OBLIQUE PROCRUSTES ROTATIONS

IN FACTOR ANALYSIS

Franklin T. Luk*

TR 82-481
January 1982

Department of Computer Science
Cornell University
Ithaca, New York 14853

*This research was supported in part by the U.S. Army Research Office under grant DAAG 29-79-C0124.

Oblique Procrustes Rotations in Factor Analysis

Franklin T. Luk*

Department of Computer Science

Cornell University

Ithaca, New York 14853

Abstract

This paper addresses the problem of rotating a factor matrix obliquely to a least squares fit to a target matrix. The target may be fully or partially specified. An iterative computing procedure is presented.

Keywords: oblique rotations, procrustes problem, factor analysis, least squares.

*This research was supported in part by the U.S. Army Research Office under grant DAAG 29-79-C0124

1. Introduction

We consider the problem of rotating a given factor matrix obliquely to a least squares fit to a prescribed pattern matrix. The target matrix may be partially or fully specified. Gruvaeus [5] solves this problem using a penalty function approach, involving a series of Fletcher and Powell [3] minimizations. Browne [2] suggests a procedure based on the application of elementary oblique rotations (see Jennrich and Sampson [6]) and reports numerical results that are at least as good as those given by Gruvaeus' method. Due to its simplicity and geometrical elegance, Browne's procedure is usually preferred over Gruvaeus' approach.

Browne [2] proposes that one applies Newton's method to compute the elementary oblique rotations. In this paper we show how these rotations can be determined fairly easily using Lagrange multipliers.

2. The Algorithm

Mathematically, our problem is to determine a nonsingular $m \times m$ matrix X, satisfying the condition:

$$(2.1) \quad \mathrm{diag}\ \{(X^T X)^{-1}\} = I,$$

that will transform a given $n \times m$ matrix A to a least squares fit to a prescribed $n \times m$ target factor pattern B. Let $I(j)$ represent the set of row indices corresponding to specified elements of column j of B, for $j=1,2,\ldots,m$. If $F = AX$, the criterion to be minimized is

(2.2) $$\sum_{j=1}^{m} \{ \sum_{i \in I(j)} (f_{ij}-b_{ij})^2 \},$$

where $F = (f_{ij})$ and $B = (b_{ij})$.

We may state here that it is still not known how to compute the global minimum of (2.2) subject to the constraint (2.1). Following Browne [2] we shall approximate X by a product of elementary oblique rotations (see Jennrich and Sampson [6]). One such rotation of the i-th primary axis over the plane defined by the i-th and j-th (say i < j) primary axes is of the form:

(2.3) $$\begin{array}{c c} \begin{array}{ccccccccc} & & & i & & j & & & \end{array} & \\ \begin{pmatrix} 1 & & & & & & & & \\ & \cdot & & & & & & 0 & \\ & & \cdot & & & & & & \\ & & & 1 & & & & & \\ & & & & \gamma & & -\delta & & \\ & & & & & 1 & & & \\ & & & & & & \cdot & & \\ & & & & & & & \cdot & \\ & & & & & & & & 1 \\ & 0 & & & & & & \cdot & \\ & & & & & & & & \cdot \\ & & & & & & & & 1 \end{pmatrix} & \begin{array}{c} \\ \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} \end{array},$$

i.e. the identity matrix with the (i,i)-th and (i,j)-th elements altered. The parameters satisfy the relation:

(2.4) $$\gamma^2 = \delta^2 + 2c_{ij}\,\delta + 1,$$

where c_{ij} $(-1 < c_{ij} < 1)$ represents the cosine of the angle between the i-th and j-th primary axes before rotation.

Designate by A^c the current rotated factor matrix at an intermediate stage, and let $A^c = (a^c_{ij}) = (\mathbf{a}^c_1, \mathbf{a}^c_2, \ldots, \mathbf{a}^c_m)$. Let us consider a rotation of the i-th primary axis over the plane defined by the i-th and j-th primary axes. Such a rotation will affect only the i-th and j-th columns of A^c and may be represented as

$$(2.5)\qquad (a_i^c, a_j^c)\begin{pmatrix} \gamma & -\delta \\ 0 & 1 \end{pmatrix}.$$

Suppose that $\alpha = |I(i)|$ and $\beta = |I(j)|$. Let $i(1), i(2), \ldots, i(\alpha)$ and $j(1), j(2), \ldots, j(\beta)$ denote the row indices of the specified elements of columns i and j, respectively, of the target B. Our goal is to determine the parameters γ and δ to minimize the euclidean length of the vector

$$(2.6)\qquad \begin{pmatrix} a^c_{i(1),i} & 0 \\ a^c_{i(2),i} & 0 \\ \vdots & \vdots \\ a^c_{i(\alpha),i} & 0 \\ 0 & -a^c_{j(1),i} \\ 0 & -a^c_{j(2),i} \\ \vdots & \vdots \\ 0 & -a^c_{j(\beta),i} \end{pmatrix} \begin{pmatrix} \gamma \\ \delta \end{pmatrix} - \begin{pmatrix} b_{i(1),i} \\ b_{i(2),i} \\ \vdots \\ b_{i(\alpha),i} \\ b_{j(1),j} - a^c_{j(1),j} \\ b_{j(2),j} - a^c_{j(2),j} \\ \vdots \\ b_{j(\beta),j} - a^c_{j(\beta),j} \end{pmatrix}.$$

In other words, we want to solve the constrained least squares problem:

(2.7) $\|F\mathbf{x} - \mathbf{g}\|$ = minimum,

where $\|\cdot\|$ denotes the euclidean vector norm, subject to the constraint:

(2.8) $x_1^2 = x_2^2 + 2cx_2 + 1$,

with $-1<c<1$. Since

(2.9) $x_2^2 + 2cx_2 + 1 = (x_2+c)^2 + (1-c^2)$,

the equation (2.8) says that

(2.10) $x_1 \neq 0$.

Using Lagrange multipliers, we obtain the matrix equation:

(2.11) $(F^TF + \lambda E)\mathbf{x} = F^T\mathbf{g} + c\lambda\mathbf{e}_2$,

where

(2.12) $E = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

and λ is a Lagrange parameter. But

(2.13) $F^TF = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$,

with $d_1 = \sum_{k=1}^{\alpha} a_{i(k),i}^2$ and $d_2 = \sum_{k=1}^{\beta} a_{j(k),i}^2$. We shall assume that $d_1 d_2 > 0$. Let

(2.14) $\mathbf{h} \equiv \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \equiv F^T\mathbf{g}$.

Equation (2.11) can therefore be written as

(2.15) $\begin{pmatrix} d_1+\lambda & 0 \\ 0 & d_2-\lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} h_1 \\ h_2+c\lambda \end{pmatrix}$.

Let us first consider the special case of a null h_1. As $x_1 \neq 0$ we get the solution:

$$
\begin{aligned}
&\lambda = d_1, \\
(2.16)\quad & x_2 = \frac{h_2+c\lambda}{d_2-\lambda} \quad \text{or} \quad \frac{h_2+cd_2}{d_2-\lambda} - c, \quad \text{and} \\
& x_1 = [x_2^2 + 2cx_2+1]^{1/2}.
\end{aligned}
$$

Note that we may also choose the negative square root for x_1. From here on we shall assume that $h_1 \neq 0$. The following lemma will be useful.

<u>Lemma 1</u> (cf. Gander [4, Theorem 1]). If $(\mathbf{x},\lambda_x)$ and $(\mathbf{y},\lambda_y)$ are solutions to the equations (2.8) and (2.11) then

$$(2.17)\quad ||F\mathbf{y} - \mathbf{g}||^2-||F\mathbf{x} -\mathbf{g}||^2 = \tfrac{1}{2}(\lambda_x-\lambda_y)[(x_1-y_1)^2-(x_2-y_2)^2].$$

<u>Proof</u>. Since $(\mathbf{x},\lambda_x)$ and $(\mathbf{y},\lambda_y)$ are solutions of (2.11) we have

$$(2.18)\quad F^TF\mathbf{x} - F^T\mathbf{g} = -\lambda_x E\mathbf{x} + c\lambda_x \mathbf{e}_2$$

and

$$(2.19)\quad F^TF\mathbf{y} - F^T\mathbf{g} = -\lambda_y E\mathbf{y} + c\lambda_y \mathbf{e}_2.$$

Premultiplying equation (2.19) by $\mathbf{y}^T$ and from it subtracting equation (2.18) premultiplied by $\mathbf{x}^T$, we obtain

$$
\begin{aligned}
(2.20)\quad & ||F\mathbf{y}||^2-||F\mathbf{x}||^2-\mathbf{g}^TF(\mathbf{y}-\mathbf{x}) \\
& = \lambda_x(\mathbf{x}^TE\mathbf{x}-cx_2)-\lambda_y(\mathbf{y}^TE\mathbf{y}-cy_2).
\end{aligned}
$$

Similarly, if we premultiply (2.18) by $\mathbf{y}^T$, premultiply (2.19) by $\mathbf{x}^T$ and from the resulting first equation subtract the second one, we get

$$(2.21) \quad -\mathbf{g}^T F(\mathbf{y}-\mathbf{x}) = \lambda_x(-\mathbf{y}^T E\mathbf{x}+cy_2) - \lambda_y(-\mathbf{y}^T E\mathbf{x}+cx_2).$$

Since

$$\|F\mathbf{y}-\mathbf{g}\|^2-\|F\mathbf{x}-\mathbf{g}\|^2 = \|F\mathbf{y}\|^2-\|F\mathbf{x}\|^2 - 2\mathbf{g}^T F(\mathbf{y}-\mathbf{x}),$$

we can add equations (2.20) and (2.21) to get

$$(2.22) \quad \|F\mathbf{y}-\mathbf{g}\|^2 - \|F\mathbf{x}-\mathbf{g}\|^2$$

$$= \lambda_x(\mathbf{x}^T E\mathbf{x}-cx_2-\mathbf{y}^T E\mathbf{x}+cy_2)$$

$$- \lambda_y(\mathbf{y}^T E\mathbf{y}-cy_2-\mathbf{y}^T E\mathbf{x}+cx_2).$$

But from (2.8) we have the relation:

$$\mathbf{x}^T E\mathbf{x}-2cx_2-1 = \mathbf{y}^T E\mathbf{y}-2cy_2-1,$$

so that

$$(2.23) \quad \mathbf{x}^T E\mathbf{x}-cx_2+cy_2 = \mathbf{y}^T E\mathbf{y}-cy_2+cx_2.$$

From (2.23) we conclude that the factors of λ_x and λ_y in (2.22) are the same. They therefore equal their arithmetic mean, which is

$$\frac{1}{2}[\mathbf{x}^T E\mathbf{x}+\mathbf{y}^T E\mathbf{y}-2\mathbf{y}^T E\mathbf{x}] = \frac{1}{2}[(x_1-y_1)^2-(x_2-y_2)^2]. \qquad \square$$

We next consider the other special case when $h_2+cd_2=0$. With the same notations as in Lemma 1 we get (cf. (2.16)) that

$$x_2 = y_2 = -c$$

and so equation (2.17) reduces to

$$(2.24)\qquad ||F\mathbf{y}-\mathbf{g}||^2-||F\mathbf{x}-\mathbf{g}||^2 = \frac{1}{2}(\lambda_x-\lambda_y)(x_1-y_1)^2.$$

From (2.8) we obtain

$$x_1^2 = y_1^2 = (-c)^2+2c(-c)+1 = 1-c^2.$$

Using (2.15) we get

$$1-c^2 = \frac{h_1^2}{(d_2+\lambda)^2} ,$$

which gives us the two possible values of λ:

$$(2.25)\qquad \lambda = -d_1 \pm (\frac{h_1^2}{1-c^2})^{1/2}.$$

From (2.24) we see that we should choose the greater value of λ. The proper solution $(\mathbf{x},\lambda)$ for this special case is therefore

$$\lambda = -d_1 + (\frac{h_1^2}{1-c^2})^{1/2} ,$$

$$(2.26)\qquad x_1 = \frac{h_1}{d_1+\lambda} , \text{ and}$$

$$x_2 = -c.$$

Henceforth we shall also assume that $h_2+cd_2\neq 0$. Let us rewrite equation (2.8) as

$$x_1^2 = (x_2+c)^2 + (1-c^2).$$

Using (2.15) to substitute for x_1 and (x_2+c), we obtain

$$(2.27) \quad \left(\frac{h_1}{d_1+\lambda}\right)^2 = \left(\frac{h_2+cd_2}{d_2-\lambda}\right)^2 + (1-c^2),$$

and we see that neither denominator can vanish. Equation (2.27) can thus be transformed into a quartic equation in λ. From Figure 1 we see that (2.27) has one real root in the open interval $(-\infty,-d_1)$, one real root in $(-d_1,d_2)$, and two distinct real or complex conjugate roots in (d_2,∞). We are only interested in the real roots and shall establish that the root in the middle interval $(-d_1,d_2)$ will minimize the criterion of (2.2).

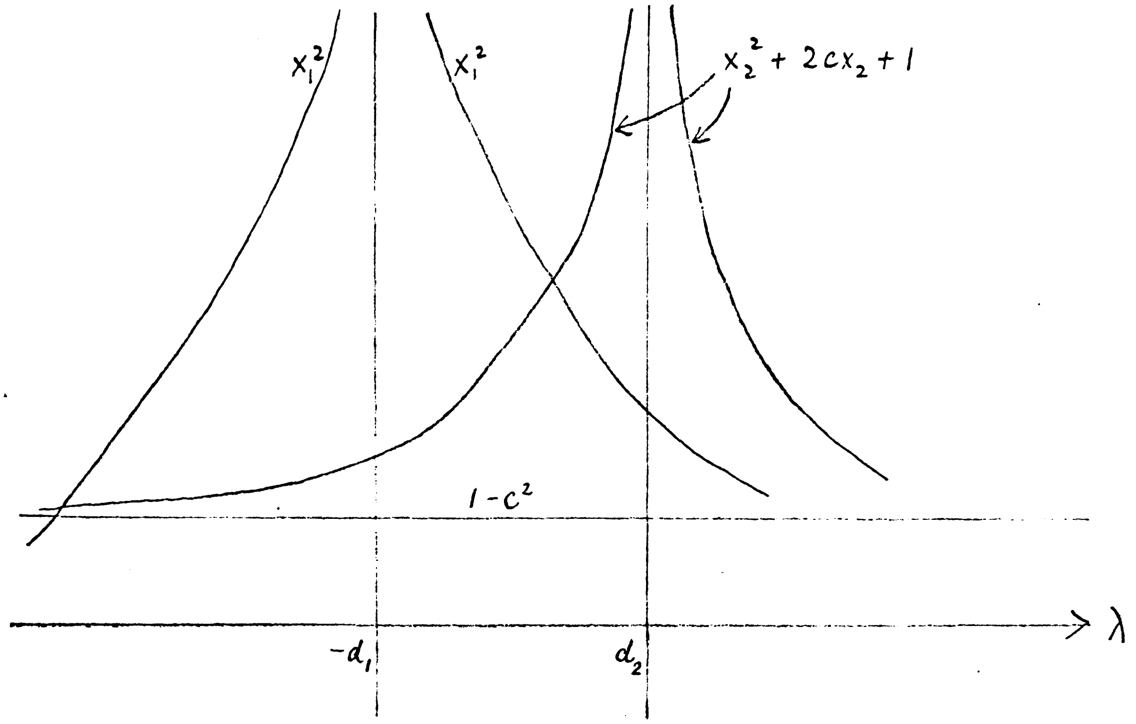

Figure 1. Graphs of x_1^2 and $x_2^2 + 2cx_2 + 1$ as functions of λ.

[illegible] 2. Let both vectors $\mathbf{x}$ and $\mathbf{y}$ satisfy equation (2.8). Then

$$(x_2+c)(y_2+c) + (1-c^2) \le |x_1y_1|, \tag{2.28}$$

where the inequality is strict if $x_2 \ne y_2$, and

$$\frac{1}{2}[(x_1-y_1)^2 - (x_2-y_2)^2] = (x_2+c)(y_2+c)+(1-c^2)-x_1y_1. \tag{2.29}$$

Proof. First, we have

$$x_1^2 = (x_2+c)^2 + (1-c^2)$$

and

$$y_1^2 = (y_2+c)^2 + (1-c^2).$$

Multiplying these two equations together, we get

$$x_1^2y_1^2 = (x_2+c)^2(y_2+c)^2 + (1-c^2)[(x_2+c)^2+(y_2+c)^2] + (1-c^2)^2$$

$$= [(x_2+c)(y_2+c)+(1-c^2)]^2$$

$$+ (1-c^2)[(x_2+c)^2+(y_2+c)^2-2(x_2+c)(y_2+c)]$$

$$= [(x_2+c)(y_2+c)+(1-c^2)]^2 + (1-c^2)(x_2-y_2)^2$$

$$\geq [(x_2+c)(y_2+c)+(1-c^2)]^2.$$

Hence

$$|x_1y_1| \geq (x_2+c)(y_2+c)+(1-c^2).$$

Note that the above inequality is strict if $x_2 \neq y_2$.

Second, we see that

$$\frac{1}{2}[(x_1-y_1)^2 - (x_2-y_2)^2]$$

$$= \frac{1}{2}[(x_1^2-x_2^2)+(y_1^2-y_2^2)-2x_1y_1+2x_2y_2]$$

$$= \frac{1}{2}[2cx_2+1+2cy_2+1-2x_1y_1+2x_2y_2]$$

$$= x_2y_2+cx_2+cy_2+c^2+(1-c^2)-x_1y_1$$

$$= (x_2+c)(y_2+c)+(1-c^2)-x_1y_1. \quad \square$$

Suppose that $(\mathbf{x},\lambda_x)$ and $(\mathbf{y},\lambda_y)$ are solutions to the equations (2.8) and (2.11) and that $\lambda_x \neq \lambda_y$. It follows that $x_2 \neq y_2$ and so the inequality (2.28) is strict. First, we consider the case when $\lambda_x > -d_1$ and $\lambda_y < -d_1$. We observe from equation (2.15) that

$$x_1 y_1 < 0.$$

Hence

$$(x_2+c)(y_2+c) + (1-c^2) - x_1 y_1 > 0$$

from (2.28), and so

$$(\lambda_x-\lambda_y)[(x_2+c)(y_2+c)+(1-c^2)-x_1 y_1] > 0.$$

Second, we let both λ_x and λ_y be greater than $-d_1$. Then

$$x_1 y_1 > 0$$

and thus

$$(x_2+c)(y_2+c) + (1-c^2) - x_1 y_1 < 0.$$

If $\lambda_x < \lambda_y$, we get

$$(\lambda_x-\lambda_y)[(x_2+c)(y_2+c)+(1-c^2)-x_1 y_1] > 0.$$

We have therefore proved the following important result.

<u>Theorem</u>. Let $(\mathbf{x},\lambda_x)$ and $(\mathbf{y},\lambda_y)$ be solutions to the equations (2.8) and (2.11). We have

(i) If $\lambda_x > -d_1$ and $\lambda_y < -d_1$, then

$$||F\mathbf{y}-\mathbf{g}||^2 - ||F\mathbf{x}-\mathbf{g}||^2 > 0.$$

(ii) If $\lambda_y > \lambda_x > -d_1$, then

$$||F\mathbf{y}-\mathbf{g}||^2 - ||F\mathbf{x}-\mathbf{g}||^2 > 0. \quad \square$$

We conclude that we need the smallest root of λ that is greater than $-d_1$, i.e. that unique root in the open interval $(-d_1, d_2)$. We can therefore apply any zero-finding procedure with assured convergence, e.g. Brent [1, Chapter 4], to the function

$$(2.30) \qquad f(\lambda) \equiv \left(\frac{h_1}{d_1+\lambda}\right)^2 - \left(\frac{h_2+cd_2}{d_2-\lambda}\right)^2 - (1-c^2)$$

to find the root in $(-d_1, d_2)$.

After the parameters δ and γ for the elementary oblique rotation have been found, the factor matrix is updated in accordance to (2.5). We modify the $(m-1)\mathrm{x}(m-1)$ matrix $C \equiv (c_{pk})$, consisting of cosines between the primary axes, as follows (see Browne [2]):

$$(2.31) \qquad c_{ki} := c_{ik} := (c_{ik} + \delta c_{jk})/\gamma$$

for $k = 1,2,\ldots,i-1,i+1,\ldots,m-1$. The other elements of C stay unchanged. The angle θ of rotation defined by δ and γ is given by

$$(2.32) \qquad \theta = \cos^{-1}[(1+\delta c_{ij})/\gamma].$$

We carry out the elementary oblique rotations in a systematic order. Transformations are made on factors 1 with 2, 1 with 3,...,1 with m, 2 with 1, 2 with 3,...,2 with m ,..., m with 1, m with 2 ,..., m with m-1 to constitute a cycle. The cycle is repeated until there is one consisting of all null rotations or the criterion of (2.2) fails to decrease. As the criterion should decrease after each nontrivial rotation and is bounded below by zero, the termination of our procedure is guaranteed.

3. Numerical Example

We were able to find only one example (Gruvaeus [5, p. 500]) for which the target matrix is partially prescribed and some specified elements are nonzero. The given factor matrix is

$$A = \begin{pmatrix} .90 & -.09 & -.03 \\ .83 & .09 & -.04 \\ .87 & -.01 & .07 \\ .55 & .79 & -.07 \\ .56 & .65 & .04 \\ .63 & .60 & .03 \\ .28 & .27 & .45 \\ .38 & .20 & .63 \\ .38 & .19 & .77 \end{pmatrix}$$

and the partially prescribed matrix is

$$B = \begin{pmatrix} .98 & 0 & 0 \\ .76 & 0 & 0 \\ x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \\ 0 & 0 & .76 \\ 0 & 0 & .93 \end{pmatrix} .$$

We regard a rotation as null if $\cos \theta \geq .9999995$. The fourth cycle consists wholly of null rotations and we obtain the final rotated matrix:

$$AX = \begin{pmatrix} .94 & -.09 & -.02 \\ .79 & .10 & .02 \\ .83 & -.04 & .11 \\ .17 & .84 & .15 \\ .20 & .66 & .24 \\ .30 & .61 & .22 \\ -.04 & .15 & .57 \\ .02 & .02 & .75 \\ -.03 & -.03 & .90 \end{pmatrix}.$$

The criterion of (2.2) decreases as follows:

Cycle	Criterion
0	.437400
1	.055413
2	.024703
3	.024691
4	.024691

References

[1] R.P. Brent, Algorithms for minimization without derivatives, Prentice-Hall, Englewood-Cliffs (1973).

[2] M.W. Browne, "Oblique rotation to a partially specified target," Br. J. math. statist. Psychol. 25 (1972), 207-212.

[3] R. Fletcher and M.J.D. Powell, "A rapidly convergent descent method for minimization," Comput. J. 2 (1963), 163-168.

[4] W. Gander, "Least squares with a quadratic constraint," Numer. Math. 36 (1981), 291-307.

[5] G.T. Gruvaeus, "A general approach to procrustes pattern rotation," Psychometrika 35 (1970), 493-505.

[6] R.I. Jennrich and P.F. Sampson, "Rotation for simple loadings," Psychometrika 31 (1966), 313-323.