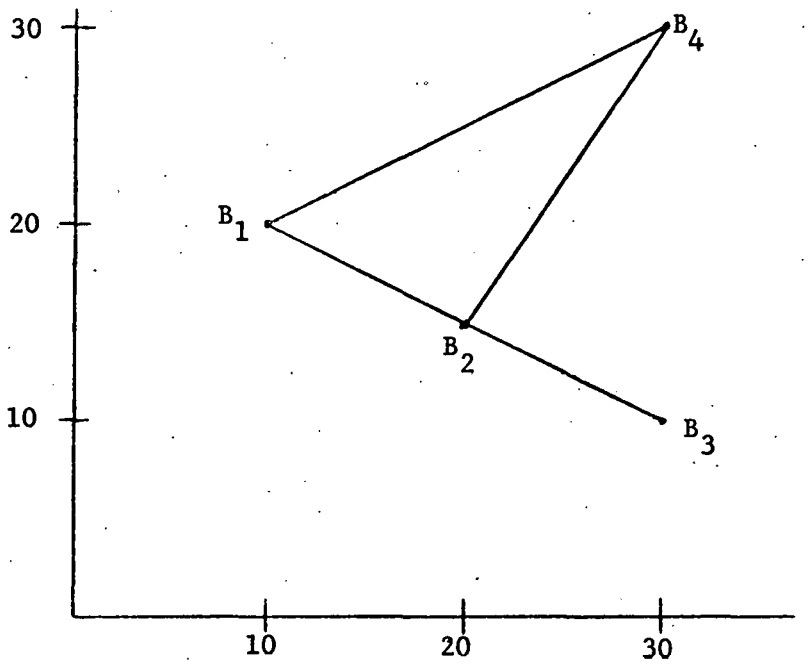


Figure 15. Degenerate and nondegenerate signature simplexes

$$- B_1 + 2B_2 = B_3 ,$$

the vectors are dependent, and B_3 is in the mixture space defined by B_1 and B_2. That the augmented vectors are dependent is clear from

$$-\begin{pmatrix} B_1 \\ -- \\ 1 \end{pmatrix} + 2\begin{pmatrix} B_2 \\ -- \\ 1 \end{pmatrix} = \begin{pmatrix} B_3 \\ -- \\ 1 \end{pmatrix}.$$

On the other hand, the signature simplex formed by B_1, B_2 and B_4 has positive 2-dimensional volume. The vectors are dependent in E^2 since

$$\frac{3}{5} B_1 + \frac{6}{5} B_2 = B_4 ,$$

but

$$\frac{3}{5} + \frac{6}{5} \neq 1,$$

and the linear combination of B_1 and B_2 that yields B_4 is unique, so B_4 is not in the mixture space of B_1 and B_2. Thus, the augmented vectors $\begin{pmatrix} B_1 \\ \hline 1 \end{pmatrix}$, $\begin{pmatrix} B_2 \\ \hline 1 \end{pmatrix}$ and $\begin{pmatrix} B_4 \\ \hline 1 \end{pmatrix}$ are linearly independent.

Now consider the augmented B'B matrices for this example. For B_1, B_2 and B_3 the B matrix is

$$\begin{bmatrix} 10 & 20 & 30 \\ 20 & 15 & 10 \end{bmatrix},$$

and the augmented B'B matrix

$$\begin{bmatrix} 500 & 500 & 500 & 1 \\ 500 & 625 & 750 & 1 \\ 500 & 750 & 1000 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

is singular since two times row 2 minus row 1 equals row 3.

In the case of B_1, B_2 and B_4, the B matrix is

$$\begin{bmatrix} 10 & 20 & 30 \\ 20 & 15 & 30 \end{bmatrix},$$

and the augmented B'B matrix

$$\begin{bmatrix} 500 & 500 & 900 & 1 \\ 500 & 625 & 1050 & 1 \\ 900 & 1050 & 1800 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

is nonsingular since its determinant is - 40,000. Note that since B_1, B_2 and B_4 are dependent in E^2, B'B is singular (three times row 1 plus six times row 2 equals five times row 3) even though the augmented B'B matrix is nonsingular.

Whenever the number of classes, m, exceeds the number of channels, n, the B_i will be linearly dependent since m vectors in an n-dimensional space cannot be independent when $m > n$. The B'B matrix will always be singular in this case.

If, however, there are no more classes than channels, the B'B matrix will be nonsingular as long as the B_i are independent. If this is the case, the simplified estimator may be computed by a closed-form solution to the least squares problem

$$\underset{\lambda}{\text{minimize}}\ (z - B\lambda)'(z - B\lambda) \tag{2.9}$$

such that

$$1 \leq J'\lambda \leq 1,$$

where J is an m x 1 vector of ones.

Klemm and Sposito (1977) have shown that for a least squares problem over an interval constraint, a closed-form solution exists. For the problem stated in (2.9), the solution is

$$\lambda^* = \hat{\lambda} + (B'B)^{-1}J(1 - J'\hat{\lambda})/J'(B'B)^{-1}J,$$

where $\hat{\lambda}$ is the unrestricted least squares estimator, viz.

$$\hat{\lambda} = (B'B)^{-1}B'z.$$

Clearly, the solution requires that the B'B matrix be nonsingular.

It can be shown that λ^* is equivalent to the estimator of λ defined by (2.8) when B'B is nonsingular. Let

$$\begin{bmatrix} B'B & J \\ \hline J' & 0 \end{bmatrix} = \begin{bmatrix} \underline{B}_{n-1} & u_n \\ \hline u'_n & b_{nn} \end{bmatrix} = \underline{B}_n .$$

Applying the formula for the inverse of a matrix by bordering (Faddeeva, 1959, p. 105), yields

$$\underline{B}_n^{-1} = \begin{bmatrix} \underline{B}_{n-1}^{-1} + \dfrac{\underline{B}_{n-1}^{-1} u_n u'_n \underline{B}_{n-1}^{-1}}{\alpha_n} & -\dfrac{\underline{B}_{n-1}^{-1} u_n}{\alpha_n} \\ \hline -\dfrac{u'_n \underline{B}_{n-1}^{-1}}{\alpha_n} & \dfrac{1}{\alpha_n} \end{bmatrix},$$

where

$$\alpha_n = b_{nn} - u'_n \underline{B}_{n-1}^{-1} u_n .$$

Substituting in terms of B and J, we find

$$\underline{B}_n^{-1} = \begin{bmatrix} (B'B)^{-1} - \dfrac{(B'B)^{-1} JJ' (B'B)^{-1}}{J'(B'B)^{-1}J} & \dfrac{(B'B)^{-1}J}{J'(B'B)^{-1}J} \\ \hline \dfrac{J'(B'B)^{-1}}{J'(B'B)^{-1}J} & -\dfrac{1}{J'(B'B)^{-1}J} \end{bmatrix}.$$

Using $\underline{B}_n^{-1}$ to solve for λ by carrying out the multiplication indicated in (2.8) gives

$$\tilde{\lambda} = (B'B)^{-1}B'z - \frac{(B'B)^{-1}JJ'(B'B)^{-1}B'z}{J'(B'B)^{-1}J} + \frac{(B'B)^{-1}J}{J'(B'B)^{-1}J}$$

$$= \hat{\lambda} - \frac{(B'B)^{-1}JJ'\hat{\lambda}}{J'(B'B)^{-1}J} + \frac{(B'B)^{-1}J}{J'(B'B)^{-1}J}$$

$$= \hat{\lambda} + \frac{(B'B)^{-1}J(1 - J'\hat{\lambda})}{J'(B'B)^{-1}J}$$

$$= \lambda^*.$$

Therefore, the closed-form solution yields the same estimate as the usual method for computing the simplified estimate in the case where B'B is nonsingular.

Since the closed-form solution is in effect taking advantage of the fact that the last row and column of the augmented B'B matrix are simply ones with a zero as the last element, the amount of calculation involved should be less than for a simplified estimation procedure which inverts the augmented B'B matrix without any shortcuts. To check whether the closed-form method is actually faster, two versions of SMPEST were compiled. The first used the usual estimation procedure, and the second utilized the closed-form solution. Both programs were run with the same data sets, and the time spent in calculating the estimates was recorded. To avoid a singular B'B matrix, all data sets consisted of data with as many bands as there were classes of material. Table 1 shows the results of the runs.

It can be seen from Table 1 that for two classes the usual solution requires one-third more time to calculate estimates than the closed-form method. The difference becomes less for three and four classes, but the closed-form method maintains an advantage. The

Table 1. Comparison of time to calculate estimates for usual and closed-form methods of simplified estimation

Data set		Time in seconds[a]	
Classes	Points	Usual	Closed form
2	100	0.032	0.024
2	100	0.034	0.025
3	100	0.043	0.042
4	100	0.060	0.053

[a]All timings were taken on an IBM 360/65.

implication is that if one knew a priori that the number of classes appearing within a single resolution element would never exceed the number of data channels, the closed-form method would be a desirable alternative to the usual simplified procedure.

C. Testing the Estimators

1. Generation of test data

To test and evaluate the different methods of proportion estimation, it was necessary to acquire a set of mixture data for which the true proportions were known. Since it is not possible to determine the precise position of the field of view of the scanner with respect to a fixed location on the ground, such as the boundary between two fields, one cannot determine the true proportions in a mixture when using actual satellite data. For this reason simulated mixture data was used for testing.

The subroutine MIXGEN was written to generate random observations from a mixture distribution which is n-dimensional normal with mean A_λ and covariance matrix M_λ. The program utilizes a random number generator from the IMSL[1] and produces random vectors by the following process.

Step 1. Perform a Cholesky decomposition on M_λ to get the matrix L such that

$$M_\lambda = LL'.$$

Step 2. Generate n independent univariate random variates x_i such that

$$X_i \sim N(0, 1), \quad i = 1, \ldots, n.$$

Then $$X = (X_1 X_2 \ldots X_n)' \sim N_n(0, I_n).$$

Step 3. Form $X^* = LX$. Then

$$X^* \sim N_n(0, M_\lambda).$$

Step 4. Take $Y = X^* + A_\lambda$.

Then $$Y \sim N_n(A_\lambda, M_\lambda).$$

Given a proportion vector λ and the mean vectors A_i and covariance matrices M_i, $i = 1, \ldots, m$, MIXGEN forms the mixture mean according to Equation (1.5) and, depending on a program option, either uses (1.6) to compute the mixture covariance or takes the average of the M_i as the mixture covariance. The specified number of random observations are then generated and stored in a file along with the true

---

[1]International Mathematical Statistical Libraries, Inc.

proportion vector λ associated with each observation.

To direct the construction of a file of simulated data, the program SIMDAT was written. SIMDAT reads several parameters that describe how the data file is to be built and calls MIXGEN to create the random observations. Two basic modes of operation are possible. In one case fixed proportions are used to generate the data, and in the other case the proportions are randomly chosen.

In the case of fixed proportions, the user specifies the proportion of each class in a mixture and the number of observations to be generated for that mixture. Several mixtures may be designated as belonging to the same group of mixtures, and the group may be generated repeatedly to produce data in a pattern that resembles physical fields of different materials.

In the case of random proportions, the user supplies certain probabilities, and the program randomly chooses classes and assigns proportions in accordance with the given probabilities. For the purpose of comparing results, the method of randomly generating the data was taken to be that used by Horwitz et al. (1974).

In considering the generation of mixtures consisting of classes of interest to the user, called user classes, and classes of material that are either unknown or of no interest to the user, called alien classes, it becomes necessary to extend the basic mixture model. Let ξ be the proportion of alien material in a pixel, and let the number of user and alien classes be u and v, respectively. Then the mean and covariance matrix of a mixture of user and alien materials are given by

$$A_\lambda = (1 - \xi) \sum_{i=1}^{u} \lambda_i A_i + \xi \sum_{i=1}^{v} \lambda_{u+i} A_{u+i}$$

$$M_\lambda = (1 - \xi) \sum_{i=1}^{u} \lambda_i M_i + \xi \sum_{i=1}^{v} \lambda_{u+i} M_{u+i} \qquad (2.10)$$

Generation of the random data proceeds through the following series of steps:

1. Select at random the proportion of alien material.

2. Randomly choose the number of user and the number of alien classes.

3. From the set of all user and alien materials pick a random subset of user materials and a random subset of alien materials according to the number of classes specified in step 2.

4. Randomly generate the proportions associated with each user class and each alien class.

5. Form the mixture mean and covariance matrix based on (2.10).

6. Generate a single random observation from the multivariate normal distribution with mean A_λ and covariance matrix M_λ, and store it along with the proportions from steps 1 and 4.

7. Repeat steps 1-6 until the required number of observations have been generated.

The proportion of alien material is selected according to the distribution

$$F(x) = \begin{cases} 0, & x < 0; \\ \alpha + (1 - \alpha - \beta)\dfrac{1 - e^{-\gamma x}}{1 - e^{-\gamma}}, & 0 \leq x < 1; \\ 1, & x \geq 1. \end{cases} \qquad (2.11)$$

Here α is the probability that a pixel contains only user material, $P(\xi = 0)$; β is the probability of only alien material, $P(\xi = 1)$; and γ is an additional parameter, which must be different from 0.

A random number from a uniform [0, 1] distribution is generated as the value of $F(x)$. If $F(x) \leq \alpha$, then the value of ξ is taken to be 0. If $F(x) \geq 1 - \beta$, then ξ is assigned the value 1. In all other cases $\alpha < F(x) < 1 - \beta$, and the middle equality of (2.11) is solved for x to yield the proportion of alien material.

The probabilities of choosing various numbers of user and alien classes are based on a consideration of the physical configuration of fields in the scene being viewed. Let τ be the ratio of the length of the edge of a resolution element on the ground to the length of the side of a "typical" field. Assuming τ to be less than 1, the probabilities ρ_i of various numbers of classes are determined from

$$\rho_1(\tau) = (1 - \tau)^2$$

$$\rho_2(\tau) = 2\tau - 2.5\tau^2$$

$$\rho_3(\tau) = \tau^2$$

$$\rho_4(\tau) = 0.5\tau^2$$

$$\rho_5(\tau) = 0.25\tau^2,$$

where 5 has been taken to be the maximum number of user or alien classes expected to be present in a single pixel. A different value of τ may be used for alien classes than is used for user classes if desired. SIMDAT will automatically normalize the probabilities if they do not already sum to 1.

To randomly choose the number of classes, a uniform [0, 1] random number, x, is generated. If $x \leq \rho_1$, the number of classes is taken to be 1; if $\rho_1 < x \leq \rho_1 + \rho_2$, the number of classes is two, etc.

Once the numbers of user and alien classes are selected, subsets of classes of each type must be picked at random. If i user (alien) classes are to be chosen, and the total number of user (alien) classes is m, then the number of subsets, S, to choose from is given by $S = \binom{m}{i}$, where $(\cdot)$ indicates the combinatorial operator. A random integer k between 1 and S is generated to designate a particular subset. The value of k is then used to determine a set of integers indicating the user (alien) classes included in the subset. These are saved in a vector for later reference.

Finally, the proportions associated with the u user and v alien classes selected are determined by generating u + v uniform [0, 1] random numbers. The first u numbers are normalized to sum to 1 and taken as the user class proportions. Similarly, the last v numbers are normalized to become the alien class proportions. All are stored in a proportion vector along with ξ for calculating the mixture mean and covariance matrix.

2. Performance criterion

Suppose that scanner responses have been recorded from N resolution elements in the region depicted in Figure 16. Associated with the ith resolution element is a response vector y_i and a vector of true proportions λ_i. The overall true proportion vector for the entire region is

$$\bar{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i .$$

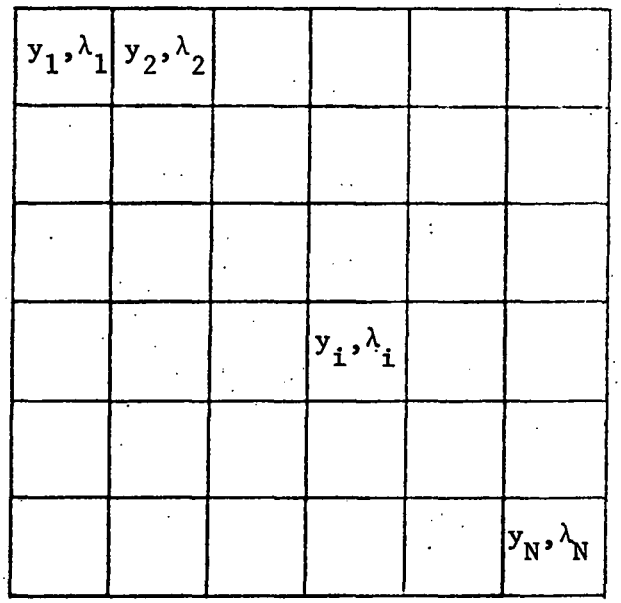


Figure 16. Layout of data for hypothetical region

Let the standard proportion estimator used to estimate λ_i be denoted by $\hat{\lambda}_i$. An estimate of the overall proportion $\bar{\lambda}$ may be obtained by computing estimates $\hat{\lambda}_i$ for each resolution element and taking their overall average to yield

$$\hat{\bar{\lambda}} = \frac{1}{N} \sum_{i=1}^{N} \hat{\lambda}_i .$$

$\bar{\hat{\lambda}}$ is known as the point-by-point standard estimator, and it gives an estimate of the proportion of the region covered by each class of material.

The basis chosen for evaluating the accuracy of a proportion estimator is its mean square error. For the standard estimator $\bar{\hat{\lambda}}$, the mean square error is

$$MSE(\bar{\hat{\lambda}}) = E||\bar{\hat{\lambda}} - \bar{\lambda}||^2 ,$$

where E denotes the expectation operator, and $||\cdot||$ represents the Euclidean norm. The bias associated with the estimator $\hat{\lambda}_i$ is

$$b_i = E(\hat{\lambda}_i) - \lambda_i ,$$

and the average bias over the entire region is

$$\bar{b} = \frac{1}{N} \sum_{i=1}^{N} b_i .$$

Horwitz et al. (1974) have shown that

$$||\bar{b}||^2 \leq MSE(\bar{\hat{\lambda}}) \leq \frac{2}{N} + ||\bar{b}||^2 ,$$

which suggests that for large regions, the mean square error of the estimator should approach the squared norm of the average bias. Unless $\bar{b}$ goes to zero as N goes to infinity, the mean square error will not tend to zero with increasing N.

In the special case where the true proportions are equal throughout the region one has

$$\lambda_i = \lambda;\ b_i = b,\ i = 1, \ldots, N.$$

It follows that each of the $\hat{\lambda}_i$ are identically distributed random variables. One also has that

$$\bar{\lambda} = \lambda;\ \bar{b} = b.$$

It is not hard to show that

$$E\|\bar{\hat{\lambda}} - \lambda\|^2 = \frac{1}{N} E\|\hat{\lambda}_i - \lambda\|^2 + \frac{N-1}{N} \|b\|^2 .$$

When N = 1, that is, when the region is just a single resolution element, this reduces to

$$E\|\bar{\hat{\lambda}} - \lambda\|^2 = E\|\hat{\lambda}_i - \lambda\|^2 ,$$

which implies that, in this case, the mean square error of $\bar{\hat{\lambda}}$ may be estimated by estimating the mean square error of $\hat{\lambda}_i$, the estimator associated with an individual resolution element.

For the simplified estimator, the development concerning mean square error is completely analogous to that presented for the standard estimator. Let $\tilde{\lambda}_i$ denote the simplified estimator for λ_i. Then

$$\bar{\tilde{\lambda}} = \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_i$$

is the point-by-point simplified estimator for the overall proportion $\bar{\lambda}$. Its mean square error is

$$MSE(\bar{\tilde{\lambda}}) = E\|\bar{\tilde{\lambda}} - \bar{\lambda}\|^2 ,$$

which has been shown to obey the bounds

$$\|\bar{\tilde{b}}\|^2 \leq \mathrm{MSE}(\bar{\tilde{\lambda}}) \leq \frac{2}{N} + \|\bar{\tilde{b}}\|^2 ,$$

where

$$\bar{\tilde{b}} = \frac{1}{N} \sum_{i=1}^{N} \tilde{b}_i$$

and

$$\tilde{b}_i = E(\tilde{\lambda}_i) - \lambda_i, \; i = 1, \ldots, N.$$

Also, in the case of equal proportions throughout the region one has

$$\mathrm{MSE}(\bar{\tilde{\lambda}}) = \frac{1}{N} \mathrm{MSE}(\tilde{\lambda}_i) + \frac{N - 1}{N} \|\tilde{b}\|^2 ,$$

where

$$\tilde{b} = \tilde{b}_i; \; \lambda = \lambda_i, \; i = 1, \ldots, N.$$

3. Results

In order to obtain some idea of how much time the standard and simplified methods take to compute proportion estimates, a timing routine was inserted into the STDEST and SMPEST programs. The timing routine measured actual elapsed CPU time and was positioned in the programs so as to measure only the time spent in the subroutines that calculate the estimates for each data point. Thus, differences in the time spent in initialization routines and in writing the estimates to disk were excluded from the results. The version of SMPEST used in this study and in the other investigations to be reported in this section employed the usual, rather than the closed-form, method of

simplified estimation. Table 2 shows the results of the timing study. It is evident from the table that the simplified method was approximately three orders of magnitude faster in most cases, but the advantage of the simplified method decreased with increasing numbers of classes.

Table 2. Proportion estimation times for standard and simplified methods

Data set		Time in seconds[a]	
Classes	Points	Standard	Simplified
2	100	44.05	0.032
2	100	39.96	0.034
3	100	41.58	0.038
3	100	46.18	0.043
4	100	48.53	0.060

[a]All timings were done on an IBM 360/65.

It must be pointed out that neither program was specifically optimized to minimize the computation time. In the case of STDEST the use of the powerful, general-purpose quadratic programming package QP360 resulted in longer execution times because more computing was being performed than was strictly necessary to solve the particular quadratic programming problem. One would expect that a quadratic programming routine specifically tailored to the problem would execute in substantially less time. This was in fact the case in the study reported in Horwitz et al. (1974), but such a program to compute the standard estimator was not written for the present investigation because the primary purpose was to examine the accuracy of proportion estimation methods rather than to compare the speed of optimized implementations.

For the first phase in comparing the accuracy of the standard and simplified estimators, a special data set was designed to illustrate some of the differences in the way the two estimators perform. The data was 2-dimensional, and three classes of user material and three of alien material were present. The arrangement of user and alien class means is shown in Figure 17, where the u's indicate user means, and the A's indicate alien means. All class covariance matrices were taken to be the diagonal matrix diag (20 20).

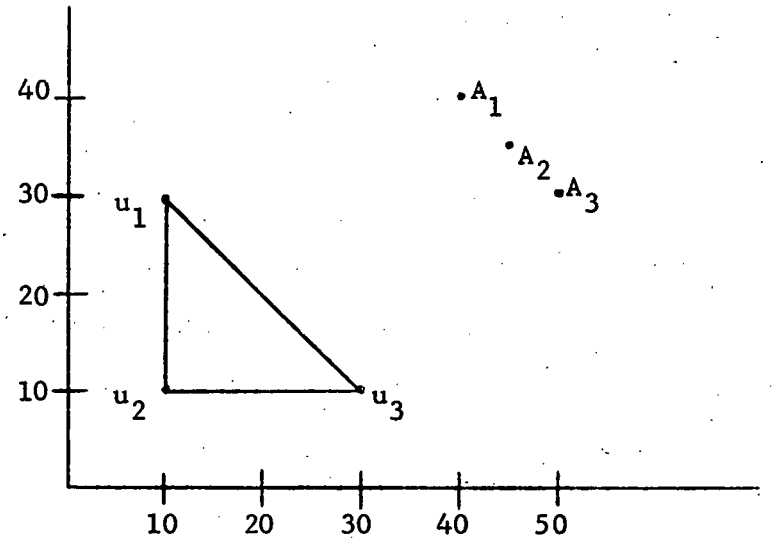


Figure 17. Class means of 2-dimensional test data

Three combinations of user classes were used in constructing mixtures: 50% class 1 - 50% class 3, 25% class 1 - 75% class 3, 100% class 3. For each combination of user classes, five mixtures were generated corresponding to different proportions of alien material. In Figure 18 the locations of the means of the three combinations of user classes are indicated by the letters a, b, and c. The numerals

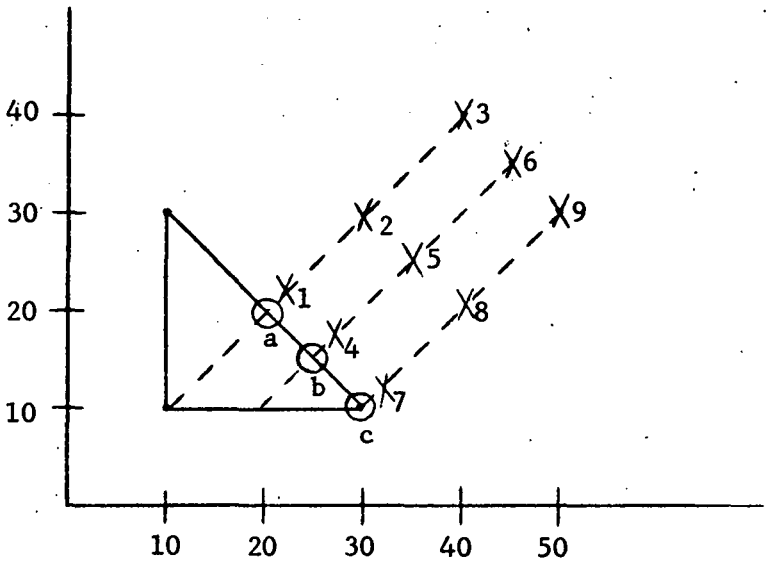


Figure 18. Location of mixture means for 2-dimensional test data

1-9 indicate the locations of the means of the mixtures formed when 10%, 50%, or 100% of each alien material is combined with the user classes. For example, b5 designates the mixture consisting of 50% alien class 2 and 50% user material, where the user material is the combination 25% class 1 - 75% class 3.

For each of the fifteen mixtures, 100 data points were generated. Thinking of the set of data points for each mixture as a separate "region," it must be noted that the true proportions associated with each point were not random. Rather, they were fixed and remained constant throughout the region. This was then an instance of the special case discussed in Section II.C.2. where all the λ_i and $\bar{\lambda}$ are equal to the common true proportion vector λ.

To estimate the mean square error of the standard estimator associated with an individual data point, $E||\hat{\lambda}_i - \lambda||^2$, the quantity

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{m} (\hat{\lambda}_{ij} - \lambda_j)^2$$

was calculated, where N is the number of points in the region, and j is summed over the classes. Similarly,

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{m} (\tilde{\lambda}_{ij} - \lambda_j)^2$$

was calculated as an estimate of the mean square error of the simplified estimator. Thus, the mean square error being estimated was not that of the estimator $\bar{\hat{\lambda}}$ (or $\bar{\tilde{\lambda}}$) formed by averaging the proportion estimates over the region, but rather the mean square error of $\hat{\lambda}_i(\tilde{\lambda}_i)$, the estimator of the proportion vector associated with an individual data point.

The results of estimating the proportions for the fifteen mixtures by using the standard and simplified estimators are given in Table 3. By choosing a well-structured pattern of class means and fixing the proportions, it was possible to build biases into the data in favor of one estimator over the other in certain cases. In all these cases except b4, the favored estimator showed a lower mean square error. Sometimes the differences were considerable as in case a6 when the mean square error for the standard estimator was eight times that for the simplified estimator and in case c9 when the error for the simplified estimator was eleven times that for the standard estimator.

Two other observations are interesting to note concerning the results. First, in cases a1-a3, there was a decided increase in the advantage enjoyed by the simplified estimator. It appears that the simplified method tended to compensate for variations in the randomly

Table 3. Estimated proportions and mean square errors of standard and simplified estimators for 2-dimensional test data

Mixture	Estimated proportions						Mean square error		
	Standard			Simplified			Standard	Simplified	Std/smp
a1	0.473	0.076	0.451	0.470	0.077	0.453	0.0945	0.0845	1.118
a2	0.454	0	0.546	0.475	0	0.525	0.0269	0.0155	1.735
a3	0.499	0	0.501	0.500	0	0.500	0.0471	0.0057	8.263
a6(+)[a]	0.252	0	0.748	0.414	0	0.586	0.1715	0.0213	8.025
a9(+)	0.086	0	0.914	0.343	0	0.657	0.3645	0.0561	6.497
b4(-)[b]	0.229	0.058	0.713	0.268	0.059	0.673	0.0800	0.0712	1.124
b5(-)	0.250	0	0.750	0.364	0	0.636	0.0392	0.0395	0.992
b6(-)	0.276	0	0.724	0.423	0	0.577	0.0457	0.0645	0.709
b3	0.501	0	0.499	0.501	0	0.499	0.1713	0.1314	1.304
b9(+)	0.069	0	0.931	0.328	0	0.672	0.0808	0.0205	3.941
c7(-)	0.053	0.031	0.916	0.100	0.057	0.843	0.0443	0.0821	0.536
c8(-)	0.073	0	0.927	0.253	0	0.747	0.0337	0.1441	0.234
c9(-)	0.056	0	0.944	0.324	0	0.676	0.0191	0.2163	0.088
c3	0.478	0	0.522	0.493	0	0.507	0.5054	0.4916	1.028
c6(-)	0.249	0	0.751	0.411	0	0.589	0.1704	0.3446	0.494
Average							0.1263	0.1193	

[a](+) – Built-in bias toward simplified estimator.

[b](-) – Built-in bias toward standard estimator.

generated data points by always projecting toward the same point u_2 to find the point on the signature simplex closest to the data point.

Secondly, in every case except those biased toward the standard estimator, the simplified estimator performed better. The overall mean square error for all fifteen mixtures was also somewhat less for the simplified estimator. It was not anticipated that the overall set of data would favor one estimator more than the other, but the simplified estimator clearly showed an advantage.

In the second phase of testing, a data file was created to simulate actual LANDSAT data. The classes used and their means and covariances were extracted from LANDSAT data in the course of an earlier study. Table 4 presents these statistics for each of the classes. Two materials, water and concrete, were taken to be alien classes, and the other five classes were designated as user classes. It is noteworthy that the signatures for the alien materials are quite dissimilar from those of the user materials; hence, one would expect the presence of a large amount of alien material in a mixture to significantly distort the response values for points associated with the mixture.

Program SIMDAT was used with randomly chosen proportions to generate a file of 2000 data points consisting of five lines with 400 points per line. The parameters used in randomly generating the observations were as follows:

γ = 1.0 parameter of distribution function of ξ

α = 0.80 probability of only user material

β = 0.05 probability of only alien material

τ = 1/7 ratio of side of resolution element to field edge.

Table 4. Class statistics for simulated LANDSAT data

Class	Mean	Covariance matrix			
Forest	27.99	1.99	1.32	1.24	0.0
	16.88	1.32	2.22	0.82	- 0.37
	61.22	1.24	0.82	13.47	7.16
	37.02	0.0	- 0.37	7.16	6.25
Urban 1	37.60	9.30	11.84	4.88	0.08
	30.25	11.84	19.01	6.79	- 0.34
	53.15	4.88	6.79	17.72	8.94
	27.58	0.08	- 0.34	8.94	6.71
Urban 2	38.38	5.57	6.56	1.52	- 0.64
	31.88	6.56	10.43	2.43	- 0.80
	43.20	1.52	2.43	12.82	6.77
	20.40	- 0.64	- 0.80	6.77	5.06
Agriculture	33.11	2.79	2.59	1.71	0.41
	23.22	2.59	4.41	0.66	- 0.63
	61.49	1.71	0.66	15.52	9.11
	34.56	0.41	- 0.63	9.11	7.40
Bare soil	47.56	19.89	29.82	16.51	4.34
	52.07	29.82	55.20	28.34	7.02
	61.19	16.51	28.34	31.47	12.57
	28.04	4.34	7.02	12.57	7.29
Concrete	64.52	18.58	22.59	10.14	2.89
	67.19	22.59	33.18	14.71	4.69
	67.73	10.14	14.71	11.29	3.83
	29.79	2.89	4.69	3.83	3.28
Water	31.50	3.65	2.47	2.13	0.74
	20.37	2.47	3.72	2.87	1.41
	18.71	2.13	2.87	13.84	7.55
	6.22	0.74	1.41	7.55	5.20

Some characteristics of the data as it was actually generated are shown in Table 5.

This test was designed to investigate how the efficiency of the standard and simplified estimators compared for various sized regions.

Table 5. Summary of simulated LANDSAT data

Average proportion of alien material	0.1165
Average user class proportions	
Forest	0.2027
Urban 1	0.1924
Urban 2	0.1938
Agriculture	0.2111
Bare soil	0.2000
Average alien class proportions	
Concrete	0.4774
Water	0.5225
Number of points with various numbers of user and alien classes present	
All user	1609
All alien	102
1 user	1434
2 user	492
3 user	42
4 user	25
5 user	7
1 alien	1519
2 alien	481

Thus, the mean square error of the point-by-point standard estimator, $MSE(\hat{\bar{\lambda}})$, and the mean square error of the point-by-point simplified estimator, $MSE(\tilde{\bar{\lambda}})$, were the criteria of interest.

Regions of size 1, 10, 50, 200, and 300 were selected in the following manner. Let N be the number of points per region. An initial region was selected by randomly picking a starting point between 1 and 401-N and taking N consecutive points from line 1. A second region was similarly selected from line 2 by picking a random starting point and taking N points. Five regions, one from each line, were selected in all for each different region size. For every region selected, STDEST and SMPEST were run to obtain the estimates $\hat{\bar{\lambda}}$ and $\tilde{\bar{\lambda}}$.

For both estimators an overall estimate of mean square error was derived by averaging the mean square errors over the five independent regions for each size of region.

Figure 19 displays the results of the test, and Table 6 presents the numerical values. It is evident that the mean square error drops rapidly with increasing region size and seems to approach a limiting value of about 0.04 in both cases. The theory discussed in the previous section indicates that this limiting value is the squared norm of the average bias over a region.

Table 6. Mean square error of standard and simplified estimators for simulated LANDSAT data

	Number of points in region				
	1	10	50	200	300
Standard	0.6038	0.0866	0.0363	0.0392	0.0376
Simplified	0.8843	0.1334	0.0572	0.0384	0.0398

4. Discussion

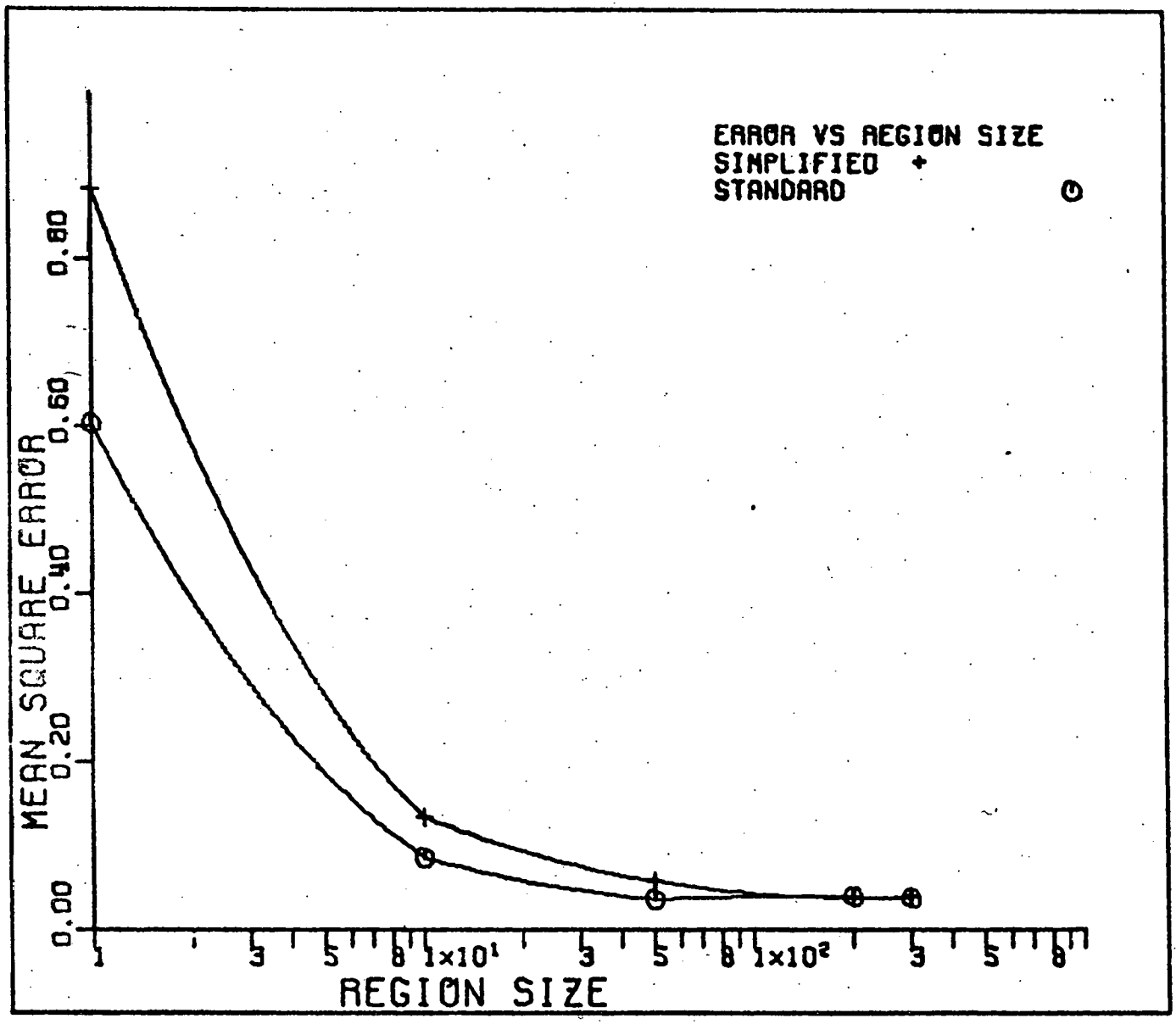
It is difficult to know how much weight to attribute to the results presented in Table 3 because of the artificial nature of the data set. One conclusion that seems to be clear, however, is that the geometric relationship of alien signatures to user signatures in the plane of the signature simplex can have a strong biasing effect on the results. This was especially evident in cases a6 and a9, which favored the simplified estimator, and cases c8 and c9, which favored the standard estimator. Since in some cases this bias gives a decided advantage to

Figure 19. Mean square error versus region size for standard and simplified estimators

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK


ERROR VS REGION SIZE
SIMPLIFIED +
STANDARD ⊙
MEAN SQUARE ERROR
0.00
0.20
0.40
0.60
0.80
1
3
5
8
1×10¹
3
5
8
1×10²
3
5
8
REGION SIZE

the simplified estimator, it is certainly possible that the simplified estimator can outperform the standard estimator under the right conditions.

A second important point about the simplified estimator is the compensating effect of projecting toward a fixed point as noted in the previous section. It appears that this reduces the variance component of the mean square error, which helps to explain why the simplified estimator performs as well as it does when it clearly has a larger bias than the standard estimator. This is illustrated quite well in case b5 of Table 3 where the proportion estimates of the simplified estimator are decidedly inferior (the standard proportion estimates are exactly correct in this case), yet the mean square errors are nearly identical. In choosing between the two estimators, the relative importance of the two components of mean square error, variance and squared bias, should probably be considered.

The results of the second phase of testing when simulated LANDSAT data was used seem to conform well to what one would expect from intuition and a consideration of the theory. In Figure 19 both estimators' mean square errors follow a rather smooth descent to a similar, if not a common limiting value. For regions of size 50 the mean square error of the simplified estimator is 58% greater than that of the standard estimator, and for regions of 100 points or larger, the difference becomes negligible.

It is difficult to judge the relative merits of the estimators because any such evaluation must take cost into account, and the implementation of the standard estimator used for this study did not

lead to reasonable cost figures. The timing results of Table 2 suggest that with the standard estimator one would be paying a 1000-fold penalty in execution time for a 58% gain in efficiency with regions of 50 points using the implementations of the present study. If the timings of the optimized implementations reported by Horwitz et al. (1974) are used instead, one has a 50% cost increase for a 58% efficiency gain in using the standard estimator with regions of size 50. For very small regions the additional cost of the standard estimator may well be worth the higher efficiency, but for large regions of 100 points or more, it seems likely that the slight increase in efficiency of the standard estimator would not be worth the added cost in most cases. There would be some region size in between where the efficiency advantage of the standard estimator would be just offset by the extra computing time required. The precise region size would depend on the relative costs of estimation inaccuracy and computer time.

III. PROPORTION ESTIMATION WITH AVERAGING

A. Data Averaging Procedure

Consider again the region shown in Figure 16. An alternate way to estimate the overall proportion $\bar{\lambda}$ is to first average the N response vectors to obtain

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i ,$$

and then compute the standard estimator based on the single data vector $\bar{y}$. This estimator may be denoted by $\hat{\bar{\lambda}}$ and is referred to as the standard estimator with data averaging as opposed to the point-by-point standard estimator defined in Section II.C.2.

Data averaging is applicable in situations where one wants to know the proportion of an entire area covered by each of several different materials. For instance, one might wish to know the total amount of oats planted in a certain section of a county in Iowa. There is a danger in averaging over too large an area, since the mixture theory employed in the estimation process allows for at most n + 1 materials to be included in a mixture, where n is the number of scanner channels.

Data averaging is not applicable in the case where one needs to estimate the proportions of materials associated with individual resolution elements, as would be the case if one were mapping the boundary of a lake.

Results concerning the mean square error of the standard estimator with data averaging have been produced similar to those mentioned

earlier for the point-by-point estimator. Horwitz et al. (1974) show that

$$MSE(\hat{\bar{\lambda}}) = E\left\|\hat{\bar{\lambda}} - \bar{\lambda}\right\|^2 \leq \frac{\beta\tau}{N} ,$$

where τ is the maximum trace of the covariance matrices M_i, $1 \leq i \leq m$, and β is a constant independent of N. This implies that the mean square error always goes to zero as N becomes large when data averaging is performed.

The simplified estimator with data averaging is defined as the simplified estimator for $\bar{\lambda}$ based on the average data vector $\bar{y}$. Let this estimator be denoted by $\tilde{\bar{\lambda}}$. The following bound has been established for the mean square error of $\tilde{\bar{\lambda}}$.

$$MSE(\tilde{\bar{\lambda}}) = E\left\|\tilde{\bar{\lambda}} - \bar{\lambda}\right\|^2 \leq \frac{m\beta\tau}{N} ,$$

where m is the number of classes, and β and τ are as defined above. Again it can be observed that in the case of data averaging, the mean square of the estimator goes to zero as N goes to infinity.

B. Estimation Results under Averaging

To study the effect of data averaging on proportion estimation, the programs STDEST and SMPEST were modified to include an option which allows for bypassing the computation of proportion estimates until all data points in a region have been read and averaged together. The calculation of proportion estimates and mean square error estimates is carried out as if the entire region consisted of a single

average data point.

The data file of simulation LANDSAT data discussed in Section II.C.3 was processed exactly as before except that data averaging was performed. The results are tabulated in Table 7 and presented in graphic form in Figure 20.

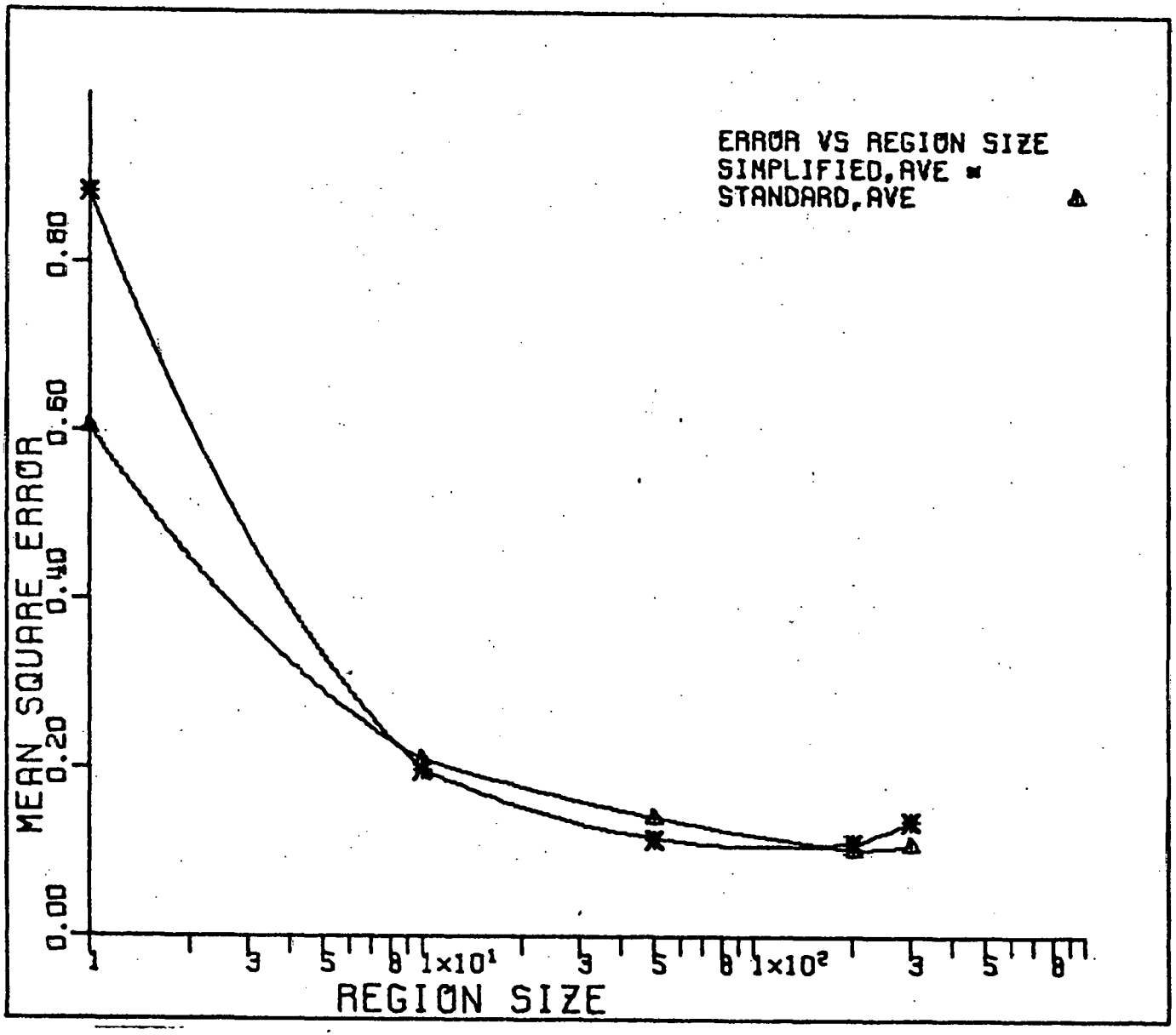
Table 7. Mean square errors for proportion estimation with data averaging

	Number of points in region				
	1	10	50	200	300
Standard, ave.	0.6038	0.2100	0.1419	0.1036	0.1097
Simplified, ave.	0.8843	0.1987	0.1170	0.1127	0.1376

The general shape of the plot in Figure 20 is similar to that obtained without averaging, but it does not follow as nice and regular a path. The theory predicts decreasing mean square error with increasing region size, but the results exhibit a tendency for the decrease in mean square error to tail off at the larger regions.

It is hypothesized that this phenomenon is due to the influence of alien material, which was spectrally quite distinct from the user classes. As more points are included in the region, the true proportion vector will tend toward the vector with all proportions equal, since user classes were selected with equal probability in generating the data. This vector of equal proportions is associated with the point at the center of the signature simplex, but the influence of alien material tends to draw the average data point outside the simplex, resulting in

Figure 20. Mean square error versus region size for standard and simplified estimators with data averaging


ERROR VS REGION SIZE
SIMPLIFIED,AVE *
STANDARD,AVE Δ
MEAN SQUARE ERROR
0.00
0.20
0.40
0.60
0.80
1
3
5
8
1x10¹
3
5
8
1x10²
3
5
8
REGION SIZE

large estimation errors.

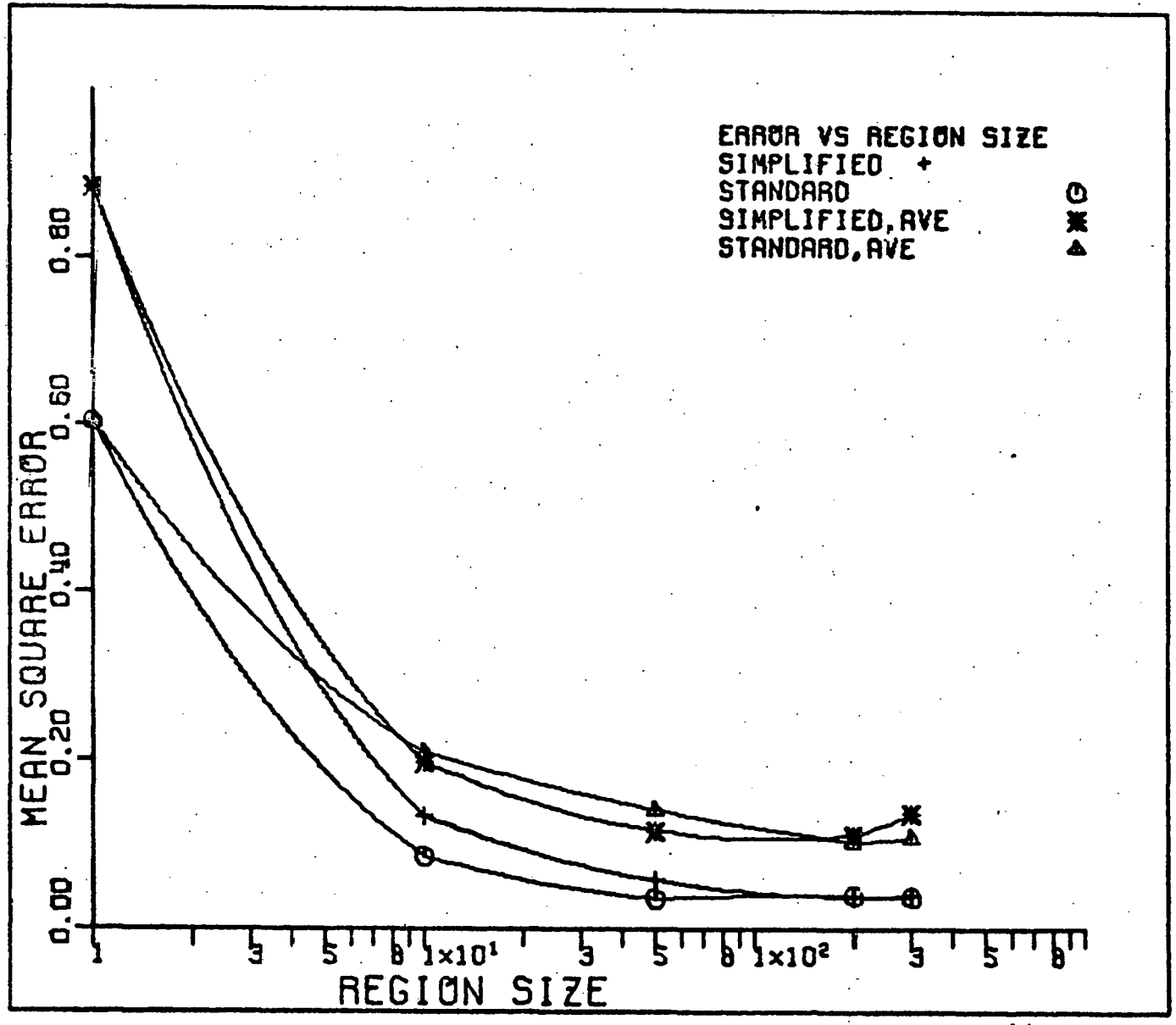
For instance, with regions of 300 points, the mean square error when the average data point happened to fall within the simplex was 1/6 to 1/8 of the mean square error incurred when the average data point was forced outside the simplex due to the influence of alien material. This can be seen in Table 8 where the average data point fell within the simplex for region 5 and outside it in the other cases. Unfortunately, time did not allow retesting the estimators with the inclusion of an alien object test in the programs which would ignore alien points in the data averaging procedure to attempt to verify the present conjecture. It does seem, however, that an alien object test should be included in any operational program employing data averaging.

The combined results for the standard and simplified estimators with and without data averaging are displayed in Figure 21. According to the theory, there should be a point where the curves for estimation with data averaging drop below the corresponding curves for estimation without data averaging. Because of the slight disturbance in the tails of the data averaging curves, it is not possible to predict from the results where the crossover point would occur. It appears that it might be necessary to take very large regions in order to observe the crossover.

Table 8. Effect of alien material on mean square errors under data averaging with regions of 300 points

Region	Standard						Simplified					
	MSE	Proportions					MSE	Proportions				
1	0.1025	0.10	0.40	0.0	0.34	0.16	0.1254	0.07	0.40	0.0	0.38	0.15
2	0.1305	0.07	0.43	0.0	0.32	0.17	0.1728	0.02	0.44	0.0	0.39	0.15
3	0.1804	0.03	0.38	0.0	0.42	0.17	0.2159	0.0	0.39	0.0	0.45	0.15
4	0.1116	0.04	0.42	0.0	0.37	0.17	0.1506	0.0	0.42	0.0	0.43	0.14
5	0.0234	0.09	0.28	0.20	0.25	0.17	0.0235	0.09	0.24	0.20	0.25	0.17
	0.1097						0.1376					

Figure 21. Mean square error versus region size for standard and simplified estimators with and without data averaging


ERROR VS REGION SIZE
SIMPLIFIED +
STANDARD
SIMPLIFIED,AVE
STANDARD,AVE
MEAN SQUARE ERROR
0.00
0.20
0.40
0.60
0.80
1
3
5
8
1x10¹
3
5
8
1x10²
3
5
8
REGION SIZE

IV. EQUAL COVARIANCE ASSUMPTION

A. Importance of Equal Covariances

To understand why the assumption of equal class covariance matrices is crucial in calculating proportion estimates, one needs to review the proportion estimation procedure. The mixture model assumes that observations are taken from a multivariate normal distribution with mean

$$A_\lambda = \sum_{i=1}^{m} \lambda_i A_i$$

and covariance matrix

$$M_\lambda = \sum_{i=1}^{m} \lambda_i M_i .$$

The maximum likelihood procedure for estimating λ leads to minimizing

$$F(\lambda) = \ln|M_\lambda| + (y - A_\lambda)' M_\lambda^{-1} (y - A_\lambda), \tag{4.1}$$

subject to the added constraints

$$\sum_{i=1}^{m} \lambda_i = 1, \ \lambda_i \geq 0, \ i = 1, \ldots, m,$$

which are imposed to insure that the estimate is a proportion vector.

Equation (4.1) expands into

$$F(\lambda) = \ln|\lambda_1 M_1 + \lambda_2 M_2 + \ldots + \lambda_m M_m|$$

$$+ (y - A\lambda)'(\lambda_1 M_1 + \ldots + \lambda_m M_m)^{-1}(y - A\lambda),$$

where A is the matrix of mean vectors. In the general case without equal covariances, the expression for $F(\lambda)$ does not lend itself to any

convenient simplification. No practical computational method has been found to minimize $F(\lambda)$ with respect to λ in this general case.

In the special case where $M_i = M$, $i = 1, \ldots, m$, the first term of (4.1) becomes a constant and may be dropped, and the second term reduces to

$$(y - A\lambda)'M^{-1}(y - A\lambda).$$

Under the transformation (2.3) the problem becomes the quadratic programming problem

$$\underset{\lambda}{\text{minimize}}\ G(\lambda) = ||Z - B_\lambda||^2 \text{ such that } \sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i \geq 0,$$

$$i = 1, \ldots, m,$$

whose solution yields the standard proportion estimate. Thus, under the present mixtures model, the equal covariance assumption is necessary to be able to employ a feasible computational procedure.

B. A Test for Equal Covariances

Since the equal covariance assumption is so vital to obtaining proportion estimates, and since its validity has sometimes been suspect, it was decided to subject the assumption to a statistical test using actual LANDSAT data. The LANDSAT frame chosen was taken on August 26, 1973 over central Iowa. A 12-section site in the scene was selected as the study area. It is an agricultural area of predominantly row crops with no urban centers or forest cover and a negligible amount of surface water.

Four crop types were chosen as the classes of interest: corn, soybeans, oats, and alfalfa. Using ground truth information, fields of each crop were selected from the 12-section site. An attempt was made to pick fields for each crop that were well-scattered throughout the scene, and field center pixels were used as much as possible. Line and column coordinates in the LANDSAT data were determined for each field with the help of a low-altitude aerial photo of the scene. A computer program was then used to read the data values associated with each field and compute the sample means and covariance matrices. These statistics along with the sample sizes are given in Table 9.

The null hypothesis to be tested was that the covariance matrices for all four classes were equal. That is, the hypothesis

$$H_o: \quad \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4$$

was tested against the alternative hypothesis

$$H_A: \quad \Sigma_i \neq \Sigma_j \qquad 1 \leq i \neq j \leq 4.$$

The test used (Morrison, 1967, p. 152) requires that the populations be normally distributed and uses a generalized likelihood-ratio criterion.

Let m and n be the number of classes and the number of dimensions, respectively. Let S_i denote the sample covariance matrix for class i, and let N_i denote the sample size for class i. Then

$$S = \sum_{i=1}^{m} \frac{(N_i - 1)S_i}{\sum_{i=1}^{m} (N_i - 1)}$$

is the pooled sample covariance matrix. Let

Table 9. Means and covariance matrices extracted from LANDSAT data for four crops

Class	Data points	Mean	Covariance matrix			
Corn	167	23.0	0.77	0.23	0.85	0.57
		14.7	0.23	0.70	0.24	0.05
		38.9	0.85	0.24	13.25	7.29
		23.4	0.57	0.05	7.29	5.23
Soybeans	159	23.5	1.07	0.15	1.07	0.22
		13.9	0.15	0.50	0.08	- 0.23
		65.1	1.07	0.08	10.42	2.55
		41.1	0.22	- 0.23	2.55	2.46
Oats	127	26.2	1.57	1.44	0.70	- 0.42
		19.8	1.44	4.72	- 7.70	- 6.08
		40.4	0.70	- 7.70	47.35	29.87
		21.6	- 0.42	- 6.08	29.87	20.44
Alfalfa	85	26.4	4.44	7.11	- 2.50	- 3.13
		18.2	7.11	15.89	- 9.81	- 8.61
		49.8	- 2.50	- 9.81	35.30	22.26
		28.3	- 3.13	- 8.61	22.26	16.42

$$M = \sum_{i=1}^{m} (N_i - 1) \ln|S| - \sum_{i=1}^{m} (N_i - 1) \ln|S_i|$$

and

$$C^{-1} = 1 - \frac{2n^2 + 3n - 1}{6(n + 1)(m - 1)} \left(\sum_{i=1}^{m} \frac{1}{(N_i - 1)} - \frac{1}{\sum_{i=1}^{m} (N_i - 1)} \right).$$

The test statistic MC^{-1} is approximately distributed as chi-squared with $n(n + 1)(m - 1)/2$ degrees of freedom for large (> 20) samples.

A small program was written to calculate the test statistic, and it was applied to the four sample covariance matrices extracted from LANDSAT data. The value of the test statistic came out to be 729.3 with 30 degrees of freedom, which was highly significant at the smallest

α level (0.005) given in the table used for the test. Thus, the hypothesis of equal covariance matrices was firmly rejected.

Looking at the matrices in Table 9, it is evident that the variances associated with oat and alfalfa fields are much larger than those of corn and soybeans. This is probably due in part to the fact that the oat and alfalfa fields tend to be smaller and more irregular in shape, making it more difficult to obtain pixels that are uncontaminated by other materials. Since the corn and soybean fields did not incur this difficulty, it was decided to apply the test to only these classes to see if the large variances for oats and alfalfa were the cause of rejecting H_o. This time the null hypothesis

$$H_o: \quad \Sigma_1 = \Sigma_2$$

was tested against

$$H_A: \quad \Sigma_1 \neq \Sigma_2.$$

The test statistic was $MC^{-1} = 81.1$ with 10 degrees of freedom, which was significant at the 0.005 level, so the equal covariance hypothesis was again rejected. It would appear that even for similar types of material (in this case, two agricultural crops) sampled from relatively large, homogeneous areas, the class covariance matrices are in fact statistically dissimilar.

C. Effect of Unequal Covariances

The results of the previous section probably come as no surprise to anyone who has worked much with LANDSAT data. In actuality covariance matrices for different classes are simply not equal. The key question then becomes whether or not it makes any difference that the covariances are unequal when one computes proportion estimates as though the matrices were all equal to the average covariance matrix.

In the first phase of examining this question, an artificial data set was constructed similar to the one described in Section II.C.3. There were three user classes, three alien classes, and two bands as depicted in Figure 22. The combinations of user classes employed were: 80% user class 1 - 20% user class 3, 50% user class 1 - 50% user class 3, and 20% user class 1 - 80% user class 3. For each combination of user classes, four different mixtures were created by adding: no alien material, 50% alien class 1, 50% alien class 2, and 50% alien class 3. The small letters in Figure 22 indicate the locations of the means of the various combinations of user classes, and the U_i and A_i indicate the locations of the means of the user and alien classes, respectively. A certain mixture will be denoted by the small letter for the user combination and the number of the alien class. For example, a3 denotes the combination of 80% user class 1 - 20% user class 3 in a 50-50 mixture with alien class 3.

Five data files were generated based on the classes shown in Figure 22. For the first data file all the covariance matrices, user and alien, were taken to be the diagonal matrix diag (40 40).

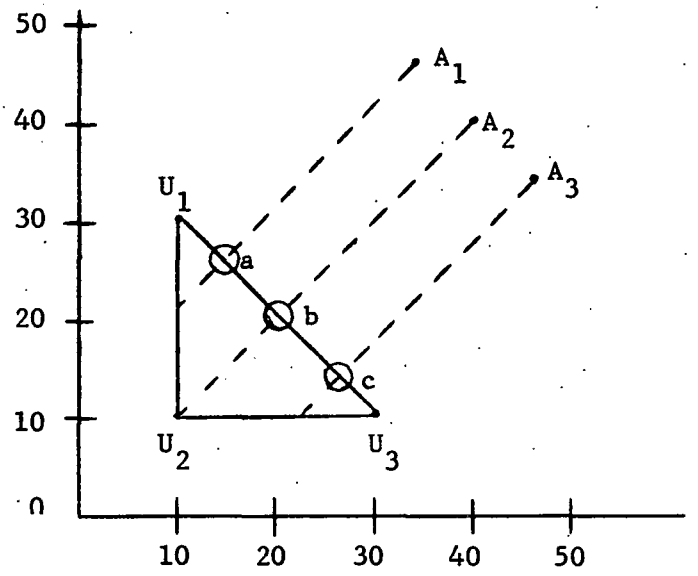


Figure 22. Arrangement of user and alien class means for unequal covariance test data

Program SIMDAT then generated 30 random observations from each of the twelve mixtures defined by the different combinations of user and alien classes. Thus, all the data in file 1 was generated using equal covariance matrices.

The other files were constructed similarly to file 1 except that in these cases the user covariance matrices were taken to be unequal (Table 10). Since the proportion associated with user class 2 was 0 in every case, its covariance matrix remained the same throughout, as did those of the alien classes. To get a rough idea of how dissimilar the user covariance matrices were, the chi-squared statistic for testing equal covariances was calculated as though a hypothetical sample of 50 observations from each user class had been drawn and yielded sample covariance matrices equal to those in Table 10. The values of the test statistic associated with each of the last four files indicate that

Table 10. Test statistic for equal covariances test using covariance matrices associated with files of artificial data

File	Covariance matrices			MC^{-1}
1	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	0
2	$\begin{pmatrix}25 & 0\\ 0 & 25\end{pmatrix}$	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}55 & 0\\ 0 & 55\end{pmatrix}$	14.56*
3	$\begin{pmatrix}5 & 0\\ 0 & 5\end{pmatrix}$	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}75 & 0\\ 0 & 75\end{pmatrix}$	139.39**
4	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}40 & 30\\ 30 & 40\end{pmatrix}$	30.41**
5	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}40 & 0\\ 0 & 40\end{pmatrix}$	$\begin{pmatrix}40 & -30\\ -30 & 40\end{pmatrix}$	30.41**

*Significant at .025 level.

**Significant at .01 level.

the covariance matrices used to generate these files were statistically dissimilar in terms of this hypothetical test.

Each data file was processed by STDEST and SMPEST, and estimates of the mean square error were obtained. The results are recorded in Table 11. For the most part the results are as expected. For files 2 and 3 the errors are smaller than for file 1 in the top third of the table where a larger proportion is associated with user class 1, which has the smaller variances, and the errors are larger in the bottom third of the table where the covariance matrix of user class 3 is weighted more heavily.

The results for files 4 and 5 are somewhat more interesting. For user class 3 there was a large positive correlation between the bands

Table 11. Mean square errors with files of equal and unequal covariance matrices using fixed proportions

Case	File:	Standard estimator					Simplified estimator				
		1	2	3	4	5	1	2	3	4	5
a		0.1660	0.1459	0.0983	0.1635	0.1823	0.1869	0.1619	0.1077	0.1970	0.1913
a1		0.0503	0.0468	0.0394	0.0471	0.0550	0.0587	0.0563	0.0547	0.0570	0.0596
a2		0.1475	0.1361	0.1209	0.1404	0.1551	0.1263	0.1221	0.1174	0.1254	0.1292
a3		0.2603	0.2456	0.2418	0.2528	0.2689	0.1955	0.1922	0.1929	0.1935	0.1985
b		0.1541	0.1541	0.1541	0.1385	0.1558	0.1421	0.1421	0.1421	0.1279	0.1441
b1		0.1011	0.1011	0.1011	0.0948	0.1062	0.0323	0.0323	0.0323	0.0301	0.0340
b2		0.0573	0.0573	0.0573	0.0508	0.0686	0.0237	0.0237	0.0237	0.0243	0.0265
b3		0.1290	0.1290	0.1290	0.1193	0.1429	0.0396	0.0396	0.0396	0.0345	0.0468
c		0.1584	0.1822	0.2190	0.1497	0.1358	0.1736	0.1997	0.2337	0.1844	0.1374
c1		0.3475	0.3591	0.3765	0.3019	0.3790	0.2293	0.2324	0.2385	0.2204	0.2355
c2		0.0611	0.0668	0.0736	0.0549	0.0706	0.0842	0.0848	0.0891	0.0847	0.0872
c3		0.0697	0.0738	0.0821	0.0520	0.0827	0.0550	0.0579	0.0629	0.0465	0.0594
Average		0.1419	0.1415	0.1411	0.1305	0.1502	0.1123	0.1121	0.1112	0.1105	0.1125

in file 4 and a large negative correlation in file 5. For the standard estimator the errors for file 4 are consistently smaller than for file 1, presumably because more of the variation in the data occurs perpendicular to the U_1U_3 line than parallel to it. Just the opposite happens with file 5, where the errors are larger than those of file 1 except in case c. In case c the estimated proportion of class 2 is lower for file 5 than for file 1, which may account for the lower mean square error.

For the simplified estimator the results are not quite so consistent. In most cases the errors associated with file 4 are lower than for the equal covariance case, but in a few instances they are slightly higher. Apparently, the simplified estimator receives less benefit from the large positive correlation than the standard estimator. For file 5 the errors are again consistently higher than for file 1 except in case c.

For the second phase of testing the effect of unequal covariance matrices, the means and covariance matrices extracted from LANDSAT data and presented in Table 9 were used to construct two simulated data files. The first file was constructed taking the average of the four covariance matrices as the common covariance matrix for each class. The second file was constructed using the different covariance matrices associated with each of the classes. Each file contained 300 points, and was generated by SIMDAT with random proportions using the parameters: $\gamma = 1.0$, $\alpha = 1.0$, $\beta = 0.0$, and $\tau = 1/7$.

Both data sets were processed by STDEST and SMPEST with and without data averaging. To examine the effect of region size, regions of 10 and

100 points were used. Overall mean square error estimates were obtained by averaging over five regions when the region size was 10 and over three regions when the region size was 100. The results are presented in Table 12 and plotted in Figures 23 and 24.

Table 12. Effect of unequal covariances on mean square errors of standard and simplified estimators using simulated LANDSAT data

Estimator	Region size	Data averaging			No averaging		
		Equal	Unequal	Unequal / equal	Equal	Unequal	Unequal / equal
Standard	10	0.1004	0.1191	1.19	0.0352	0.0288	0.82
	100	0.0451	0.0404	0.90	0.0100	0.0125	1.25
Simplified	10	0.1506	0.1526	1.01	0.0576	0.0479	0.83
	100	0.0451	0.0404	0.90	0.0123	0.0095	0.77

No clear patterns seem to show up in the results. In some cases the error is larger with unequal covariance matrices than with equal ones, and in other cases it is smaller. This inconsistency happens both with and without data averaging, for the standard and for the simplified estimator, and for small regions and large ones. If one looks at the regions, it is evident that the random variation between regions of the same test case is much greater than differences between corresponding equal and unequal covariance cases. If there is any effect due to unequal covariances, it does not appear to be significant enough to show through the sampling error.

On the basis of the results of the unequal covariance tests, it appears that the proportion estimation procedure is fairly robust with regard to the equal covariance assumption. If one estimates proportions as though all the covariance matrices were equal to a common average

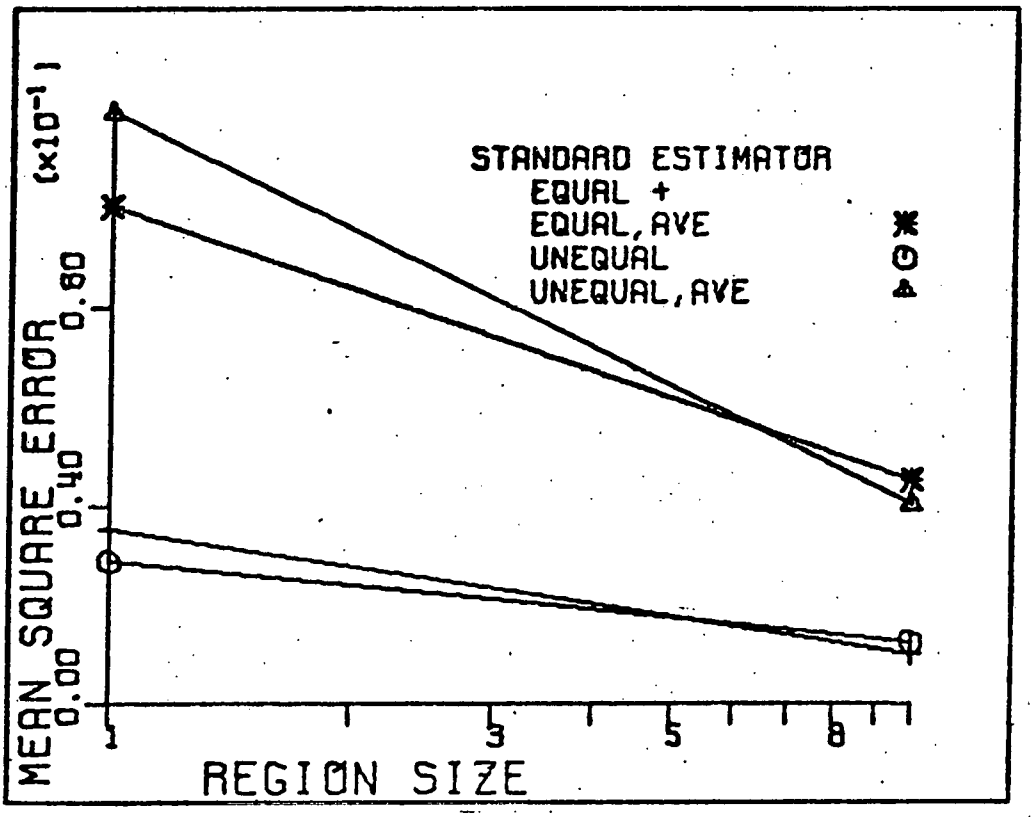
Figure 23. Effect of unequal covariances with simulated LANDSAT data using standard estimator

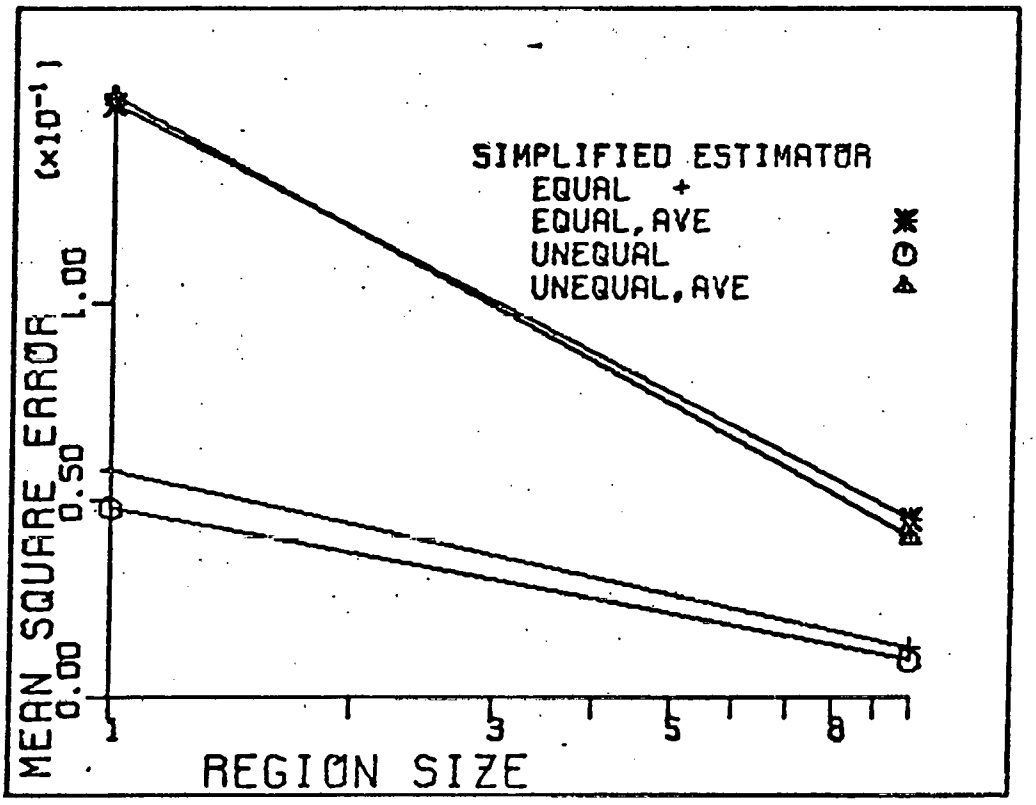
Figure 24. Effect of unequal covariances with simulated LANDSAT data using simplified estimator

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK


STANDARD ESTIMATOR
EQUAL +
EQUAL, AVE
UNEQUAL
UNEQUAL, AVE
MEAN SQUARE ERROR (×10⁻¹)
0.00
0.40
0.80
1
3
5
8
REGION SIZE


SIMPLIFIED ESTIMATOR
EQUAL +
EQUAL, AVE
UNEQUAL
UNEQUAL, AVE
MEAN SQUARE ERROR (×10⁻¹)
0.00
0.50
1.00
1
3
5
8
REGION SIZE

matrix, it does not seem to matter much if the actual data came from normal distributions with unequal covariance matrices.

The results of the first phase of tests using fixed proportions show that a high degree of interband correlation can have an effect on the accuracy of the estimates, depending on the geometry of the signature simplex. However, if one merely redistributes the variation by increasing the variances associated with one class and decreasing the variances of another class, the errors seem to average out, so that overall there is no net effect.

The only conclusion that can be drawn from the tests using simulated LANDSAT data with random proportions is that even when the covariance matrices are significantly different, the accuracy of the proportion estimates does not appear to be measurably affected.

V. USE OF L_1 NORM IN CLASSIFICATION

A. Motivation for Using L_1 Norm

1. Normality assumption

Throughout all of the developments of the previous chapters, the assumption was made that the spectral responses for each type of material were normally distributed. The probable reasons that this assumption is usually made include tradition, and mathematical tractability. By taking the classes to follow normal distributions, one can completely specify the density function for each class by estimating only the first- and second-order moments from training data. Also, normality leads to a reasonable form for the likelihood function, which can be evaluated by straightforward computation.

If one examines in detail multispectral data actually taken from natural scenes, one will most likely observe various departures from strict normality, such as outliers or pronounced peakedness or flatness in histograms of the data. Figure 15 is a histogram of the responses obtained by LANDSAT over a soybean field in central Iowa. The symmetry of the distribution is apparent, but the peakedness at the center suggests possible nonnormality. Histograms for other fields in the same LANDSAT image were found to exhibit a variety of shapes, but the general shape shown in Figure 25 is typical of many of the histograms.

Some researchers have subjected the normality assumption to statistical examination. Crane et al. (1972) used data from airborne multispectral scanners flown over two different agricultural sites.

BAND 7

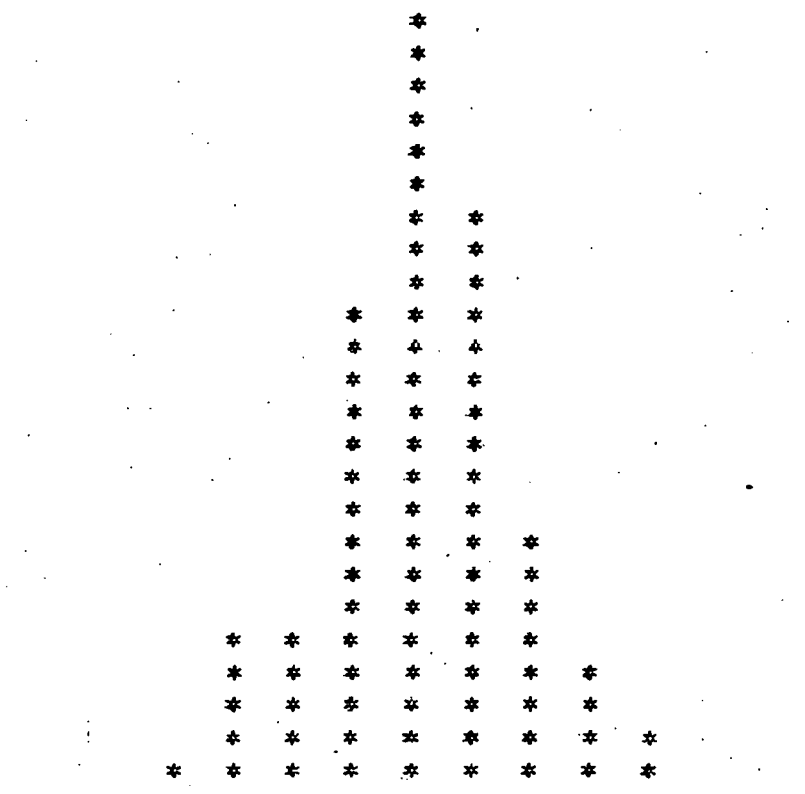


Figure 25. Histogram of LANDSAT data from soybean field

They applied a chi-squared goodness-of-fit test for normality on the data from each of 54 fields, where each field contained but one crop, and boundary pixels were excluded. They looked at both the original spectral channels and channels transformed by a principal components transformation.

It was found that none of the fields examined tested to be multivariate normal at the 1% significance level. All had at least one nonnormal spectral channel. Overall, 65% of the untransformed channels tested nonnormal. Their conclusion was that multispectral data of the type they studied was definitely nonnormal in character.

Given that the normal model does not truly reflect the real data, two questions arise:

Can another model be used instead of the normal model?

If so, how does a classifier based on this alternate model perform compared to a least squares classifier?

The remainder of the chapter considers these two questions in more detail.

2. Basis for using L_1 norm

In Chapter I the Bayes discriminant procedure was shown to be the procedure that assigns an unknown observation x to class k if

$$\sum_{\substack{i=1 \\ i\neq k}}^{m} q_i p(x|\pi_i)\, C\,(k|i) \leq \sum_{\substack{i=1 \\ i\neq j}}^{m} q_i p(x|\pi_i)\, C\,(j|i),\ j = 1, \ldots, m.$$

Also, if the costs of misclassification are equal, and the prior probabilities are equal, the procedure reduces to choosing class k such that

$$p(x|\pi_k) \geq p(x|\pi_j),\ j = 1, \ldots, m,$$

which yields the maximum likelihood solution.

Under the assumption that the classes are normally distributed with means μ_i and covariance matrices Σ_i, the likelihood function is

$$p(x|\pi_i) = \frac{1}{(2\pi)^{n/2}|\Sigma_i|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)}$$

and

$$\ln(p(x|\pi_i)) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i),$$
$$i = 1, \ldots, m.$$

Finding the maximum likelihood solution involves evaluating m quadratic forms of the type

$$Q = (x-\mu)'\Sigma^{-1}(x-\mu).$$

If Σ is positive definite, it may be decomposed as

$$\Sigma = LL',$$

where L is lower triangular, and the observation vectors may be transformed by

$$y = L^{-1}(x-\mu).$$

Then the quadratic form reduces to

$$Q = y'y = \sum_{i=1}^{m} y_i^2 .$$

Let the L_p norm of a vector be defined as

$$\|y\|_p = \left(\sum_{i=1}^{n} |y_i|^p \right)^{1/p} .$$

Then

$$\|y\|_p^p = \sum_{i=1}^{n} |y_i|^p ,$$

and it follows that

$$Q = \|y\|_2^2 .$$

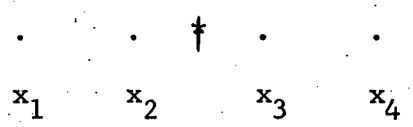
Thus, Q is simply the squared Euclidean, or L_2, norm of the vector y. There is no essential reason why norms other than the Euclidean norm cannot be considered in connection with the classification problem. Choices for p besides p = 2 lead to estimation procedures having different (and, in some cases, superior) properties than procedures based on the L_2 norm. In this chapter attention will be focused on the L_1 norm as an alternative to least squares.

One of the problems that often arises in assuming that the errors in one's data are normally distributed is the presence of extreme data points, or outliers. If the errors were truly normal, there would almost never be any outliers, yet outliers can and do occur in real data. It has been shown that the presence of extreme points in the data can seriously degrade the performance of an estimator based on the L_2 norm.

On the other hand, the L_1 norm is much less sensitive to outliers in the data. The L_1 norm, when used to select the point whose distance from the collection of points in a data set is a minimum,

will pick the median of the data points. It ignores how far an extreme point is from the center of the data and considers only on which side of the center it falls.

Consider two simple examples in one dimension. Let the data points be represented by x_i, $i = 1, \ldots, 4$. The problem is to find the point p such that the $\|y\|$ is minimized, where $y_i = x_i - p$. Let the location of the L_2 estimate and the L_1 estimate be designated by "*" and "|" respectively. The sample data points and the two estimates are depicted below.



As a second example suppose another point, x_5, is added to the data set. The revised set of points and the new L_1 and L_2 estimates now become as shown.

It is evident that the addition of the extreme point x_5 has affected the L_2 estimate much more drastically than it has affected the L_1 estimate. Had x_5 been placed even farther to the right, the L_2 estimate would have shifted more in that direction, while the L_1 estimate would have remained at x_3.

To see how the L_1 and L_2 norms could give rise to different results in a classification situation, consider the following example with two classes and three bands. For simplicity assume the covariance matrices

are both equal to the identity matrix and let the means be

$$\mu_1 = \begin{pmatrix} 20 \\ 50 \\ 50 \end{pmatrix} \qquad \mu_2 = \begin{pmatrix} 40 \\ 35 \\ 35 \end{pmatrix}.$$

Now suppose the response vector

$$x = \begin{pmatrix} 0 \\ 30 \\ 30 \end{pmatrix}$$

was observed, where the response in band 1 was not recorded properly for some reason. The goal is to classify the observation x into either class 1 or 2 depending on which class is "closest" to x. Using the L_2 norm one finds

$$\|x - \mu_1\|_2^2 = (x - \mu_1)'(x - \mu_1) = 1200$$

$$\|x - \mu_2\|_2^2 = (x - \mu_2)'(x - \mu_2) = 1650.$$

Using the L_1 norm one has

$$\|x - \mu_1\|_1 = \sum_{j=1}^{3} |x_j - \mu_{1j}| = 60$$

$$\|x - \mu_2\|_1 = \sum_{j=1}^{3} |x_j - \mu_{2j}| = 50.$$

Thus, by taking the L_2 norm as the measure of distance, one would classify x into class 1, but using the L_1 norm would result in x being classified into class 2. It can be seen that the anomalous response in band 1 caused x to be closer to μ_1 in terms of the L_2

measure, but the L_1 norm was not affected as much by the one extreme value in the response. This of course does not mean that the L_1 norm will always result in a better classification than the L_2 norm, but it does illustrate that the L_1 norm is less sensitive to extreme points in the data.

As an alternative to the normal model and L_2 norm, Chhikara and Odell (1973b) proposed what they termed normed exponential density functions. The general form of an r-normed exponential density function is

$$f^{(r)}(y) = K_r e^{-c\|y\|_r^r},$$

where

$$K_r = (2\int_0^\infty e^{-cu^r}du)^{-n}.$$

Here n is the dimension of y, and c is a positive constant determined such that $E(yy') = I$. The density function corresponding to the L_1 norm is

$$f^{(1)}(y) = \frac{1}{2^{n/2}} e^{-\sqrt{2}\sum_{k=1}^{n}|y_k|}, \quad -\infty < y_k < \infty, \tag{5.1}$$

which is the multivariate analog of the double exponential density.

Now consider how the L_1 model can be utilized in the discriminant problem. Let μ_i be the mean and Σ_i the covariance matrix for class i, i = 1, ..., m. Assume each Σ_i is positive definite and may thus be factored as

$$\Sigma_i = S_i S_i' ,$$

and the let the observation vector x be transformed according to

$$y = S_i^{-1}(x - \mu_i).$$

The inverse transformation is

$$x = S_i y + \mu_i ,$$

and the Jacobian of the inverse transformation is $|S_i^{-1}|$. Making the $y \to x$ transformation in (5.1) and changing notation slightly, the density function for class i may be written

$$p(x|\pi_i) = \frac{1}{2^{n/2}|S_i|} e^{-\sqrt{2}\sum_{k=1}^{n} |S_{i(k)}^{-1}(x-\mu_i)|}, \quad i = 1, \ldots, m,$$

where $S_{i(k)}^{-1}$ is the kth row of S_i^{-1}.

For simplicity assume equal prior probabilities and equal costs of misclassification. Then the Bayes procedure chooses class j if

$$p(x|\pi_j) \geq p(x|\pi_i), \quad i = 1, \ldots, m. \tag{5.2}$$

Writing out the density functions, (5.2) becomes

$$2^{-n/2}|S_j^{-1}| e^{-\sqrt{2}\sum_{k=1}^{n} |S_{j(k)}^{-1}(x-\mu_j)|} \geq 2^{-n/2}|S_i^{-1}| e^{-\sqrt{2}\sum_{k=1}^{n} |S_{i(k)}^{-1}(x-\mu_i)|^2} .$$

Simplifying and taking logarithms of both sides yields

$$\sum_{k=1}^{n} |S_{i(k)}^{-1}(x - \mu_i)| - \sum_{k=1}^{n} |S_{j(k)}^{-1}(x - \mu_j)| \geq \frac{1}{\sqrt{2}} \ln \frac{|S_j|}{|S_i|} .$$

Thus the Bayes discriminant region for class j is

$$R_j = \left\{ x: \sum_{k=1}^{n} |S_{i(k)}^{-1}(x - \mu_i)| - \sum_{k=1}^{n} |S_{j(k)}^{-1}(x - \mu_j)| \geq \frac{1}{\sqrt{2}} \ln \frac{|S_j|}{|S_i|} , \right.$$
$$\left. i = 1, \ldots, m \right\}, \; j = 1, \ldots, m. \qquad (5.3)$$

One of the advantages of the L_1 norm in classification is that the boundaries between discriminant regions defined by (5.3) are piecewise linear (in x). This makes evaluation of the probability of misclassifying an observation from π_i into π_j a problem in integrating over linear planes since the probability is given by

$$p(j|i) = \int_{R_j} p(x|\pi_i)dx, \; j = 1, \ldots, m, \; j \neq i.$$

An exact evaluation of the probabilities of misclassification is thus possible under the L_1 norm. In the case of the L_2 norm, the evaluation of the probabilities of misclassification involves the integration of multivariate normal density functions over quadratic regions, prohibiting an exact evaluation of the probabilities.

Another advantage of L_1 is its computational efficiency. With the L_2 norm one must compute the quadratic forms

$$(x - \mu_i)' \Sigma_i^{-1} (x - \mu_i), \; i = 1, \ldots, m,$$

corresponding to each of the m classes in order to classify the observation x. The computation in the case of the L_1 norm involves evaluating

$$\sum_{k=1}^{n} \left| S_{i(k)}^{-1}(x - \mu_i) \right|, \quad i = 1, \ldots, m.$$

To compare the amount of computation required for each norm, assume an image is to be classified by a Bayes discriminant procedure and let

n = number of spectral bands

m = number of classes

r = number of rows in the image

c = number of columns in the image.

Chhikara and Odell (1973b) have calculated the number of orderings, O, multiplications, M, and additions, A, necessary to carry out the computations indicated above for each of the two norms. For L_1 they found

$$O_1 = mrc$$

$$M_1 = mrc(n^2 + n)/2$$

$$A_1 = mrc(n^2 + 3n - 2)/2,$$

and for L_2 they got

$$O_2 = mrc$$

$$M_2 = mrc(n^2 + 3n)/2$$

$$A_2 = mrc(n^2 + 3n - 2)/2.$$

The computational savings in using L_1 arises in the number of multiplications required.

As an example, let n = 4, m = 10, r = 100, and c = 100. Then the number of orderings, multiplications, and additions would be

$$O_1 = O_2 = 10^5$$

$$A_1 = A_2 = 13 \times 10^5$$

$$M_1 = 10^6$$

$$M_2 = 14 \times 10^5.$$

It is seen that, in this case, L_1 offers a savings of 5:7 in terms of the number of multiplications necessary to classify the image, or 400,000 fewer multiplications.

B. Implementation and Testing of L_1 Classifier

To implement a classifier using the L_1 norm, a previously written program for maximum likelihood classification based on least squares, CLSSFY, was modified to employ a discriminant function based on the L_1 norm. Using the notation of the previous section, the discriminant function to be maximized under L_2 was

$$D_i = -\frac{1}{2} \ln|\Sigma_i| - \frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i),$$

and under L_1 the discriminant function is

$$D_i = \ln|S_i^{-1}| - \sqrt{2} \sum_{k=1}^{n} |S_{i(k)}^{-1} (x - \mu_i)|.$$

Other aspects of the program were essentially unchanged by the switch to the L_1 norm.

CLSSFY reads the means and covariance matrices of the classes and the coordinates of the region to be classified. For each point in the region, the value of D_i is computed corresponding to each class and the class k is determined such that

$$D_k \geq D_i \ , \ i = 1, \ldots, m.$$

The class index k and the value of D_k of the discriminant function are stored on disk for subsequent mapping or tabulation of the classification results.

In comparing the results obtained under L_1 with those obtained under L_2, it was necessary to select a criterion for comparison. Since the program SIMDAT has the capability of generating data sets for which the exact proportions of all classes are known for each data point, a mean square error criterion based on the errors in the estimated proportions of each class in a region was chosen. Program CLSSFY classifies each data point into exactly one class. Therefore, the "estimated proportions" for a data point are always 1.0 for the class selected and 0.0 for all other classes. Using such vectors of estimated proportions for each point, the mean square error for the estimated proportions of classes in a region may be estimated by the method discussed in Chapter II. Other criteria could certainly be used to evaluate classification accuracy. The number of points correctly classified, assuming each point's true identity consists of only one class, divided by the total number of points in the region is commonly used, but the mean square error criterion seems to correspond more closely with the methods used in previous chapters.

Several data sets were constructed in order to evaluate the accuracy of classification with the L_1 norm. A 2-dimensional data set was constructed specifically to show how the L_1 norm can give different results than the L_2 norm in the presence of outliers. The geometric arrangement of the means is depicted in Figure 26, where the U_i are the user class means, and Z represents the data point that would result if some recording error were to cause the value of the response in band 2 to not be recorded for an observation from class 2. Data file A was generated to consist of points from classes 1 and 2, a mixture of classes 1 and 2, and some points in the vicinity of Z. Table 13 contains the results of running the L_1 and L_2 classifiers on this data set as well as the data sets discussed below.

Two data sets were constructed with three user classes and two bands. The class means were:

$$A_1 = \begin{pmatrix} 10 \\ 10 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 30 \\ 10 \end{pmatrix}, \text{ and } A_3 = \begin{pmatrix} 30 \\ 20 \end{pmatrix}.$$

The covariance matrix for each class was taken to be the diagonal matrix diag (50 50). Instead of generating the data from a multivariate normal distribution, the data for each file were taken from a contaminated normal distribution.

For data file B1 the normal data were contaminated by introducing a certain percentage (20%) of extreme points at random into the data. The extreme points were formed by setting the response in one of the bands to 100.

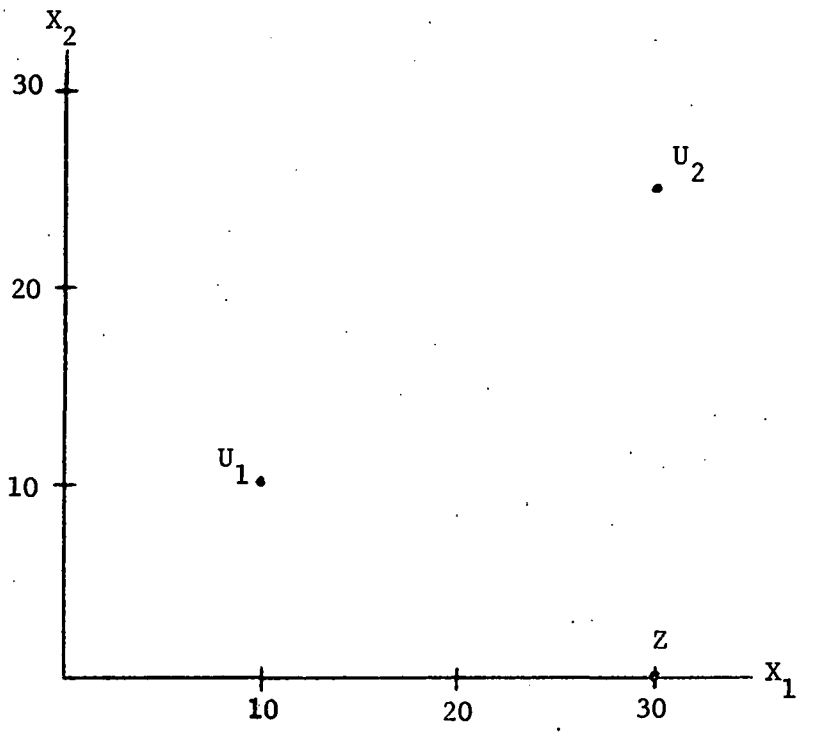


Figure 26. Arrangement of class means and extreme point used in data file A for testing L_1 classifier

Table 13. Comparison of mean square errors and processing times for L_1 and L_2 classifiers

File	Total points	Mean square error (*10^2)		Processing time (sec)	
		L_1	L_2	L_1	L_2
A	100	0.07	2.84	0.05	0.05
B1	400	0.5025	0.5309	0.27	0.31
B2	400	0.7828	0.7504	0.27	0.40
C1	300	1.040	1.189	0.52	0.59
C2	300	1.009	1.018	0.48	0.48
D1	300	1.67	1.76	0.59	0.66
D2	300	0.76	0.73	0.59	0.64

The second file based on these three classes and two bands, B2, was generated as multivariate normal data contaminated by a Cauchy distribution. Instead of generating n univariate N(0, 1) variates and constructing a $N_n(\mu, \Sigma)$ observation as described in Chapter II, the univariate deviates were generated from a Cauchy distribution. A relatively high contamination rate of 50% resulted in marginal distributions with a sharp peak at the mean.

Two data sets were generated similar to ones discussed in earlier chapters except that a contamination factor was introduced in producing the observations. One data file, C1, contained observations from the user and alien classes listed in Table 4. The data were generated with random mixtures of classes as before, but deviates from a Laplace distribution were used to contaminate the data. The second data set, C2, was based on the classes presented in Table 9. No alien material was present in this case, and mixtures of the user classes were randomly generated. The data were contaminated by introducing 10% extreme points.

The final two files listed in Table 13, D1 and D2, were generated using the same file parameters as for C1 and C2, respectively, except that no contamination was used. These last two files were included to see if the L_1 classifier would perform any worse than the L_2 classifier on truly normal data.

C. Results

The results in Table 13 are encouraging, but not as conclusive as one might wish for. The mean square errors for file A should be taken

as indicating the potential improvement that can be realized with the L_1 classifier since the data was constructed specifically to illustrate how L_1 can do better with extreme points in the data.

Files B1 and B2 present ambiguous results. The differences are small in both cases, probably because there was a great deal of overlap in the discriminant regions defined by the two classifiers.

Files C1 and C2 both favored the L_1 classifier, although the difference was very slight for file C2, where the contamination rate was only 10%. Since these data sets both consisted of LANDSAT-type data, one with alien material and the other without, these results are perhaps the most encouraging ones as far as the use of L_1 is concerned.

The results for files D1 and D2 indicate that there does not appear to be much penalty for using the L_1 classifier when the data are really normal. In fact, file D1 demonstrates that L_1 can actually do better in some cases.

The timing results seem to be as one would expect from a consideration of the computations involved. On the average the L_1 classifier required 10% less computing time, which could be highly significant where extremely large volumes of data are involved, as can be the case in processing multispectral data.

While the results presented in Table 13 are merely a preliminary investigation, they do seem to indicate that the L_1 norm merits further attention. It would be helpful to have a clearer understanding of the kinds of perturbations in the data that L_1 can handle better than L_2. Also, the fact that the L_1 norm is simply one norm from the general

class of L_p norms suggests that there may be other norms worth examining in this context.

D. Application to Mixtures

In Section A of this chapter it was shown how a normed exponential model based on the L_1 norm could be utilized in classifying observations of multispectral data. An avenue for further research would be to consider how one might approach the mixtures problem given a normed exponential model for the data. In this section the groundwork for such an approach is presented, and some suggestions are made for continuing the development.

The mixtures model to be presented here parallels the presentation given in Chapter I except that the L_1-normed exponential density is taken as the model for the data. Suppose there are m classes of material and n spectral channels. Assume a resolution element consists of N cells, with N_i cells containing material i. Let X_{ij} be the random variable associated with the response of the jth cell containing material i, and assume X_{ij} and X_{ik}, $1 \leq j \neq k \leq N_i$, are independent for all $i = 1, \ldots, m$. Let μ_i^* and Σ_i^* denote the mean and covariance matrix of X_{ij}, and assume $\Sigma_i^* = S_i^* S_i^{*\prime}$. Then under the L_1 model, the density function of X_{ij} is

$$f_i(x) = \frac{|S_i^{*-1}|}{2^{n/2}} e^{-\sqrt{2} \sum_{k=1}^{n} |S_{i(k)}^{*-1}(x-\mu_i^*)|} \tag{5.4}$$

Let Y_i be the random variable associated with the total response obtained from the cells of class i, $i = 1, \ldots, m$. Then

$$Y_i = \sum_{j=1}^{N_i} X_{ij} ,$$

$$E(Y_i) = N_i \mu_i^* ,$$

and

$$V(Y_i) = N_i \Sigma_i^* .$$

If the entire resolution element contained only material from class i, then one would have

$$E(Y_i) = N\mu_i^* \equiv \mu_i$$

and

$$V(Y_i) = N\Sigma_i^* \equiv \Sigma_i .$$

Let Y denote the random variable representing the observed response from the entire resolution element, and let λ_i be the proportion of the resolution element containing material i. If the random variables for cells from different classes are assumed to be independent, then

$$Y = \sum_{i=1}^{m} Y_i = \sum_{i=1}^{m} \sum_{j=1}^{N_i} X_{ij} ,$$

$$E(Y) = \sum_{i=1}^{m} N_i \mu_i^* = \sum_{i=1}^{m} \lambda_i N \mu_i^* = \sum_{i=1}^{m} \lambda_i \mu_i .$$

and

$$V(Y) = \sum_{i=1}^{m} N_i \Sigma_i^* = \sum_{i=1}^{m} \lambda_i N \Sigma_i^* = \sum_{i=1}^{m} \lambda_i \Sigma_i .$$

Y is a linear combination of multivariate random variables each having an L_1-normed exponential distribution. It remains to be determined what the density function is for Y.

Thus, a starting point for further investigation would be to refer to results from multivariate distribution theory to establish the density function $g(y; \mu, \Sigma, \lambda)$ for Y. Once one has the distribution of Y, the problem becomes one of estimating λ given an observation on Y. For a maximum likelihood solution, one would have to solve

$$\underset{\lambda}{\text{maximize}}\ g(y; \mu, \Sigma, \lambda)$$

such that

$$\sum_{i=1}^{m} \lambda_i = 1 \text{ and } \lambda_i \geq 0,\ i = 1, \ldots, m.$$

It is quite possible that certain simplifying assumptions would be necessary to obtain a solution with a feasible amount of effort.

The following questions are given to suggest points to be pursued in considering the application of the L_1 model to mixtures.

1. What is the distribution of Y, the random variable associated with the mixture?

2. If g(Y) is known, can a computationally efficient algorithm be found to obtain estimates of the λ_i?

3. How good are the estimates of λ?

4. How do the properties of the L_1 norm affect the values of the estimates, especially when alien material is present?

It is felt that alternatives to the normal model and least squares, such as those presented above based on the L_1 norm, should be given

further consideration in view of the potential computational simplifications and the nonnormal nature of much multispectral data. However, more should be learned about these alternatives to accurately and completely assess the tradeoffs involved in using them in lieu of normality and the L_2 norm.

VI. BIBLIOGRAPHY

Anderson, T. W. 1958. An introduction to multivariate statistical analysis. John Wiley & Sons, Inc., New York. 374 pp.

Aspiazu, Celestino. 1977. Application of LANDSAT-1 data to crop area and yield estimates, forest identification, and soil classification in Iowa. Ph.D. Thesis. Iowa State University, Ames, Iowa. 138 pp.

Bauer, M. E., and J. E. Cipra. 1973. Identification of agricultural crops by computer processing of ERTS MSS data. Pages 205-212 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Symposium on significant results obtained from the Earth Resources Technology Satellite-1, vol. 1, sec. A. National Aeronautics and Space Administration NASA SP-327.

Bernstein, R. 1974. Scene correction (precision processing) of ERTS sensor data using digital image processing techniques. Pages 1909-1928 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Third Earth Resources Technology Satellite-1 symposium, vol. 1, sec. B. National Aeronautics and Space Administration NASA SP-351.

Boeckel, J. H. 1974. ERTS-1 system performance overview. Pages 1-12 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Third Earth Resources Technology Satellite-1 symposium, vol. 1, sec. A. National Aeronautics and Space Administration NASA SP-351.

Chen, C. 1973. Statistical pattern recognition. Hayden Book Co., Rochell Park, N.J. 236 pp.

Chhikara, R. S., and P. L. Odell. 1973a. Estimation of proportions of objects and determination of training sample-size in a remote sensing application. Pages 4B-16-4B-24 in Symposium proceedings, machine processing of remotely sensed data. Laboratory for Applications of Remote Sensing, West Lafayette, Ind.

Chhikara, R. S., and P. L. Odell. 1973b. Discriminant analysis using certain normed exponential densities with emphasis on remote sensing application. Pattern Recognition 5(3): 259-272.

Colvocoresses, A. P. 1973. Unique characteristics of ERTS. Pages 1523-1525 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Symposium on significant results obtained from the Earth Resources Technology Satellite-1, vol. 1, sec. B. National Aeronautics and Space Administration NASA SP-327.

Colwell, R. N. 1973. Remote sensing as an aid to the management of earth resources. American Scientist 61(2): 175-183.

Crane, R. B., W. A. Malila, and W. Richardson. 1972. Suitability of the normal density assumption for processing multispectral scanner data. IEEE Trans. on Geoscience Electronics 10(4): 158-164.

Day, N. E. 1969. Estimating the components of a mixture of normal distributions. Biometrika 56(3): 463-474.

Detchmendy, D. M., and W. H. Pace. 1972. A model for spectral signature variability for mixtures. Pages 596-620 in F. Shahrokhi, ed. Remote sensing of earth resources, vol. 1. University of Tennessee Space Institute, Tullahoma, Tenn.

Faddeeva, V. N. 1959. Computational methods of linear algebra. Dover Publications, Inc., New York. 252 pp.

Fu, K. S., D. A. Landgrebe, and T. L. Phillips. 1969. Information processing of remotely sensed agricultural data. Proc. of the IEEE 57(4): 639-653.

Fukunaga, K. 1972. Introduction to statistical pattern recognition. Academic Press, New York. 369 pp.

Hadley, G. 1964. Nonlinear and dynamic programming. Addison-Wesley Publishing Company, Reading, Mass. 484 pp.

Hajic, E. J., and D. S. Simonett, 1976. Comparisons of qualitative and quantitative image analysis. Pages 374-411 in J. Lintz and D. S. Simonett, eds. Remote sensing of environment. Addison-Wesley Publishing Company, Reading, Mass.

Hallum, C. R. 1972. On a model for optimal proportions estimates for category mixtures. Pages 951-958 in Proceedings of the eighth international symposium on remote sensing of environment. Environmental Research Institute of Michigan, Ann Arbor, Mich.

Hasselblad, V. 1966. Estimation of parameters for a mixture of normal distributions. Technometrics 8(3): 431-444.

Hasselblad, V. 1969. Estimation of finite mixtures of distributions from the exponential family. J. Amer. Stat. Assoc. 64(328): 1459-1471.

Holmes, R. A., and R. B. MacDonald. 1969. The physical basis of system design for remote sensing in agriculture. Proc. of the IEEE 57(4): 629-639.

Horton, M. L., and J. L. Heilman. 1973. Crop identification using ERTS imagery. Pages 27-33 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Symposium on significant results obtained from the Earth

Resources Technology Satellite-1, vol. 1, sec. A. National Aeronautics and Space Administration NASA SP-327.

Horwitz, H. M., R. F. Nalepka, P. D. Hyde, and J. P. Morgenstern. 1971. Estimating the proportions of objects within a single resolution element of a multispectral scanner. Pages 1307-1320 in Proceedings of the seventh international symposium on remote sensing of environment. Environmental Research Institute of Michigan, Ann Arbor, Mich.

Horwitz, H. M., P. D. Hyde, and W. Richardson. 1974. Improvements in estimating proportions of objects from multispectral data. ERIM 190100-25-T. (Environmental Research Institute of Michigan, Ann Arbor, Mich.)

Klemm, R. J., and V. A. Sposito. 1977. Least squares solutions over interval restrictions. Unpublished technical paper, Department of Statistics, Iowa State University, Ames, Iowa.

Kunzi, H. P., W. Krella, and W. Oettli. 1966. Nonlinear programming. Blaisdell Publishing Company, Waltham, Mass. 240 pp.

Landgrebe, D. A. 1971. Systems approach to the use of remote sensing. Pages 139-154 in International workshop on earth resources survey systems, vol. 1. National Aeronautics and Space Administration NASA SP-283.

Landgrebe, D. A. 1976. Machine processing of remotely acquired data. Pages 349-373 in J. Lintz and D. S. Simonett, eds. Remote sensing of environment. Addison-Wesley Publishing Company, Reading, Mass.

Lintz, J., and D. S. Simonett. 1976. Sensors for spacecraft. Pages 323-343 in J. Lintz and D. S. Simonett, eds. Remote sensing of environment. Addison-Wesley Publishing Company, Reading, Mass.

Malila, W. A., and R. F. Nalepka. 1973. Atmospheric effects in ERTS-1 data, and advanced information extraction techniques. Pages 1097-1104 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Symposium on significant results obtained from the Earth Resources Technology Satellite-1, vol. 1, sec. B. National Aeronautics and Space Administration NASA SP-327.

Malila, W. A., and R. F. Nalepka. 1974. Advanced processing and information extraction techniques applied to ERTS-1 MSS data. Pages 1743-1772 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Third Earth Resources Technology Satellite-1 symposium, vol. 1, sec. B. National Aeronautics and Space Administration NASA SP-351.

Mendel, J. M., and K. S. Fu. 1970. Adaptive, learning and pattern recognition systems; theory and applications. Academic Press, New York. 444 pp.

Morrison, D. F. 1967. Multivariate statistical methods. McGraw-Hill Book Company, New York. 338 pp.

Nalepka, R. F., and P. D. Hyde. 1972. Classifying unresolved objects from simulated space data. Pages 935-949 in Proceedings of the eighth international symposium on remote sensing of environment. Environmental Research Institute of Michigan, Ann Arbor, Mich.

National Academy of Sciences. 1970. Remote sensing with special reference to agriculture and forestry. National Academy of Sciences, Washington, D.C. 424 pp.

National Aeronautics and Space Administration (NASA). 1972. ERTS-1 data user's handbook. Goddard Space Center, Greenbelt, Md.

Odell, P. L., and J. P. Basu. 1975. Concerning several methods for estimating crop acreages using remote sensing data. Pages 1-25 in Statistical theory and methodology for remote sensing data analysis with special emphasis on LACIE. University of Texas at Dallas, Dallas, Texas.

Pace, W. H., and D. M. Detchmendy. 1973. A fast algorithm for the decomposing of multispectral data into mixtures. Pages 831-848 in F. Shahrokhi, ed. Remote sensing of earth resources, vol. 2. University of Tennessee Space Institute, Tullahoma, Tenn.

Reeves, R. G., editor-in-chief. 1975. Manual of remote sensing. American Society of Photogrammetry, Falls Church, Va. 2 vols.

Salvato, P. 1973. Iterative techniques to estimate signature vectors for mixture processing of multispectral data. Pages 3B-48-3B-62 in Symposium proceedings, machine processing of remotely sensed data. Laboratory for Applications of Remote Sensing, West Lafayette, Ind.

Sebestyen, G. S. 1962. Decision making processes in pattern recognition. Macmillan, New York. 162 pp.

Smedes, H. W., Hulstrom, R. L., and K. J. Ranson. 1975. The mixture problem in computer mapping of terrain: Improved techniques for establishing spectral signatures, atmospheric path radiance, and transmittance. Pages 1099-1159 in Proceedings of the NASA earth resources survey symposium, Houston, Texas. National Aeronautics and Space Administration. NASA X-58168.

Sposito, V. A. 1976. Introduction to linear and nonlinear programming. Iowa State University Press, Ames, Iowa. 269 pp.

Swain, P. H. 1972. Pattern recognition: A basis for remote sensing data analysis. LARS Information Note 111572. (Laboratory for Applications of Remote Sensing, West Lafayette, Ind.)

Teicher, H. 1961. Identifiability of mixtures. Ann. Math. Stat. 32(1): 244-248.

Thomson, F. J. 1973. Crop species recognition and mensuration in the Sacramento Valley. Pages 181-188 in S. C. Freden, E. P. Mercanti, and M. A. Becker, eds. Symposium on significant results obtained from the Earth Resources Technology Satellite-1, vol. 1, sec. A. National Aeronautics and Space Administration NASA SP-327.

Wolfe, J. H. 1970. Pattern clustering by multivariate mixture analysis. Mult. Behav. Res. 5(3): 329-350.

VII. ACKNOWLEDGMENTS

The author wishes to express appreciation to Professor Richard Carlson of the Agronomy Department for his helpful discussions and for generously lending data tapes and other materials. Thanks also go to Professor R. J. Lambert for his unlimited cooperation as Ames Lab Group Leader and senior administrator in the production of this paper, and especially to Professor V. A. Sposito, without whose help and guidance this work would never have been completed. Finally, the efficiency and perseverance of Letha Osmundson in the typing of this manuscript is gratefully acknowledged.