PSYCHOMETRIKA—VOL. 46, NO. 2.
JUNE, 1981

COMPUTATIONAL ASPECTS OF THE GREATEST LOWER BOUND TO THE RELIABILITY AND CONSTRAINED MINIMUM TRACE FACTOR ANALYSIS

JOS M. F. ten BERGE, TOM A. B. SNIJDERS AND FRITS E. ZEGERS

UNIVERSITY OF GRONINGEN

In the last decade several algorithms for computing the greatest lower bound to reliability or the constrained minimum-trace communality solution in factor analysis have been developed. In this paper convergence properties of these methods are examined. Instead of using Lagrange multipliers a new theorem is applied that gives a sufficient condition for a symmetric matrix to be Gramian. Whereas computational pitfalls for two methods suggested by Woodhouse and Jackson can be constructed it is shown that a slightly modified version of one method suggested by Bentler and Woodward can safely be applied to any set of data. A uniqueness proof for the solution desired is offered.

Key words: communality, internal consistency, Heywood case, positive semidefinite.

1. Introduction

In the last decade several authors have adressed the problems of finding the Gramian matrix Σ_t that has the minimal trace while either satisfying the constraints

$$\Sigma_x = \Sigma_t + (\Sigma_x - \Sigma_t) = \Sigma_t + \Sigma_e \tag{1}$$

where Σ_e is diagonal, or satisfying the constraints

$$\Sigma_x = \Sigma_t + (\Sigma_x - \Sigma_t) = \Sigma_t + \Sigma_e \tag{2}$$

for Σ_e both diagonal and Gramian. In either case, Σ_x is an observed covariance matrix that is to be decomposed in a common (true-score) and a unique (error) part. We shall refer to the minimum-trace Σ_t that satisfies (1) as the minimum-trace factor analysis (MTFA) solution.

The minimum-trace Σ_t that satisfies (2) is referred to as the constrained minimum-trace factor analysis (CMTFA) solution in this paper. The solutions differ only if the Σ_e in the MTFA solution has one or more negative entries. The solutions have also implications for reliability theory. The CMTFA solution corresponds to the greatest lower bound (g.l.b.) to reliability defined by Jackson and Agunwamba [1977] as

$$\rho = 1 - \frac{\text{tr } \Sigma_e^*}{1'\Sigma_x 1} \tag{3}$$

where Σ_e^* is the Σ_e that has the maximal trace subject to the constraints (2). The MTFA solution has implications for reliability theory as well, cf. Bentler [1972] and Bentler & Woodward [1980], but we shall ignore all lower bounds to reliability except the greatest.

The concept of minimum trace factor analysis has been discussed in the early work of Ledermann [1939] who noted that minimum-trace and minimum-rank solutions do not

The authors are obliged to Charles Lewis and Dirk Knol for helpful comments, and to Frank Brokken and Henk Camstra for developing computer programs.

Requests for reprints should be sent to Jos M. F. ten Berge, Subfaculteit voor de Psychologie der Rijksuniversiteit, Oude Boteringestraat 34, 9712 GK Groningen. THE NETHERLANDS.

0033-3123/81/0600-2933$00.75/0

always coincide. (The authors are obliged to Jan de Leeuw for pointing out this and other sources). Computational methods for MTFA, however, seem not to have been considered before Bentler's work [Bentler, 1972]. The MTFA solution offered by Bentler is referred to as "Bentler's method" throughout this paper. Woodhouse and Jackson [1977] offered a solution to the CMTFA-problem. This solution is labelled the Woodhouse-Jackson procedure (WJP) here. In addition, Woodhouse and Jackson [1977, p. 591] outlined a partialling procedure that may be used to extend Bentler's method into a CMTFA-solution. Bentler and Woodward [1980, eq. 3.24] gave a similar partialling procedure. They also derived a second CMTFA method (eq. 3.26) as an alternative to their extension of Bentler's method. This second method is referred to as the Bentler-Woodward procedure (BWP).

In Section 2 some basic facts needed in later sections are summarized. The WJP is discussed in Section 3, Bentler's method and the two partialling procedures are discussed in Section 4, and the BWP is discussed in Section 5, along with a slightly modified procedure that is referred to as a "modified Bentler-Woodward procedure" (MBWP).

The methods can be treated either as reliability methods or as communality methods; we shall use both interpretations interchangeably, taking Σ_t for either a common-part covariance matrix or a true score covariance matrix, and taking Σ_e either for a uniqueness matrix or for a matrix of error variances.

2. *Some General Principles*

This section contains a number of principles, stated as lemmas, that do not relate to any of the computational methods specifically.

Lemma 2.1. Every CMTFA Solution Is Generically Singular

We shall call a linearly dependent set of variables "generically linearly dependent" if each variable in the set is a linear composite of the other variables, and "specifically linearly dependent" otherwise. The covariance matrix associated with the set is called "generically singular" or "specifically singular", respectively. It can be shown that a CMTFA solution is generically singular.

Proof. Suppose we have a Σ_t that is not a generically singular matrix. Then any set of variables that has Σ_t as covariance matrix contains at least one variable that has a non-vanishing residual in the regression on the other variables. Removing this residual yields a new covariance matrix equal to Σ_t except that the variance of the residual has been subtracted from one particular diagonal entry. Hence the trace of Σ_t was not minimal. This shows that generic singularity is necessary for a CMTFA solution. It is not sufficient, however; cf. the data presented in (21) and the text below it. Generic singularity of the MTFA solution can be shown by the same proof. □

It can be demonstrated that Σ_t is generically singular if and only if

$$\Sigma_t a = 0 \tag{4}$$

for some vector a without any zero elements.

Proof. If Σ_t is generically singular then there exist vectors $b_1, b_2, \ldots, b_n$ with $\Sigma_t b_i = 0$ and $b_{ii} = 1$ for all i. Let a be a vector in the null-space of Σ_t with the minimal number of zero elements. Suppose that $a_i = 0$ for some i. If $c \neq a_j/b_{ij}$ for every j with $b_{ij} \neq 0$ then $a + cb_i$ would be a vector in the null-space of Σ_t with fewer zero elements than a. It follows that $a_i \neq 0$ for all i. The proof of the converse is obvious. □

Lemma 2.2. A Set of Variables Is Generically Linearly Dependent If and Only If the G.L.B. of Their Sum Is Unity

Clearly, the g.l.b. is unity if and only if $\Sigma_e = 0$ is the only Σ_e compatible with (2). Let Σ_x be generically singular; then from (4) we have

$$a'\Sigma_x a = 0 \tag{5}$$

for some vector a with $a_i \neq 0, i = 1, \ldots, n$. For every $\Sigma_e \neq 0$ we have

$$a'(\Sigma_x - \Sigma_e)a < 0 \tag{6}$$

implying that $(\Sigma_x - \Sigma_e)$ is not Gramian. Therefore, Σ_e must be the null matrix of order n, and the g.l.b. is unity. Conversely, if Σ_x is not generically singular, then by Lemma 2.1. it cannot correspond to a CMTFA solution and hence the g.l.b. is less than unity.

If the g.l.b. is unity, then principal component analysis, image analysis and common-factor analysis with squared multiple correlations (covariances) on the diagonal of Σ_t coincide with CMTFA. A conventional lower bound to the reliability that is unity here is Guttman's λ_6 [Guttman, 1945]. If the g.l.b. is not unity then λ_6 still implies diagonal values of Σ_t that correspond to squared multiple covariances in common-factor analysis [Nicewander, 1975]. The g.l.b. implies diagonal values of Σ_t that are in between the observed variances and the squared multiple covariances. In this respect, CMTFA is a compromise between principal components analysis and common-factor analysis with squared multiple covariances on the diagonal of the covariance matrix.

Lemma 2.3. The Set of the Σ_e for Which $(\Sigma_x - \Sigma_e)$ Is Gramian, Is Convex

Let θ be the vector that contains the diagonal entries of Σ_e, and let θ vary over R^n. Then the set

$$C = \{\theta \mid (\Sigma_x - \Sigma_e) \text{ is Gramian}\} \tag{7}$$

is a convex set. Also the subset of C

$$S = \{\theta \mid (\Sigma_x - \Sigma_e) \text{ and } \Sigma_e \text{ Gramian}\} \tag{8}$$

is convex. The proof is straightforward and has been given by Woodhouse and Jackson [1977, p. 582]. The MTFA and CMTFA solutions correspond to the vector θ having the maximal sum of elements in C and S, respectively.

Lemma 2.4. The CMTFA Solution Is Unique

Suppose there exist two "best" (i.e., having the maximal sum of elements) points θ_1 and θ_2 in S, with associated true score covariance matrices $\Sigma_t^{(1)}$ and $\Sigma_t^{(2)}$. The convexity of S implies that $\theta^* = \lambda\theta_1 + (1 - \lambda)\theta_2 (0 < \lambda < 1)$ is also in S, with associated true score covariance matrix $\Sigma_t^* = \lambda\Sigma_t^{(1)} + (1 - \lambda)\Sigma_t^{(2)}$. Clearly, θ^* is also a "best" point. By Lemma 2.1. Σ_t^* is generically singular. Hence there is a vector a without zero elements with $\Sigma_t^* a = 0$. The null-space of a sum of Gramian matrices is the intersection of the null-spaces of these matrices. Hence $\Sigma_t^{(1)} a = 0$ and $\Sigma_t^{(2)} a = 0$ and

$$(\Sigma_e^{(1)} - \Sigma_e^{(2)})a = (\Sigma_t^{(2)} - \Sigma_t^{(1)})a = 0, \tag{9}$$

where $\Sigma_e^{(i)}$ has the elements of θ_i on its diagonal, $i = 1, 2$. Since a has no zero elements, (9) yields $\theta_1 = \theta_2$. This concludes the proof of the uniqueness of the CMTFA solution. The same line of proof can be applied for establishing the uniqueness of the MTFA solution.

The uniqueness of the CMTFA solution has previously been shown in a very complicated manner by Hakim, Lochard, Olivier and Térouanne [Note 2], who also proved that the CMTFA solution depends continuously on Σ_x. Della Riccia and Shapiro [Note 1] have independently arrived at a uniqueness proof closely related to ours.

Lemma 2.5. A Sufficient Condition for a Symmetric Matrix to be Positive Semi-Definite

In this section a theorem is derived that gives a sufficient condition for a symmetric matrix to be Gramian. The theorem plays an essential role in later developments.

Define the set

$$M(D) = \{T \mid (\text{diag } TT') = D^2\} \tag{10}$$

where D is some positive definite diagonal matrix of order n and T can be any matrix of order $n \times r$ ($r \leq n$). Clearly, $M(D)$ is the set of all $n \times r$ matrices with row sums of squares as specified by D.

Theorem 1. If G is a $n \times n$ symmetric matrix for which there exists a $T_0 \in M(D)$ with

$$GT_0 = 0 \tag{11}$$

and which satisfies

$$\text{tr } T'GT \geq 0 \tag{12}$$

for all $T \in M(D)$ then G is positive semi-definite.

Proof. It can be assumed that T_0 is a square matrix (if not, we can add zero columns to make it square) with rank less than n (if not, then the theorem holds trivially). Let $t'_1, t'_2, \ldots, t'_n$ denote the n rows of T_0. There is a unit-length vector a orthogonal to these rows. Let x be an arbitrary vector in R^n. The theorem will be proven by showing that $x'Gx \geq 0$. Construct the matrix T_ε for any ε such that

$$T_\varepsilon = \begin{pmatrix} d_1(d_1^2 + \varepsilon^2 x_1^2)^{-1/2}(t'_1 + \varepsilon x_1 a') \\ \vdots \\ d_n(d_n^2 + \varepsilon^2 x_n^2)^{-1/2}(t'_n + \varepsilon x_n a') \end{pmatrix}. \tag{13}$$

It is clear that $T_\varepsilon \in M(D)$; hence (12) implies

$$\text{tr } T'_\varepsilon GT_\varepsilon \geq 0. \tag{14}$$

For $\varepsilon \neq 0$ define the matrix

$$S_\varepsilon = \varepsilon^{-1}(T_\varepsilon - T_0). \tag{15}$$

From (11), (14) and (15) it follows that

$$\text{tr } S'_\varepsilon GS_\varepsilon = \varepsilon^{-2} \text{ tr } T'_\varepsilon GT_\varepsilon \geq 0. \tag{16}$$

Row i of S_ε is the vector

$$\varepsilon^{-1}(d_i(d_i^2 + \varepsilon^2 x_i^2)^{-1/2} - 1)t'_i + (d_i^2 + \varepsilon^2 x_i^2)^{-1/2} d_i x_i a' \tag{17}$$

which tends to $x_i a'$ as ε tends to zero. Thus

$$\lim_{\varepsilon \to 0} S_\varepsilon = xa'. \tag{18}$$

Now tr $S'_\varepsilon GS_\varepsilon$ is a continuous function of the elements of S_ε. Therefore from (17) and (18) we have

$$\begin{aligned} \lim_{\varepsilon \to 0} \text{tr } S'_\varepsilon GS_\varepsilon &= \text{tr } \left(\lim_{\varepsilon \to 0} S_\varepsilon\right)' G \left(\lim_{\varepsilon \to 0} S_\varepsilon\right) = \text{tr } ax'Gxa' \\ &= \text{tr } a'ax'Gx = \text{tr } x'Gx = x'Gx. \end{aligned} \tag{19}$$

From (16) and (19) it follows that

$$x'Gx \geq 0, \tag{20}$$

which completes the proof of the theorem.

3. *The Woodhouse-Jackson Procedure*

The Woodhouse-Jackson procedure (WJP) can be pictured geometrically as a method for moving from the origin to the point in S (see Lemma 2.3.) that has the largest sum of coordinates. It can also be described in algebraic terms, as an iterative sequence of steps of two types: The first step is subtraction of the smallest eigenvalue of Σ_t from its diagonal elements, which produces a better Σ_t^* unless Σ_t is singular. The second step is the change in any two diagonal entries of Σ_t such that their sum remains constant while the resulting Σ_t^* corresponds to a point in S and is again nonsingular.

The first step decreases all diagonal elements of Σ_t by the same amount. The second step does not affect tr Σ_t but merely serves to make a new first step possible. If the best point has been attained, no more steps of either type are possible. Woodhouse and Jackson infer this from the convexity of S. One might think that the reverse holds as well. However, as will be shown below, the impossibility of moving by either the first or second step is not sufficient for a minimum of tr Σ_t.

The second step of the WJP, changing the ith and jth diagonal entries of Σ_t, can be described in the following way. Consider a true score covariance matrix $\Sigma_t^{(1)}$, corresponding to the point θ_1 in S. Let $t_1, \ldots, t_n$ be a set of true score variables having $\Sigma_t^{(1)}$ as covariance matrix. Then t_i and t_j can each be expressed as the sum of the linear regression on the remaining $n-2$ true score variables plus the residual from that regression. A new pair of variables, t_i^* and t_j^*, can be constructed by exchanging the residuals of t_i and t_j. After replacing t_i and t_j by t_i^* and t_j^* we obtain a new set of n variables with a covariance matrix $\Sigma_t^{(2)}$ that equals $\Sigma_t^{(1)}$ except for the ith and jth diagonal element, where the residual variance parts have been exchanged. Since $\Sigma_t^{(2)}$ is a covariance matrix it corresponds to a point in C. It can be verified that $\Sigma_t^{(1)}$ and $\Sigma_t^{(2)}$ have the same trace, rank and determinant. Let θ_2 be the point in C corresponding with $\Sigma_t^{(2)}$. Then the second step of the WJP consists of moving half-way from θ_1 to θ_2, provided that the new point is in S (if not, then the WJP takes a boundary point of S as new point). This can be interpreted as "going half-way to the point of exchanging residuals." This interpretation is obscured in the computational approach of the WJP. However, it can readily be derived from Woodhouse & Jackson [1977, eq. (5.2)] after partitioning the determinant in the same way as was done on page 583 of their paper.

The regression interpretation of the second step of the WJP proves in a nongeometric way that this step must yield a new Σ_t that is again Gramian. It also shows that the WJP need not converge to the best point in S. If the WJP attains a nonoptimal θ for which all residuals vanish (implying a rank of $\Sigma_t \leq n-2$) then the WJP has converged to a nonoptimal point. For example, consider the matrix

$$\Sigma_x = \begin{pmatrix} \frac{5}{3} & 1 & 2 \\ 1 & \frac{5}{2} & 3 \\ 2 & 3 & 7 \end{pmatrix} \tag{21}$$

with smallest eigenvalue 1. Reducing the diagonal entries by 1, as the first step of the WJP prescribes, will yield a Σ_t of rank one, with tr $\Sigma_t = 8\frac{1}{6}$. Now all residuals vanish, and the WJP breaks down at a nonoptimal point. (The best point has diagonal elements 1, 2 and 5 for Σ_t.)

This flaw in the WJP can also be demonstrated in the terminology of Woodhouse and Jackson [1977]. Basic to the step of type (ii) in their algorithm is the solution of their equation (5.2.), which reads

$$\det(\Sigma_t^{(1)} - kU) = 0, \tag{22}$$

where U is a matrix with $u_{ii} = 1$, $u_{jj} = -1$ and remaining elements zero. This is a quadratic equation in k with roots $k_1 = 0$ (the matrix $\Sigma_t^{(1)}$ is singular) and k_2. The matrix $\Sigma_t^{(2)}$ introduced above is given by $\Sigma_t^{(2)} = \Sigma_t^{(1)} - k_2 U$.

It follows from elementary properties of the determinant function that

$$\det(\Sigma_t^{(1)} - kU) = \det \Sigma_t^{(1)} - kM_{ii} + kM_{jj} - k^2 M_{iijj}, \tag{23}$$

where M_{ii} and M_{jj} are minors and M_{iijj} is the subdeterminant of $\Sigma_t^{(1)}$ obtained by deleting the ith and jth rows and columns. Now suppose that $\Sigma_t^{(1)}$ has rank $n - 2$ or less; then

$$\det(\Sigma_t^{(1)} - kU) = -k^2 M_{iijj}. \tag{24}$$

Hence if $M_{iijj} > 0$, then the only solution of (22) is $k = 0$, and the algorithm is stuck in a point which is not necessarily optimal; if $M_{iijj} = 0$, then every k is a solution of (22) and it is not clear how the algorithm proceeds.

A second type of flaw in the WJP has to do with the procedure by which the WJP avoids negative error variances, namely by never leaving the set S. Consider the 3×3 covariance matrix with all elements 1, except $\sigma_3^2 = 2$. The set S now consists of only a line segment. The WJP needs at least two dimensions to leave the origin, hence it produces a g.l.b. of 1 instead of the correct value 0.9.

Although it may be possible to remedy these and other flaws in the WJP, once they are recognized, there are fundamental problems with the WJP that would seem to remain. Firstly, the WJP has no general criterion for determining after convergence whether or not the best point in S has been attained. This is especially threatening since the WJP approaches the g.l.b. from above. Nonoptimal points for the WJP do not correspond to lower bounds to the reliability at all. Secondly, as the number of variables increases, computing time for the WJP increases sharply. For these reasons, the WJP is not very attractive. We shall examine alternative methods in the next two sections.

4. Bentler's Method, and Partialling Procedures

Bentler's method aims at finding the MTFA solution. In his derivation Bentler [1972] applies calculus to the problem of minimizing tr Σ_t. Define $H^2 = \text{diag } \Sigma_t$ and $\Sigma_0 = \Sigma_x - \text{diag } \Sigma_x$. Taking first derivatives with respect to the elements of H^2 and setting these equal to zero yields the equation

$$(\Sigma_0 + H^2)(I + \Lambda) = 0 \tag{25}$$

where Λ is a symmetric matrix of Lagrange multipliers with zero diagonal entries. Bentler proceeds by claiming that $(I + \Lambda)$ is "clearly Gramian" [Bentler, 1972, p. 348] without offering any proof. If $(I + \Lambda)$ is Gramian, it can be factored as

$$(I + \Lambda) = TT' \tag{26}$$

where T is a matrix with rows of length one. From (25) and (26) Bentler arrives at the equation

$$\Sigma_0 T = -H^2 T \tag{27}$$

as a necessary condition for the minimum of tr H^2. In order to solve this equation he suggests taking an arbitrary T with $\text{diag}(TT') = I$ and iteratively replacing row i of T by

$$t_i' = -(R_i' TT' R_i)^{-1/2} R_i' T \tag{28}$$

where R_i' is row i of Σ_0. (Bentler and Woodward, 1980, have shown that (28) monotonely reduces tr $T'\Sigma_x T$, or equivalently, tr $T'\Sigma_0 T$, over the set L defined below.) The procedure stops if and only if each row of T satisfies (28), which can be summarized as

$$\Sigma_0 T = -\Gamma^2 T \tag{29}$$

for some nonnegative diagonal matrix Γ^2, which is a candidate for H^2. If $(\Sigma_0 + \Gamma^2)$ is Gramian then it is taken as the MTFA solution. Formula (29) represents a necessary condition for the minimum of tr $T'\Sigma_x T$.

Bentler and Woodward [1980] essentially repeated the deficient derivation of equations (26) and (27), cf. their page 257. They have not shown whether or not Bentler's method must always yield a set of values that, if inserted in the diagonal of Σ_0, yield a Gramian matrix. Nor have they shown that this matrix, if Gramian, must be the MTFA solution. Their classical calculus approach in fact can only yield necessary conditions for the solution wanted. Stronger results can be obtained from a convex analysis approach, cf. Della Riccia & Shapiro [Note 1], or from a relatively simple noncalculus approach that shall be pursued here.

Define the set L by

$$L = M(I) \tag{30}$$

[see (10)] where I is the $n \times n$ identity matrix. Then for every $T \in L$ and every diagonal matrix H^2 for which $(\Sigma_0 + H^2)$ is Gramian, we have

$$\text{tr } T'(\Sigma_0 + H^2)T \geq 0. \tag{31}$$

It can be concluded that

$$\text{tr } H^2 = \text{tr } TT'H^2 = \text{tr } T'\underline{H^2 T} \geq -\text{tr } T'\Sigma_0 T. \tag{32}$$

Denote the minimum possible value of tr H^2 by m_1 and the minimum of tr $T'\Sigma_0 T$ for $T \in L$ by m_2. Then (32) implies

$$m_1 \geq -m_2. \tag{33}$$

If a $T_m \in L$ and an H^2 can be found for which

$$\text{tr } H^2 = -\text{tr } T_m'\Sigma_0 T_m \tag{34}$$

while $(\Sigma_0 + H^2)$ is Gramian then $m_1 = -m_2$ and the minimum of tr H^2 has been found. This shows that a Gramian $(\Sigma_0 + \Gamma^2)$ is a sufficient condition for having attained the global minimum. The following theorem shows that this condition is also necessary and that (33) holds as an equality.

Theorem 2. At the global minimum of tr $T'\Sigma_0 T$ over L the associated $(\Sigma_0 + \Gamma^2)$ is Gramian.

Proof. Let the global minimum of tr $T'\Sigma_0 T$ be attained for $T = T_m$. Then by definition,

$$\text{tr } T_m'\Sigma_0 T_m \leq \text{tr } T'\Sigma_0 T \tag{35}$$

for all $T \in L$, and from (29) we have

$$(\Sigma_0 + \Gamma_m^2)T_m = 0 \tag{36}$$

for some nonnegative diagonal matrix Γ_m^2. Premultiplying (36) by T_m' and taking the trace yields

$$\text{tr } T_m'\Sigma_0 T_m = -\text{tr } T_m'\Gamma_m^2 T_m = -\text{tr } \Gamma_m^2. \tag{37}$$

From (35) and (37) it follows that

$$-\text{tr } \underline{\Gamma_m^2} \leq \text{tr } T'\Sigma_0 T \tag{38}$$

for all $T \in L$ or, equivalently,

$$\text{tr } T'(\Sigma_0 + \Gamma_m^2)\underline{T} \geq 0 \tag{39}$$

for all $T \in L$. From Theorem 1 applied to (36) and (39), taking $D = I$, it follows that $(\Sigma_0 + \Gamma_m^2)$ is Gramian. This completes the proof of Theorem 2.

It can be concluded that Bentler's method will, if it avoids local minima, yield the MTFA solution or, equivalently, the best point in the set C. From the uniqueness of this point it follows that the global minimum of tr $T'\Sigma_0 T$ generates a unique matrix Γ_m^2, whereas T_m is at best unique up to a rotation. Bentler's method is caught at a local minimum if and only if the associated $(\Sigma_0 + \Gamma^2)$ is non-Gramian. Hence, local minima can be identified by computing the smallest eigenvalue of this matrix. When a local minimum is attained restarting the iterations from a new initial matrix T is mandatory. This may occur when the rank of the former starting matrix was too small. The authors have been using starting matrices of rank $n - s$, where s is the number of nonnegative eigenvalues of Σ_0. This number is a lower bound to the rank of $(\Sigma_0 + H^2)$, cf. Mulaik [1972, p. 142]. Hence $n - s$ is an upper bound to the rank of T_m. In our experience, a rank of $n - s$ for the starting matrix is large enough to avoid local minima. Although our choice of starting matrices differs from the one suggested by Bentler and Woodward [1980, p. 261], both their approach and ours give rapid convergence to the global minimum.

It should be remembered that Bentler's method yields the MTFA solution, not necessarily the CMTFA solution. That is, negative error variances may occur. In fact, they do occur quite often. The associated reliability coefficient is then still a lower bound to the reliability, although no longer the greatest lower bound. In the context of MTFA one would have a so-called Heywood case. Woodhouse and Jackson [1977, Section 3] have suggested dealing with this situation by setting all negative error (unique) variances to zero, partialling out the corresponding observed variables from Σ_x, and restarting the computations from the residual covariance matrix for the remaining variables, a procedure that obviously may have to be repeated in order to attain a solution free of negative error variances. However, this partialling method is not correct. The following counterexample shows this quite clearly.

Consider the covariance matrix for three variables,

$$\Sigma_x = \begin{pmatrix} 7.1 & 10 & 20 \\ 10 & 17 & 30 \\ 20 & 30 & 60 \end{pmatrix}. \tag{40}$$

Bentler's method here yields a matrix $T_m' = (1 \quad 1 \quad -1)$ and a Γ_m^2 with diagonal entries 10, 20 and 50, which implies error variances of -2.9, -3, and 10. Setting all negative error variances to zero and restarting the computations produces new error variances 0, 0 and $\frac{520}{207}$, with tr $\Sigma_e = 2.512$. However, if only the first error variance is set to zero we obtain the better error variances 0, $\frac{77}{71}$, and $\frac{430}{71}$, with tr $\Sigma_e = 2.915$. This shows that the partialling method suggested by Woodhouse and Jackson is not correct.

A more promising approach to the partialling problem has been offered by Bentler and Woodward [1980, eqs. (3.23), (3.24)]. Their approach also starts by partialling out the variables that have negative error variances in the MTFA solution, and restarting the computations from the residual matrix. Again, variables having negative error variances in the partial MTFA solution are partialled out. However, variables that previously had been partialled out will return into the computations if the corresponding row of the matrix T defined by the Bentler & Woodward equations (3.23) and (3.24) has a sum of squares less than one. The partialling procedure is terminated if all rows of this matrix T have length one or greater and all negative error variances have disappeared. Bentler and Woodward have not shown that their partialling procedure must converge. However, if it does converge then it generates a nonnegative diagonal matrix Δ. From the results to be derived in Section 5 below [equations (44) and (57)] it can be shown that this matrix can be taken as

$\Delta = \Sigma_e$ in the CMTFA solution if and only if $(\Sigma_x - \Delta)$ is a Gramian matrix. Thus, if the Bentler-Woodward partialling method converges then it can be verified whether or not the CMTFA solution has been attained.

There are two problems with the Bentler-Woodward partialling method. First, a convergence proof has not been given. Second, the iterative use of Bentler's method and partialling may require quite laborious computations. A much faster method with more completely derived convergence properties will be discussed in the next section.

5. *The Bentler-Woodward Procedure and a Modified Version*

In Section 4 it was shown that Bentler's method is highly attractive as an MTFA method. Fortunately, a CMTFA solution, very similar to Bentler's method, has been offered by Bentler and Woodward [1980, eq. (3.26)]. A theoretical justification for this Bentler-Woodward procedure (BWP) can be facilitated by slightly modifying it into a modified Bentler-Woodward procedure (MBWP). To simplify the presentation we first derive the MBWP, examine its properties and then compare it to the BWP. Again, as in Section 4, Lagrange multipliers are not needed.

Define the set

$$M = \{T \mid TT'_{ii} \geq 1,\ i = 1, 2, \ldots, n\} \tag{41}$$

where T is of order $n \times r, r \leq n$. For any Σ_e compatible with (2) we have

$$\text{tr } T'(\Sigma_x - \Sigma_e)T \geq 0 \tag{42}$$

for all $T \in M$, or, equivalently

$$\text{tr } TT'\Sigma_e \leq \text{tr } T'\Sigma_x T \tag{43}$$

for all $T \in M$. From (41) and (43) it follows that

$$\text{tr } \Sigma_e \leq \text{tr } T'\Sigma_x T \tag{44}$$

for all $T \in M$. In particular we have

$$\max \text{tr } \Sigma_e \leq \min_M \text{tr } T'\Sigma_x T. \tag{45}$$

If we find a Σ_e, compatible with (2), and a $T \in M$ for which (44) holds as an equality, then that Σ_e gives the CMTFA-solution.

Obviously, we shall now want to examine the minimization of tr $T'\Sigma_x T$ over M. As in Bentler's method, each row of a starting matrix chosen from M can be iteratively replaced by the best row, keeping the other rows fixed. Specifically, we can write, for a given i, in the notation of Section 4,

$$\begin{aligned} \text{tr } T'\Sigma_x T = \text{tr } TT'\Sigma_x &= \sum_{g,h} t'_g t_h \sigma_{gh} \\ &= t'_i t_i \sigma_i^2 + 2 \sum_{\substack{h=1 \\ h \neq i}}^{n} t'_i t_h \sigma_{ih} + c_i \\ &= t'_i t_i \sigma_i^2 + 2t'_i T'R_i + c_i \end{aligned} \tag{46}$$

where the scalar c_i and the vector $T'R_i$ are constant with respect to t_i. The problem is to minimize (46) as a function of t_i, subject to $t'_i t_i \geq 1$.

First we consider the case $(R'_i TT'R_i) > 0$. Necessary for a minimum is that $t'_i T'R_i$ is minimal given the value of $t'_i t_i$. From the Schwarz inequality it follows that t_i must be of the form

$$t'_i = -\lambda R'_i T \tag{47}$$

for some $\lambda > 0$. This implies that we need only minimize the function

$$f(\lambda) = \lambda^2 \sigma_i^2 (R_i' TT' R_i) - 2\lambda (R_i' TT' R_i) + c_i \tag{48}$$

which is a parabola with a unique minimum for $\lambda = \sigma_i^{-2}$. The constraint $t_i' t_i \geq 1$ can be expressed as

$$\lambda \geq (R_i' TT' R_i)^{-1/2}. \tag{49}$$

Clearly, if $\sigma_i^2 \leq (R_i' TT' R_i)^{1/2}$ then this constraint is inactive and the optimal λ is $\lambda = \sigma_i^{-2}$, which implies that we must take

$$t_i' = -\sigma_i^{-2} R_i' T \qquad \text{if } 0 < \sigma_i^2 \leq (R_i' TT' R_i)^{1/2}. \tag{50a}$$

If $\sigma_i^2 > (R_i' TT' R_i)^{1/2}$ then the function (48) is monotonely increasing for λ satisfying (49). Hence the minimum is attained for $\lambda = (R_i' TT' R_i)^{-1/2}$ which implies that we must take

$$t_i' = -(R_i' TT' R_i)^{-1/2} R_i' T \qquad \text{if } 0 < (R_i' TT' R_i)^{1/2} < \sigma_i^2. \tag{50b}$$

Finally, if $(R_i' TT' R_i) = 0$ then (46) reduces to $\sigma_i^2 t_i' t_i + c_i$. If $\sigma_i^2 > 0$ then any t_i with $t_i' t_i = 1$ is optimal; if $\sigma_i^2 = 0$ then any t_i with $t_i' t_i \geq 1$ is optimal. We suggest taking

$$t_i' = (s_i' s_i)^{-1/2} s_i' \qquad \text{if } (R_i' TT' R_i) = 0, \tag{50c}$$

where s_i' denotes the previous version of row i of T that is to be replaced by t_i'.

Algorithm (50), applied to the rows of a starting matrix from the set M [see (41)], is the MBWP. It gives the best replacement of row s_i' by t_i' subject to the constraint $t_i' t_i \geq 1$. Since s_i itself is also a candidate for t_i, replacing s_i by t_i according to the MBWP can never increase the value of tr $T'\Sigma_x T$. The latter function is bounded from below. Therefore, the MBWP must converge if it is terminated when no further significant reduction of tr $T'\Sigma_x T$ has been obtained in a set of n row replacements. Convergence occurs if and only if each row of T satisfies (within limits of accuracy)

$$R_i' T = -\sigma_i^2 t_i' \qquad \text{with } t_i' t_i \geq 1, \tag{51a}$$

or

$$R_i' T = -(R_i' TT' R_i)^{1/2} t_i' \qquad \text{with } t_i' t_i = 1. \tag{51b, c}$$

We then have, for some $T \in M$

$$\Sigma_0 T = -\Gamma^2 T \tag{52}$$

where the diagonal matrix Γ^2 has diagonal elements

$$\gamma_i^2 = \sigma_i^2 \qquad \text{with } t_i' t_i \geq 1 \tag{53a}$$

or

$$\gamma_i^2 = (R_i' TT' R_i)^{1/2} < \sigma_i^2 \qquad \text{with } t_i' t_i = 1. \tag{53b, c}$$

Formula (52) gives a necessary condition for the minimum of tr $T'\Sigma_x T$. The diagonal matrix

$$\Delta = [(\text{diag } \Sigma_x) - \Gamma^2] \tag{54}$$

is Gramian and satisfies

$$(\Sigma_x - \Delta) T = 0. \tag{55}$$

If the ith diagonal element of Δ is positive then row i of T has length one [see (53b, c)]. Thus we have

$$\text{tr } \Delta = \text{tr } TT'\Delta. \tag{56}$$

From (55) and (56) we have

$$\text{tr } \Delta = \text{tr } T'\Sigma_x T. \tag{57}$$

If Δ leaves $(\Sigma_x - \Delta)$ Gramian then from (44) and (57) it follows that we can take $\Sigma_e = \Delta$ and thus have obtained the CMTFA solution. In other words, a Gramian $(\Sigma_x - \Delta)$ is sufficient for the global minimum of tr $T'\Sigma_x T$ over M. Theorem 3 will show that it is also necessary or, equivalently, that (45) holds as an equality. Thus the MBWP generates the CMTFA solution if and only if it attains the global minimum of tr $T'\Sigma_x T$.

Theorem 3. At the global minimum of tr $T'\Sigma_x T$ over M the associated $(\Sigma_0 + \Gamma^2) = (\Sigma_x - \Delta)$ is Gramian.

Proof. Let the global minimum of tr $T'\Sigma_x T$ over M be attained for a $T_m \in M$ with associated Δ_m. Define the set

$$N = \{T \mid (\text{diag } TT') = (\text{diag } T_m T_m')\}. \tag{58}$$

where T is of order $n \times r$ ($r \leq n$). From (57) and the definition of T_m we have

$$\text{tr } \Delta_m = \text{tr } T_m'\Sigma_x T_m \leq \text{tr } T'\Sigma_x T \tag{59}$$

for all $T \in M$, and from (56) and (58) it follows that

$$\text{tr } \Delta_m = \text{tr } T_m T_m'\Delta_m = \text{tr } TT'\Delta_m = \text{tr } T'\Delta_m T \tag{60}$$

for all $T \in N$. Since N is a subset of M we obtain from (59) and (60)

$$\text{tr } T'(\Sigma_x - \Delta_m)T \geq 0 \tag{61}$$

for all $T \in N$, while (55) implies

$$(\Sigma_x - \Delta_m)T_m = 0. \tag{62}$$

Note that $T_m \in N$.

From Theorem 1 applied to (61) and (62), taking $M(D) = N$, it follows that $(\Sigma_x - \Delta_m)$ is Gramian. This completes the proof of Theorem 3.

It can readily be verified that T_m has at least one row of length one, unless $\Sigma_x T_m = 0$, which occurs if and only if Σ_x is generically singular. From the uniqueness of the CMTFA solution it follows that there is a unique matrix Δ_m associated with the global minimum of tr $T'\Sigma_x T$.

Having established local and global convergence properties for the MBWP we now compare it to the BWP. In our notation, the BWP is essentially given by

$$t_i' = s_i' \qquad \text{if } (R_i' TT' R_i) = 0, \tag{63c}$$

$$t_i' = -(R_i' TT' R_i)^{-1/2} R_i' T \qquad \text{if } 0 < (R_i' TT' R_i)^{1/2} < \sigma_i^2 \qquad \text{and} \qquad g^{(i)} \geq g^{(i+1)} \tag{63b}$$

where $g^{(i)}$ is the value of tr $T'\Sigma_x T$ before replacing row i of T and $g^{(i+1)}$ is the value after, and

$$t_i' = -\sigma_i^{-2} R_i' T \text{ otherwise,} \tag{63a}$$

cf. Bentler & Woodward [1980, eq. (3.26)]. In the BWP a starting matrix is taken from the set M and when (63) converges to a T that is not in M then the rows with $t't < 1$ are normalized and (63) is rerun [Bentler & Woodward, 1980, p. 259].

Bentler and Woodward have shown that (63) monotonically reduces tr $T'\Sigma_x T$ and that it converges to a T that satisfies a necessary condition for a minimum of tr $T'\Sigma_x T$. They have not shown that the BWP will converge in a finite number of reruns if rows with $t't < 1$ occur. Nor have they shown that (63) will generate a set of values that, if inserted in the diagonal of Σ_x, will yield a Gramian matrix. Nor have they shown that such a matrix, if Gramian, must be the CMTFA solution. Thus it appears that from a theoretical point of view the MBWP is to be preferred to the BWP. We shall now show, however, that the BWP and MBWP are nearly identical as far as practical computations are concerned. First, it is clear that (63a) and (63b) jointly are equivalent to (50a) and (50b) except for the case where $0 < (R_i' TT'R_i)^{1/2} < \sigma_i^2$ and $g^{(i)} < g^{(i+1)}$. However, this exception cannot occur: It was shown above that (50b) gives the best replacement of s_i by t_i if $0 < (R_i' TT'R_i)^{1/2} < \sigma_i^2$. This best replacement cannot increase tr $T'\Sigma_x T$ because then s_i would be better than t_i. Hence the additional condition $g^{(i)} \geq g^{(i+1)}$ in (63b) is redundant. Removing it from (63b) does not affect the results. Similarly, the rerun procedure can be deleted from the BWP. Since neither (63a) nor (63b) nor (63c) generates rows with $t't < 1$ it is not possible to obtain such a row in the final solution.

The only difference between the BWP and the MBWP that remains is that between (50c) and (63c). If $\sigma_i^2 > 0$, $(R_i' TT'R_i) = 0$ and $s_i's_i > 1$ then (63c) produces a nonoptimal row t_i' whereas (50c) still gives an optimal row. Since such a case will rarely if ever occur in practice, and even if it did occur, would hardly survive subsequent iterations, it seems that the BWP and MBWP are identical procedures for all practical purposes. Nevertheless, the theoretical differences are nontrivial. If a final matrix T would contain a nonoptimal row generated by (63c), then this matrix would not satisfy (56) and (57) nor would Theorem 3 apply. Thus the global convergence properties of the BWP seem to be much less tractable than those of the MBWP. For this reason the MBWP should be preferred to the BWP.

Basic to our derivation of the MBWP is the inequality (44), which holds for all matrices T in the set defined by (41). Although Bentler and Woodward have implicitly made reference to such a set, and have derived an inequality (2.15) parallel to our (44), this inequality is not equivalent to our (44). Whereas their inequality (2.15) requires knowledge of the variables that have zero error variances in the CMTFA solution, our (44) does not. This explains why our development has produced stronger results than were obtained by Bentler and Woodward. Nevertheless, an algorithm identical to the MBWP—except for step (50c)—was first suggested in a prepublication draft of the Bentler & Woodward [1980] paper. The authors have kindly been allowed to see that draft.

It was pointed out in the previous section that convex analysis can be applied to attain necessary and sufficient conditions for the MTFA solution. The same holds for the CMTFA solution, cf. Della Riccia & Shapiro [Note 1]. Their results are in complete agreement with our inequality (44) and Theorem 3.

The MBWP is a highly attractive method for computing the g.l.b. or CMTFA solution. Heywood cases cannot occur with the method and local minima can be detected and dealt with as in Bentler's method, cf. Section 4. Practical experience indicates that local minima are not encountered if the starting matrix is taken according to the rule developed for Bentler's method in Section 4. Bentler and Woodward [1980, p. 261] also report favorable results with their method of choosing a starting matrix. Also, convergence to the global minimum appears to be rapid with both starting options.

REFERENCE NOTES

1. Della Riccia, G. & Shapiro, A. *Minimum rank and minimum trace of covariance matrices.* Beer-Sheva, Israel: Dept. of Mathematics, Ben Gurion University of the Negev, 1980.
2. Hakim, M., Lochard, E. O., Olivier, J. P. & Térouanne, E. *Sur les traces de Spearman (I)* (Cahiers du bureau universitaire du recherche opérationelle, Serie recherche no. 25). Paris: Université Pierre et Marie Curie, 1976.

REFERENCES

Bentler, P. M. A lower-bound method for the dimension-free measurement of internal consistency. *Social Science Research*, 1972, *1*, 343–357.

Bentler, P. M. & Woodward, J. A. Inequalities among lower-bounds to reliability: With applications to test construction and factor analysis. *Psychometrika*, 1980, *45*, 249–267.

Guttman, L. A basis for analyzing test-retest reliability. *Psychometrika*, 1945, *10*, 255–282.

Jackson, P. H. & Agunwamba, C. C. Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I. Algebraic lower bounds. *Psychometrika*, 1977, *42*, 567–578.

Ledermann, W. On a problem concerning matrices with variable diagonal elements. *Proceedings of the Royal Society of Edinburgh*, 1939, *60*, 1–17.

Mulaik, S. A. *The foundations of factor analysis*. New York: McGraw-Hill Inc., 1972.

Nicewander, W. A. A relationship between Harris factors and Guttman's sixth lower bound to reliability. *Psychometrika*, 1975, *40*, 197–203.

Woodhouse, B. & Jackson, P. H. Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II. A search procedure to locate the greatest lower bound. *Psychometrika*, 1977, *42*, 579–591.

Manuscript received 1/28/80
First revision received 7/7/80
Second revision received 10/30/80
Final version received 3/2/81