A GENERALIZATION OF TAKANE'S ALGORITHM FOR DEDICOM

HENK A. L. KIERS AND JOS M. F. TEN BERGE

UNIVERSITY OF GRONINGEN

YOSHIO TAKANE

MCGILL UNIVERSITY

JAN DE LEEUW

UNIVERSITY OF CALIFORNIA, LOS ANGELES

An algorithm is described for fitting the DEDICOM model for the analysis of asymmetric data matrices. This algorithm generalizes an algorithm suggested by Takane in that it uses a damping parameter in the iterative process. Takane's algorithm does not always converge monotonically. Based on the generalized algorithm, a modification of Takane's algorithm is suggested such that this modified algorithm converges monotonically. It is suggested to choose as starting configurations for the algorithm those configurations that yield closed-form solutions in some special cases. Finally, a sufficient condition is described for monotonic convergence of Takane's original algorithm.

Key words: DEDICOM, least squares fitting, majorization.

DEDICOM is a model proposed by Harshman (1978) for the analysis of asymmetric data. For an extensive description of this model we refer to Harshman, Green, Wind, and Lundy (1982). A brief description of the model will be given here. According to the DEDICOM model a square data matrix X, containing entries x_{ij} representing the (asymmetric) relation of object i to object j is decomposed as

$$X = ARA' + N, \tag{1}$$

where A is an n by p ($p < n$) matrix of weights for the n objects on p dimensions or aspects, R is a square matrix of order p, representing (asymmetric) relations among the p dimensions, and N is an error matrix with entries n_{ij} representing the part of the relation of object i to object j that is not explained by the model. The objective of fitting this model to the data is to explain the data by means of relations among as small a number of dimensions as possible. These dimensions can be considered as "aspects" of the objects. The "loadings" of the objects on these aspects are given by matrix A. The entries in matrix A indicate the importance of the aspects for the objects. The dimensionality of R and A, and hence the number of aspects to be determined, is to be based on some external criterion, defined by the user.

Several algorithms have been developed for fitting the DEDICOM model. A comparison of most of these has been given by Harshman and Kiers (1987). Kiers (1989) has discussed properties of a number of these algorithms and concludes that his columnwise alternating least squares algorithm is preferable from various points of view.

One of the methods discussed by Harshman and Kiers (1987) is a method proposed by Takane (1985). His method appeared to be very efficient in most practical cases, but turned out to be inadequate in some cases. Moreover, no general convergence properties are known for this method. Therefore, this method has not been recommended for general use. In the present paper, it will be shown that a slight modification of Takane's method is sufficient to overcome these problems.

Before describing the resulting modified algorithm, a brief description of the DEDICOM model will be given. Next, a new type of algorithm will be discussed from which Takane's method can be derived as a special case. Finally, it will be described how Takane's method is to be modified in order to obtain an efficient algorithm that does converge monotonically.

A Monotonically Converging Algorithm for DEDICOM

The DEDICOM model has to be fit in the least squares sense over matrices A and R of order n by p, and p by p, respectively. Without loss of generality matrix A is constrained to be column-wise orthonormal. The loss function that is to be minimized can be written as

$$\sigma(A, R) = \|X - ARA'\|^2. \tag{2}$$

Because $A'A = I_p$, the minimum of σ over R for fixed A is given by $R = A'XA$. Minimizing (2) over A, for fixed R, is equivalent to maximizing

$$f(A) = \operatorname{tr} A'XAA'X'A, \tag{3}$$

over matrix A, subject to the constraint $A'A = I_p$. The algorithm to be presented here is based on the following results.

Result 1. Let X and A be fixed matrices of appropriate orders, and let E be any matrix of the same order as A, then we have

$$\operatorname{tr} E'XEA'X'A \geq -\alpha \operatorname{tr} E'E, \tag{4}$$

if α is some scalar not smaller than the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$, where $\otimes$ refers to the Kronecker product of matrices.

Proof. In order to prove (4) we rewrite the left hand side of (4) as follows. Let e denote Vec (E), the vector with the elements of E strung out row-wise into a column-vector, then

$$\operatorname{tr} E'XEA'X'A = -\operatorname{tr} E'(-X)EA'X'A = -e'(-X \otimes A'XA)e. \tag{5}$$

It should be noted that $(-X \otimes A'XA)$ is not generally symmetric.

As is readily verified, for any square matrix C of appropriate order, we have

$$e'Ce = e'[\tfrac{1}{2} (C + C')]e = e'C_s e, \tag{6}$$

where C_s denotes the symmetric part of C. Let α be the largest eigenvalue of C_s, then it is well-known that

$$e'C_s e \leq \alpha e'e. \tag{7}$$

Combining (6) and (7) we have

$$e'Ce \leq \alpha e'e. \tag{8}$$

Obviously, this property is equally valid for any larger α. On the other hand, since the upper bound in (7) can be attained for certain vectors $\mathbf{e}$, there is no lower α for which (8) holds for every $\mathbf{e}$. Finally, substituting $(-X \otimes A'XA)$ in (8) for C yields the inequality

$$\operatorname{tr} E'(-X)EA'X'A = \mathbf{e}'(-X \underline{\otimes} A'XA)\mathbf{e} \le \alpha \mathbf{e}'\mathbf{e} = \alpha \operatorname{tr} E'E, \qquad (9)$$

from which (4) follows immediately. □

Result 2. If B contains p columns that form an orthonormal basis for the column space of $(XAA'X'A + X'AA'XA + 2\alpha A)$, where α is larger than or equal to the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$, then $f(B) \ge f(A)$.

Proof. Let $g(A, B)$ be defined as $g(A, B) \equiv \operatorname{tr} B'XBA'X'A$. It will first be proven that for B_0, a certain choice of B to be defined later, we have $g(A, B_0) \ge f(A)$. Next, we will prove that $g(A, B_0) \ge f(A)$ implies $f(B_0) \ge f(A)$. Finally, it will be shown that $f(B_0) = f(B)$ for the choice of B as stated in Result 2.

We write B as $B = (A + E)$, where $E = (B - A)$. Then $g(A, B)$ can be rewritten as

$$g(A, B) = \operatorname{tr} (A + E)'X(A + E)A'X'A$$

$$= \operatorname{tr} A'XAA'X'A + \operatorname{tr} E'XEA'X'A + \operatorname{tr} E'(XAA'X'A + X'AA'XA). \qquad (10)$$

Using Result 1 for the second term in the right hand side of (10) yields

$$g(A, B) \ge \operatorname{tr} A'XAA'X'A - \alpha \operatorname{tr} E'E + \operatorname{tr} E'(XAA'X'A + X'AA'XA). \qquad (11)$$

Let the right-hand side of (11) be defined as the function $h(B)$. Upon substitution of $E = (B - A)$ in the right-hand side of (11) it follows that $h(B)$ can be expressed as

$$h(B) = f(A) - \alpha \operatorname{tr} (B - A)'(B - A) + \operatorname{tr} (B - A)'(XAA'X'A + X'AA'XA). \qquad (12)$$

Function $h(B)$ can be elaborated as

$$h(B) = f(A) - \alpha \operatorname{tr} B'B + 2\alpha \operatorname{tr} B'A - \alpha \operatorname{tr} A'A$$

$$+ \operatorname{tr} B'(XAA'X'A + X'AA'XA) - \operatorname{tr} A'(XAA'X'A + X'AA'XA)$$

$$= f(A) - \alpha p + 2\alpha \operatorname{tr} B'A - \alpha p + \operatorname{tr} B'(XAA'X'A + X'AA'XA) - 2f(A)$$

$$= -f(A) - 2\alpha p + \operatorname{tr} B'(XAA'X'A + X'AA'XA + 2\alpha A). \qquad (13)$$

Let a singular value decomposition of $(XAA'X'A + X'AA'XA + 2\alpha A)$ be given by

$$(XAA'X'A + X'AA'XA + 2\alpha A) = PDQ'. \qquad (14)$$

Define $B_0 \equiv PQ'$. Then, as is well-known (Green, 1969; ten Berge, 1983), B_0 maximizes $h(B)$ subject to $B'B = I_p$. Hence we have $h(B_0) \ge h(A)$. Also, from (12) it follows immediately that $h(A) = f(A)$. Combining these results with (11) yields

$$g(A, B_0) \ge f(A). \qquad (15)$$

Next it will be proven that $f(B_0) \ge f(A)$ by using (15) and the inequality

$$\|B_0'XB_0 - A'XA\|^2 \ge 0. \qquad (16)$$

Expanding (16) yields

$$f(B_0) + f(A) \ge 2 g(A, B_0). \qquad (17)$$

Combining (15) and (17) yields $f(B_0) \geq f(A)$.

Above it has been proven that $f(B_0) \geq f(A)$. From the definition of B_0 it follows that B_0 contains columns that form an orthonormal basis for the column space of $(XAA'X'A + X'AA'XA + 2\alpha A)$. From the definition of $f(A)$ it follows that $f(A) = f(AT)$ for any orthonormal matrix T. Because B and B_0 contain orthonormal basis vectors for the same space, we have $f(B) = f(B_0)$. Hence we have $f(B) \geq f(A)$, which was to be proven.   □

Based on Result 2, an algorithm for maximizing $f(A)$ subject to $A'A = I_p$ is readily given. Updating matrix A as the Gram-Schmidt orthonormalized version of $(XAA'X'A + X'AA'XA + 2\alpha A)$ for any α larger than the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$ always increases function f. Moreover, it follows from $\sigma(A, R) \geq 0$ that, subject to the constraint $A'A = I_p$, function $f(A)$ has an upper bound equal to tr $X'X$. Therefore, iteratively updating A in this way monotonically increases $f(A)$ and must converge to a stable function value of f.

Modification of Takane's Algorithm

The algorithm proposed earlier by Takane is closely related to the algorithm described above. Takane's algorithm consists of updating matrix A by choosing the new A as the Gram-Schmidt orthonormalization of $(XAA'X'A + X'AA'XA)$. Obviously, Takane's algorithm is the special case of the algorithm described above with α chosen equal to zero. However, when the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$ is larger than zero, there is no guarantee that $f(A)$ will increase by means of Takane's method. As has been noted by Harshman and Kiers (1987), Takane's algorithm tends to be very efficient but does not always converge monotonically.

The parameter α in the procedure for updating A can be seen as a parameter that might slow down convergence, because, the larger α, the more the update will resemble its predecessor. In order to take maximal advantage of the efficiency of Takane's method we propose the following algorithm. First compute the update B for A according to Takane's procedure; then evaluate $f(B)$; only in case $f(B) \leq f(A)$ compute the update for A based on the generalized algorithm; repeat this procedure until $f(B) \approx f(A)$. This modified algorithm will also converge monotonically.

In the following sections we will discuss some choices concerning the specific form of the algorithm to be used.

Cases With Closed-Form Solutions for Maximizing $f(A)$

Above we have described a procedure for maximizing $f(A)$ by means of an iterative procedure. In some special cases, however, closed-form solutions for maximizing $f(A)$ exist. These are the case where A has only one column, the case where X is symmetric, the case where X is skew-symmetric and the case where $(X'X + XX')$ has rank p.

In case the number of columns of A is 1, matrix A can be replaced by a vector $\mathbf{a}$, and we have $f(A) = (\mathbf{a}'X\mathbf{a})(\mathbf{a}'X'\mathbf{a}) = (\mathbf{a}'X\mathbf{a})^2 = (\frac{1}{2}\mathbf{a}'(X + X')\mathbf{a})^2$, which is maximal when $\mathbf{a}$ is the eigenvector corresponding to the largest absolute eigenvalue of $(X + X')$.

When X is symmetric the problem of fitting the DEDICOM model basically reduces to the problem of finding a reduced rank approximation of X. A closed-form solution for a reduced rank approximation of any matrix has been given by a theorem of Eckart and Young (1936). In case X is symmetric, this approximation $\hat{X}$ of X can always be written as $\hat{X} = ARA'$, where R is the diagonal matrix containing the p eigenvalues of X that are the p largest eigenvalues of X in the absolute sense, on its diagonal and A contains the corresponding p eigenvectors of X.

When X is skew-symmetric it is useful to distinguish the case where p is even and the case where p is odd. When p is even a closed-form solution exists which is based on the fact that all nonzero singular values of a skew-symmetric matrix have even multiplicity, and every pair of left-hand singular vectors associated with a nonzero singular value consists of the right-hand singular vectors associated with the same singular value, in reversed order, and with reversed sign (Gower, 1977; Harshman, 1981). According to Eckart and Young (1936) a reduced rank approximation of X is given by $\hat{X} = P_p D_p Q_p'$, where P_p and Q_p are column-wise orthonormal matrices containing the first p left and right singular vectors of X, respectively, and D_p is the diagonal matrix containing the largest p singular values of X. Because all succeeding pairs of columns of P_p and Q_p contain the same vectors, but in reversed order, and with reversed sign, a rotation matrix T (consisting of 2×2 blocks along the diagonal with a final unit diagonal element when n is odd) exists such that $Q_p = P_p T$. Choosing $A = P_p$, and $R = D_p T'$, shows that the reduced rank approximation can be written in the form of the DEDICOM model, $\hat{X} = ARA'$. Therefore, the reduced rank approximation yields a closed-form solution for fitting the DEDICOM model to skew-symmetric data when A has an even number of columns.

When X is skew-symmetric and p is odd, the solution for A with $p - 1$ columns can never be improved by a solution for A with p columns. This follows from the fact that skew-symmetric matrices have even rank (Gower, 1977), and that the DEDICOM approximation to a skew-symmetric data matrix X, given by $\hat{X} = AA'XAA'$, is skew-symmetric itself and hence has even rank (Harshman, 1981). Therefore, the rank of $\hat{X}$ is smaller than or equal to $p - 1$, and the best rank p DEDICOM approximation of X is equal to the best rank $p - 1$ DEDICOM approximation of X. In conclusion, when p is odd, it suffices to fit the DEDICOM model for $p - 1$, because the approximation of X can never be improved by using the p-dimensional DEDICOM model.

Finally, when $(X'X + XX')$ has rank p, we have the following closed-form solution. Let $(X'X + XX') = P\Lambda P'$ be an eigen decomposition of $(X'X + XX')$, where $P'P = I_p$, and Λ is a positive definite diagonal $p \times p$ matrix. Then the DEDICOM solution is given by $A = P$, and $R = P'XP$ and yields a perfect fit of the data matrix (de Leeuw, 1983; see Takane, 1985).

The Choice of α

In the description of the generalized algorithm we mentioned that the algorithm converges monotonically whenever α is larger than the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$. However, this does not imply that α is to be chosen equal to this eigenvalue. It should be noted that this eigenvalue depends on the current version of A. Therefore, in order to set α to this value, the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$ should be recomputed in every iteration where α is needed. To avoid this computationally expensive procedure one might use a different choice for α. That is, choosing any α greater than the largest eigenvalue of the symmetric part of $(-X \otimes X)$ guarantees monotonic convergence as well, because the largest eigenvalue of the symmetric part of $(-X \otimes X)$ is never smaller than the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$. The latter can be readily verified by noting that the largest eigenvalue of the symmetric part of any matrix C is equal to the maximum of $e'Ce$ over e subject to $e'e = 1$, see (6) and (7).

The above alternative for choosing α is not the only possibly useful one. Some choices for α are computationally even simpler. Among these the largest singular value of $(X \otimes A'XA)$ and the largest singular value of $(X \otimes X)$ are of particular interest. The computation of these singular values is facilitated by the fact that the largest singular

value of a Kronecker product of matrices is equal to the product of the largest singular values of these matrices. It is readily verified that the values mentioned here are never smaller than the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$.

Apart from these choices of α, the choice $\alpha = 0$ is of particular interest. This is the case where the modified Takane algorithm and Takane's algorithm coincide. As has been remarked above, choosing α equal to 0 may jeopardize monotonic convergence. Although monotonic convergence may occur when $\alpha = 0$, a sufficient condition for monotonic convergence is that the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$ is smaller than or equal to zero. The largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$ is smaller than or equal to zero if and only if the symmetric part of $(X \otimes A'XA)$ is positive semi-definite. Therefore, a sufficient condition for monotonic convergence of Takane's algorithm is that the symmetric part of $(X \otimes A'XA)$ is positive semi-definite for every A that emerges during the iteration process.

This sufficient condition for monotonic convergence of Takane's algorithm is not very useful, because it can only be evaluated during the iterations. Above, it has been remarked that the largest eigenvalue of the symmetric part of $(-X \otimes A'XA)$ is always smaller than or equal to the largest eigenvalue of the symmetric part of $(-X \otimes X)$. Therefore, monotonic convergence of Takane's method is also guaranteed when the largest eigenvalue of the symmetric part of $(-X \otimes X)$ is smaller than or equal to zero. As a result, a sufficient condition for monotonic convergence of Takane's method is that the symmetric part of $(X \otimes X)$ be positive semi-definite. This condition implies that $(X \otimes X) + (X' \otimes X')$ is positive semi-definite. The latter condition may be used to assess in advance whether or not modifying Takane's algorithm is necessary.

The Choice of a Starting Configuration and the Performance of the Algorithm

The closed-form solutions described above are useful for choosing a starting configuration for the matrix of "loadings" A for the DEDICOM algorithm. We propose choosing a starting configuration for A which coincides with the closed-form solutions whenever they are available. Hence in cases with a closed-form solution no iterations will take place. This is achieved by choosing as a start for the matrix A either the matrix whose first p columns contain the eigenvectors of $(X + X')$ corresponding to the p largest absolute eigenvalues of $(X + X')$, or the matrix whose first p columns contain the eigenvectors of $(X'X + XX')$ corresponding to the p largest eigenvalues of $(X'X + XX')$, which are all nonnegative. Obviously, the former start yields the closed-form solution for the cases where matrix X is symmetric or a 1-dimensional solution is required, and the latter start yields the closed-form solution when matrix X is symmetric or $(X'X + XX')$ has rank p. The latter start also yields the closed-form solution when X is skew-symmetric, as will be shown below.

For the case where X is skew-symmetric, the singular value decomposition of X can be given by $X = P_n DT'P_n'$, where T is an orthonormal matrix with blocks of order 2×2 along its diagonal and a final unit diagonal element if n is odd, D is a diagonal matrix containing the singular values of X, and P_n is a column-wise orthonormal matrix (Gower, 1977). It follows that $(X'X + XX') = (P_n TDP_n'P_n DT'P_n') + (P_n DT'P_n'P_n TDP_n') = (P_n TD^2 T'P_n') + (P_n D^2 P_n')$. Because of the special structure of T and the fact that the nonzero singular values of X and hence the nonzero elements of D have even multiplicities (Gower) T commutes with D. Because D has nonnegative elements only, T commutes with D^2 also, hence we have $TD^2T' = TT'D^2 = D^2$. It follows that $(X'X + XX') = 2P_n D^2 P_n'$. This implies that the eigenvectors of $(X'X + XX')$ are given by P_n. Therefore, the first p (p even) singular vectors of X are equal to the first p eigenvectors of $(X'X + XX')$.

The choice of starting configurations that coincide with closed-form solutions when these are known in not very useful in itself. In such cases it suffices to compute the closed-form solution directly. However, such starts may be very useful in cases where the conditions for existence of closed-form solutions are almost satisfied, for example, when X is nearly symmetric or skew-symmetric, or when $(X'X + XX')$ has nearly rank p.

The algorithm has been programmed on a CDC-Cyber main-frame computer. The algorithm has been tested by analyzing 17 random data sets and two empirical data sets (the "Word Association Data" and the "Car Switching Data"; Harshman et al., 1982). In all of these, using the two starting configurations described above resulted in quick convergence that did not even need the modification suggested for Takane's algorithm. There was no consistent difference in computation time depending on whether eigenvectors of $(X + X')$ or of $(X'X + XX')$ were used, but overall the latter start resulted in slightly smaller computation times.

In all the example data sets tested so far, there was no need for modification of Takane's algorithm when iterations started with either of the two above mentioned starting configurations ("rational starts"). However, when other starting configurations are used modification of Takane's algorithm is, in certain cases, necessary to obtain monotonicity. An example of such a (contrived) case is the following. Let matrix X and the starting configuration for A be given by

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & -2 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The value of σ is 9.00 at the start, 6.30 after one iteration and 6.72 after two iterations. Because of this increase of the function value, the modification procedure is needed in this step and, with α chosen equal to the first singular value of $(X \otimes A'XA)$, it yields $\sigma = 5.80$; then again in the next iteration Takane's procedure results in an increase of σ, $\sigma = 6.13$, which is improved by the modified approach ($\sigma = 5.70$); the process continues in this way until convergence. It can be concluded that the modification is necessary in this case to guarantee monotonic convergence and is efficient. It should be noted, however, that using both of the rational starts results in quick monotone convergence of Takane's algorithm, for this contrived data.

Discussion

The algorithm described in the present paper is based on a technique that is called "majorization" (e.g., de Leeuw & Heiser, 1980). This technique essentially consists of minimizing a function by means of approximating it in each iteration by a different function which is always greater than or equal to the original function and coincides with it in at least one point. It can be shown that our algorithm can be presented as a method in which the loss function σ is minimized over A by means of minimizing a function of A that majorizes σ.

The modification of Takane's algorithm by means of taking the Gram-Schmidt orthonormalized version of $(XAA'X'A + X'AA'XA + 2\alpha A)$ instead of that of $(XAA'X'A + X'AA'XA)$ is related to accelerating techniques such as discussed by Ramsay (1975) for slowly converging iterative processes. Ramsay's technique is based on adding a scalar (a damping parameter θ) times the matrix to be updated (A_{old}) to the matrix update based on the original procedure (A_{un}). In case of DEDICOM this procedure would not work without making some adjustments, because ($A_{un} + \theta A_{old}$) is not

necessarily column-wise orthonormal. Obviously, Ramsay's procedure differs from our procedure in that in our procedure $(XAA'X'A + X'AA'XA)$ is used instead of A_{un}.

In the present paper a modification has been proposed for Takane's algorithm. However, with the use of either of the rational starts discussed above, we have not seen any cases where this modification was actually needed. In practice, Takane's original algorithm appears to converge monotonically when rational starts are taken. However, no proof is available that Takane's algorithm will *always* converge monotonically when any of the rational starts is used. The main objective of our modification has been to provide an algorithm with guaranteed monotonic convergence in case Takane's original algorithm would not converge monotonically.

References

de Leeuw, J. (1983). *Notes on DEDICOM*. Unpublished manuscript.

de Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis V* (pp. 501–522). Amsterdam: North-Holland.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika, 1*, 211–218.

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In J. Barra et al. (Eds.) *Recent Developments in Statistics* (pp. 109–123). Amsterdam: North-Holland.

Green, B. F. (1969). Best linear composites with a specified structure. *Psychometrika, 34*, 301–318.

Harshman, R. A. (1978, August). *Models for analysis of asymmetrical relationships among N objects or stimuli*. Paper presented at the First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology, Hamilton, Ontario.

Harshman, R. A. (1981). *DEDICOM multidimensional analysis of skew-symmetric data. Part 1. Theory*. Unpublished technical memorandum, Bell Laboratories, Murray Hill, NJ.

Harshman, R. A., Green, P. E., Wind, Y., & Lundy, M. E. (1982). A model for the analysis of asymmetric data in marketing research. *Marketing Science, 1*, 205–242.

Harshman, R. A., & Kiers, H. A. L. (1987, June). *Algorithms for DEDICOM analysis of asymmetric data*. Paper presented at the European Meeting of the Psychometric Society, Enschede.

Kiers, H. A. L. (1989). An alternating least squares algorithm for fitting the two- and three-way DEDICOM model and the IDIOSCAL model. *Psychometrika, 54*, 515–521.

Ramsay, J. O. (1975). Solving implicit equations in psychometric data analysis. *Psychometrika, 40*, 337–360.

Takane, Y. (1985). Diagonal Estimation in DEDICOM. *Proceedings of the 1985 Annual Meeting of the Behaviormetric Society* (pp. 100–101). Sapporo.

ten Berge, J. M. F. (1983). A generalization of Kristof's theorem on the trace of certain matrix products. *Psychometrika, 48*, 519–523.