

**GENERALIZED
SET COMPONENT ANALYSIS :
A TOOLBOX FOR MULTIPLE SETS ANALYSIS**

Dré Nierop

**Department of Data Theory
University of Leiden**

	Abstract	1
1	Introduction	2
2	Generalized Principal Component Analysis (GPCA)	4
2.1	Upper bounds of variance accounted for	5
2.2	Different ways for defining balance among sets	5
2.3	Balancing of sets related to variance accounted for	6
3	Generalized Canonical Correlation Analysis (GCCA)	7
3.1	Canonical Correlation Analysis (CCA)	8
3.2	Relation between CCA and GCCA	8
4	Generalized Set Component Analysis (GSCA)	9
4.1	Subspace restrictions for prediction	10
4.2	Different filters in one analysis	11
4.2.1	Redundancy Analysis (RA)	11
4.2.2	Generalized Redundancy Analysis (GRA)	12
4.2.3	Generalized Reduced Rank Regression (GRRR)	12
4.3	Compound filters	13
4.3.1	Shrinkage techniques	14
5	Quadratic filters	15
5.1	Fixed Set Component Analysis (FSCA)	16
5.2	Set Component Analysis (SCA)	16
5.2.1	Correlations of set variates with proportional weights	17
5.2.2	Penalty on little variance accounted for	18
5.3	Computation	18
	References	19

Abstract

We present Generalized Set Component Analysis as a tool box for composing many types of multiple sets analysis. Basic components are filters for each set that model the eigenvalue structure of these sets. Many examples are given to illustrate the concept. For instance we generalize well-known models for multiple sets, such as *Principal Component Analysis*, *Canonical Correlation Analysis*, *Redundancy Analysis* and *Reduced Rank Regression*. One new model called Set Component Analysis is discussed extensively. It is introduced as a stabilized form of *Generalized Canonical Correlation Analysis*.

Keywords: canonical correlation analysis, eigenvalues, filters, multiple sets, multivariate analysis, prediction, principal component analysis, reduced rank regression, redundancy analysis, set component analysis, shrinkage techniques, space restrictions.

1 Introduction

Within a larger theoretical framework to be discussed later, we introduce *Set Component Analysis (SCA)*, a new method for the exploratory analysis of two or multiple sets of variables. It can be seen as a stabilized form of *Generalized Canonical Correlation Analysis (GCCA)* or as a generalization of *Principal Component Analysis (PCA)* for multiple sets. We elaborate these two viewpoints in more detail.

The *stability* of SCA compared to GCCA (Carroll, 1968) is treated in two parts. First we discuss the *stability of the common latent variables* and next the *stability of the weights*. In GCCA we maximize squared correlations between canonical variates and a common latent variable, and in SCA we maximize squared correlations between set variates and a common latent variable. The only difference is that in SCA we use a filter that makes a solution less attractive when there is little variance accounted for by the common latent variable. So it is very unlikely to find common latent variables with little variance accounted for, which are usually unstable. In GCCA, by contrast, there is in principle not any relation between the fit of the solution and the variance accounted for. As to the stability of the weights that constitute the set variates in SCA we expect the weights to be more stable compared to GCCA, because the common latent variables are more stable. The relation between stable weights and stable common latent variables is understood through an intermediate step. Stable common latent variables imply that the correlations of variables with the common latent variables will also be more stable. These correlations we call the *structure correlations*. (Gittins (1985) describes the structure correlation as the variable/ canonical variate correlation, but in this multiple sets report we change it into the variable/ common latent variate correlation.) The stability of the weights now follows from the fact that in SCA the weights that constitute the canonical variate of a set are *proportional* to the corresponding structure correlations. In this report we assume the variables and the common latent variables to be in deviation of the mean and to be unit-normalized; otherwise, the weights should be proportional to 'structure covariances', i.e. the covariances of variables with the common latent variables. In GCCA the weights are proportional to the structure correlations only under very specific conditions.

Our second point of view for introducing SCA was to discuss this method as a *generalization* of PCA for *multiple sets*. A simple way to generalize PCA for multiple sets seems to be a PCA of all sets taken together in one large partitioned matrix. In order to balance the influence of the sets on the solution equally we normalize the sets in a preprocessing step. A very common procedure is to divide the variables of each set by a corresponding set specific *balancing constant*. Usually this balancing constant is the sum of squares of the variables of the set, which results in an equal total variance for all sets. It is a

misconception that this strategy always equally balances the influence of the sets, as will be shown by deriving the upper bounds of the variance accounted for under several conditions. In principle the upper bounds should be equal for all sets. It will be clarified why the balancing of sets is also not always achieved satisfactorily by choosing the balancing constant equal to the first eigenvalue of the corresponding sets. Next alternative normalizations are formulated by using the upper bounds of the variance accounted for. Because we want to control not only the variance accounted for, but also the correlations of set variates, we have decided to balance the solution by analyzing correlations of set variates and to filter the set variance. So SCA is only indirectly a generalization of PCA, in the sense that it tries to improve the GCCA solution with respect to the variance accounted for by relocating the canonical variates. These relocated canonical variates we call *set variates*.

The usefulness of SCA depends on the structure of the data. The GCCA solution is equal to the generalized PCA solution, if the variables are unit-normalized and uncorrelated within each set (van de Geer, 1984). Under these conditions SCA is identical to GCCA and generalized PCA. The results of the three methods can differ more when the intercorrelations within one or more sets gradually deviate more from zero. Therefore, SCA is expected to be especially useful for the analysis of the relation between two or multiple sets, when *high intercorrelations* between the variables occur within one or more sets. These high intercorrelations can frequently be found, for instance, in data with many variables in one or more sets. The GCCA solution for this kind of data can capitalize on only a very small part of the data. This effect will more easily occur when the number of variables within one or more sets increases. The reason for this tendency can be found in the number of eigenvalues of a set with relative small values. The eigenvectors corresponding to these small eigenvalues occupy a larger part of the space spanned by some highly intercorrelated set, when we add more highly correlated variables to a set. The larger proportion of small eigenvalues increases the probability to find little variance accounted for. Such little variance solutions will generally not be representative for the data and not easily interpretable, because the structure correlations are all very low. It is possible to circumvent these problems with highly intercorrelated sets by first replacing these sets by low rank approximations. In a second step we analyze the low rank approximations for all sets together. In SCA we can directly analyze the raw data in one step without the mentioned problems. Another consideration of using SCA can be that SCA favours a solution with an *equally balanced* influence of sets. This equal balancing prevents to a certain extent the disappearance of one or more sets in the solution. A last reason to use SCA can be that a researcher is interested to have more *stable weights* for the variables compared to GGCA in situations where high intercorrelations within sets occur.

In section 2 we present straightforward generalizations of PCA for multiple sets (GPCA) and discuss the upper bounds for the variance accounted for under several conditions. The GCCA model is given in section 3 and we formulate space restrictions for the common latent variable in order to obtain ordinary 2 sets CCA. In section 4 we show how GPCA and GCCA can be understood as techniques that filter the eigenvalues of sets. The general formulation of filtering eigenvalues is described as Generalized Set Component Analysis (GSCA). We illustrate how space restrictions can generally be used for prediction purposes. GPCA, GCCA, Redundancy Analysis (van den Wollenberg, 1977), generalizations of RA and generalizations of Reduced Rank Regression are formulated as special cases of GSCA, as well as the use of low rank approximations of sets. The new filter of SCA is described in section 5.

2 Generalized Principal Component Analysis (GPCA)

Suppose the data to be analyzed are collected in the matrix \mathbf{H} , partitioned into K sets: $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_k, \dots, \mathbf{H}_K)$ with n rows (objects) and m_k columns (variables) for set \mathbf{H}_k . We assume the data to be in deviation of the mean and to be unit normalized, so the columns have sum of squares equal to 1. These assumptions imply that $\mathbf{H}_k' \mathbf{H}_k$ is a correlation matrix between the variables of set k . The generalization of PCA for multiple sets is described by maximizing the following function

$$\text{GPCA}(\mathbf{X}) = \text{tr} \sum_{k=1}^K w_k^{-1} \mathbf{X}' \mathbf{H}_k \mathbf{H}_k' \mathbf{X} = \text{tr} \sum_{k=1}^K \mathbf{X}' \mathbf{P}_k (\Phi_k^2 w_k^{-1}) \mathbf{P}_k' \mathbf{X}, \quad (1)$$

where $n \mathbf{X}_p = \mathbf{x}_1, \dots, \mathbf{x}_s, \dots, \mathbf{x}_p$ denote the common latent variables with $\mathbf{X}' \mathbf{X} = \mathbf{I}$ and $w_1, \dots, w_k, \dots, w_K$ denote fixed balancing constants for set k .

The last part of equation (1) is derived from the Singular Value Decomposition (SVD) of each \mathbf{H}_k . This SVD for set k is given by $\mathbf{H}_k = \mathbf{P}_k \Phi_k \mathbf{Q}_k'$, where \mathbf{P}_k ($n \times p_k$) and \mathbf{Q}_k ($m_k \times p_k$) denote orthonormal eigenvector matrices and Φ_k denotes a diagonal matrix with p_k nonzero singular values in descending order. So $\mathbf{H}_k \mathbf{H}_k' = \mathbf{P}_k \Phi_k^2 \mathbf{P}_k'$ and the eigenvalues Φ_k^2 of the symmetric matrix $\mathbf{H}_k \mathbf{H}_k'$ are equal to the squared singular values of \mathbf{H}_k . The last part of (1) is added to get used to the idea that the balancing of sets can also be described by a rescaling of the eigenvalues of these sets.

Because we are maximizing balanced variance accounted for, we can relate the balancing constants with the variance accounted for. In the next sections we will first derive for each set separately the upper bounds of the variance accounted for, then we describe some examples of balancing in GPCA.

2.1 Upper bounds of variance accounted for

Our rationale for the treatment of sets in GPCA is to control the influence of these sets on the solution with respect to their variance. A measure for the influence of set k is the variance of \mathbf{H}_k accounted for by the p common latent variables \mathbf{X} . In matrix formulation this Variance Accounted For is equal to

$$\text{VAF}(p,k) = \text{tr } \mathbf{X}'\mathbf{H}_k\mathbf{H}_k'\mathbf{X} = \text{tr } \mathbf{X}'\mathbf{P}_k\Phi_k^2\mathbf{P}_k'\mathbf{X} = \text{tr } \Phi_k^2 \mathbf{D}_{\mathbf{X}'\mathbf{P}_k\mathbf{P}_k'\mathbf{X}}, \quad (2)$$

where $\mathbf{D}_{\mathbf{X}'\mathbf{P}_k\mathbf{P}_k'\mathbf{X}}$ denotes a diagonal matrix with the diagonal elements of $\mathbf{X}'\mathbf{P}_k\mathbf{P}_k'\mathbf{X}$. These diagonal elements give the squared scalar products between the common latent variables \mathbf{X} and the eigenvectors \mathbf{P}_k , and each diagonal element has an upper bound of 1, because \mathbf{X} and \mathbf{P}_k have unit normalized columns. The upper bound for set k of the variance accounted for in p dimensions is now given by the sum of the first p eigenvalues

$$\max \text{VAF}(k) = \sum_{s=1}^p \phi_{sk}^2, \quad (3)$$

with the eigenvalues arranged in descending order. We also derive the upper and lower bounds of $\max \text{VAF}(k)$, if the eigenvalue structure Φ_k^2 of set k is unknown. The upper bound is reached, if \mathbf{H}_k is of deficient rank in the sense that the $m_k - p$ smallest eigenvalues are all equal to zero and $\max \text{VAF}(k) = \text{tr } \Phi_k^2 = m_k$. The lower bound is reached if \mathbf{H}_k is orthonormal and therefore all eigenvalues are equal to 1. Under this condition the $m_k - p$ smallest eigenvalues are as large as possible, and $\min \text{VAF}(k) = p$. Summarizing we write for set k

$$p \leq \max \text{VAF}(k) \leq m_k. \quad (4)$$

2.2 Different ways for defining balance among sets

We formulate some straightforward types of GPCA by fixing the balancing constants w_k in a simple way. In the next section we elaborate the relation between balancing sets and the variance accounted for. The first type of analysis is GPCAI with $w_k=1 \forall k$. So the rescaled eigenvalues are *identical* to the original eigenvalues. After substitution of $w_k=1 \forall k$ in (1) we have to maximize

$$\text{GPCAI}(\mathbf{X}) = \text{tr } \mathbf{X}'\mathbf{H}\mathbf{H}'\mathbf{X} = \text{tr } \mathbf{X}'\mathbf{P}\Phi^2\mathbf{P}'\mathbf{X}, \quad (5)$$

with $\mathbf{X}'\mathbf{X}=\mathbf{I}$. The matrices \mathbf{P}_k are collected in one partitioned matrix $\mathbf{P}=(\mathbf{P}_1,\dots,\mathbf{P}_K)$ and the corresponding singular values are collected in one $(\sum_k p_k) \times (\sum_k p_k)$ diagonal matrix Φ . We emphasize that $\mathbf{P}\Phi^2\mathbf{P}'$ doesn't give the eigenvalue decomposition of matrix $\mathbf{H}\mathbf{H}'$, because \mathbf{P} is not necessarily an orthonormal matrix. Only the \mathbf{P}_k are orthonormal for each set k separately. The type of set balancing presented above amounts to calculating the first p principal components of $\mathbf{P}\Phi^2\mathbf{P}'$ and the way the variables are grouped in sets has no influence on the solution, because $\mathbf{P}\Phi^2\mathbf{P}'=\mathbf{H}\mathbf{H}'$.

The second type of analysis in this section we call GPCAt with $w_k = (\text{tr } \Phi_k^2) \forall k$. So in this case the balancing constant is the *trace* of $\mathbf{H}_k \mathbf{H}_k'$, which is equal to the sum of the eigenvalues Φ_k^2 and equal to the sum of squares of \mathbf{H}_k . So with unit normalized variables $\text{tr } \Phi_k^2 = m_k$. In the format of (1) we maximize

$$\text{GPCAt}(\mathbf{X}) = \text{tr} \sum_{k=1}^K \mathbf{X}' \mathbf{P}_k (\Phi_k^2 / \text{tr } \Phi_k^2) \mathbf{P}_k' \mathbf{X} \quad (6)$$

with $\mathbf{X}'\mathbf{X}=\mathbf{I}$. For GPCAt this amounts to performing a principal component analysis on the datamatrix \mathbf{H} after the sum of squares of each set is normalized to 1. We choose GPCAt if we want the mean proportion of variance accounted for as large as possible.

The third type of analysis is GPCAf with $w_k = \phi_{1k}^2 \forall k$. The balancing constant in (1) is now the *first eigenvalue* ϕ_{1k}^2 of $\mathbf{H}_k \mathbf{H}_k'$, with the eigenvalues arranged in descending order. We maximize

$$\text{GPCAf}(\mathbf{X}) = \text{tr} \sum_{k=1}^K \mathbf{X}' \mathbf{P}_k (\Phi_k^2 / \phi_{1k}^2) \mathbf{P}_k' \mathbf{X} \quad (7)$$

with $\mathbf{X}'\mathbf{X}=\mathbf{I}$. The reasons for choosing this criterion will be discussed in the next section.

2.3 Balancing of sets related to variance accounted for

In the previous section we defined several types of set balancing. For the identical balancing in GPCAi the maximum variance accounted for of set k is between p and m_k , depending on the structure of set \mathbf{H}_k as derived in (4). We can find considerable differences of *maxVAF* between sets, if the number of variables of one or more sets is large compared to the number of dimensions. From our point of view the ideal balancing for set k in (1) would be to take $w_k = \text{maxVAF}(k) \forall k$, as formulated in (3), in order to obtain an equally balanced influence of the sets on the solution with respect to their variance. The *maxVAF* balancing gives another type of GPCA

$$\text{GPCAm}(\mathbf{X}) = \text{tr} \sum_{k=1}^K \mathbf{X}' \mathbf{P}_k (\Phi_k^2 / \sum_{s=1}^p \phi_{sk}^2) \mathbf{P}_k' \mathbf{X}, \quad (8)$$

with $\mathbf{X}'\mathbf{X}=\mathbf{I}$ and with p equal to the number of columns of \mathbf{X} . The balancing of GPCAt (6) and GPCAf (7) appear to be special cases of GPCAm. They only depend on the number of dimensions we want to compute. For $p = \max(m_k)$ we find the equality $\text{GPCAm} = \text{GPCAt}$ and for $p = 1$ the equality $\text{GPCAm} = \text{GPCAf}$. The solutions of GPCAm are not nested. (Solutions are nested if successive computation of p dimensions always gives the same results as simultaneous computation of p dimensions for all possible p .) This property can be seen as a drawback compared to GPCAt and GPCAf. If one wants to choose only between *trace* balancing or *first eigenvalue* balancing, the number of dimensions compared to the maximum number of variables in a set has to be decisive.

An even more strict equality concept can be formulated by requiring an equal balance of the influence of sets on *each* dimension with respect to variance. In case of successive one dimensional solutions the GPCAm criterion function can be adapted to this concept by using *deflation*. After each successive step the original matrices \mathbf{H}_k are replaced by their antiprojections on the previous dimensions of \mathbf{x}_r . The resulting \mathbf{X} will be orthogonal. In fact we apply first eigenvalue balancing, because in each step there is only one dimension to find the maximum of

$$dGPCAf(\mathbf{X}) = \sum_{s=1}^p \sum_{k=1}^K \mathbf{x}_s' \mathbf{P}_{(k)s} (\Phi_{(k)s}^2 / \phi_{1(k)s}^2) \mathbf{P}_{(k)s}' \mathbf{x}_s, \quad (9)$$

where the deflation enters the problem by defining

$$\begin{aligned} \mathbf{P}_{(k)s} \Phi_{(k)s}^2 \mathbf{P}_{(k)s}' &= \mathbf{P}_k \Phi_k^2 \mathbf{P}_k' && \text{for } s = 1 \quad \forall k \\ \mathbf{H}_{(k)s} \mathbf{H}_{(k)s}' &= \mathbf{P}_{(k)s} \Phi_{(k)s}^2 \mathbf{P}_{(k)s}' = (\mathbf{I} - \mathbf{x}_{s-1} \mathbf{x}_{s-1}') \mathbf{H}_{(k)s-1} \mathbf{H}_{(k)s-1}' (\mathbf{I} - \mathbf{x}_{s-1} \mathbf{x}_{s-1}') \\ &= (\mathbf{I} - \mathbf{x}_{s-1} \mathbf{x}_{s-1}') \mathbf{P}_{(k)s-1} \Phi_{(k)s-1}^2 \mathbf{P}_{(k)s-1}' (\mathbf{I} - \mathbf{x}_{s-1} \mathbf{x}_{s-1}'). && \text{for } s = 2, \dots, p \quad \forall k \end{aligned}$$

In each successive dimension the datamatrices $\mathbf{H}_{k(r)}$ change and therefore the eigenvectors and eigenvalues of the sets.

The techniques formulated in this section emphasize an equally balanced influence on the solution with respect to set variance, but they don't affect the correlations of set variates. In the next section we will discuss a technique that balances the influence on the solution only with respect to correlations of weighted sums of set variables.

3 Generalized Canonical Correlation Analysis (GCCA)

In GCCA as formulated by Carroll (1968) we maximize squared correlations between canonical variates and p common latent variables

$$GCCA(\mathbf{X}, \mathbf{Z}_1, \dots, \mathbf{Z}_k, \dots, \mathbf{Z}_K) = \sum_{s=1}^p \sum_{k=1}^K (\mathbf{x}_s' \mathbf{z}_{(k)s})^2, \quad (10)$$

where $n \mathbf{X}_p = \mathbf{x}_1, \dots, \mathbf{x}_s, \dots, \mathbf{x}_p$ denote the common latent variables with $\mathbf{X}'\mathbf{X} = \mathbf{I}$ and

$n(\mathbf{Z}_k)_p = \mathbf{z}_{(1)k}, \dots, \mathbf{z}_{(k)s}, \dots, \mathbf{z}_{(k)p}$ denote the unit normalized canonical variates for set k and dimension s ,

with $\mathbf{z}_{(k)s} = \mathbf{H}_k \mathbf{t}_{(k)s} = \mathbf{P}_k \Phi_k \mathbf{Q}_k' \mathbf{t}_{(k)s} = \mathbf{P}_k \mathbf{v}_{(k)s}$

and $\mathbf{z}_{(k)s}' \mathbf{z}_{(k)s} = \mathbf{t}_{(k)s}' \mathbf{H}_k' \mathbf{H}_k \mathbf{t}_{(k)s} = \mathbf{v}_{(k)s}' \mathbf{P}_k' \mathbf{P}_k \mathbf{v}_{(k)s} = \mathbf{v}_{(k)s}' \mathbf{v}_{(k)s} = 1. \quad \forall k, s$

As in the previous sections the SVD for set k is given by $\mathbf{H}_k = \mathbf{P}_k \Phi_k \mathbf{Q}_k'$, where \mathbf{P}_k ($n \times p_k$) and \mathbf{Q}_k ($m_k \times p_k$) denote orthonormal eigenvector matrices and Φ_k denotes a diagonal matrix with p_k nonzero singular values in descending order. The canonical weights $\mathbf{t}_{(k)s}$ can be derived from the weights $\mathbf{v}_{(k)s}$ by $\mathbf{t}_{(k)s} = \mathbf{Q}_k \Phi_k^{-1} \mathbf{v}_{(k)s} \quad \forall k, s$.

For our purpose we need a description of GGCA with only the \mathbf{X} as unknown parameters. Therefore, we first substitute $\mathbf{z}_{(k)s} = \mathbf{P}_k \mathbf{v}_{(k)s}$ with $\mathbf{v}_{(k)s}' \mathbf{v}_{(k)s} = 1 \quad \forall k, s$ in (10) and compute a partial maximum for

$$\begin{aligned} \text{GCCA}(\mathbf{x}_s, \mathbf{c}_{(k)s}, \mathbf{v}_{(k)s}) &= (\mathbf{x}_s' \mathbf{P}_k \mathbf{v}_{(k)s})^2 + \mathbf{c}_{(k)s} \\ &= \mathbf{v}_{(k)s}' \mathbf{P}_k' \mathbf{x}_s \mathbf{x}_s' \mathbf{P}_k \mathbf{v}_{(k)s} + \mathbf{c}_{(k)s} \end{aligned} \quad \forall k, s \quad (11)$$

where \mathbf{x}_s denotes the fixed common latent variable \mathbf{x}_s
and $\mathbf{c}_{(k)s}$ denotes the sum for all other fixed parameters. $\forall k, s$

The partial maximum of $\text{GCCA}(\mathbf{v}_{(k)s})$ is reached if $\mathbf{v}_{(k)s}$ is equal to the eigenvector associated with the largest eigenvalue of matrix $\mathbf{P}_k' \mathbf{x}_s \mathbf{x}_s' \mathbf{P}_k$. Because matrix $\mathbf{P}_k' \mathbf{x}_s \mathbf{x}_s' \mathbf{P}_k$ is of rank one, we have only one nonzero eigenvalue with eigenvector $\mathbf{v}_{(k)s} = \mathbf{P}_k' \mathbf{x}_s (\mathbf{x}_s' \mathbf{P}_k \mathbf{P}_k' \mathbf{x}_s)^{-1/2}$. Now we simplify (10) by substituting $\mathbf{z}_{(k)s} = \mathbf{P}_k \mathbf{P}_k' \mathbf{x}_s (\mathbf{x}_s' \mathbf{P}_k \mathbf{P}_k' \mathbf{x}_s)^{-1/2}$

$$\begin{aligned} \text{GCCA}(\mathbf{X}) &= \sum_{s=1}^p \sum_{k=1}^K (\mathbf{x}_s' \mathbf{P}_k \mathbf{P}_k' \mathbf{x}_s (\mathbf{x}_s' \mathbf{P}_k \mathbf{P}_k' \mathbf{x}_s)^{-1/2})^2 \\ &= \text{tr} \sum_{k=1}^K \mathbf{X}' \mathbf{P}_k \mathbf{P}_k' \mathbf{X} = \text{tr} \mathbf{X}' \mathbf{P} \mathbf{P}' \mathbf{X}, \end{aligned} \quad (12)$$

with $\mathbf{X}' \mathbf{X} = \mathbf{I}$ and the matrices \mathbf{P}_k collected in one partitioned matrix $\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_K)$.

3.1 Canonical Correlation Analysis (CCA)

Ordinary 2-sets CCA is usually not defined in terms of the common latent variables \mathbf{X} . Rather we simply maximize the sum of the canonical correlations between the canonical variates of two sets. Using exactly the same notation as in the previous section this can be expressed as the maximization of

$$\text{CCA}(\mathbf{Z}_1, \mathbf{Z}_2) = \sum_{s=1}^p \mathbf{z}_{1(r)}' \mathbf{z}_{2(r)} = \text{tr} \mathbf{Z}_1' \mathbf{Z}_2 = \text{tr} \mathbf{V}_1' \mathbf{P}_1' \mathbf{P}_2 \mathbf{V}_2, \quad (13)$$

with $\mathbf{Z}_k = \mathbf{H}_k \mathbf{T}_k = \mathbf{P}_k \Phi_k \mathbf{Q}_k' \mathbf{T}_k = \mathbf{P}_k \mathbf{V}_k$
and $\mathbf{Z}_k' \mathbf{Z}_k = \mathbf{T}_k' \mathbf{H}_k' \mathbf{H}_k \mathbf{T}_k = \mathbf{V}_k' \mathbf{P}_k' \mathbf{P}_k \mathbf{V}_k = \mathbf{V}_k' \mathbf{V}_k = \mathbf{I}$. for $k = 1, 2$

The solution for \mathbf{V}_1 and \mathbf{V}_2 can be found by taking respectively the p principal left and right eigenvectors of matrix $\mathbf{P}_1' \mathbf{P}_2$. The singular values give the canonical correlations.

3.2 Relation between CCA and GGCA

We can derive CCA from GGCA by imposing *subspace restrictions* on the common latent variables \mathbf{X} . We want the solution of \mathbf{X} to be in the subspace associated with \mathbf{H}_c , spanned by the orthonormal basis \mathbf{P}_c . In other words we require

$$\mathbf{X} = \mathbf{P}_c \mathbf{P}_c' \mathbf{X}. \quad (14)$$

The same restriction is obtained if we require $\mathbf{X} = \mathbf{P}_c \mathbf{V}_c$ and therefore $\mathbf{V}_c = \mathbf{P}_c' \mathbf{X}$. We use the notation 'c' especially for subspace restrictions in order to avoid confusion with 'k' in subsequent sections.

Substitution of respectively $\mathbf{X}_1 = \mathbf{P}_1 \mathbf{V}_1$ and $\mathbf{X}_2 = \mathbf{P}_2 \mathbf{V}_2$ in (12) for $K = 2$ results in the maximization of

$$\begin{aligned} \text{GCCA}(\mathbf{V}_1) &= \text{tr } \mathbf{V}_1' \mathbf{V}_1 + \mathbf{V}_1' \mathbf{P}_1' \mathbf{P}_2 \mathbf{P}_2' \mathbf{P}_1 \mathbf{V}_1 \\ \text{and } \text{GCCA}(\mathbf{V}_2) &= \text{tr } \mathbf{V}_2' \mathbf{V}_2 + \mathbf{V}_2' \mathbf{P}_2' \mathbf{P}_1 \mathbf{P}_1' \mathbf{P}_2 \mathbf{V}_2, \end{aligned} \quad (15)$$

with $\mathbf{X}_1' \mathbf{X}_1 = \mathbf{V}_1' \mathbf{V}_1 = \mathbf{X}_2' \mathbf{X}_2 = \mathbf{V}_2' \mathbf{V}_2 = \mathbf{I}$. The solutions for \mathbf{V}_1 and \mathbf{V}_2 can be found by taking the p principal eigenvectors of respectively matrix $\mathbf{P}_1' \mathbf{P}_2 \mathbf{P}_2' \mathbf{P}_1$ and $\mathbf{P}_2' \mathbf{P}_1 \mathbf{P}_1' \mathbf{P}_2$. These eigenvectors are equal to respectively the p principal left and right eigenvectors of matrix $\mathbf{P}_1' \mathbf{P}_2$, which we recognize from (13). The eigenvalues of $\mathbf{P}_1' \mathbf{P}_2 \mathbf{P}_2' \mathbf{P}_1$ and $\mathbf{P}_2' \mathbf{P}_1 \mathbf{P}_1' \mathbf{P}_2$ are equal to the squared singular values of $\mathbf{P}_1' \mathbf{P}_2$ and therefore equal to the squared canonical correlations. The canonical variates of set 1 are given by \mathbf{X}_1 and the canonical variates of set 2 by \mathbf{X}_2 , because $\mathbf{X}_1 = \mathbf{P}_1 \mathbf{V}_1$ and $\mathbf{X}_2 = \mathbf{P}_2 \mathbf{V}_2$, see (13).

4 Generalized Set Component Analysis (GSCA)

The general Set Component Analysis criterion for describing all the models in this report is a function of the common latent variables \mathbf{X} . We maximize

$$\text{GSCA}(\mathbf{X}) = \text{tr } \sum_{k=1}^K \mathbf{X}' \mathbf{P}_k \Omega_k(\Phi_k^2) \mathbf{P}_k' \mathbf{X}, \quad (16)$$

with $\mathbf{X}' \mathbf{X} = \mathbf{I}$, where \mathbf{P}_k is given by the SVD for set k , $\mathbf{H}_k = \mathbf{P}_k \Phi_k \mathbf{Q}_k'$, and where $\Omega_k(\Phi_k^2)$ denotes a diagonal matrix with the filtered eigenvalues of set k . The filter Ω maps in a formal way the values of some matrix \mathbf{A} into matrix $\Omega(\mathbf{A})$. The filter is indexed with k , so every set has its own filter Ω_k .

By specifying between brackets the appropriate filters we describe the techniques we have discussed in the previous chapters. We start with the GPCA filter

$$\text{GPCA}\{\Omega_k(\Phi_k^2) = \Phi_k^2 w_k^{-1}\}, \quad \forall k \quad (17)$$

As we can verify easily, substitution of this filter in (16) results in (1). All different types of GPCA formulated in section 2 are described by substituting the corresponding balancing constants in (17). The names of the corresponding filters are: *identity* filter for $w_k = 1$, *trace* filter for $w_k = (\text{tr } \Phi_k^2)$, *first eigenvalue* filter for $w_k = \phi_{1k}^2$ and *maxVAF* filter for $\sum_p \phi_{sk}^2$.

For the GCCA filter we have

$$\text{GCCA}\{\Omega_k(\Phi_k^2) = \mathbf{I}\}, \quad \forall k \quad (18)$$

where \mathbf{I} is an identity matrix of appropriate size $m_k \times m_k$. Substitution of this *constant* filter in (16) results in (12). By applying the constant filter we replace the Pythagorean distances between the rows of a certain set by the Mahalanobis distances (Meulman, 1986).

For the description of CCA we have to add subspace restrictions to (18)

$$\text{CCA}\{\Omega_k(\Phi_k^2) = \mathbf{I}\}, \mathbf{X}_c = \mathbf{P}_c \mathbf{P}_c' \mathbf{X}_c. \quad \text{for } k = 1, 2 \text{ and } c = 1 \text{ or } 2 \quad (19)$$

The canonical variates of set c are given by \mathbf{X}_c . This is explained in section 3.2.

Many other techniques can be described in GSCA by specifying filters, whether or not in combination with subspace restrictions. We will indicate some possibilities in the next sections.

4.1 Subspace restrictions for prediction

In general subspace restrictions introduce an asymmetry in the analysis concerning the location of the common latent variables in one particular set. The motivation for this asymmetry can be prediction. For instance the prediction of one or more sets of criterion variables by a weighted sum of predictor variables. The common latent variables of GSCA can only be expressed as a weighted sum of predictor variables, if the space spanned by the predictor set also includes the common latent variables. For predictor set c this is achieved by requiring $\mathbf{X}_c = \mathbf{P}_c \mathbf{P}_c' \mathbf{X}_c$. The \mathbf{X} are labeled with c , in order to discriminate between the solutions that we obtain after imposing subspace restrictions on different sets c . We give two examples of applying subspace restrictions for prediction purposes.

The introduction of prediction in GCCA is formulated as

$${}^c\text{GCCA}\{\Omega_k(\Phi_k^2) = \mathbf{I}\}, \mathbf{X}_c = \mathbf{P}_c \mathbf{P}_c' \mathbf{X}_c, \quad \forall k \quad (20)$$

where the upper left superscript c of ${}^c\text{GCCA}$ indicates that we are dealing with a subspace restriction on set c in GCCA. By maximizing this criterion we find common latent variables that are a linear combination of set c and have the highest sum of squared canonical correlations with all other sets. In this case set c is a predictor set in a 'variate' sense, because it maximizes the relations with the set variates and not with the set variables in terms of variance accounted for. The maximization of ${}^c\text{GCCA}$ for $K = 2$ and $c = 1$ or 2 gives the CCA solution, with the canonical variates of set c equal to \mathbf{X}_c . The two maximization problems in principle lead to the same eigenvalue problem (section 3.2). If we know the prediction for one set, we can easily derive from this solution the prediction for the other set. For more than two sets we cannot reduce the K maximization problems for $c = 1, \dots, K$ to one single eigenvalue problem. Each prediction problem has to be solved by itself, unless some of the \mathbf{P}_c matrices are exactly the same. In Gifi (1981) a related version of ${}^c\text{GCCA}$ is used for multivariate analysis of variance.

In an analogous way we introduce prediction of set variables in GPCA by

$${}^c\text{GPCA}\{\Omega_k(\Phi_k^2) = \Phi_k^2 w_k^{-1}\}, \mathbf{X}_c = \mathbf{P}_c \mathbf{P}_c' \mathbf{X}_c, \quad \forall k \quad (21)$$

with a subspace restriction on set c in GPCA. By maximizing this criterion for $w_k=1 \quad \forall k$ we find common latent variables that are a linear combination of set c and have the largest variance accounted for of all sets. For $w_k = (\text{tr } \Phi_k^2) \quad \forall k$ we maximize the mean proportion of variance of all sets accounted for by the predictor set.

In another context we will find in the next sections examples of applying subspace restrictions for prediction purposes.

4.2 Different filters in one analysis

In the previous sections we applied in GSCA always the same filter for all sets. We now describe other techniques by combining different filters in one analysis. In this way we introduce next to subspace restrictions a second kind of asymmetry in the analysis concerning the type of filters. First we formulate Redundancy Analysis as an example of combining two different filters in one analysis with two sets. Next we give some generalizations of Redundancy Analysis for multiple sets and generalizations of Reduced Rank Regression.

4.2.1 Redundancy Analysis (RA)

We begin with a description of Redundancy Analysis (RA) and next to this we show how RA can be conceived of as a two set GSCA with different filters. In RA we maximize the variance of the criterion set that is accounted for by the canonical variates of the predictor set. Some authors (e.g., van den Wollenberg, 1977) divide the variance accounted for by the number of criterion variables, which are also referred to as criteria. For each dimension r of the solution the variance accounted for is called the *redundancy* of the criteria. The sum of the redundancies is called the *overall redundancy*. We indicate the predictor set with c and the criterion set with k and maximize the overall redundancy of the criteria as follows

$$\begin{aligned} \text{RA}_k(\mathbf{Z}_c) &= \sum_{s=1}^p \eta_{(k)s}^2 = \text{tr } \mathbf{Z}_c' \mathbf{H}_k \mathbf{H}_k' \mathbf{Z}_c \\ &= \text{tr } \mathbf{V}_c \mathbf{P}_c' \mathbf{P}_k \Phi_k^2 \mathbf{P}_k' \mathbf{P}_c \mathbf{V}_c, \end{aligned} \quad (22)$$

where $\mathbf{Z}_c = \mathbf{H}_c \mathbf{T}_c = \mathbf{P}_c \Phi_c \mathbf{Q}_c' \mathbf{T}_c = \mathbf{P}_c \mathbf{V}_c$

denote the canonical variates of the predictor set with $\mathbf{Z}_c' \mathbf{Z}_c = \mathbf{V}_c' \mathbf{V}_c = \mathbf{I}$,

$\mathbf{H}_k = \mathbf{P}_k \Phi_k \mathbf{Q}_k'$ denotes the SVD of the criterion set

and $\eta_{(k)s}^2$ is the redundancy of the m_k criteria for dimension s .

By specifying the filters and subspace restrictions we formulate RA as a two sets GSCA and maximize

$${}^c\text{RA} \quad \left\{ \begin{array}{l} \Omega_k(\Phi_k^2) = \Phi_k^2 \\ \Omega_c(\Phi_c^2) = \mathbf{I}, \mathbf{X}_c = \mathbf{P}_c \mathbf{P}_c' \mathbf{X}_c, \end{array} \right. \quad \text{for } K = 2, c = 1 \text{ or } 2 \text{ and } k = K - c \quad (23)$$

In this way we denote the use of different filters in one analysis accompanied by subspace restrictions.

We still have to show that (23) does the same job as the ordinary formulation of RA. We substitute (23) with $\mathbf{P}_c' \mathbf{X}_c = \mathbf{V}_c$ in (16) and compare it with (22). For $c = 1, k = 2$ and $c = 2, k = 1$ we obtain the equality ${}^c\text{RA} = \mathbf{I} + \text{RA}_k$. Therefore maximization of these two functions gives the same result.

4.2.2 Generalized Redundancy Analysis (GRA)

Generalizations of RA for multiple sets can be formulated by combining two different types of filters in the following way

$${}^c\text{GRA} \quad \left\{ \begin{array}{l} \Omega_k(\Phi_k^2) = \Phi_k^2 w_k^{-1} \\ \Omega_c(\Phi_c^2) = \mathbf{I}, \mathbf{X}_c = \mathbf{P}_c \mathbf{P}_c' \mathbf{X}_c, \end{array} \right. \quad \forall k \neq c \quad (24)$$

where c denotes the predictor set and

$w_1, \dots, w_k, \dots, w_K$ denote fixed balancing constants for criterion set $k \quad \forall k \neq c$

With this model we maximize the variance of the criterion sets that is accounted for by linear combinations of the predictor set. The choice of the balancing constants is discussed extensively in section 2.

4.2.3 Generalized Reduced Rank Regression (GRRR)

We introduce briefly the Reduced Rank Regression model. Originally Izenman (1975) described RRR for the multivariate linear model. Estimation of this model gives the same results as RA (section 4.2.1). We refer to RRR models in this section in a wider sense following De Leeuw & Bijleveld (1987). Other names for the same models are redundancy analysis models, growth curve models, MIMIC (Multiple Indicators/ Multiple Causes) models, or errors-in-variables models. The RRR model is basically a two sets model, where one set of variables, the input set, influences another set of variables, the output set. Other names for the input variables are exogenous or independent variables and for the output variables endogenous or dependent variables. The influence of the input set on the output set is mediated by unobserved latent variables.

We give the generalized model for multiple sets followed by one example. The Generalized Reduced Rank Regression (GRRR) model resembles the GRA model. The subspace

restrictions are omitted compared to GRA and there are several input sets instead of one predictor set. We give

$$\begin{aligned} \text{GRRR } \{ \Omega_k(\Phi_k^2) = \Phi_k^2 w_k^{-1} \} & \quad \text{for } l = 1, \dots, L, \\ \{ \Omega_l(\Phi_l^2) = \mathbf{I} \}, & \quad k = (L+1), \dots, K \quad L < K \end{aligned} \quad (25)$$

where $1, \dots, l, \dots, L$ denote the input sets and

$w_{L+1}, \dots, w_k, \dots, w_K$ denote fixed balancing constants for output set k .

With this model we mediate the influence of the input sets on the output sets by common latent variables \mathbf{X} . The appropriate so called \mathbf{G} and \mathbf{H} matrices for respectively input and output sets can be derived from \mathbf{X} , but this is outside the scope of this report. For $w_k=1 \forall k$ the GRRR model is one of the generalizations of RA for multiple sets suggested by Van de Geer (1984).

One example with two sets is given by taking $L = 1, K = 2$ and $w_2 = 1$. Substituting in (16) we write

$$\text{RRR}(\mathbf{X}) = \text{tr } \mathbf{X}'\mathbf{P}_1\mathbf{P}_1'\mathbf{X} + \text{tr } \mathbf{X}'\mathbf{P}_2\Phi_2^2\mathbf{P}_2'\mathbf{X}. \quad (26)$$

This model is the same as the reformulated reduced rank regression model described by De Leeuw & Bijleveld (1987), with the first set as input variables and the second set as output variables. In fact they create a family of solutions by introducing a weight α^2 for the left part, $\text{tr } \alpha^2 \mathbf{X}'\mathbf{P}_1\mathbf{P}_1'\mathbf{X}$. For the limiting cases of α they prove that (26) is equal to a principal component analysis of the output variables if $\alpha=0$, which can be easily verified by omitting the left part in (26). For $\alpha \rightarrow \infty$ they prove that (26) is equal to RA, which can be understood by realizing that the left part has a absolute maximum of p if the common latent variables are in the space of the input variables, so if $\mathbf{X}_1 = \mathbf{P}_1\mathbf{P}_1'\mathbf{X}_1$. After substitution with $\mathbf{X}_1 = \mathbf{P}_1\mathbf{P}_1'\mathbf{X}_1 = \mathbf{P}_1\mathbf{V}_1$ we recognize in the right part of (26) the formulation of RA in (22). In Van der Burg (1988) our RRR model is described in a comparable way as a generalization of RA by releasing the subspace restriction of the predictor set.

4.3 Compound filters

In this section we discuss the possibility of constructing compound filters by combining two filters in one, separated by a threshold value. We show how preprocessing steps can be incorporated in the analysis by applying this kind of filters. The concept is illustrated by elaborating the practice of replacing a set of variables by an approximation of lower rank. When these low rank approximations are used, we refer to this preprocessing as *shrinkage* techniques.

4.3.1 Shrinkage techniques

Shrinkage techniques are usually applied within the CCA model. The purpose is to eliminate the possibility of finding CCA solutions with very little variance accounted for by the canonical variates. This is achieved by literally eliminating from each set the part of the information that projects on eigenvectors with small eigenvalues. In other words, we are reducing the rank of the set k by replacing $\mathbf{H}_k = \mathbf{P}_k \Phi_k \mathbf{Q}_k'$ by $\underline{\mathbf{H}}_k = \mathbf{P}_k \underline{\Phi}_k \mathbf{Q}_k'$, where $\underline{\Phi}_k$ denotes a diagonal matrix equal to Φ_k , but with eigenvalues below a certain threshold value made equal to zero. After this preprocessing step a CCA is performed in a second step on the matrices $\underline{\mathbf{H}}_k$.

The two step method for the CCA model described above can be compressed in one GSCA step by filtering the eigenvalues of set k in such a way that all eigenvalues below a certain threshold become equal to zero and above this threshold become equal to one. The resulting filter is represented in figure 1 as the *shrinkage constant* filter, together with the *identity* filter and the *first eigenvalue* filter. For the definitions of these filters see (17) and (18).

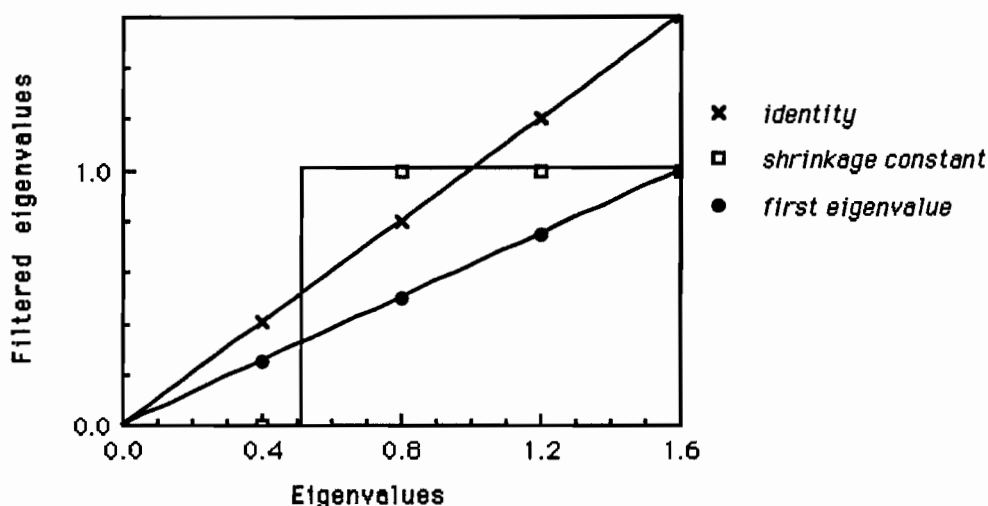


Figure 1 Shrinkage of *constant* filter

The *eigenvalues* are on the horizontal axis and the filtered eigenvalues are on the vertical axis. The identity filter is given by a sloping line with an arbitrary chosen largest eigenvalue of 1.6. The shrinkage constant filter consists of two parts, separated by a threshold. In figure 1 we took a value for the threshold of $0.33 \times \Phi_1^2$. Eigenvalues beneath this threshold are transformed to 0 and above this threshold to 1. The left part approximates the identity filter and the right part is equal to the constant filter. In this way the left part eliminates the small variances and the right part gives the relevant spatial information.

Compound two step filters can also be used for reducing the rank of sets by adapting identity, trace or first eigenvalue filters. It can be expected that the thresholds for this filters have to be relative higher compared to the identity filter to find substantial differences in the solutions, because the maximum shrinkage shift of near zero eigenvalues is for constant filters always equal to 1, while for identity, trace and first eigenvalue filters this maximum is almost zero.

A major drawback in the application of shrinkage techniques is the arbitrariness of the threshold. There are many different methods to find a reasonable value for the threshold. This creates the problem of choosing the appropriate method, maybe even different methods for different sets. It is possible to approach this problem in another way by replacing the two step shrinkage constant filter by a one step filter without a threshold that approximates the shrinkage constant filter. Such a one step filter is described in the next section.

5 Quadratic filters

The GSCA filters described in the previous section can be classified according to the polynomial function used to transform the eigenvalues. We classify the constant filter as a zero degree filter and the identity, trace, first eigenvalue and maxVAF filters as first degree filters. In this section we describe second degree or quadratic filters. We discuss two quadratic filters in the next sections. In fixed SCA we describe a quadratic first eigenvalue filter that approximates the shrinkage constant filter. In SCA we apply a free quadratic filter that results in the maximization of correlations of set variates with variable weights proportional to structure correlations.

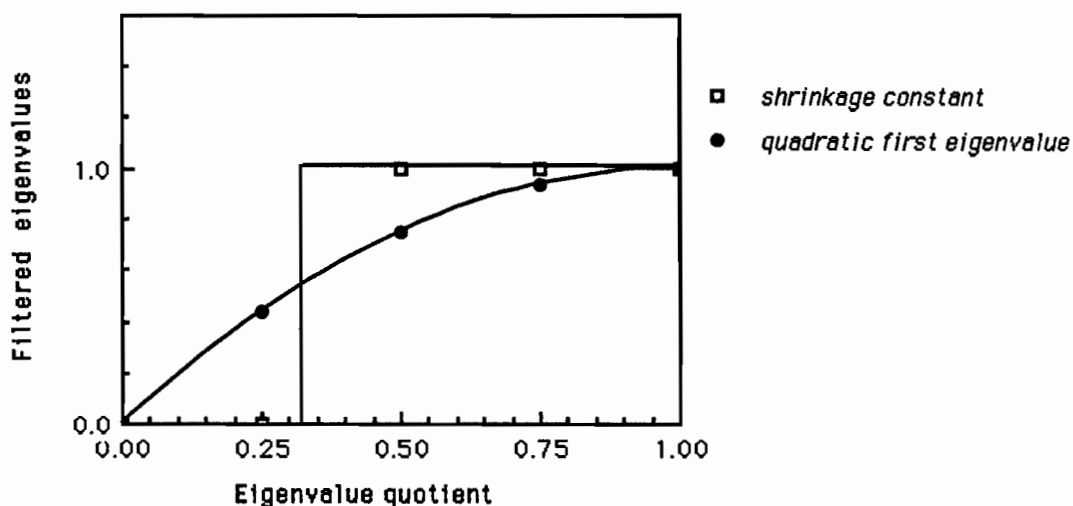


Figure 2 *Quadratic first eigenvalue filter*

5.1 Fixed Set Component Analysis (FSCA)

In FSCA the term 'Fixed' indicates the fact that we actually use a fixed form of the SCA filter. The GSCA filter for FSCA is a *quadratic first eigenvalue* filter

$$\text{FSCA}\{\Omega_k(\Phi_k^2) = \mathbf{I} - (\mathbf{I} - \Phi_k^2 / \phi_{1k}^2)^2\}, \quad \forall k \quad (27)$$

In figure 2 we represent the quadratic first eigenvalue filter to show how this filter approximates the shrinkage constant filter. The representation of the horizontal axis is more general than in figure 1, because the *eigenvalue quotient* gives the eigenvalues divided by the largest eigenvalue. It is clear that the approximation of this specific shrinkage constant filter is rather crude, but we have to bear in mind that the location of the threshold is variable. We require for a general approximation that the filtered eigenvalues are near one at the right side and steeply go down at the left side. In other words we can summarize this technique by calling it a gradual shrinkage technique.

5.2 Set Component Analysis (SCA)

In this section we give a description of SCA by specifying the appropriate GSCA filter. In the next sections we show how SCA maximizes correlations of set variates with proportional weights and how SCA incorporates a penalty on little variance accounted for.

The GSCA filter for SCA is a *free quadratic* filter, which is defined for each set k and each dimension s

$$\text{SCA}\{\Omega_{(k)s}(\Phi_k^2) = \mathbf{I} - (\mathbf{I} - \Phi_k^2 w_{(k)s}^{-1})^2\}, \quad \forall k, s \quad (28)$$

where $w_{(1)s}, \dots, w_{(k)s}, \dots, w_{(K)s}$ denote free balancing factors for set k and dimension s .

The $w_{(k)s}$ are free in the sense that the optimal $w_{(k)s}$ have to be found by maximizing the SCA criterion. After substitution of (28) in (16) we have introduced in the GSCA(\mathbf{X}) function extra parameters to estimate by incorporating the free balancing factors in the filters. By fixing the \mathbf{X} and setting the first derivative equal to zero we find conditional maxima for the balancing factors, which are a function of \mathbf{X} . The balancing factors are made equal to the resulting conditional maxima

$$w_{(k)s} = \frac{\mathbf{x}_s' \mathbf{P}_k \Phi_k^4 \mathbf{P}_k' \mathbf{x}_s}{\mathbf{x}_s' \mathbf{P}_k \Phi_k^2 \mathbf{P}_k' \mathbf{x}_s} = \frac{\mathbf{b}_{(k)s}' \Phi_k^2 \mathbf{b}_{(k)s}}{\mathbf{b}_{(k)s}' \mathbf{b}_{(k)s}}, \quad \forall k, s \quad (29)$$

with $\mathbf{b}_{(k)s} = \Phi_k \mathbf{P}_k' \mathbf{x}_s \quad \forall k, s$. The upper bound of $w_{(k)s}$ is equal to the largest eigenvalue of matrix Φ_k^2 , which is the first eigenvalue ϕ_{1k}^2 . In FSCA we fixed the balancing factors equal to this upper bound. The lower bound of $w_{(k)s}$ is equal to zero, because Φ_k^2 is a positive semi-definite matrix. Resuming we have

$$0 \leq w_{(k)s} \leq \phi_{1k}^2. \quad \forall k, s \quad (30)$$

In figure 3 we represent the free quadratic filter for $w_{(k)s} = 0.33 \times \phi_{1k}^2$ and for the upper bound $w_{(k)s} = \phi_{1k}^2$.

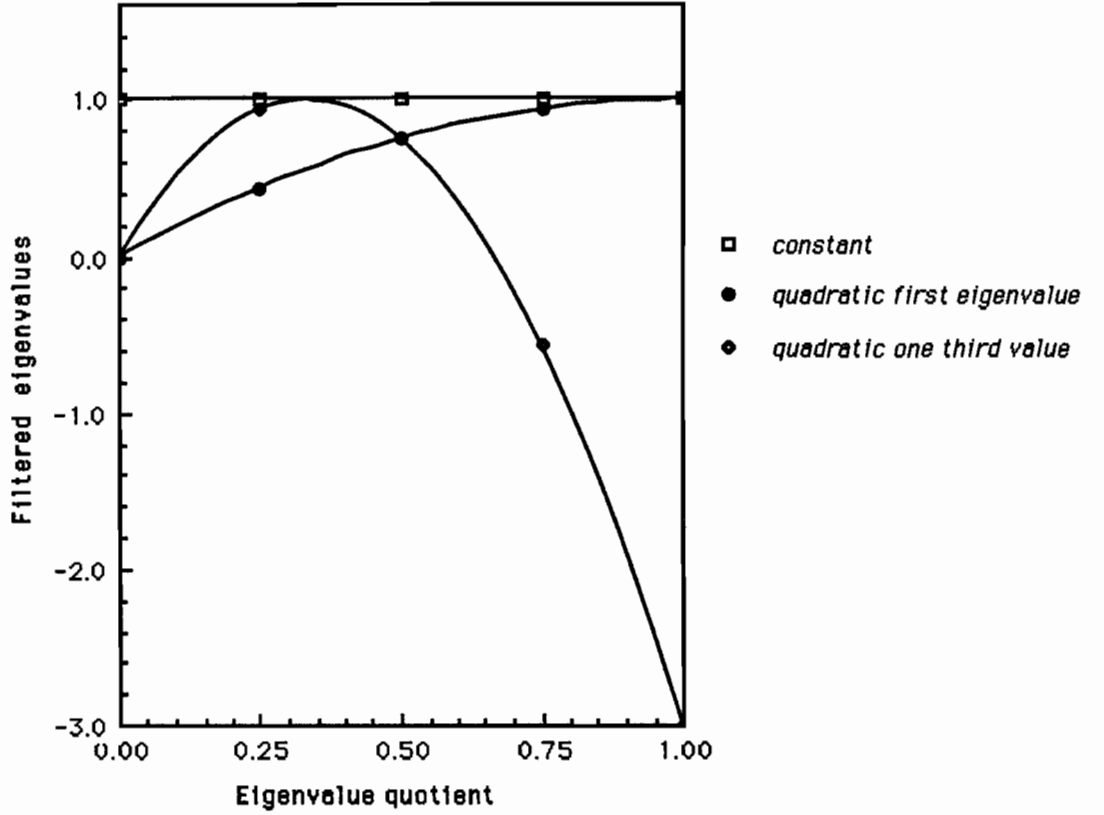


Figure 3 Free quadratic filter

The filtered quadratic eigenvalue in figure 3 is maximal, when the eigenvalue quotient is equal to $w_{(k)s} / \phi_{1k}^2$.

5.2.1 Correlations of set variates with proportional weights

We show how SCA maximizes correlations of set variates by substituting (29) in (28) and then in (16). We obtain

$$\begin{aligned}
 \text{SCA}(\mathbf{x}_s) &= \sum_{s=1}^p \sum_{k=1}^K \frac{(\mathbf{x}_s' \mathbf{P}_k \Phi_k^2 \mathbf{P}_k' \mathbf{x}_s)^2}{\mathbf{x}_s' \mathbf{P}_k \Phi_k^4 \mathbf{P}_k' \mathbf{x}_s} \\
 &= \sum_{s=1}^p \sum_{k=1}^K \frac{(\mathbf{x}_s' \mathbf{H}_k \mathbf{H}_k' \mathbf{x}_s)^2}{\mathbf{x}_s' \mathbf{H}_k \mathbf{H}_k' \mathbf{H}_k \mathbf{H}_k' \mathbf{x}_s} \\
 &= \sum_{s=1}^p \sum_{k=1}^K (\mathbf{x}_s' \mathbf{z}_{(k)s})^2, \tag{31}
 \end{aligned}$$

with $\mathbf{z}_{(k)s} = \mathbf{H}_k \mathbf{H}_k' \mathbf{x}_s / (\mathbf{x}_s' \mathbf{H}_k \mathbf{H}_k' \mathbf{H}_k \mathbf{H}_k' \mathbf{x}_s)^{1/2} \forall k, s$. So we maximize the sum the squared set correlations of the unit normalized set variates $\mathbf{z}_{(k)s}$ with the common latent variables \mathbf{x}_s .

If we compare SCA in (31) with GCCA in (10), the only difference we observe is the definition of $\mathbf{z}_{(k)s} = \mathbf{H}_k \mathbf{t}_{(k)s} \forall k, s$. In SCA are for each set k the weights of variables $\mathbf{H}_k' \mathbf{x}_s / (\mathbf{x}_s' \mathbf{H}_k \mathbf{H}_k' \mathbf{H}_k \mathbf{H}_k' \mathbf{x}_s)^{1/2}$ proportional to the structure correlations $\mathbf{H}_k' \mathbf{x}_s$. In GCCA we don't have these restrictions for the weights $\mathbf{t}_{(k)s}$.

The GCCA model gives the upper bounds for the SCA, as we can show by defining the the right part of (28) as a penalty function

$$\text{PEN}\{\Omega_{(k)s}(\Phi_k^2) = (\mathbf{I} - \Phi_k^2 \mathbf{w}_{(k)s}^{-1})^2\} \quad \forall k, s \quad (32)$$

By substitution of the appropriate filters in (16) we obtain $\text{SCA} = \text{GCCA} - \text{PEN}$, where PEN gives a penalty related with the size of the relocation step of the set variates needed to obtain a larger variance accounted for.

5.2.2 Penalty on little variance accounted for

In this section we explain how SCA gradually prevents the occurrence of little variance accounted for by *differential weighting* of the projected latent variable onto the subspaces spanned by the sets. For each set k we express the variance accounted for $\text{VAF}(k)$ as a product of the correlations of set variates and the balancing factors by substituting (29) in (31). We write

$$\text{VAF}(k) = \rho_{(k)s}^2 \mathbf{w}_{(k)s}, \quad \forall k, s \quad (33)$$

where $\rho_{(k)s}^2 = (\mathbf{x}_s' \mathbf{z}_{(k)s})^2$ and $\text{VAF}(k) = \mathbf{x}_s' \mathbf{H}_k \mathbf{H}_k' \mathbf{x}_s \forall k, s$.

If we want to prevent little variances accounted for, we must not only maximize the correlations of set variates $\rho_{(k)s}^2$, but we must also prevent small values for the balancing factors $\mathbf{w}_{(k)s}$. The latter is obtained by differential weighting of the projected latent variable onto the subspaces spanned by the sets. The available projection space of the common latent variable \mathbf{x}_s for high $\mathbf{w}_{(k)s}$ values is gradually reduced if $\mathbf{w}_{(k)s}$ gets smaller, because the penalty for projecting on eigenvectors with large eigenvalues is increasing fast. As we see in figure 3 for $\mathbf{w}_{(k)s} = 0.33 \times \phi_{1k}^2$ the penalty weight for projection on the first eigenvector is already -4. For smaller balancing factor values this penalty weight increases fast.

5.3 Computation

The estimation of parameters in all presented models gives no major problems. For all models we need to find largest eigenvalue solutions. Only in case of the SCA model we have to formulate an iterative procedure, where we alternate between the estimation of \mathbf{X} in (28) and the estimation of the balancing factors $\mathbf{w}_{(k)s}$ in (29). This iteration scheme will be described in a forthcoming publication.

References

- Carroll, J. D. (1968). A generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th annual convention of the American Psychological Association*, 3, 227-228.
- de Leeuw, J. & Bijleveld, C. (1987). *Fitting reduced rank regression models by alternating least squares*. (Report RR-87-05). Leiden: University of Leiden, Department of Data Theory.
- Gifi, A. (1981). *Nonlinear multivariate analysis*. Leiden: University of Leiden, Department of Data Theory.
- Gittins, R. (1985). *Canonical analysis, a review with applications in ecology*. Berlin: Springer-Verlag.
- Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5, 248-264.
- Meulman, J. J. (1986). *A distance approach to nonlinear multivariate analysis*. Leiden: DSWO Press.
- van de Geer, J. P. (1984). Linear relations among K sets of variables. *Psychometrika*, 49, 79-94.
- van den Wollenberg, A. L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, 42, 207-219.
- van der Burg, E. (1988). *Nonlinear canonical correlation and some related techniques*. Leiden: DSWO Press.