NOVEL USES OF QUADRATIC SURFACES FOR MEDICAL DIAGNOSIS*

N. DECLARIS, O. IJAOLA, R. W. NEWCOMB
University of Maryland, College Park, Maryland 20742

ABSTRACT

There are situations in medical diagnostics where conventional clustering and pattern recognition techniques are impractical or all together they fail to apply. For these cases the concept of "minimal ellipsoids" is introduced together with an approach for its utlization in disease classification and illness identification in a given patient.

I. Introduction

The use of information processing machines in medicine to process patient data, to recognize patterns and to screen and diagnose disease(1, 2) is increasing. Generally it is based on two broad approaches(3):
a) pattern recognition using probabilistic models and statistical analysis,
b) automatic classification schemes emphasizing accuracy. However, there are situations in clinical diagnosis where no collection of statistical data can provide the diagnostic precision needed to correctly identify the illness for therapeutic purposes(4). The reasons why in these situations statistical techniques are unsatisfactory can be well understood(5, 6). A deterministic approach is outline here for separating clinical and paraclinical data of patients into classes appropriate for diagnostic and therapeutic purposes. The heart of the approach is a novel use of quadratic surfaces(7, 8) for "fine classification" of diseases and their subsequent use to identify "specific illness" of a given "host".

We show that the approach is: (a) well formulateable mathematically (in terms of well posed problems for which solutions do exist) and (b) analytically numerically and computationally tractable.

The following is a problem which arises in medical (differential) diagnosis. From confirmed clinical past cases it is known that certain physiological measurements (symptom, observations, laboratory tests, etc.) receive particular sets of values for an identified "disease A" (abnormal state of the physiological system) and a different set of values for another distinct "disease B". Given that these data are available, how does a physician go about deciding whether a given patient whose phys-

---

*Supported in part by U.S. Public Health Service Grant, No. RRO-7042-11.

iological measurements he just obtained is "host" to "disease A" or to "disease B" or he is facing a new case that can not be decided upon by the present measurements and the past data?

The motivation and fundamental theoretical considerations which led to our mathematical formulation of this problem are best visualized geometrically in two dimensions. Thus our discussion is initially made with reference to the x-y plane, however it should perhaps be amphasized from the beginning that the practical cases of interest (for medical diagnosis) involve situations whose representations requires much higher dimensionality, of the order 20 to 200.

Thus, given N cases of known "disease A" they may be viewed as N points $(x_i^{\alpha}, y_i^{\alpha})$ on the x-y plane forming the set A while M cases of known "disease B" cases may be viewed as points (x_j^b, y_j^b) of a second set B. That is

$$(x_i^{\alpha}, y_i^{\alpha}) \epsilon A \qquad i = 1, 2, \ldots N \tag{1.a}$$

and

$$(x_j^b, y_j^b) \epsilon B \qquad j = 1, 2, \ldots M \tag{1.b}$$

under the assumption that sets A and B cluster near points

$$(x_o^{\alpha}, y_o^{\alpha}) \text{ and } (x_o^b, y_o^b)$$

respectfully, for convenience.

A simple and frequently mentioned technique is to enclose sets A and B each with a circle C_A and C_B defined by the radii r_A and r_B respectively. Then to decide whether a new case - a point (x^t, y^t) - is: 1) inside C_A, 2) inside C_B, or 3) outside C_A and of C_B.

Mathematically this amounts to the use of the quatratic form:

$$(x^t - x_o)^2 + (y^t - y_o)^2 \begin{cases} > r_A^2 & \text{outside } C_A \\ \leq r_A^2 & \text{inside } C_A \end{cases} \tag{2.a}$$

where

$$r_A^2 = \max_{\forall_i} \left[(x_i^{\alpha} - x_o)^2 + (y_i^{\alpha} - y_o)^2 \right] \tag{2.b}$$

to decide wheter the given point (x^t, y^t) is inside or outside the circle of radius r_A. This simple approach is sueful (affording much computational convenience) when the circles C_A and C_B do not intersect on the plane, Figure 1. Clearly the approach is of limited applicability if the two circles happen to intersect while the given data do not; Figure 2. While it is possible to pre-stretch the axis through a linear transformation, and then define the circles C_A and C_B this (eigenvalue scheme) does not always work for the same basic reason.

A much better appraoch is to enclose, if possible, the given data points (sets A and B) with non-intersecting ellipses, E_A and E_B, Figure 3, and then use these quadratic forms to test the point (x^t, y^t). The question then in two dimensions is: given a set of points $N \geq 3$, is it possible to find an ellipse E, containing all points and enclosing minimum area (points on E are considered as contained by the ellipse)?

It is not hard to show that this is equivalent to using the quadratic form (in the z-w plane, a transformation of the x-y plane):

$$\left(\frac{z}{d_1}\right)^2 + \left(\frac{w}{d_2}\right)^2 \begin{cases} > 1 & \text{outside } E \\ \leq 1 & \text{inside } E \end{cases} \qquad (3)$$

where d_1 and d_2 are the minimum values possible and defined by three suitable points of the given set.

In the next two sections we outline the much harder problems of:

1. establishing that this basic approach can be extended to n-dimensions, and
2. formulating practical algorithms (for computer implementations) for finding the "minimal ellipsoid" from given sets od data and then utilizing it for diagnosing a particular case under question.

There are still several other aspects that must be considered before the effectiveness of the approach can be taken full advantage of for general medical diagnosis purposes, such as updating the data, modification by learning, etc. However these problems, under active considerations are not taken up in this paper.

II. Mathematical Formulation

The data to be classified are considered, in the usual fashion[7,8] as n-vectors in the real domain R^n. If the scalar product of two vectors x, y is written as $\langle x, y \rangle$ and K is an nxn positive definite (real symmetric) matrix, the equation

$$\langle x, Kx \rangle \leq 1 \qquad x \in R^n \qquad (4)$$

defines an n-dimensional ellipsoid. The main idea is, given a set of data vectors:

$$\alpha^i \in R^n \qquad i = 1, 2, 3, \ldots, N \geq n \qquad (5)$$

there exists a "minimal ellipsoid" which contains the data and which has a maximum number of data points on its boundary. In fact it can be shown[9] that the minimal ellipsoid is completely determined by n+1 data points which must lie on its surface.

The situation is somewhat simplified by finding an orthongonal matrix

A which diagonalizes K so that equation (5) becomes:

$$\langle y, D^{-1} y \rangle \le 1 \tag{6.a}$$

where

$$AKA^{tr} = D^{-1} \tag{6.b}$$

and

$$D = \text{diag.}[d_i] \qquad i = 1, 2, \dots, n \tag{6.c}$$

That is we need to find suitable orthogonal coordinates in R^n (with reference to the data) to identify d_i. This is accomplished by an original axis translation and successive axis rotation (about the origin) ensuring that all given data vectors fall within the volume defined by the ellipsoid.

Thus given a set of α^i a new set is generated

$$\phi^i = \alpha^i - c_o \tag{7.a}$$

and it is reordered from

$$||\phi^i|| = n_i \tag{7.b}$$

to satisfy $n_1 \ge n_i$ $i = 1, 2, 3, \dots, N$, where $||\phi^i||$ denotes the usual norm. C_o is also chosen as to minimize the max n_i for all i. Next rotation of the axis takes place to align the first axis along ϕ_1. Then through suitable rotation and the Gram-Schmidt procedure a square matrix R is found such that

$$RR^{tr} = 1 \qquad \text{and} \tag{8.a}$$

$$R\phi^1 = y_1 = \begin{matrix} n_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{matrix} \tag{8.b}$$

we choose $d_1 = n_1^2$ and proceed in the same mannor to find the other coordinates. However, we must be careful to rotate all remaining components (after we extract the vector components along the selected axis) about the just selected axis and test using the sub ellipsoid (quadratic surface) for the coordinates selected. This is accomplished through the equation:

$$d_2 = \max_i \left\{ \frac{(n_i^{(2)})^2}{1 - [\phi_{i1}^{(2)})^2 / d_i]} \right\} \tag{9}$$

This process repeats itself by successive rotations and gives explicit expression for the overall transformation and the matrix $D^{(9)}$. The notation does get a bit complicated but the problem is straightforward.

III. Algorithmic Aspects

In this section an algorithm for constructing "minimal ellipsoids" from a given set of data is outlined in much simplified from in order to illuminate the basic features of the approach. Some other desirable features are also present, such as provisions for identifying possible multi-ellipsoidal surfaces and/or learning capability (perception like), but they are not discussed here.

The first part of the algorithm sets up suitable notation, Figure 4, where:

α^i represent confirmed patients as vectors with reference to measurement axis x_j, $j = 1, 2, \ldots (N+1)$ and $i = 1, 2, \ldots$ $M > N+1$

y_k represent unit vectors coordinate axis defining with reference to the minimal ellipsoid, $k = 1, 1, \ldots N$

β_k^i represent the confirmed patient vector data projected on the y_k defined axis $i = 1, 2, \ldots N+1$ and $k = 1, 2, \ldots N$.

Remark: The equation for the interior of the minimal ellipsoid is selected so as to satisfy:

$$\sum_{k=1}^{N} \left(\frac{\beta_k^i}{d_k} \right)^2 \leq 1 \qquad \text{for all } j \tag{10}$$

with d_k having the smallest possible values. The procedure is to select one y_k, test all β_k^i $(i = 1, 2, \ldots N+1)$ and choose d_k to satisfy the above inequality, then go back and select another y_{k+1} (orthogonal to y_k). This process is repeated until all y_k have been found.

The first axis (the major axis) of the ellipsoid is found by constructing the difference vectors

$$v^r = \alpha^i - \alpha^j \qquad r = 1, 2, \ldots, \frac{N(N+1)}{1}$$

The vector v^R is given by:

$$||v^R||^2 = (2d_1)^2 \geq ||v^r|| \text{ for all } r$$

establishes $y_1 = \frac{1}{2d_1} v^R$ and of course d_1.

The second part of the algorithm, Figure 5, has two branches. One branch performs the expansion

$$
\begin{aligned}
C^j &= \beta_1^j y_1 + z_1^j \\
z_1^j &= \beta_{2y_2}^j + z_2^j \\
&\cdots\cdots\cdots \\
z_{N+1}^j &= \beta_N^j y_N + z_N^j
\end{aligned} \tag{11}
$$

where $\langle y_k \cdot z_k^j \rangle = 0$. The other branches main subroutine is the SHRINK TEST where each z_{k-1}^j is tested via the inequality:

$$
\sum_{i=1}^{k-1} \left(\frac{\beta_i^j}{d_i}\right)^2 + \left(\frac{\|z_{k-1}^j\|}{(1-\rho\Delta)d_{k-1}}\right)^2 \begin{cases} > 1 & \text{outside} \\ \leq 1 & \text{inside} \end{cases} \tag{12}
$$

and ρ is increased one step at a time $\rho = 1, 2, \ldots, R$.

The $\hat{z}_{k-1}^j$ which ensures the maximum value of ρ_{mx} and all data points inside the ellipsoid determines y_k and d_k. That is:

$$
y_k = \frac{1}{\|\hat{z}_{k-1}^j\|} \hat{z}_{k-1}^j \tag{13}
$$

and

$$
d_k = (1 - \rho_{mx}\Delta)\, d_{k-1} \tag{14}
$$

Remark: $\Delta = \frac{d_1}{R}$ is a predetermined unit of axis "shrinkage", and R the total number of "shrinkage steps".

IV. Concluding Remarks

Various parts and slightly different versions of the above algorithm have been successfully tested computationally. The computer implementation and actual medical use of this approach involves additional considerations (computational efficiency, storage, accuracy, etc.) beyond the scope of this paper.

V. References

1. E. R. Gabrieli, "The Use of Data Mechanization and Computers in Clinical Medicine", N. Y., Acad. of Sci., 1969.

2. E. A. Patrick, et. al., "Review of Pattern Recognition in Medical Diagnosis and Consulting Relative to a New System Model", IEEE Syst., Man., Cybern., January 1974.

3. C. A. Kulikowski, "Pattern Recognition Approach to Medical Diagnosis", IEEE Trans. Syst., Sci., Cybern., July 1970.

4. A. R. Feinstein, "Clinical Judgement", The Williams and Wilkins Co., Baltimore, 1969.

5. L. Kanal, "Patterns in Pattern Recognition: 1968-1974", IEEE Trans. on Information Theory, November 1974.

6. K. Fukunaga, "Introduction to Statistical Pattern Recognition", Academic Press, New York, 1972.

7. N. J. Nilson, "Learning Machines", McGraw-Hill.

8. R. O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis", Wiley, New York, 1973.

9. O. Ijaola, Ph. D. Thesis, Graduate School, University of Maryland, (under preparation).

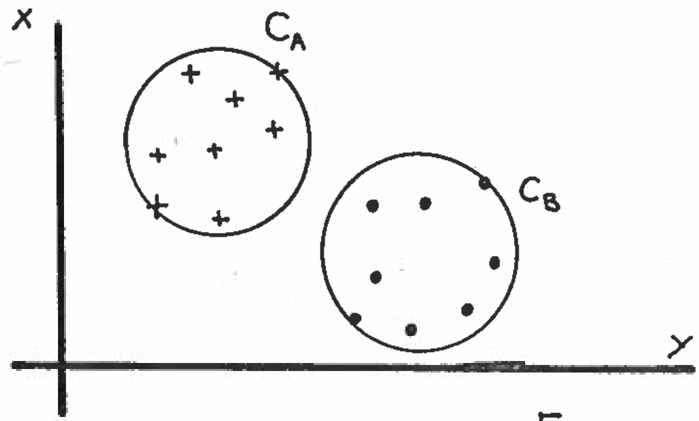


Figure 1

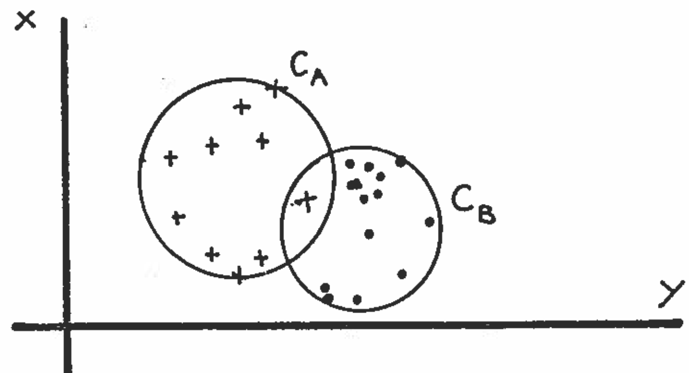


Figure 2

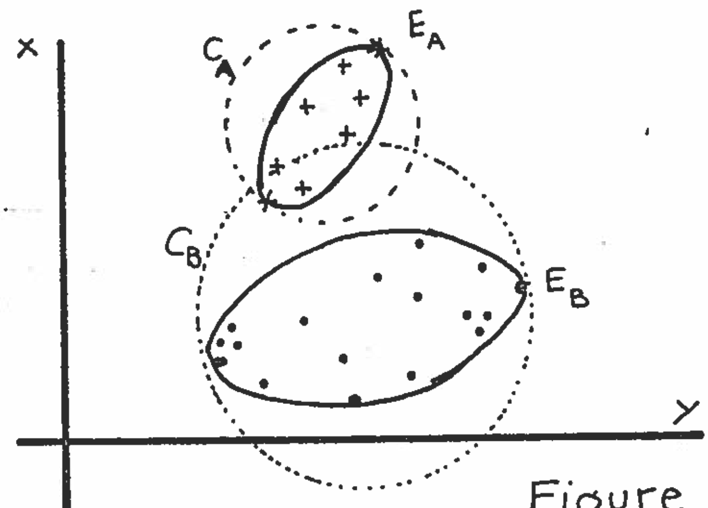


Figure 3

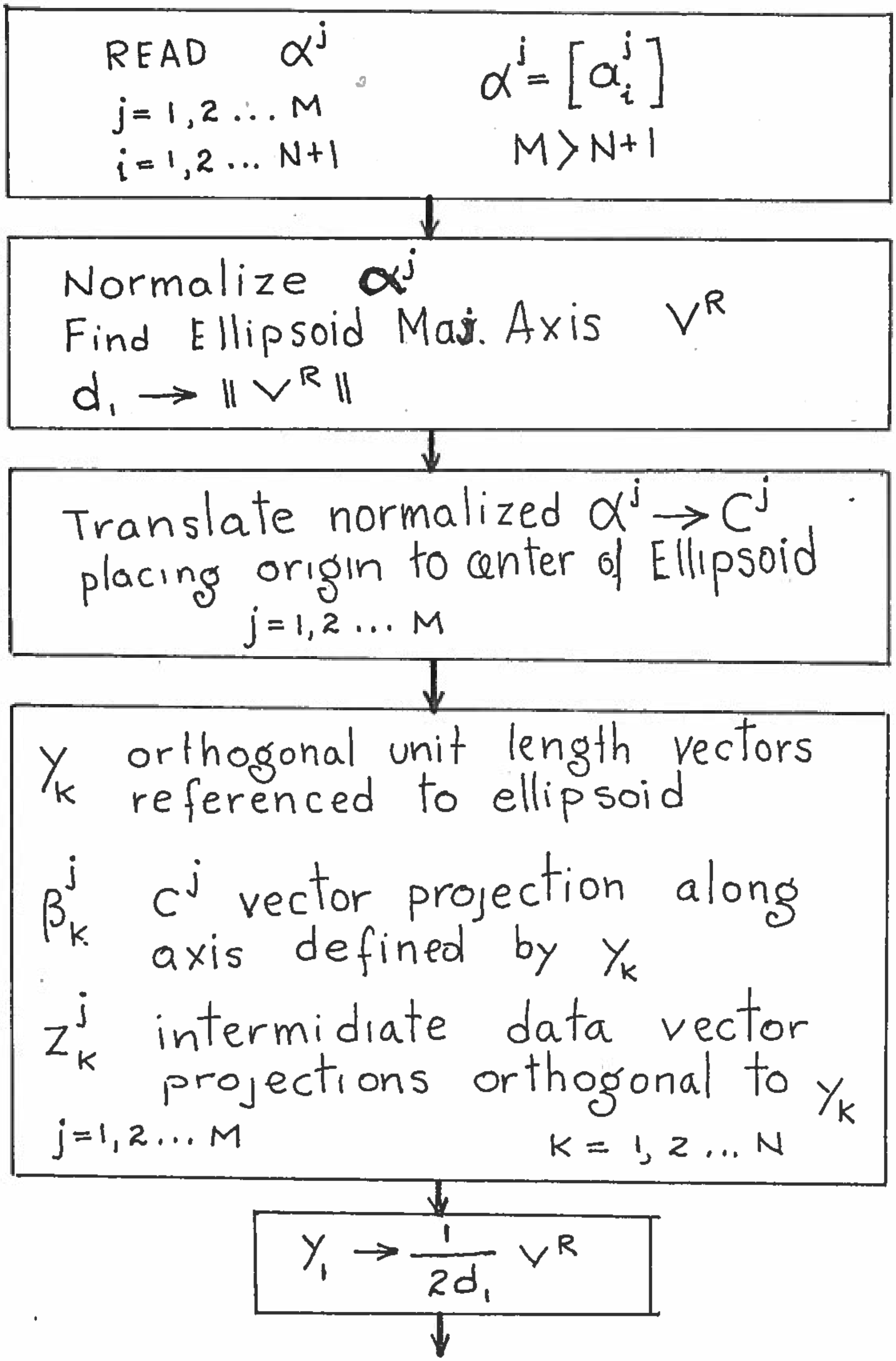
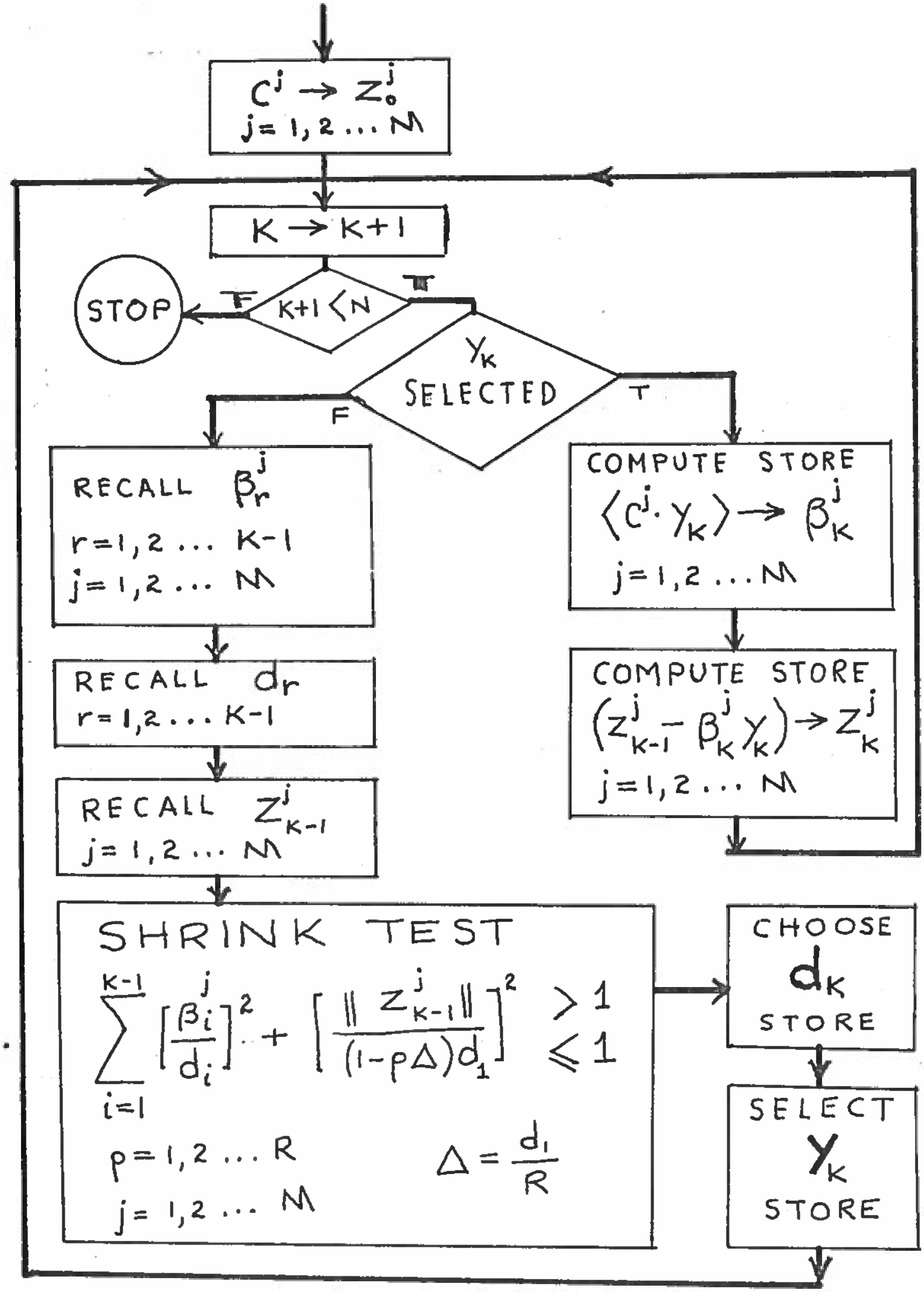

READ α^j
$j = 1, 2 \ldots M$
$i = 1, 2 \ldots N+1$
$\alpha^j = [a_i^j]$
$M > N+1$
Normalize α^j
Find Ellipsoid Maj. Axis V^R
$d_1 \rightarrow \| V^R \|$
Translate normalized $\alpha^j \rightarrow C^j$
placing origin to center of Ellipsoid
$j = 1, 2 \ldots M$
Y_k orthogonal unit length vectors referenced to ellipsoid
β_k^j C^j vector projection along axis defined by Y_k
Z_k^j intermidiate data vector projections orthogonal to Y_k
$j = 1, 2 \ldots M$
$K = 1, 2 \ldots N$
$Y_1 \rightarrow \frac{1}{2d_1} V^R$

Figure 4



$C^j \to Z_0^j$
$j = 1, 2 \ldots M$
$K \to K+1$
STOP
F
$K+1 < N$
T
Y_K SELECTED
F
T
RECALL β_r^j
$r = 1, 2 \ldots K-1$
$j = 1, 2 \ldots M$
COMPUTE STORE
$\langle C^j \cdot Y_K \rangle \to \beta_K^j$
$j = 1, 2 \ldots M$
RECALL d_r
$r = 1, 2 \ldots K-1$
COMPUTE STORE
$(Z_{K-1}^j - \beta_K^j Y_K) \to Z_K^j$
$j = 1, 2 \ldots M$
RECALL Z_{K-1}^j
$j = 1, 2 \ldots M$
SHRINK TEST
$\sum_{i=1}^{K-1} \left[\frac{\beta_i^j}{d_i}\right]^2 + \left[\frac{\| Z_{K-1}^j \|}{(1-\rho\Delta)d_1}\right]^2 \begin{matrix} > 1 \\ \leqslant 1 \end{matrix}$
$\rho = 1, 2 \ldots R$
$\Delta = \frac{d_1}{R}$
$j = 1, 2 \ldots M$
CHOOSE
d_K
STORE
SELECT
Y_K
STORE

Figure 5