The Procrustes criterion is a common measure for the distance between two matrices \( X \) and \( Y \), and can be interpreted as the sum of squares of the Euclidean distances between their respective column vectors. Often a weighted Procrustes criterion, using, for example, a weighted sum of the squared distances between the column vectors, is called for. This paper describes and analyzes the performance of an algorithm for rotating a matrix \( X \) such that the column-weighted Procrustes distance to \( Y \) is minimized. The problem of rotating \( X \) into \( Y \) such that an aggregate measure of Tucker's coefficient of congruence is maximized is also discussed.

Key words: rotation, matching, Procrustes analysis, Tucker's coefficient of congruence.

**Introduction**

"Procrustes analysis" is a popular term for the problem, often encountered in factor analysis and multidimensional scaling, of rotating a matrix \( X \) so that it matches a matrix \( Y \) as closely as possible. Let \( Y \) and \( X \) be \( n \times p \) matrices and \( H \) be a \( p \times p \) orthogonal matrix. A measure of closeness frequently employed is the so-called Procrustes or least-squares criterion. Denoting by \( \text{tr}(\cdot) \) the trace of a square matrix, this criterion can be written as

\[
\text{tr} (Y-XH)'(Y-XH). \tag{1}
\]

Various writers have discussed the problem of finding, for a given \( X \) and \( Y \), an orthogonal \( H \) that minimizes \( (1) \) (see, e.g., Gower, 1984, pp. 761-771). The unweighted least-squares criterion \( (1) \) is appropriate when the residuals have equal variance and hence should be weighted equally. If, on the other hand, one wishes to weight the residuals differently, a weighted least-squares criterion is more appropriate. Often it is reasonable to use equal weights across columns but not across rows, and vice versa. The corresponding least-squares criteria are then

\[
\text{tr} (Y-XH)'D_D(Y-XH), \tag{2}
\]

and

\[
I(H) = \text{tr} (Y-XH)'D_D(Y-XH)', \tag{3}
\]

We wish to thank Richard A. Harshman and C. F. Jeff Wu for valuable discussions in the early stages of this work. We would also like to thank Jos ten Berge, John Gower, and the Editor, Associate Editor, and referees whose comments and suggestions greatly improved this paper.

Requests for reprints and code should be sent to Deborah F. Swayne, Bellcore, 445 South Street, Room 2L-331, Box 1910, Morristown, NJ 07960-1910, or by electronic mail to dfs@bellcore.com (Internet) or ...dfs!bellcore (uucp).
where $D_n$ and $D_p$ are diagonal matrices that reflect information about the relative magnitude of the variances. The problem of finding an orthogonal $H$ that minimizes (2) or (3) has been discussed by Lissitz, Schönemann, and Lingoes (1976). The first problem is easy, since it corresponds to minimizing (1) with $Y$ and $X$ replaced by $D_n Y$ and $D_n X$. The second problem is considerably more difficult. Lissitz et al. suggest the replacement of the condition that $H$ be orthogonal by the condition that $HD_p$ be orthogonal, which is a problem that is easier to solve mathematically but does not help solve their original problem. This article introduces an algorithm for finding an orthogonal $H$ that minimizes (3). This algorithm is easy to implement and works well in practice. (It should be pointed out that our algorithm can also be used if $H$ is a $p \times q$ matrix of orthonormal columns; that is, $H'H = I_q$. In the interest of notational and semantic simplicity, we retain the assumption that $p = q$, but the results reported in this paper easily extend to the general case of $q < p$.)

The algorithm is described in the next section, and its convergence properties thereafter. In the last section, we describe several related issues. These include the problem of finding a matrix $H$ that is the product of an orthogonal and a diagonal matrix and that minimizes (1), as well as the problem of finding an orthogonal $H$ that maximizes an aggregate measure of Tucker’s (1951) coefficient of congruence. An algorithm for solving the first problem using planar rotations was recently described by Mooijaart and Commandeur (1990), while a problem similar to the second one was discussed by Brokken (1983), who presented an algorithm derived from the Lagrangian equation. The implementation of our proposed algorithm in any statistical computing environment that allows for the singular value decomposition of matrices should not pose any difficulties; tested implementations of the algorithm in both the S and the New S statistical programming languages (Becker & Chambers, 1984; Becker, Chambers, & Wilks, 1988) can be obtained from the authors.

The Algorithm

The proposed algorithm is based on the idea of embedding the general problem into a specific one for which a solution is easy to find. Whenever $X$ is a matrix whose column vectors are pairwise orthogonal and of equal Euclidean length $p$, it is not hard to find an orthogonal $H$ that minimizes (3). In this case,

$$I(H) = \text{tr} (Y - XH)D_p^2(Y - XH)' = \text{tr} (YD_p^2Y') - 2\text{tr} (XHD_p^2Y') + \text{tr} (D_p^2H'X'XH)$$

$$= \text{tr} (YD_p^2Y') - 2\text{tr} (D_p^2Y'XH) + \rho^2 \text{tr} (D_p^2).$$

Hence, minimizing (3) is equivalent to maximizing

$$\text{tr} (D_p^2Y'XH),$$

which is very similar to the unweighted least-squares problem presented in Gower (1984, pp. 761–771). Writing the singular value decomposition of $D_p^2Y'X$ as $U\Gamma V'$, its solution may be found as

$$H = VU'.$$

Whenever the column vectors of $X$ are no longer orthogonal or of equal Euclidean length, the above argument obviously does not work, but here the link to the special case already solved can be made in the following manner. Given the $n \times p$ matrices $X$ and $Y$, consider the $(n + p) \times p$ matrices