CHAPTER 11
FACTOR TRANSFORMATIONS: ANALYTIC TRANSFORMATIONS

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 11

FACTOR TRANSFORMATIONS : ANALYTIC TRANSFORMATIONS

There has been a long felt need for automated factor transformation procedures both to eliminate the rather tedious graphical methods described in Chapter 10 and to provide objectivity in the results obtained. While the computations involved in the graphical transformation procedure can be programed readily for a computer and the factor graphs can be made mechanically, there remains the series of subjective judgments for each rotation. thus, the procedure had a stop and go characteristic which consumed considerable analysts time. Further, these judgments were sophisticated so that many individuals felt that they were not capable of making these judgments. A further and more serious complaint was the possible lack of objectivity in the results obtained. An important consideration is that the obtained results should not depend on the individual performaing the analysis. The same transformation for a particular factor matrix should be obtained by different analysts. That is: the transformation should not depend upon the analyst; it should be independent of the analyst so as to be repeatable at different laboratories. A variety of automated procedures will be discussed in this chapter.

Thurstone (1947) described several preliminary ideas for analytic approaches to simple structure. However, after some experimental trials these suggestions were not followed up. Several of the other developments refer to five criteria given by Thurstone in a section on "Uniqueness of simple structure in a given correlation matrix" (1947, pages 334-340). These criteria were to be applied after a transformation had been achieved to judge the uniqueness of the given solution. Further, attributes having complex factorial compositions were to be excluded from consideration. A preferable reference would be to Thurstone's suggsted equation for a simple structure (1947, pages 354-356). In Thurstone's notation, this equation is:

$$\phi = \sum_{j=1}^{n} \prod_{p=1}^{r} v_{jp}^{2} \tag{11.1}$$

where the v's , for Thurstone, are the projections on the normals. However, in terms of subsequent developments, the v's could be interpreted either as projections on the normals (structure loadings) in our matrix G or as factor weights (pattern loadings) in our matrix B . In order for zero to be the least value considered, the coefficients v are squared. A product is obtained for each attribute of the squared coefficients of that attribute. Then, a sum is obtained of these products over all attributes. Thus, a minimum of the function is such as to maximize the number of near zero factor coefficients. This equation may be considered, with adaptation, as the basis for a number of proposed analytic procedures. Carroll (1953) proposed a criterion which can be thouhgt of as an adaptation of equation (11.1). Instead of using the product of all coefficients for each attribute, Carroll used products of pairs of coefficients and obtained a sum of these products for each attribute. These sums of paired products were summed over all attributes. Let ϕ^* be Carroll's criterion, then:

$$\phi^* = \sum_{i=1}^{n} \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} v_{ij}^2 v_{jk}^2 \quad . \tag{11.2}$$

Other interpreted properties of simple structure were used by some of the proposed techniques. The general concern was to maximize the zero or near zero factor coefficients in accorance with the concept of simple structure.

11.1 Orthogonal Transformations

Several transformation procedures restrict the transfomations to being orthogonal. For this class of transformations the matrix of normals is restricted to being orthonormal so that:

$$FF' = I \quad . \tag{11.3}$$

The matrix of projections on the normals is given by:

$$AF' = G \quad . \tag{7.9}$$

With the orthogonality restriction, the communality of each attribute i is given by:

$$h_i^2 = \sum_{j=1}^{r} a_{ij}^2 = \sum_{k=1}^{r} g_{ik}^2 \quad . \tag{11.4}$$

Carroll's criterion may be written in terms of the projections on the normals (structure loadings):

$$\phi^* = \sum_{i=1}^{n} \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} g_{ij}^2 g_{ik}^2 \quad . \tag{11.5}$$

Several algebraic operations yield an interesting result which relates to other transformation criteria to be considered. The pairs of products of g's in equation (11.5) are as if the pairs below a diagonal of an $r \times r$ matrix. When all pairs of such a matrix, both sides of the diagonal as well as the diagonal, are considered equation (11.5) may be written as:

$$\phi^* = \frac{1}{2} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{r} \sum_{k=1}^{r} g_{ij}^2 g_{jk}^2 - \sum_{j=1}^{r} g_{ij}^4 \right\} \tag{11.6}$$

with equation (11.4) this equation may be simplified to:

$$\phi^* = \frac{1}{2} \left\{ \sum_{i=1}^{n} (h_i^2)^2 - \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^4 \right\} \tag{11.7}$$

A constant C and function Q can be defined as below:

$$C = \sum_{i=1}^{n} (h_i^2)^2 \quad . \tag{11.8}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^4 \quad . \tag{11.9}$$

Equation (11.7) becomes:

$$\phi^* = \tfrac{1}{2}C - \tfrac{1}{2}Q \quad . \qquad (11.10)$$

Since the communalities are fixed by the initial factor matrix, C is a constant. Then, Carroll's criterion ϕ^* is a minimum when function Q is a maximum. This relation equates the solution for Carroll's criterion to the Quartimax criterion which is to be discussed next.

Shortly following Carroll's development, Saunders (1935) proposed a criterion for orthogonal transformations of factors based on the statistical concept of kurtosis of a distribution. First, Saunders constructed an extended factor matrix by appending vertically the given factor matrix with the negative of the given factor matrix. Thus, for any factor, original or transformed, the distribution of factor coefficients would be symmetric with mean zero. Each attribute would have two coefficients on any factor, a positive coefficient and a negative coefficient. Since such a distribution would be symmetric, the distribution would have zero skewness. Saunders observed for simple structure there would be a large number of attribute coefficients near zero with a smaller number of attribute coefficients trailing out to the tails of this distribution. Such a distribution should have a high value of kurtosis. Saunders proposed criterion S as:

$$S = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^4 \right\} \Bigg/ \left\{ \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^2 \right\}^2 \quad . \qquad (11.11)$$

With the preceding equations Saunders' criterion becomes:

$$S = Q/C \qquad (11.12)$$

which, since C is a constant, is maximized by maximizing function Q .

Neuhaus and Wrigley (1954) took a different statistical approach from that of Saunders. In order to eliminate the algebraic signs of the factor coefficients they considered the squared coefficients. They suggested that the transformation should result in a large number of small squared coefficients and a number of large squared coefficients. This result would occur with a large variance of the distribution of transformed, squared coefficients. Their function may be designated by the letter N and can be expressed as: .

$$N = \tfrac{1}{nr} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{r} (g_{ij}^2)^2 - \tfrac{1}{nr} \left[\sum_{i=1}^{n} \sum_{j=1}^{r} (g_{ij}^2) \right]^2 \right\} \qquad (11.13)$$

With the preceding equations the Neuhaus and Wrigley function becomes:

$$N = \{(nr)Q - C\} \big/ (nr)^2 \quad . \qquad (11.14)$$

Since C is a constant, N is maximized by maximizing function Q .

Ferguson (1954) discussed factor transformations in terms of information theory and proposed that a function equivalent to Q would lead to a parsimonious result.

Since function Q involves the fourth power of the transfomed factor coefficients it became known as the Quartimax criterion. Computer programs were developed to maximize this function and experimental trials were run with a variety of original factor matrices. When the simple structure was strong and almost orthogonal the results were quite acceptable. However, when the configuration of attribute vectors was relatively complex and oblique the Quartimax solution was not acceptable. In thess types of cases the quartimax axes tended to be near the principal axes of the configuration. These conditions are illustrated by a small, artificial example. This example starts from a constructed factor weight matrix B given in Table 11.1 . There is a moderate simple structure with attributes 1 and 2 loading only on factor 1 while attributes 3 and 4 load only on factor 2 . Attributes 3 , 4 , and 5 are complex and span the space from factor 1 to factor 2 . Several examples are produced by choosing different correlations between the factors. This correlation is ϕ_{12} and is given values of .0 , .3 , .6 and .9 for the examples run. For the given matrix B and each of the chosen correlations between the factors, a transformation of the configuration was made to the principal axes of the configuration. This resulted in a matrix A for each of the examples. Table 11.2 presents the matrix A and the Quartimax transformed matrix for each of the chosen values of ϕ_{12} . For ϕ_{12} equal to .0 the Quartimax solution is quite near the input matrix B . When ϕ_{12} is .3 the Quartimax solution approaches the input matrix A . The trend increases markedly as the value of ϕ_{12} increases to .9 in which case the Quartimax matrix is almost identical with the input matrix A .

Following the experience with the Quartimax solution Kaiser (1958) proposed the Varimax orthogonal transformation. Kaiser pointed out that each of the Quartimax criteria worked to simplify the rows of the factor matrix. He proposed to concentrate on the columns of the transformed matrix with a criterion for each column j of the transformed matrix:

$$v_j = \left\{ n\sum_{i=1}^{n}(g_{ij}^2)^2 - \left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \Big/ n^2 \qquad (11.15)$$

For a general function Kaiser summed the single column functions v_j over all columns of the transformed matrix to obtain function V .

$$V = \sum_{j=1}^{r}\left\{ n\sum_{i=1}^{n}(g_{ij}^2)^2 - \left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \Big/ n^2$$

or

$$V = \left\{ n\sum_{j=1}^{r}\sum_{i=1}^{n} g_{ij}^4 - \sum_{j=1}^{r}\left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \Big/ n^2 \quad . \qquad (11.16)$$

A further suggestion by Kaiser was that the rows of the input factor matrix should be normalized to unit length vectors before applying the Varimax solution. He noted that some transformed

Table 11.1

Input Matrix B for Artificial Example
for Orthogonal Transformations

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

Table 11.2

Quartimax Transformation for Artificial Example

	$\phi_{12} = .0$		$\phi_{12} = .3$		$\phi_{12} = .6$		$\phi_{12} = .9$	
Input Matrix A								
	1	2	1	2	1	2	1	2
1	.54	.45	.58	.39	.64	.30	.68	.15
2	.38	.32	.42	.28	.45	.21	.49	.11
3	.38	-.46	.47	-.38	.53	-.28	.58	-.14
4	.26	-.31	.31	-.25	.35	-.19	.39	-.09
5	.51	.17	.57	.15	.63	.12	.68	.06
6	.56	-.05	.64	-.03	.72	-.02	.78	-.01
7	.62	-.27	.72	-.21	.80	-.16	.88	-.08
Quartimax Transformed Matrix								
	1	2	1	2	1	2	1	2
1	.70	.06	.41	.57	.63	.31	.68	.15
2	.50	.04	.29	.41	.45	.23	.49	.11
3	-.05	.60	.57	-.19	.54	-.27	.58	-.14
4	-.04	.40	.38	-.13	.36	-.18	.39	-.09
5	.48	.24	.48	.34	.63	.14	.68	.07
6	.36	.43	.61	.20	.72	.00	.78	.00
7	.25	.62	.75	.05	.81	-.13	.88	-.07

factors had few high loadings which resulted in reduced weighting for these factors. In order to counter this underweighting he suggested the normalization prodcedure in order to equalize the weights for the various transformed factors. To accomplish this step let H be a diagonal matrix containing the square roots of the communalities of the attributes. Let $\tilde{A}$ contain the normalized input factor matrix determined by:

$$\tilde{A} = H^{-1}A \quad . \tag{11.17}$$

Application of the Varimax procedure to $\tilde{A}$ results in a matrix $\tilde{G}$ for the normalized vectors. Return to the original lengthed vectors can be accomplished by:

$$G = H\tilde{G} \quad . \tag{11.18}$$

Transformation without the normalization procedure is termed the Raw Varimax solution. When the normalization procedure is used the solution is termed the Normal Varimax procedure.

Table 11.3 gives the results for the artificial example for the Raw Varimax and the Normal Varimax solutions. For the value of ϕ_{12} equal to .0 the Raw Varimax solution had a small rotation from the input B matrix while the Normal Varimax solution exactly reproduced the input B matrix. At ϕ_{12} equal to .3 the Raw varimax solution interchanged the two transformed factors. However, these factors are quite recognizable by low loadings for attributes 1 and 2 for factor 1 and for attributes 3 and 4 for factor 2 . Note that the obliqueness has resulted in the input loadings of zero now being positive for the transformed factors. Common practice is to interpret small loadings as zero. For the Normal Varimax the loadings of these previously zero loadings are small positives. However, the transformed factors are quite identifiable to the input matrix B factors. By the time ϕ_{12} becomes .6 the previously small loadings have increased. For ϕ_{12} equals .9 the Raw Varimax solution has approached the principal factors of the input matrix A . The Normal Varimax solution balances the two input B factors with the previously zero loadings having become moderately positive. This balance of the factors for the Normal Varimax is shown by attribute 6 , which started out in the input matrix B in Table 11.1 having equal loadings on the two factors. For the Normal Varimax solution for all values of ϕ_{12} this attribute 6 has equal loadings on the two transformed factors.

Figure 11.1 shows the locations of the transformed axes for $\phi_{12} = .6$ for all three orthogonal methods of tranformation discussed. The configuration of the seven attribute vector terminals are plotted. Note that the first Quartimax axis, Q_1 , is almost identical with the first principal axis, A_1. This is the property of the Quartimax transformation discussed previously. The configuration of attribute points lies inside the two Normal Varimax axes. The Raw Varimax axes have drifted from the Normal Varimax axes toward the input principal axes. Axis R_1 has moved from axis N_1 toward input axis A_1. This figure illustrates some of the relative properties of the three methods of orthogonal transformations discussed in this section.

Table 11.3

Varimax Transformations for Artificial Example

	$\phi_{12} = .0$		$\phi_{12} = .3$		$\phi_{12} = .6$		$\phi_{12} = .9$	
	Input Matrix A							
	1	2	1	2	1	2	1	2
1	.54	.45	.58	.39	.64	.30	.68	.15
2	.38	.32	.42	.28	.45	.21	.49	.11
3	.38	-.46	.47	-.38	.53	-.28	.58	-.14
4	.26	-.31	.31	-.25	.35	-.19	.39	-.09
5	.51	.17	.57	.15	.63	.12	.68	.06
6	.56	-.05	.64	-.03	.72	-.02	.78	-.01
7	.62	-.27	.72	-.21	.80	-.16	.88	-.08
	Raw Varimax Transformed Matrix							
	1	2	1	2	1	2	1	2
1	.70	.06	.17	.68	.31	.63	.67	.20
2	.50	.04	.12	.49	.22	.45	.48	.14
3	-.05	.60	.60	.04	.59	.11	.59	-.10
4	-.04	.40	.40	.03	.39	.07	.39	-.07
5	.48	.24	.32	.50	.42	.48	.68	.11
6	.36	.43	.50	.41	.57	.43	.78	.04
7	.25	.62	.67	.33	.72	.38	.88	-.02
	Normal Varimax Transformed Matrix							
	1	2	1	2	1	2	1	2
1	.70	.00	.69	.11	.66	.22	.59	.37
2	.50	.00	.49	.08	.47	.16	.42	.27
3	.00	.60	.09	.59	.19	.57	.32	.51
4	.00	.40	.06	.40	.13	.38	.21	.34
5	.50	.20	.52	.28	.54	.35	.53	.44
6	.40	.40	.46	.46	.51	.51	.55	.55
7	.30	.60	.39	.64	.48	.66	.57	.67

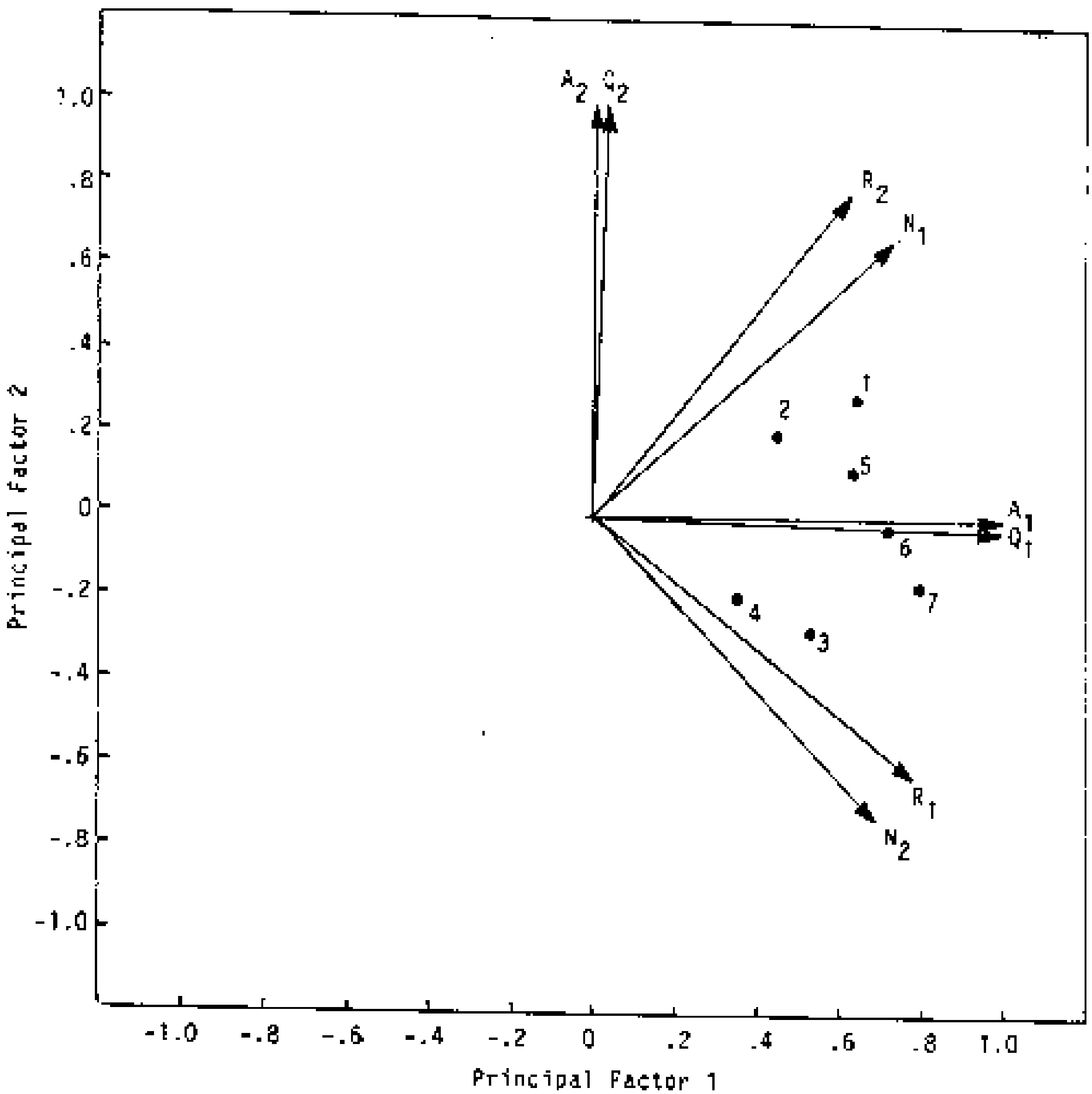

Figure 11.1 Factor Plot for Artificial Example, $\phi_{12} = .6$, showing axes:

A_1, A_2 for Input Matrix A;

N_1, N_2 for Normal Varimax Transformation;

Q_1, Q_2 for Quartimax Transformation;

R_1, R_2 for Raw Varimax Transformation.

Results for transformation of the Nine Mental Tests Example by the analytic orthogonal transformation methods are given in Table 11.4. These results may be compared to the projections on the normals obtained by least squares hyperplane fitting in Table 10.11. Since the present results are restricted to orthogonal transformations they can not be as clean as the LSQHYP results that are not so restricted. However, the three transformed factors are quite recognizable. Even so, the interpretations can not be as clean and sharp as with the oblique transformation.

With the addition of a parameter γ the Quartimax and Varimax functions can be combined to a single function ξ which we may term <u>Orthomax</u> .

$$\xi = \left\{ n\sum_{j=1}^{r}\sum_{i=1}^{n} g_{ij}^4 - \gamma\sum_{j=1}^{r}\left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \quad . \tag{11.19}$$

The denominator n^2 of the Varimax function has been dropped as unessential. for the Quartimax, γ equals 0 ; for the Varimax, γ equals 1 . There have been suggestions for other values of γ ; however, these have not lead to practical use. Computational procedures are based on function ξ .

An iterative type plan is followed in the computations which transform orthogonally each pair of factors in turn. These computations cycle through all pairs of factors solving for the minimum ξ for each pair. In each cycle of computations, a minimum is found for each pair of factors. These transformations are considered as conditional since, for any pair of factors, the transformations of other pairs will affect the solution for the given pair. Consequently, the cycles must be repeated until there are no significant transformations for all pairs. Consider the pair of factors k and m . The function may be written for this pair as:

$$\xi_{km} = n\left\{\sum_{i=1}^{n} g_{ik}^4 + \sum_{i=1}^{n} g_{im}^4\right\} - \gamma\left\{\left[\sum_{i=1}^{n} g_{ik}^2\right]^2 + \left[\sum_{i=1}^{n} g_{im}^2\right]^2\right\} \quad . \tag{11.20}$$

An orthogonal transformation of factors k and m is accomplished through an angle θ . The new factor coefficients are:

$$\tilde{g}_{ik} = g_{ik}\cos\theta + g_{im}\sin\theta \quad ; \tag{11.21a}$$

$$\tilde{g}_{im} = -g_{ik}\sin\theta + g_{im}\cos\theta. \tag{11.21b}$$

The transformed function is:

$$\tilde{\xi}_{km} = n\left\{\sum_{i=1}^{n} \tilde{g}_{ik}^4 + \sum_{i=1}^{n} \tilde{g}_{im}^4\right\} - \gamma\left\{\left[\sum_{i=1}^{n} \tilde{g}_{ik}^2\right]^2 + \left[\sum_{i=1}^{n} \tilde{g}_{im}^2\right]^2\right\} \quad . \tag{11.22}$$

After algebraic and trigonometric operations this function may be written as:

$$\tilde{\xi}_{km} = K + \tfrac{1}{4}C \bullet \cos 4\theta + \tfrac{1}{4}E \bullet \sin 4\theta \tag{11.23}$$

where K , C and E are temporary coefficients defined by:

Table 11.4

Orthogonal Analytic Transformations
Nine Mental Tests Example

Transformed Factor Matrices									
	Quartimax			Raw Varimax			Normal Varimax		
	1	2	3	1	2	3	1	2	3
1	.59	.09	-.15	.61	.08	.08	.61	.07	.12
2	.73	.00	.00	.69	-.03	.26	.67	-.05	.30
3	.55	.36	-.05	.53	.34	.19	.53	.32	.23
4	.03	.71	-.03	.04	.71	.06	.05	.71	.09
5	.06	.79	.02	.05	.78	.13	.06	.78	.16
6	.09	.72	.01	.08	.71	.13	.09	.70	.15
7	.56	.07	.34	.40	.01	.53	.37	-.01	.55
8	.33	.43	.40	.16	.37	.54	.14	.35	.56
9	.44	.31	.38	.28	.25	.54	.25	.23	.57

$$K = n\left\{ \tfrac{1}{2}\sum_{i=1}^{n}[g_{ik}^2 + g_{im}^2]^2 + \tfrac{1}{4}\sum_{i=1}^{n}[g_{ik}^2 - g_{im}^2]^2 + \sum_{i=1}^{n}[g_{ik}g_{im}]^2 \right\}$$
$$- \gamma\left\{ \tfrac{1}{2}\left[\sum_{i=1}^{n}(g_{ik}^2 + g_{im}^2)\right]^2 + \tfrac{1}{4}\left[\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)\right]^2 + \left[\sum_{i=1}^{n}g_{ik}g_{im}\right]^2 \right\} ; \qquad (11.24)$$

$$C = n\left\{ \sum_{i=1}^{n}[g_{ik}^2 + g_{im}^2]^2 - 4\sum_{i=1}^{n}g_{ik}^2 g_{im}^2 \right\} - \gamma\left\{ \left[\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)\right]^2 - 4\left[\sum_{i=1}^{n}g_{ik}g_{im}\right]^2 \right\}; \qquad (11.25)$$

$$E = 4\left\{ n\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)(g_{ik}g_{im}) - \gamma\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)\sum_{i=1}^{n}g_{ik}g_{im} \right\} \qquad (11.26)$$

The desired maximum of $\tilde{\xi}_{km}$ is obtained by setting the first derivative of $\tilde{\xi}_{km}$ equal to zero.

$$\frac{d\tilde{\xi}_{km}}{d\theta} = -C \bullet \sin 4\theta + E \bullet \cos 4\theta = 0 \quad . \qquad (11.27)$$

However, equation (11.27) only gives the condition for an optimum solution which may be either a minimum or a maximum. The second derivative must be investigated to determine the maximum solution.

$$\frac{d^2\tilde{\xi}_{km}}{d\theta^2} = -4C \bullet \cos 4\theta - 4E \bullet \sin 4\theta = -4[C \bullet \cos 4\theta + E \bullet \sin 4\theta] \quad .$$

For a maximum this second derivative must be negative which results in the following inequality.

$$[c \bullet \cos 4\theta + E \bullet \sin 4\theta] > 0 \quad . \qquad (11.28)$$

The desired solution must satisfy equations (11.27) and (11.28).

There are four conditions to be considered in terms of whether or not each of the coefficients C and E equals zero.

1) If $C = 0$ and $E = 0$,
   The angle of rotation is indeterminate and may be set at 0. Then:
   $\cos\theta = 1$; $\sin\theta = 0$.
2) If $C = 0$ and $E \neq 0$,
   from (11.27), $\cos 4\theta = 0$ so that $\sin 4\theta = \pm 1$; by (11.28), when $E > 0$, $\sin 4\theta = +1$ which results in $4\theta = \pi/2$ and $\theta = \pi/8$ so that:
   $\cos\theta = .923880$; $\sin\theta = .382683$.
   When $E < 0$, $\sin 4\theta = -1$ which results in $4\theta = -\pi/2$ and $\theta = -\pi/8$ so that:
   $\cos\theta = .923880$; $\sin\theta = -.382683$.
3) If $C \neq 0$ and $E = 0$,
   from (11.27), $\sin 4\theta = 0$ so that $\cos 4\theta = \pm 1$; by (11.28), when $c > 0$, $\cos 4\theta = +1$ which results in $4\theta = 0$ and $\theta = 0$ so that:
   $\cos\theta = 1$; $\sin\theta = 0$.

When $C < 0$, $\cos 4\theta = -1$ which results in $4\theta = \pi$ and $\theta = \pi/4$ so that:

$\cos\theta = .707107$; $\sin\theta = .707107$.

4) If $C \neq 0$ and $E \neq 0$,

algebraic operations on equation (11.27) yields:

$E \bullet \sin 4\theta = C \bullet \sin 4\theta$

so that:

$$\frac{E}{C} = \frac{\sin 4\theta}{\cos 4\theta} = \tan 4\theta \quad . \qquad (11.30)$$

The inequality of equation (11.28) needs to be investigated in selection of a maximum solution. From (11.27)

$$\sin 4\theta = \frac{E}{C}\cos 4\theta$$

which may be substituted in (11.28) to obtain:

$$\frac{E^2}{C}\cos 4\theta + C \bullet \cos 4\theta > 0$$

or

$$\frac{(C^2 + E^2)}{C}\cos 4\theta > 0 \quad . \qquad (11.31)$$

Since $(C^2 + E^2)$ must be positive, the $\cos 4\theta$ must have the same algebraic sign as coefficient C . This rule helps pick out the maximum solution. Let T be a temporary coefficient defined by the positive value of:

$$T = \sqrt{1 + \tan^2 4\theta} \quad .$$

If $C > 0$, $\cos 4\theta = \frac{1}{T}$; (11.32a)

if $C < 0$, $\cos 4\theta = \frac{-1}{T}$. (11.32b)

The remainder of the solution follows. Note that positive square roots are appropriate in the following steps.

$$\sin 4\theta = \cos 4\theta \bullet \tan 4\theta \quad ; \qquad 11.33)$$

$$\cos 2\theta = \sqrt{\frac{1 + \cos 4\theta}{2}} \quad ; \qquad (11.34)$$

$$\sin 2\theta = \frac{\sin 4\theta}{2 \bullet \cos 2\theta} \quad ; \qquad (11.35)$$

$$\cos\theta = \sqrt{\frac{1 + \cos 2\theta}{2}} \quad ; \qquad (11.36)$$

$$\sin\theta = \frac{\sin 2\theta}{2 \bullet \cos\theta} \quad . \qquad (11.37)$$

The desired transformation is given in equations (11.36) and (11.37). The foregoing steps are readily programmed into an efficient computer subroutine.

11.2 Oblique Algebraic Transformations, OBLIMIN

Oblique algebraic transformations started with Carroll's (1953) proposal which was expressed as equations (11.2) and (11.5) without the orthogonality restriction. Dur to the relation with the Quartimax procedure this unrestricted propossal has become termed the Quartimin procedure. Kaiser in his unpublished Ph.D. dissertation (University of California at Berkeley, 1956) suggested an oblique counterpart to his Varimax criterion which is termed the Covarimin criterion. In experimental trials by Carroll and Kaiser the Quartimin procedure yielded transformed factors that were overly oblique for the trial factor matrices they concidered. In contrast, the Covarimin procedure yielded transformed factors that were not adequately oblique. A compromize suggested by Carroll (1957) was to take a transformation that was half way between the Quartimin and the Covarimin transformations. This lead to equation (11.38) with the use of parameter γ in a manner similar to equation (11.19).

$$\xi = \sum_{k=1}^{r-1} \sum_{m=k+1}^{r} \left\{ n\sum_{i=1}^{n} g_{ik}^2 g_{im}^2 - \gamma \left[\sum_{i=1}^{n} g_{ik}^2\right] \left[\sum_{i=1}^{n} g_{im}^2\right] \right\} \quad . \tag{11.38}$$

The three values of γ considered are given below along with interpretations.

Quartimin: $\gamma = 0$ said to be most oblique;
Biquartimin; $\gamma = .5$ said to be less oblique;
Covarimin: $\gamma = 1.0$ said to be least oblique.

Solutions for these three oblique criteria for the Nine Mental Tests example are given in Tables 11.5 and 11.6. The computations follow an iterative scheme to be discussed later may be started from any selected initial transformation. In Table 11.5 an intial transformation was performed with a Normal Varimax solution. Note for $\gamma = 0.0$ the cosines of the angles of normals 1 and 2 with normal 3 are very negative; that is: very oblique. This obliqueness has reduced for $\gamma = 0.5$. However, for $\gamma = 1.0$ the cosine of the angless among the normals have turned positive so that the trend is not from most oblique to least oblique but from most negative cosines of angles among the normals to positive cosines among the angles. Table 11.6 illustrates a common problem experienced early with a variety of studies. Note that the cosine of the angle between transformed factor 1 and 2 has become 1.00 in the iteration. These two tranformed factors have become identical. This is a most unsatisfactory result. This problem occurs also for $\gamma = 1.0$. Another property of the solution is illustrated with these two tables. The obtained results depend upon the starting position. As a consequence of the problems illustrated here, this approach to oblique algebraic transformations is not considered for practical use.

An elemental step in the solution to minimize ξ is to transform normal k (row k of F) in the direction of normal m (row m of F) so that ξ is reduced as much as possible. A cycle of computations will involve taking each normal in turn as k and to transform it in the direction of each other normal in turn as normal m . These cycles are continued until there are no significant transformations for any normal. The oblimin criterion can be written for normal k as;

Table 11.5

Structure OBLIMIN Transformation
Nine Mental Tests Example
Start from Normal Varimax Solution

$\gamma = 0.0$ Quartimin	$\gamma = 0.5$ Biquartimin	$\gamma = 1.0$ Covarimin

Projections on Normals

	$\gamma = 0.0$ Quartimin			$\gamma = 0.5$ Biquartimin			$\gamma = 1.0$ Covarimin		
	1	2	3	1	2	3	1	2	3
1	.46	.07	-.12	.60	.04	-.05	.61	.09	.28
2	.42	-.10	.03	.67	-.09	.12	.72	-.01	.46
3	.34	.24	-.03	.50	.30	.05	.55	.35	.38
4	.02	.59	-.04	-.01	.70	-.02	.04	.71	.12
5	-.01	.62	.02	-.01	.77	.05	.07	.79	.20
6	.02	.56	.01	.03	.69	.04	.09	.71	.20
7	.04	-.22	.37	.36	-.05	.44	.48	.05	.63
8	-.16	.07	.41	.11	.32	.46	.25	.41	.59
9	-.07	-.03	.39	.23	.20	.45	.36	.29	.62

Cosines of Angles Between Normals

	$\gamma = 0.0$ Quartimin			$\gamma = 0.5$ Biquartimin			$\gamma = 1.0$ Covarimin		
	1	2	3	1	2	3	1	2	3
1	1.00	.36	-.79	1.00	-.11	-.25	1.00	.01	.47
2	.36	1.00	-.59	-.11	1.00	-.16	.01	1.00	.13
3	-.79	-.59	1.00	-.25	-.16	1.00	.47	.13	1.00

$\gamma = 0.0$ Quartimin	$\gamma = 0.5$ Biquartimin	$\gamma = 1.0$ Covarimin
8 Cycles	14 Cycles	14 Cycles
$\xi = .24025$	$\xi = -1.10194$	$\xi = -4.15831$

Table 11.6

Structure OBLIMIN Transformation
Nine Mental Tests Example
Start from Principal Fators

	γ = 0.0 Quartimin				γ = 0.5 Biquartimin				γ =1.0 Covarimin		

Projections on Normals

	1	2	3		1	2	3		1	2	3
1	.33	.32	-.06	1	.34	.34	.05	1	.61	.09	.28
2	.23	.22	.14	2	.24	.25	.27	2	.72	-.01	.46
3	.21	.21	-.08	3	.23	.23	-.02	3	.55	.35	.38
4	.03	.04	-.31	4	.05	.04	-.39	4	.04	.71	.12
5	-.01	.00	-.29	5	.01	.00	-.37	5	.07	.79	.20
6	.01	.01	-.26	6	.03	.02	-.33	6	.09	.71	.20
7	-.15	-.16	.39	7	-.14	-.14	.43	7	.48	.05	.63
8	-.29	-.29	.25	8	-.27	-.27	.21	8	.25	.41	.59
9	-.23	-.23	.30	9	-.21	-.21	.29	9	.36	.29	.62

Cosines of Angles Between Normals

	1	2	3		1	2	3		1	2	3
1	1.00	1.00	-.78	1	1.00	1.00	-.62	1	1.00	.01	.47
2	1.00	1.00	-.80	2	1.00	1.00	-.60	2	.01	1.00	.13
3	-.78	-.80	1.00	3	-.62	-.60	1.00	3	.47	.13	1.00

50 Cycles*	50 Cycles*	18 Cycles
ξ = .50114	ξ = .16493	ξ = -4.15831

* Convergence was not achieved in 50 cycles.

$$\xi_k = \sum_{\substack{j=1 \\ j\neq k}}^{r} \left\{ n\sum_{i=1}^{n} g_{ik}^2 g_{ij}^2 - \gamma \left[\sum_{i=1}^{n} g_{ik}^2\right] \left[\sum_{i=1}^{n} g_{ij}^2\right] \right\} \quad . \tag{11.39}$$

Let q_{ik} be a coefficient for each attribute i in terms of normal k :

$$q_{ik} = \sum_{\substack{j=1 \\ j\neq k}}^{r} g_{jk}^2 \quad . \tag{11.40}$$

Then equation (11.39) may be written as:

$$\xi_k = n \sum_{i=1}^{n} [g_{i\ ik}^2 q_{ik}] - \gamma \left\{\sum_{i=1}^{n} g_{ik}^2\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad . \tag{11.41}$$

Let $\underline{f}_k$ be row k of matrix F (the row for normal k) and $\underline{f}_m$ be row m of matrix F (the row for normal m). As in graphical transformations, an unnormalized, transformed normal k may be noted as $\underline{f}_k^*$ which is obtained by:

$$\underline{f}_k^* = \underline{f}_k + z_{km}\underline{f}_m \tag{11.42}$$

where z_{km} is the raw shift coefficient. Normalization of $\underline{f}_k^*$ to the new, normalized normal, $\tilde{\underline{f}}_k$ yields:

$$\tilde{\underline{f}}_k = \{\underline{f}_k + z_{km}\underline{f}_m\} \Big/ \sqrt{1 + 2z_{km}p_{km} + z_{km}^2} \tag{11.43}$$

where z_{km} is the cosine of the angle between normals k and m . The projections on the tranformed normal are:

$$\tilde{g}_{ik} = \{g_{ik} + z_{km}g_{im}\} \Big/ \sqrt{1 + 2z_{km}p_{km} + z_{km}^2} \quad . \tag{11.44}$$

Coefficient z_{km} is the km cell of matrix P defined by:

$$P = FF' \quad . \tag{11.45}$$

The transformed criterion is $\tilde{\xi}_{km}$ which is obtained by substitution of $\tilde{g}_{ik}$ for the g_{ik} in equation (11.41):

$$\tilde{\xi}_{km} = n\sum_{i=1}^{n} [\tilde{g}_{ik}^2 q_{ik}] - \gamma \left\{\sum_{i=1}^{n} \tilde{g}_{ik}^2\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad . \tag{11.46}$$

To obtain a minimum of $\tilde{\xi}_{km}$ the derivative with respect to z_{km} is set equal to zero; this yields:

$$\frac{d\tilde{\xi}_{km}}{dz_{km}} = \frac{a + bz_{km} + cz_{km}^2}{[1 + 2p_{km}z_{km} + z_{km}^2]^2} = 0 \tag{11.47}$$

where:

$$a = n\sum_{i=1}^{n} [g_{im}g_{ik} - g_{ik}^2 p_{km}]q_{ij} - \gamma \left\{\sum_{i=1}^{n} [g_{im}g_{ik} - g_{ik}^2 p_{km}]\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad ; \tag{11.48}$$

$$b = n\sum_{i=1}^{n} [g_{im}^2 - g_{ik}^2]q_{ik} - \gamma \left\{\sum_{i=1}^{n} [g_{im}^2 - g_{ik}^2]\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad ; \tag{11.49}$$

$$c = n\sum_{i=1}^{n} [g_{im}^2 p_{km} - g_{ik}g_{im}]q_{ik} - \gamma \left\{\sum_{i=1}^{n} [g_{im}^2 p_{km} - g_{ik}g_{im}]\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad . \tag{11.50}$$

Since the denominator of (11.47) is necessarily positive (being a square) and approaches infinity only when z_{km} approaches infinity (which would make (11.47) approach zero), attention is given to the numerator equalling zero for the derivative to equal zero. Let y designate the numerator:

$$y = a + bz_{km} + cz_{km}^2 \quad . \tag{11.51}$$

If $c = 0$,

$$z_{km} = -a/b \quad . \tag{11.52}$$

If $c \neq 0$, the roots of equation (11.51 are given by the quadratic equation:

$$z_{km} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \quad . \tag{11.53}$$

There is a problem as to which sign to use. Investigation of the nature of the function in equation (11.46) indicates that the positive sign is always correct so that:

$$z_{km} = \frac{-b + \sqrt{b^2 - 4ac}}{2c} \quad . \tag{11.54}$$

The preceding material provides a basis for a computer program to obtain a solution for the Oblimin function. However, the obtained result may depend on the starting position as illustrated in Tables 11.5 and 11.6. There is no guarantee that there is a single minimum.

11.3 Oblique Transformations, DIRECT OBLIMIN

The DIRECT OBLIMIN transformations were initiated by Jennrich and Sampson (1966) in their article on rotation to simple structure in which they adapted Carrol's criterion to involving the squared factor weights (pattern loadings) in our matrix B. Let the Jennrich and Sampson criterion be designated J and be:

$$J = \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} \sum_{i=1}^{n} b_{ij}^2 b_{ik}^2 \quad . \tag{11.55}$$

This criterion is to be minimized. As described previously, the Oblimin procedures yielded unsatisfactory results with some data matrices in that the transformation yielded singular normals matrices and factor matrices. A dimension had been lost. Jennrich and Sampson overcame this problem in that the b's approach infinity any time the factor correlation matrix approaches singularity. Jennrich and Sampson preferred to apply a transfomration, T , directly to the factor scores which results in applying the inverse of this transformation to the factor weight matrix; thus:

$$B = AT^{-1} \quad . \tag{7.4}$$

Other relevant equations from Chapter 7 folow.

$$R_{bb} = TT' \quad . \tag{7.1}$$

$$D = \{\text{Diag } R_{bb}^{-1}\}^{-1/2} \quad . \tag{7.8}$$

Then:

$$B = GD^{-1} \quad . \tag{11.56}$$

When the factor correlation matrix becomes singular due to highly oblique factors, the diagonal entries in the inverse become very large as do the diagonal entries in D^{-1}. Then, as indicated previously, the entries in the factor weight matrix B also become very large. This works against the minimization of the direct oblimin function

The Direct Oblimin function has been expanded to involve a parameter γ similar to the expansion of the Quartimin criterion to the Oblimin criterion given in equation (11.38). The expanded criterion may be designated ψ and is given below:

$$\psi = \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} \left\{ n \sum_{i=1}^{n} b_{ij}^2 b_{ik}^2 - \gamma \left[\sum_{i=1}^{n} b_{ij}^2 \right] \left[\sum_{i=1}^{n} b_{ik}^2 \right] \right\} \quad . \tag{11.57}$$

Table 11.7 gives results of the Direct Oblimin solution for the Nine Mental Tests Example. Values of γ range from -1.0 to 0.5 . In contrast to the Oblimin situation in which positive values of γ gave the better results, preferable results are obtained for the Direct Oblimin criterion with negative values of γ . In fact, the computations broke down with a value of γ of $+1.0$. With γ equal to 0.5, there were very few near zero factor loadings which was accompanied by high factor correlations. Jennrich (1979), in a study on the admissible values of γ in direct oblimin rotation, indicated that there might be trouble in trying to use a positive value of γ . A minimum value of ψ might not exist. He recommended that positive values of γ be avoided. The simple loadings of Jennrich and Sampson have γ equal to 0 . Note that there are only small changes for this example in the near zero loadings as γ progresses from 0 to -1.0 . However, the factor correlations decrease in this range. A phenomena not shown in this example is that when very negative values of γ are used the transformed factors approach the principal factors. A recomendation is that several small negative values of γ be tried and a selection made for the best results.

Comments on analysis procedures follow. In equation (11.57) ψ is the sum, as it were, of entries on one side of a square matirx. A useful alternative equation takes one half of the sum of all cells of such a matrix less the diagonal cells. Thus, an alternative equation for ψ is:

$$\psi = \tfrac{1}{2} \left\{ n \left[\sum_{i=1}^{n} \left\{ \sum_{k=1}^{r} b_{ik}^2 \right\}^2 - \sum_{i=1}^{n} \sum_{k=1}^{r} b_{ik}^4 \right] - \gamma \left[\left\{ \sum_{i=1}^{n} \sum_{k=1}^{r} b_{ik}^2 \right\}^2 - \sum_{k=1}^{r} \left\{ \sum_{i=1}^{n} b_{ik}^2 \right\}^2 \right] \right\} \tag{11.58}$$

An elemental step in solution to minimize ψ is to transform trait vestor k (row k of T) in the direction of trait vector m (row m of T) so as to reduce ψ as much as possible. There will be a number of cycles with each trait vector in turn being k which is paired with each other vector m in turn. The series of cycles is continued until only trivial transformations are performed

Table 11.7

Direct OBLIMIN Transformations
Nine Mental Tests Example

$\gamma = -1.0$				$\gamma = -.5$				$\gamma = 0.0$				$\gamma = 0.5$			
Factor Weight Matrices B															
	…1	2	3		1	2	3		1	2	3				
	…2	.06	-.02	1	.63	.06	-.03	1	.65	.06	-.08	1	1.08	-.08	-.50
	…5	-.10	.18	2	.65	-.10	.17	2	.65	-.11	.15	2	1.02	-.44	-.04
	…0	.30	.09	3	.50	.30	.08	3	.51	.29	.05	3	.73	.24	-.28
	…1	.72	-.01	4	.01	.72	-.01	4	.02	.72	-.03	4	-.15	1.05	-.33
	…01	.77	.06	5	-.01	.77	.06	5	-.01	.77	.05	5	-.24	1.09	-.20
	…3	.70	.06	6	.03	.70	.05	6	.03	.70	.04	6	-.15	.97	-.20
	…5	-.10	.53	7	.24	-.12	.54	7	.19	-.17	.57	7	.08	-.56	.98
	…01	.27	.55	8	-.02	.25	.56	8	-.08	.20	.61	8	-.48	.04	1.04
	…1	.14	.55	9	.10	.13	.56	9	.05	.08	.60	9	-.23	-.18	1.00
Factor Correlation Matrices R_{bb}															
	…1	2	3		1	2	3		1	2	3		1	2	3
	…00	.11	.45	1	1.00	.12	.48	1	1.00	.13	.59	1	1.00	.73	.91
	…1	1.00	.29	2	.12	1.00	.32	2	.13	1.00	.40	2	.73	1.00	.84
	…5	.29	1.00	3	.48	.32	1.00	3	.59	.40	1.00	3	.91	.84	1.00
$\psi = 5.667$				$\psi = 3.278$				$\psi = 0.800$				$\psi = -4.534$			

.

within a cycle.

Let $\underline{t}_k$ and $\underline{t}_m$ be trait vectors k and m . Let $\tilde{\underline{t}}_k$ be the new trait vector k which is defined by:

$$\tilde{\underline{t}}_k = s_{kk}\underline{t}_k + s_{km}\underline{t}_m \tag{11.59}$$

where s_{kk} and s_{km} are the transformation coefficients. A restriction that $\tilde{\underline{t}}_k$ be a unit vector leads to the following restriction on the trasformation coefficients.

$$s_{kk}^2 + 2s_{kk}s_{km}\phi_{km} + s_{km}^2 = 1 \tag{11.60}$$

where ϕ_{km} is the correlation between factors k and m , that is, the km cell of matrix R_{bb} given in equation (7.1). Let $\tilde{T}$ be the new traits matrix given by:

$$\tilde{T} = ST \tag{11.61}$$

where the transformation matrix S has s_{kk} in the kk cell and s_{km} in the km cell with all other off-diagonal cells being zero and all other diagonal cells being unity. Let $\tilde{B}$ be the new matrix of factor weights which may be obtained by an equation similar to equation (7.4).

$$\tilde{B} = A\tilde{T}^{-1} \quad . \tag{11.62}$$

Algebraic operations yield:

$$\tilde{B} = BS^{-1} \quad . \tag{11.63}$$

Let $\underline{t}^k$ and $\underline{t}^m$ be the k'th and m'th columns of T^{-1} . Similarly, let $\tilde{\underline{t}}^k$ and $\tilde{\underline{t}}^m$ be the k'th and m'th columns of $\tilde{T}^{-1}$. Then equation (11.62) yields:

$$\tilde{\underline{t}}^k = \frac{1}{s_{kk}}\underline{t}^k \quad ; \tag{11.64}$$

$$\tilde{\underline{t}}^m = \frac{-s_{km}}{s_{kk}}\underline{t}^k + \underline{t}^m \quad . \tag{11.65}$$

Let $\underline{b}_k$ and $\underline{b}_m$ be the k'th and m'th columns of matrix B. Similarly, let $\tilde{\underline{b}}_k$ and $\tilde{\underline{b}}_m$ be the k'th and m'th columns of matrix $\tilde{B}$. Then, from equation (11.63);

$$\tilde{\underline{b}}_k = \frac{1}{s_{kk}}\underline{b}_k \quad ; \tag{11.66}$$

$$\tilde{\underline{b}}_m = \frac{-s_{km}}{s_{kk}}\underline{b}_k + \underline{b}_m \quad . \tag{11.67}$$

For $j \neq k, m$:

$$\tilde{\underline{b}}_j = \underline{b}_j \quad . \tag{11.68}$$

Let:

$$z_{km} = \frac{s_{km}}{s_{kk}} \quad . \tag{11.69}$$

From equations (11.60) and (11.69):

$$\frac{1}{s_{kk}^2} = 1 + 2z_{km}\phi_{km} + z_{km}^2 \quad . \tag{11.70}$$

From the preceding equations:

$$\tilde{b}_{ik}^2 = b_{ik}^2\{1 + 2z_{km}\phi_{km} + z_{km}^2\} \quad ; \tag{11.71}$$

$$\tilde{b}_{im}^2 = z_{km}^2 b_{ik}^2 - 2z_{km}b_{ik}b_{im} + b_{im}^2 \quad . \tag{11.72}$$

In terms of the transformed criterion $\tilde{\psi}_{km}$ can be written as:

$$\tilde{\psi}_{km} = \tfrac{1}{2}\{nu_n - \gamma u_\gamma\} \tag{11.73}$$

where:

$$u_n = \sum_{i=1}^{n}\left[\tilde{b}_{ik}^2 + \tilde{b}_{im}^2 + \sum_{\substack{j=1 \\ j\neq k,m}}^{r} b_{ij}^2\right]^2 - \sum_{i=1}^{n}\left[\tilde{b}_{ik}^4 + \tilde{b}_{im}^4 + \sum_{\substack{j=1 \\ j\neq k,m}}^{r} b^4 ij\right] \quad ; \tag{11.74.n}$$

$$u_\gamma = \left\{\sum_{i=1}^{n}\left[\tilde{b}_{ik}^2 + \tilde{b}_{im}^2 + \sum_{\substack{j=1 \\ j\neq k,m}}^{r} b_{ij}^2\right]\right\}^2 - \left\{\left[\sum_{i=1}^{n}\tilde{b}_{ik}^2\right]^2 + \left[\sum_{i=1}^{n}\tilde{b}_{im}^2\right]^2 + \sum_{\substack{j=1 \\ j\neq k,m}}^{r}\left[\sum_{i=1}^{n} b_{ij}^2\right]^2\right\} \quad . \tag{11.74.γ}$$

With substitution from equations (11.71) and (11.72) into equation (11.73) the criterion $\tilde{\psi}_{km}$ after algebraic operations may be written as:

$$\tilde{\psi}_{km} = c_0 + c_1 z_{km} + c_2 z_{km}^2 + c_3 z_{km}^3 + c_4 z_{km}^4 \quad . \tag{11.75}$$

Coefficient c_0 equals the untransformed criterion obtained from equations (11.73), (11.74.n) (11.74.ψ) using the untransformed b's instead of the transformed $\tilde{b}$'s. Coefficients c_1 through c_4 are obtained by the following equations. For notational conveience in the present context let q_i be defined by:

$$q_i = \sum_{\substack{j=1 \\ j\neq k,m}}^{r} b_{ij}^2 \quad , \tag{11.76}$$

For $p = 1, 2, 3,$ and 4, let:

$$c_p = \tfrac{1}{2}\{nc_{pn} + \gamma c_{p\gamma}\} \quad . \tag{11.77}$$

$$c_{1n} = 4\phi_{km}\left\{\sum_{i=1}^{n} b_{ik}^2 b_{im}^2 + \sum_{i=1}^{n} b_{ik}^2 q_i\right\} - 4\left\{\sum_{i=1}^{n} b_{ik}^3 b_{im} + \sum_{i=1}^{n} b_{ik} b_{im} q_i\right\} \quad . \tag{11.78.1n}$$

$$c_{1\gamma} = 4\phi_{km}\left\{\left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{im}^2\right] + \left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} q_i\right]\right\} - 4\left\{\left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{ik} b_{im}\right] + \left[\sum_{i=1}^{n} b_{ik} b_{im}\right]\left[\sum_{i=1}^{n} q_i\right]\right\} \quad . \tag{11.78.1γ}$$

$$c_{2n} = 2\left\{\sum_{i=1}^{n} b_{ik}^4 + \sum_{i=1}^{n} b_{ik}^2 b_{im}^2 + 2\sum_{i=1}^{n} b_{ik}^2 q_i - 4\phi_{km}\sum_{i=1}^{n} b_{ik}^3 b_{im}\right\} \quad ; \tag{11.78.2n}$$

$$c_{2\gamma} = 2\left\{\left[\sum_{i=1}^{n} b_{ik}^2\right]^2 + \left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{im}^2\right] + 2\left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} q_i\right]\right. \tag{11.78.2γ}$$

$$c_{3n} = 4\left\{\phi_{km}\sum_{i=1}^{n} b_{ik}^4 - \sum_{i=1}^{n} b_{ik}^3 b_{im}\right\} \quad ; \tag{11.78.3n}$$

$$c_{3\gamma} = 4\left\{\phi_{km}\left[\sum_{i=1}^{n} b_{ik}^2\right]^2 - \left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{ik} b_{im}\right]\right\} \tag{11.78.3γ}$$

$$c_{4n} = 2\sum_{i=1}^{n} b_{ik}^4 \quad ; \tag{11.78.4n}$$

$$c_{4\gamma} = 2\left[\sum_{i=1}^{n} b_{ik}^2\right]^2 \quad . \tag{11.78.4γ}$$

The minimum of $\tilde{\psi}_{km}$ may be obtained by equating the dreivative os $\tilde{\psi}_{km}$ with respect to z_{km} equal to zero and finding the roots of the resulting equation.

$$\tilde{\psi}_{km} = c_1 + \tfrac{1}{2}c_2 z_{km} + \tfrac{1}{3}c_3 z_{km}^2 + \tfrac{1}{4}c_4 z_{km}^3 = 0 \quad . \tag{11.79}$$

A first special case to be considered involves a negative value for c_4 which, from inspection of equations (11.77), (11.78.4n) and (11.78.4γ), will occur only for positive values of γ. As previously noted, undesirable transformations may occur with positive values of γ. Therefore, a conclusion should be reached to exclude the solution whenever c_4 is negative. Otherwise, there are three zero roots to equation (11.79) whith the middle root being for a maximum $\tilde{\psi}_{km}$ and the least root and greatest root being for minimum values of $\tilde{\psi}_{km}$. The value of $\tilde{\psi}_{km}$ is to be determined for these two roots and the one yielding the lesser value of $\tilde{\psi}_{km}$ is to be chosen as the desired solution .

11.4 Orthoblique Transformations

Harris and Kaiser (1964) made a most interesting and stimulating contribution to factor transformations in their article on Oblique Factor Analysis Solutions by Orhtogonal Transformations. In this contribution they presented a general theory with two special cases that might be used with given original factor matrices. They showed that all transformations, orthogonal or oblique, could be generated with orthogonal transformations. In describing this contribution, as much as reasonable of their notation will be used.

Some preliminary matters are to be considered first. Let A_0 be any original factor matrix obtained by any of the factor extraction methods. Harris and Kaiser start from a factor matrix with the principal factors identification. It is not necessary for A_0 to be a principal factors matrix. These preliminary matters concern a transformation from any A_0 to a matrix A with the principal factors identification. Two procedures are to be considered: first the procedure followed by Harris and Kaiser and second an equivalent procedure which facilitates the application of any transformation applied to A to be applicable to the original matrix A_0 . Harris and Kaiser used matrix R^* defined by:

$$A^* = A_0 A_0' \quad . \tag{11.80}$$

The principal factors solution is obtained for R^* as in Chapter 8 by obtaining an eigen solution which is symbolized here as, folowing the Harris and Kaiser notation:

$$R^* = QM^2Q' \quad . \tag{11.81}$$

Q is an $n \times r$ matrix of eigenvectors and, thus is orthonormal by columns and M^2 is a diagonal matrix of eigenvalues. Note the use of the square for the eigenvalues; this is to facilitate following equations. Then, the principal factors matrix A is given by:

$$A = QM \quad . \tag{11.82}$$

An alternative procedure yields a transformation of A_0 to A . This is desirable to be able to apply all transformations on A to be applicable to A_0 . An eigen solution is obtained for the product matrix $A_0'A_0$:

$$A_0'A_0 = V_0\Lambda_0V_0' \quad . \tag{11.83}$$

Then:

$$A = A_0V_0 \quad ; \tag{11.84}$$

$$M^2 = \Lambda_0 \quad . \tag{11.85}$$

It can be shown that:

$$AA' = R^* \quad ;$$

$$A'A = \Lambda_0 = M^2 \quad . \tag{11.86}$$

Then:

$$Q = AM^{-1} \quad . \tag{11.87}$$

A following relation is that:

$$Q'Q = I \quad .$$

These relations are necessary for the results of this procedure to yield identical results to the first procedure.

Let $\tilde{T}$ be any $r \times r$, nonsingular transformation matrix in the set of all possible such matrices. Note that the rows of $\tilde{T}$ are not normalized. Let, in a more general sense than equation (7.4):

$$A\tilde{T}^{-1} = \tilde{B} \quad . \tag{11.88}$$

Note that:

$$A_0V_0\tilde{T}^{-1} = \tilde{B} \quad . \tag{11.89}$$

Let:

$$V_0\tilde{T}^{-1} = \tilde{T}_0^{-1} \quad ;$$

$$\tilde{T}_0 = \tilde{T}V_0' \quad . \tag{1190}$$

Then:

$$A_0\tilde{T}_0^{-1} = \tilde{B} \quad . \tag{11.91}$$

Since V_0 is square, nonsingular, $\tilde{T}_0$, as given in equation (11.90), is an element in the set of all possible $r \times r$, nonsingular matrices. Any generated $\tilde{T}$ yields a $\tilde{T}_0$ which may be applied to A_0 to yield $\tilde{B}$, as per equation (11.91).

Let $\tilde{\Phi}$ be a generalized factor covariance matrix defined, in parrallel whith equation (7.1), by:

$$\tilde{T}\tilde{T}' = \tilde{\Phi} \quad . \tag{11.92}$$

The rows of $\tilde{T}$ are normalized to T by:

$$T = D_1^{-1}\tilde{T} \tag{11.93}$$

where D_1 is defined by:

$$D_1^2 = \mathrm{Diag}(\tilde{\Phi}) \quad . \tag{11.94}$$

In the present context, the factor correlation matrix is symbolized by Φ and is given by equation (7.1) as:

$$TT' = \Phi \quad . \tag{11.95}$$

From (11.92) and (11.93):

$$\tilde{\Phi} = D_1 \Phi D_1 \quad . \tag{11.96}$$

The factor weight matrix for normalized factors is given in equation (7.4):

$$AT^{-1} = B \quad . \tag{7.4}$$

With equations (11.88) and (11.93), equation (7.4) yields:

$$\tilde{B} = BD_1^{-1} \quad . \tag{11.97}$$

The preceding discussion has focused on the trnaformational relations involving an original factor matrix A_0 and a principal factors matrix A as well as generalized transformations in $\tilde{T}$ with row wise normalized transformations in T . Next to be considered are the orthoblique transformations suggested by Harris and Kaiser. They presented the esscence of their suggextion in equation (2) (1964 page 350). We find that this understanding this development is enhanced by considering an Eckart-Young decomposition (Eckart and Young 1936; Johnson 1963) of matrix $\tilde{T}$:

$$\tilde{T} = T_1' D_2^{-2} T_2' \tag{11.98}$$

where T_1 and T_2 are r x r orthonormal matrices; D_2 is an r x r , positive, nonsingular diagonal matrix. In the following developments it is important to remember with T_1 and T_2 being square orthogonal matrices that:

$$T_1T_1' = T_1'T_1 = I \quad ; \tag{11.99.1}$$

$$T_2T_2' = T_2'T_2 = I \quad . \tag{11.99.2}$$

A reversal of directions is considered to develop $\tilde{T}$ from matrices T_1 , T_2 and D_2 . To develop all possible $\tilde{T}$, orthonormal matrices T_1 and T_2 range independently over all elemenst in the set of all possible r x r orthonormal matrices. Also, diagonal matrix D_2 ranges over all possible postive, nonsingular, r x r diagonal matrices. For special cases, one or two of these matrices may be set at prescribed values. Matrix $\tilde{\Phi}$ is obtained from equations (11.92) and (11.98):

$$\tilde{\Phi} = T_1' D_2^{-1} T_2' T_2 D_2^{-1} T_1 = T_1' D_2^{-1} T_1 \quad . \tag{11.100}$$

Matrices D_1 , T and Φ may be obtained from equations (11.94), (11.93) and (11.96). The foregoing material provides the essence of the Harris-Kaiser orthoblique transformations in that all transformations, orthogonal or oblique, can be generated from orthogonal transformations T_1 and T_2 along with diagonal matrix D_2 . Special cases are to be considered next.

A special category of transformations is that of orthogonal transformations. To produce orthogonal transformations matrix D_2 is set to an identity matrix. Equation (11.100) reduces to:

$$\tilde{\Phi} = T_1' I T_1 = I \quad . \tag{11.101}$$

Then:

$$\tilde{T} = T_1' T_2' \quad . \tag{11.102}$$

Since the product of orthonormal matrices is also an orthonormal matrix, $\tilde{T}$ is an orthonormal matrix. Since $\tilde{\Phi}$ is an identity matrix, D_1 is an identity matrix so that T equals $\tilde{T}$ and Φ equals an identity matrix. Thus, the obtained matrix T may be used in equation (7.4) to yield orthogonally transformations to factor weight matrix B .

Harris and Kaiser suggested that a special class of solutions might yield interesting practical procedures. In this class the number of parameters in equation (11.98) is reduced by defining:

$$T_2 = I \quad ; \tag{11.103}$$

$$D_2^{-1} = M^p \quad . \tag{11.104}$$

Then, from equation (11.98):

$$\tilde{T} = T_1' M^p \tag{11.105}$$

and from equation (11.100):

$$\tilde{\Phi} = T_1' M^{2p} T_1 \quad . \tag{11.106}$$

Also, from equations (11.88) and (11.105):

$$\tilde{B} = A M^{-p} T_1 = Q M M^{-p} T_1 \quad . \tag{11.107}$$

With these definitions, harris and Kaiser considered several cases generated from definitions of p .

A first case involves $p = 0$. With the convention that $M^0 = I$, $\tilde{T} = T_1'$ and $\tilde{\Phi} = I$. Thus, this case reduces to the orthogonal transformations discussed in a preceding paragraph.

A second case involves $p = 1/2$. From equations (11.105), (11.106) and (11.107):

$$\tilde{T} = T_1' M^{1/2} \quad ; \tag{11.108}$$

$$\tilde{\Phi} = T_1' M T_1 \quad ; \tag{11.109}$$

$$\tilde{B} = A M^{-1/2} T_1 \quad . \tag{11.110}$$

For this case, the product matrix $B'B$ is of interest. From equation (11.110)

$$\tilde{B}'\tilde{B} = T_1' M^{-1/2} A' A M^{-1/2} T_1$$

which becomes with equation 11.86)

$$\tilde{B}'\tilde{B} = T_1' M T_1 \quad . \tag{11.111}$$

Note that $\tilde{B}'\tilde{B}$ equals $\tilde{\phi}$. Using equations (11.96) and (11.97), equation (11.111) yields:

$$B'B = D_1^2 \Phi D_1^2 \quad . \tag{11.112}$$

This case can be termed the Φ proportional to $B'B$ case. (Note: we use B where Harris and Kaiser used A.) Examples of transformation using this case are described in a subsequent paragraph.

A third case involves p = 1 . From equations (11.105), (11.106) and (11.107):

$$\tilde{T} = T_1' M \quad ; \tag{11.113}$$

$$\tilde{\Phi} = T_1' M^2 T_1 \quad ; \tag{11.114}$$

$$\tilde{B} = A M^{-1} T_1 \quad . \tag{11.115}$$

As for the preceding case, the product matrix $\tilde{B}'\tilde{B}$ is of interest. From equation (11.115):

$$\tilde{B}'\tilde{B} = T_1' M^{-1} A' A M^{-1} T_1$$

which yields with equations (11.86) and ((11.99.1):

$$\tilde{B}'\tilde{B} = I \tag{11.116}$$

so that:

$$B'B = D_1^2 \tag{11.117}$$

which is a diagonal matrix. Note, from equation (11.82):

$$\tilde{B} = Q T_1 \quad . \tag{11.118}$$

so that the orthogonal tranxformation T_1 is applied directly to matrix Q of eigenvectors. Thus, for all possible orthogonal transformations, $B'B$ will be a diagonal matrix. As discussed in a subsequent paragraph, this case is termed the Independent Clusters case.

Examples of the Orthoblique transformations are given in Tables 11.8 and 11.9. Table 11.8 gives results for the artificial example used to illustrate the Orthomax tranformations with the parameter matrix B given in Table 11.1. The illustrations in Table 11.8 use the parameter ϕ_{12} = .6. The Quartimax transformation was chosen to yield matrix T_1 for each of the cases of Orthoblique transformations . The input parameter matrix B is repeated from Table 11.1 at the left of Table 11.8 for comparison purposes. With p = 0 the Orthogonal transformation is identical to that given in Table 11.2 for ϕ_{12} = .6 . The parameter ϕ_{12} = .6 was chosen to provide an example of the change in results using the Φ Proportional to $B'B$ case for which p = 1/2 . The results obtained are moderately near to the input parameters. There has been a distinct improvement over the Orthogonal transformation results. For p = 1, the independent clusters case, the results have deteriorated. This is not surprising since the input configuration was not a

Table 11.8

Orthoblique Transformations for Artificial Example, $\phi_{12} = .6$
Using Quartimax Transformations

	Input Parameters		p = 0 Orthogonal		p = 1/2 Φ Proportional To B'B		p = 1 Independent Clusters	

Factor Weight Matrices B

	1	2	1	2	1	2	1	2
1	.70	.00	.63	.31	.61	.15	.77	-.09
2	.50	.00	.45	.23	.44	.11	.55	-.06
3	.00	.60	.54	-.27	-.04	.62	-.20	.75
4	.00	.40	.36	-.18	-.03	.41	-.14	.50
5	.50	.20	.63	.14	.42	.32	.48	.19
6	.40	.40	.72	.00	.32	.50	.30	.45
7	.30	.60	.81	-.13	.22	.69	.13	.71

Matrices B'B

	1	2	1	2	1	2	1	2
1	1.24	.44	2.56	.07	1.41	.55	1.57	.00
2	.44	1.08	.07	.29	.55	.90	.00	1.28

Matrices $R_{bb} = \Phi$

	1	2	1	2	1	2	1	2
1	1.00	.60	1.00	.00	1.00	.49	1.00	.80
2	.60	1.00	.00	1.00	.49	1.00	.80	1.00

Table 11.9

Orthoblique Transformations for Nine Mental Tests Example
Using Normal Varimax Transformation

p = 0.0 Orthogonal	p = 1/2 Φ Proportional To B'B	p =1.0 Independent Clusters

Factor Weight Matrices B

	1	2	3	1	2	3	1	2	3
1	.61	.07	.12	.61	.05	.00	.70	.01	-.13
2	.67	-.05	.30	.65	-.10	.19	.68	-.18	.12
3	.53	.32	.23	.50	.29	.12	.54	.25	.02
4	.05	.71	.09	.01	.71	.02	.02	.73	-.05
5	.06	.78	.16	.00	.77	.09	-.01	.78	.04
6	.09	.70	.15	.04	.69	.09	.04	.70	.03
7	.37	-.01	.55	.27	-.10	.53	.15	-.22	.61
8	.14	.35	.56	.01	.27	.55	-.14	.18	.66
9	.25	.23	.57	.13	.14	.55	.00	.04	.64

Matrices B'B

	1	2	3	1	2	3	1	2	3
1	1.32	.41	.84	1.13	.13	.41	1.28	.00	.00
2	.41	1.88	.68	.13	1.76	.33	.00	1.80	.00
3	.84	.68	1.15	.41	.33	.95	.00	.00	1.26

Matrices $R_{bb} = \Phi$

	1	2	3	1	2	3	1	2	3
1	1.00	.00	.00	1.00	.09	.39	1.00	.24	.69
2	.00	1.00	.00	.09	1.00	.25	.24	1.00	.46
3	.00	.00	1.00	.39	.25	1.00	.69	.46	1.00

independent clusters situation. Note, matrix $B'B$ is a diagonal matrix for $p = 1$. This agrees with equation 11.117. In proposing the Φ Proportional to $B'B$ case, $p = 1/2$, Harris and Kaiser (1964) noted that this was approximately true for a large number of examples of real data. This case works moderately well in a variety of studies.

An extreme case of simple structure has a cluster of attributes for each transformed factor with no intermediate attributes Thus, since each attribute has a loading on one and only one transformed factor, the product of paired loadings for each pair of transformed factors will eaual zero so that the off-diagonal entries in $B'B$ will be zero so that $B'B$ will be a diagonal matrix. This corresponds to the result in equation (11.117) so that the third special case applies. The transformation problem is to develop transformation matrix T_1 . Any of the Orthomax procedures described earlier appear to be usable. A Quartimax transformation was used in Table 11.8.

Table 11.9 presents results for the Nine Mental Tests Example. A Normal Varimax transformation was used to generate matrices T_1 . Results may be compaired with the Matrix B of Table 10.10. While the transformed factors are quite recognizable for the Φ Proportional to $B'B$ case, the definitions of the hyperplanes may not be as clean for near zero loadings as was true for the graphical transformations. There are several undesirable negative loadings in the B matrix for the Independent Clusters solution, a result that should not be surprising since the Nine Mental Tests Example did not possess an independent clusters configuration.

11.5 PROMAX Transformations

An intuitively appealing suggestion was made by Hendrickson and White (1964) as a procedure to transform an orthogonal tranformation to an oblique solution. They named this procedure as the PROMAX transformation. Starting from an orthogonal transformation to loadings x_{ij} , n oblique solution was desired which would reduce smaller loadings toward zero while increasing the absolute value of the higher loadings. They proposed to generate target loadings, y_{ij} , dependent on the x_{ij} which increased the relative ratio of the absolute values of the high loadings to the low loadings. Once having such a target, a least squares fitting procedure would be applied to obtain the oblique transformation. To accomplish this, Hendrickson and White suggested the following power type function. Let:

$$\delta_{ij} = +1 \quad \text{if } x_{ij} \geq 0 \quad ; \tag{11.119a}$$

$$\delta_{ij} = -1 \quad \text{if } x_{ij} < 0 \quad . \tag{11.119b}$$

then Hendrickson and white suggest that:

$$y_{ij} = \delta_{ij}|x_{ij}^p| \tag{11.120}$$

where p is a power parameter to be greater than unity. This function raises the x_{ij} to the power p and assigns the algebraic sign of x_{ij} to the y_{ij} . There remains a question as to the value of the power p . A least squares type fit of the oblique loadings to the y_{ij} is the next step. Let $\tilde{F}'$ be a non-normalized transformation to matrix $\tilde{G}$:

$$A\tilde{F}' = \tilde{G} \quad . \tag{11.121}$$

This transformation is taken as non-normalized to allow the derived factor loadings to have a proportional relation to the best fit to the target matrix Y with elements y_{ij} . Consider:

$$E = Y - \tilde{G} = Y - A\tilde{F}' \quad . \tag{11.122}$$

By standard least squares theory:

$$\tilde{F}' = (A'A)^{-1}A'Y \quad . \tag{11.123}$$

Matrix $\tilde{F}'$ is normalized by columns to matrix F' of normals to the hyperplanes To accomplish this:

$$D^{-2} = \mathrm{diag}(\tilde{F}\tilde{F}') \tag{11.124}$$

$$F' = \tilde{F}'D \quad . \tag{11.125}$$

Hendrickson and White provided three examples of the use of Promax; Harman 24 Tests Example (1960), Cattell and Dickman Ball Problem (1962) and Thurstone Box Problem (1947). In the first two of these examples Hendrickson and White found that a power coefficient, p, equal to 4 worked best. For the Thurstone Box Problem their finding was that p equal to 2 worked best. Use of Promax for the Nine Mental Tests Example is given in Table 11.10. The Least Squares Hyperplane Fitting, LSQHYP, from Chapter 10, are given at the left for comparison. Results for two values of p are given: p = 2 and p = 4. For factors 1 and 3 the results for p = 4 appear to be superior, being quite close to the LSQHYP results. For factor 2, the results for p = 2 appear to be better, the quite negative loadings for attributes 2 and 7 on factor 2 for p = 4 are undesirable. Maybe, an analyst should try both p = 2 and p = 4 and select the preferable results. It appears that this selection might be made factor by factor. General experience has indicated that Promax yields good to excellent results.

11.6 Weighted Least Squares Hyperplane Fitting

The general form for least squares hyperplane fitting is similar to Thurstone's (1947) suggested equation for simple structure given in equation (11.1). The equation for this form is:

$$\xi = \sum_{i=1}^{n}\sum_{j=1}^{r} v_{ij}^2 w_{ij} \quad . \tag{11.126}$$

where the v_{ij}'s may be either the projections on normals (structure loadings) in our matrix G or the factor weights (pattern loadings) in our matrix B . The new feature consists of the weights w_{ij} which may be determined by various judgments and/or techniques. Given the weights the

Table 11.10

Promax Transformations for Nine Mental Tests Example
Using Normal Varimax to Develop Target

	LSQHYP Transformation			Promax Transformation p = 2			Promax Transformation p = 4		
	Normal Matrices F'								
	1	2	3	1	2	3	1	2	3
1	.30	.68	.26	.35	.55	.36	.27	.48	.28
2	.50	-.73	.19	.56	-.77	.26	.51	-.79	.21
3	.82	.08	-.95	.75	.32	-.90	.82	.39	-.94
	Projections on Normals, Matrices G								
	1	2	3	1	2	3	1	2	3
1	.53	.05	-.09	.56	.04	-.01	.53	.03	-.07
2	.54	-.06	.07	.59	-.11	.16	.53	-.14	.09
3	.42	.32	.01	.45	.27	.09	.40	.24	.03
4	.01	.71	-.03	.00	.68	-.01	-.01	.66	-.03
5	-.01	.78	.02	-.01	.73	.06	-.03	.70	.03
6	.02	.71	.02	.02	.66	.05	.00	.63	.02
7	.17	.02	.40	.24	-.11	.46	.16	-.17	.42
8	-.06	.39	.43	-.01	.25	.48	-.07	.18	.44
9	.05	.27	.42	.11	.12	.48	.04	.06	.43

analytic problem is to establish the factor transformation to minimize ξ. The analytic procedure for the projections on the normals was described in Chapter 10, section 10.3. A special case for the weights was discussed there for which the weights were restricted to being 0 or 1 . With this restriction a procedure LSQHYP was discussed. A suggestion was made that these weights could be defined by the analyst as a result of judgments from graphical rotatations. Other determinations of weights for the projections on normals case will be described subsequently. Development of an anlytic solution for the case of factor weights is discussed next.

For the case of the factor weights, b_{ij} , equation (11.126) is written as:

$$\xi = \sum_{i=1}^{n}\sum_{j=1}^{r} b_{ij}^2 w_{ij} \quad . \tag{11.127}$$

There appears to be no simple, general solution for the transformation matrix T to minimize function ξ. Instead, a series of transformations of each factor with respect to each other factor is employed for each of a series of cycles. This is similar to the solution for the Direct Oblimin criterion. Consider factor k which, successively, takes on the value of each factor. For each value of k , each of the other factors is taken as factor m and a series of solutions is computed for transforming k in the direction of each m . For each pair of factors, k and m , the criterion may be written as:

$$\tilde{\xi}_{km} = \sum_{i=1}^{n}\tilde{b}_{ik}^2 w_{ik} + \sum_{i=1}^{n}\tilde{b}_{im}^2 w_{im} + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq k,m}}^{r} b_{ij}^2 w_{ij} \quad . \tag{11.128}$$

The transformation is defined in equation (11.59) as was done for the Direct Oblimin solution. The restriction on coefficients s_{kk} and s_{km} is given in equation (11.60). Parameter z_{km} is defined in terms of s_{kk} and s_{km} in equation (11.69). The transformed $\tilde{b}_{ik}^2$ and $\tilde{b}_{im}^2$ are given in equations (11.71) and (11.72). Substitution fromequations (11.71) and (11.72) into equation (11.128) yields:

$$\tilde{\xi}_{km} = c_0 + c_1 z_{km} + c_2 z_{km}^2 \tag{11.129}$$

where:

$$c_0 = \left\{ \sum_{i=1}^{n} b_{ik}^2 w_{ik} + \sum_{i=1}^{n} b_{im}^2 w_{im} + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq k,m}}^{r} b_{ij}^2 w_{ij} \right\} \quad ; \tag{11.130.0}$$

$$c_1 = 2\left\{ \sum_{i=1}^{n} b_{ik}^2 w_{ik}\phi_{km} - \sum_{i=1}^{n} b_{ik} b_{im} w_{im} \right\} \quad ; \tag{11.130.1}$$

$$c_2 = \left\{ \sum_{i=1}^{n} b_{ik}^2 w_{ik} + \sum_{i=1}^{n} b_{ik}^2 w_{im} \right\} \quad . \tag{11.130.2}$$

With the weights w_{ik} and w_{im} being limited to being non-negative, coefficient c_2 must be non-negative. The possibility that c_2 equalling zero is so unlikely that this possibility may be ignored. Since equation (11.129) is a quadratic the solution for an optimum is:

$$z_{km} = -c_1/2c_2 \quad . \tag{11.131}$$

With c_2 taken as positive, equation (11,131) yields the desired minimum value of $\tilde{\xi}_{km}$. Then, from equation (11.70):

$$s_{kk} = 1 \Big/ \sqrt{1 + 2z_{km}\phi_{km} + z_{km}^2} \quad . \tag{11.132}$$

From equation (11.69):

$$s_{km} = s_{kk}z_{km} \quad . \tag{11.133}$$

Rvised values of t_{hk} , h = 1,r, can be obtained from equation (1159) as:

$$\tilde{t}_{hk} = s_{kk}t_{hk} + s_{km}t_{hm} \quad . \tag{11.133}$$

Revised values of ϕ_{hk} are obtained from:

$$\tilde{\phi}_{hk} = \tilde{\phi}_{kh} = s_{kk}\phi_{hk} + s_{km}\phi_{hm} \quad \text{for } h \neq k \tag{11.134}$$

$$\tilde{\phi}_{kk} = 1 \quad . \tag{11.135}$$

Revised values of b_{ik} and b_{im} may be obtained from eauations (11.66) and (11.67). A computer program can start from initial matrices T , Φ , and B ; cycle through the factors as k and, within each k , cycle through all factors except k as m . Such a program would continue until none of the transformations within a cycle are greater than some minimum value. A possible good starting transformation can be obtained from the LSQHYP solution described in Chapter 10. There remains the problem of determining the weights, w_{ij} to be used.

Examples of least squares hyperplane fitting are given in Table 11.11 with the weights being 0 or 1 as selected from graphical rotations as in the LSQHYP procedure discussed in Chapter 10, section 10.3. The least weighted least squares solution for the projections on the normals is given on the left of Table 11.11 with the weights for the projections being given as superscripts. These results are the same as given in Table 10.11. The resulting factor weights are given at the lower right of Table 11.11 for comparison with results for the weighted least squares solutionfor the factor weights which are given at the lower right. Again, the weights used are given as superscripts to the factor weights. These weights are identical to those used on the projections on the normals. In this example there are only small changes in the factor weights from the weighted least squares solution on the projections to the solution on the factor weights. In general, there may be only slight differences between these two solutions so that the weighted least squares solution on the projections on the normals might be recommended when the weights are given before the solution is accomplished.

Weighted least squares hyperplane fitting provides a mechanism to automate in a simulated fashion some of the detailed judgments made by analysts in graphical rotations of

Table 11.11
Weighted Least Squares Hyperplane Fitting
Nine Mental Tests Example

Projections on Normals (Structure Loadings)

	Least Squares Projections			Least Squares Factor Weights		
	1	2	3	1	2	3
1	.53[0]	.05[1]	-.09[1]	.52	.05	-.08
2	.54[0]	-.06[1]	.07[1]	.54	-.06	.08
3	.42[0]	.32[0]	.01[1]	.41	.32	.01
4	.01[1]	.71[0]	-.03[1]	.00	.71	-.04
5	-.01[1]	.78[0]	.02[1]	-.01	.78	.02
6	.02[1]	.71[0]	.02[1]	.02	.71	.02
7	.17[0]	.02[1]	.40[0]	.18	.03	.41
8	-.06[1]	.39[0]	.43[0]	-.04	.40	.44
9	.05[1]	.27[0]	.42[0]	.06	.27	.42

Factor Weights (Pattern Loadings)

	Least Squares Projections			Least Squares Factor Weights		
	1	2	3	1	2	3
1	.68	.05	-.12	.66[0]	.05[1]	-.10[1]
2	.68	-.06	.09	.68[0]	-.06[1]	.10[1]
3	.53	.32	.01	.52[0]	.32[0]	.02[1]
4	.01	.72	-.04	.00[1]	.72[0]	-.04[1]
5	-.01	.79	.03	-.02[1]	.79[0]	.03[1]
6	.03	.71	.02	.02[0]	.72[0]	.02[1]
7	.22	.02	.50	.23[0]	.03[1]	.49[0]
8	-.07	.40	.55	-.05[1]	.40[0]	.53[0]
9	.06	.27	.53	.08[1]	.28[0]	.52[0]

[0] Weights = 0.

[1] Weights = 1.

factors. An experienced analyst develops judgments as to the likelihood that an attribute vector will be in a particular hyperplane. Such a judgment may be termed the analyst's personal probability Tucker and Finkbeiner (1981) proposed a simulated representation of such judgments based on the loadings on a factor at some point in the graphical rotation procedure. They termed this an artificial personal probability and suggested a convenient equation:

$$APP_{ij} = 1 \Big/ \{1 + a|v_{ij}|^{c}\} \qquad (11.136)$$

where APP_{ij} is the value of the artificial personal probability for attribute i on factor j and v_{ij} is the loading of attribute i on factor j. Parameters of this function are a and c to be set by the analyst from properties of the data being analysed. Values of a and c are suggested here for the analysis of correlation matrices. Figure 11.2 illustrates the form of this function in terms of the factor weights b_{ij} as the v_{ij} for what Tucker and Finkbeiner termed the two sided condition, this being the most general condition when meaningful factor loadings may be either positive or negative. There is a short distance about a factor loading of zero when the value of APP is approximately unity, this representing a judgment that attributes having such small loadings would be in the hyperplane. On either side of the central interval the APP function slopes off until becoming approximately zero. This function as illustrated is different from a step function that symbolize judgments that attributes are either in a hyperplane or out of a hyperplane. Tucker and Finkbeiner suggested that there should be a transition between in a hyperplane and being out of a hyperplane.

Tucker and Finkbeiner considered the case of what has been called the positive manifold for which meaningful factor loadings are judged to be zero or positive. Such a case arrises for ability measures where factors should not have negative effects on the performance of any attribute. They termed this the one sided case with the form of the APP function illustrated in Figure 11.3. In order to force loadings away from being negative the valuse of APP are set to unity. A judgment is required from the analyst before using the transformation procedure as to whether a one sided or a two sided function is to be used. This judgment should depend on the nature of the data being analysed. One point is that all measures should be made in a positive direction for which a higher value is better. Attributes, such as reaction time, for which low values are better should be reversed in direction by changing the signs of the loadings for such an attribute in the original factor matrix. On the positive side, the APP function is the same for the one sided case as for the two sided case.

Values of the parameters a and c of the APP function of equation 11.136 may be related to the value of v for $APP = .5$ and $APP = .1$, these values being selected for analyses involving correlation matrises. Let $v_{.5}$ be the positive value of v for $APP = .5$ and $v_{.1}$ be the

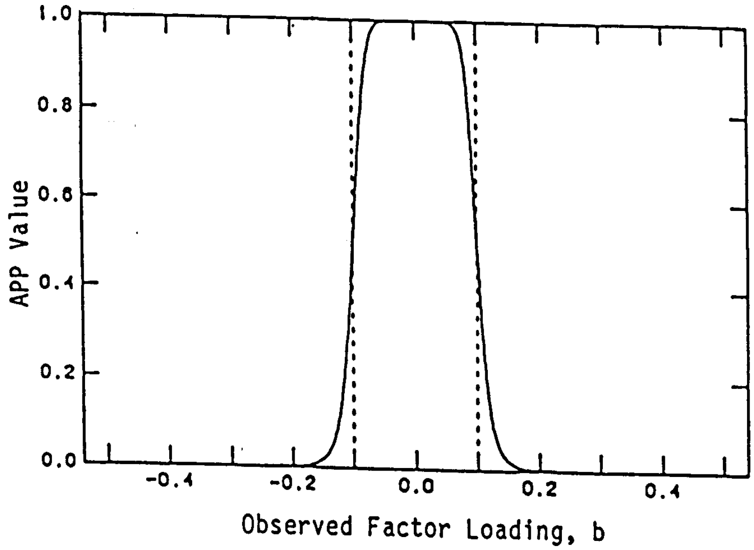

Figure 11.2. Illustration of two sided APP function: $a = 10^9$; $c = 9$.

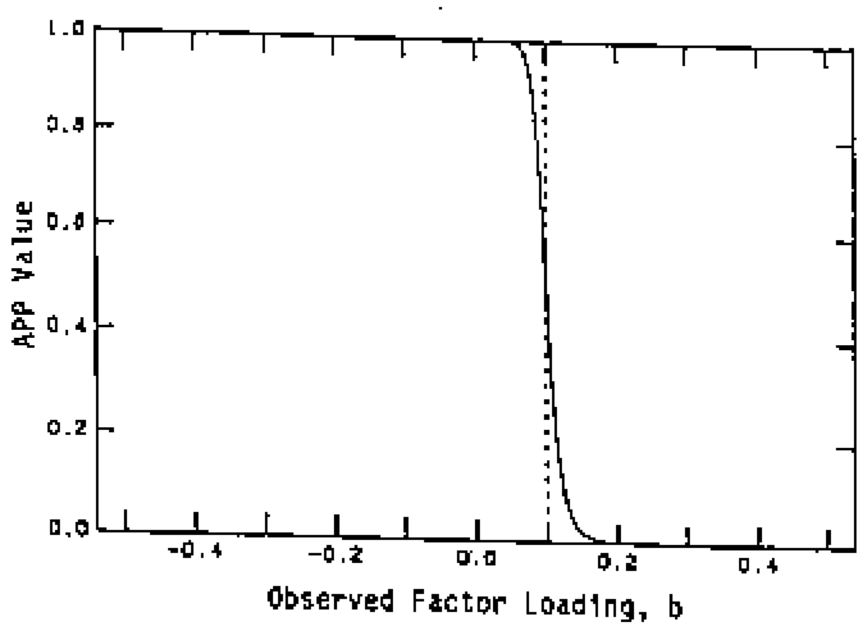

Figure 11.3. Illustration of one sided APP function:

$a = 10^9$; $c = 9$; if $b < 0$, APP = 1 .

positive value of v for $APP = .1$. The ratio of $b_{.1}$ to b_5 is related to the slope of the APP function. At a ratio of unity the function would be a step function; as the ratio increases, the slope of the function flattens out. The following equations give the relations of the parameters to these values of v :

$$c = \ln(9) \Big/ [\ln(v_{.1}) - \ln(v_{.5}] \quad ; \qquad (11.137)$$

$$a = (v_{.5})^{-c} \quad . \qquad (11.138)$$

Tucker and Finkbeiner found with many experimental trials the a value of $c = 9$ gave best results. The value of $v_{.5}$ can be selected by the analyst and the value of c computed by equation 11.138. A computational method suggested by Tucker and Finkbeiner will be discussed subsequently.

A first attempt to use APP in an iterative fashion involved the projections on the normals. In this procedure the v_{ij} were defined as the projections on the normals, g_{ij} . A starting transformation was necessary which might be obtained by a VARIMAX transformation. Each of the factors was transformed independently with trial weights being the APP values computed from the trial projections on the normals. This was a successive trial type procedure with new weights being computed as the APP values after each solution for new normals and projections on these normals by the least squares hyperplane fitting procedure described previously. While there was no explicit criterion to be minimized or maximized, each normal appeared to come to a stable position after a few trials. With some bodies of data the results obtained appeared excellent; there was, however, a problem with some other bodies of data. This procedure suffered the same problem as other single plane methods: two or more factors arrived at the same stable position even though they had different starting positions. Consequently, this procedure was discarded.

Tucker and Finkbeiner reported excellent results with the APP applied to the factor weights. They termed this procedure DAPPFR for direct artificial probability factor rotation. Results for the nine mental tests example are given in Table 11.12. Since these attributes are ability tests, a prior judgment is indicated that a one sided solution would be appropriate. This solution is shown on the left of Table 11.12 and, indeed, these results are very close to the LSQHYP solution given in Table 11.11. For a contrast, a two sided solution was run with results given at the right of Table 11.12. While factors 1 and 3 results are close to the LSQHYP results, the results for factor 2 are quite unsatisfactory. The negative loading for attribute 7 indicates that these results should be rejected for ability measures. Tucker and Finkbeiner (personal correspondence) reported that the results from extensive simulation and Monte Carlo studies that DAPPFR gave, on the average, the best of any automated transformation procedure tried including Direct Oblimin. Second place was PROMAX with a power constant of 2. Such

Table 11.12
Dappfr Transformations Factor Weights
Nine Mental Tests Example

	One Sided Solution			Two Sided Solution		
	1	2	3	1	2	3
1	.66	.05	-.10	.65	.24	-.13
2	.67	-.05	.10	.68	.01	.11
3	.52	.33	.01	.52	.42	.00
4	.00	.72	-.04	.00	.74	-.05
5	-.02	.80	.03	-.02	.78	.03
6	.02	.71	.02	.02	.70	.02
7	.24	.02	.49	.29	-.25	.59
8	-.06	.40	.54	.00	.05	.64
9	.08	.28	.52	.14	-.04	.62

judgments are possible with simulation and Monte Carlo studies since a criterion is available from the generation of the data. Trials with real data can only lead to a judgment that the obtained solutions are liked or are not liked. But, there is no concrete criterion for a judgment of superiority. The DAPPFR procedure appears to work very well both for real data and simulated data.

Computational procedures start from some initial location of the trait vectors in matrix T and factor weights in matrix B . Then, a series of cycles of weighted least squares hyperplanes fitting are performed with the APP values used as weights. These APP values are computed by equation 11.136 using the factor weights b_{ij} as the v_{ij} . Initial values of APP are computed from the initial matrix B . These weights are revised after each cycle. While there is no explicit algebraic function to be optimized the procedure does arrive at a stable position, usually in 20 cycles or fewer. Tucker and Finkbeiner commented that a value of $b_{.5}$ equal to .15 gave good results when data being analysed were correlation matrices and attributes have high communalities. They suggested an alternative computational procedure dependent on the mean absolute value of the entreis in the B matrix. Let $\bar{b}$ be the mean absolute value of the entries in the trial matrix B . For the first cycle use:

$$a = (\bar{b})^{-9} \quad . \qquad (11.139)$$

For all subsequent cycles use:

$$a = \tfrac{r}{6}(\bar{b})^{-9} \qquad (11.140)$$

where r is the number of factors. A good starting transformation involves VARIMAX followed by PROMAX with a power constant of 2 .