PSYCHOMETRIKA—VOL. 56, NO. 2, 309–315
JUNE 1991

A NUMERICAL APPROACH TO THE APPROXIMATE AND THE EXACT MINIMUM RANK OF A COVARIANCE MATRIX

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

UNIVERSITY OF GRONINGEN

A concept of approximate minimum rank for a covariance matrix is defined, which contains the (exact) minimum rank as a special case. A computational procedure to evaluate the approximate minimum rank is offered. The procedure yields those proper communalities for which the unexplained common variance, ignored in low-rank factor analysis, is minimized. The procedure also permits a numerical determination of the exact minimum rank of a covariance matrix, within limits of computational accuracy. A set of 180 covariance matrices with known or bounded minimum rank was analyzed. The procedure was successful throughout in recovering the desired rank.

Key words: communality estimation, minimum rank, factor analysis.

The communality problem in factor analysis can be stated as follows: For a given $n \times n$ covariance matrix Σ, find a diagonal matrix U^2 of unique variances such that $(\Sigma - U^2)$ has the smallest possible rank, subject to the constraint

$$\Sigma \geq U^2 \geq 0, \tag{1}$$

which means that both $(\Sigma - U^2)$ and U^2 have to be Gramian (nonnegative definite). The diagonal elements of $(\Sigma - U^2)$ are communalities, called proper if they satisfy (1), and improper otherwise.

There is a long history of attempts to solve the communality problem by determining the minimum rank of Σ, that is, the smallest possible rank of $(\Sigma - U^2)$ such that U^2 satisfies (1). Knowing the minimum rank of Σ does not necessarily imply that an associated matrix U^2 has been found, but it is a first step. In addition, the minimum rank is of theoretical interest in its own right.

The problem of determining the minimum rank by mathematical arguments has been solved for $n \times n$ matrices with $n \leq 5$, albeit the computations required can be rather tedious (P. A. Bekker, personal communication, July 6, 1988). For $n > 5$, there are necessary and sufficient conditions for having a minimum rank 1 or $(n - 1)$, and there are some conditions that are either necessary or sufficient for intermediate values (Bekker & de Leeuw, 1987) but the problem of determining the minimum rank in general has not been solved.

A more fundamental problem with the minimum rank approach to communalities is that the minimum rank tends to be closer to n than to 1. Although cases of high minimum rank are of interest in certain fields like structural regression analysis (see Bekker & de Leeuw, 1987, sec. 5), they are disturbing for practitioners in psychology, who like to have a low minimum rank to obtain an efficient condensation of common variance in a low-dimensional space. For this purpose they are, in fact, willing to ignore small eigenvalues of $(\Sigma - U^2)$, and therefore, the minimum rank is of little practical value.

The authors are obliged to Paul Bekker for stimulating and helpful comments.

Requests for reprints should be sent to Jos M. F. ten Berge, Vakgroep Psychologie, Rijksuniversiteit Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, THE NETHERLANDS.

0033-3123/91/0600-90198$00.75/0

Practitioners interested in low-rank approximations typically set out to find a matrix of factor loadings of order $n \times r$ for some predetermined value of r, that optimizes some criterion-of-fit, for instance, by Lisrel or Minres, and obtain the communalities as a by-product. This approach does yield low-rank approximations, but at the cost of violating (1). Although nonnegativity of U^2 and of the communalities can be enforced in most factor analysis procedures, the resulting $(\Sigma - U^2)$ typically has a number of small negative eigenvalues, reflecting a violation of (1). In practice, this violation often goes unnoticed because factors with small eigenvalues, including negative ones, are usually ignored. Occasionally, however, this yields embarrassing results nonetheless, in that the factors maintained explain more than 100% variance. Ignoring factors associated with small eigenvalues would make more sense if these eigenvalues could be constrained to be nonnegative. It is the purpose of the present paper to give a method of communality estimation that permits low-rank approximations without violating the constraint (1). This method is based on a definition of the approximate minimum rank of Σ, and an algorithm that puts this definition to practice. In addition, the (exact) minimum rank of Σ will be treated as a special case of approximate minimum rank, with perfect approximation. It follows that any algorithm for approximate minimum rank is also an algorithm for exact minimum rank itself, which can thus be approached numerically.

The Approximate Minimum Rank of a Covariance Matrix

It has been argued above that it is desirable to have proper communalities on the one hand, and many zero eigenvalues of $(\Sigma - U^2)$ on the other. Such communalities generally do not exist, but an approximation of this ideal is feasible. That is, we may find proper communalities that minimize the sum of those eigenvalues ignored in low-rank approximations of the common variance. Specifically, it is possible to minimize, for arbitrary but fixed r, the function

$$f(U^2) = \sum_{i=r+1}^{n} \lambda_i(\Sigma - U^2), \tag{2}$$

subject to (1) where $\lambda_i(\Sigma - U^2)$ is the i-th largest eigenvalue of $(\Sigma - U^2)$. The function f has a well-known interpretation in terms of least-squares theory. From Eckart and Young (1936) it follows that f measures the squared distance between the matrix of common parts of the variables to the best-fitting rank-r matrix, and minimizing f amounts to finding a matrix of common parts that is as close as possible to a rank-r matrix. The special case $r = 0$ has been examined before. That is, minimizing the sum of all eigenvalues of $(\Sigma - U^2)$ is equivalent to minimizing the sum of the communalities, which is known as (constrained) minimum trace factor analysis (Bentler & Woodward, 1980; Della Riccia & Shapiro, 1982; Ten Berge, Snijders, & Zegers, 1981).

A computational method for minimizing (2) subject to (1) has been outlined by ten Berge and Kiers (1988), and will be specified below. The existence of such a method motivates the following definition of approximate minimum rank.

Definition 1. The approximate minimum rank (amr) of a covariance matrix Σ is the smallest value of r ($r = 1, 2, \ldots, n - 1$) such that the minimum of $f(U^2)$, subject to $\Sigma \geq U^2 \geq 0$, is below or equal to a tolerance parameter $\delta \geq 0$.

The tolerance parameter δ can be interpreted as the maximum amount of unexplained common variance that one is willing to tolerate in low-rank approximations. To eval-

uate amr(Σ), one may first try an arbitrary value of r, and then try a higher value of r if the obtained minimum of $f(U^2) > \delta$, or a lower value if $f(U^2) \leq \delta$.

Once amr(Σ) has been determined, one also has a solution for U^2, which, unfortunately, need not be unique. The communalities implied by U^2 are proper, which is attractive from a mathematical point of view, and those eigenvalues ignored have the smallest possible sum, which is attractive from a practical point of view. It is furthermore clear from Definition 1 that amr(Σ) coincides with the exact minimum rank mr(Σ) if $\delta = 0$. Implications of this observation will be examined below. First, however, a computational solution for the problem of minimizing (2) subject to (1) will be reviewed.

A Computational Method for Minimizing $f(U^2)$

It is well-known that for any symmetric matrix S, the minimum of tr $X'SX$, subject to $X'X = I_p$, is obtained when X is the $n \times p$ matrix containing the last p eigenvectors of S, and that this minimum is the sum of the p smallest eigenvalues of S; see, for instance, Takeuchi, Yanai, and Mukherjee (1982, p. 26). It follows that the minimum of $f(U^2)$ coincides with the minimum of the function

$$g(X,U^2) = \text{tr } X'(\Sigma - U^2)X, \tag{3}$$

where X is an $n \times (n - r)$ columnwise orthonormal matrix. It is possible to minimize g for fixed U^2 as a function of X only, by computing a matrix of eigenvectors, associated with the $(n - r)$ smallest eigenvalues of $(\Sigma - U^2)$. It is also possible to minimize g for fixed X as a function of U^2 only, subject to (1), by maximizing tr $X'U^2X$. The latter problem is a weighted minimum trace factor analysis problem (Shapiro, 1982b), which can be solved by various iterative algorithms. By alternating between optimizing X and U^2, the function $g(X,U^2)$ is reduced monotonically. An application of this method to factor analysis has been given by ten Berge and Kiers (1988). The remaining part of the present paper will be focused on evaluating mr(Σ) by means of this method. The specific algorithm used for this purpose will be explained in a later section.

Minimum Rank as a Special Case of AMR

The minimum rank mr(Σ) of a covariance matrix Σ is the special case of amr(Σ), where δ is required to be zero:

Definition 2. The minimum rank mr(Σ) is the smallest value of r for which $f(U^2)$ has a minimum of zero, subject to $\Sigma \geq U^2 \geq 0$.

Clearly, Definition 2 is equivalent to the definition of minimum rank that has been around for half a century, and the latter definition has merely been rephrased as Definition 2 for the purpose of interpreting it as the special case of amr(Σ) where $\delta = 0$.

The practical utility of mr(Σ) stems from fields outside psychology. For instance, in structural regression analysis (Bekker & de Leeuw, 1987) it is important to determine the maximal number of linear dependencies that may exist between the structural (common) parts of the variables. This can be done by finding mr(Σ), and examining the null-space of $(\Sigma - U^2)$.

From a numerical point of view, mr(Σ) is a well-behaved quantity because it is usually stable if it is at or above the Ledermann bound (Shapiro, 1982a, Theorem 2.3.). That is, slightly changing the elements of Σ will not affect mr(Σ). On the other hand, a numerical approach to mr(Σ) has a fundamental limitation in that whatever algorithm is applied, it has to be terminated in a finite number of iterations. As a result, one may

encounter certain eigenvalues of $(\Sigma - U^2)$ differing slightly from zero merely because of lack of computational accuracy. When the smallest $(n - r)$ eigenvalues are very close to zero but not exactly zero, then one cannot decide whether this has to be attributed to inaccuracy or to an underestimation of the r needed. What can be done, however, is to evaluate the accuracy of the algorithm in cases where $\text{mr}(\Sigma)$ is known or at least has a known upper bound. Accordingly, simulation studies were conducted where $\text{mr}(\Sigma)$ was known or bounded, and it was tested whether f became close enough to zero when the correct or maximal value of r was used. To obtain satisfactory results, it was necessary to refine the algorithm used by ten Berge and Kiers (1988) considerably. The main obstacle was the property of the modified Bentler-Woodward procedure for constrained minimum trace factor analysis (Bentler & Woodward, 1980; ten Berge, Snijders, & Zegers, 1981) to yield some very small negative eigenvalues, implying a minor violation of (1), upon termination of the iterations. In addition, it was deemed necessary to have an increased accuracy in cases where f becomes close to zero. Various checks and restart options had to be inserted before a well-behaved algorithm could be obtained. The resulting algorithm is the topic of the next section.

An Algorithm for Minimizing $g(X,U^2)$

For a given $n \times n$ matrix Σ and a fixed rank parameter r, we have to minimize (3) subject to (1). A monotonically convergent algorithm can be obtained by iteratively optimizing X for fixed U^2, and U^2, subject to (1), for fixed X. These two steps comprise the main iterations. Updating X for fixed U^2 is a straightforward eigenvector problem that requires no further discussion. However, optimizing Σ subject to (1) for fixed X is rather cumbersome. This part of the main iterations requires an iterative procedure (inner iterations) that will be discussed shortly.

In terms of the main iterations, the algorithm used proceeds as follows. First, U^2 is initialized as a zero matrix. This simple choice for U^2 has proven more successful as a start than anything else tried. X is obtained as the $n \times (n - r)$ matrix of eigenvectors of $(\Sigma - U^2)$, associated with the smallest $(n - r)$ eigenvalues. From then on, U^2 and X are iteratively updated, until the absolute difference between the previous value of g, evaluated upon updating X, and the current value of g, is smaller than $\alpha 10^{-4}$, where α is a positive parameter that is initialized as unity, but decreased when g approaches zero (i.e., when $g < .01$, continue with $\alpha = .01$; when $g < .001$, use $\alpha = .001$ from then on). Upon convergence of the main iterations, g is evaluated. If $g < .000001$, computations are terminated, and g is considered to have converged to zero. Otherwise, the entire procedure is restarted, initializing U^2 as $\sigma_1^2 e_1 e_1'$, where σ_1^2 is the observed variance of the first variable, and e_1 is the first column of the $n \times n$ identity matrix. Again, g is evaluated upon convergence, and if $g < .000001$, computations are terminated. Otherwise, another restart, now with $U^2 = \sigma_2^2 e_2 e_2'$, is carried out, and so on. If necessary, n restarts of this form are used. If these fail to yield a value of $g < .000001$, the smallest value of g obtained is considered to be nonzero for the particular value of r adopted.

Thus far, we have not discussed the inner iterations that optimize U^2 for fixed X, which will be taken up now. Let the diagonal matrix W^2 be defined as

$$W^2 = \text{Diag } (XX'). \tag{4}$$

It is clear from (3) that minimizing $g(X,U^2)$ for fixed X is the same as maximizing

$$h(U^2) = \text{tr } X'U^2X = \text{tr } W^2U^2, \tag{5}$$

subject to $\Sigma \geq U^2 \geq 0$, or equivalently, subject to

$$W\Sigma W \geq W^2U^2 \geq 0. \tag{6}$$

Maximizing (5) subject to (6) is a weighted minimum trace factor analysis (Shapiro, 1982b) problem for Σ, that can be treated as an unweighted problem (Bentler & Woodward, 1980; ten Berge et al., 1981) for $W\Sigma W$. Accordingly, the algorithm updates W^2U^2 iteratively. We adopted the modified Bentler-Woodward procedure (MBWP; ten Berge et al., 1981) for these inner iterations. Essentially, the MBWP seeks the global minimum of the function

$$l(T) = \text{tr } TT'C, \tag{7}$$

where T is an $n \times n$ matrix constrained by Diag $(TT') \geq I$, and C is a given $n \times n$ covariance matrix, which is $W\Sigma W$ in the present context. T is initialized randomly and orthonormalized by Gram-Schmidt, in the second main iteration. In subsequent main iterations, T is initialized as the last computed T in the previous main iteration.

The parameter α, used in the stopping criterion for the main iterations, also plays a role in the inner iterations. The inner iterations are terminated as soon as the difference between the previous value of l and the current value of l becomes less than $\alpha 10^{-4}$. Then the smallest eigenvalue of $(W\Sigma W - W^2U^2)$ is evaluated. In theory it should be zero, but in practice in turns out to be slightly negative. If this eigenvalue is smaller than $-\alpha 10^{-5}$, the inner iterations are continued until it is equal to or above this value. If necessary, T is restarted entirely from scratch. Next, U^2 is evaluated from W^2U^2, and the algorithm moves to updating X. The accuracy of the algorithm may appear to be wasteful as a procedure for determining amr(Σ). However, the algorithm was especially designed to be accurate in determining mr(Σ).

A Simulation Study

To test the efficiency of our procedure in determining mr(Σ), two types of matrices were constructed. First, 90 matrices were constructed with known minimum rank. This was done by constructing Σ^{-1} rather than Σ itself. Specifically, if Σ^{-1} has an $(n - k) \times (n - k)$ positive diagonal submatrix in the upper left-hand corner, and a $k \times k$ submatrix of strictly positive elements in the lower right-hand-corner, and if the lower left off-diagonal $k \times (n - k)$ submatrix has at least one column of strictly positive elements, then mr(Σ) = k (P. A. Bekker, personal communication, April 20, 1989). Matrices of this pattern were constructed for various values of n and k (see the upper half of Table 1) by randomly filling a lower-triangular $n \times n$ matrix with elements from the uniform [−.5, .5] distribution, except for the off-diagonal elements of a $(n - k) \times (n - k)$ diagonal submatrix that were set to zero. Next, the nonzero elements below the diagonal were copied in the cells above the diagonal to enforce symmetry, and the signs of certain negative elements were reflected to obtain the pattern specified by Bekker. Finally, the smallest eigenvalue minus .5 was subtracted from each diagonal element to enforce positive definiteness, and the resulting matrix was inverted to obtain Σ.

In addition, 90 matrices with bounded mr(Σ) were constructed by filling an $n \times k$ matrix A randomly with elements from the uniform [−1, 1] distribution, and adding elements from the uniform [0, 1] distribution to the diagonal of AA'. The resulting Σ has an explicit communality solution for $r = k$. The 90 matrices of this type are recorded in the lower half of Table 1.

For the 180 constructed matrices, our algorithm, using $r = k$, should yield a value zero for $g(X,U^2)$ and hence for f. In all cases, it turned out that g was less than .000001

TABLE 1

Number of Cases Where Only One Start Was Needed

order	mr(Σ)	r	replications	one start only
7×7	4	4	20	16
8×8	5	5	20	15
8×8	6	6	20	18
10×10	8	8	20	20
10×10	6	6	10	8
7×7	$\underline{\leq}4$	4	20	20
8×8	$\underline{\leq}5$	5	20	18
8×8	$\underline{\leq}6$	6	20	20
10×10	$\underline{\leq}8$	8	20	20
10×10	$\underline{\leq}6$	6	10	6

upon convergence. Also, the first starting option was already successful in a majority of cases, as can be seen from Table 1. Typically, computation times (for each start separately) were between one and two minutes on a personal computer. For the matrices with known mr(Σ), it is clear that our procedure should yield a nonzero value of g when $r = k - 1$ is used. This was tested in 12 cases. Throughout these cases, the obtained value of g was clearly above .000001. In fact, all 12 values were above .01, indicating the discriminatory power of our procedure.

Discussion

The simulation results presented indicate that a reliable method for determining mr(Σ) has been obtained. It should be noted, however, that the criterion of zeroness (<.000001) used in our algorithm has proven adequate only for matrices with elements of similar size as used in the simulations. That is, the first 90 matrices had a largest eigenvalue 2, implying that the elements in these matrices are in the range [−2, 2].

Similarly, the remaining 90 matrices were constrained to have elements in the range $[-n, n]$. If the algorithm is to be used for matrices with larger elements, then either the criterion of zeroness may have to be relaxed, or the matrices can be rescaled to have smaller elements, for instance, by using the correlation matrix instead of the covariance matrix Σ. The simulation studies do not imply that the algorithm is efficient in determining amr(Σ). It should be noted that extremely high standards of computational accuracy were imposed to distinguish between $g = 0$ and $g > 0$. When zeroness is no longer the main issue, computational accuracy can be relaxed considerably.

References

Bekker, P. A., & de Leeuw, J. (1987). The rank of reduced dispersion matrices. *Psychometrika, 52*, 125–135.

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

Della Riccia, G., & Shapiro, A. (1982). Minimum rank and minimum trace of covariance matrices. *Psychometrika, 47*, 443–448.

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

Shapiro, A. (1982a). Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika, 47*, 187–199.

Shapiro, A. (1982b). Weighted minimum trace factor analysis. *Psychometrika, 47*, 243–264.

Takeuchi, K., Yanai, H., & Mukherjee, B. N. (1982). *The foundations of multivariate analysis*. New Delhi: Wiley.

ten Berge, J. M. F., & Kiers, H. A. L. (1988). Proper communality estimates minimizing the sum of the smallest eigenvalues for a covariance matrix. In M. G. H. Jansen & W. H. van Schuur (Eds.), *The many faces of multivariate analysis. Proceedings of the SMABS-88 Conference in Groningen* (pp. 30–37). Groningen: RION.

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

Manuscript received 1/3/90
Final version received 5/11/90