LOWER BOUNDS FOR CANONICAL CORRELATIONS OR REDUNDANCIES DERIVED FROM AN OVERALS SOLUTION

John P. Van de Geer

Rijksuniversiteit Leiden
Vakgroep Datatheorie
Middelstegracht 4
2312 TW Leiden

INDEX

ABSTRACT

An OVERALS solution for K sets of variables does not directly reveal what the canonical correlation would be if those variables were divided into two sets, by taking one of the K sets as one set, opposed to a second set containing the variables in all other $(K - 1)$ sets.

However, it is possible to derive from the OVERALS solution a lower bound for the canonical correlation.

The same reasoning applies for a redundancy solution based on one of the K sets, and contrasted with all other $(K - 1)$ sets taken together.

1 INTRODUCTION

OVERALS is a method of data analysis that applies to variables divided into K sets of variables. The result of an OVERALS analysis gives linear compounds of the variables of each set, in such a way that these K linear compounds correlate as much as possible with their average vector.

Users of OVERALS often are inclined to interpret OVERALS results as if they answer the question whether one particular set is related to all other sets taken together. The proper answer to such a question would be to divide the variables not into K sets, but into just 2 sets. Further analysis of relations between those two sets then could be a canonical correlation analysis, or a redundancy analysis.

In Section 2 a more precise definition is given of canonical analysis (CA), and of redundancy analysis (RA). Section 3 introduces a simplification, and Section 4 shows expressions for CA and RA based on that simplification. Section 5 shows when CA and RA produce identical solutions. Section 6 shows what the CA criterion will be when one takes the RA solution, and *vice versa*. Section 7 gives details of the OVERALS solution. Section 8 shows how lower bounds for CA and RA can be developed from the OVERALS solution. Section 9 is a short comment on lower bounds derived from a more-dimensional OVERALS solution, and Section 10 is an, also short, comment on lower bounds derived from a "non-linear" OVERALS solution.

2 DEFINITIONS OF METHODS

2.1 Notation

In the sequel we use the abbreviation SSQ, which for any arbitrary matrix $\mathbf{Y}$ with n rows is defined by

$$\mathrm{SSQ}(\mathbf{Y}) = \mathbf{Y}'\mathbf{Y}/n.$$

Furthermore, we shall have to do with data collected in a matrix $\mathbf{X}$, with n rows (n = number of objects), and as many columns as there are variables. The assumption will be that columns of $\mathbf{X}$ are standardized (i.e., have zero mean, and have sum of squared elements equal to n, so that each column has variance 1). It then follows that

$$\mathrm{SSQ}(\mathbf{X}) = \mathbf{X}'\mathbf{X}/n = \mathbf{R},$$

where $\mathbf{R}$ is a correlation matrix.

Also, $\mathbf{X}$ will be partitioned into K sets:

$$\mathbf{X} = (\mathbf{X}_1,...,\mathbf{X}_k,...,\mathbf{X}_K),$$

where k is a running index to indicate some particular set ($k = 1,...,K$).

2.2 Canonical analysis CA

In CA there are only K=2 sets $\mathbf{X}_1$ and $\mathbf{X}_2$. The objective of CA is to identify weights $\mathbf{v}_1$ and $\mathbf{v}_2$, in such a way that the correlation between the weighted sums $\mathbf{X}_1\mathbf{v}_1$ and $\mathbf{X}_2\mathbf{v}_2$ is maximized. This correlation is called a *canonical correlation*, and will be denoted by the symbol ω.

A solution in more than one dimension can be obtained. In that case CA maximizes ω_s, the canonical correlation on dimension s, under the condition that $\mathbf{X}_1\mathbf{v}_{1s}$ and $\mathbf{X}_2\mathbf{v}_{2s}$ are uncorrelated with all preceding solutions $\mathbf{X}_1\mathbf{v}_{1r}$ and $\mathbf{X}_2\mathbf{v}_{2r}$ ($r<s$).

2.3 Redundancy analysis RA

Given again that there are two sets $\mathbf{X}_1$ and $\mathbf{X}_2$, the RA solution requires weights $\mathbf{v}_2$ such that $\mathbf{X}_2\mathbf{v}_2$ "explains as much as possible of the variance of $\mathbf{X}_1$".

The last expression means that $\mathbf{v}_2$ must be chosen in such a way that the correlations between $\mathbf{X}_2\mathbf{v}_2$ and $\mathbf{X}_1$ have a maximum sum of squares.

Assuming that weights $\mathbf{v}_2$ are chosen in such a way that $\mathrm{SSQ}(\mathbf{X}_2\mathbf{v}_2)=1$, the correlation between $\mathbf{X}_1$ and $\mathbf{X}_2\mathbf{v}_2$ are given in the column

$$\mathbf{X}_1'\mathbf{X}_2\mathbf{v}_2/n,$$

with sum of squares equal to

$$\mathbf{v}_2'\mathbf{X}_2'\mathbf{X}_1\mathbf{X}_1'\mathbf{X}_2\mathbf{v}_2/n^2 = \delta^2,$$

so that the objective of RA is to maximize δ^2, called "redundancy".

Solutions in more than one dimension can be defined by requiring that δ_s^2 must be maximized under the condition that $\mathbf{X}_2\mathbf{v}_{2s}$ is uncorrelated with all preceding solutions $\mathbf{X}_2\mathbf{v}_{2r}$ $(r<s)$.

2.4 OVERALS

We will come back to OVERALS in Section 7. As a short introduction it can be said that, given K sets $\mathbf{X}_k$ $(k=1,..,K)$ the objective of OVERALS is to find weights $\mathbf{v}_k$, in order to form K weighted sumvectors $\mathbf{X}_k\mathbf{v}_k$. Weights $\mathbf{v}_k$ should be chosen in such a way that the K vectors $\mathbf{X}_k\mathbf{v}_k$ have a maximum sum of squared correlations with their average vector $\mathbf{Xv}/K = \Sigma\mathbf{X}_k\mathbf{v}_k/K$.

Subsequent dimensions of the OVERALS solution are defined by maximizing the same criterion, under the condition that $\mathbf{Xv}_s$ is uncorrelated with $\mathbf{Xv}_r$ $(r<s)$.

3 SIMPLIFICATION OF METHODS

The three methods described in Section 2 can be greatly simplified by using for each matrix $\mathbf{X}_k$ a *singular value decomposition* (SVD).

Let $\mathbf{X}_k$ be a matrix with n rows, m_k columns, and rank r_k. The SVD of X_k then is defined by

$$X_k = \boldsymbol{P}_k\boldsymbol{\Phi}_k\mathbf{Q}_k',$$

where

$\mathbf{P}_k$ is an n x r_k matrix that satisfies $\mathbf{P'P}/n = \mathbf{I}$;

$\mathbf{Q}_k$ is an m_k x r_k matrix that satisfies $\mathbf{Q}_k'\mathbf{Q}_k = \mathrm{I}$;

$\boldsymbol{\Phi}_k$ is a diagonal r_k x r_k matrix, with non-zero diagonal elements, in descending order.

The basic reason for this simplification is that a weighted sum $\mathbf{X}_k\mathbf{v}_k$ also can be written as

$$\mathbf{X}_k\mathbf{v}_k = \mathbf{P}_k\boldsymbol{\Phi}_k\mathbf{Q}_k'\mathbf{v}_k = \mathbf{P}_k\mathbf{t}_k,$$

with

$$\mathbf{t}_k = \Phi_k \mathbf{Q}_k' \mathbf{v}_k.$$

In other words, a weighted sum $\mathbf{X}_k \mathbf{v}_k$ is always equivalent to a weighted sum $\mathbf{P}_k \mathbf{t}_k$.

Cautionary note: we shall also use the expression "SVD" for a decomposition of other matrices, in which the number of columns is not equal to the number of objects n. An example would be the matrix $\mathbf{X}_1'\mathbf{X}_2/n$. In those cases the SVD solution has the same definition as above, with the difference that **P** now stands for a matrix that satisfies $\mathbf{P}'\mathbf{P} = \mathbf{I}$.

4 RE-DEFINITION OF CA AND RA

4.1 Canonical analysis CA

The simplification suggested in Section 3 makes it possible to re-define CA by taking the SVD solution of

$$\mathbf{P}_1'\mathbf{P}_2/n = \mathbf{T}_1 \Omega \mathbf{T}_2'$$

with $\mathbf{T}_1'\mathbf{T}_1 = \mathbf{I}$, and $\mathbf{T}_2'\mathbf{T}_2 = \mathbf{I}$, whereas Ω is a diagonal matrix of canonical correlations. In fact, we then have the result that

$$\mathrm{SSQ}(\mathbf{P}_1\mathbf{T}_1) = \mathbf{I}$$

$$\mathrm{SSQ}(\mathbf{P}_2\mathbf{T}_2) = \mathbf{I}$$

whereas correlations between $\mathbf{P}_1\mathbf{T}_1$ and $\mathbf{P}_2\mathbf{T}_2$ are given by

$$\mathbf{T}_1'\mathbf{P}_1'\mathbf{P}_2\mathbf{T}_2/n = \Omega.$$

4.2 Redundancy analysis RA

Similarly, RA is simplified to solving for the SVD
$\mathbf{\Phi}_1\mathbf{P}_1'\mathbf{P}_2/n = \mathbf{S}_1\Delta\mathbf{S}_2'$
where $\mathbf{S}_1'\mathbf{S}_1 = \mathbf{I}$ and $\mathbf{S}_2'\mathbf{S}_2 = \mathbf{I}$, and where Δ is a diagonal matrix with diagonal elements equal to the square roots of the "redundancies".

To show why this result corresponds to the RA, take the correlations between $\mathbf{X}_1$ and $\mathbf{P}_2\mathbf{S}_2$. They are given in

$$\mathbf{X}_1'\mathbf{P}_2\mathbf{S}_2/n = \mathbf{Q}_1\mathbf{\Phi}_1\mathbf{P}_1'\mathbf{P}_2\mathbf{S}_2/n = \mathbf{Q}_1\mathbf{S}_2\Delta,$$

with sums of squares on the diagonal of

$$\Delta\mathbf{S}_2'\mathbf{Q}_1'\mathbf{Q}_1\mathbf{S}_2\Delta = \Delta^2.$$

It must be pointed out that the RA solution is not symmetric. Reversion of roles of $\mathbf{X}_1$ and $\mathbf{X}_2$ leads to a different SVD solution: the SVD of $\mathbf{P}_1'\mathbf{P}_2\mathbf{\Phi}_2$, with different results for weights and redundancies.

4.3 Symmetry of CA versus asymmetry of RA

The symmetry of the CA solution also can be shown in a different way. Take the projection of $\mathbf{P}_1\mathbf{t}_1$ on $\mathbf{P}_2$. It becomes

$$\mathbf{P}_2\mathbf{P}_2'\mathbf{P}_1\mathbf{t}_1/n = \mathbf{P}_2\mathbf{t}_2\omega$$

and this projection therefore is proportional to $\mathbf{P}_2\mathbf{t}_2$.
Conversely, take the projection of $\mathbf{P}_2\mathbf{t}_2$ on $\mathbf{P}_1$. It is:

$$\mathbf{P}_1\mathbf{P}_1'\mathbf{P}_2\mathbf{t}_2/n = \mathbf{P}_1\mathbf{t}_1\omega.$$

So the projection is "back and forth": the projection of $\mathbf{P}_1\mathbf{t}_1$ is proportional to $\mathbf{P}_2\mathbf{t}_2$, whereas the projection of $\mathbf{P}_2\mathbf{t}_2$ is proportional to $\mathbf{P}_1\mathbf{t}_1$.
On the other hand, take the RA solution. The projection of $\mathbf{P}_1\mathbf{\Phi}_1\mathbf{s}_1$ on $\mathbf{P}_2$ becomes

$$\mathbf{P}_2\mathbf{P}_2'\mathbf{P}_1\mathbf{\Phi}_1\mathbf{s}_1/n = \mathbf{P}_2\mathbf{s}_2\delta$$

and therefore is proportional to $\mathbf{P}_2\mathbf{s}_2$. But the projection of $\mathbf{P}_2\mathbf{s}_2$ on $\mathbf{P}_1\mathbf{\Phi}_1$ is

$$\mathbf{P}_1\mathbf{\Phi}_1\mathbf{\Phi}_1^{-2}\mathbf{\Phi}_1\mathbf{P}_1'\mathbf{P}_2\mathbf{s}_2/n = \mathbf{P}_1\mathbf{\Phi}_1^{-1}\mathbf{s}_1\delta$$

and therefore is not proportional to $\mathbf{P}_1\mathbf{\Phi}_1\mathbf{s}_1$.

In fact, RA primarily asks for solutions $\mathbf{S}_2$ (such that columns of $\mathbf{P}_2\mathbf{S}_2$ successively explain as much of the variance of $\mathbf{X}_1$ as possible). But the SVD solution above also gives a matrix $\mathbf{S}_1$, although the RA criterion never asked for it. Still, solutions $\mathbf{P}_1\mathbf{\Phi}_1\mathbf{S}_1$ now are revealed as the linear compounds which project on $\mathbf{P}_2$ as the vectors $\mathbf{P}_2\mathbf{S}_2\Delta$, proportional to $\mathbf{P}_2\mathbf{S}_2$.

5 WHEN ARE SOLUTIONS FOR CA AND RA IDENTICAL?

5.1 Trivial identity

Solutions for CA and RA become the same, obviously, if $\mathbf{\Phi}_1 = \mathbf{I}$. This will happen if all columns of $\mathbf{X}_1$ are uncorrelated with each other; $\mathbf{X}_1'\mathbf{X}_1/n = \mathbf{I}$.

5.2 Non-trivial identity

A non-trivial identity occurs when the solution for $\mathbf{s}_{2g}$ (solution for $\mathbf{s}_2$ on dimension g of RA) is the same as $\mathbf{t}_{2h}$ (solution for $\mathbf{t}_2$ on dimension h of CA). The possibility that $g = h$ is included. It then follows that $\mathbf{P}_1\mathbf{\Phi}_1\mathbf{s}_{1g}$ and $\mathbf{P}_1\mathbf{t}_{1h}$ have projections on $\mathbf{P}_2$ which are proportional. These projections are $\mathbf{P}_2\mathbf{s}_{2g}\delta_g$ and $\mathbf{P}_2\mathbf{t}_{2h}\omega_h$, respectively. But this also implies that the vectors $\mathbf{P}_1\mathbf{\Phi}_1\mathbf{s}_{1g}$ and $\mathbf{P}_1\mathbf{t}_{1h}$ are proportional, from which it follows that

$$\mathbf{\Phi}_1\mathbf{s}_{1g} = \mathbf{t}_{1h}.$$

Assuming that the diagonal elements in $\mathbf{\Phi}_1$ are different, such a proportionality is possible only if all elements in $\mathbf{s}_{1g}$ are zero's, except one element, on position j of $\mathbf{s}_{1g}$ and $\mathbf{t}_{1h}$. The latter element then must be an element 1 (because $\mathbf{s}_{1g}$ and $\mathbf{t}_{1h}$ have elements with sum of squares equal to 1).

In words, the g^{th} RA solution is identical to the h^{th} CA solution, if it happens that $\mathbf{P}_1\mathbf{t}_{1h}$ coincides with some principal component (the j^{th} principal component) of $\mathbf{X}_1$. Such coincidence may happen more than once; we then have more than one solution of CA identical to the RA solution. Such coincidence might happen for all dimensions. This would mean that $\mathbf{S}_1$ and $\mathbf{T}_1$ contain permutations of the columns of the identity matrix **I**, but not necessarily the same permutations. It also might happen that $\mathbf{S}_1$ and $\mathbf{T}_1$ are the same permutations of **I**, so that $\mathbf{S}_1 = \mathbf{T}_1$. It even might happen that $\mathbf{S}_1 = \mathbf{T}_1 = \mathbf{I}$; so that the first CA solution is the same as the first RA solution, the second CA solution the same as the second RA solution, and so on.

We thus have quite a range of possibilities for identity between CA and RA. The possibilities listed above are in increasing order of identity. At the same time, such possible identities become increasingly implausible for empirical data.

6 COMPARISON OF CA AND RA BY INTERCHANGING CRITERIA

6.1 Redundancy produced by CA

Given the CA solution, how much of the variance of $\mathbf{X}_1$ is explained by $\mathbf{P}_2\mathbf{t}_2$? Correlations between $\mathbf{X}_1$ and $\mathbf{P}_2\mathbf{t}_2$ are given in the column

$$\mathbf{X}_1'\mathbf{P}_2\mathbf{t}_2/n = \mathbf{Q}_1\mathbf{\Phi}_1\mathbf{P}_1'\mathbf{P}_2\mathbf{t}_2/n = \mathbf{Q}_1\mathbf{\Phi}_1\mathbf{t}_1\omega$$

with sum of squares

$$\mathbf{t}_1'\mathbf{\Phi}_1\mathbf{Q}_1'\mathbf{Q}_1\mathbf{\Phi}_1\mathbf{t}_1\omega^2 = \mathbf{t}_1'\mathbf{\Phi}_1^2\mathbf{t}_1\omega^2.$$

One can show that the latter result is equal to the amount of variance of $\mathbf{X}_1$, explained by $\mathbf{P}_1\mathbf{t}_1$, multiplied by the squared cosine of the angle between $\mathbf{P}_1\mathbf{t}_1$ and $\mathbf{P}_2\mathbf{t}_2$. The cosine of this angle is equal to the canonical correlation. Correlations between $\mathbf{X}_1$ and $\mathbf{P}_1\mathbf{t}_1$ are:

$$\mathbf{X}_1'\mathbf{P}_1\mathbf{t}_1/n = \mathbf{Q}_1\mathbf{\Phi}_1\mathbf{P}_1'\mathbf{P}_1\mathbf{t}_1/n = \mathbf{Q}_1\mathbf{\Phi}_1\mathbf{t}_1,$$

with sum of squares

$$\mathbf{t}_1'\mathbf{\Phi}_1\mathbf{Q}_1'\mathbf{Q}_1\mathbf{\Phi}_1\mathbf{t}_1 = \mathbf{t}_1'\mathbf{\Phi}_1^2\mathbf{t}_1,$$

so that variance of $\mathbf{X}_1$ explained by $\mathbf{P}_2\mathbf{t}_2$ becomes

$$\mathbf{t}_1'\Phi_1^2\mathbf{t}_1\omega^2.$$

Note that a good CA solution does not imply that $\mathbf{t}_1'\Phi_1^2\mathbf{t}_1$ must be large. In fact, if it would happen that $\mathbf{P}_1\mathbf{t}_1$ coincides with the last principal component of $\mathbf{X}_1$ (with smallest value of Φ_1^2), the amount of variance of $\mathbf{X}_1$ explained by $\mathbf{P}_1\mathbf{t}_1$ will be minimal, although $\mathbf{P}_1\mathbf{t}_1$ would be the best canonical solution.

6.2 CA solution produced by RA

Conversely, the cosine of the angle between $\mathbf{P}_1\Phi_1 s_1$ and its projection $\mathbf{P}_2 s_2\delta$ cannot be larger than the best canonical correlation. The correlation (cosine of the angle) is

$$s_1'\Phi_1\mathbf{P}_1'\mathbf{P}_2 s_2/n(s_1'\Phi_1^2 s_1)^{1/2} = \delta/(s_1'\Phi_1^2 s_1)^{1/2},$$

which therefore gives a lower bound for ω.

However, the angle between $\mathbf{P}_2 s_2$ and its projection $\mathbf{P}_1\Phi_1^{-1}s_1$ will be smaller, and therefore gives a better lower bound.

The correlation between $\mathbf{P}_2 s_2$ and $\mathbf{P}_1\Phi_1^{-1}s_1$ is

$$s_2'\mathbf{P}_2'\mathbf{P}_1\Phi_1^{-1}s_1/n(s_1'\Phi_1^{-2}s_1)^{1/2} = \delta(s_1'\Phi_1^{-2}s_1)^{1/2}$$

and this correlation is larger than the correlation between $\mathbf{P}_2 s_2$ and $\mathbf{P}_1\Phi_1 s_1$ (unless it happens that the CA and RA solutions are identical, in which case the two correlations are both equal to ω).

(*Proof.* The correlation between $\mathbf{P}_1\Phi_1 s_1$ and $\mathbf{P}_1\Phi^{-1}s_1$ is

$$(s_1'\Phi_1^2 s_1)^{-1/2}s_1'\Phi_1\mathbf{P}_1'\mathbf{P}_1\Phi_1^{-1}s_1(s_1'\Phi_1^{-2}s_1)^{-1/2} =$$

$$(s_1'\Phi_1^2 s_1)^{-1/2}(s_1'\Phi_1^{-2}s_1)^{-1/2}.$$

The squared correlation can never be larger than 1, and therefore:

$$(s_1'\Phi_1^2 s_1)(s_1'\Phi_1^{-2}s_1) > 1.)$$

7 OVERALS

This section is restricted to a *numerical* OVERALS solution, based on the given quantification of the data matrix **X**. (Non-numerical OVERALS will be discussed in Section 10.)

Using the simplification introduced in Section 3, the numerical OVERALS solution is based on the matrix

$$\mathbf{P} = (\mathbf{P}_1,\ldots,\mathbf{P}_k,\ldots,\mathbf{P}_K),$$

where k is used as a running index for sets: $k = 1,\ldots,K$. The OVERALS solution is defined by the stationary equations

$$\mathbf{P}_k'\mathbf{P}\mathbf{t}/n = \mathbf{t}_k\gamma K.$$

These equatations can be summarized as

$$\mathbf{P}'\mathbf{P}\mathbf{t}/n = \mathbf{t}\gamma K,$$

which shows that $\mathbf{t}$ is an eigenvector of $\mathbf{P}'\mathbf{P}/n$, and that γK is the corresponding eigenvalue. For normalization purposes, vector $\mathbf{t}$ will be normalized in such a way that

$$\mathbf{t}'\mathbf{P}'\mathbf{P}\mathbf{t}/n = \mathbf{t}'\mathbf{t}\gamma K = \gamma^2K^2$$

so that $\mathbf{t}'\mathbf{t} = \gamma K$.

Correlations between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{P}\mathbf{t}$ then are found to be

$$\begin{aligned}(\mathbf{t}_k'\mathbf{P}_k'\mathbf{P}_k\mathbf{t}_k)^{-1/2}\mathbf{t}_k'\mathbf{P}_k'\mathbf{P}\mathbf{t}(\mathbf{t}'\mathbf{P}'\mathbf{P}\mathbf{t})^{-1/2} &= (\mathbf{t}_k'\mathbf{t}_k)^{-1/2}\mathbf{t}_k'\mathbf{t}_k\gamma K(\gamma^2K^2)^{-1/2} \\ &= (\mathbf{t}_k'\mathbf{t}_k)^{1/2} = \mathrm{a}_k\end{aligned}$$

so that $\Sigma \mathrm{a}_k^2 = \Sigma(\mathbf{t}_k'\mathbf{t}_k) = \mathbf{t}'\mathbf{t} = \gamma K$. The value a_k^2 is in OVERALS terminology called the discrimination of the solution for set k. The last equation shows that the eigenvalue is the average discrimination. The discrimination a_k^2 also is equal to the variance of $\mathbf{P}_k\mathbf{t}_k$. The objective of OVERALS is to determine weights $\mathbf{t}$ in such a way that the average discrimination is maximized. (Subsequent solutions have the same criterion, under the condition that subsequent solutions for $\mathbf{P}\mathbf{t}$ must be uncorrelated with earlier solutions for $\mathbf{P}\mathbf{t}$.)

The value a_k corresponds to the cosine of the angle between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{Pt}$. This angle will be indicated as α_k, so that

$$a_k = \cos \alpha_k.$$

8 OVERALS RELATED TO CA AND RA

8.1 Introduction

OVERALS is sometimes applied by a user who has the general idea that the OVERALS solution will reveal to what extent some set $\mathbf{X}_k$ is related to all other (K-1) sets. This general idea is not wrong, but it is rather unspecific. The user may have in mind that the OVERALS solution answers the question: what is the (best) canonical correlation between set $\mathbf{X}_k$ and all other sets? Or: how much of the variance of all other sets is explained by $\mathbf{X}_k$? Or: how much of the variance of $\mathbf{X}_k$ is explained by all other sets?

If a user is really interested in such specific questions, the appropriate answer will be found by applying CA or RA, after a partitioning of $\mathbf{X}$ in only two sets: $\mathbf{X}_k$ as one set, and all other (K-1) sets taken together as the second set.

The OVERALS solution does not give specific answers to the questions raised above. However, the OVERALS solution provides *lower bounds* for CA and RA solutions. Such lower bounds will be developed in the following paragraphs of this section, on the assumption that only the first (the best) OVERALS solution will be considered. We shall therefore omit the index for the dimension of the solution.

8.2 OVERALS and CA

OVERALS yields a_k^2; the squared correlation between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{Pt}$. The value of a_k^2 does not give a lower bound for the squared canonical correlation, the reason being that $\mathbf{Pt}$ is a weighted sum of the variables in all K sets, and not a weighted sum of the variables in sets other than $\mathbf{X}_k$. This could be remedied by taking

$$\mathbf{v}_k = \mathbf{Pt} - \mathbf{P}_k\mathbf{t}_k,$$

which is a weighted sum of variables not in set $\mathbf{X}_k$. It then follows that the correlation between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{v}_k$ must be a lower bound for the canonical correlation.

We shall indicate the angle between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{v}_k$ by the symbol ζ_k. It then follows that

$$\omega^2 \geq \cos^2\zeta_k.$$

Figure 1 shows the relations between $\mathbf{Pt}$, $\mathbf{P}_k\mathbf{t}_k$, and $\mathbf{v}_k$ for the situation where $\gamma K \geq 1$. Obviously, the two vectors $\mathbf{Pt}$ and $\mathbf{P}_k\mathbf{t}_k$ are located in a plane. Their difference vector $\mathbf{v}_k = \mathbf{Pt} - \mathbf{P}_k\mathbf{t}_k$ therefore is also in the same plane. Vector $\mathbf{P}_k\mathbf{t}_k$ has length a_k, whereas $\mathbf{Pt}$ has length γK. The projection of $\mathbf{Pt}$ on $\mathbf{P}_k\mathbf{t}_k$ is $\mathbf{P}_k\mathbf{t}_k\gamma K$, with length $a_k\gamma K$. The projection of $\mathbf{v}_k$ on $\mathbf{P}_k\mathbf{t}_k$ is $\mathbf{P}_k\mathbf{t}_k(\gamma K\text{-}1)$, with length $a_k(\gamma K\text{-}1)$. It follows from Figure 1 that

$$\text{cotg } \zeta_k = (\text{cotg } \alpha_k)(\gamma K - 1)/\gamma K.$$

This equation can be converted into an expression of $(\cos \zeta_k)$ in terms of γK and $(\cos \alpha_k)$. However, such an expression becomes more complicated, whereas the relation between the two cotangents is quite simple.

Figure 2 shows a similar figure as Figure 1, but now for the situation that $\gamma K < 1$. It illustrates that in such a situation a small value of $(\cos \alpha_k)$ may produce a relatively large lower bound $(\cos \zeta_k)$.

The latter results perhaps need some clarification. Suppose we have the extreme situation that $\gamma = 0$. This means that $\mathbf{Pt}$ is a zero vector. Assuming that the weights $\mathbf{t}_k$ are not all zero's, it then follows that

$$\mathbf{v}_k = \mathbf{0} - \mathbf{P}_k\mathbf{t}_k = -\mathbf{P}_k\mathbf{t}_k,$$

and this implies a correlation of -1 between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{v}_k$. But it also implies that the correlation between $\mathbf{P}_k\mathbf{t}_k$ and $-\mathbf{v}_k$ is unity, so that there must be a canonical correlation of unity. In this extreme situation, with zero eigenvalue, the correlation between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{Pt}$ becomes 0/0 and is undefined. And therefore it will no longer be valid that $\mathbf{t}_k'\mathbf{t}_k = a_k^2$ represents such a correlation. However, if γ is close to zero, Figure 2 remains valid, and it then also remains valid that $\gamma K = \Sigma a_k^2$. This implies that a_k^2 never can be larger than γK. In the limiting case, where $a_k^2 = \gamma K$, the result will be that $\cos^2\zeta_k = (1 - \gamma K)$.

Table 1 shows results for $(\cos \zeta_k)$, given $a_k = (\cos \alpha_k)$, and some selected values of γK. We have the following comments.

(a) When $\gamma K = 1$, all lower bounds are zero. Note that a value $\gamma K = 1$ will be found if all columns of $\mathbf{P}$ are uncorrelated.

(b) When $\gamma K = .5$, the values of $(\cos \zeta_k)$ are equal to $(\cos \alpha_k)$.

(c) When $\gamma K > .5$, the lower bounds ($\cos \zeta_k$) are smaller than $a_k = (\cos \alpha_k)$.

(d) When $\gamma K < .5$, the lower bounds are larger than a_k.

(e) When $\gamma K > .5$, lower bounds are identical with those of $(\gamma K)/(2\gamma K - 1)$.

(f) Columns of Table 1 are interrupted by horizontal dotted lines which indicate that values of ($\cos \zeta_k$) and their corresponding ($\cos \alpha_k$) are impossible for values of $\gamma K < 1$, because it always must be true that
$(\cos^2 \alpha_k) \leq \gamma K$.

But this is a restriction only if $\gamma K < 1$.

8.3 OVERALS and RA

The amount of variance of $\mathbf{X}_k$ explained by $\mathbf{v}_k$ will be a lower bound of the redundancy in the RA solution. The projection of $\mathbf{v}_k$ on $\mathbf{P}_k\mathbf{t}_k$ is proportional to $\mathbf{P}_k\mathbf{t}_k$ itself. It follows that the amount of variance of $\mathbf{X}_k$ explained by $\mathbf{v}_k$ must be equal to the amount of variance of $\mathbf{X}_k$ explained by $\mathbf{P}_k\mathbf{t}_k$, multiplied by the squared cosine of the angle ζ_k between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{v}_k$.
Correlations between $\mathbf{P}_k\mathbf{t}_k$ and $\mathbf{X}_k$ are:

$$\mathbf{Q}_k\mathbf{\Phi}_k\mathbf{P}_k'\mathbf{P}_k\mathbf{t}_k/n(\mathbf{t}_k'\mathbf{t}_k)^{-1/2} = \mathbf{Q}_k\mathbf{\Phi}_k\mathbf{t}_k/a_k$$

with sum of squares $\mathbf{t}_k'\mathbf{\Phi}_k^2\mathbf{t}_k/a_k^2$. The amount of variance of $\mathbf{X}_k$ explained by $\mathbf{Pt}$ then is found by multiplying the expression above by $\cos^2 \alpha_k = a_k^2$, with result

$$B_k = \mathbf{t}_k'\mathbf{\Phi}_k^2\mathbf{t}_k.$$

Here, B_k is the sum of the squared component loadings of the original variables in $\mathbf{X}_k$ with respect to the first OVERALS solution. The OVERALS program prints such component loadings, so that B_k then easily can be computed. The amount of variance of $\mathbf{X}_k$ explained by $\mathbf{v}_k$ becomes

$$(\mathbf{t}_k'\mathbf{\Phi}_k^2\mathbf{t}_k/a_k^2)(\cos^2 \zeta_k) = B_k(\cos^2 \zeta_k/\cos^2 \alpha_k)$$

and this expression gives a lower bound for the RA solution, given that it is a RA solution where a weighted sum of all variables not in $\mathbf{X}_k$ should explain most of the variance in $\mathbf{X}_k$.

But the problem can be reversed: what is the lower bound for a RA solution where a weighted sum of $\mathbf{X}_k$ should explain most of the variance in the other sets? The amount of variance of **X** explained by **Pt** can be shown to be equal to $\mathbf{t}'\Phi^2\mathbf{t}$. The amount of variance explained in sets other than $\mathbf{X}_k$ is then found by subtraction:

$$\mathbf{t}'\Phi^2\mathbf{t} - \mathbf{t}_k'\Phi_k{}^2\mathbf{t}_k.$$

The amount of variance in the other sets explained by $\mathbf{P}_k\mathbf{t}_k$ then is found by multiplying the last expression by $\cos^2 \alpha_k = a_k{}^2$:

$$(\mathbf{t}'\Phi^2\mathbf{t} - \mathbf{t}_k'\Phi_k{}^2\mathbf{t})a_k{}^2.$$

Remember that we used B_k for the sum of the squared component loadings of the variables in $\mathbf{X}_k$. Therefore,

$$\mathbf{t}'\Phi\mathbf{t} = \Sigma B_k$$

and stands for the sum of the squared component loadings of all variables.

9 BOUNDS DERIVED FROM MORE-DIMENSIONAL OVERALS

9.1 Bounds derived from other OVERALS dimensions than the first

Similar lower bounds as derived from the first OVERALS dimension, can be calculated on the basis of the second, third, etc., OVERALS dimensions. These bounds are also lower bounds for the best canonical correlation, or the best RA solution. They may be larger than the bounds derived from the first dimensions.

E.g., the lower bound for the canonical correlation depends for a great deal upon the value of a_k. Although the sum of the values $a_k{}^2$ is maximized in the first OVERALS dimension, this does not exclude that *some* a_k in the first dimension may be relatively small, and becomes much larger in a subsequent dimension.

For RA there is no reason at all to expect especially good lower bounds from the first dimension. The OVERALS criterion is not interested in component loadings, and it may happen that component loadings in the first dimension are all relatively small.

9.2 Bounds for more dimensions in CA or RA

Suppose we have an OVERALS solution in two dimensions. Is it possible to derive bounds for the *two* best canonical correlations or two best RA solutions? The answer is that this is possible. But it requires almost the same amount of computation as solving for CA or RA directly.

To illustrate, we take bounds for CA. Given the OVERALS solution, each set can be compressed to a set with only two variables in it:

$$\mathbf{Z}_k = (\mathbf{P}_k\mathbf{t}_{k1}, \mathbf{P}_k\mathbf{t}_{k2}) = \mathbf{P}_k\mathbf{T}_k.$$

Lower bounds for the canonical correlations can be found by applying CA to these compressed sets, by taking $\mathbf{Z}_k$ as one set, and all other $\mathbf{Z}_{h\neq k}$ together as the second set. The equations needed for such a "compressed CA" can be somewhat simplified (because the matrices $\mathbf{Z}_k$ are OVERALS results), but not to an extent that is very useful. For one thing, the columns of $\mathbf{Z}_k$ are not uncorrelated, and the vectors $\mathbf{v}_{k1}$ and $\mathbf{v}_{k2}$ also will not be orthogonal.

10 NON-NUMERICAL OVERALS

10.1 Definition

If the variables in $\mathbf{X}$ are categorical, the numerical OVERALS solution is based on a quantification of the categories which is specified *a priori*. The OVERALS results depend only on the identification of *optimal weights* $\mathbf{t}$. A non-numerical OVERALS, not only identifies optimal weights, but also an *optimal category* quantification.

In the same way we can have a non-numerical CA, or a non-numerical RA. E.g., the non-numerical RA solution in one dimension would specify a category quantification by which the redundancy is maximized. Any other category quantification would result in smaller redundancy.

Non-numerical solutions can be *single* or *multiple*. A single solution implies that the category quantification must be the *same* for all dimensions of the solution. A multiple solution allows the possibility that categories are quantified *differently* for each dimension of the solution.

E.g., a non-numerical single RA in two dimensions will quantify categories in such a way that this quantification maximizes the sum of the two redundancies. A multiple RA first quantifies categories such that the first redundancy is maximized, and thereafter maximizes the second redundancy by taking a different quantification.

10.2 Bounds

Let a non-numerical OVERALS solution be given. Bounds for CA and RA can be derived from it, in the same way as shown for the numerical solution.

Suppose we take bounds for CA from the first non-numerical OVERALS dimension. They are based on the OVERALS category quantification, and therefore are bounds valid for the CA solution based on this particular category quantification.

But even if a proper CA analysis were performed on the variables with OVERALS quantification, the resulting canonical correlation would only be a lower bound again for the non-numerical CA. The reason is obvious: non-numerical CA may take a different category quantification by which the "non-numerical canonical correlation" becomes larger than that based on the OVERALS quantification.

The bound based on OVERALS remains valid. However, one could say that it is only a lower bound for the canonical correlation that would be obtained if the OVERALS result is treated numerically. And this canonical correlation, in turn, is only a lower bound for the non-numerical canonical correlation.

11 CONCLUSIONS

We may repeat the argument of Section 8.1. If a researcher has the general idea that OVERALS will reveal to what extent individual sets are related to all other sets, then this notion is not incorrect, but it is rather unspecific. It could be made more specific in terms of CA or RA criteria. OVERALS does not give direct answers to CA or RA questions, but it does give lower bounds.

Section 9 shows that such lower bounds can better be limited to the best CA or RA, even if there are more OVERALS dimensions. Section 10 shows that the lower bounds have certain limitations for non-numerical CA and RA. The practical conclusion, of course, is that lower bounds are just what the word says. A small lower bound for a canonical correlation does not exclude the possibility that the correlation itself is very large.

Lower bounds for CA are more trustworthy than lower bounds for RA. The reason is that the OVERALS criterion is not at all concerned with component loadings, or "explained variance". This also can be seen from the fact that the OVERALS solution depends on the matrices $\mathbf{P}_k$ and ignores the eigenvalue matrices Φ_k^2. In theory it may happen that the best OVERALS solution for **Pt** explains very little of the variance of the original variables. Derived lower bounds for RA then give hardly any indication about what the real RA solution might be.

REFERENCES

Van de Geer, J.P. (1984), Linear relations among *k* sets of variables. *Psychometrika,* 1984, 49, 79-94.

Van de Geer, J.P. (1986), *Introduction to linear multivariate data analysis*. Leiden, DSWO Press.

Van de Geer, J.P. (1988), *Algebra and Geometry of OVERALS*. Report (in press), Dept. of Data Theory, University of Leiden.

TABLE 1

Relations between (cos ζ_k) and (cos α_k) for some selected values of γK.

γK	10	5	3	2	1.5	1	.4	.2	.05
	.526	.556	.6	.666	.75				
cos α_k									
.05	.045	.040	.033	.025	.017	0	.074	.197	.689
.10	.090	.080	.067	.050	.033	0	.149	.378	.886
.15	.135	.120	.101	.076	.051	0	.222	.519	.945
.20	.181	.161	.135	.102	.068	0	.293	.632	.968
.25	.226	.202	.170	.128	.086	0	.361	.718	.980
.30	.272	.244	.205	.155	.104	0	.427	.783	.986
.35	.319	.286	.242	.184	.124	0	.489	.831	.990
.40	.366	.330	.279	.213	.144	0	.547	.868	.993
.45	.413	.374	.318	.244	.166	0	.603	.896	.995
.50	.461	.419	.359	.277	.189	0	.655	.918	.996
.55	.510	.466	.402	.313	.214	0	.703	.935	.997
.60	.510	.446	.402	.313	.214	0	.703	.935	.997
.60	.559	.514	.447	.351	.243	0	.747	.949	.998
.65	.610	.565	.495	.393	.274	0	.789	.960	.998
.70	.662	.617	.547	.440	.331	0	.827	.969	.999
.75	.714	.672	.603	.493	.354	0	.862	.977	.999
.80	.768	.730	.664	.555	.406	0	.894	.983	.999
.85	.824	.791	.732	.628	.474	0	.924	.988	.9995
.90	.881	.855	.809	.718	.567	0	.952	.993	.9997
.95	.939	.925	.897	.836	.712	0	.976	.997	.9999

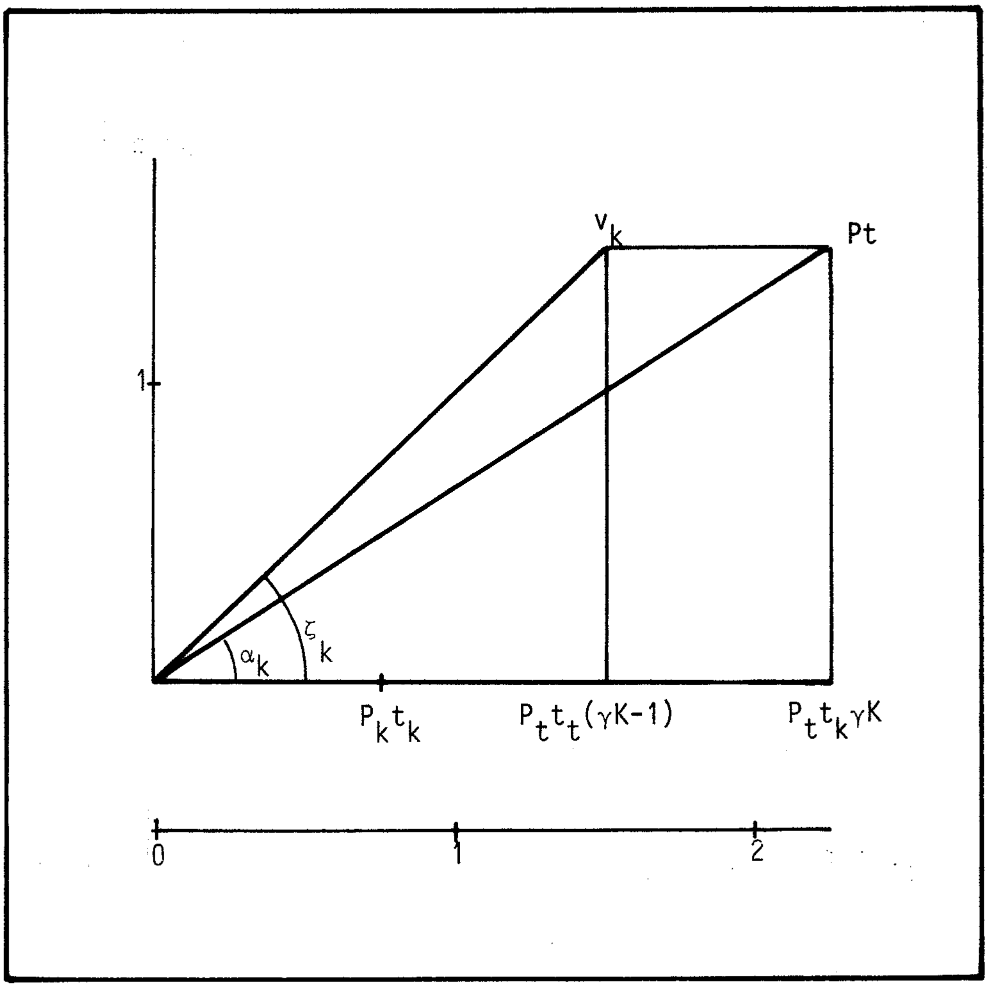

Figure 1.
Graph of vectors with $\gamma K = 3$, $(\cos \alpha_k) = .75$, and $(\cos \zeta_k) = .603$.

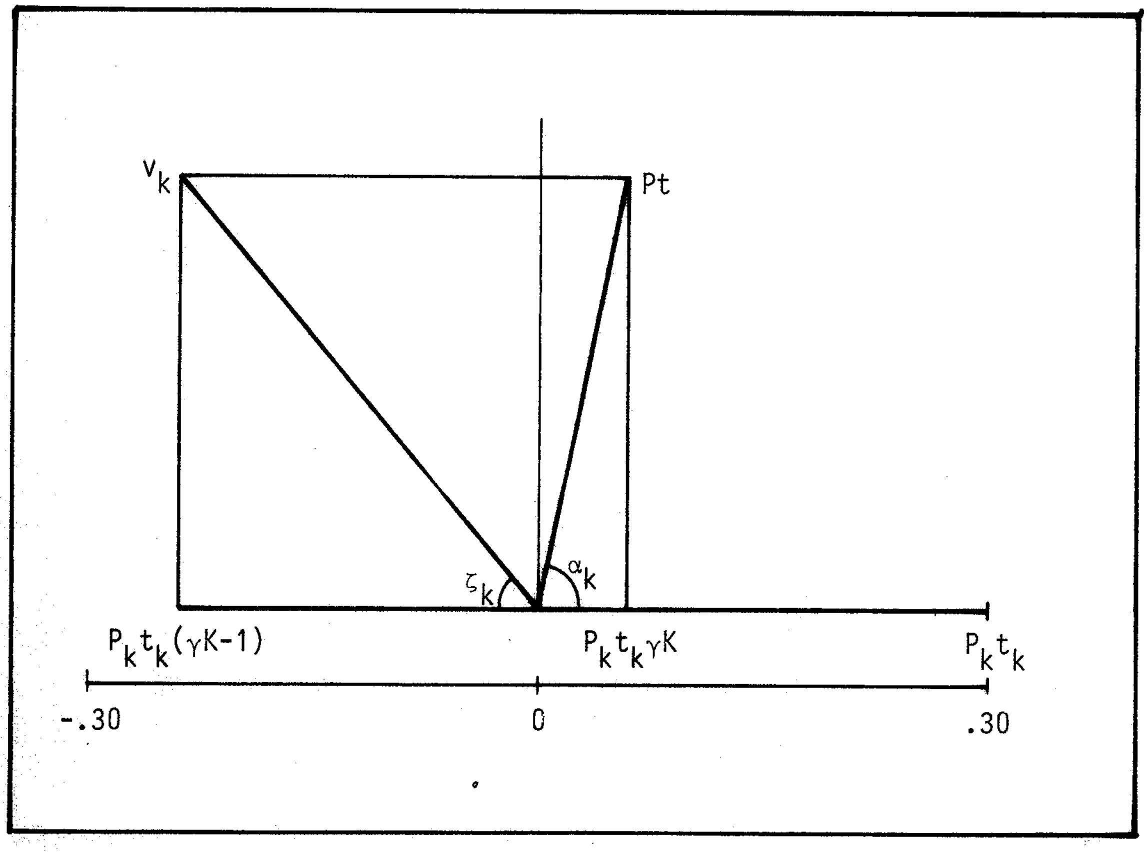

Figure 2.
Graph of vectors with γK = .2, (cos α_k) = .30, and (cos ζ_k) = .632.