PSYCHOMETRIKA—VOL. 48, NO. 4.
DECEMBER, 1983

A GENERALIZATION OF KRISTOF'S THEOREM ON THE TRACE OF CERTAIN MATRIX PRODUCTS

JOS M.F. TEN BERGE

UNIVERSITY OF GRONINGEN

Kristof has derived a theorem on the maximum and minimum of the trace of matrix products of the form $X_1\hat{\Gamma}_1X_2\hat{\Gamma}_2 \cdots X_n\hat{\Gamma}_n$ where the matrices $\hat{\Gamma}_i$ are diagonal and fixed and the X_i vary unrestrictedly and independently over the set of orthonormal matrices. The theorem is a useful tool in deriving maxima and minima of matrix trace functions subject to orthogonality constraints. The present paper contains a generalization of Kristof's theorem to the case where the X_i are merely required to be submatrices of orthonormal matrices and to have a specified maximum rank. The generalized theorem contains the Schwarz inequality as a special case. Various examples from the psychometric literature, illustrating the practical use of the generalized theorem, are discussed.

Key words: constrained least-squares problems, constrained matrix trace optimization, matrix inequality.

Kristof [1970] has derived the following theorem on the trace of certain matrix products.

Theorem 1. (Kristof's theorem). Let X_i be an orthonormal matrix, $\hat{\Gamma}_i$ be a fixed diagonal matrix, Γ_i be the diagonal matrix obtained from $\hat{\Gamma}_i$ by arranging the absolute values of the elements in the diagonal of $\hat{\Gamma}_i$ in (weakly) descending order, $i = 1, \ldots, n$, all matrices of order $m \times m$ with $m \geq 2$. Then, under unrestricted and independent variation of X_i,

$$-\text{tr}\, \Gamma_1\Gamma_2 \cdots \Gamma_n \leq \text{tr}\, X_1\hat{\Gamma}_1X_2\hat{\Gamma}_2 \cdots X_n\hat{\Gamma}_n \leq \text{tr}\, \Gamma_1\Gamma_2 \cdots \Gamma_n. \tag{1}$$

The limits can be attained.

Proof. See Kristof [1970]. □

Kristof has demonstrated the practical utility of Theorem 1 in various psychometric applications. Typically, these applications involve optimization problems of traces of matrix products under orthonormality constraints, for which closed-form solutions exist. Using Kristof's theorem, one can find global optima at once, without having to resort to partial differentiation with Lagrange multipliers; see also Levin [1979].

The present paper is concerned with generalizing Kristof's theorem. The need for this is apparent from a number of psychometric problems in which the matrices $X_1, \ldots, X_n$ of Kristof's theorem are required to be only semi-orthonormal, i.e., row-wise or column-wise orthonormal. Clearly, it is desirable to relax the orthonormality conditions of Kristof's theorem to the effect that semi-orthonormal matrices can be handled. This will be done in Theorem 2 below. In fact, this Theorem offers even greater generality to the matrices $X_1, \ldots, X_n$. That is, they need merely be submatrices of orthonormal matrices. The reason for seeking such generality is theoretical: Theorems should be derived in the greatest possible generality.

For later reference, the following definitions and lemmas will be convenient.

The author is obliged to Frits Zegers and Dirk Knol for critically reviewing a previous draft of this paper.
Requests for reprints should be sent to Jos ten Berge, Subfaculteit Psychologie R.U. Groningen, Grote Markt 32, 9712 HV Groningen, The Netherlands.

0033-3123/83/1200-5026$00.75/0
© 1983 The Psychometric Society

Definition 1. A matrix is suborthonormal if it is a submatrix of an orthonormal matrix. The set of suborthonormal matrices is called U.

Definition 2. A matrix is semi-orthonormal if it is row-wise and/or column-wise orthonormal. The set of semi-orthonormal matrices is called V.

Let W denote the set of orthonormal matrices. Then $W \subset V \subset U$.

Lemma 1. Every suborthonormal matrix can be augmented to a semi-orthonormal matrix by adding only rows and by adding only columns to it.

Proof. Let X be a suborthonormal matrix. Construct an orthonormal matrix which has X as a submatrix and the Lemma is evident. □

Lemma 2. A matrix is suborthonormal if and only if its singular values are in the range [0, 1].

Proof. Let X be an $n \times k$ matrix, $n \geq k$, with Eckart-Young or singular value decomposition (SVD) or basic structure, cf. Green [1969, 314–316]

$$X = PDQ', \tag{2}$$

where $P'P = Q'Q = QQ' = I_k$ and D is diagonal with diagonal elements $d_1 \geq d_2 \geq \cdots \geq d_k \geq 0$. The diagonal elements of D are the singular values of X. Singular values are defined to be nonnegative and arranged in (weakly) descending order throughout the present paper.

Let it be given that $d_1 \leq 1$. Construct the semi-orthonormal matrix

$$X_+ = [X \vdots P(I - D^2)^{1/2} \vdots P_*] \tag{3}$$

where P_* is an orthonormal completion of P. Clearly, $X_+ X'_+ = I_n$ which implies that $X \in U$. Conversely, let $X \in U$. By Lemma 1 there is an X_* such that $[X \vdots X_*]$ is row-wise orthonormal. For every w with $w'w = 1$ we have

$$1 = w'w = w'Iw = w'(XX' + X_* X'_*)w \geq w'XX'w \tag{4}$$

which implies that X has no singular values greater than 1. □

Lemma 3. A matrix is semi-orthonormal if and only if its singular values are unity.

Proof. Let X be an $n \times k$ matrix, $n \geq k$, with SVD

$$X = PDQ' \tag{5}$$

where $P'P = QQ' = I$. If $D = I$ then $X'X = QP'PQ' = I$. Conversely, if $X'X = I$ then the eigenvalues of $X'X$ are unity, hence the singular values of X are unity. □

Lemma 4. The product of two suborthonormal matrices is suborthonormal.

Proof. Let $A \in U$ and $B \in U$, where A has k columns and B has k rows. Expand A into a column-wise orthonormal A_+ and B into a row-wise orthonormal B_+. The nonzero eigenvalues of $A_+ B_+ B'_+ A'_+ = A_+ A'_+$ are the eigenvalues of $A'_+ A_+ = I$. Using Lemma 2, this implies that $A_+ B_+$ is suborthonormal. Since AB is a submatrix of $A_+ B_+$, it must also be suborthonormal. □

Theorem 2. (Generalized Theorem 1). Let X_i be a suborthonormal matrix with rank $\leq r_i$, $\hat{\Gamma}_i$ be a fixed diagonal matrix, Γ_i be the diagonal matrix obtained from $\hat{\Gamma}_i$ by arranging the absolute values of the diagonal elements of $\hat{\Gamma}_i$ in (weakly) descending order, $i = 1, \ldots, n$. Let m be the largest number of rows or columns of the X_i, and Δ_i be the diagonal $m \times m$ matrix containing Γ_i in its upper left corner and zeroes elsewhere, $i = 1, 2, \ldots, n$.

Then

$$-\text{tr } \Delta_1\Delta_2 \cdots \Delta_n E_r \leq \text{tr } X_1\hat{\Gamma}_1 X_2\hat{\Gamma}_2 \cdots X_n\hat{\Gamma}_n \leq \text{tr } \Delta_1\Delta_2 \cdots \Delta_n E_r \tag{6}$$

where $r = \min(r_i)$, and E_r is the $m \times m$ matrix containing I_r in its upper left corner and zeroes elsewhere.

Proof. Define Y_i as the $m \times m$ matrix containing X_i in its upper left corner and zeroes elsewhere, $i = 1, \ldots, n$. Let Y_i have the SVD

$$Y_i = P_i D_i Q_i' \tag{7}$$

where P_i and Q_i are orthonormal and D_i is diagonal. Defining

$$f(X) \overset{d}{=} \text{tr } X_1\hat{\Gamma}_1 X_2\hat{\Gamma}_2 \cdots X_n\hat{\Gamma}_n$$

we have

$$f(X) = \text{tr } Y_1\hat{\Delta}_1 Y_2\hat{\Delta}_2 \cdots Y_n\hat{\Delta}_n \tag{8}$$

where $\hat{\Delta}_i$ is the $m \times m$ diagonal matrix containing $\hat{\Gamma}_i$ in its upper left corner and zeroes elsewhere. From (7) and (8) we have

$$f(X) = \text{tr } P_1 D_1 Q_1'\hat{\Delta}_1 P_2 D_2 Q_2'\hat{\Delta}_2 \cdots P_n D_n Q_n'\hat{\Delta}_n. \tag{9}$$

Applying Theorem 1 to (9) yields

$$-\text{tr } D_1\Delta_1 D_2\Delta_2 \cdots D_n\Delta_n \leq f(X) \leq \text{tr } D_1\Delta_1 D_2\Delta_2 \cdots D_n\Delta_n. \tag{10}$$

For some Y_i we have rank $Y_i \leq r$. Hence there must be a D_i with at most r nonzero elements and $D_i = D_i E_r$. Therefore, (10) can be reduced to

$$-\text{tr } D\Delta E_r \leq f(X) \leq \text{tr } D\Delta E_r \tag{11}$$

where $D = D_1 D_2 \cdots D_n$ and $\Delta = \Delta_1\Delta_2 \cdots \Delta_n$. Clearly, tr $D\Delta E_r$ only depends on the first r diagonal elements of D. Noting that the X_i are suborthonormal and every Y_i has the same nonzero singular values as X_i we can infer from Lemma 2 that the first r elements in the diagonal of D are in the interval [0, 1]. As a result (11) implies

$$-\text{tr } \Delta E_r \leq f(X) \leq \text{tr } \Delta E_r. \tag{12}$$

This completes the proof of Theorem 2.

Theorem 2 differs from Theorem 1 in several respects. The most striking difference is that the X_i are no longer required to be orthonormal. Second, the X_i need no longer be square. Third, the assumption that the X_i vary independently and unrestrictedly (except for their ranks) over the set of suborthonormal matrices is not made. The reason for this is that the assumption is not met in some of the intended applications, see below. In the absence of the assumption, the statement "the limits can be attained" had to be omitted from Theorem 2. However, it can readily be verified that the limits of (6) can be attained if the X_i of Theorem 2 are varying independently and (except for rank) unrestrictedly over the set of suborthonormal matrices.

In various applications some of the X_i are restricted to have rank $X_i = r_i$ instead of rank $X_i \leq r_i$. Obviously, this modification does not affect the validity of (6).

Kristof [1970, p. 522] derived a second equivalent version of Theorem 1 in which the diagonal matrices $\hat{\Gamma}_i$ are replaced by arbitrary square matrices A_i. The special case $n = 2$ of this version had been derived earlier by Von Neumann [1937] as was noted by Kristof. Green [1969, p. 317] applied Von Neumann's result and pointed out that it also holds if X_1 and X_2 are orthonormals of different orders. Since orthonormal matrices are suborthonormal, this generalization of Von Neumann's result is a special case of the $n = 2$ version of Theorem 2.

It may be worth noting that the well-known Schwarz inequality is also a special case of the $n = 2$ version of Theorem 2. To verify this, let x and y be nonzero vectors of order k. The suborthornormal "matrices" $x'(x'x)^{-1/2}$ and $y(y'y)^{-1/2}$ can be taken as X_1 and X_2, respectively, with $r = 1$ and $\hat{\Gamma}_1 = I_k$ and $\hat{\Gamma}_2 = I_1$. Then Theorem 2 implies

$$-1 \leq x'(x'x)^{-1/2}y(y'y)^{-1/2} \leq 1 \tag{13}$$

or, equivalently,

$$-(x'x)^{1/2}(y'y)^{1/2} \leq x'y \leq (x'x)^{1/2}(y'y)^{1/2} \tag{14}$$

which is the Schwarz inequality.

Applications of Theorem 2

This section contains three applications of Theorem 2 to psychometric problems. The first problem has also been treated by Kristof [1970], using Theorem 1. Our treatment of this problem allows the elimination of a restrictive assumption. The remaining two applications are the principal components problem and canonical correlation problem, respectively. Theorem 2 permits very direct and highly general solutions to these problems.

Application 1. Multiple regression (Kristof's Example 4). Let X be an $s \times t$ matrix ($s \geq t$) of rank t, containing predictor variables, and Y be an $s \times u$ matrix of criterion variables. It is desired to minimize $\eta(B) = \text{tr } (Y - XB)(Y - XB)'$, where B is a $t \times u$ matrix of regression weights.

Noncalculus solutions for this problem are well-known, e.g., Bock [1975, p. 170–171] and do not require Theorem 1. However, in order to illustrate the range of possible applications of Theorem 1, Kristof applied this Theorem to this problem.

A remarkable feature of Kristof's derivation is that it rests on the unnatural assumption that $t = u$. When Theorem 2 is applied, this assumption can be omitted. This can be seen from the key inequality $\text{tr } (T'V)M(W'U)\Lambda \leq \text{tr } M\Lambda$, cf. Kristof [1970, p. 526]. For instance, if $t > u$ then $T'V$ is suborthonormal of rank $\leq u$; $W'U$ is orthonormal of rank u, and M and Λ are diagonal, nonnegative, of order $u \times u$, with diagonal elements arranged in (weakly) descending order. Therefore, Theorem 2 gives $\text{tr } (T'V)M(W'U)\Lambda \leq \text{tr } M\Lambda E_u = \text{tr } M\Lambda$. The remaining part of the derivation is analogous to Kristof's, and need not be repeated here.

Application 2. Principal components analysis. Consider the problem of jointly determining p principal components for a $k \times k$ correlation matrix R. Typically, this problem is treated as determining a $k \times p$ matrix B which maximizes the sum of squared loadings $\text{tr } B'R^2B$ subject to the orthogonality constraint $B'RB = I_p$. Let R have rank $r \geq p$ and eigendecomposition $R = K\Lambda K'$, $K'K = I_r$, Λ diagonal, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. The problem is to maximize

$$\eta(B) = \text{tr } B'R^2B = \text{tr } B'K\Lambda^2K'B \tag{15}$$

subject to $B'RB = B'K\Lambda K'B = I_p$. This constraint implies that $B'K\Lambda^{1/2}$ is suborthonormal of rank p. Hence Theorem 2 yields

$$\eta(B) = \text{tr } (B'K\Lambda^{1/2})\Lambda(\Lambda^{1/2}K'B) \leq \text{tr } \Lambda E_p. \tag{16}$$

This upper bound can be attained for

$$B_* = K\Lambda^{-1/2} \begin{bmatrix} I_p \\ \cdots \\ 0 \end{bmatrix} N,$$

where N is an arbitrary $p \times p$ orthonormal matrix. B_* is the well-known solution for the principal components problem.

Application 3. Canonical correlations. Kristof [1970, p. 529] mentioned the possibility of solving the canonical correlation problem by Theorem 1. However, it is not clear how this can be done if the two sets of variables involved have different numbers of variables or if less than the maximal number of canonical variates are determined. It will be shown here that Theorem 2 permits a fully general treatment.

Let X_1 be a standardized $n \times k$ matrix of rank r_1 and X_2 be a standardized $n \times m$ matrix of rank r_2, and let $r_1 \geq r_2$. Suppose it is desired to determine r pairs of canonical variates from X_1 and X_2, $r \leq r_2$. Then the problem is to determine a $k \times r$ matrix B_1 and an $m \times r$ matrix B_2 which maximize

$$\eta(B_1, B_2) = \text{tr } B_1'X_1'X_2B_2 \tag{17}$$

subject to the constraints

$$B_1'X_1'X_1B_1 = B_2'X_2'X_2B_2 = I_r. \tag{18}$$

Define the Eckart-Young decompositions $X_1 = P_1D_1Q_1'$, $X_2 = P_2D_2Q_2'$ and $P_1'P_2 = U\Gamma V'$, with $P_1'P_1 = Q_1'Q_1 = I_{r_1}$, $P_2'P_2 = Q_2'Q_2 = I_{r_2}$, $U'U = V'V = VV' = I_{r_2}$, D_1, D_2 and Γ diagonal. From (18) it is clear that $C_1' = B_1'Q_1D_1$ and $C_2' = B_2'Q_2D_2$ are semi-orthonormal of rank r. Now $\eta(B_1, B_2)$ can be written as

$$\eta(B_1, B_2) = \text{tr } B_1'Q_1D_1P_1'P_2D_2Q_2'B_2 = \text{tr } C_1'U\Gamma V'C_2. \tag{19}$$

Since $C_1'U$ is suborthonormal of rank $\leq r$ and $V'C_2$ is suborthonormal of rank r, Theorem 2 implies

$$\eta(B_1, B_2) \leq \text{tr } \Gamma E_r. \tag{20}$$

This bound can be attained by taking

$$B_1 = Q_1D_1^{-1}U \begin{bmatrix} N \\ \cdots \\ 0 \end{bmatrix} \quad \text{and} \quad B_2 = Q_2D_2^{-1}V \begin{bmatrix} N \\ \cdots \\ 0 \end{bmatrix} \tag{21}$$

where N is an arbitrary orthonormal $r \times r$ matrix. This is sufficient to show that B_1 and B_2 as given in (21) maximize the sum of the first r canonical correlations.

Discussion

The applications discussed above clearly demonstrate the practical utility of Theorem 2. Many other applications could be cited. For instance, the proofs in Kristof's Example 3 and 5 [Kristof, 1970] can be simplified if Theorem 2 is used instead of Theorem 1. In fact Kristof's proofs for these Examples can be interpreted as implicit generalizations of Theorem 1 in the direction of Theorem 2.

REFERENCES

Bock, R. D. *Multivariate statistical methods in behavioral research.* New-York: McGraw-Hill, 1975.

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

Kristof, W. A theorem on the trace of certain matrix products and some applications. *Journal of Mathematical Psychology*, 1970, *7*, 515–530.

Levin, J. Applications of a theorem on the traces of certain matrix products. *Multivariate Behavioral Research*, 1979, *1*, 109–113.

Von Neumann, J. Some matrix-inequalities and metrization of matrix-space. *Tomsk University Review*, 1937, 1, 286–300. Reprinted in A. H. Taub (Ed.), *John von Neumann collected works* (Vol. IV). New-York: Pergamon, 1962.

Manuscript received 9/21/82
Final version received 3/16/83