PSYCHOMETRIKA—VOL. 47, NO. 3.
SEPTEMBER, 1982

WEIGHTED MINIMUM TRACE FACTOR ANALYSIS

ALEXANDER SHAPIRO

DEPARTMENT OF MATHEMATICS
BEN-GURION UNIVERSITY OF THE NEGEV

In the last decade several authors discussed the so-called minimum trace factor analysis (MTFA), which provides the greatest lower bound (g.l.b.) to reliability. However, the MTFA fails to be scale free. In this paper we propose to solve the scale problem by maximization of the g.l.b. as the function of weights. Closely related to the primal problem of the g.l.b. maximization is the dual problem. We investigate the primal and dual problems utilizing convex analysis techniques. The asymptotic distribution of the maximal g.l.b. is obtained provided the population covariance matrix satisfies sone uniqueness and regularity assumptions. Finally we outline computational algorithms and consider numerical examples.

Key words: reliability, reduced rank, sample estimates.

1. Introduction

Consider the well-known factor analytic decomposition of a $p \times p$ population covariance matrix

$$\Sigma = (\Sigma - \Psi) + \Psi \tag{1.1}$$

where Ψ is diagonal and $\Sigma - \Psi$ is Gramian. In canonical factor analysis matrix Ψ is chosen to minimize the rank of $\Sigma - \Psi$. In the last decade several authors considered the problem of finding Ψ that minimize the trace of $\Sigma - \Psi$. We shall refer to this approach as the minimum-trace factor analysis (MTFA). The concept of MTFA has been discussed by Ledermann [1939], who also considered its relation to the minimum-rank factor analysis. Bentler [1972] applied the MTFA approach to the reliability theory. He defined the coefficient ρ that is a lower bound to reliability

$$\rho = 1 - \frac{1'\Psi^*1}{1'\Sigma 1} \tag{1.2}$$

where Ψ^* is the MTFA solution and 1 is the vector of unit weights. Bentler [1972] also was the first to propose the computational algorithm for the MTFA (see also Woodhouse and Jackson [1977], Bentler and Woodward [1980] and Ten Berge, Snijders and Zegers [1981]). However, the MTFA solution could lead to negative entries of Ψ (Heywood case). Jackson and Agunwamba [1977], Woodhouse and Jackson [1977] and independently Bentler and Woodward [1980] considered the MTFA with further constraint on Ψ to be Gramian. This approach will be referred to as the constrained minimum-trace factor analysis (CMTFA). The CMTFA provides the greatest lower bound (g.l.b.) to reliability, which is defined as in (1.2) except the condition for Ψ^* to be Gramian.

One of the difficulties with the MTFA is that it is not scale-independent. Bentler [1968] discussed the problem of reliability maximization as the function of weights for a fixed matrix Ψ. In this paper we propose a scale-free approach that solves the weight

I wish to express my gratitude to Dr. A. Melkman for the idea of theorem 3.3.

Requests for reprints should be sent to: A. Shapiro, Department of Statistics & O.R., University of South Africa, P. O. Box 392, PRETORIA 0001, South Africa.

0033-3123/82/0900-2939-$00.75/0

© 1982 The Psychometric Society

problem by maximization of the g.l.b. to reliability. The organization of this paper will be as follows. In Section 2 we introduce the constrained minimum-trace criterion and recall some known results. The relation between the MTFA and the MRFA is considered in Section 3. In the cases when the minimal reduced rank is one or $p-1$ we give a complete description of the problem. The primal and dual problems related to maximization of the g.l.b. are considered in Section 4. In Section 5 we discuss the sampling theory of the WMTFA and in Section 6 some approaches to numerical computations are proposed.

The following notations and definitions will be used. R^p denotes the p-dimensional vector space of column vectors. For diagonal matrices $D = \text{diag}(d_i)$, $W = \text{diag}(w_i)$ we use small letters to denote vectors $d = (d_1, \ldots, d_p)'$, $w = (w_1, \ldots, w_p)'$ etc. The term by term product of two vectors $d = (d_1, \ldots, d_p)'$ and $y = (y_1, \ldots, y_p)'$ will be denoted by d^*y, i.e., $d^*y = (d_1y_1, \ldots, d_py_p)'$. For $x = (x_1, \ldots, x_p)' \in R^p$ we denote $x^2 = x^*x = (x_1^2, \ldots, x_p^2)'$, $|x| = (|x_1|, \ldots, |x_p|)'$ and $\|x\| = (x_1^2 + \cdots + x_p^2)^{1/2}$.

For a $p \times m$ matrix A we denote $N(A')$ the null space of A', i.e., $N(A') = \{x \in R^p: A'x = 0\}$. $|B|$ denotes the determinant of a $p \times p$ matrix B. For a differentiable function f: $R^p \to R$ we denote $\nabla f(x)$ the gradient of f at x, i.e.,

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_p}\right)'.$$

conv$\{\cdot\}$ denotes the convex hull and $\overline{\text{conv}}\{\cdot\}$ the topological closure of the convex hull of a subset of R^p.

A property is said to hold almost everywhere (on the set of symmetric matrices) if it holds everywhere except for a set of (Lebesgue) measure zero.

2. *The Constrained Minimum-Trace Criterion*

Let $S = [s_{ij}]$ be a $p \times p$ Gramian matrix (covariance matrix) and $w = (w_1, \ldots, w_p)'$ be a chosen vector of weights. Consider the following problem

$$\underset{D}{\text{minimize}} \ \text{tr}\ W(S - D)W \tag{2.1}$$

subject to D diagonal and $d \in L(S)$, where $L(S)$ denotes the set of vectors $d \in R^p$ such that $S - D$ and D are Gramian. If all weights w_i, $i = 1, \ldots, p$ are different from zero we shall call it the constrained minimum-trace (CMT) problem with weight vector w. It was proven by Hakim, Lochard, Olivier and Térouanne [1976, Note 1] and independently by Della Riccia and Shapiro [1980, Note 2] and Ten Berge, Snijders and Zegers [1981] that the CMT problem has a unique solution vector d^*. It is clear that d^* is the function of S and w and we denote it by $d^* = t(w, S)$.

Now consider a scale transformation that is given by a nonsingular diagonal matrix Q. We obtain

$$W(S - D)W = W_1(S_1 - D_1)W_1 \tag{2.2}$$

where $S_1 = QSQ$ is a new covariance matrix and $W_1 = Q^{-1}W$, $D_1 = Q^2D$. It follows that the original CMT problem is equivalent to the CMT problem with the covariance matrix S_1 and the weight vector w_1, $w_1 = Q^{-1}w$. Furthermore we have that $d_1^* = Q^2d^*$, where $d_1^* = t(w_1, S_1)$ and $d^* = t(w, S)$. If the weight vector varies according to the formula $w_1 = Q^{-1}w$ together with the scale transformation Q we say that the CMT criterion is invariant under scale changes. Choosing $Q = W$ we clearly obtain unit weights in this case.

Example

Consider $w_i = \sigma_i^{-1}$, $i = 1, \ldots, p$, where σ_i^{-2} is the i-th diagonal element of the matrix S^{-1}. It can be easily verified that for this choice of weights the CMT criterion is invariant under scale changes. It is well known that σ_i is the standard error of multiple regression of the i-th variable in S on the remaining $p - 1$ variables. It was shown by Guttman [1956] that if $d \in L(S)$ then the following inequality holds

$$1 \geq \frac{1}{p} \sum_{i=1}^{p} \frac{d_i}{\sigma_i^2} = \frac{1}{p} \operatorname{tr} WDW \geq 1 - \frac{m}{p} \tag{2.3}$$

where m is the rank of $S - D$. Therefore if m is small relative to p then tr $W(S - D)W$ will be close to its minimal value.

For further considerations the following known results will be needed.

Theorem 2.1 (Della Riccia & Shapiro [1980, Note 2])

Let r^* and r be, respectively, the ranks of Gramian matrices $S - D^*$ and $S - D$. If: (a) $r^* + r < p$ and (b) all principal minors of $S - D^*$ of size r^* are different from zero, then $d_i^* \geq d_i$ for all $i = 1, \ldots, p$ and $r^* \leq r$. (A principal minor is a minor obtained by deleting a certain number of rows and columns having the same indices).

Bentler [1972] gave necessary conditions for a solution of the MTFA. Della Riccia and Shapiro [1980, Note 2] and independently Ten Berge, Snijders and Zegers [1981] proved that similar conditions are necessary and sufficient for MTFA and CMTFA.

Theorem 2.2 (necessary and sufficient conditions)

The point $d^* \in L(S)$ is the solution of the CMT problem if and only if the weight vector w can be represented in the following form

$$w^2 = \sum_{i=1}^{k} f_i^2 - \sum_{j \in I(d^*)} \alpha_j h_j \tag{2.4}$$

where vectors f_i, $i = 1, \ldots, k$, belong to the null space $N(S - D^*)$; $I(d) = \{i\colon d_i = 0, 1 \leq i \leq p\}$; h_j is the j-th coordinate vector and $\alpha_j, j \in I(d^*)$ are nonnegative numbers.

We note that for MTFA or if D^* is positive definite the sum $\sum \alpha_j h_j$ in (2.4) has to be omitted.

3. *Relation Between the MTFA and the MRFA*

This section will be concerned with the question in what case the minimum rank solution coincides with the minimum trace one. This question has been discussed in the early work of Ledermann [1939], who also pointed out that these two solutions can differ essentially even in the case of reduced rank one. Therefore it may be interesting to find conditions for the equivalence of the minimum rank and minimum trace approaches. We are willing to demonstrate that if the minimal reduced rank is small relative to p, then the MR and MT solutions usually coincide. It can be shown that the rank of the *sample* covariance matrix cannot be reduced below the bound $(2p + 1 - \sqrt{8p + 1})/2$ almost surely [Shapiro, in press]. In other words a reduced rank of the sample covariance matrix cannot be small relative to p. However, that is usually supposed for the *population* covariance matrix. Because the sample covariance matrix converges in probability to the population covariance matrix as $n \to \infty$ we can expect that the MR solution of the sample covariance matrix will be close to the MT solution if they coincide for the population covariance matrix. This effect has been observed by Bentler and Woodward [1980] when they compared the MT and MR approaches.

In the following theorem we describe the case of a reduced rank one. This result originally belongs to Ledermann [1939] and we propose a short proof of it.

Theorem 3.1

Let Σ be reducible to rank one, i.e., $\Sigma - D^* = aa'$, where $a = (a_1, \ldots, a_p)'$ is a p-dimensional column vector. Then d^* is the MT solution if and only if

$$|a_i| \leq \sum_{j \neq i} |a_j|; \qquad i = 1, \ldots, p \tag{3.1}$$

Proof: Without loss of generality we can assume that $a_i > 0$, $i = 1, \ldots, p$. Let $g = t(1, \Sigma)$ and $g \neq d^*$. Then it follows from Theorem 2.1 that the matrix $\Sigma - G$ has rank $p - 1$ and from Theorem 2.2 we have that $(\Sigma - G)v = 0$ where v is a p-dimensional sign vector, $v_i = \pm 1$.

We can write $\Sigma - G = (\Sigma - D^*) + Q$ where $Q = \text{diag}(q_i)$ is a diagonal matrix with eigenvalues $q_{i_1} \geq \cdots \geq q_{i_p}$. Because the matrix $\Sigma - D^*$ is Gramian and of rank one we have [e.g., Bellman, 1960]

$$q_{i_{p-1}} \geq \lambda_p \geq q_{i_p} \tag{3.2}$$

where $\lambda_1 \geq \cdots \geq \lambda_p$ are the eigenvalues of the matrix $Q + (\Sigma - D^*)$. From $\lambda_p = 0$ and (3.2) follows that all numbers q_i, except one, are nonnegative. From $(\Sigma - G)v = 0$ we have

$$\sigma_{ii} - g_i = -v_i \sum_{j \neq i} v_j \sigma_{ij}; \qquad i = 1, \ldots, p \tag{3.3}$$

and then

$$q_i = d_i^* - g_i = (\sigma_{ii} - a_i^2) - \left(\sigma_{ii} + v_i \sum_{j \neq i} v_j a_i a_j\right)$$

$$= -v_i a_i \sum_{j=1}^{p} v_j a_j. \tag{3.4}$$

Because all numbers q_i, except one, are nonnegative and from (3.4) we obtain that all v_i, except one, have the same sign. Without loss of generality we assume that $v_1 = -1$ and $v_2 = \cdots = v_p = 1$. Then from (3.4) follows that $a_1 - (a_2 + \cdots + a_p) > 0$ and this proves the sufficiency of (3.1).

On the other hand suppose that $a_1 - (a_2 + \cdots + a_p) > 0$. Then we can take $g_i = d_i^* - q_i$, where $q_i = a_i(a_1 - a_2 - a_3 - \cdots - a_p) > 0$ for $i \geq 2$ and $q_1 = -a_1(a_1 - a_2 - a_3 - \cdots - a_p) < 0$. It follows that λ_{p-1} is greater or equal to the smallest of the numbers q_i, $i \geq 2$, and thus $\lambda_{p-1} > 0$. It can be easily checked that $(\Sigma - G)v = 0$, where $v = (-1, 1, 1, \ldots, 1)'$ and therefore the matrix $\Sigma - G$ has zero eigenvalue. But $\lambda_{p-1} > 0$ and hence we obtain that $\lambda_p = 0$ and therefore $\Sigma - G$ is Gramian. Furthermore $\sum_{i=1}^{p} q_i = -(a_1 - a_2 - a_3 - \cdots - a_p)^2 < 0$ and hence $\text{tr}(\Sigma - G) < \text{tr}(\Sigma - D^*)$. □

Using the scale transformation $W(\Sigma - D^*)W = (Wa)(Wa)'$ we can formulate the previous theorem in the following form.

Theorem 3.1(a)

Let $\Sigma - D^* = aa'$, then $d^* = t(w, \Sigma)$ if and only if

$$|w_i a_i| \leq \sum_{j \neq i} |w_j a_j|, \qquad i = 1, \ldots, p.$$

Furthermore, if $|w_i a_i| - \sum_{j \neq i} |w_j a_j| = q > 0$ for some i, $1 \leq i \leq p$, then the MT solution $g = t(w, \Sigma)$ is given by

$$g_j = d_j^* - \frac{v_j a_j q}{w_j} \tag{3.5}$$

where $v_j = 1$ for $j \neq i$ and $v_j = -1$ for $j = i$. And $\operatorname{tr} W(\Sigma - D^*)W - \operatorname{tr} W(\Sigma - G)W = q^2$.

Now we shall need the following definition.

Definition 3.1

We say that a $p \times p$ matrix S is irreducible if it is impossible to rearrange rows and corresponding columns of S in such a way that

$$S = \left[\begin{array}{c|c} S_1 & 0 \\ \hline 0 & S_2 \end{array}\right],$$

where S_1 is a $p_1 \times p_1$ and S_2 is a $p_2 \times p_2$ matrix, $p_1 + p_2 = p$ and 0 is a zero matrix.

Guttman [1958] proposed an example wherein it is impossible to reduce the rank below $p - 1$. In the following theorem we give simple conditions that describe all such cases.

Theorem 3.2

Let S be a $p \times p$ Gramian matrix. Then the rank of a reduced (Gramian) matrix $S - D$ is greater than or equal to $p - 1$ for any choice of diagonal matrix D if and only if S is irreducible and there exist numbers $v_i = \pm 1$, $i = 1, \ldots, p$ such that

$$\operatorname{sign} s_{ij} = -v_i v_j \tag{3.6}$$

for each $s_{ij} \neq 0$; $i \neq j = 1, \ldots, p$. ($\operatorname{sign} a = a/|a|$ for a nonzero number a).

Proof: Suppose that the matrix S satisfies the conditions of the theorem. By the scale transformation VSV we obtain a matrix with nonpositive off-diagonal elements. It is known as Perron-Frobenius theorem [e.g., Gantmacher, 1959] that an irreducible matrix with nonnegative (nonpositive) elements has the maximal (the minimal) eigenvalue of multiplicity one. This proves the first part of the theorem.

On the other hand suppose that S does not satisfy the conditions. If S is not irreducible then we can reduce the rank of S to $p - 2$ simply by reducing S_1 to the rank $p_1 - 1$ and S_2 to $p_2 - 1$. Now suppose S does not satisfy (3.6). Then for $p = 3$ it is well known that the rank can be reduced to one. For $p > 3$ we can rearrange rows and corresponding columns of S in such a way that the $(p - 1) \times (p - 1)$ matrix S_0 will not satisfy (3.6), where S_0 is obtained from S by deleting the first row and first column. Consider the $(p - 1) \times (p - 1)$ matrix $S^* = S_0 - s\alpha^{-1}s'$, where $s = (s_{12}, s_{13}, \ldots, s_{1p})'$ and α is some positive number. Choosing α large enough we obtain the off-diagonal elements of the matrix $s\alpha^{-1}s'$ small enough such that they do not change the sign of the nonzero off-diagonal elements of S_0. It follows that there exists such α that the matrix S^* does not satisfy (3.6). Using induction on p we conclude that S^* can be reduced to the rank $(p - 1) - 2$ and therefore S to the rank $p - 2$. □

As a consequence of Theorem 3.2 we obtain the following result.

Theorem 3.3

If a Gramian matrix S satisfies conditions (3.6) then the solution d^* of MTFA (with weight vector w) is given by

$$d^* = Z^{-1}Sz \tag{3.7}$$

where

$$z = \mathrm{v}|w|.$$

Proof. Let S be irreducible and let it satisfy (3.6). Then from theorem 3.2 follows that $S - D^*$ has rank $p - 1$. Therefore using theorem 2.2 we obtain that d^* is the solution of the MT problem if and only if there exists a vector $z \in N(S - D^*)$ such that $w^2 = z^2$. Furthermore $(S - D^*)z = 0$ is equivalent to $Sz = Zd^*$ and then to $d^* = Z^{-1}Sz$. It remains to choose the signs of z in such a way that $S - D^*$ will be Gramian. It can be easily verified that the maximal values of diag $(S - D^*)$ we obtain for $z = v|w|$. Then because $S - D^*$ has to be Gramian for some choice of signs of z it will be Gramian for $z = v|w|$.

The case when S is not irreducible can be treated in a similar way. □

So far we discussed the cases of reduced rank one or $p - 1$. Now we are going to develop simple necessary conditions for the general case.

Definition 3.2

Denote $\lambda_1(d) \geq \cdots \geq \lambda_p(d)$ the eigenvalues of $S - D$. We say that d is a stationary point (of the MT problem with weight vector w) if:

(a) $\lambda_p(d) = 0$

(b) $\lambda_p(d - \varepsilon t_i) \leq 0$ for each $\varepsilon \geq 0$ and $i = 1, \ldots, p$; where vector t_i is given by

$$t_i = 1 - \left(w_i^{-2} \sum_{j=1}^{p} w_j^2\right) h_i, \qquad i = 1, \ldots, p$$

and h_i is the i-th coordinate vector.

We note that $t_i' w^2 = 0$, i.e., w^2 is orthogonal to vectors t_i, $i = 1, \ldots, p$. Hence if $\lambda_p(d) = 0$ and $\lambda_p(d_0) > 0$, where $d_0 = d - \varepsilon t_i$ for some ε and i, then tr $W(S - D)W =$ tr $W(S - D_0)W$ and $S - D_0$ is positive definite. In other words if d^* is the solution of the MT problem then d^* is stationary.

Lemma 3.1

The point d^* is stationary if and only if

(a) $\lambda_p(d^*) = 0$

(b) $\varepsilon w_i^{-2} \sum_{j=1}^{p} w_j^2 \geq K(\varepsilon)/K_i(\varepsilon)$ for all $\varepsilon > 0$ and $i = 1, \ldots, p$; where $K(\varepsilon) = |S - D^* + \varepsilon I_p|$, $K_i(\varepsilon) = |S_i - D_i^* + \varepsilon I_{p-1}|$ and $S_i - D_i^*$ is obtained from $S - D^*$ by deleting the i-th row and the i-th column (I_p denotes the $p \times p$ identity matrix).

Proof: Let $\lambda_p(d)$ be positive for some d then it is well known that $\lambda_p(d + \alpha_i h_i) = 0$, where α_i is the reciprocal of the corresponding i-th main diagonal element of $S - D$ or equivalently $\alpha_i = |S - D|/|S_i - D_i|$. Consider $d = d^* - \varepsilon 1$ for some $\varepsilon > 0$. We have $\lambda_p(d) = \varepsilon > 0$ and

$$d^* - \varepsilon t_i = d + \left(\varepsilon w_i^{-2} \sum_{j=1}^{p} w_j^2\right) h_i.$$

Then (b) follows by the concavity [e.g., Bellman, 1960] of $\lambda_p(d)$. We illustrate the situation in Figure 1. □

Theorem 3.4

Let $S - D^*$ be Gramian of rank $m < p$. Then d^* is stationary if and only if

$$\sum_{j=1}^{m} e_{ij}^2 \leq 1 - \frac{w_i^2}{\sum_{j=1}^{p} w_j^2}; \qquad i = 1, \ldots, p \tag{3.8}$$

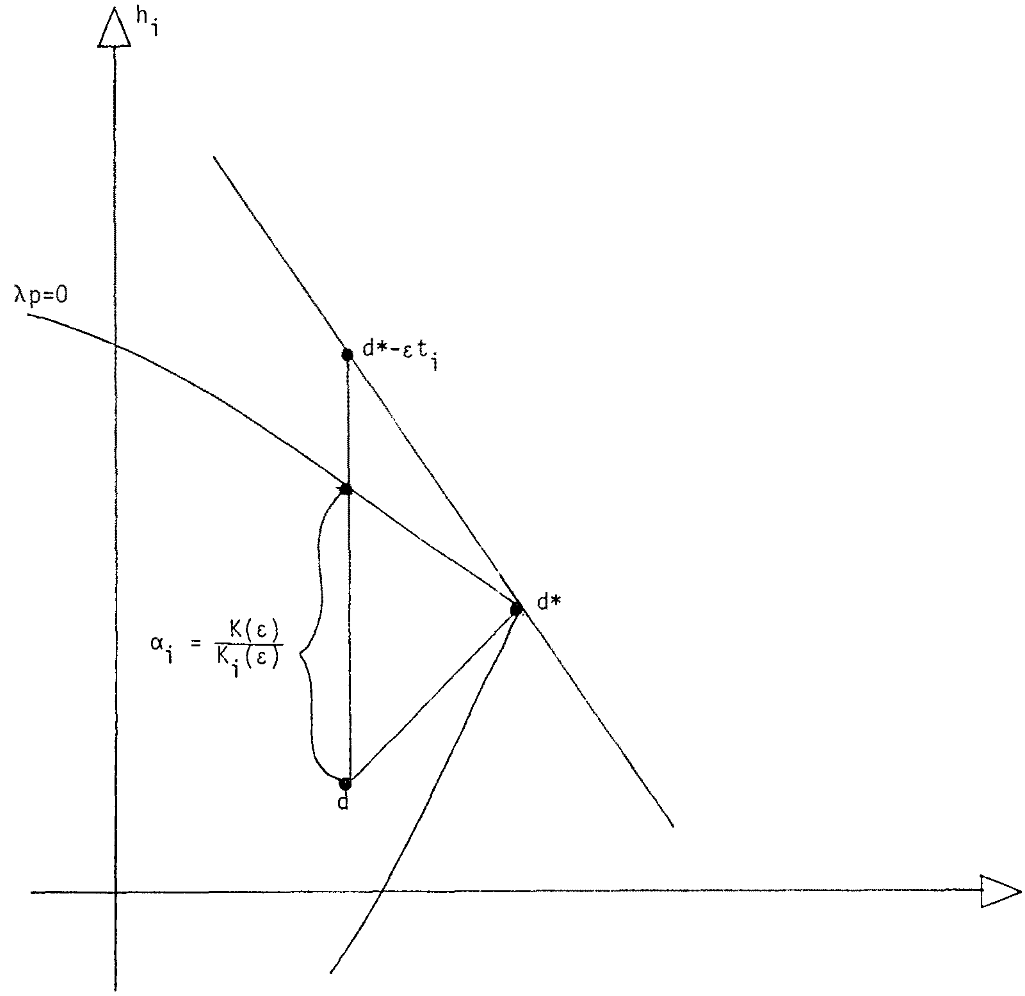

FIGURE 1

where e_{ij} is the i-th component of the orthonormal eigenvector of $S - D^*$ corresponding to the eigenvalue $\lambda_j(d^*)$.

Proof: It follows from Lemma 3.1 that d^* is stationary if and only if

$$\frac{w_i^2}{\sum_{j=1}^{p} w_j^2} \leq \limsup_{\varepsilon \downarrow 0} \frac{\varepsilon K_i(\varepsilon)}{K(\varepsilon)}. \tag{3.9}$$

Denote $\beta_1 \geq \cdots \geq \beta_m \geq \beta_{m+1} = \cdots = \beta_{p-1}$ the eigenvalues of the matrix $S_i - D_i^*$. We have

$$K(\varepsilon) = \varepsilon^{p-m} \prod_{j=1}^{m} (\lambda_j(d^*) + \varepsilon) \tag{3.10}$$

$$K_i(\varepsilon) = \varepsilon^{p-m-1} \prod_{j=1}^{m} (\beta_j + \varepsilon). \tag{3.11}$$

and then

$$\lim_{\varepsilon\to 0}\frac{\varepsilon K_i(\varepsilon)}{K(\varepsilon)}=\frac{\left(\prod_{j=1}^{m}\beta_j\right)}{\left(\prod_{j=1}^{m}\lambda_j(d^*)\right)}. \tag{3.12}$$

On the other hand $S - D^* = AA'$, where A is a $p \times m$ matrix. The matrix A can be chosen such that the $m \times m$ matrix $B = A'A$ is diagonal and $B = \operatorname{diag}(\lambda_j(d^*))$.

Then

$$|B| = \prod_{j=1}^{m}\lambda_j(d^*). \tag{3.13}$$

Furthermore

$$\prod_{j=1}^{m}\beta_j = |a_1'a_1 + \cdots + a_p'a_p - a_i'a_i| = |B - a_i'a_i| \tag{3.14}$$

where a_j is the j-th row vector of A. From (3.13) and (3.14) follows

$$\begin{aligned}\frac{\left(\prod_{j=1}^{m}\beta_j\right)}{\left(\prod_{j=1}^{m}\lambda_j(d^*)\right)} &= |B - a_i'a_i||B^{-1}|\\ &= |I_m - a_i'a_iB^{-1}| = |B^{-1/2}(I_m - a_i'a_iB^{-1})B^{1/2}|\\ &= |I_m - b_i'b_i| = 1 - \operatorname{tr} b_i'b_i\end{aligned} \tag{3.15}$$

where $b_i = a_iB^{-1/2} = [e_{i1}, e_{i2}, \ldots, e_{im}]$. From (3.9), (3.12) and (3.15) follows that d^* is stationary if and only if

$$\frac{w_i^2}{\sum_{j=1}^{p} w_j^2} \le 1 - \sum_{j=1}^{m} e_{ij}^2. \qquad \square$$

The average of the numbers $\Sigma_{j=1}^{m} e_{ij}^2$ is equal to $(1/p)\sum_{i=1}^{p}\left(\sum_{j=1}^{m} e_{ij}^2\right) = m/p$ and the average of the numbers $1 - w_i^2/\sum_{j=1}^{p} w_j^2$ is $(p-1)/p$. So we can conclude from theorem 2.2, 3.1, and 3.4 that, if the rank m of a reduced matrix $\Sigma - D^*$ (of the population covariance matrix Σ) is small relative to p, then $d^* = t(w, \Sigma)$ for a fairly large set of weight vectors w. And if for a given weight vector w the inequalities (3.8) are "well satisfied" for each $i = 1, 2, \ldots, p$, then we have a good reason to believe that the Stationary point d^* is also the solution of the corresponding MT problem.

4. *The Weighted Minimum-Trace Factor Analysis*

Throughout this section, we shall be dealing with a fixed $p \times p$ covariance matrix S which will be assumed positive definite. We consider the following functions:

$$\rho(d, w) = 1 - \frac{w'Dw}{w'Sw} \tag{4.1}$$

$$\rho_*(w) = \min_{d\in L}\rho(d, w) \tag{4.2}$$

$$\rho^*(d) = \max_{w\neq 0}\rho(d, w) \tag{4.3}$$

where $L = L(S) = \{d: S - D \text{ and } D \text{ are Gramian}\}$. The coefficient $\rho_*(w)$ can be recognized as the g.l.b. to reliability for a chosen weight vector w. It can easily be seen that an equivalent definition of $\rho(d, w)$ is as follows:

$$\rho(d, w) = \frac{w'(S - D)w}{w'Sw} \tag{4.4}$$

Therefore if w belongs to the null space $N(S - D)$ for some choice of $d \in L$ then the g.l.b. $\rho_*(w)$ will be zero! It seems that, although such a situation will rarely, if ever, occur in practice, some systematic procedure for weights selection has to be developed. In this section we discuss how to choose the weights in such a way that the g.l.b. will be as big as possible. So our primal problem will be the maximization of the function $\rho_*(w)$. We shall refer to it as the primal weighted minimum-trace (PWMT) problem and denote

$$\rho_* = \max_{w \neq 0} \rho_*(w) = \max_{w \neq 0} \min_{d \in L} \rho(d, w). \tag{4.5}$$

Closely related to the primal problem is the dual problem of minimizing the function $\rho^*(d)$, $d \in L$ (the DWMT problem).

We denote

$$\rho^* = \min_{d \in L} \rho^*(d) = \min_{d \in L} \max_{w \neq 0} \rho(d, w). \tag{4.6}$$

Definition 4.1

We say that (d_0, w_0), $d_0 \in L$, $w_0 \neq 0$, is a saddle point if

$$\rho(d_0, w) \leq \rho(d_0, w_0) \leq \rho(d, w_0) \tag{4.7}$$

for all $d \in L$ and $w \neq 0$. In the following proposition we summarize some well-known results from the minimax theory [e.g., Zangwill, 1967, p. 45].

Proposition 4.1

$\rho_*(w) \leq \rho^*(d)$ for all $d \in L$, $w \neq 0$ and therefore $\rho_* \leq \rho^*$. The equality $\rho_* = \rho^*$ holds if and only if there exists a saddle point. Furthermore, if (d_0, w_0) is a saddle point then w_0 solves the primal and d_0 solves the dual problems. On the other hand if $\rho_*(w_0) = \rho^*(d_0)$ then (d_0, w_0) is a saddle point.

We investigate the primal and dual problems utilizing convex analysis techniques. Following Pshenichyi [1971] we introduce the useful definition.

Definition 4.2

A function $f: R^p \to R$ is said to be upper quasi-differentiable (u.q.d.) at $x \in R^p$ if the directional derivative

$$f'(x; q) = \lim_{\varepsilon \downarrow 0} \frac{f(x + \varepsilon q) - f(x)}{\varepsilon}$$

exists for every direction q and there exists a convex, closed, bounded set $\partial f(x) \subseteq R^p$ such that

$$f'(x; q) = \max_{y \in \partial f(x)} y'q \tag{4.8}$$

The set $\partial f(x)$ will be called the subgradient of f at x and the function $f'(x; \cdot)$ will be called the subdifferential of f at x. We say that f is lower quasi-differentiable (l.q.d.) at x, if $-f$ is u.q.d., i.e., min replaces max in (4.8).

It is known [Rockafellar, 1970] that any convex (concave) function is u.q.d. (l.q.d).

Furthermore, a convex function f is differentiable at x_0 if and only if $\partial f(x_0) = \{\nabla f(x_0)\}$, i.e., the set $\partial f(x_0)$ contains only one vector.

Now let us consider the primal problem. We have by the definition that

$$\rho_*(w) = 1 - \frac{h(w)}{w'Sw} \tag{4.9}$$

where

$$h(w) = \max_{d \in L} w'Dw. \tag{4.10}$$

Lemma 4.1

The function $h(w)$ is convex, differentiable and

$$\nabla h(w) = 2D^*w \tag{4.11}$$

where

$$d^* = t(w, S).$$

The function $\rho_*(w)$ is continuously differentiable and

$$\nabla \rho_*(w) = \alpha Sw - \beta D^*w \tag{4.12}$$

where

$$\alpha = \frac{2w'D^*w}{(w'Sw)^2}$$

and

$$\beta = \frac{2}{w'Sw}.$$

Proof: The function $h(w)$ is the maximum of convex functions $w'Dw$ (D is Gramian) and therefore convex. We have $\nabla w'Dw = 2Dw$ and then using max-description (4.10) of $h(w)$ we obtain [e.g., Pshenichnyi, 1971].

$$\partial h(w) = \overline{\text{conv}}\{2Dw\colon d = t(w, S)\}. \tag{4.13}$$

We know that if w has no zero components then the corresponding CMT solution d^* is unique. Any way it can be shown [cf., Dalla Riccia and Shapiro, 1980, Note 2] that the components of d^* corresponding to the nonzero coordinates of w are unique. So vector D^*w is always unique. Hence $\partial h(w)$ contains only one vector $2D^*w$ and (4.11) follows. It is known [e.g., Rockafellar, 1970] that the gradient of a differentiable convex function is continuous. Therefore $\rho_*(w)$ is continuously differentiable and (4.12) follows from (4.9) and (4.11) by simple calculus. □

Theorem 4.1

The function $\rho_*(w)$ attains its maximum at some point w^* which satisfies the following equation (necessary condition)

$$D^*w^* = \gamma Sw^* \tag{4.14}$$

where $d^* = t(w^*, S)$ and $\gamma > 0$.

Proof: It can easily be seen that $h(\alpha w) = \alpha^2 h(w)$ and hence

$$\frac{h(\alpha w)}{(\alpha w)'S(\alpha w)} = \frac{h(w)}{w'Sw}. \tag{4.15}$$

Then the maximization of $\rho_*(w)$ is equivalent to the minimization of $h(w)$ subject to $w'Sw = 1$. Because the equation $w'Sw = 1$ defines the closed, bounded set and $h(w)$ is continuous we obtain that $h(w)$ attains its minimum on this set. Then (4.14) follows from the necessary condition $\nabla\rho_*(w) = 0$ and (4.12). □

We are able to say more about the dual problem. In fact we shall obtain necessary and sufficient conditions for the optimal solution of the DWMT problem. From definition (4.3) we have

$$\rho^*(d) = 1 - \min_{w \neq 0} (w'Dw/w'Sw). \tag{4.16}$$

This problem was studied by several authors [cf., Bentler, 1968]. Denote $\gamma_1(d) \geq \cdots \geq \gamma_p(d)$ solutions of the equation

$$Dw = \gamma Sw. \tag{4.17}$$

It can be seen that $\gamma_i(d)$ are the eigenvalues of the matrix $S^{-1}D$ or equivalently of the symmetric matrix $S^{-1/2}DS^{-1/2}$. The minimum in (4.16) is equal to $\gamma_p(d)$ and is attained at w satisfying (4.17) for $\gamma = \gamma_p(d)$ [e.g., Gantmacher, 1959]. If the matrix D possesses zero or negative diagonal elements then $\gamma_p(d)$ will be zero or negative and thus $\rho^*(d) > 1$. In other words in the *DWMT problem Heywood case* will *never occur*.

Lemma 4.2

The matrix $S - D$ is Gramian if and only if $\gamma_1(d) \leq 1$.

Proof: $S - D$ is Gramian if and only if $S^{-1/2}(S - D)S^{-1/2} = I_p - S^{-1/2}DS^{-1/2}$ is Gramian. The last matrix has eigenvalues $1 - \gamma_i(d)$. Then $S - D$ is Gramian if and only if $1 - \gamma_1(d) \geq 0$. □

Now we can formulate the DWMT problem as follows:

$$\max_d \gamma_p(d), \quad \text{subject to} \quad \gamma_1(d) \leq 1. \tag{4.18}$$

We denote by $\Pi(d)$ and $\pi(d)$ the vector spaces of eigenvectors of $S^{-1}D$ corresponding to the eigenvalues $\gamma_1(d)$ and $\gamma_p(d)$ respectively.

Lemma 4.3

$\gamma_1(d)$ is a convex and $\gamma_p(d)$ is a concave function

$$\partial\gamma_1(d) = \operatorname{conv}\{z\colon z = y^2,\ y \in \Pi(d),\ y'Sy = 1\} \tag{4.19}$$

$$\partial\gamma_p(d) = \operatorname{conv}\{z\colon z = y^2,\ y \in \pi(d),\ y'Sy = 1\}. \tag{4.20}$$

Proof: $\gamma_1(d)$ as the maximal eigenvalue of $S^{-1/2}DS^{-1/2}$ can be described in the following way [e.g., Bellman, 1960].

$$\gamma_1(d) = \max_x x'S^{-1/2}DS^{-1/2}x, \qquad x'x = 1. \tag{4.21}$$

It follows that $\gamma_1(d)$ is convex as the maximum of linear (by d) functions. Furthermore the gradient of these functions is given by $\nabla x'S^{-1/2}DS^{-1/2}x = y^2$, where $y = S^{-1/2}x$. Then using max-description (4.21) we obtain

$$\partial\gamma_1(d) = \text{conv}\{z: z = y^2, \gamma_1(d) = y'Dy, x'x = y'Sy = 1\}. \quad (4.22)$$

The maximum in (4.21) is attained at the eigenvector of $S^{-1/2}DS^{-1/2}$ corresponding to the maximal eigenvalue $\gamma_1(d)$. Then we have that $S^{-1/2}DS^{-1/2}x = \gamma_1(d)x$ is equivalent to $Dy = \gamma_1(d)Sy$ and this with (4.22) proves (4.19). The concavity of $\gamma_p(d)$ and (4.20) can be shown in a similar way. □

Theorem 4.2 (necessary and sufficient conditions)

A point d^* solves the DWMT problem if and only if $\gamma_1(d^*) = 1$ and there exist vectors $e_i \in \pi(d^*), i = 1, \ldots, s; f_j \in \Pi(d^*), j = 1, \ldots, r$; such that

$$\sum_{i=1}^{s} e_i^2 = \sum_{j=1}^{r} f_j^2. \quad (4.23)$$

The solution d^* is unique if vector $\sum_{i=1}^{s} e_i^2$ of equality (4.23) have no zero coordinates.

Proof. Let us consider the DWMT problem in form (4.18). The maximization of the concave function $\gamma_p(d)$ on the convex body $\{d: \gamma_1(d) \leq 1\}$ can be considered as the convex programming problem. It is clear that the maximum of $\gamma_p(d)$ is attained on the boundary $\{d: \gamma_1(d) = 1\}$.

The optimality conditions in convex programming are well known [e.g., Rockafellar, 1970]: in order that d^*, $\gamma_1(d^*) = 1$, be a solution to problem (4.18) it is necessary and sufficient that the cones spanned by the sets $\partial\gamma_1(d^*)$ and $\partial\gamma_p(d^*)$ have a non-empty intersection. Then (4.23) follows from (4.19) and (4.20).

It is known [Rockafellar, 1970] that vector e belongs to the subgradient $\partial\gamma_p(d^*)$ if and only if

$$\gamma_p(d^*) + e'(d - d^*) \geq \gamma_p(d)$$

for all $d \in R^p$. Then the uniqueness of d^* can be proven as the uniqueness of the MT solution for nonzero weights [cf., Della Riccia & Shapiro, 1980, Note 2]. □

Denote $\mu_i(d) = \gamma_i(d^2)$, $i = 1, \ldots, p$, i.e., $\mu_i(d)$ are the eigenvalues of the symmetric matrix $DS^{-1}D$. Using (4.18) and the property $\mu_i(\alpha d) = \alpha^2\mu_i(d)$, $i = 1, \ldots, p$, one can transfer the DWMT problem to the following unconstrained problem [cf., Moss, Note 3].

$$\rho^* = 1 - \max_d \mu_p(d)\mu_1^{-1}(d). \quad (4.24)$$

Using lemma 3.4 it is not difficult to show that $\mu_p(d)\mu_1^{-1}(d)$ is the l.q.d. function and

$$\partial[\mu_p(d)\mu_1^{-1}(d)] = \text{conv}\{z: z = 2\mu_p(d)\mu_1^{-1}(d)D^{-1}(f_{\min}^2 - f_{\max}^2)\} \quad (4.25)$$

where $f_{\max}$ ($f_{\min}$) is a unit-length eigenvector of $DS^{-1}D$ corresponding to the maximal (minimal) eigenvalue.

Now we return to the question of a saddle point. We have that (d_0, w_0) is a saddle point if and only if it solves the primal and dual problems. The next result follows.

Lemma 4.4

A point (d_0, w_0) is saddle if and only if

$$d_0 = t(w_0, S) \quad (4.26)$$

and

$$w_0 \in \pi(d_0). \tag{4.27}$$

Condition (4.26) is equivalent for positive definite D_0 (see 2.4) to: $\gamma_1(d_0) = 1$ and there exist vectors $f_j \in \Pi(d_0), j = 1, \ldots, r$, such that

$$w_0^2 = \sum_{j=1}^{r} f_j^2. \tag{4.28}$$

If a saddle point exists, it must be looked for between the solutions of the primal and dual problems. The following theorem gives the necessary and sufficient conditions for the existence of a saddle point.

Theorem 4.3

Let w^* be a solution of the PWMT problem and $d^* = t(w^*, S)$. Then a saddle point exists if and only if $w^* \in \pi(d^*)$. Or equivalently, let d^* be a solution of the DWMT problem. Then a saddle point exists if and only if s in (4.23) can be taken equal to one.

Proof: The theorm follows at once from Proposition 4.1, Lemma 4.4 and Theorem 4.2. □

Corollary 4.1

Let w^* be some weight vector and $d^* = t(w^*, S)$. If $w^* \in \pi(d^*)$ then:

(a) w^* solves the primal and d^* solves the dual problem (so necessary condition (4.14) is sufficient if $\gamma = \gamma_p(d^*)$)
(b) $\rho^* = \rho_*$
(c) Furthermore, if in addition dim $\pi(d^*) = 1$ (i.e., the minimal eigenvalue of $S^{-1}D$ has multiplicity one) and w^* has no zero components then the solution w^* is unique up to multiplication by a nonzero number.

Corollary 4.2

Let d^* be a solution of the DWMT problem and dim $\pi(d^*) = 1$, then a saddle point exists. In particular it follows from Perron-Frobenious theorem [e.g., Gantmacher, 1959] that if S is irreducible and there exist signs $v_i = \pm 1, i = 1, \ldots, p$, such that

$$\text{sign } s_{ij} = v_i v_j \tag{4.29}$$

for all $s_{ij} \neq 0$, $i \neq j, \ldots, p$, then the maximal eigenvalue (the minimal eigenvalue) of $D^{-1}SD^{-1}$ (of $DS^{-1}D$) has multiplicity one and therefore a saddle point exists.

Corollary 4.2 provides the convenient sufficient conditions. If a solution d^* of the dual problem is available, then the existence of a saddle point can be verified simply by examination of the multiplicity of the minimal (of the maximal) eigenvalue of the matrix $S^{-1}D^*$ (of the matrix SD^{*-1}). Even more, in many practical cases the existence of a saddle point can be stated in advance if the signs of off-diagonal elements of S satisfy equalities (4.29).

Corollary 4.3

Let $S - D = aa'$, $a \in R^p$, and D be positive definite. Then S has a saddle point. This saddle point is given by (d, w), where $w = D^{-1}a$, if and only if

$$a_i^2 d_i^{-1} \leq \sum_{j \neq i} a_j^2 d_j^{-1}, \qquad \text{for all } i = 1, \ldots, p. \tag{4.30}$$

If condition (4.30) holds, then

$$\rho^* = \rho_* = \frac{\xi}{(1+\xi)} \tag{4.31}$$

where $\xi = a'D^{-1}a$.

Proof: The existence of a saddle point follows from Corollary 3.2 (S satisfies (4.29)). Condition (4.30) follows from Theorem 3.1(a) and Lemma 4.4. Formula (4.31) can be verified by a simple calculation. □

Theorem 4.4

Let S be a block diagonal matrix, $S = \text{diag}(S_i)$, where S_i is a $p_i \times p_i$ covariance matrix, $i = 1, \ldots, k$; $p_1 + \cdots + p_k = p$. Suppose that (d_*, w_*), $(d_{*1}, w_{*1}), \ldots, (d_{*k}, w_{*k})$ solve the primal problem for matrices $S, S_1, \ldots, S_k$, respectively, and $\rho_*, \rho_{*1}, \ldots, \rho_{*k}$ are the corresponding coefficients.

(d^*, w^*), $(d_1^*, w_1^*), \ldots, (d_k^*, w_k^*)$ and $\rho^*, \rho_1^*, \ldots, \rho_k^*$ are defined analogously for the dual problem. Then

$$\rho_* = \max_{1 \le i \le k} \rho_{*i} \tag{4.32}$$

d_* (generally not unique) can be taken as

$$d'_* = (d'_{*1}, \ldots, d'_{*k}) \tag{4.33}$$

and if $\rho_{*1} > \rho_{*i}$ for $i = 2, \ldots, k$, then

$$w'_* = (w'_{*1}, 0, \ldots, 0) \tag{4.34}$$

Analogously for the dual problem.

Proof: Let $(\alpha w)' = (\alpha_1 w'_1, \ldots, \alpha_k w'_k)$ be some p-dimensional weight vector, w_i be p_i-dimensional weight vectors and $\alpha_1, \ldots, \alpha_k$ numbers. Then

$$\rho_*(\alpha w) = 1 - \frac{\left(\sum_{i=1}^{k} \alpha_i^2 a_i\right)}{\left(\sum_{i=1}^{k} \alpha_i^2 b_i\right)} \tag{4.35}$$

where $a_i = w'_i D_i w_i$, $d_i = t(w_i, S_i)$ and $b_i = w'_i S_i w_i$, $i = 1, \ldots, k$. Equality (4.35) can be written as

$$\rho_*(\alpha w) = 1 - \frac{\alpha' A \alpha}{\alpha' B \alpha} \tag{4.36}$$

where $\alpha = (\alpha_1, \ldots, \alpha_k)'$, $A = \text{diag}(a_i)$ and $B = \text{diag}(b_i)$. Then we obtain that

$$\max_{\alpha} \rho_*(\alpha w) = 1 - \min_{1 \le i \le k} a_i b_i^{-1} = 1 - a_{i_0} b_{i_0}^{-1} \tag{4.37}$$

for some i_0, $1 \le i_0 \le k$. And the maximum is attained when $\alpha_i = 0$ for $i \ne i_0$, $i = 1, \ldots, k$ and $\alpha_{i_0} \ne 0$. This proves (4.32), (4.33) and (4.34).

Consider the dual problem. Let $D = \text{diag}(D_i)$, where D_i are $p_i \times p_i$ diagonal matrices $i = 1, \ldots, k$. Then $S^{-1}D = \text{diag}(S_i^{-1}D_i)$ and it follows that

$$\gamma_p(d) = \min_{1 \le i \le k} \gamma_{p_i}(d_i)$$

where $\gamma_{p_i}(d_i)$ is the minimal eigenvalue of the matrix $S_i^{-1}D_i$. We have

$$\rho^* = 1 - \max_d \gamma_p(d) = 1 - \max_d \min_{1 \leq i \leq k} \gamma_{p_i}(d_i) = 1 - \min_{1 \leq i \leq k} \gamma_{p_i}(d_i^*)$$

subject to $S_i - D_i$ is Gramian, $i = 1, \ldots, k$. Therefore it follows

$$\rho^* = \max_{1 \leq i \leq k} \rho_i^* . \tag{4.38}$$

□

We discuss the numerical aspects of the existence of a saddle point in Section 6. Now we are going to give an example when a saddle point does not exist.

With any nonsingular covariance matrix S is associated another nonsingular covariance matrix S_1, defined by $S_1 = S^{-1}$. It is clear from (4.24) that $\rho^*(S_1) = \rho^*(S)$. Furthermore, necessary and sufficient conditions (4.23) of the DWMT problem are symmetric, i.e., D_1 is a solution of the dual problem for the matrix S_1 if and only if $D^* = \alpha D_1^{*-1}$ is a solution for the matrix S, where α is some nonzero number. Not so for the PWMT problem. In fact if $s = 1$ in equality (4.23) then S has a saddle point, but if $r > 1$, then S_1 will not. Let for example $S = I_p + aa'$, where a is a p-dimensional vector satisfying the following inequalities

$$a_i^2 \leq \sum_{j \neq i} a_j^2, \qquad 1, \ldots, p. \tag{4.39}$$

From (4.30) we have that $D = I_p$ is the solution of the DWMT problem for the matrix S. Hence the matrix $S^{-1} = I_p - (1 + a'a)^{-1}aa'$ has the solution $D_1 = (1 + a'a)^{-1}I_p$. Using lemma 4.4 we obtain that D_1 provides a saddle point for S_1 if and only if there exist signs $v_i = \pm 1$ such that

$$\sum_{i=1}^{p} v_i a_i^2 = 0. \tag{4.40}$$

Therefore if condition (4.39) holds and (4.40) does not, then the matrix S^{-1} has no saddle point.

5. *Sampling Considerations*

In this section we discuss the sampling theory of the WMTFA. Let Σ be the $p \times p$ population covariance matrix and S be a sample covariance matrix based on N $(>p)$ observations. We consider the g.l.b. $\rho_*(w)$ and the maximal g.l.b. ρ_* as the functions of S, denoted $\rho_*(w, S)$ and $\rho_*(S)$ respectively. We are concerned with the sample estimates $\hat{\rho}_*(w) = \rho_*(w, S)$ and $\hat{\rho}_* = \rho_*(S)$ of the parameters $\rho_*(w) = \rho_*(w, \Sigma)$ and $\rho_* = \rho_*(\Sigma)$. We discuss the differential properties of the functions $\rho_*(w, S)$ and $\rho_*(S)$ and show that one can serve as an approximation to the other provided Σ satisfies some uniqueness conditions. If in addition some regularity conditions hold, we obtain the asymptotic sampling distribution of $\hat{\rho}_*$.

The result of the following lemma was suggested (for unit weights) by Bentler [1972], Bentler and Woodward [1980] and proven by Ten Berge et al [1981].

Lemma 5.1

$$\rho_*(w, S) = 1 - \min_F \frac{\operatorname{tr} F'SF}{w'Sw} \tag{5.1}$$

subject to $\operatorname{diag}(FF') \geq W^2$, where F is a $p \times k$ matrix. The minimum is attained at $F_0 = [f_1, \ldots, f_k]$, where column vectors f_i, $i = 1, \ldots, k$ satisfy the necessary and sufficient conditions of Theorem 2.2.

Proof: Let $S - D$ and D be Gramian and $\operatorname{diag} FF' \geq W^2$. Then we have

$$\operatorname{tr} F'SF = \operatorname{tr} F'(S - D)F + \operatorname{tr} \underline{D}FF' \geq \operatorname{tr} \underline{D}FF' \geq \operatorname{tr} DW^2 = w'Dw \tag{5.2}$$

and the equality holds only if $(S - D)F = 0$ and $\operatorname{tr} DFF' = w'Dw$. The existence of such a matrix F is ensured by Theorem 2.2. □

We denote $\Omega(S)$ the set of weight vectors w that solve the PWMT problem for the matrix S and $\Phi(w, S)$ denotes the set of matrices $Y = FF'$ such that the minimum in (5.1) is attained at F. The next lemma follows at once from lemma 5.1 and (4.5).

Lemma 5.2

$$\rho_*(S) = \max_{Y,\, w} \left(1 - \frac{\operatorname{tr} SY}{w'Sw}\right) \tag{5.3}$$

subject to $\operatorname{diag}(Y) \geq W^2$, Y Gramian and $w'w = 1$. The maximum is attained at $w_0 \in \Omega(S)$ and $Y \in \Phi(w_0, S)$.

Theorem 5.1

$\rho_*(S)$ and $\rho_*(w, S)$ (for fixed w) are continuous and u.q.d. functions of S and

$$\partial\rho_*(w, S) = \operatorname{conv}\{Z\colon Z = \alpha ww' - \beta Y,\ Y \in \Phi(w, S)\} \tag{5.4}$$

$$\partial\rho_*(S) = \operatorname{conv}\{Z\colon Z = \alpha ww' - \beta Y,\ w \in \Omega(S),\ Y \in \Phi(w, S)\} \tag{5.5}$$

where $\alpha = (\operatorname{tr} SY)(w'Sw)^{-2}$ and $\beta = (w'Sw)^{-1}$.

Proof: It can be seen that the function

$$f(S) = 1 - \frac{\operatorname{tr} SY}{w'Sw}$$

is twice continuously differentiable and

$$\nabla f(S) = \alpha ww' - \beta Y \tag{5.6}$$

Then the theorem follows from max-descriptions (5.1) and (5.3) (see e.g., Dem'yanov and Malozemov [1974], Pshenichyi [1971]). □

Corollary 5.1

Let $w_0 \in \Omega(S)$, then $\partial\rho_*(w_0, S) = \partial\rho_*(S)$ if and only if $\Omega(S) = \{\alpha w_0\}$. In other words the subdifferentials of $\rho_*(S)$ and $\rho_*(w_0, S)$ coincide at S if and only if w_0 is the unique solution (up to multiplication by a nonzero number) of the PWMT problem for the matrix S. (For sufficient conditions of such uniqueness see corollary 4.1).

Corollary 5.2

The function $\rho_*(S)$ is differentiable at S if and only if $\Omega(S) = \{\alpha w_0\}$ and $\Phi(w_0, S) = \{Z_0\}$, i.e., the PWMT problem has the unique solution w_0 and the matrix $Z_0 = F_0 F_0'$ is unique, where F_0 is a matrix at which the minimum in (5.1) is attained.

In the asymptotic theory one substitutes the real increment of a function by its subdifferential approximation, i.e.,

$$\rho_*(S) - \rho_*(\Sigma) \cong \rho_*'(\Sigma, \Delta\Sigma) \tag{5.7}$$

where $\Delta\Sigma = S - \Sigma$ and $\rho_*'(\Sigma, \cdot)$ is the subdifferential of ρ_* at Σ. Then the next theorem follows from Corollary 5.1.

Theorem 5.2

The sample statistics $\hat{\rho}_*(w_0)$ and $\hat{\rho}_*$ asymptotically coincide if and only if w_0 is the unique (up to multiplication by a nonzero number) solution of the PWMT problem for the matrix Σ.

A sample behaviour of $\hat{\rho}_*(w)$ has been studied by Shapiro [in press]. The following definition will be useful in this respect.

Definition 5.1

Let S be a $p \times p$ Gramian matrix, $d = t(w, S)$ and $e_1, \ldots, e_m$ be a vector basis of the null space $N(S - D)$. We say that S is MTFA regular for the weight vector w if the set of $m(m+1)/2$ vectors $e_i^* e_j$, $i \leq j = 1, \ldots, m$, is linearly independent.

Shapiro [in press] showed that if Σ is MTFA regular, then $\rho_*(w, S)$ is differentiable at Σ, and derived the asymptotic distribution of $\hat{\rho}_*(w)$ in this case. We recall this result taking advantage of Theorem 5.2. It will be supposed that the sample is drawn from a normally distributed population with positive definite covariance matrix Σ.

Theorem 5.3

Let w_0 be the unique solution of the PWMT problem (for the matrix Σ) and let Σ be MTFA regular for the weight vector w_0. Then $n^{1/2}(\hat{\rho}_* - \rho_*)$ has an asymptotically normal distribution with zero mean and variance

$$\frac{2}{\gamma}\left[\sum_{i,j=1}^{k} (f_i'\Psi f_j)^2 - \frac{2\delta}{\gamma}\sum_{i=1}^{k}(f_i'\Psi w_0)^2 + \delta^2\right] \tag{5.8}$$

where $n = N - 1$, $\Psi = t(w_0, \Sigma)$, vectors $f_1, \ldots, f_k$ satisfy the necessary and sufficient conditions of theorem 2.2 (for the matrix Σ and the weight vector w_0), $\gamma = w_0' \Sigma w_0$ and $\delta = w_0' \Psi w_0$.

An interesting consequence follows from Corollary 5.2. It can be shown that $\rho_*(S)$ is a locally Lipschitz function. It is known [e.g., Stein, 1970] that a locally Lipschitz function is differentiable almost everywhere. It follows that almost everywhere the PWMT solution is unique (up to multiplication by a nonzero number).

6. *Computational Algorithms*

In this section we outline some computational approaches to a numerical solution of the PWMT and DWMT problems for a chosen matrix S. These problems are equivalent provided the existence of a saddle point. We discuss the saddle point problem at the end of the section.

Now consider the PWMT problem. From Lemma 5.2 follows

$$\rho_*(S) = 1 - \min_{F, w} \operatorname{tr} \frac{F'WSWF}{w'Sw} \tag{6.1}$$

subject to $\operatorname{diag}(FF') \geq I_p$ and $w'w = 1$.

We propose to solve Problem (6.1) by a Gauss-Seidel type algorithm, that iteratively replaces F by a best solution, keeping w fixed, then replaces w by a best solution for fixed F and so on. To implement such an algorithm we have to know how to solve the following two problems.

$$\min_{F} \operatorname{tr} F'WSWF, \qquad \operatorname{diag}(FF') \geq I_p \tag{6.2}$$

and

$$\min_{w} \frac{\mathrm{tr}F'WSWF}{w'Sw}, \qquad w'w = 1 \tag{6.3}$$

Problem (6.2) is equivalent to the CMT problem (for the matrix $S_1 = WSW$ and unit weights) and can be solved by Bentler's method [Bentler (1972), Bentler & Woodward (1980), Ten Berge et al. (1981)].

Let us consider Problem (6.3). We have

$$\mathrm{tr}\ F'WSWF = \sum_{i=1}^{k} f'_i WSWf_i = w'Qw \tag{6.4}$$

where $f_1, \ldots, f_k$ are the column vectors of F and $Q = \sum_{i=1}^{k} F_i SF_i$. Then a solution of (6.3) is given by an eigenvector of $S^{-1}Q$ corresponding to the minimal eigenvalue.

The algorithm will converge to a stationary point F^*, w^* that solves problems (6.2) and (6.3), i.e.,

$$(S - D^*)W^*F^* = 0 \qquad \text{and} \qquad \mathrm{diag}(F^*F^{*\prime}) \geq I_p \tag{6.5}$$

$$Qw^* = \gamma Sw^* \tag{6.6}$$

where $d^* = t(w^*, S)$, $Q = \sum_{i=1}^{k} F_i^* SF_i^*$, $F^* = [f_1, \ldots, f_k]$ and γ is the minimal eigenvalue of $S^{-1}Q$. From (6.5) we have $(S - D^*)F_i^* w^* = 0$ and thus $F_i^* SF_i^* w = F_i D^* F_i w^*$. Then

$$Qw^* = \left(\sum_{i=1}^{m} D^*F_i^{*2}\right)w^* = D^*w^* \tag{6.7}$$

and using (6.6)

$$D^*w^* = \gamma Sw^*. \tag{6.8}$$

Hence the necessary conditions of theorem 4.1 hold. Note that γ is not necessarily the minimal eigenvalue $\gamma_p(d^*)$ of the matrix $S^{-1}D^*$. However, if $\gamma = \gamma_p(d^*)$ then (w^*, d^*) is a saddle point, w^* solves the PWMT and d^* solves the DWMT problems (see Lemma 4.4 and Corollary 4.1). But if $\gamma \neq \gamma_p(d^*)$ we cannot be sure that the algorithm has not been trapped in a local minimum.

The DWMT problem is more promising in this sense. Because of the convex nature of the dual problem we were able to obtain necessary and sufficient conditions (4.23). We consider the dual problem in unconstrained form (4.24). Note that $\lambda_i^{-1}(D) = \mu_{p-i}(D^{-1})$, where $\lambda_1(d) \geq \cdots \geq \lambda_p(d)$ are the eigenvalues of the matrix DSD, and hence (4.24) is equivalent to the maximization of the function $g(d) = \lambda_p(d)\lambda_1^{-1}(d)$. The main difficulty is that g is not necessarily everywhere differentiable or more precisely g is differentiable at d if and only if $\lambda_1(d) \neq \lambda_2(d)$ and $\lambda_{p-1}(d) \neq \lambda_p(d)$. It is a known phenomenon that gradient-direction methods for such functions can lead to nonconvergence and jamming [Zangwill (1967), Den'yomov & Malozemov (1974)]. When jamming occurs, the algorithm generates a sequence of points $\{d_k\}$ which converges to some nondifferentiable point d that is not optimal. To overcome the jamming phenomenon we considered $g(d)$ in a neighbourhood of a nondifferentiable point as

$$g(d) = \min_{i,j} g_{ij}(d) \tag{6.9}$$

where

$$g_{ij}(d) = \lambda_i(d)\lambda_j^{-1}(d).$$

Then we utilized Madsen and Schjoer-Jacobson's [1978] approach to the minimax optimization of the function $g(d)$. "The basic idea (of the algorithm) is to generate, by means of first derivatives, a sequence of linear approximations to the nonlinear problem and to solve the linear systems in the minimax sense ...". The linear approximations are given by:

$$g_{ij}(d_k + h) \cong g_{ij}(d_k) + \nabla g_{ij}(d_k)'h \tag{6.10}$$

where the gradient of g_{ij} is

$$\nabla g_{ij}(d) = 2\lambda_i(d)\lambda_j^{-1}(d)(g_i^2 - g_j^2) \tag{6.11}$$

and $g_1, \ldots, g_p$ are the orthonormal eigenvectors of the matrix DSD corresponding to the eigenvalues $\lambda_1(d), \ldots, \lambda_p(d)$.

We report and discuss the numerical results of the above approach in the next section. Unfortunately the jamming phenomenon is not always overcome by this algorithm. We expect that more stronger results can be obtained by an application of an ε-subgradient type approach recently developed in convex and nondifferentiable programming [Bertsekas & Mitter (1971), Dem'yanov & Malozemov (1974), Goldstein (1977)]. The discussion and the proof of convergence of such an algorithm, applied to a problem very similar to ours, can be found in Cullumn, Donath & Wolfe [1975]. We point out that this is in some sense an ε-steepest descent (ε-steepest ascent) method. The direction of ε-steepest ascent at d is given by $y = x/\|x\|$, where x is the element in $\partial g(d, \varepsilon)$ closest to the origin. And (cf., (4.25))

$$\partial g(d, \varepsilon) = \operatorname{conv}\{z\colon z = 2\lambda_p(d)\lambda_1^{-1}(d)D^{-1}(g_{\min}^2 - g_{\max}^2)\} \tag{6.12}$$

where $g_{\max}$ ($g_{\min}$) is a unit-length vector in the vector space spanned by eigenvectors $g_1, \ldots, g_k$ ($g_p, \ldots, g_{p-m}$) such that

$$|\lambda_1(d) - \lambda_k(d)| \le \varepsilon, \qquad (|\lambda_p(d) - \lambda_{p-m}(d)| \le \varepsilon).$$

It is clear that $\partial g(d, 0) = \partial g(d)$ and it can be shown (see necessary and sufficient conditions (4.23)) that d^* solves the DWMT problem if and only if $x = 0$, i.e., $0 \in \partial g(d^*)$. The element x can be determined computationally if a contact point of $\partial g(d, \varepsilon)$ can be computed [Gilbert (1966)]. A point $u_0 \in A$ is a contact point of the compact set A corresponding to the direction y, if

$$y'u_0 = \max_{u \in A} y'u \tag{6.13}$$

We denote the set of contact points $C(A \mid y)$ and remark that

$$C(\operatorname{conv}\{A\} \mid y) = \operatorname{conv}\{C(A \mid y)\} \quad \text{and} \quad C(A_1 + A_2 \mid y) = C(A_1 \mid y) + C(A_2 \mid y).$$

The following proposition demonstrates that contact points can be easily computed in our case [cf., Cullumn et al, (1975), Lemma 7.1].

Proposition 6.1

Consider vector space $V \subseteq R^p$ spanned by orthonormal vectors $f_1, \ldots, f_m$ and the set A, $A = \{f^2 : f \in V, f'f = 1\}$. Then

$$C(A \mid y) = \{f^2 : f = \alpha_1^* f_1 + \cdots + \alpha_m^* f_m\} \tag{6.14}$$

where $\alpha^* = (\alpha_1^*, \ldots, \alpha_m^*)'$ is a unit-length eigenvector of the $m \times m$ matrix $F'YF$ corresponding to the maximal eigenvalue and $F = [f_1, \ldots, f_m]$.

Proof: From $f \in V$ and $f'f = 1$ follows that $f = \alpha_1 f_1 + \cdots + \alpha_m f_m$ and $\alpha_1^2 + \cdots + \alpha_m^2 = 1$. Then

$$f^2 = \sum_{i,j=1}^{m} \alpha_i \alpha_j f_i^* f_j$$

and

$$\max_{u \in A} y'u = \max_{\alpha'\alpha=1} y'\left(\sum_{i,j=1}^{m} \alpha_i \alpha_j f_i^* f_j \right)$$

$$= \max_{\alpha'\alpha=1} \sum_{i,j=1}^{m} \alpha_i \alpha_j f_i' Y f_j = \max_{\alpha'\alpha=1} \alpha' F' Y F \alpha. \tag{6.15}$$

It is well known that the last maximum in (6.15) is attained at an eigenvector of $F'YF$ corresponding to the maximal eigenvalue [Bellman, 1960]. □

Finally, we consider the saddle point problem. Suppose that we computed a point d^* that solves the DWMT problem. Since Perron-Frobenius theorem it is not surprising that the most frequent case will be when $\dim \pi(d^*) = 1$ and a saddle point exists (see Corollary 4.2). However, if $\dim \pi(d^*) > 1$ we usually do not expect the existence of a saddle point. Let $h_1, \ldots, h_m$ $(q_1, \ldots, q_k)$ be a vector basis of the space $\pi(d^*)$ (of the space $\Pi(d^*)$). It can be shown [Shapiro, in press] that necessary and sufficient conditions (4.23) of the DWMT problem are equivalent to

$$\sum_{i,j=1}^{m} \alpha_{ij} h_i^* h_j = \sum_{i,j=1}^{k} \beta_{ij} q_i^* q_j \tag{6.16}$$

where matrices $A = [\alpha_{ij}]$ and $B = [\beta_{ij}]$ are Gramian. And (see Theorem 4.3) a saddle point exists if and only if matrix A can be taken of rank one. If the intersection of the vector spaces spanned by vectors $h_i^* h_j$, $1 \le i, j \le m$, and vectors $q_i^* q_j$, $1 \le i, j \le k$ has dimension one and vectors $h_i^* h_j$, $1 \le i \le j \le m$ are linearly independent then the matrix A is unique up to multiplication by a nonzero number. In this case the existence of a saddle point can be easily verified. Shapiro [in press] showed that vectors $h_i^* h_j$, $1 \le i \le j \le m$, are linearly independent almost surely. So it seems that in most practical cases the existence of a saddle point can be verified in a simple way.

7. *Numerical Examples*

We implemented the DWMT algorithm (Madsen & Schjoer-Jacobson approach) to examples that were reported by Bentler [1972] (Social Class Data and WAIS Intelligence Test example). In each example the procedure converged to a solution point that was also a saddle point. This is not surprising since in both covariance matrices the off-diagonal elements are positive (see Corollary 4.2). In the Social Class Data example we obtained uniqueness estimates $d^* = (0.033, 0.023, 0.103, 0.061, 0.073, 0.038)$, weights 1.360, 1.960, 0.399, 0.684, 0.565, 1.032 and $\rho_* = \rho^* = 0.991$ (to compare with $\rho = 0.969$ for the unit weights).

In WAIS Intelligence Test example the procedure converged to a solution-saddle point with uniqueness estimates 1.303, 2.775, 3.113, 2.323, 3.011, 1.380, 3.172, 2.624, 3.036, 2.473, 2.728 weights 2.041, 0.856, 0.722, 1.052, 0.617, 1.976, 0.675, 0.852, 0.738, 0.825, 0.646 and $\rho_* = \rho^* = 0.961$ ($\rho = 0.949$ for the unit weights).

In the examples above ρ_* is greater than ρ by a small amount because ρ is close to unit.

Now we consider the artificial example with covariance matrix

$$\begin{array}{llll} 1.00 & & & \\ -.05 & 1.00 & & \\ -.35 & -.20 & 1.00 & \\ -.40 & -.05 & -.05 & 1.00 \end{array}$$

that has negative off-diagonal elements. It follows from theorem 3.3 that the g.l.b. of this matrix is zero (unit weights). On the other hand in this example the procedure converged to a solution-saddle point d^* (dim $\pi(d^*) = 1$ and $f^2_{\min} = f^2_{\max}$) with uniqueness estimates $d^* = (0.409, 0.362, 0.397, 0.396)$, weights 1.377, 0.522, -1.058, -1.043 and $\rho_* = \rho^* = 0.736$.

In all examples above the uniqueness assumptions of Theorem 4.2 and Corollary 4.1(c) were satisfied.

REFERENCE NOTES

1. Hakim, M., Lochard, E. O., Olivier, J. P. & Térouanne, E. *Sur les traces de Spearman.* Cahiers du bureau universitaire du recherche opérationelle, Université Pierre et Marie Curie, Paris, 1976.
2. Della Riccia, G. & Shapiro, A. *Minimum rank and minimum trace of covariance matrices.* Department of Mathematics, Ben-Gurion University of the Negev, Beer Sheva, Israel, 1980.
3. Moss, A. G. *Theoretical aspects of weight-vector maximization of composite test reliability.* Institute of Educational Technology, The Open University, Walton Hall, England, 1977.

REFERENCES

Bentler, P. M. Alpha-maximized factor analysis (alphamax): its relation to alpha and canonical factor analysis. *Psychometrika*, 1968, *33*, 335–345.

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.

Bertsekas, D. P. & Mitler, S. K. Steepest descent for optimization problems with nondifferentiable cost functions. *Proceedings of the 5th Annual Princeton Conference on Information Sciences and Systems*, 1971.

Bellman, R. *Introduction to matrix analysis.* McGraw-Hill, 1960.

Cullum, J., Donath, W. E. & Wolfe, P. The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices. *Mathematical Programming Study 3*, North-Holland Publishing Company, 1975.

Dem'yanow, V. F. & Malozemov, V. N. *Introduction to minimax.* Wiley, New York, 1974.

Guttman, L. Best possible systematic estimates of communalities. *Psychometrika*, 1956, *21*, 273–285.

Guttman, L. To what extent can communalities reduce rank? *Psychometrika*, 1958, *23*, 297–307.

Gantmacher, F. R. *The theory of matrices.* New York, 1959.

Gilbert, E. An iterative procedure for computing the minimum of a quadratic form on a convex set. *SIAM Journal Control*, 1966, *4*, 61–80.

Goldstein, A. A. Optimization of Lipshitz continuous functions. *Mathematical Programming*, 1977, *13*, 14–22.

Harman, H. *Modern factor analysis.* Chicago, 1976.

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 Edinburgh*, 1939, *60*, 1–17.

Madsen, K. & Schijoer-Jacobson. Linearly constrained minimax optimization. *Mathematical Programming*, 1978, *17*, 208–223.

Pshenichnyi, B. N. *Necessary conditions for an extremum.* Marcel Dekker, Inc., New York, 1971.

Rockafellar, R. T. *Convex analysis.* Princeton, 1970.

Stein, F. M. *Singular integrals and differentiability properties of functions.* Princeton Mathematics Serial no. 30, 1970.

Shapiro, A. Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *In press.*

Ten Berge, J. M. F., Snijders, T. A. B., & Zegers, F. E. Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis. *Psychometrika*, 1981, *46*, 201–213.

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

Zangwill, W. J. *Nonlinear programming, a unified approach*. Prentice-Hall, 1967.

Manuscript received 2/22/80
First revision received 8/10/81
Final version received 2/24/82