Analysis of $K$ sets of data, with differential emphasis on agreement between and within sets

Mohamed Hanafi$^a$,*, Henk A.L. Kiers$^b$

$^a$Unité Mixte de Recherche en sensométrie et chimiométrie (ENITIAA-INRA), Département SMAD, rue de la Géraudiére, BP 82225, 44322 Nantes Cedex 3, France
$^b$Heymans Institute, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands

Received 18 July 2005; received in revised form 14 April 2006; accepted 14 April 2006
Available online 12 May 2006

Abstract

A general class of methods for (partial) rotation of a set of (loading) matrices to maximal agreement has been available in the literature since the 1980s. It contains a generalization of canonical correlation analysis as a special case. However, various other generalizations of canonical correlation analysis have been proposed. A new general class of methods for each such alternative generalization of canonical correlation is proposed. Together, these general classes of methods form a superclass of methods that strike a compromise between explaining the variance within sets of variables and explaining the agreement between sets of variables, as illustrated in some examples. Furthermore, one general algorithm for finding the solutions for all methods in all general classes is offered. As a consequence, for all methods in the superclass of methods, algorithms are available at once. For the existing methods, the general algorithm usually reduces to the standard algorithms employed in these methods, and thus the algorithms for all these methods are shown to be related to each other.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Generalizations of canonical correlation analysis; MAXBET; MAXDIFF; Iterative majorization

1. Introduction

Hotelling (1935, 1936) proposed canonical correlation analysis as a method for describing the relation between the scores of a set of observation units on two groups of variables. Let $X_i$ denote the matrix with columnwise centered scores of $n$ observation units on $m_i$ variables, $i=1,2$. Then canonical correlation analysis consists of finding linear combinations $X_1t_1$, of the columns of $X_1$, and $X_2t_2$, of the columns of $X_2$, such that their correlation is maximal. Thus, canonical correlation analysis consists of maximizing:

$$\tau(t_1,t_2) = \text{corr}(X_1t_1, X_2t_2),$$

over $t_1$ and $t_2$, where $\text{corr}(.,.)$ denotes the product moment correlation coefficient. Having found a first set of canonical weight vectors $t_1^1$ and $t_2^1$ defining the first set of canonical variates, one may continue the search for canonical variates...
by maximizing \( \tau \) over \( t_1 \) and \( t_2 \), subject to the constraints that \( X_1 t_1 \) and \( X_2 t_2 \) are orthogonal to the canonical variates that have already been found (which, in the first round, are \( X_1 t_1^1 \) and \( X_2 t_2^1 \), respectively).

Several generalizations of canonical correlation analysis have been proposed for handling situations with more than two sets of variables (e.g., see Kettenring, 1971, for an overview). Here we limit ourselves to the generalizations that consist of maximizing the following functions:

\[
\text{SUMCOR} : \sum_{i \neq j, j=1}^{K} \text{corr} (X_i t_i, X_j t_j),
\]

\[
\text{MAXVAR} : \lambda_i^2 \left( \left( \text{corr} (X_i t_i, X_j t_j) \right) \right),
\]

\[
\text{SSQCOR} : \sum_{i \neq j, j=1}^{K} \text{corr}^2 (X_i t_i, X_j t_j),
\]

over \( t_1, \ldots, t_K \), where \( \lambda(.) \) denotes the largest eigenvalue of the matrix in parentheses.

The SUMCOR criterion, has been proposed by Horst (1965). The two other criteria have been proposed by Kettenring (1971), where the criterion MAXVAR has been considered by Kettenring as a unified form of two old methods: the rank one analysis proposed by Horst (1961) and Carroll’s (1968) generalized canonical correlation.

The above generalizations of canonical correlation analysis (GCA) are all based on maximizing a function of the correlation between canonical variates. Various other methods for analyzing relations among sets of variables have been proposed. For the case of two groups of variables, interbattery factor analysis (Tucker, 1958) is an important case in point. An important difference between (generalized) interbattery factor analysis and GCA is that the former does not only focus on optimally describing the relation between sets of variables, but in addition requires that the variance within sets of variables is explained well by the components used. Redundancy analysis (Van den Wollenberg, 1977) and PLS (Wold, 1985; see also Löhmöller (1987), and Tenenhaus et al. (1995) for overviews) also go beyond merely explaining the relation between two sets of variables, but the first does so in an asymmetric way, and the second in a successive, dimension by dimension approach. In the present paper, we focus on symmetric treatment of the data sets, and on approaches finding all dimensions simultaneously.

A third kind of method for analyzing relations between two or more sets of variables consists of techniques for (partial) rotation of several matrices to maximal agreement. A number of these techniques have been collected in various classes of techniques by Van de Geer (1984), called MAXBET, MAXDIFF, MAXRAT and MAXNEAR. The first two classes of these techniques have been further generalized by Ten Berge (1986, 1988). Like GCA techniques, the MAXBET/MAXDIFF techniques search for weights to form linear combinations that optimally agree with each other, but as in interbattery factor analysis, besides agreement, some emphasis is placed on explaining the variance within the sets of variables. Interestingly, this class of techniques has one of the above mentioned GCA techniques (namely, SUMCOR, see Van de Geer, 1984) as a special case. It is the purpose of the present paper to propose a general framework comprising the general class of techniques proposed by Ten Berge, as well as the three GCA techniques mentioned earlier. Specifically, like the general class of techniques that has SUMCOR as a special case, here we will propose two general classes of techniques that have, respectively, MAXVAR and SSQCOR as special cases.

The proposal of two new classes of techniques leads to various new methods, which combine ingredients underlying the existing methods for agreement rotation and GCA. Thus, new possibilities arise for striking compromises between emphasis on explaining the variance within sets of variables and explaining the agreement between sets of variables. By treating all methods in a general framework, it will be made clear how each class of methods, and each member of these classes, chooses this compromise, and it will be shown that this leads to attractive new possibilities for combining the optimization of agreement and of explaining of variance within sets.

As a spin-off of treating all methods in a general framework it was found that all methods can be carried out by means of a single very general algorithm. This algorithm is based on the Iterative Majorization approach (Heiser, 1995). As a consequence, for all new methods arising from our framework, algorithms are available at once. For the existing methods, the general algorithm usually reduces to the standard algorithms employed in these methods, and thus the algorithms for all these methods are shown to be related to each other. Thus, the second purpose of the present paper is to offer this general algorithm for the analyses involved in all methods in all three general classes of techniques proposed here.
In Section 2, we will describe the general class of techniques proposed by Van de Geer (1984) and extended by Ten Berge (1986, 1988), and we will focus on the three characteristics distinguishing the techniques in this general class. In Section 3, by using the same distinguishing characteristics, we will propose general classes of techniques based on the two other GCA methods, and the ensuing general framework of methods, will be described. Next, we will show that these classes themselves are distinguished by a fourth characteristic of the methods, in Section 4, and several special cases will be illustrated. Finally, in Section 5, a general algorithm will be proposed for the analyses in all methods.

In the sequel, the following notation is adopted. The matrix $X = [X_1 | X_2 | \cdots | X_K]$ is a data matrix with $n$ rows, and $\sum_{i=1}^{K} m_i$ columns, partitioned in $K$ submatrices $X_i$ of order $n \times m_i$. It is, from now on, assumed that all $X_i$ are centered, so as to ensure that interpretations in terms of variance bias, variance explained, etc., as used later, are indeed warranted. Furthermore, we have $T' = [T'_1 | T'_2 | \cdots | T'_K]$ with $T_i$ of order $m_i \times r$ and $r \leq \min (m_i)$, $i = 1, 2, \ldots, K$.

2. Van de Geer and Ten Berge’s general class of methods, revolving around SUMCOR

Van de Geer (1984) has considered a family of criteria in terms of linear relations among $K$ sets of variables. Van de Geer only considered the situation where the data matrices have the same numbers of columns: $m_1 = m_2 = \cdots = m_K$. This family is based on three independent basic choices (Van de Geer, 1984, p.80.), which serve as distinguishing characteristics in the general class of methods proposed by Van de Geer, which was later extended by Ten Berge (1986, 1988):

(i) What to Analyze? Analyze either the original data matrices $X_1, X_2, \ldots, X_K$ or analyze matrices that give columnwise orthonormal bases $P_1, P_2, \ldots, P_K$ for the original data matrices. The first analysis is denoted “Analysis of $X$” and the latter “Analysis of $P$”. In the description of our general class of methods, we will use the same symbol $X_i$ to denote either the original or the columnwise orthonormal bases matrices, depending on the context.

(ii) Fairness (or Balance). The fairness issue deals with the question (Van de Geer, p. 80) “will a solution be accepted as a good one even if it is dominated by only a few of the sets, ignoring the other sets?”. In case of one single linear combination per set, the weight vectors $t_1, t_2, \ldots, t_K$ may be constrained to have an equal sums of squares $t'_i t_i = t'_2 t_2 = \cdots = t'_K t_K = 1$. Using this constraint will lead to more fairness in the sense that all sets have equal importance in the solution. Orthonormality is related to this in the sense that, if more than one linear combination is used for each set of variables, using orthonormality constraints ensures fairness for the solution as a whole.

As far as more than one linear combination per set is concerned, Van de Geer only considers taking $m$ mutually orthogonal dimensions, implemented by taking the $m \times m$ matrices $T_i$ orthonormal. Overall, Van de Geer thus considered three possibilities: using a single linear combination per set with and without equal sums of squares for the weight vectors, and using $m$ linear combinations per set, with mutually orthogonal weights. Ten Berge (1986, 1988) generalized some of Van de Geer’s methods by allowing for an arbitrary number ($r \leq m$) of linear combinations per set, and requiring columnwise orthonormality for the $m \times r$ matrices $T_i$. Taking this into account, we have the following three types of constraints:

(c1) $T$ has only one column with sum of squares equal to $K$.
(c2) $T$ has only one column; each subcolumn $T_i$ has sum of squares equal to 1.
(c3) $T_i$ is constrained to be columnwise orthonormal.

(iii) Variance bias. Van de Geer (1984) reviewed four criteria with different “variance biases”, that is, different emphasis on the aim to explain variance of the variables in each set by means of the linear combinations for each set. The best known of these are MAXBET and MAXDIFF, of which the former puts most emphasis on explaining variance within sets. The MAXBET and MAXDIFF problems consist of maximizing the functions:

$$\text{MAXBET} : \sum_{i,j=1}^{K} \text{tr} \left( T'_i X'_i X'_j T_j \right),$$

$$\text{MAXDIFF} : \sum_{i \neq j, i,j=1}^{K} \text{tr} \left( T'_i X'_i X'_j T_j \right),$$

over $T_1, \ldots, T_K$, subject to either of three constraints: (c1)–(c3).

The main difference between MAXBET and MAXDIFF is the absence of the terms with $i = j$, which are the very terms that pertain to the within-sets variances. Thus, in this way MAXBET puts more emphasis on describing within-sets variance than does MAXDIFF.
It may be noted that the decision as to what to analyze may have a direct impact on variance bias. Specifically, by choosing to analyze $P_i$ rather than $X_i$, one removes the within-sets structure in the data. Hence by analyzing $P_i$, the difference between MAXBET and MAXDIFF vanishes (neither method has a variance bias), as can also be seen from the fact that, in this case, the difference between the criteria is a constant.

When analyzing $P_i$, MAXBET equals MAXDIFF, hence only three different options remain, corresponding to the three constraints mentioned above. Van de Geer (1984, p. 83) mentions that using constraint (c1) leads to the MAXVAR problem, and using (c2) leads to SUMCOR. Clearly, using (c3) leads to a variant of SUMCOR, where more than one component per set is obtained simultaneously.

Ten Berge (1986) has unified Van de Geer’s general approach, as far as MAXBET is concerned. Specifically, Ten Berge demonstrated that the three different constraints can be handled by one general procedure, upon considering the three constraints as special cases of the constraint $T_i' T_i = I_r$, where $r \leq \min (m_i)$, as follows. MAXBET with (c1) comes down to the $K=1$ case applied to the supermatrix $X$; (c2) and (c3) are trivially subsumed under the constraint $T_i' T_i = I_r$. Ten Berge has offered a monotonically converging algorithm for solving this general MAXBET problem. The parallel MAXDIFF problem had been solved earlier by Ten Berge and Knol (1984, see also Ten Berge, 1988), but is slightly less general because it does not subsume the constraint (c1). Ten Berge (1988) has focussed on unifying the MAXBET and the MAXDIFF problems as far as possible, and has generalized them even further by an approach for successively finding sets of linear combinations.

To summarize, the latest unification consists of general algorithms for the methods MAXBET and MAXDIFF (the two most interesting possibilities for the variance bias choice), handling two or three of the constraints chosen under the Fairness choice, and useful for both, the analysis of $X_i$ and of $P_i$ (“What to Analyze?”). As has been mentioned above, the analysis of $P$ by MAXBET and MAXDIFF leads to the same procedures, which, depending on the constraints used are SUMCOR and MAXVAR. Thus, the general MAXBET and MAXDIFF approaches can be viewed as generalizations of these GCA methods, that allow for taking into account the within-sets variance. However, the MAXVAR criterion can be generalized in a more straightforward way, and similar generalizations of the criterion SSQCOR can be formulated. These will be considered in the next section, and will later be unified in one general framework, where, moreover, all constraints mentioned above can be used.

### 3. Other generalizations of GCA criteria

In the present section, it will be shown that each of the GCA criteria mentioned in Section 1, SUMCOR, MAXVAR and SSQCOR, can be seen as analysis of $P$ variants of the following general criteria (the last four of which are new):

\[
\text{MAXDIFF} : \quad \sum_{i \neq j, i, j=1}^K \text{tr} \left( T_i' X_i X_j T_j \right),
\]

\[
\text{MAXBET} : \quad \sum_{i, j=1}^K \text{tr} \left( T_i' X_i X_j T_j \right),
\]

\[
\text{MAXDIFF A} : \quad \sum_{i=1}^K \left( \alpha_i \text{tr} \left( T_i' T_i \right) + \sum_{j \neq i, j=1}^K \alpha_j \text{tr} \left( T_i' X_j T_j \right) \right)^2,
\]

\[
\text{MAXBET A} : \quad \sum_{i=1}^K \left( \sum_{j=1}^K \alpha_j \text{tr} \left( T_i' X_j T_j \right) \right)^2,
\]

\[
\text{MAXDIFF B} : \quad \sum_{i \neq j, i, j=1}^K \left( \text{tr} \left( T_i' X_i X_j T_j \right) \right)^2,
\]

\[
\text{MAXBET B} : \quad \sum_{i, j=1}^K \left( \text{tr} \left( T_i' X_i X_j T_j \right) \right)^2,
\]
Specifically, SUMCOR will be seen to be an analysis of $P$ variant of MAXBET. As has already been mentioned, in the case of analysis of $P$, MAXBET and MAXDIFF coincide, hence SUMCOR is also an analysis of $P$ variant of MAXDIFF. Thus, MAXBET and MAXDIFF generalize the SUMCOR criterion. Note that, whereas they generalize the same GCA criterion, they do not coincide in general: In the analysis of $X$ (rather than of $P$) the criteria differ in that the former gives more emphasis to within-sets variance than does the latter. Analogously, MAXBET A and MAXDIFF A will be shown to generalize the MAXVAR criterion (with MAXBET A putting more emphasis on within-sets variance than does MAXDIFF A). Finally, MAXBET B and MAXDIFF B both generalize the SSQCOR criterion (with MAXBET B putting more emphasis on within-sets variance than does MAXDIFF B).

In the following, we consider how these criteria are extensions of the different GCA criteria (SUMCOR, MAXVAR, SSQCOR). The latter criteria all referred to single column matrices $T_i$ (i.e., $r = 1$), and employed the correlation rather than the inner product. In the next subsection, we will consider the single column cases of the new criteria, and will explain how they extend the original criteria. In Section 3.3, the general case, $r \geq 1$, will be considered.

3.1. The case $r = 1$, distinction by choice ‘What to Analyze?’

The three GCA variants SUMCOR, MAXVAR and SSQCOR have all been expressed as finding vectors $t_1, t_2, \ldots, t_K$ such that $X_i t_1, X_i t_2, \ldots, X_i t_K$ optimally resemble each other in some sense. It will now be shown that optimization of these three criteria is equivalent to optimization of the criteria:

$$
\text{SUMCOR} : \quad \sum_{i \neq j, i, j = 1}^K \left( u_i' P_i P_j u_j \right),
$$

$$
\text{MAXVAR} : \quad \lambda^2 \left( \left\{ u_i' P_i P_j u_j \right\} \right),
$$

$$
\text{SSQCOR} : \quad \sum_{i \neq j, i, j = 1}^K \left( u_i' P_i P_j u_j \right)^2,
$$

over $u_1, \ldots, u_K$, where again $\lambda(.)$ denotes the largest eigenvalue of the matrix in parentheses.

The vector $u_i$ is taken such that $P_i u_i$ has unit sum of squares, which implies the constraint $u_i' u_i = 1, i = 1, \ldots, K$. The equivalence of the maximization of the above criteria with maximizing the original functions over (arbitrary) vectors $t_1, \ldots, t_K$, follows from the fact that, without loss of generality, we may require $X_i t_i$ to have unit length, $i = 1, \ldots, K$, in the original criteria (because this does not affect the correlation), hence $\text{corr} \left(X_i t_i, X_j t_j\right) = \left(t_i' X_i X_j t_j\right)$ for thus constrained vectors $t_i$. Furthermore, $X_i \left( X_i X_i / n \right)^{-1/2} \left( X_i X_i / n \right)^{1/2} t_i = P_i u_i$ with $P_i = X_i \left( X_i X_i / n \right)^{-1/2}$ and $u_i = \left(X_i X_i / n \right)^{1/2} t_i$, so that $u_i' u_i = \left(t_i' X_i X_j t_j\right) = 1$, according to the constraint imposed above. Hence maximization of the above criteria, involving the terms $\left(u_i' P_i P_j u_j\right)$, subject to $u_i' u_i = 1$, is equivalent to maximizing the analogous criteria where $\left(u_i' P_i P_j u_j\right)$ is replaced by $\text{corr} \left(X_i t_i, X_j t_j\right)$. Thus, the correlation-based criteria can all be seen as analysis of $P$ methods.

From these new descriptions of SUMCOR, MAXVAR and SSQCOR, it is at once clear that they do not take within-sets variance into account, simply because in the matrices $P_i$ variances are normalized, and covariances are 0. Furthermore, it is of interest to note that maximization of criterion MAXVAR subject to $u_i' u_i = 1, i = 1, \ldots, K$ can be written more explicitly as the maximization of criterion:

$$
\text{MAXVAR} : \quad \sum_{i=1}^K \left( \sum_{j=1}^K \lambda_j u_i' P_i P_j u_j \right)^2 = \sum_{i=1}^K \left( \lambda_i u_i' u_i + \sum_{j \neq i}^{K} \lambda_j u_i' P_i P_j u_j \right)^2,
$$

over $a$ and $u_1, \ldots, u_K$, subject to $a' a = 1$ and $u_i' u_i = 1, i = 1, \ldots, K$; where $\lambda_i$ is the $i$th element of $a$; this follows from the fact that, with $[R]_{ij} = \left(u_i' P_i P_j u_j\right)$, we can write $\sum_{i=1}^K \left( \sum_{j=1}^K \lambda_j u_i' P_i P_j u_j \right)^2 = a' R^2 a$, the maximum of which over $a$, subject to $a' a = 1$, equals $\lambda^2(R)$. 

Table 1
General framework for the analysis of $K$ sets of Data

<table>
<thead>
<tr>
<th>Analysis of $P$</th>
<th>Analysis of $X$ without variance bias</th>
<th>Analysis of $X$ with variance bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAXDIFF = MAXBET ($r = 1$: SUMCOR)</td>
<td>MAXDIFF</td>
<td>MAXBET</td>
</tr>
<tr>
<td>MAXDIFF A = MAXBET A ($r = 1$: MAXVAR)</td>
<td>MAXDIFF A</td>
<td>MAXBET A ($r = 1$: MCOA)</td>
</tr>
<tr>
<td>MAXDIFF B = MAXBET B ($r = 1$: SSQCOR)</td>
<td>MAXDIFF B</td>
<td>MAXBET B</td>
</tr>
</tbody>
</table>

The equivalent methods are given in brackets.

Clearly, by replacing $P$ by $X$ in the above criteria, we obtain the analysis of $X$ variants for these methods. Surprisingly, only one of these variants coincides with a method proposed in the literature so far: replacing $P_i$ by $X_i$ in SUMCOR leads to the MAXDIFF criterion. The same replacement in the other criteria leads to new analysis of $X$ methods (see Table 1). Here, it is thus proposed to extend the possibilities for the analysis of $X$ to the criteria MAXVAR, SSQCOR, and the resulting criteria (including the MAXDIFF) are the following:

- MAXDIFF ($r = 1$):
  \[ \sum_{i \neq j, i, j = 1}^{K} t'_i X'_i X'_j t_j, \]

- MAXDIFF A ($r = 1$):
  \[ \sum_{i = 1}^{K} \left( x_i t'_i t_i + \sum_{j \neq i, j = 1}^{K} x_j t'_i X'_j t_j \right)^2, \]

- MAXDIFF B ($r = 1$):
  \[ \sum_{i \neq j, i, j = 1}^{K} \left( t'_i X'_i X'_j t_j \right)^2. \]

Clearly, these criteria are the $r = 1$ cases of MAXDIFF, and the new criteria MAXDIFF A and MAXDIFF B proposed in the beginning of Section 3.

The resulting taxonomy of methods for the $r = 1$ case is presented in the first two columns of Table 1. It should be noted that, whereas the methods in the left column ignore the within-sets variance, the methods in the second column of Table 1 do take this into account, as was mentioned before. This can also be seen from the fact that by going from the first to the second one actually replaces the correlation between $X_i t_i$ and $X_j t_j$ by the covariance between these components. From the fact that

\[ \text{cov} (X_i t_i, X_j t_j) = \sqrt{\text{var} (X_i t_i) \text{var} (X_j t_j)} \text{corr} (X_i t_i, X_j t_j), \]

it is clear that the criteria using covariances rather than correlations also take into account within-sets variances (as expressed in $\text{var} (X_i t_i)$), whereas this is ignored in the criteria based on correlations, as are the analysis of $P$ criteria.

3.2. The case $r = 1$, distinction by variance bias

The criteria SUMCOR, MAXVAR and SSQCOR and their generalizations proposed in Section 3.1 do not have a high “variance bias”, because they exclude combinations where $i = j$. With $P$, it was only natural to leave these terms out, because they were constant anyway. With $X$, however, taking these terms into consideration leads to new criteria. The MAXDIFF criterion then changes into the MAXBET criterion, which puts more emphasis on explaining the internal structure than does MAXDIFF. When $r = 1$, the MAXBET criterion can be written as

\[ \text{MAXBET} (r = 1) : \sum_{i = 1}^{K} t'_i X'_i t_i + \sum_{i \neq j, i, j = 1}^{K} t'_i X'_j t_j, \]

which shows that the MAXBET criterion is in fact the MAXDIFF criterion to which some extra emphasis on internal structure (as expressed by $\sum_{i = 1}^{K} t'_i X'_i t_i$), is added. When $r = 1$, we may analogously put more emphasis on internal
structure in the two criteria MAXDIFF A and MAXDIFF B, to obtain:

\[
\text{MAXBET A} \ (r = 1) : \sum_{i=1}^{K} \left( \sum_{j=1}^{K} \alpha_{j} t_{i} t_{j} X_{i}^{T} X_{j} \right)^{2},
\]

\[
\text{MAXBET B} \ (r = 1) : \sum_{i,j=1}^{K} \left( t_{i} t_{j} X_{i}^{T} X_{j} \right)^{2}.
\]

Clearly, these criteria are the \( r = 1 \) cases of MAXBET A and MAXBET B. These methods are presented in the third column of Table 1. Thus, the methods for the analysis of \( X \) can be put in the scheme presented as the right two columns of Table 1.

It is of interest to note that the MAXBET A criterion is equivalent to the multiple co-inertia analysis (MCOA) criterion, proposed by Chessel and Hanafi (1996).

### 3.3. The general case \( r \geq 1 \)

The criteria described above for \( r = 1 \) are all special cases of the criteria with \( r \geq 1 \) mentioned at the beginning of Section 3. In Table 1 distinctions are made with respect to “What to Analyze” and “variance bias”. A further distinction can be made in terms of the Fairness choice: All criteria can be maximized subject to either \( T^{T} T = I \) or \( T_{i}^{T} T_{i} = I \). This completes the new framework which is summarized in Table 1 (noting that for all methods both constraints can be used).

From Table 1 it is clear that the MAXBET and MAXDIFF criteria are \( r > 1 \) generalizations of the SUMCOR criterion, and that similar generalizations of MAXVAR, SSQCOR now follow from the general framework. Also, it can be seen that optimization of MAXBET A criterion, for \( r > 1 \) leads to a generalization of MCOA. Finally, we remark that, whereas in generalized canonical correlation analysis, the matrices \( X_{i} \) are usually the original centered data matrices, in the general framework they can be any matrices. In fact, important examples of the analysis of \( X \) are the comparisons of loading matrices, or configuration matrices, after rotation to optimal agreement. Examples of the analysis of \( P \) are the comparison of indicator matrices for categorical variables, as considered by Meyer (1989, 1992).

### 4. Studying differences between models in the general framework, and some applications

Of the six classes of methods in the general framework, MAXBET and MAXDIFF, with the possibility for applying the criteria in the analysis of \( X \) rather than of \( P \), have been well established in the literature (e.g., Van de Geer, 1984, 1986, vol. 2, 1986, vol. 2), and interesting applications (e.g., in the context of matching rotation) are commonly used. The other classes of methods are new. Therefore, in the present section, attention will be paid to the interpretation of these methods, and some applications of the new methods. Thus, here it is considered how a user might choose from the abundance of methods offered by the general framework.

For the sake of simplicity, we will focus on the case where \( r = 2 \) in all interpretations and applications; for \( r > 2 \), similar interpretations hold. We will first compare methods that appear in the same row of Table 1, and choose to do so for the last row. The methods in the last row are based on the criteria MAXDIFF B and MAXBET B.

In addition, we will compare the interpretations of different criteria appearing in the same column. For the first column this would amount to a comparison of GCA criteria SUMCOR, MAXVAR and SSQCOR, which has, however, to some extent already been done by Kettenring (1971). Here we will compare criteria in the third column: MAXBET, MAXBET A and MAXBET B.

In practice, the choices to be made correspond to how one wishes to describe relations between sets of variables. The choices on What to Analyze and variance bias pertain both to the emphasis on finding components that describe within-sets variance well, and the Fairness choice pertains to the influence of the different sets of the variables (fair or unfair) on the outcome.

#### 4.1. Comparison of interpretations of methods in the same row of Table 1

In the present section, we will compare the criteria MAXDIFF B of \( P \), MAXDIFF B of \( X \), and MAXBET B of \( X \), all appearing in the last row of Table 1. As far as the choice What to Analyze is concerned, MAXDIFF B of \( P \) corresponds
Table 2
Evaluation of eight port wines by four assessors

<table>
<thead>
<tr>
<th>Assessors</th>
<th>Variables</th>
<th>( w_1 )</th>
<th>( w_2 )</th>
<th>( w_3 )</th>
<th>( w_4 )</th>
<th>( w_5 )</th>
<th>( w_6 )</th>
<th>( w_7 )</th>
<th>( w_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1</td>
<td>Red</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Brown</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Soft</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Plum</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>A_2</td>
<td>Ruby</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Tawny</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Intensity</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>A_3</td>
<td>Red</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Blue</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Brown</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Intensity</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>A_4</td>
<td>Depth</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Fresh</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Brightness</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3
Correlations between components for the four sets, for the three different methods

<table>
<thead>
<tr>
<th>MAXDIFF B of ( P )</th>
<th>MAXDIFF B</th>
<th>MAXBET B</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>( A_2 )</td>
<td>( A_3 )</td>
</tr>
<tr>
<td>1.00</td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>0.90</td>
<td>1.00</td>
<td>0.95</td>
</tr>
<tr>
<td>0.95</td>
<td>0.95</td>
<td>1.00</td>
</tr>
<tr>
<td>0.92</td>
<td>0.90</td>
<td>0.89</td>
</tr>
</tbody>
</table>

to the analysis of \( P \), and the other criteria correspond to the analysis of \( X \); as far as Fairness is concerned, here we only consider the case with more than one component, and use \( T'_i T_i = I_r \), \( i = 1, \ldots, K \); in all cases; as far as the variance bias is concerned, the MAXDIFF B method corresponds to the case without variance bias, whereas the MAXBET B method corresponds to the case with variance bias. We will now study what the differences between these criteria actually amount to, by means of a practical application.

The data to be used to illustrate the differences between methods consist of a subset of the data from sensory evaluation of port wines reported by Williams and Langron (1984) and refer to the assessment of the appearance of \( n = 8 \) port wines (denoted \( w_1, \ldots, w_8 \)) by \( K = 4 \) assessors (denoted \( A_1, A_2, A_3, A_4 \)). The data sets are reproduced in Table 2. In an attempt to capture only the common information (agreement) between assessors, and to illustrate how the methods mentioned above differ in doing this, we decided to center (by variables), and to analyze the four data sets by means of three above-mentioned techniques, using \( r = 2 \), and \( T'_i T_i = I_2 \), \( i = 1, 2, 3, 4 \). The most important aspects of the results are presented in Table 3, giving the agreement between assessors (i.e. the correlation between vec \((X_iT_i)\) for different \( i \), where vec(.) denotes the vector containing all the elements of the matrix strung out rowwise into a column vector), demonstrating to what extent the different methods take into account the agreement between sets, and Table 4, giving the sums of explained variances by the components \( X_iT_i \), showing to what extent the within-sets variances are described. Here and in the sequel, as usual, explained variance per variable is defined as the amount of variance explained in the regressions of a variable in \( X_i \) on the set of components in \( X_iT_i \); the total explained variance per set is obtained by summing these explained variances. Note that, also in the analysis of \( P \) variants, the regression of the original variable (in \( X_i \)) on the components is used, not that in \( P_i \).

In the first method, based on the analysis of \( P \), all emphasis is on agreement between sets, and none on within-sets variance. The second method is based on the analysis of \( X \), and for that reason does take within sets variance into account, but the third method does so to a greater extent. This is reflected in the outcomes: The agreement between
Table 4
Explained variances of the components for the four sets, for the three different methods

<table>
<thead>
<tr>
<th></th>
<th>MAXDIFF B of P</th>
<th>MAXDIFF B</th>
<th>MAXBET B</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>12.10</td>
<td>12.21</td>
<td>12.27</td>
</tr>
<tr>
<td>A2</td>
<td>6.92</td>
<td>6.93</td>
<td>6.93</td>
</tr>
<tr>
<td>A3</td>
<td>8.14</td>
<td>8.17</td>
<td>8.18</td>
</tr>
<tr>
<td>A4</td>
<td>2.37</td>
<td>2.69</td>
<td>3.03</td>
</tr>
<tr>
<td>Total</td>
<td>29.53</td>
<td>30.00</td>
<td>30.41</td>
</tr>
</tbody>
</table>

sets (see Table 3) are highest for the results from the first method, and lowest for the results from the third method. Conversely, the average explained variances by the components (see Table 4) is lowest for the first method, and highest for the third method.

4.2. Comparison of interpretations of methods in the same column of Table 1

We will now consider what differences are entailed by moving vertically in Table 1 by comparing the criteria MAXBET, MAXBET A and MAXBET B. Clearly, these cases correspond to the same options as the choices pertaining to What to Analyze and variance bias as are concerned: analysis of $X_i$ with variance bias.

In the present section, we will demonstrate that the differences between methods in different rows of Table 1 entail differences in “fairness”. These will be illustrated on our example data sets. Thus, here it is seen that one may vary fairness by varying constraints imposed, but that this is not the only way of doing so.

As far as the fairness choice is concerned, a “fair” solution can be defined as an overall $\sum_{i=1}^{k} X_i T_i$ component, where the components for all sets take equal parts in the total solution. An unfair solution then is of the form $\sum_{i=1}^{k} z_i X_i T_i$, where different sets of variables have different contributions to the solution, as indicated by the weights $z_i$, $i = 1, \ldots, K$.

Van de Geer considered only the case $r = 1$, and proposed the case where $t t' = 1$, as an instance of unfairness because, upon taking $z_i$ as the norm of $t_i$ and normalizing the subvectors $t_i$ to unit sums of squares, the component $X_t$ will be of the unfair form $\sum_{i=1}^{k} z_i X_i t_i$. Specifically, optimizing the $r = 1$ case of the criterion MAXBET subject to the constraint $t t' = 1$ gives such an unfair treatment to the different sets. Optimizing the $r = 1$ case of the criterion MAXBET subject to the constraints $t_i t_i = 1$, instead gives a fair treatment. It should be emphasized that, as Van de Geer mentioned, “unfair” does not necessarily mean unwanted. In fact, it is very well possible that an unfair solution is preferred over a fair one, in cases where “natural” differences in importance should indeed be revealed by the methods for analyzing agreement among sets.

The question of (un)fairness is not covered completely by Van de Geer, for at least two reasons: First, the proposition of normalization is valid only in the case $r = 1$. In the general case $r > 1$, the question of (un)fairness requires a new definition. Second, the above two possibilities for constraining one’s solution pertain to criterion MAXBET. In fact, however, maximization of MAXBET subject to $t t' = 1$ is equivalent to maximization of MAXBET A subject to $t_i t_i = 1$, $i = 1, \ldots, K$ as proven by Chessel and Hanafi (1996, pp. 42–45), hence in case of criterion MAXBET A one cannot vary Fairness by varying the constraint.

As will be explained now, a different way of dealing with (un)fairness while considering the general case ($r \geq 1$) is by choosing one’s criterion function: MAXBET, MAXBET A, or MAXBET B. Optimization of these criteria subject to $T_i T_i = I_i, i = 1, \ldots, K$ (and the added constraint $a a = 1$ only for the optimization of the criterion MAXBET A) can be written alternatively as maximizing functions involving:

$$\hat{S}_T = [ \text{vec}(X_1 T_1) | \text{vec}(X_2 T_2) | \cdots | \text{vec}(X_K T_K) ],$$

where again vec(.) denotes the vector containing all the elements of the matrix strung out rowwise into a column vector. Together with auxiliary parameters denoting weights attached to the different sets, this can be done as follows. The criterion MAXBET can be written as

$$\sum_{i=1}^{K} \left( (\text{vec}(X_i T_i))' \hat{S}_T 1 \right).$$
where $\mathbf{1}$ is a vector with $K$ unit elements only; the criterion MAXBET A can be written as

$$
\sum_{i=1}^{K} \left( \text{vec}(\mathbf{X}_i \mathbf{T}_i) \mathbf{S}_{Ta} \right)^2.
$$

Finally, maximizing the criterion MAXBET subject to $\mathbf{T}_i' \mathbf{T}_i = \mathbf{I}_r$, $i = 1, \ldots, K$, is equivalent to maximizing the criterion

$$
\sum_{i=1}^{K} \left( \text{vec}(\mathbf{X}_i \mathbf{T}_i) \mathbf{S}_{Ta} \right)^2,
$$

subject to $\mathbf{T}_i' \mathbf{T}_i = \mathbf{I}_r$, $a_i'a_i = 1$, $i = 1, \ldots, K$. All these equivalences are proven in the appendix.

In the criteria MAXBET, MAXBET A and MAXBET B, we encounter a “compromise” vector which is a (weighted) sum of vectors $\text{vec}(\mathbf{X}_i \mathbf{T}_i)$, $i = 1, \ldots, K$, that is, the columns of $\mathbf{S}_{T}$. To study differences on “fairness” of different solutions, we will study to what extent these compromises are based equally on all groups (i.e., with equal weights in forming the compromise from the columns of $\mathbf{S}_{T}$), or unequally (i.e., with different weights for the columns of $\mathbf{S}_{T}$). We see at once that MAXBET involves a fair compromise, with equal weights attached to all columns of $\mathbf{S}_{T}$, whereas in MAXBET A and MAXBET B the compromises are defined by linear combinations $\mathbf{S}_{Ta}, \mathbf{S}_{Ta_i}$, where the weights will, in practice be unequal, thus leading to unfair compromises. In fact, in MAXBET B, $K$ compromises are used, one for each group.

Next, comparing MAXBET A and MAXBET B and considering $\mathbf{a}$ or $\mathbf{a}_i$, fixed, we see that maximization of MAXBET A implies the search of matrices $\mathbf{T}_i$ such that all $\text{vec}(\mathbf{X}_i \mathbf{T}_i)$, $i = 1, \ldots, K$, have maximal cross-products with the single vector $\mathbf{S}_{Ta}$; maximization of MAXBET B, on the other hand, implies the search of matrices $\mathbf{T}_i$ such that each $\text{vec}(\mathbf{X}_i \mathbf{T}_i)$, has a maximal cross-product with the associated vector $\mathbf{S}_{Ta_i}$, for $i = 1, \ldots, K$.

As a consequence, given $\mathbf{a}$ or $\mathbf{a}_i$, $i = 1, \ldots, K$, maximization of MAXBET B will lead to larger differences between $\text{vec}(\mathbf{X}_1 \mathbf{T}_1)$, $\text{vec}(\mathbf{X}_2 \mathbf{T}_2)$, \ldots, $\text{vec}(\mathbf{X}_K \mathbf{T}_K)$ than maximization of MAXBET A. Now, maximization of MAXBET B over each vector $\mathbf{a}_i$ implies the search of a linear combination of the columns of $\mathbf{S}_{T}$, hence of, $\text{vec}(\mathbf{X}_1 \mathbf{T}_1)$, $\text{vec}(\mathbf{X}_2 \mathbf{T}_2)$, \ldots, $\text{vec}(\mathbf{X}_K \mathbf{T}_K)$, that has a maximal cross-product with $\text{vec}(\mathbf{X}_i \mathbf{T}_i)$, $i = 1, \ldots, K$. Thus, it can be expected that, in each vector $\mathbf{a}_i$, the $i$th element, being the weight for $\text{vec}(\mathbf{X}_i \mathbf{T}_i)$, is relatively high. In contrast, maximization of MAXBET A over $\mathbf{a}$, implies the search of a single linear combination of the columns of $\mathbf{S}_{T}$. This vector will not give very different weights to the columns of $\mathbf{S}_{T}$, because it contains the first principal component for the columns of $\mathbf{S}_{T}$, the elements of which have a tendency to be relatively equal, especially, since its columns are relatively similar. Therefore, it can be expected that the different compromises involved in MAXBET B will be unfairer than the single one present in MAXBET A.

It thus follows that the criteria MAXBET, MAXBET A and MAXBET B are ordered such that they lead to increasingly “unfair” solutions. We can state this less evaluative as follows: The criteria MAXBET, MAXBET A and MAXBET B are ordered from emphasis on equal treatment of all sets to emphasis on optimal interest in a limited number of sets with the strongest interrelations as well as within variances, at the cost of relations between other sets.

To illustrate these relations, we return to our example data sets, where again we consider the case $r = 2$, with the constraint $\mathbf{T}_i' \mathbf{T}_i = \mathbf{I}_2$, $i = 1, \ldots, K$. Thus, we optimize criteria MAXBET, MAXBET A and MAXBET B. Optimizing MAXBET leads to a fair compromise. Optimizing MAXBET A and MAXBET B lead to unfair compromises. This is clearly demonstrated by the results of our analyses: Table 5 presents the weights attached to the different sets (i.e., the weights applied to the columns of $\mathbf{S}_{T}$).

It can be seen that, indeed, the weights in the MAXBET A analysis are different and it is hence indeed unfair compared to the maximization of MAXBET (for which the weights are equal by definition). The MAXBET B leads to four different sets of weights, for each different compromise. For these compromises, the weights vary even more than for the compromises obtained with MAXBET A. It can be seen from Table 5 that the weights attached to the set that plays the role of reference set are, not surprisingly, by far the largest, and other weights differ considerably from this as well as from each other. Clearly, these “compromises” are very much dominated by one set (in this case, the reference set or reference assessor).

It can be concluded that the criteria MAXBET, MAXBET A and MAXBET B differ from each other in the way how they, implicitly, define “summarizers” for the components from the different sets. The separate components are
chosen such that they optimally resemble the summarizer, so, for the definition of the components it is crucial how the summarizer is defined. The first method uses the “fairest” summarizers (using the mean of the standardized mean of the components); the following methods use summarizers that are increasingly “unfair”, or stated more neutrally, are increasingly more differentially influenced by the different components. Thus, if it is desired to have a solution in which all sets are treated in a balanced way, one should choose the first criterion. If, on the other hand, there is an interest in focusing on the strongest relationships and, if applicable, within-sets variances, one can take one of the criteria that allows for differential treatment of sets, in varying degrees, as desired. Thus, by choosing one’s criterion one can vary the degree of “fairness” desired. It can be concluded that, thus, in addition to varying fairness by the choice of constraint, one can modify fairness by the choice of the criterion.

5. A general algorithm for maximizing all criteria in the general framework, subject to all constraints

Above, a general framework has been proposed for the analysis of a number of data sets (or their orthogonal basis matrices), consisting of six criteria, MAXBET, MAXDIFF, MAXBET A, MAXDIFF A, MAXBET B and MAXDIFF B, which can be optimized over matrices, $T_i$, $i = 1, 2, \ldots, K$, subject to two kinds of constraints, $T'T = I_r$ or $T'_iT_i = I_r$, $i = 1, 2, \ldots, K$. This thus leads to $2 \times 6 \times 2$ methods. Fortunately, the optimizations involved in all these methods can be performed by a single general algorithm. This general algorithm will be described in the present section. The general form of its description makes clear that its applicability is not limited to the methods considered in the present paper.

The general algorithm to be presented here is meant for the maximization of functions $f(T)$ of a $m \times r$ matrix $T$ (with submatrices $T_i$, $i = 1, 2, \ldots, K$) for which holds

$$f(T) = g(T, T),$$

where function $g$ is defined as

$$g(T, W) = \text{tr} \left( T'A^{(W)}T \right),$$

with $W$ an $m \times r$ matrix, and $A^{(W)}$ is symmetric and positive definite matrix of order $m \times m$, which depends continuously on $W$. In addition, we suppose that the function $g$ satisfies the following inequality:

$$g(T, W) \leq \sqrt{g(T, T)g(W, W)}.$$

We note that in general, the fact that $g$ satisfies (2) does not imply that (3) holds automatically. In fact, inequality (3) is a necessary condition on which our general algorithm is based.

The general algorithm to be proposed here is meant for maximization of $f(T)$ over $T$ subject to the constraint or $T'T = I_r$ or $T'_iT_i = I_r$, $i = 1, 2, \ldots, K$. In the first subsection, we will develop the announced general algorithm for the constrained maximization of functions for which (1)–(3) hold. In the next subsection, we will show that the criteria MAXBET, MAXDIFF, MAXBET A, MAXDIFF A, MAXBET B and MAXDIFF B, proposed above all belong to the class of functions to which our general algorithm applies.
5.1. Monotonically convergent algorithm for maximizing the general function \( f(T) \)

In the present section an algorithm will be proposed for maximizing the general function \( f(T) \) subject to the constraint \( T'T = I_r, i = 1, 2, \ldots, K \). It will first be noted that the problem of maximizing \( f(T) \) subject to \( T'T = I_r \) can be seen as a special case of the problem of maximizing \( f(T) \) subject to the constraints \( T_i'T_i = I_r, i = 1, 2, \ldots, K \), by simply considering this as the case where \( K = 1 \), and hence where the whole matrix equals the one and only submatrix of \( T \), and, similarly, the submatrices of \( A^{(W)} \), which will be used in our algorithm, in this case simply reduce to the one and only submatrix equal to \( A^{(W)} \) itself.

We will now propose a monotonically convergent algorithm for maximizing \( f(T) \) subject to the constraints \( T_i'T_i = I_r, i = 1, 2, \ldots, K \). Here, the algorithm is said to be monotonically convergent if and only if there is a continuous and bounded function \( F \) such that the sequence \( F(Q^{(s)}) \) is monotone, where \( Q^{(s)} \) is a sequence generated by the algorithm.

Many papers have dealt with the problem of finding general principles from which families of algorithms can be constructed (e.g., for multidimensional scaling, see De Leeuw and Heiser, 1980; Meulman, 1986; De Leeuw, 1988; Groenen, 1993; Meyer, 1992; for Weighted Procrustes Analysis, see Kiers, 1990; and Kiers and Ten Berge, 1992; for extensions of correspondence analysis, see Meyer, 1991). These works share the use of a general technique for constructing convergent algorithms called \textit{Iterative Majorization}. \textit{Iterative Majorization} consists of solving simpler subproblems by covering the function to be optimized with a series of simple auxiliary functions according to three principles (detailed in Heiser, 1995). The approach presented in the present paper can be considered as a special case of \textit{Iterative Majorization}.

An iterative algorithm that increases the general function \( f(T) \) monotonically subject to the constraints \( T_i'T_i = I_r, i = 1, 2, \ldots, K \), can be constructed by looking for an update \( Z' = [Z_1' \mid Z_2' \mid \cdots \mid Z_K'] \) of \( \tilde{T}' = [\tilde{T}_1' \mid \tilde{T}_2' \mid \cdots \mid \tilde{T}_K'] \) a current matrix satisfying \( \tilde{T}_i'T_i = I_r, i = 1, 2, \ldots, K \), so that \( f(\tilde{T}') \leq f(Z) \). This can be done by using the solution of the following problem:

Maximize \( \text{tr} \left( T_i'A_i^{(T)}\tilde{T}^t \right) \) subject to the constraint \( T_i'T_i = I_r \), \hspace{1cm} (4)

where \( i \) is fixed, and \( A_i^{(T)} = \left[ A_{i1}^{(T)} \mid A_{i2}^{(T)} \mid \cdots \mid A_{iK}^{(T)} \right] \) a \((m_i, m)\) matrix containing the \( i \)th set of rows of the partitioned matrix

\[
A^{(\tilde{T})} = \begin{bmatrix}
A_{11}^{(\tilde{T})} & A_{12}^{(\tilde{T})} & \cdots & A_{1K}^{(\tilde{T})} \\
A_{21}^{(\tilde{T})} & A_{22}^{(\tilde{T})} & \cdots & A_{2K}^{(\tilde{T})} \\
\vdots & \vdots & \ddots & \vdots \\
A_{K1}^{(\tilde{T})} & A_{K2}^{(\tilde{T})} & \cdots & A_{KK}^{(\tilde{T})}
\end{bmatrix}
\]

The singular value decomposition of the \((m_i, r)\) matrix \( A_i^{(\tilde{T})}\tilde{T} = P_iD_iQ_i^t \), with \( P_i'P_i = I_r, Q_i'Q_i = Q_iQ_i^t = I_r \) and \( D_i \) is a diagonal matrix, nonnegative and with diagonal elements in weakly descending order. Then the solution of problem (4) is the matrix \( Z_i = P_iQ_i^t \) (see Cliff, 1966). As consequence, we have for \( i = 1, 2, \ldots, K \):

\[
\text{tr} \left( T_i'A_i^{(T)}\tilde{T}^t \right) \leq \text{tr} \left( Z_i'A_i^{(T)}\tilde{T}^t \right),
\]

for all \( T_i \) for which \( T_i'T_i = I_r \), and upon summing over \( i \), we have:

\[
\text{tr} \left( T'A^{(T)}\tilde{T}^t \right) \leq \text{tr} \left( Z'A^{(T)}\tilde{T}^t \right),
\]

for all \( T \) satisfying the present constraints \( T_i'T_i = I_r, \hspace{1cm} i = 1, 2, \ldots, K \), hence also for \( T = \tilde{T} \).
Using this and the Cauchy–Schwarz inequality gives:
\[ \text{tr} \left( \tilde{T}' A^{(T)} \tilde{T} \right) \leq \text{tr} \left( Z' A^{(T)} Z \right) \leq \sqrt{\text{tr} \left( Z' A^{(T)} Z \right) \text{tr} \left( T' A^{(T)} \tilde{T} \right) }, \]
from which it follows that:
\[ f (\tilde{T}) \leq g (Z, \tilde{T}). \]

Finally, using inequality (3) we find
\[ f (\tilde{T}) \leq g (Z, \tilde{T}) \leq \sqrt{g(Z, Z)} \sqrt{g(\tilde{T}, \tilde{T})} = \sqrt{f(Z)} \sqrt{f(\tilde{T})}, \]

from which it is immediate that \( f (\tilde{T}) \leq f (Z) \).

Iteratively updating \( T \) as described above, we increase the function \( f (T) \) monotonically. Convergence is attained if and only if the above inequalities become equalities which implies \( T = Z \). Hence, at convergence we have \( T = Z \) which implies
\[ A_i^{(T)} \tilde{T} = P_i D_i Q_i' = \tilde{T}_i Q_i D_i Q_i = TH_i \quad (5) \]
with \( H_i \) symmetric positive semi-definite. Hence Eq. (5) is a necessary condition for a maximum of \( f (T) \). It is noted that the derivation giving the necessary condition (5) is similar to the one in Ten Berge et al. (1988).

The algorithm derived above can be summarized as

Choose \( \tilde{T} = \begin{bmatrix} \tilde{T}_1 & \tilde{T}_2 & \cdots & \tilde{T}_K \end{bmatrix} \) (e.g., randomly, such that, \( \tilde{T}_i \tilde{T}_j = I_r, i = 1, \ldots, K \)), and \( \varepsilon \) (e.g., 0.00001)

While \( (f(Z) - f(\tilde{T})) < \varepsilon \)

Calculate \( A_{\tilde{T}_i} \), make it symmetric and positive definite

For \( i = 1, 2, \ldots, K \),

consider the singular value decomposition: \( A_{\tilde{T}_i} \tilde{T} = PDQ' \)

Set \( Z_i = P_i Q_i' \)

End

Set \( Z = [Z_1' Z_2' \cdots Z_k'] \)

End

5.2. Applying our general algorithm to the optimization of criteria MAXBET, MAXDIFF, MAXBET A, MAXDIFF A, MAXBET B and MAXDIFF B

We will now show that the problems of maximizing criteria MAXBET, MAXDIFF, MAXBET A, MAXDIFF A, MAXBET B and MAXDIFF B, subject to the considered constraints: \( T'T = I_r, \) and \( T_i T_i = I_r \), can all be considered as special cases of the general function \( f (T) \) subject to the constraints \( T_i T_i = I_r \).

In the following we denote, respectively, the criteria MAXBET, MAXDIFF, MAXBET B and MAXDIFF B by \( f_1, f_2, f_3 \) and \( f_4 \). For the criteria MAXBET A and MAXDIFF A, we consider the vector \( a \) fixed and we denote MAXBET A and MAXDIFF A by \( f_5 \) and \( f_6 \), respectively; the problem of finding the optimal \( a \) is considered separately.

Hence, we have to associate to functions \( f_q(T) \) functions \( g_q(T, W), q = 1, 2, \ldots, 6 \) and show that these functions are special cases of the general functions \( g(T, W) \). For \( q = 1, 2, \ldots, 6 \), let be \( g_q(T, W) \) be functions defined as
\[ g_1(T, W) = \sum_{i,j=1}^{K} \text{tr} \left( T_i' X_i' X_j T_j \right), \]
\[ g_2(T, W) = \sum_{i \neq j, j=1}^{K} \text{tr} \left( T_i' X_i' X_j T_j \right), \]
g_3(T, W) = \sum_{i,j=1}^{K} \text{tr}(T'_i X_j X_j T_j) \text{tr}(W'_i X'_i X'_j W_j),

g_4(T, W) = \sum_{i \neq j, j=1}^{K} \text{tr}(T'_i X'_j X_j T_j) \text{tr}(W'_i X'_i X_j W_j),

g_5(T, W) = \sum_{i=1}^{K} \text{tr} \left( T'_i X'_j \sum_{j=1}^{K} x_j X_j T_j \right) \text{tr} \left( W'_i X'_i \sum_{j=1}^{K} x_j X_j W_j \right),

g_6(T, W) = \sum_{i=1}^{K} \text{tr} \left( x_i T'_i T_i + \sum_{i \neq j=1}^{K} x_j T'_i X'_j X_j T_j \right) \text{tr} \left( x_i W'_i W_i + \sum_{i \neq j=1}^{K} x_j W'_i X'_i X_j W_j \right).

If, for q = 1, 2, ..., 6, functions g_q(T, W), q = 1, 2, ..., 6, satisfy properties (1), (2) and (3), then these functions g_q(T, W) can be considered as special cases of g(T, W). Then it follows that functions f_q(T), q = 1, 2, ..., 6, are special cases of the general function f(T). In all six expressions above, property (1) is satisfied immediately upon substituting T for W, yielding, in fact, for q = 1, 2, ..., 6, f_q(T) = g_q(T, T). To demonstrate that functions g_q(T, W), q = 1, 2, ..., 6, verify property (2), we have to find matrices that play the role of A^{(W)} in (2). We will only consider the function g_3(T, W) in detail. The function g_3(T, W) is defined as

\[ g_3(T, W) = \sum_{i,j=1}^{K} \text{tr}(T'_i X_j X_j T_j) \text{tr}(W'_i X'_i X_j W_j), \]

which is equivalent to

\[ g_3(T, W) = \sum_{i,j=1}^{K} \text{tr}(T'_i (\text{tr}(W'_i X'_i X_j W_j) X'_j X_j) T_j). \]

Let A^{(W)}_{g_3} be the super matrix with submatrices defined as \( [A^{(W)}_{g_3}]_{ij} = [c_{ij} X'_i X_j], \) with \( c_{ij} = \text{tr}(W'_i X'_i X_j W_j). \)

Then \( g_3(T, W) \) can be written as

\[ g_3(T, W) = \text{tr} \left( T' A^{(W)}_{g_3} T \right), \]

Hence the matrix defined as A^{(W)}_{g_3} plays the role of A^{(W)} in (2) for the present function. A similar reasoning allows to show that functions g_q(T, W)q = 1, 2, ..., 6 can also be written as

\[ g_q(T, W) = \text{tr} \left( T' A^{(W)}_{g_q} T \right), \]

where the matrices A^{(W)}_{g_q} q = 1, 2, ..., 6 are given as follows:

\[ [A^{(W)}_{g_1}]_{ij} = X'_i X_j, \]

\[ [A^{(W)}_{g_2}]_{ij} = \begin{cases} X'_i X_j, & i \neq j, \\ 0, & i = j, \end{cases} \]

\[ [A^{(W)}_{4}]_{ij} = \begin{cases} \text{tr}(W'_i X'_i X_j W_j) X'_i X_j, & i \neq j, \\ 0, & i = j, \end{cases} \]
\[
[A_{5W}]_{ij} = [z_i \beta_j X_i | X_j], \\
[A_{6W}]_{ij} = \begin{cases} 
[\sigma_{ij}^2 | X_i | X_j], & i \neq j, \\
[\sigma^2_{ii}], & i = j.
\end{cases}
\]

The matrices \(A_{6W}^{(W)}\) for \(q = 1, 2, \ldots, 6\) are in general, asymmetric and are not positive definite. However, the criteria do not change when \(A_{6q}^{(W)}\) for \(q = 1, 2, \ldots, 6\) are replaced by their symmetric parts, and these criteria change only by a constant (and hence the optimization problem does not change) when we add \(\sigma^{(W)} I\) to these matrices. Thus by replacing the matrices by their symmetric part and/or add a matrix \(\sigma^{(W)} I\), with \(\sigma^{(W)}\) chosen large than \(-K\) times the smallest eigenvalue of the symmetric matrix we can assure that the resulting matrix is positive definite. Thus, we have found a positive definite matrix fulfilling the role of \(A^{(W)}\) in (2), as required.

It has thus been shown that the functions \(f_q(T), q = 1, 2, \ldots, 4\), are special cases of the general function \(f(T)\). It follows that for maximization of the criteria MAXBET, MAXDIFF, MAXBET B, MAXDIFF B subject to the constraints \(T^T T = I_r\) or \(T_i^T T_i = I_r\), \(i = 1, 2, \ldots, K\) subject to the constraints \(T^T T = I_r\) and \(a a^T = 1\), an iterative algorithm which increases the two criteria MAXBET A and MAXDIFF A monotonically subject to the constraints \(T^T T = I_r\) or \(T_i^T T_i = I_r\), \(i = 1, 2, \ldots, K\) and \(a a^T = 1\) can easily be constructed by using the general algorithm iteratively. We will only consider maximization of MAXBET A in detail. The problem consists of finding updates \(Z' = [Z'_1 | Z'_2 | \cdots | Z'_K]\) and \(b\) in terms of a current matrix \(\tilde{T} = [\tilde{T}_1 | \tilde{T}_2 | \cdots | \tilde{T}_K]\) and a current vector \(\tilde{a}\) satisfying \(\tilde{T}_i^T \tilde{T}_i = I_r, \ i = 1, 2, \ldots, K\) and \(\tilde{a} \tilde{a}^T = 1\), such that:

\[
\Phi(\tilde{T}, \tilde{a}) \leq \Phi(Z, a),
\]

where \(\Phi\) denotes the criterion MAXBET A.

This problem is solved easily by using our general algorithm because, for \(\tilde{a}\) fixed, the criterion \(\Phi\) is a particular case of the general function \(f(T)\). We denote by \(Z\) the solution obtained after convergence of our general algorithm then the following relation is satisfied:

\[
\Phi(\tilde{T}, \tilde{a}) \leq \Phi(Z, a).
\]

Next, for \(Z\) fixed, we consider the problem:

Maximize \(\Phi(Z, a)\) subject to \(a a^T = 1\).

This problem has as solution \(b\) the first eigenvector of the matrix \(S\) with elements \(s_{ij} = \text{tr}(Z_i^T X_i | X_j)\), because \(\Phi(Z, a) = a^T S^2 a\). Hence \(\Phi(\tilde{T}, \tilde{a}) \leq \Phi(Z, a) \leq \Phi(Z, b)\).

Iteratively updating \(T\) by the general algorithm and updating \(a\) as the eigenvector associated with the largest eigenvalue of the matrix \(S\), we increase the MAXBET A criterion monotonically. A similar argument can be used to increase the criterion MAXDIFF A monotonically where we again iteratively update \(T\) by the general algorithm applied to the criterion MAXDIFF A, and update \(a\) as the eigenvector associated with the largest eigenvalue of the matrix \(W^2\), where \(W\) is defined as

\[
[W]_{ij} = \begin{cases} 
s_{ij}, & i \neq j, \\
\text{tr}(Z_i^T Z_i), & i = j.
\end{cases}
\]

In both cases, rather than finding the optimal \(Z\) in each main iteration, one may just use a single update of \(Z\), then update \(a\), then \(Z\) again, etc.

The general algorithm proposed in this part can be used for obtaining solutions for the class of methods defined by the criteria MAXBET, MAXDIFF, MAXBET A, MAXDIFF A, MAXBET B and MAXDIFF B and the generalized constraints and \(T^T T = I_r\) and \(T_i^T T_i = I_r\). In this way, a unified treatment of all these methods is provided. This unified treatment generalizes the Ten Berge (1988) approach, which dealt only with the criteria MAXBET and MAXDIFF and only with the treatment of the constraint \(T_i^T T_i = I_r\). In addition, the present approach generalizes the approach proposed by Meyer (1989, 1991, 1992). In fact, the criteria considered by Meyer are special cases of the general function \(f(T)\), with the submatrices \(X_i\) replaced by orthonormal basis matrices \(P_i\), and \(r = 1\).
6. Discussion

The class of methods proposed in the first part of this paper offers extensions of Van de Geer’s (1984) and Ten Berge’s (1986, 1988) general classes of methods. Their methods generalize one particular GCA criterion. Here we similarly generalized two other GCA criteria by taking into account the options concerning the three choices: (i) What to Analyze?, (ii) fairness and orthogonality, (iii) variance bias. The procedures can be used in various contexts, like in comparison of sets of variables by (variants of) GCA, or in rotation of factor loading matrices or configurations matrices to maximal agreement. In both cases it is possible to vary the emphasis on within-sets variance and on fairness.

The general algorithm proposed in the third part of the present paper converges monotonically to a stationary point, but convergence to the global optimum is not guaranteed. Therefore, it is recommended to run the algorithm several times for the same data sets, using different initial configurations. The stationary points are characterized by the necessary conditions (5). For most problems, the necessary conditions (5) are new; for the problems treated before by Ten Berge (1988) and Meyer (1991), which have been noted to be special cases of our general class of optimization problems, these stationary equations reduce to the ones given by these authors. Specifically, the stationary equation established by Ten Berge is the special case of (5) in the special case where the matrices \( A^T \) do not depend on \( T \) (hence the matrices \( A^T \) are constant). Similarly, the stationary conditions established by Meyer are special cases of condition (5) with \( r = 1 \). The general algorithm allows us to compute solutions for all methods considered in the first part of the paper. Computation speed of the algorithm is good: All analyses for the examples studied were carried out within a few seconds, and experience with bigger data matrices has shown that computation time does not increase prohibitively.

Ten Berge’s (1988) framework also contained an approach for successively finding subsets of components, given the previous ones. This approach can be followed here for all methods in a completely analogous fashion. On the other hand, Ten Berge (1988) did not handle the case where MAXDIFF is optimized subject to the constraint \( t^T t = 1 \). Clearly, this gap has now been filled.

As a final remark, we wish to note that the general algorithm proposed here is not always the most efficient one. In fact, for various special cases closed-form solutions, based on eigenvectors are available. The general algorithm, on the other hand, does solve various problems for which neither closed-form solution nor other monotonically convergent algorithms seem to be available as yet.

Acknowledgement

The authors are obliged to J.M.F Ten Berge for critical comments on a previous draft.

Appendix

Some equivalences used in Section 4.2 will be proven. For later use we define \( v_i \equiv \hat{S}_T (\text{vec}(X_i T_i)) \), hence

\[
v_i' = [\text{tr} (T_i' X_i' X_i T_1) \quad \text{tr} (T_i' X_i' X_2 T_2) \quad \ldots \quad \text{tr} (T_i' X_i' X_K T_K)].
\]

We consider the criteria

\[
\hat{\Psi} (T_1, T_2, \ldots, T_K) = \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \hat{S}_T 1 \right),
\]

\[
\hat{\Phi} (T_1, T_2, \ldots, T_K, a) = \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \hat{S}_T a \right)^2,
\]

\[
\hat{\Theta} (T_1, T_2, \ldots, T_K, a_1, a_2, \ldots, a_K) = \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \hat{S}_T a_i \right)^2.
\]

In the present appendix, the aim is to prove the following equivalencies:

1. Maximizing MAXBET subject to \( T_i' T_i = 1_i \), \( i = 1, 2, \ldots, K \), is equivalent to maximizing \( \hat{\Psi} \) subject to \( T_i' T_i = 1_r \), \( i = 1, 2, \ldots, K \).
(2) Maximizing \( \text{MAXBET A} \) subject to \( T_i' T_i = I_r, i = 1, 2, \ldots, K \) is equivalent to maximizing \( \hat{\Phi} \) subject to \( a' a = 1, T_i' T_i = I_r, i = 1, 2, \ldots, K \).

(3) Maximizing \( \text{MAXBET B} \) subject to \( T_i' T_i = I_r, i = 1, 2, \ldots, K \) is equivalent to maximizing \( \hat{\Theta} \) subject to \( T_i' T_i = I_r, a_i' a_i = 1, i = 1, 2, \ldots, K \).

Proof of equivalence (1). The criterion \( \hat{\Psi} \) can be written as

\[
\hat{\Psi} (T_1, T_2, \ldots, T_K) = \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \hat{\Sigma}_T 1 \right) = \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \left( \sum_{j=1}^{K} \text{vec} (X_j T_j) \right) \right) 
\]

\[
= \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \left( \sum_{j=1}^{K} \text{vec} (X_j T_j) \right) \right) 
\]

\[
= \sum_{i=1}^{K} \sum_{j=1}^{K} \text{tr} \left( T_i' X_i' X_j T_j \right).
\]

Proof of equivalence (2). The criterion \( \hat{\Phi} \) can be written as

\[
\hat{\Phi} (T_1, T_2, \ldots, T_K, a) = \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \hat{\Sigma}_T a \right)^2 
\]

\[
= \sum_{i=1}^{K} \left( (\text{vec} (X_i T_i))' \left( \sum_{j=1}^{K} \alpha_j \text{vec} (X_j T_j) \right) \right)^2 
\]

\[
= \sum_{i=1}^{K} \left( \sum_{j=1}^{K} \alpha_j \text{tr} \left( T_i' X_i' X_j T_j \right) \right)^2.
\]

Proof of equivalence (3). The criterion \( \hat{\Theta} \) can be written as

\[
\hat{\Theta} (T_1, T_2, \ldots, T_K, a_1, a_2, \ldots, a_K) = \sum_{i=1}^{K} \left( \left( \hat{\Sigma}_T \text{vec} (X_i T_i) \right)' a_i \right)^2 = \sum_{i=1}^{K} (v_i' a_i)^2.
\]

The solution of maximizing \( (v_i' a_i)^2 \) subject to \( a_i' a_i = 1 \) is the vector defined as

\[
a_i = \frac{v_i}{\sqrt{v_i' v_i}}.
\]

This implies that

\[
\hat{\Theta} \left( T_1, T_2, \ldots, T_K, \frac{v_1}{\sqrt{v_1' v_1}}, \frac{v_2}{\sqrt{v_2' v_2}}, \ldots, \frac{v_K}{\sqrt{v_K' v_K}} \right) = \sum_{i=1}^{K} v_i' v_i.
\]

Upon noting that

\[
v_i' v_i = \sum_{j=1}^{K} \left( \text{tr} \left( T_i' X_i X_j T_j \right) \right)^2,
\]

\[
\sum_{i=1}^{K} v_i' v_i = \sum_{i=1}^{K} \left( \sum_{j=1}^{K} \left( \text{tr} \left( T_i' X_i X_j T_j \right) \right)^2 \right).
\]
we find

\[ \hat{\Theta} \left( T_1, T_2, \ldots, T_K, \frac{v_1}{\sqrt{v_1}v_1}, \frac{v_2}{\sqrt{v_2}v_2}, \ldots, \frac{v_K}{\sqrt{v_K}v_K} \right) = \sum_{i,j=1}^{K} (\text{tr} (T_i'X_jX_j'X_i))^2. \]

Maximizing MAXBET B subject to $T_i'T_i = I_r, \quad i = 1, 2, \ldots, K$, is equivalent to maximizing $\hat{\Theta}$ subject to $T_i'T_i = I_r, a_i'a_i = 1, \quad i = 1, 2, \ldots, K$.

References


