Common and distinct components in data fusion

Smilde, A.K.; Mage, I.; Naes, T.; Hankemeier, T.; Lips, M.A.; Kiers, H.A.L.; ... ; Bro, R.

Common and Distinct Components in Data Fusion

Age K. Smilde[1†], Ingrid Måge[2], Tormod Naes[2], Thomas Hankemeier[3], Mirjam A. Lips[4], Henk A.L. Kiers[5], Evrim Acar[6] and Rasmus Bro[6]

[1] Biosystems Data Analysis, Faculty of Sciences, University of Amsterdam, Science Park 904, P.O. Box 94215, 1090 GE Amsterdam, The Netherlands

[2] Nofima, Ås, Norway

[3] LACDR, Leiden University, Netherlands

[4] Leiden University Medical Center, Department of Endocrinology and Metabolism, Netherlands

[5] Heymans Institute, University of Groningen, Netherlands

[6] Department of Food Science, University of Copenhagen, Denmark

Abstract

In many areas of science multiple sets of data are collected pertaining to the same system. Examples are food products which are characterized by different sets of variables, bio-processes which are on-line sampled with different instruments, or biological systems of which different genomics measurements are obtained. Data fusion is concerned with analyzing such sets of data simultaneously to arrive at a global view of the system under study. One of the upcoming areas of data fusion is exploring whether the data sets have something in common or not. This gives insight into common and distinct variation in each data set, thereby facilitating understanding the relationships between the data sets. Unfortunately, research on methods to distinguish common and distinct components is fragmented, both in terminology as well as in methods: there is no common ground which hampers comparing methods and understanding their relative merits. This paper provides a unifying framework for this subfield of data fusion by using rigorous arguments from linear algebra. The most frequently used methods for distinguishing common and distinct components are explained in this framework and some practical examples are given of these methods in the areas of (medical) biology and food science.

Keywords: DISCO, JIVE, O2PLS, GSVD

---

[1] Author for correspondence. e-mail a.k.smilde@uva.nl

1

1 Introduction and Motivation

1.1 Data fusion

Simultaneous analysis of several data blocks has been proposed a long time ago [20, 60], but today we can see a renewed interest fueled by the strongly increasing needs in many sciences. A number of different methods have been put forward [3, 59, 67, 24, 2] all of them with a common interest of either understanding relations better or obtaining better prediction results. The methodologies are known under different names in different disciplines, important examples being data fusion, data integration, multi-block analysis, multi-set analysis and multi-mode analysis (for definitions, see [69, 22]). Some of the methods are rather straightforward generalizations of standard methods for one or two data sets such as concatenated PCA and PLS regression [70], while others are explicitly developed for handling multi-block data focusing on a number of concepts unique for such applications. In the latter group one can find methods such as SO-PLS and PO-PLS [31], DISCO-SCA [40], O2PLS [59] and GSVD [16].

This paper will focus on one particular aspect that appears crucial in data fusion, namely the distinction between common and distinct information in the blocks. The main aim is to provide concrete definitions of the concepts and to discuss how these definitions relate to the most well known methods in the area. Main attention will be given to interchangeable data blocks sharing the row-mode which usually consists of samples or subjects; thus multi-block predictive methods such as SO-PLS, PO-PLS [31] are not discussed. We will also restrict ourselves to direct analysis in contrast to indirect analysis such as analyzing covariance or correlation matrices. Focus will be on definitions based on column spaces: the spaces spanned by object scores on the variables but interpretation in terms of variable loadings (i.e. the row-space) will also be given some attention. Selected methods will be illustrated by real data sets. Situations with two blocks as well as situations with more than two blocks will be discussed. As an integral part of the discussion, we will also incorporate relative measures of fit of the different parts of the blocks.

1.2 Motivating examples

1.2.1 Food Science

In food product development we are typically interested in understanding how product formulations (ingredients etc.) of a set of product prototypes are related to the descriptive sensory properties of the product and also possibly to the consumer liking of the product. A typical situation might be that one is interested in substituting one of the ingredients by a cheaper one and is interested in seeing whether this change has any noticeable effect on the smell, the taste, the texture or all of them. Another typical situation is in new product development where the developer wants to understand how two important sensory modalities such as smell and taste are affected by the ingredients used. In both cases, it is crucial for further product optimization to know how this happens, for instance whether the smell and taste have a joint source of variability and/or what is influencing only one of them.

1.2.2 Biology

An important class of health problems is Diabetes Mellitus Type II (DM2). Consider measurements performed on a group of DM2 patients using a metabolomics platform (e.g. LC-MS), clinical measurements (such as insulin resistance, fasting glucose levels, blood pressure) and life-style variables. Then these measurements will have parts in common and have distinctive parts. The common part between the metabo-lomics and clinical measurements may reflect the relation between branched amino-acids and insulin resistance [28]; there may also be common parts between the life-style variables and the clinical measurements, such as exercise and blood pressure. Some of the metabolites, such as bile acids, may not be directly related to insulin resistance and life-style and will, hence, be distinct. Since all measurements pertain to the same system (DM2) it is worthwhile exploring and understanding the complete data set in a holistic way.

1.2.3 General idea

The above two examples show common features which are summarized in Figure 1. Knowledge is required of a complex system (first and upper layer; e.g. DM2). Measurements are performed on this system resulting in three blocks of data $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ (second layer; e.g. metabolomics, clinical

3

and life-style measurements in the DM2 example; smell, taste and consumer liking in the food science example). These measurements are preferably collected in such a way that *diversity* is increased ([42, 22]. Although diverse and information-rich data is obtained, the problem is that the data blocks contain partly overlapping contributions of parts A, B and C of the system (e.g. A is insulin-glucose-amino-acid metabolism and B reflects cardiocasvular complications in the DM2 example; sweetness (A) in the case of taste and smell in the food science example) and also irrelevant variation and noise. The idea behind finding common and distinct variation in the three data blocks is to separate and quantify the different sources of variation which are spread across all data blocks (third layer). Interpreting the different sources of variation will then lead to a reconstruction of the system (fourth and bottom layer; e.g. the etiology of DM2). In our paper, we will mainly describe moving from the second to the third layer (the boxed part), and will only touch upon moving from the third to the fourth layer. In Section 4 we will present some real-life examples which were already introduced above.

2   General mathematical framework

For the definition of the basic concepts we will start with two data matrices or blocks $\mathbf{X}_1$ of size $(I \times J_1)$ and $\mathbf{X}_2$ of size $(I \times J_2)$ and afterwards discuss how these concepts can be extended to three or several blocks of data. It is assumed that the two matrices share the first mode (the I-mode, [51]) usually representing samples or objects and the data have been column-centered throughout. Note that the two data sets may have different number of columns, usually representing variables, which means that they may (and often will) contain different types of measurements.

This section will be devoted to precise definitions of common and distinct components for the blocks in the data set. All these definitions are inspired by and related to previous definitions, but the main aim here is to make the definitions precise and unambiguous and therefore better suited for comparing methodologies. The definitions will be made in terms of subspaces, but later on we will expand to discuss the same concepts in terms of components which are basis vectors, chosen in one way or another, for the subspaces.

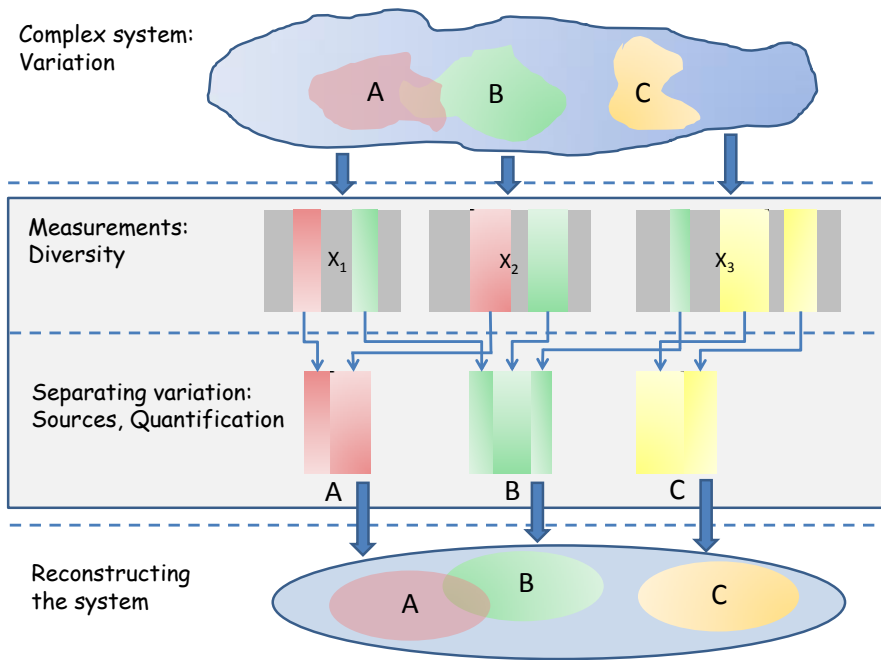The mathematical framework represents the idealized situation of noiseless

Figure 1: Measurements are performed on a complex system probing parts A, B and C of that system. The resulting data blocks $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ contain mixed variation which has to be separated in sources (red/green is common; yellow is distinct; grey is irrelevant variation and noise). These quantified sources are then used to reconstruct the system. This paper concerns mainly the box within the blue lines.

data. In practice, of course, this never happens. Hence, in later Sections we are also going to discuss which kind of compromises and choices have to be made in real-life situations. In that context, we also discuss several existing methods for finding common and distinct subspaces as used in the psychometrics, bioinformatics, chemometrics, computer science, data analysis and statistics literature.

2.1 Description of the framework

2.1.1 The two-block case

The two spaces spanned by the columns of $\mathbf{X}_1$ and $\mathbf{X}_2$ ($R(\mathbf{X}_1)$ and $R(\mathbf{X}_2)$) are located in the same I-dimensional column-space R^I, see Figure 2 for an illustration in three dimensional space. Each variable is a vector in this co-

ordinate system indicating the level of that variable for each sample (row). These variables are not explicitly shown in this figure but will lie within the space indicated by the blue and green column-spaces.
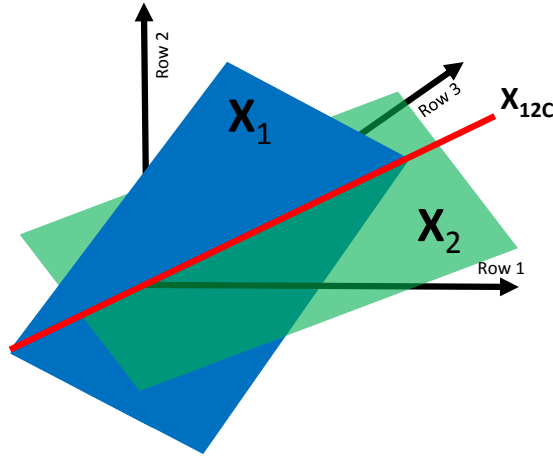

Figure 2: The I-dimensional space having $R(\mathbf{X}_1)$ (blue) and $R(\mathbf{X}_2)$ (green) as subspaces. Only three axis of this I-dimensional space are drawn. The red line $\mathbf{X}_{12C}$ represents the common subspace. For the sake of illustration the dimensions of both column-spaces are equal (two). This is not necessarily always the case.

If the two column-spaces intersect non-trivially (the zero is always shared), then the intersection space is called the common space. In Figure 2 there is only one common direction (i.e. the common space is one-dimensional), but there can be more or none. The common subspace will be called $R(\mathbf{X}_{12C})$ where the subscript C stands for 'Common'. Note that $R(\mathbf{X}_{12C}) \subseteq R(\mathbf{X}_1)$ and $R(\mathbf{X}_{12C}) \subseteq R(\mathbf{X}_2)$. The common part of the two blocks will in most cases not span the whole of $R(\mathbf{X}_1)$ and $R(\mathbf{X}_2)$. Some definitions regarding the rest of these spaces are therefore needed. As will be discussed later, it is useful to distinguish between different ways of representing these subspaces, depending on choices regarding orthogonality. In all cases, these subspaces representing the rest after identification of the common part will be called "distinct" subspaces. The requirement is that the space spanned by the columns in a block $\mathbf{X}_k(k = 1, 2)$ is a direct sum of the common space and

6

the distinct space within that block. Hence, these two parts within a block are linearly independent (two subspaces are linearly independent if no vector in one subspace can be written as a linear combination of the vectors of the other and vice versa).
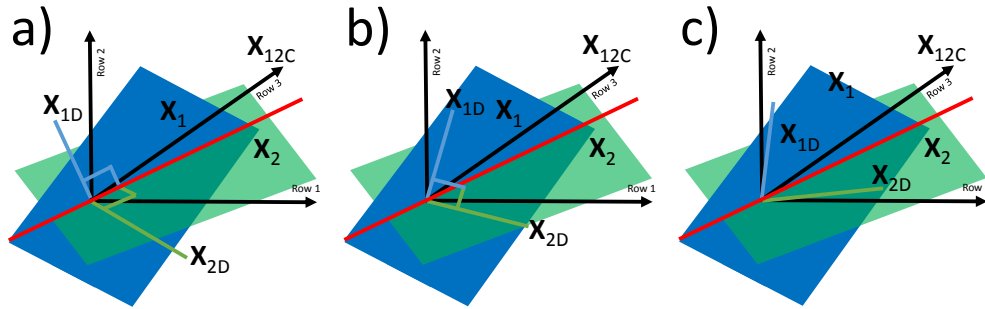

Figure 3: See also legend Figure 2. The distinct subspaces (1-dimensional in this case) are spanned by $\mathbf{X}_{1D}$ and $\mathbf{X}_{2D}$ for $R(\mathbf{X}_1)$ and $R(\mathbf{X}_2)$, respectively; a) both distinct subspaces are chosen orthogonal to the common subspace, b) both distinct subspaces are chosen mutually orthogonal, c) no orthogonality.

These subspaces are called $R(\mathbf{X}_{1D})$ and $R(\mathbf{X}_{2D})$ where the subscript D stands for 'Distinct'. The choice whether or not to choose orthogonality depends on the application. In Figure 3 three possibilities are shown, namely making the distinct subspaces orthogonal to the common subspace or making the distinct subspaces orthogonal to each other or imposing no orthogonality at all. In general, it is not possible to combine the orthogonalities of Figure 3a and b.

What we have accomplished now is decomposing $R(\mathbf{X}_1)$ and $R(\mathbf{X}_2)$ into

7

direct sums of spaces:

$$
\begin{aligned}
R(\mathbf{X}_1) &= R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{1D}) \\
R(\mathbf{X}_2) &= R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{2D})
\end{aligned}
\tag{1}
$$

because $R(\mathbf{X}_{12C}) \cap R(\mathbf{X}_{1D}) = \{0\}$ and $R(\mathbf{X}_{12C}) \cap R(\mathbf{X}_{2D}) = \{0\}$ [39]. Hence, it also holds that

$$
\begin{aligned}
dimR(\mathbf{X}_1) &= dimR(\mathbf{X}_{12C}) + dimR(\mathbf{X}_{1D}) \\
dimR(\mathbf{X}_2) &= dimR(\mathbf{X}_{12C}) + dimR(\mathbf{X}_{2D})
\end{aligned}
\tag{2}
$$

If the distinct-orthogonal-to-common option is chosen (see Figure 3a), then additionally it holds that $R(\mathbf{X}_{12C}) \perp R(\mathbf{X}_{1D})$ and $R(\mathbf{X}_{12C}) \perp R(\mathbf{X}_{2D})$. Note that for this case, given the common space, the decomposition is unique because then $R(\mathbf{X}_{1D})$ is the orthogonal complement of $R(\mathbf{X}_{12C})$ within $R(\mathbf{X}_1)$ and likewise for $R(\mathbf{X}_{2D})$ (but not necessarily the basis within the subspaces if these have dimension higher than one). In the non-orthogonal case, the distinct part can be defined by any set of linearly independent vectors that are in the original spaces, but not in the common space. For a thorough description of direct sums of spaces, see [72].

We can take it one step further by also decomposing both the distinct subspaces $R(\mathbf{X}_{1D})$ and $R(\mathbf{X}_{2D})$ in two parts:

$$
\begin{aligned}
R(\mathbf{X}_{1D}) &= R(\mathbf{X}_{1DO}) \oplus R(\mathbf{X}_{1DNO}) \\
R(\mathbf{X}_{2D}) &= R(\mathbf{X}_{2DO}) \oplus R(\mathbf{X}_{2DNO})
\end{aligned}
\tag{3}
$$

where $R(\mathbf{X}_{1DO})$ is the "distinct-orthogonal (DO)" part and the other part will be called "distinct-non-orthogonal" (DNO); where $R(\mathbf{X}_{1DO}) \perp R(\mathbf{X}_{2DO})$ and $R(\mathbf{X}_{1DNO})$ is the remaining part of $R(\mathbf{X}_{1D})$ after removing $R(\mathbf{X}_{1DO})$ and likewise for $R(\mathbf{X}_{2DNO})$. Again, by the definition of direct sum we have $R(\mathbf{X}_{1DO}) \cap R(\mathbf{X}_{1DNO}) = \{0\}$ and $R(\mathbf{X}_{2DO}) \cap R(\mathbf{X}_{2DNO}) = \{0\}$. The argument for the split of Eqn. 3 is that one may be interested in looking at the parts of the blocks that have no correlation with (parts of) each other at all. Note that such an additional split can only be performed when the dimensions of the subspaces allow so, e.g., in Figure 3 both distinct subspaces have only dimension one and thus cannot be decomposed further. Depending on the dimensions of the distinct subspaces and their relative positioning in space, different possibilities can be distinguished. A choice has to be made

8

by the user and is application dependent. A summary of alternatives is presented in the Appendix (Section 7.1) but one example is the following. If $\mathbf{X}_1$ and $\mathbf{X}_2$ contain measurements of two instruments, then choosing $R(\mathbf{X}_{2DO})$ orthogonal to the whole of $R(\mathbf{X}_1)$ can be interpreted as the unique contribution of instrument 2 relative to instrument 1 or, stated differently, what is the gain by adding instrument 2?

Summarizing, we arrive at the following direct sum decompositions of the column-spaces of $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$
\begin{aligned}
R(\mathbf{X}_1) &= R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{1D}) = R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{1DO}) \oplus R(\mathbf{X}_{1DNO}) \quad (4) \\
R(\mathbf{X}_2) &= R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{2D}) = R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{2DO}) \oplus R(\mathbf{X}_{2DNO})
\end{aligned}
$$

representing our general definition of the basic concepts of common (C), distinct (D), distinct-orthogonal (DO) and distinct-non-orthogonal (DNO) subspaces. This decomposition is unique meaning that when the decomposition of Eqn. 4 is chosen, then every vector in $R(\mathbf{X}_1)$ can be written uniquely as a sum of three vectors in the three different subspaces $R(\mathbf{X}_{12C})$, $R(\mathbf{X}_{1DO})$ and $R(\mathbf{X}_{1DNO})$ and likewise for $R(\mathbf{X}_2)$, if the dimensions allow so.

The decomposition of Eq. 4 gives also a break-down of the dimensions of the separate subspaces:

$$
\begin{aligned}
dimR(\mathbf{X}_1) &= dimR(\mathbf{X}_{12C}) + dimR(\mathbf{X}_{1D}) \quad (5) \\
&= dimR(\mathbf{X}_{12C}) + dimR(\mathbf{X}_{1DO}) + dimR(\mathbf{X}_{1DNO}) \\
dimR(\mathbf{X}_2) &= dimR(\mathbf{X}_{12C}) + dimR(\mathbf{X}_{2D}) \\
&= dimR(\mathbf{X}_{12C}) + dimR(\mathbf{X}_{2DO}) + dimR(\mathbf{X}_{2DNO}).
\end{aligned}
$$

2.1.2 Generalizations to three blocks

The generalization to three blocks of data goes as follows. Consider the sets $\mathbf{X}_1(I \times J_1)$, $\mathbf{X}_2(I \times J_2)$ and $\mathbf{X}_3(I \times J_3)$. We can define again a part which is in common between all three column-spaces, $R(\mathbf{X}_{123C})$ with obvious notation. Next, we can define a part in common between $R(\mathbf{X}_1)$ and $R(\mathbf{X}_2)$ which is not intersecting with $R(\mathbf{X}_3)$, $R(\mathbf{X}_{12C})$, and likewise we can define $R(\mathbf{X}_{13C})$ and $R(\mathbf{X}_{23C})$. The complete part of $R(\mathbf{X}_1)$ which is shared with the other blocks can then be written as $R(\mathbf{X}_{123C}) \oplus R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{13C})$ with the properties that $R(\mathbf{X}_{123C}) \cap R(\mathbf{X}_{12C}) = \{0\}$, $R(\mathbf{X}_{123C}) \cap R(\mathbf{X}_{13C}) = \{0\}$

9

and $R(\mathbf{X}_{12C}) \cap R(\mathbf{X}_{13C}) = \{0\}$.

The distinct part of $R(\mathbf{X}_1)$ can again be defined as the part of $R(\mathbf{X}_1)$ linearly independent of $R(\mathbf{X}_{123C}) \oplus R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{13C})$. This leads to the following decomposition:

$$R(\mathbf{X}_1) = R(\mathbf{X}_{123C}) \oplus R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{13C}) \oplus R(\mathbf{X}_{1D}) \qquad (6)$$

and the distinct part $R(\mathbf{X}_{1D})$ can again be broken down in several parts. The first part may be chosen to be the subspace of $R(\mathbf{X}_{1D})$ orthogonal to $R(\mathbf{X}_2) \cup R(\mathbf{X}_3)$ with obvious notation $R(\mathbf{X}_{1DO23})$. Then there is a part orthogonal to only $R(\mathbf{X}_2)$, $R(\mathbf{X}_{1DO2})$, and a part only orthogonal to only $R(\mathbf{X}_3)$, $R(\mathbf{X}_{1DO3})$, where again $R(\mathbf{X}_{1DO23}) \cap R(\mathbf{X}_{1DO2}) = \{0\}$ and $R(\mathbf{X}_{1DO23}) \cap R(\mathbf{X}_{1DO3}) = \{0\}$. Hence, the full decomposition of $R(\mathbf{X}_1)$ becomes

$$R(\mathbf{X}_1) = R(\mathbf{X}_{123C}) \oplus R(\mathbf{X}_{12C}) \oplus R(\mathbf{X}_{13C}) \oplus R(\mathbf{X}_{1DO23}) \qquad (7)$$
$$\oplus R(\mathbf{X}_{1DO2}) \oplus R(\mathbf{X}_{1DO3}) \oplus R(\mathbf{X}_{1DNO})$$

that represents the most elaborate decomposition of $R(\mathbf{X}_1)$ if all dimensions allow so with different possibilities for orthogonalities. Because of the direct sum properties the dimensions add up in the same way as in Eq. 2 and 5. Similar decompositions can be made for $R(\mathbf{X}_2)$ and $R(\mathbf{X}_3)$. Schematically, the decomposition of Eqn. 7 is shown in Figure 4.

Equations 4, 6 and 7 show an increasing degree of complexity. We give here the full decompositions to be complete, but it is important to mention that in most practical cases, one is not interested in all these subspaces making the actual practical decomposition simpler. This is even more so in cases with more than three blocks.

2.2 Theoretical considerations

In practice, data always contain noise and therefore we cannot expect to find a decomposition that satisfies all the idealistic requirements described above. Also, which decomposition to make under which constraints depends very much on the type of application. Before showing how various types of already existing methods try to solve this challenge, we will here discuss some of the major issues that have to be taken into account. These issues represent choices which have to be made regarding the nature of the common
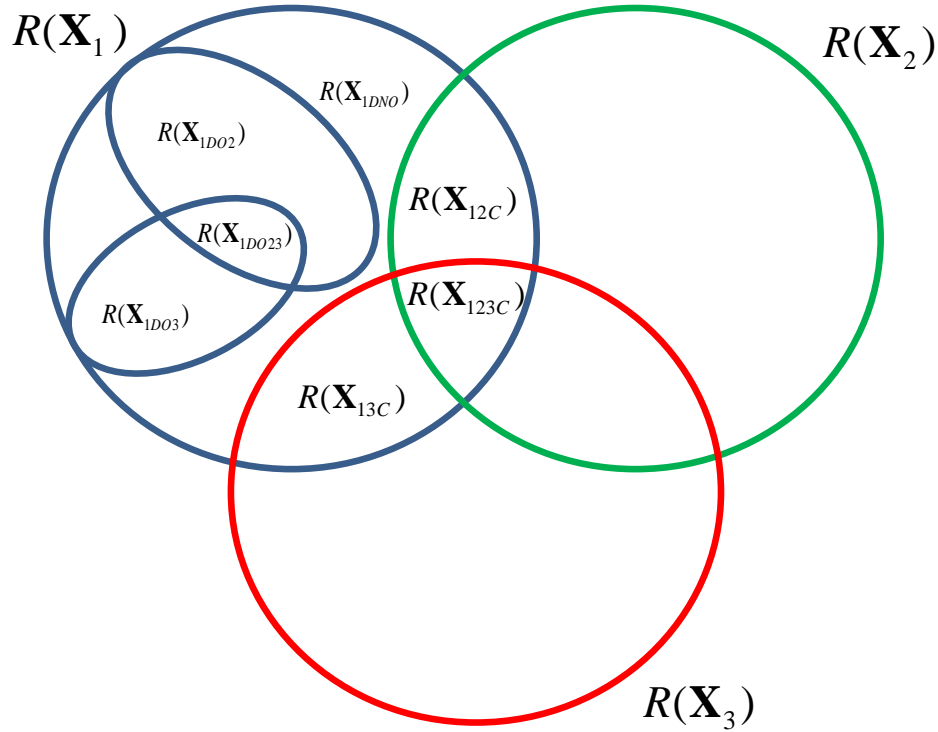
Figure 4: The decomposition of $R(\mathbf{X}_1)$ in the three block situation.

and distinct components, diagnostic tools such as explained sum-of-squares, scaling of the variables and the data sets.

2.2.1 Fundamentally different choices of common components.

Of particular importance here is the concept of common variation because it can be considered as a starting point of the decomposition. Since practical implementations are usually based on extracting components or basis vectors for the different spaces, most of the following discussions will be related to components rather than to general vector spaces as was the case above.

In noisy data, the situation as shown in Figure 2 does not usually hold: there is no common space in mathematical terms (an intersection) because the column-spaces have changed due to the noise. There are two fundamen-
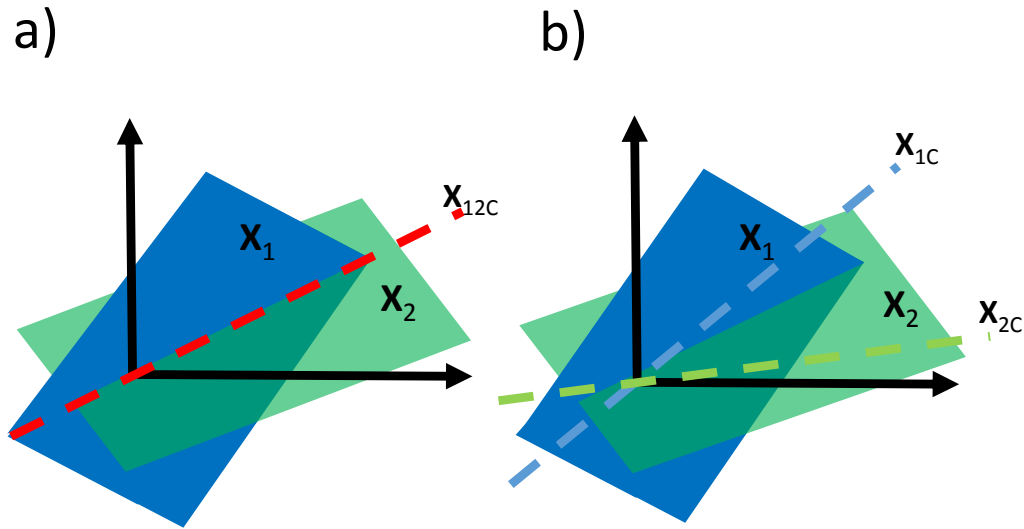
Figure 5: The common subspace under noisy conditions. It can be chosen as a compromise in-between $R(\mathbf{X}_1)$ and $R(\mathbf{X}_2)$ but not a part of neither of those (red dashed line, a) or as parts of their respective column-spaces but unequal to each other (blue and green lines, b).

tally different categories of approaches and these are present in the methods that are discussed in Section 3. In the first category, a common component is found as the best compromise solution between the two column-spaces: vector $\mathbf{X}_{12C}$ in Figure 5a (although it is customary to use a bold-lowercase character for a vector, we keep the notation using a matrix to stress the fact that we are generally discussing subspaces). This vector is neither in the column-space of $\mathbf{X}_1$ nor in the column-space of $\mathbf{X}_2$. In the second category, a different choice is made. The common component is estimated separately in each column-space. Hence, rather than one common component, two separate ones are found but generally in a manner that seeks them to be as similar as possible(although $\mathbf{X}_{1C} \neq \mathbf{X}_{2C}$; Figure 5b) and thus they can be

seen as representing a common component. Both choices are made in the methods to be discussed and both approaches have their pros and cons (see Table 3 for more details). Note the change in notation of the common parts to emphasize this difference.

2.2.2 Sums of Squares and Explained Variation.

When it comes to assessing the importance of a subspace in the decomposition there are at least two aspects that have to be taken into account; the dimension of the subspaces identified and variances explained by those subspaces in the original data. The former relates to how many linearly independent components are estimated to form the subspace. The latter relates to the contributions of the subspaces to the total variation in a block. If orthogonality is used in the decomposition when defining the distinct space (Figure 3a), it is easy to show that the total sum of squares (SS) for a block can be split in one contribution from the common space ($R(\mathbf{X}_{12C})$) and one for the orthogonal distinct contribution ($R(\mathbf{X}_{1D})$):

$$\|\mathbf{X}_1\|^2 = \|\mathbf{X}_{12C}\|^2 + \|\mathbf{X}_{1D}\|^2 \tag{8}$$

where we use the symbol $\|.\|^2$ to indicate the squared Frobenius norm of a matrix. An analogous equation can be written for $\mathbf{X}_2$. If orthogonality is not imposed between the common and distinct parts (Figure 3b), a decomposition of SS is still possible, but the interpretation of the last term is different. In that case, it is simply defined as the additional variation that is explained by adding the distinct components, i.e. as $\|\mathbf{X}_1\|^2 = \|\mathbf{X}_{12C}\|^2 + \|\widetilde{\mathbf{X}}_{1D}\|^2$, where $\widetilde{\mathbf{X}}_{1D}$ is the part of $R(\mathbf{X}_{1D})$ orthogonal to $R(\mathbf{X}_{12C})$. This is sometimes called Extra Sum of Squares (ESS; see also [34]). Note that the order in which the terms are calculated in Eqn. 8 matters in the non-orthogonal case. For the orthogonal case, the two interpretations coincide.

When decomposing the distinct part further into an orthogonal and a non-orthogonal part, the resulting (E)SS can be written as

$$\|\mathbf{X}_1\|^2 = \|\mathbf{X}_{12C}\|^2 + \|\mathbf{X}_{1DO}\|^2 + \|\mathbf{X}_{1DNO}\|^2 \tag{9}$$

and the interpretation depends on the orthogonality properties. In the most extreme case, all subspaces $R(\mathbf{X}_{12C})$, $R(\mathbf{X}_{1DO})$ and $R(\mathbf{X}_{1DNO})$ are orthogonal to each other, then Eqn. 9 can be interpreted in terms of sums of squares

of contributions of each block. In all other cases, Eqn. 9 has an ESS interpretation, e.g., when $R(\mathbf{X}_{1DO})$ and $R(\mathbf{X}_{1DNO})$ are not orthogonal, then $\|\widetilde{\mathbf{X}}_{1DNO}\|^2$ is the ESS of the distinct-non-orthogonal part where $R(\widetilde{\mathbf{X}}_{1DNO})$ is orthogonalized relative to $R(\mathbf{X}_{1DO})$. Explained variation of common or distinct components within a block can now be calculated and expressed as percentages of the total variation in that block. Note that this process is analogous to the Type I ANOVA where focus is on additional contribution of variables in explaining a Y-variable and note also the similarity to SO-PLS in a multi-block regression context [31].

The issue of variance explained by common components in a block is visualized in Figure 6 for the second category of methods. The column vectors making up the column-spaces of $\mathbf{X}_1$ and $\mathbf{X}_2$ are explicitly drawn in the figure. In the left (a) part all these column vectors (i.e. variables) are close to the common components within each block. Hence, the common components are representative of their respective column-spaces: they are embedded well and explain a high amount of variation in each block. This is not the case for the right (b) part of the figure: the common component $\mathbf{X}_{1C}$ is not well embedded in $\mathbf{X}_1$. Usually, explaining within-block variation and having between-block correlation cannot be achieved simultaneously and a good account of this trade-off is given elsewhere [60].

2.2.3 High-dimensional data.

High-dimensional data need some extra considerations. This type of data is abundant in modern scientific fields such as genomics, e.g., when considering gene-expression data where the number of genes (variables) is much larger than the number of samples. In our framework, there is now *necessarily* a common subspace simply due to the dimensions. For instance, if $\mathbf{X}_1$ has size 20×1000 and $\mathbf{X}_2$ has size 20×10000 both of rank 20, then they trivially share the same 20-dimensional column-space which is thus $R(\mathbf{X}_{12C})$. In such cases, calculating for instance canonical correlations is problematic and some type of regularization is necessary. Without such regularization, the chances are in most cases high that only uninteresting, trivial and noisy components are identified. One way of trying to solve the problem with many variables and few objects is to use PCA for each block separately, in this way reducing both the noise and the dimensionality (see [62, 29]). Whether this approach is preferable depends on a number of properties (ranks of the different sub-
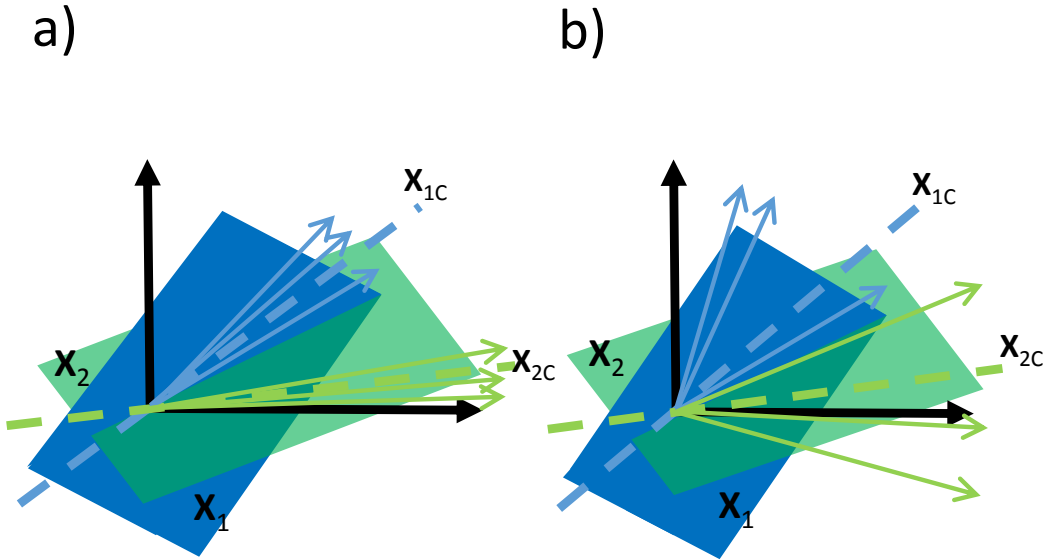
Figure 6: Well-embedded common component (a) and poorly embedded common direction (b).

spaces, noise characteristics etc.) and other approaches will be discussed later in this paper.

2.2.4 Scaling issues.

Another general aspect that is important to discuss is the scaling of the blocks and variables within blocks. We will refer to this as between-block scaling and within-block scaling. One example of the latter is known as auto-scaling in chemometrics, standardization in psychometrics and normalization in statistics. We assumed already centered data and auto-scaling on top of that also divides every column of a matrix by its standard deviation. Hence, the data is analyzed in correlation mode. The between-block scaling is related to the total variation of a block. It is often natural to do some type of

overall scaling of the blocks in order to avoid too much dominance of one of the blocks. For instance, in cases where one of the blocks has only a few variables and another one has many variables, the joint approach could put almost all emphasis on trying to model the larger data set. This may lead to solutions where one is not modeling the joint variation, but merely within block variability, which is clearly not the intention in data fusion. A possible way to counter this is to divide each block by the Frobenius norm prior to analysis. General guidelines for centering and scaling are available [7, 61] and there is also literature on scaling in multi-block data analysis [66, 43, 71, 55].

3   How established methods relate to the definitions

In this section, we will discuss how a number of already existing methods aiming for identifying common and distinct components are related to the definitions given in Section 2. We will discuss these methods mostly by using two blocks of data and more than two blocks if clarity allows to do so (we will index the blocks by $k = 1, ..., K$). Tables 1 and 2 summarize properties of these discussed methods. The methods to be discussed originate from different fields of science and thus use different notations. We will try to harmonize this by using as much as possible a uniform notation based on the familiar PCA model:

$$\mathbf{X} = \mathbf{X}\mathbf{W}\mathbf{P}^T + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{10}$$

where the matrix of weights $\mathbf{W}$ defines linear combinations of the columns of $\mathbf{X}$, generating scores $\mathbf{T}$ and loadings $\mathbf{P}$ which are the regression coefficients of $\mathbf{X}$ on $\mathbf{T}$. In PCA, the matrix $\mathbf{W}$ will be identical to $\mathbf{P}$, but this is not necessarily so for all methods. To arrive at a consistent terminology for all the methods to be discussed, we will use the terms and corresponding symbols weights, scores and loadings in the following.

3.1   Simultaneous Component Analysis, Generalized Canonical Correlation and a compromise

The two most different ways of defining common variability are probably PCA on the concatenated matrix $[\mathbf{X}_1|...|\mathbf{X}_K]$ which focuses on explaining

the simultaneous variation in all blocks and Canonical Correlation or its generalized form (GCA; see below) which only focuses on correlation between the blocks. The PCA model on concatenated data goes under various names as will be explained in the next section.

3.1.1 Simultaneous Component Analysis

The optimization criterion for Simultaneous Component Analysis (SCA) is

$$\min_{(\mathbf{T},\mathbf{P}_k)} \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{T}\mathbf{P}_k^T\|^2 \tag{11}$$

where the simultaneous components are represented by $\mathbf{T}(I \times R)$ and the loadings $\mathbf{P}_k(I \times J_k)$ measure how these components are related to the original data. This model is known under different names: SUM-PCA in chemometrics [47], Simultaneous Component Analysis (SCA-P) in psychometrics [56] and Tucker1 in three-way analysis [45]. The underlying idea of using this model is that $\mathbf{T}$ represents as much as possible the variation in all data blocks simultaneously. Hence, a model of each block can be written as

$$\mathbf{X}_k = \mathbf{T}\mathbf{P}_k^T + \mathbf{E}_k; k = 1, ..., K \tag{12}$$

and several properties of SCA are described in Tables 1 and 2. The optimization problem of Eqn. 11 is stated as a least-squares problem, but can also be formulated as the problem of finding the eigenvectors of

$$\mathbf{Z}_{SCA} = \sum_{k=1}^{K} \mathbf{X}_k \mathbf{X}_k^T \tag{13}$$

and selecting the R eigenvectors belonging to the R largest eigenvalues. Alternatively, the components $\mathbf{T}$ can be found using the SVD of $[\mathbf{X}_1|..|\mathbf{X}_K]$ and choosing the R left singular vectors corresponding to the R largest singular values (i.e. a PCA on the concatenated matrix $[\mathbf{X}_1|...|\mathbf{X}_K]$). Hence, the components $\mathbf{T}$ are in the column-space of $[\mathbf{X}_1|..|\mathbf{X}_K]$ and not necessarily in the column-spaces of any of the individual matrices $\mathbf{X}_k$. The matrix $\mathbf{T}$ represents both common and distinct variation according to the definitions given above and, hence, the model is not separating common and distinct sources of variation. Moreover, the term *simultaneous* component analysis suggest a focus on common components which is not the case. Nevertheless,

we present SCA here since it is much used in multi-set analysis and a starting point of other methods. Note that the least squares property does not hold per data block, but only across all blocks simultaneously. However, given $\mathbf{T}$, Eqn. 12 is a least squares model for the set of all $\mathbf{X}_k$.

Without loss of generality, the simultaneous components $\mathbf{T}$ can be chosen to be orthogonal due to rotational freedom of the model. The subspace spanned by $\mathbf{T}$ is unique like in ordinary PCA. The residuals $\mathbf{E}_k$ are orthogonal to the model part of $\mathbf{X}_k$ (which is $\mathbf{T}\mathbf{P}_k^T$) and thus a break-down of sum-of-squares can be calculated. Note, however, that due to the fact that $\mathbf{T}$ is not necessarily in the range of $\mathbf{X}_k$ neither is $\mathbf{E}_k$. SCA is sensitive to between- and within-block scaling.

Whereas SCA is a simultaneous method for data fusion, there is a history of sequential methods in chemometrics which serve the same purpose. These methods are known under different names and versions (Hierarchical PCA, Consensus PCA, Multiblock PCA). Due to their sequential nature, it is sometimes difficult to assess their properties, but some results exist [70, 47].

SCA has been used in several areas of science and is a special case of a much broader method in data mining called Collective Matrix Factorization [44]. In metabolomics and process chemometrics it is used in conjunction with multilevel data analysis and as a step after an initial ANOVA [46, 19, 15]. It is also used in spectroscopy [4, 50, 57, 41] and in sensory science [32, 11, 6].

3.1.2 Generalized Canonical Correlation Analysis (GCA)

The goal of GCA is to identify linear combinations of the blocks, $\mathbf{X}_k\mathbf{W}_k$, which fit as well as possible to a set of orthogonal common components $\mathbf{T}$. This is done by minimizing the criterion

$$\min_{(\mathbf{T},\mathbf{W}_k)} \sum_{k=1}^{K} \|\mathbf{X}_k\mathbf{W}_k - \mathbf{T}\|^2 \tag{14}$$

with respect to $\mathbf{T}(\mathbf{T}^T\mathbf{T} = \mathbf{I})$ and $\mathbf{W}_k(k = 1, ..., K)$ [63]. The number of columns in $\mathbf{T}$, A, must be smaller than or equal to the number of columns in the $\mathbf{X}_k$ with the smallest number of columns. If the number of samples, I, is smaller than all $J_k(k = 1, ..., K)$, then $A = I$ is the maximum number

of components. Note that the same solution can be obtained by maximizing a sum of correlations between linear combinations of the X blocks, which is the typical formulation for the situation with only two X-blocks [18]. In that case, this is usually referred to as canonical correlation analysis. In practice, the actual solution $\mathbf{T}$ is found as the eigenvectors of the matrix

$$\mathbf{Z}_{GCA} = \sum_{k=1}^{K} \mathbf{X}_k (\mathbf{X}_k^T \mathbf{X}_k)^+ \mathbf{X}_k^T \tag{15}$$

where again the $+$ means the Moore-Penrose (pseudo-)inverse. The $\mathbf{W}_k's$ can then be found by regressing $\mathbf{T}$ on $\mathbf{X}_k$: $\mathbf{W}_k = \mathbf{X}_k^+ \mathbf{T}$.

If there are A common components in the X-blocks according to the definition given in Section 2.1, the criterion in Eqn. 14 will exactly be equal to 0. In the two-block case the common components will correspond to components with a canonical correlation equal to one.

The solution $\mathbf{T}$ in Eqn. 14 is not necessarily within any of the column-spaces of the $\mathbf{X}_k's$ but it is in the column space of $[\mathbf{X}_1|...|\mathbf{X}_K]$ (for a proof, see the Appendix). Although this is not the goal of GCA, when needed a model of $\mathbf{X}_k$ can be obtained by regressing $\mathbf{X}_k$ on $\mathbf{T}_k = \mathbf{X}_k \mathbf{W}_k$ giving loadings $\mathbf{P}_k$ from which also explained variances can be calculated(see Table 1).

Since GCA only concentrates on correlation and gives no emphasis on within block variability (thereby potentially poorly embedded and hence unstable), several methods have been developed for balancing the two aspects. One particular solution is obtained by defining a continuum of solutions between SCA and GCA using a ridge regression type of formulation joining Eqns. 13 and 15 in one single formula [11]. Enhancing stability of the GCA components can also be obtained by using PCA on the individual data blocks data before using GCA [62] or by regularization [53]. The solution using PCA as a first step will be called PCA-GCA in the example section.

It is possible to also obtain distinct components using PCA-GCA. This can be done by regressing each block on its own common components. The residuals from these regressions represent two distinct subspaces $R(\mathbf{X}_{1D})$ and $R(\mathbf{X}_{2D})$ which can subsequently be subjected to a PCA for each subspace. Note that in this case $R(\mathbf{X}_{1D})$ is orthogonal to $R(\mathbf{X}_{1C})$ and likewise $R(\mathbf{X}_{2D})$ is orthogo-

nal to $R(\mathbf{X}_{2C})$, but $R(\mathbf{X}_{1D})$ is not necessarily orthogonal to $R(\mathbf{X}_{2D})$. Hence, we are in the situation of Figure 3a. GCA does not depend on within-block and between-block scaling and thus the distinct subspaces also do not depend on that. However, performing a PCA on the distinct subspaces depends of course on the within-scaling of the distinctive matrices. Examples of the use of GCA can be found, e.g., in sensory science [11, 63]. Also in signal processing GCA-type methods are used [10] based on the work in biometrics [20].

3.2 O2PLS

There seem to be three different implementations of O2PLS [58, 59, 27]. The last implementation is a generalization of O2PLS to OnPLS (for more than two blocks). The O2(n)PLS methods are usually described in terms of iterative algorithms rather than through formal definitions of well-defined criteria which makes their properties difficult to assess. We describe the implementation of Lofstedt [25].

The starting point for O2PLS is the SVD of the covariance matrix $\mathbf{X}_2^T\mathbf{X}_1$:

$$\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{X}_2^T\mathbf{X}_1 \tag{16}$$

and collecting the R singular vectors corresponding to the R largest singular values of Eqn. 16 in $\mathbf{U}_R$ (left-singular vectors) and $\mathbf{V}_R$ (right-singular vectors), respectively, as weights for the (preliminary) common components. This SVD is known as the product SVD (PSVD) and is a member of a broad class of generalizations of the ordinary SVD [14, 13] and Eqn. 16 is actually also the first step of Bookstein's version of PLS [5]. Define $\mathbf{F}_1 = \mathbf{X}_1 - \mathbf{X}_1\mathbf{V}_R\mathbf{V}_R^T$ and $\mathbf{F}_2 = \mathbf{X}_2 - \mathbf{X}_2\mathbf{U}_R\mathbf{U}_R^T$ then due to the truncation to R components, the (preliminary) common components $\widetilde{\mathbf{T}}_{1C} = \mathbf{X}_1\mathbf{V}_R$ still share some variation with $\mathbf{F}_1$ and likewise $\mathbf{X}_2\mathbf{U}_R$ with $\mathbf{F}_2$. This part can be calculated by solving

$$\max_{\|\mathbf{z}_1\|=1} \|\widetilde{\mathbf{T}}_{1C}^T\mathbf{F}_1\mathbf{z}_1\|^2 \tag{17}$$

which maximizes the shared variation of $\widetilde{\mathbf{T}}_{1C}$ and $\mathbf{F}_1$ in one (orthogonal) component (indexed by $l = 1, ..., L$). A deflation procedure then subsequently regresses $\mathbf{X}_1$ on this component $\mathbf{F}_1\mathbf{z}_1$ and gives the residuals $\mathbf{X}_{res1}$. A similar

procedure can be used for $\mathbf{X}_2$, and the number of orthogonal components has to be chosen (or found). Then (final) common components between the deflated matrices can be extracted one by one by using the MAXDIFF criterion [17]:

$$\max_{\mathbf{w}_1, \mathbf{w}_2} tr(\mathbf{w}_1^T \mathbf{X}_{res2}^T \mathbf{X}_{res1} \mathbf{w}_2) = tr(\mathbf{t}_{2C}^T \mathbf{t}_{1C}); \mathbf{W}_k^T \mathbf{W}_k = \mathbf{I}(k = 1, 2) \qquad (18)$$

where the matrices $\mathbf{T}_{1C}$ and $\mathbf{T}_{2C}$ collect the vectors $\mathbf{t}_{1C}$ and $\mathbf{t}_{2C}$, respectively, and the matrix $\mathbf{W}_k; k = 1, 2$ contain the weights for the different dimensions. This will result in the following models for $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\begin{aligned}
\mathbf{X}_1 &= \mathbf{T}_{1C}\mathbf{W}_1^T + \mathbf{T}_{1D}\mathbf{P}_{1D}^T + \mathbf{E}_1 = \mathbf{X}_{1C} + \mathbf{X}_{1D} + \mathbf{E}_1 \qquad (19) \\
\mathbf{X}_2 &= \mathbf{T}_{2C}\mathbf{W}_2^T + \mathbf{T}_{2D}\mathbf{P}_{2D}^T + \mathbf{E}_2 = \mathbf{X}_{2C} + \mathbf{X}_{2D} + \mathbf{E}_2
\end{aligned}$$

where $\mathbf{T}_{1D}$ collects the orthogonal components $\mathbf{F}_1 \mathbf{z}_l$ (hence the name O(rthogonal)2PLS) and $\mathbf{P}_{1D}$ its loadings, and likewise for $\mathbf{T}_{2D}$ and $\mathbf{P}_{2D}$. The orthogonality properties between the different matrices are shown in Table 2 (see [64]). This means that O2PLS takes the viewpoint of Figure 3a (apart from the fact that each block has its own common component). Eqn. 19 shows that also O2PLS fits in our framework but calculating explained variances is hampered by the orthogonality properties. Note again that we changed notation of the common parts to emphasize that $R(\mathbf{X}_{1C}) \neq R(\mathbf{X}_{2C})$. O2PLS is within-block scale dependent but between-block scale independent.

The O2PLS method has been used amongst others in spectroscopy [30, 9, 21, 35], in the plant sciences [8, 49] and its extension to more than two blocks (OnPLS) has been used in genomics [48, 26]. The latter paper also describes an implementation of the multi-block problem as shown in Figure 4 showing the complexity of such a decomposition. There is an interesting relationship of O2PLS with Procrustes analysis, as explained in the Appendix.

3.3 DIStinct and COmmon-Simultaneous Component Analysis (DISCO-SCA)

Also the DISCO-SCA (or DISCO, for short) method [40, 67] can be posed in terms of our framework. The first step in DISCO is to solve an SCA problem to find scores $\widetilde{\mathbf{T}}(I \times R)$ and loadings $\widetilde{\mathbf{P}}((J_1 + J_2) \times R)$ of the concatenated matrix $[\mathbf{X}_1|\mathbf{X}_2]$. The loading matrix $\widetilde{\mathbf{P}}$ can be partitioned in $\widetilde{\mathbf{P}}_1(J_1 \times R)$

and $\widetilde{\mathbf{P}}_2(J_2 \times R)$. Subsequently, the matrix $\widetilde{\mathbf{P}}$ is orthogonally rotated to a simple structure reflecting distinct and common components. For the sake of illustration, assume that $R = 3$; there are one common and two distinct components (one for each block). Then $\widetilde{\mathbf{P}}$ is orthogonally rotated to a structure $\mathbf{P}_{target}$ according to

$$\min_{\mathbf{Q}^T\mathbf{Q}=\mathbf{I}} \left\| \mathbf{V} \odot (\widetilde{\mathbf{P}}\mathbf{Q} - \mathbf{P}_{target}) \right\|^2 \qquad (20)$$

with $\mathbf{V}$ a matrix of zero's and one's selecting the elements across which the minimization occurs, the symbol $\odot$ indicates the Hadamard or elementwise product and

$$\mathbf{P}_{target} = \begin{vmatrix} x & 0 & x \\ x & 0 & x \\ x & 0 & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{vmatrix} \qquad (21)$$

where the symbol x means an arbitrary value not necessarily zero and $\mathbf{P} = \widetilde{\mathbf{P}}\mathbf{Q} = [\mathbf{P}_1^T|\mathbf{P}_2^T]^T$. This will result in the first component being distinct for $\mathbf{X}_1$, the second component distinct for $\mathbf{X}_2$ and the third component will be the common one. After finding the optimal $\mathbf{Q}$, the scores $\widetilde{\mathbf{T}}$ are counter-rotated resulting in $\mathbf{T} = \widetilde{\mathbf{T}}\mathbf{Q} = [\mathbf{t}_1\mathbf{t}_2\mathbf{t}_3]$ and the following decomposition is obtained:

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{T}\mathbf{P}_1^T = \mathbf{t}_1\mathbf{p}_{11}^T + \mathbf{t}_2\mathbf{p}_{12}^T + \mathbf{t}_3\mathbf{p}_{13}^T + \mathbf{E}_1 \qquad (22) \\ \mathbf{X}_2 &= \mathbf{T}\mathbf{P}_2^T = \mathbf{t}_1\mathbf{p}_{21}^T + \mathbf{t}_2\mathbf{p}_{22}^T + \mathbf{t}_3\mathbf{p}_{23}^T + \mathbf{E}_2 \end{aligned}$$

where $\mathbf{p}_{11}$ gives loadings for the distinct component for $\mathbf{X}_1$; $\mathbf{p}_{22}$ for the distinct component for $\mathbf{X}_2$ and $\mathbf{p}_{13}, \mathbf{p}_{23}$ for the common component. If $\mathbf{p}_{12}$ is not close to zero then there is a distinct non-orthogonal part in the decomposition of $\mathbf{X}_1$ (the red colored $\mathbf{X}_{1DNO}$ in Table 1). The SCA solution has orthogonal columns in $\widetilde{\mathbf{T}}$ and rotates orthogonally afterwards thus these columns remain orthogonal. Hence, $\mathbf{T}$ is orthogonal and it holds that $\mathbf{X}_{1DNO}$ is orthogonal to both $\mathbf{X}_{1DO}$ and $\mathbf{X}_{1C}$, but it is clearly not orthogonal to $\mathbf{X}_{2DO} = \mathbf{t}_2\mathbf{p}_{22}^T$ (see Table 3 and Figure 3a). Minimizing the sum of squared elements of $\mathbf{p}_{12}$ and $\mathbf{p}_{21}$ is exactly what the above mentioned rotation tries to do, thereby

minimizing the sizes of these distinct non-orthogonal parts and defining those parts as being distinct non-orthogonal (see the $\mathbf{P}_{target}$ in Eqn. 21). Thus, this is yet another implementation of the general decomposition scheme where the vectors can be matrices when more than one common and distinct components are present. Contrary to O2PLS, the common parts in DISCO span the same column-space. Because DISCO starts with an SCA and subsequently utilizes a rotation, both $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ are in the column-space of the combined $[\mathbf{X}_1|\mathbf{X}_2]$ rather than the individual parts. The uniqueness properties of DISCO are unknown. Due to the orthogonality of the scores matrix $\mathbf{T}$ explained variances can be calculated based on Eqn. 22. DISCO is within- and between-block scale dependent and has been used in metabolomics [67] and in gene-expression analysis [65].

3.4   Generalized Singular Value Decomposition (GSVD)

A method used in gene-expression data to separate common from distinct components is the Generalized Singular Value Decomposition (GSVD) [3] which is also a generalization of the SVD known as the Quotient SVD (QSVD) [14]. The mathematics of the GSVD dates back already some time [68, 33]. The original GSVD is used for fusion of data sharing the same columns but this problem can be transposed to our situation. The original GSVD is a matrix decomposition method and does not have least squares properties. To repair its sensitivity to noise, we follow the implementation of the Adapted GSVD which comes down to first filtering the data with an SCA step [67]. For the two-block case the model is

$$
\begin{aligned}
\mathbf{X}_1 &= \widehat{\mathbf{X}}_1 + \mathbf{E}_1 = \mathbf{TD}_1\mathbf{V}_1^T + \mathbf{E}_1 \\
\mathbf{X}_2 &= \widehat{\mathbf{X}}_2 + \mathbf{E}_2 = \mathbf{TD}_2\mathbf{V}_2^T + \mathbf{E}_2
\end{aligned}
\tag{23}
$$

with $\widehat{\mathbf{X}}_k$ is the filtered data; $\mathbf{V}_k^T\mathbf{V}_k = \mathbf{I}(k = 1, 2)$, $\mathbf{D}_k(k = 1, 2)$ diagonal and such that $\mathbf{D}_1^2 + \mathbf{D}_2^2 = \mathbf{I}$, and $\mathbf{T}$ a full-rank matrix but not necessarily orthogonal. Due to the latter constraint it is possible to divide the generalized singular values (the elements of $\mathbf{D}_k(k = 1, 2)$) in three groups: if $d_{1R}^2 \approx 1$ the corresponding component is distinctive for $\mathbf{X}_1$, if $d_{2R}^2 \approx 1$ the corresponding component is distinctive for $\mathbf{X}_2$ and if $d_{1R}^2 \approx d_{2R}^2$ the corresponding component is common. Obviously, there is a certain amount of arbitrariness in

these choices. Once such a choice is made, Eqn. 23 can be written as

$$\mathbf{X}_1 = \mathbf{T}_1\mathbf{D}_{11}\mathbf{V}_{11}^T + \mathbf{T}_2\mathbf{D}_{12}\mathbf{V}_{12}^T + \mathbf{T}_3\mathbf{D}_{13}\mathbf{V}_{13}^T + \mathbf{E}_1 \tag{24}$$
$$\mathbf{X}_2 = \mathbf{T}_1\mathbf{D}_{21}\mathbf{V}_{21}^T + \mathbf{T}_2\mathbf{D}_{22}\mathbf{V}_{22}^T + \mathbf{T}_3\mathbf{D}_{23}\mathbf{V}_{23}^T + \mathbf{E}_2$$

which fits our framework. Upon assuming that $\mathbf{T}_1\mathbf{D}_{11}\mathbf{V}_{11}^T = \mathbf{X}_{1DO}$, $\mathbf{T}_2\mathbf{D}_{22}\mathbf{V}_{22}^T = \mathbf{X}_{2DO}$ and $\mathbf{T}_3\mathbf{D}_{13}\mathbf{V}_{13}^T$, $\mathbf{T}_3\mathbf{D}_{23}\mathbf{V}_{23}^T$ are the common components, then $\mathbf{T}_2\mathbf{D}_{12}\mathbf{V}_{12}^T = \mathbf{X}_{1DNO}$ and $\mathbf{T}_1\mathbf{D}_{21}\mathbf{V}_{21}^T = \mathbf{X}_{2DNO}$. Due to the orthogonality of both $\mathbf{V}_1$ and $\mathbf{V}_2$ it holds that $\mathbf{X}_{1DO}$, $\mathbf{X}_{1DNO}$ and $\mathbf{X}_{1C}$ are mutually orthogonal, and likewise for block $\mathbf{X}_2$. However, $\mathbf{X}_{1DNO}$ is not orthogonal to $\mathbf{X}_{2DO}$ and similarly $\mathbf{X}_{2DNO}$ is not orthogonal to $\mathbf{X}_{1DO}$. This is the same as for DISCO and is again the situation of Figure 3a.

For an invertible $\mathbf{X}_2$ the GSVD equals the SVD of $\mathbf{X}_1\mathbf{X}_2^{-1}$ which explains the term Quotient SVD. For these cases, the uniqueness properties of the GSVD are the same as those of the SVD. For non-invertible $\mathbf{X}_2$ the uniqueness properties are not clear. GSVD is within- and between-block scale dependent and has been used in gene-expression analysis [3] and has been extended for more than two blocks in different ways [12, 36].

3.5   Joint and Individual Variances Explained (JIVE)

The method of Joint and Individual Variances Explained (JIVE [24]) goes as follows. For two blocks, it derives directly a decomposition according to:

$$\mathbf{X}_1 = \mathbf{T}_C\mathbf{P}_{1C}^T + \mathbf{T}_{1D}\mathbf{P}_{1D}^T + \mathbf{E}_1 = \mathbf{X}_{1C} + \mathbf{X}_{1D} + \mathbf{E}_1 \tag{25}$$
$$\mathbf{X}_2 = \mathbf{T}_C\mathbf{P}_{2C}^T + \mathbf{T}_{2D}\mathbf{P}_{2D}^T + \mathbf{E}_2 = \mathbf{X}_{2C} + \mathbf{X}_{2D} + \mathbf{E}_2$$

which fits in our framework. Note that we use the notation $\mathbf{T}_C$ to stress that the common scores are the same. In estimating this decomposition, the following constraints are used

$$\mathbf{X}_{1C}^T\mathbf{X}_{1D} = 0; \mathbf{X}_{1C}^T\mathbf{X}_{2D} = 0; \mathbf{X}_{2C}^T\mathbf{X}_{1D} = 0; \mathbf{X}_{2C}^T\mathbf{X}_{2D} = 0 \tag{26}$$

and, thus, the distinct part in a block is orthogonal to the common parts in all blocks but the distinct parts in different blocks are not necessarily orthogonal. This is again an implementation of Figure 3a. The (low) ranks of all common and distinct matrices involved are determined by permutation tests. Since $\mathbf{E}_k$ is not necessarily orthogonal to neither $\mathbf{X}_{kC}$ nor $\mathbf{X}_{kD}$, separating sums-of-squares (and variances, despite the name) is not easy for JIVE. For other

24

properties, see Tables 1 and 2. JIVE is within- and between-block scale dependent and has been applied in gene-expression analysis [24].

3.6 Structure Revealing Data Fusion

In Structure Revealing Data Fusion [2] an approach is chosen based on penalties. The method is developed for fusing two-way and three-way arrays but can equally well be used for fusing two-way arrays. Starting point is the model

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{TD}_1\mathbf{V}_1^T + \mathbf{E}_1 = \mathbf{TP}_1^T + \mathbf{E}_1 \\ \mathbf{X}_2 &= \mathbf{TD}_2\mathbf{V}_2^T + \mathbf{E}_2 = \mathbf{TP}_2^T + \mathbf{E}_2 \end{aligned} \tag{27}$$

where the matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal and the diagonals of $\mathbf{T}^T\mathbf{T}$, $\mathbf{V}_1^T\mathbf{V}_1$ and $\mathbf{V}_2^T\mathbf{V}_2$ consist of ones (i.e. the columns of $\mathbf{T}$, $\mathbf{V}_1$ and $\mathbf{V}_2$ have length one). The components are now estimated under an L_1 penalty [54]:

$$\min_{\mathbf{T},\mathbf{V}_1,\mathbf{V}_2,\mathbf{D}_1,\mathbf{D}_2} \left\|\mathbf{X}_1 - \mathbf{TD}_1\mathbf{V}_1^T\right\|^2 + \left\|\mathbf{X}_2 - \mathbf{TD}_2\mathbf{V}_2^T\right\|^2 + \lambda(\|diag(\mathbf{D}_1)\|_1 + \|diag(\mathbf{D}_2)\|_1)$$

$$\tag{28}$$

where $\lambda \geq 0$ is the penalty parameter to be set by the user, the symbol $\|.\|_1$ represents the L_1-norm and $diag(\mathbf{D}_k)$ is the vector carrying the diagonal of $\mathbf{D}_k$. Increasing the penalty value $\lambda \geq 0$ will force more elements in $\mathbf{D}_1$ and $\mathbf{D}_2$ to become zero. From the patterns of these zero's the common and distinct components are defined. This type of approach - albeit in fusing three-way and two-way data - has been used in metabolomics [2, 1]. Structure Revealing Data Fusion is within- and between-block scale dependent.

A special class of Structure Revealing Data Fusion methods are the multivariate curve resolution methods as used in chemometrics [51]. This class of methods performs data fusion mostly using hard constraints on the parameters based on chemical information. There are very many applications of this method in different fields of chemistry.

4 Examples

To illustrate some of the methods falling under our framework and their relationships, we will show some real data examples that were already introduced

shortly in Section 1. Two real data sets from medical biology and food science will be used to show aspects related to practical use of the methods. All data will be analyzed by the methods PCA-GCA (see Section 3.1.2) and DISCO (see Section 3.3). These methods are selected because they represent different orthogonality constraints (referring to Figure 3a) and b), respectively). They also represent different choices of common components as discussed in Section 2.2: PCA-GCA estimates separate common components for each data block while DISCO estimates a best compromise which is in the column space of the concatenated data blocks.

For all methods, the first step is to decide the dimensionalities of the subspaces. This is not a trivial task and different strategies exist for the different methods. The strategies for PCA-GCA and DISCO will be explained briefly in the examples below, but a thorough discussion of the model selection is not within the scope of this paper. Once the dimensions are decided, it is straightforward to estimate basis vectors (or components) for each of the subspaces.

4.1 Sensory example

The sensory example focuses on one of the typical aspects of a product development process: The product developer is interested in understanding how well two important modalities of the descriptive sensory profile relate to the ingredients in the recipe. A typical issue of interest for being able to optimize product quality is whether the recipe influences both smell and taste and in which way this happens. In particular, one is interested in knowing what aspects of smell and taste that are common and what is unique in the two sensory profiles.

This example consists of descriptive sensory attributes of flavored water samples and is a subset of a larger data set [29]. The 18 water samples are created according to a full factorial experimental design with two flavor types (A and B), three flavor doses (0.2, 0.5 and 0.8 g/l) and three sugar levels (20, 40 and 60 g/l). A trained sensory panel consisting of 11 assessors evaluated the samples first by smelling (9 descriptors) and then by tasting (14 descriptors), using an intensity scale from 1 to 9. Two data blocks (SMELL and TASTE) were constructed by averaging across assessors. The blocks were mean-centered and block-scaled to sum-of-squares one prior to analysis.

A crucial aspect of the decomposition is to decide the dimensions of the common and distinct subspaces. For DISCO, this is a two-step process: first, the number of SCA components is selected. This number represents the sum of the dimensions of all subspaces, i.e. $dimR(\mathbf{X}_{12C}) + dimR(\mathbf{X}_{1D}) + dimR(\mathbf{X}_{2D})$. Then, the most appropriate target matrix $\mathbf{P}_{Target}$ (Eqn. 21) is sought by evaluating the non-congruence value (Eqn. 20) for all possible allocations of common and distinct components. Since there are three independent design factors in this experiment (flavor type, flavor dose and sugar level), we choose to keep three SCA components even if the third component explain very little variance (Figure 7a). The lowest non-congruence value is approximately equal for models with one and two common components (Figure 7b), but after a closer inspection of the scores we choose the model with one common component and one disticnt component per block.

For real data the non-congruence value is never zero, meaning that the zeros in the target matrix are not exactly zero in the rotated loadings, which means that the distinct component for one block also explains some variance in the other block. The latter is the distinct-non-orthogonal subspace. The DISCO decomposition for this data set is then:

$$
\begin{aligned}
\mathbf{X}_S &= \mathbf{X}_{C,S}(70.0\%) + \mathbf{X}_{D,T}(3.4\%) + \mathbf{X}_{D,S}(9.4\%) + \mathbf{E}_S \qquad (29) \\
\mathbf{X}_T &= \mathbf{X}_{C,T}(26.8\%) + \mathbf{X}_{D,T}(55.5\%) + \mathbf{X}_{D,S}(1.9\%) + \mathbf{E}_T
\end{aligned}
$$

where each subspace is of dimension one; S and T stand for Smell and Taste, respectively and between brackets is the amount of explained variation in the block. Note that both distinct-non-orthogonal subspaces are very small in this case, and probably consist of noise only.

For PCA-GCA, the dimension selection is also a stepwise procedure: First, an appropriate number of principal components is selected for each data block, corresponding to $dimR(\mathbf{X}_1)$ and $dimR(\mathbf{X}_2)$ in Eqn 2. Next, the correlation coefficients and explained variances from GCA are evaluated in order to decide the number of common components, $dimR(\mathbf{X}_{12C})$. The number of distinct components is then given as the difference between $dimR(\mathbf{X}_k)$ and $dimR(\mathbf{X}_{12C})$. In this example, we choose to keep three components for each block, following the same argument as for DISCO (three design factors). Figure 7c shows that the canonical correlation together with the explained variances clearly suggest one common component (correlation =
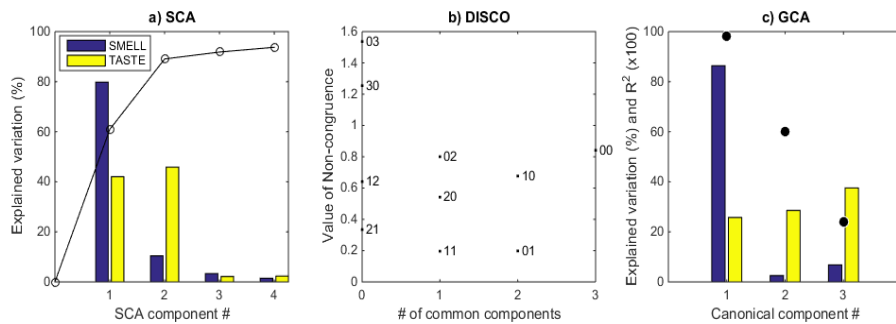
Figure 7: Plots for selecting numbers of components for the sensory example. a) SCA. The curve represents cumulative explained variance for the concatenated data blocks. The bars show how much variance each component explain in the individual blocks. b) DISCO. Each point represent the non-congruence value for a given target (model). The plot includes all possible combinations of common and distinct components based on a total rank of three. The horizontal axis represents the number of common components and the numbers in the plot represent the number of distinct components for SMELL and TASTE respectively. c) PCA-GCA: black dots represent the canonical correlation coefficient (x100) and the bars show how much variance the canonical components explain in each block.

0.98), which means that the distinct subspaces are two-dimensional. The distinct subspaces can be split into an orthogonal and non-orthogonal part as for DISCO, but that is not done here. The decomposition from PCA-GCA

28

is then:

$$\begin{aligned}
\mathbf{X}_S &= \mathbf{X}_{C,S}(86.4\%) + \mathbf{X}_{D,S}(9.3\%) + \mathbf{E}_S \qquad (30) \\
\mathbf{X}_T &= \mathbf{X}_{C,T}(25.7\%) + \mathbf{X}_{D,T}(66.2\%) + \mathbf{E}_T
\end{aligned}$$

where the common part has dimensionality one and both distinct parts have dimensionality two.

The subspaces found by PCA-GCA and DISCO are very similar. The correlation between the common DISCO component ($\mathbf{T}_{12C}$) and the common PCA-GCA components ($\mathbf{T}_{1C}$ and $\mathbf{T}_{2C}$) are 0.98 for both blocks. The correlation between the distinct (orthogonal) SMELL component from DISCO ($\mathbf{T}_{1DO}$) and the first distinct SMELL component from PCA-GCA (first column of $\mathbf{T}_{2D}$) is 0.74. The corresponding number for the distinct TASTE components is 0.99. Figure 8 shows biplots from PCA-GCA for each of the two blocks. It is clear that the common component distinguishes between flavor type (A and B). This component explains 86% of the SMELL variation and 26% of the TASTE variation. As a validation of the commonness, note that the sensory attributes that span this subspace are the same both for smelling and tasting: synthetic/lactonic/oral for flavor A versus ripe/tropical/sulfurous for flavor type B. The first distinctive SMELL component explains 7% of the variation and is related to the flavor dose, showing that the lowest dose tend to give a more lactonic smell. The first distinctive TASTE component explains 63% of the variation and describes differences in sugar level. The attributes that span this component are sweet/ripe versus sour/synthetic/skin/dry.

This example shows that both methods are able to separate common and distinct subspaces in a similar way. The subspaces that explain a large proportion of the variance (common and distinct TASTE) are practically equal for both methods (correlations > 0.98), while there is less agreement regarding the weaker distinct SMELL component (correlation = 0.74).

4.2 Medical biology example

The data set is a subset of a larger study on the effects of gastric bypass surgery on obese and diabetic subjects [23]. Here, we focus on 14 obese patients with Diabetes Mellitus Type II (DM2) who underwent gastric bypass
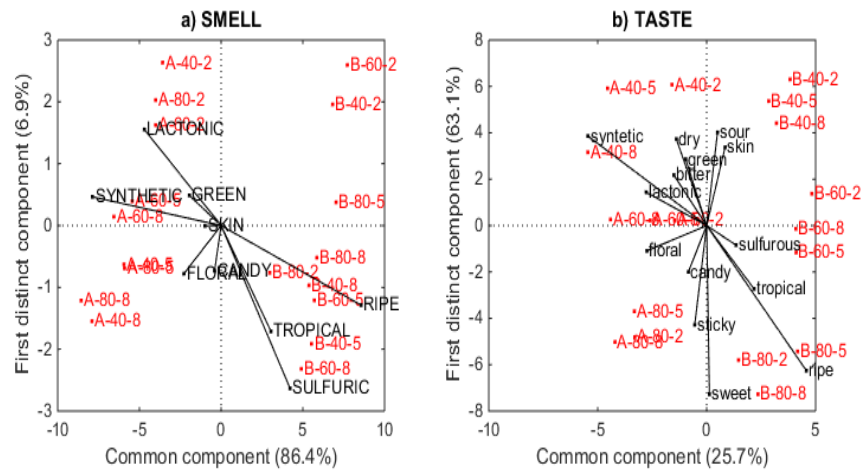
Figure 8: : Biplots from PCA-GCA, showing the variables as vectors and the samples as points. The samples are labeled according to the design factors flavor type (A/B), sugar level (40,60,80) and flavor dose (2,5,8). The plots show the common component (horizontal) against the first distinct component for each of the two blocks.

surgery. Blood samples were taken four weeks before and three weeks after surgery and on each occasion samples were taken both before and after a meal. The blood samples were then analyzed on multiple analytical platforms for the determination of amines, lipids and oxylipins. The three data blocks Amines (A), Lipids (L) and Oxylipins (O) consist of 14 subjects x 4 samples = 56 rows, and 34, 243 and 32 variables respectively. All variables in all three blocks were square-root transformed, in order to obtain more evenly distributed data. Individual differences between subjects were removed by subtracting each subjects' average profile. All variables were then scaled to unit variance. The blocks were also scaled to unit norm prior to SCA, to normalize scale differences between blocks.
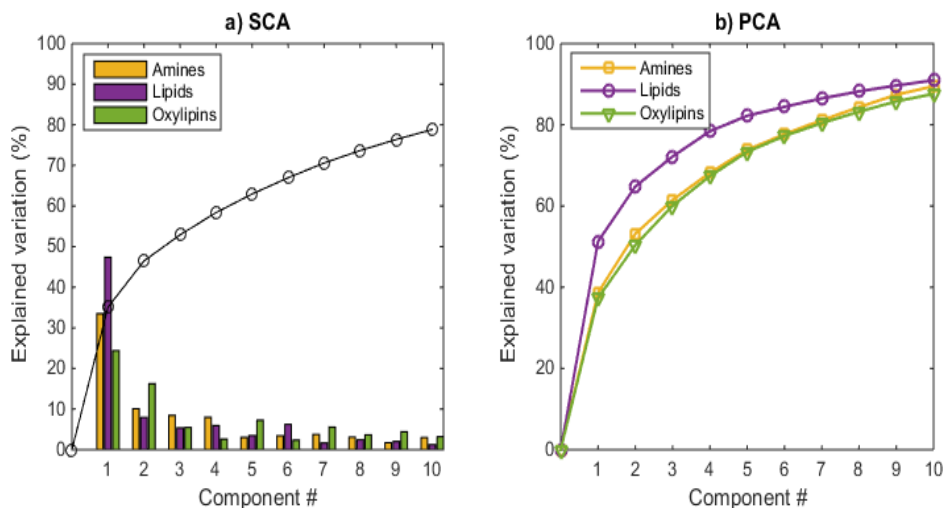
30

Figure 9: Explained variances for a) SCA. The bars represent variances within each block, and the curve represent cumulative explained variance in all blocks combined. b) PCA on each block separately.

Selecting the dimensions of the subspaces is more complicated when the numbers of blocks increase. In this three-block example, we need to decide the dimensions of seven subspaces: $\mathbf{X}_{123C}$, $\mathbf{X}_{12C}$, $\mathbf{X}_{13C}$, $\mathbf{X}_{23C}$, $\mathbf{X}_{1D}$, $\mathbf{X}_{2D}$, and $\mathbf{X}_{3D}$. For DISCO, we start by deciding the sum of all the dimension, i.e. the number of SCA components. Explained variance as a function of components for SCA is given in Figure 9a. The curve of cumulative variance does not have a clear bend, which makes it hard to decide the cutoff between structure and noise. To allocate the common and distinct components we

31

need to fix the number of SCA components and then compare the fit values of Eqn. 20. The computations are time-consuming, as there are e.g. 462 possible target matrices for the 5-component model. To illustrate the complexity in selecting the dimensions for the subspaces, we have calculated all possible rotations for models with 3-5 SCA components, and the results for the four best-fit values are given in Table 3. The values are very similar, making it hard to conclude which rotation gives the best fit. Looking further into the actual rotated score vectors, we discover that many of the models agree on some of the subspaces. These are marked with colors in Table 3. We choose to interpret the 5-component model with fit value 0.24 (the best 5-component model), since this model includes all the agreed upon subspaces. The model contains one component that is common across all three blocks, two components common for A and L, and one distinct component from both A and O. The decomposition of each block is illustrated by pie charts in Figure 10a-c. Notice that there is a substantial contribution of one of the C-AL components also in the O block (7%), which implies that this component could perhaps also be regarded as common across all three blocks.

In PCA-GCA, the number of principal components need to be set for each block separately before performing GCA. Explained variations for the three PCA models are shown in Figure 9b. As for SCA, it is not clear how many components to keep for each block. To investigate how the choice affect the GCA, we ran GCA on all combinations of 5-8 components from each block (64 combinations in total). The canonical correlation coefficient for cases with more than two blocks is defined as the average correlation between all pairs of components from different blocks. Using 0.7 as correlation threshold for commonness in the GCA, we found that 85% of the models had two common components across all blocks, and one common component across A and L. The model based on five components for each block is illustrated in Figure 10d-f. Closer investigation of the components revealed that the second common component across all three blocks is very similar to the one of the C-AL DISCO component mentioned above, which explained 7% of the variation in O. This illustrates the complexity of splitting common and distinct components in noisy and complex data.

To interpret the different subspaces, we plot the scores and loadings from the DISCO model. Figure 10 shows the one-dimensional subspace that is common for all three blocks (C-ALO), which accounts for 19%, 28% and
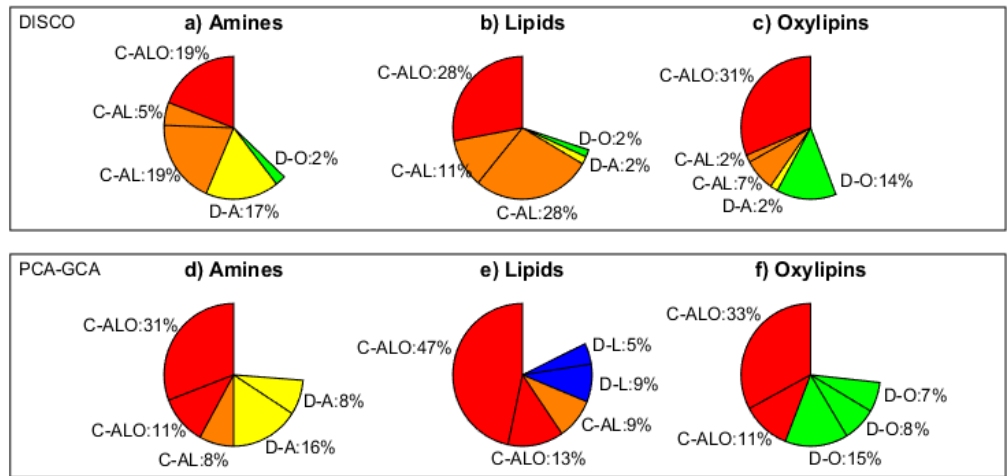
Figure 10: Subplots a), b) and c) show the decomposition by DISCO for blocks A, L and O respectively, while d), e) and f) show the corresponding decomposition by PCA-GCA. Each segment represent a component (dimension).

31% of the variation in A, L and O respectively. The scores are shown in the top panel of Figure 10. It is clear that the component contains information both related to surgery and meal; the scores are increasing after surgery and decreasing after the meal. The variables spanning this dimension in each of the three blocks are shown in the bar plots of Figure 10 (bottom). The most striking observation is that the branched chain amino acids leucine, valine (and to a lesser extend leucine) and L-2-aminoadipic acid (closely related to branched chain amino acids) are down regulated after surgery, which confirms earlier findings [23]. There is more in common between amines and lipids
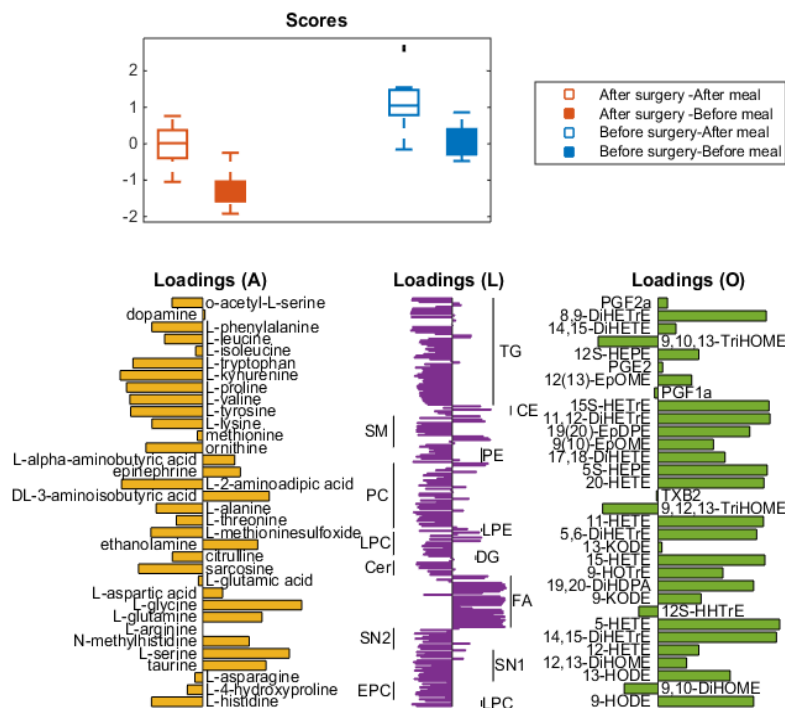
Figure 11: Scores and loadings for the one-dimensional DISCO subspace common between all three blocks.

than oxylipids; both amines and lipids are involved in central carbon and energy metabolism and therefore they may show higher correlation among some amino acids and some lipid groups (as reflected by common subspace).

The two-dimensional subspace common between A and L is shown in Figure 12. These two components together account for 24% and 39% in the A and L blocks respectively, and they even explain 9% in the O block. Here also, we see groupings according to both surgery and meal, especially in the vertical dimension. Note that the two groups that were overlapping in the C-ALO
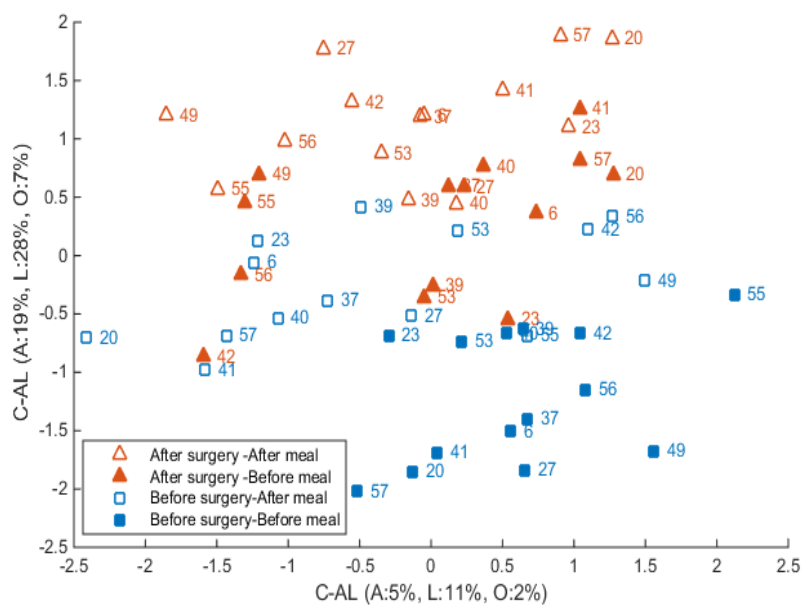
Figure 12: Scores for the two-dimensional DISCO subspace common between the amine and lipids blocks.

component ("before surgery-before meal" versus "after surgery-after meal") are completely separated in this subspace. Plots of the distinct components (not shown) did not reveal clear patterns related to the factors treatment and meal. Hence, all effects are seen in the common parts meaning that a large part of the metabolism is affected simultaneously by these two factors.

5 Discussion

5.1 Revisiting the framework

After having given a short tour of methods for finding common and distinct components, it is worthwhile to recapitulate the general mathematical framework as presented in Eqns. 4 and 7. A summary of the models underlying the presented methods is given in the column 'Model' of Table 2. It appears that for the two-block case the general model is

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{X}_{1C} + \mathbf{X}_{1D} + \mathbf{E}_1 \\ \mathbf{X}_2 &= \mathbf{X}_{2C} + \mathbf{X}_{2D} + \mathbf{E}_2 \end{aligned} \tag{31}$$

with different properties of the matrices $\mathbf{X}_{kC}$, $\mathbf{X}_{kD}$ and $\mathbf{E}_k$. Some methods do not estimate $\mathbf{X}_{kD}$ (GCA) and some methods do not distinguish between common and distinct (SCA). Different choices are made regarding the positioning of the column-spaces of $\mathbf{X}_{kC}$ and $\mathbf{X}_{kD}$ (see Table 2 under 'Subspace properties'). Also different (although sometimes implicit) choices are made regarding orthogonality (see Table 2 under 'Orthogonality') resulting in differences in (E)SS. None of the methods makes a rigorous direct sum decomposition as in Eqns. 4 and 7. The cases for more than two blocks shows an even wider variety of possibilities. The choice of orthogonality constraints depends on the application. It may well be that in most practical applications only the common and orthogonal distinct parts are the most informative.

5.2 Finding common and distinct subspaces

There is an interesting difference in the way the various methods find common and distinct subspaces. Some methods work clearly in the column-spaces of the matrices involved (GCA, JIVE), some methods work through the row-spaces (DISCO, O2PLS) and some methods work in both types of spaces simultaneously (GSVD and Structure Revealing Data Fusion). Whether or not this has consequences for the interpretation of the results of the different models is an open question.

5.3 Open issues and future work

There are obviously many open issues in this field of research. We have only briefly touched upon the issue of explained variances, but there are many

nontrivial aspects that need attention. Also the problem of interpretation, that is, moving from layer three to four in Figure 1, needs attention. This is a very important issue because interpretation is one of the raisons-d'être for data fusion methods. Our framework primarily considers the column space of the data matrices but interpretation is done mostly in the row-space. How to investigate this depends on the scope of the analysis and the type of data available and it is not possible to set up a completely general procedure. There are, however, some general tools that can be useful: One important possibility is to simply project the original data blocks onto the estimated subspaces. For instance, for $R(\mathbf{X}_{12C})$ one simply regresses $\mathbf{X}_1$ onto a suitable basis for the space $R(\mathbf{X}_{12C})$. In this way one obtains information about how the original data are related to the basis for each subspace. Also moving to analyzing more than two blocks simultaneously is not trivial. Many choices have to be made and no clear guidelines exist on how to perform this. Model selection becomes an even more important issue then and possibly Bayesian factor analysis methods with automated model selection can be of use in this context [37].

6   Acknowledgements

7   Appendix

7.1   Possibilities of orthogonal decompositions

7.1.1   Choices for Distinct-Orthogonal (DO) spaces.

There are several possibilities for choosing orthogonality in the decompositions of Eqn. 4. These will be outlined and explained below. We will focus attention on $R(X_{1DO})$ but analogous results hold for $R(X_{2DO})$; we will consider $R(X_{1DNO})$ as a 'rest' term and not consider this space explicitly. The first level to discuss possibilities is regarding the status of $R(X_{1D})$. The alternatives are:

A0: no orthogonality restrictions for $R(X_{1D})$.

A1: $R(X_{1D})\perp R(X_{12C})$ (see Figure 3a).

A2: $R(X_{1D})\perp R(X_{2D})$ (see Figure 3b).

A3: $R(X_{1D})\perp R(X_2)$ which implies **A1** and **A2**.

and, as said earlier, alternative **A3** is not always possible. The status of $R(X_{1DO})$ is nested in alternatives **A0-A3**, since $R(X_{1DO})$ is a part of $R(X_{1D})$.

The alternatives under **A0** are:

A01: $R(X_{1DO})\perp R(X_{2DO})$.

A02: $R(X_{1DO})\perp R(X_{2D})$.

A03: $R(X_{1DO})\perp R(X_2)$.

which shows increasing degrees of orthogonality.

The alternatives under **A1** are:

A10: $R(X_{1DO})\perp R(X_{12C})$ which follows from **A1**.

A11: $R(X_{1DO})\perp R(X_{12C})$ and $R(X_{1DO})\perp R(X_{2DO})$.

A12: $R(X_{1DO})\perp R(X_{12C})$ and $R(X_{1DO})\perp R(X_{2D})$ which implies $R(X_{1DO})\perp R(X_2)$ and is the same as alternative **A03**.

which shows again increasing degrees of orthogonality.

The alternatives under **A2** are:

A20: $R(X_{1DO})\perp R(X_{2D})$ which follows from **A2** and is the same as alternative **A02**.

A21: $R(X_{1DO})\perp R(X_{2D})$ and $R(X_{1DO})\perp R(X_{12C})$ which is again the same as alternative **A03**.

and under alternative **A3** there is only one option namely $R(X_{1DO})\perp R(X_2)$ which is again the same as option **A03**. Concluding, for the two block case there are five different alternatives to select $R(X_{1DO})$: **A01**, **A02**, **A03**, **A10** or **A11**. Whether these alternatives are available for a specific application depends on the dimensions and positioning of the subspaces. An example of this and how to analyze such situations is presented in the next Subsection 7.1.2.

7.1.2 A specific example.

As an example of using rigorous linear algebra results to explore possibilities for (non-)orthogonal decompositions consider the example of two distinct subspaces both of dimension two (see Section 2.1.1). It can be proven that if $R(\mathbf{X}_{1D})$ and $R(\mathbf{X}_{2D})$ are both two-dimensional and not orthogonal, then for every vector $\mathbf{x}$ in $R(\mathbf{X}_{1D})$ there is exactly one vector $\mathbf{y}$ in $R(\mathbf{X}_{2D})$ orthogonal to $\mathbf{x}$. This goes as follows. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are both orthogonal matrices serving as bases for $R(\mathbf{X}_{1D})$ and $R(\mathbf{X}_{2D})$, respectively. Assume also that $r(\mathbf{A}^T\mathbf{B}) = 2$ where $r(.)$ means the rank of a matrix. This implies:

- $R(\mathbf{X}_{1D})$ is not orthogonal to $R(\mathbf{X}_{2D})$; otherwise $\mathbf{A}^T\mathbf{B} = 0$

- $R(\mathbf{X}_{1D})$ does not contain a vector orthogonal to the whole of $R(\mathbf{X}_{2D})$ (this vector $\mathbf{g}$ could be written as $\mathbf{g} = \mathbf{A}\mathbf{h}$ and then $\mathbf{h}^T\mathbf{A}^T\mathbf{B} = 0$, which contradicts $r(\mathbf{A}^T\mathbf{B}) = 2$)

- $R(\mathbf{X}_{2D})$ does not contain a vector orthogonal to the whole of $R(\mathbf{X}_{1D})$ (analogously as above)

now there is for any nonzero vector $\mathbf{x} \in R(\mathbf{X}_{1D})$ exactly one nonzero vector $\mathbf{y} \in R(\mathbf{X}_{2D})$ such that $\mathbf{x}^T\mathbf{y} = 0$.

Proof: write $\mathbf{x} = \mathbf{A}\mathbf{u}$, $\mathbf{y} = \mathbf{B}\mathbf{v}$ and $\mathbf{B} = \mathbf{A}\mathbf{U} + \mathbf{A}^\perp\mathbf{V}$. Find a vector $\mathbf{v}$ such that $\mathbf{x}^T\mathbf{y} = 0$, or, such that $\mathbf{u}^T\mathbf{A}^T(\mathbf{A}\mathbf{U} + \mathbf{A}^\perp\mathbf{V})\mathbf{v} = 0$ which equals $\mathbf{u}^T\mathbf{U}\mathbf{v} = 0$ because of the orthogonality of $\mathbf{A}$ and the definition of the orthogonal complement $\mathbf{A}^\perp$. Then $\mathbf{U} = \mathbf{A}^T\mathbf{B}$ which follows by pre-multiplying $\mathbf{B} = \mathbf{A}\mathbf{U} + \mathbf{A}^\perp\mathbf{V}$ with $\mathbf{A}^T$; by defining $\mathbf{t} = [t_1|t_2]^T = \mathbf{U}^T\mathbf{u}$ which is a unique nonzero vector (because $r(\mathbf{U} = \mathbf{A}^T\mathbf{B}) = 2$) it follows that $\mathbf{v} = [t_2| - t_1]^T$ is the vector which makes $\mathbf{u}^T\mathbf{U}\mathbf{v} = 0$. Hence, there is exactly one vector $\mathbf{y} = \mathbf{B}\mathbf{v}$ in $R(\mathbf{X}_{2D})$ which is orthogonal to $\mathbf{x}$.

7.2 GCA proof

In the main text it was stated that the solution $\mathbf{T}$ in Eqn. 14 is in the column space of $[\mathbf{X}_1|...|\mathbf{X}_K]$. This will be proven now for the two-block situation for simplicity, but is easily generalized to the more-than-two block situation. Eqn. 15 can also be written as

$$\sum_{k=1}^{2} \mathbf{X}_k(\mathbf{X}_k^T\mathbf{X}_k)^+\mathbf{X}_k^T = [\mathbf{X}_1|\mathbf{X}_2][\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^+|\mathbf{X}_2(\mathbf{X}_2^T\mathbf{X}_2)^+]^T = \mathbf{T}\mathbf{S}\mathbf{T}^T \quad (32)$$

where the full eigenvalue decomposition (i.e. $\mathbf{S} > 0$) is used. Post-multiplying both sides of Eqn. 32 by $\mathbf{TS}^{-1}$ gives now

$$[\mathbf{X}_1|\mathbf{X}_2][\mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^+|\mathbf{X}_2(\mathbf{X}_2^T\mathbf{X}_2)^+]^T\mathbf{TS}^{-1} = \mathbf{T} \tag{33}$$

or

$$[\mathbf{X}_1|\mathbf{X}_2]\mathbf{Q} = \mathbf{T} \tag{34}$$

which shows that $\mathbf{T}$ (and also its first columns when only those are used) is in the range of $[\mathbf{X}_1|\mathbf{X}_2]$. This argument is easily extended to more than two blocks.

7.3 Relationship between O2PLS and Procrustes Analysis

An interesting relationship exists between O2PLS and Procrustes Analysis. The Procrustes problem can be stated as

$$\min_{\mathbf{R}^T\mathbf{R}=\mathbf{I}} \|\mathbf{X}_2\mathbf{R} - \mathbf{X}_1\|^2 \tag{35}$$

and the solution of this problem is $\mathbf{R} = \mathbf{U}\mathbf{V}^T$ where $\mathbf{U}$ and $\mathbf{V}$ are from the SVD of $\mathbf{X}_2^T\mathbf{X}_1 = \mathbf{U}\mathbf{S}\mathbf{V}^T$ [38]. Then post-multiplying both $\mathbf{X}_2\mathbf{R}$ and $\mathbf{X}_1$ with $\mathbf{V}$ gives $\mathbf{X}_2\mathbf{R}\mathbf{V} = \mathbf{X}_2\mathbf{U}\mathbf{V}^T\mathbf{V} = \mathbf{X}_2\mathbf{U}$ and $\mathbf{X}_1\mathbf{V}$ which are the same quantities as obtained for O2PLS. Note that the Procrustes problem of Eqn. 35 is equivalent to

$$\min_{\mathbf{R}_k^T\mathbf{R}_k=\mathbf{I}} \|\mathbf{X}_2\mathbf{R}_2 - \mathbf{X}_1\mathbf{R}_1\|^2 \tag{36}$$

which is the symmetric formulation of the problem with solution $\mathbf{R}_2 = \mathbf{U}$ and $\mathbf{R}_1 = \mathbf{V}$ [52].

8 Tables

Table 1: **Fundamental aspects of the fusion methods**

Abbreviations: C is common; D is distinct. Colors: brown is mixed subspace; green is common subspace; red/blue are distinct subspaces.

Methods	Model	Uniqueness
SCA	$\mathbf{X}_k = \mathbf{T}\mathbf{P}_k^T + \mathbf{E}_k$ or $\mathbf{X}_k = \widetilde{\mathbf{T}}_k\mathbf{P}_k^T + \mathbf{F}_k$ $(\widetilde{\mathbf{T}}_k = \mathbf{X}_k\mathbf{X}_k^+\mathbf{T})$	$R(\mathbf{T})$ is unique
GCA	$\mathbf{X}_k = \mathbf{X}_k\mathbf{W}_k\mathbf{P}_k^T + \mathbf{E}_k = \mathbf{T}_k\mathbf{P}_k^T + \mathbf{E}_k = \mathbf{X}_{kC} + \mathbf{E}_k$	$R(\mathbf{T})$ is unique
O2PLS	$\mathbf{X}_1 = \mathbf{T}_{1C}\mathbf{W}_1^T + \mathbf{T}_{1D}\mathbf{P}_{1D}^T + \mathbf{E}_1 = \mathbf{X}_{1C} + \mathbf{X}_{1D} + \mathbf{E}_1$ $\mathbf{X}_2 = \mathbf{T}_{2C}\mathbf{W}_2^T + \mathbf{T}_{2D}\mathbf{P}_{2D}^T + \mathbf{E}_2 = \mathbf{X}_{2C} + \mathbf{X}_{2D} + \mathbf{E}_2$	??
DISCO	$\mathbf{X}_1 = \mathbf{T}_1\mathbf{P}_{11}^T + \mathbf{T}_2\mathbf{P}_{12}^T + \mathbf{T}_3\mathbf{P}_{13}^T + \mathbf{E}_1 = \mathbf{X}_{1C} + \mathbf{X}_{1DO} + \mathbf{X}_{1DNO} + \mathbf{E}_1$ $\mathbf{X}_2 = \mathbf{T}_1\mathbf{P}_{21}^T + \mathbf{T}_2\mathbf{P}_{22}^T + \mathbf{T}_3\mathbf{P}_{23}^T + \mathbf{E}_2 = \mathbf{X}_{2C} + \mathbf{X}_{2DNO} + \mathbf{X}_{2DO} + \mathbf{E}_2$	$R(\mathbf{T})$ is unique
GSVD	$\mathbf{X}_1 = \mathbf{T}\mathbf{D}_1\mathbf{V}_1^T + \mathbf{E}_1 = \mathbf{T}_1\mathbf{D}_{11}\mathbf{V}_{11}^T + \mathbf{T}_2\mathbf{D}_{12}\mathbf{V}_{12}^T + \mathbf{T}_3\mathbf{D}_{13}\mathbf{V}_{13}^T + \mathbf{E}_1 = $ $\mathbf{X}_{1DO} + \mathbf{X}_{1C} + \mathbf{X}_{1DNO} + \mathbf{E}_1$ $\mathbf{X}_2 = \mathbf{T}\mathbf{D}_2\mathbf{V}_2^T + \mathbf{E}_2 = \mathbf{T}_1\mathbf{D}_{21}\mathbf{V}_{21}^T + \mathbf{T}_2\mathbf{D}_{22}\mathbf{V}_{22}^T + \mathbf{T}_3\mathbf{D}_{23}\mathbf{V}_{23}^T + \mathbf{E}_2 = $ $\mathbf{X}_{2DNO} + \mathbf{X}_{2C} + \mathbf{X}_{2DO} + \mathbf{E}_2$	See Section 3.4
JIVE	$\mathbf{X}_k = \mathbf{T}(\mathbf{P}_{kC}^T) + \mathbf{T}_k(\mathbf{P}_{kD})^T + \mathbf{E}_k = \mathbf{X}_{kC} + \mathbf{X}_{kD} + \mathbf{E}_k$	Subspaces unique
SRDF	$\mathbf{X}_k = \mathbf{T}\mathbf{D}_k\mathbf{V}_k^T + \mathbf{E}_k = \mathbf{X}_{kC} + \mathbf{X}_{kD} + \mathbf{E}_k$	??

Table 2: **Some properties of the fusion methods**

Methods	Subspace properties	Orthogonality
SCA	$R(\mathbf{T}) \subseteq R[\mathbf{X}_1\|...\|\mathbf{X}_K]$; $R(\mathbf{T}) \nsubseteq R[\mathbf{X}_k]$; $R(\widetilde{\mathbf{T}}_k) \subseteq R(\mathbf{X}_k)$	$\mathbf{T}^T\mathbf{T} = \mathbf{I}$
GCA	$R(\mathbf{T}) \subseteq R[\mathbf{X}_1\|...\|\mathbf{X}_K]$; $R(\mathbf{T}) \nsubseteq R[\mathbf{X}_k]$; $R(\widetilde{\mathbf{T}}_k) \subseteq R(\mathbf{X}_k)$	$\mathbf{T}^T\mathbf{T} = \mathbf{I}$
O2PLS	$R(\mathbf{X}_{kC}) \subseteq R(\mathbf{X}_k)$; $R(\mathbf{X}_{kD}) \subseteq R(\mathbf{X}_k)$	$\mathbf{X}_{kC}^T\mathbf{X}_{kD} = 0$; $\mathbf{E}_k^T\mathbf{X}_{kD} = 0$; $\mathbf{E}_k^T\mathbf{X}_{kC} \neq 0$; $\mathbf{X}_{kC}^T\mathbf{X}_{k'D} \neq 0 (k \neq k')$; $\mathbf{X}_{kD}^T\mathbf{X}_{k'D} \neq 0 (k \neq k')$
DISCO	$R(\mathbf{T}) \subseteq R[\mathbf{X}_1\|...\|\mathbf{X}_K]$ $R(\mathbf{T}_l) \nsubseteq R(\mathbf{X}_k); l = 1,2,3; \forall k$	All orthogonal except: $(\mathbf{X}_{1DO})^T\mathbf{X}_{2DNO} \neq 0$; $(\mathbf{X}_{2DO})^T\mathbf{X}_{1DNO} \neq 0$
GSVD	$R(\mathbf{T}) \subseteq R[\mathbf{X}_1\|\mathbf{X}_2]$; $R(\mathbf{T}_l) \nsubseteq R(\mathbf{X}_k)$; $l = 1,2,3; k = 1,2$	$\mathbf{T}^T\mathbf{T} \neq 0$; $\mathbf{X}_{1DO}$, $\mathbf{X}_{1DNO}$ and $\mathbf{X}_{1C}$ mutually orthogonal; $\mathbf{X}_{2DO}$, $\mathbf{X}_{2DNO}$ and $\mathbf{X}_{2C}$ mutually orthogonal
JIVE	$R(\mathbf{T}) \nsubseteq R(\mathbf{X}_k)$; $R(\mathbf{T}_k) \subseteq R(\mathbf{X}_k)$; $R(\mathbf{T}) \subseteq R[\mathbf{X}_1\|...\|\mathbf{X}_K]$	$(\mathbf{X}_{k'C})^T\mathbf{X}_{kD} = 0; \forall k, k'$; $(\mathbf{X}_{k'D})^T\mathbf{X}_{kD} \neq 0 (k \neq k')$
SRDF	??	No orthogonality

Table 3: Overview of DISCO models for the medical biology example. The table shows models for the four lowest fit-values for rotations based on 3-5 SCA-components. Components are labeled as common (C-) or distinct (D-). The colored components are subspaces that are the same across several models and the correlation between these subspaces are given in Table 3. The framed model is selected for further interpretation.

SCA comp	Increasing Fit Values								ExplVar
	1		2		3		4		
3	C-AL C-AL C-ALO	0.13	C-AL C-AO C-LO	0.15	C-AO C-ALO C-ALO	0.16	C-ALO C-ALO C-ALO	0.20	53%
4	C-ALO C-ALO C-ALO C-ALO	0.19	D-A C-AL C-AL C-ALO	0.20	D-A D-O C-ALO C-ALO	0.23	C-AL C-AL C-AL C-ALO	0.24	58%
5	D-A D-O C-AL C-AL C-ALO	0.24	C-AL C-AO C-ALO C-ALO C-ALO	0.26	D-A D-O C-AL C-ALO C-ALO	0.28	D-A D-O C-LO C-ALO C-ALO	0.28	63%

References

[1] E. Acar, R. Bro, and A.K. Smilde. Data fusion in metabolomics using coupled matrix and tensor factorizations. *Proceedings of the IEEE*, 103(9):1602–1620, 2015.

[2] E. Acar, E.E. Papalexakis, G. Grdeniz, M.A. Rasmussen, A.J. Lawaetz, M. Nilsson, and R. Bro. Structure-revealing data fusion. *BMC Bioinformatics*, 15:239, 2014.

[3] O. Alter, P.O. Brown, and D. Botstein. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proceedings of the National Academy of Sciences of the United States of America*, 100:3351–3356, 2003.

[4] M. Bevilacqua, R. Bucci, S. Materazzi, and F. Marini. Application of near infrared (nir) spectroscopy coupled to chemometrics for dried egg-pasta characterization and egg content quantification. *Food Chemistry*, 140(4):726, 734 2013.

[5] F.L. Bookstein. Partial least squares: a dose response model for measurement in the behaviioral and brain sciences. *Psycoloquy*, 5(23):1, 1994.

[6] R. Bro, E. M. Qannari, H. A. L. Kiers, T. Naes, and M. B. Frst. Multi-way models for sensory profiling data. *J.Chemom.*, 22:36–45, 2008.

[7] R. Bro and A.K. Smilde. Centering and scaling in component analysis. *Journal of Chemometrics*, 17:16–33, 2003.

[8] M. Bylesjo, D. Eriksson, M. Kusano, T. Moritz, and J. Trygg. Data integration in plant biology: the o2pls method for combined modeling of transcript and metabolite data. *Plant Journal*, 52(6):1181–1191, 2007.

[9] R. Consonni, L.R. Cagliani, M. Stocchero, and S. Porretta. Evaluation of the production year in italian and chinese tomato paste for geographical determination using o2pls models. *Journal of Agricultural and Food Chemistry*, 58(13):7520, 7525 2010.

[10] N. Correa, T. Adali, and V.D. Calhou. Canonical correlation analysis for data fusion and group iinference: examining applications of medical imaging data. *IEEE Signal Processing Magazine*, 27:39–50, 2010.

[11] T. Dahl and T. Naes. A bridge between tucker-1 and carroll's generalized canonical analysis. *Computational Statistics & Data Analysis*, 50(11):3086–3098, 2006.

[12] L. De Lathauwer. An extension of the generalized svd for more than two matrices. Internal Report 09-206, ESAT-SISTA, KU Leuven (Leuven, Belgium), 2009.

[13] B. De Moor. On the structure and geometry of the product singular value decomposition. *Linear Algebra and its Applications*, 168:95–136, 1992.

[14] B. De Moor and H. Zha. A tree of generalizations of the ordinary singular value decomposition. *Linear Algebra and its Applications*, 147:469–500, 1991.

[15] O.E. de Noord and E.H. Theobald. Multilevel component analysis and multilevel pls of chemical process data. *Journal of Chemometrics*, 19(5-7):301–307, 2005.

[16] G. H. Golub and C. Van Loan. *Matrix Computations*. John Hopkins University Press (Third Edition), 1996.

[17] M. Hanafi and H.A.L. Kiers. Analysis of k sets of data, with differential emphasis on agreement between and within sets. *Computational Statistics & Data Analysis*, 51(3):1491–1508, 2006.

[18] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[19] J.J. Jansen, H.C.J. Hoefsloot, J van der Greef, M.E. Timmerman, and A.K. Smilde. Multilevel component analysis of time-resolved metabolic fingerprinting data. *Analytica Chimica Acta*, 530(2):173–183, 2005.

[20] J.R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58:433–460, 1971.

[21] G.M. Kirwan, T. Hancock, K. Hassell, J.O. Niere, D. Nugegoda, S. Goto, and M.J. Adams. Nuclear magnetic resonance metabonomic profiling using to2pls. *Analytica Chimica Acta*, 781:33, 40 2013.

[22] D. Lahat, T. Adali, and C. Jutten. Multimodal data fusion: an overview of methods, challenges and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, 2015.

[23] M.A. Lips, J.B. Van Klinken, V. Van Harmelen, H.K. Dharuri, P.A.C. 't Hoen, J.F.J. Laros, G.J. Van Ommen, I.M. Janssen, B. Van Ramshorst, B.A. Van Wagensveld, D.J. Swank, F. Van Dielen, A. Dane, A. Harms, R. Vreeken, T. Hankemeier, J.W.A. Smit, H. Pijl, and K. Willems van Dijk. Roux-en-y gastric bypass surgery, but not calorie restriction, reduces plasma branched-chain amoni acids in obese women independent of weight loss or the presence of type 2 diabetes mellitus. *Diabetes Care*, 37(12):3150–3156, 2014.

[24] E.F. Lock, K.A. Hoadley, J.S. Marron, and A.B. Nobel. Joint and individual variation explained (jive) for integrated analysis of multiple data types. *Ann Appl Stat*, 7(1):523–542, 2013.

[25] T. Lofstedt. *OnPLS*. PhD thesis, Umea University, 2012.

[26] T. Lofstedt, D. Hoffman, and J. Trygg. Global, local and unique decompositions in onpls for multiblock data analysis. *Analytica Chimica Acta*, 791:13–24, 2013.

[27] T. Lofstedt and J. Trygg. Onpls-a novel multiblock method for the modelling of predictive and orthogonal variation. *Journal of Chemometrics*, 25(8):441–455, 2011.

[28] C.J. Lynch and S.H. Adams. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nature Reviews Endocrinology*, 10(723-736), 2014.

[29] I. Mage, E. Menichelli, and T. Naes. Preference mapping by po-pls: Separating common and unique information in several data blocks. *Food Quality and Preference*, 24(1):8–16, 2012.

[30] E. Mattarucchi, M. Stocchero, J.M. Moreno-Rojas, G. Giordano, F. Reniero, and C. Guillou. Authentication of trappist beers by lc-ms fingerprints and multivariate data analysis. *Journal of Agricultural and Food Chemistry*, 58(23):12089–12095, 2010.

[31] T. Naes, O. Tomic, N.K. Afseth, V. Segtnan, and I. Mage. Multi-block regression based on combinations of orthogonalisation, pls-regression and canonical correlation analysis. *Chemometrics and Intelligent Laboratory Systems*, 124:32–42, 2013.

[32] J. Pages. Collection and analysis of perceived product inter-distances using multiple factor analysis: Application to the study of 10 white wines from the loire valley. *Food Quality and Preference*, 16(7):642–649, 2005.

[33] C.C. Paige and M.A. Saunders. Towards a generalized singular value decomposition. *SIAM J. Numer. Anal.,*, 18(3):398–405, 1981.

[34] P.R. Peres-Neto, P. Legendre, S. Dray, and D. Borcard. Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology*, 87(10):2614–2625, 2006.

[35] E.A. Petrakis, L.R. Cagliani, M.G. Polissiou, and R. Consonni. Evaluation of saffron (crocus sativus l.) adulteration with plant adulterants by h-1 nmr metabolite fingerprinting. *Food Chemistry*, 173(890):896, 2015.

[36] S.P. Ponnapalli, M.A. Saunders, C.F. Van Loan, and O. Alter. A higher-order generalized singular value decomposition for comparison of global mrna expression from multiple organisms. *Plos One*, 6(12):e28072, 2011.

[37] P. Ray, L.L. Zheng, J. Lucas, and L. Carin. Bayesian joint analysis of heterogeneous genomics data. *Bioinformatics*, 30(10):1370–1376, 2014.

[38] P.H. Schonemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31:1–10, 1966.

[39] J.R. Schott. *Matrix analysis for Statistics*. Wiley and Sons, 1997.

[40] M. Schouteden, K. Van Deun, S. Pattyn, and I. Van Mechelen. Sca with rotation to distinguish common and distinctive information in linked data. *Behavior Research Methods*, 45(3):822–833, 2013.

[41] R.F. Shan, Y. Zhao, M.L. Fan, X.W. Liu, W.S. Cai, and X.G. Shao. Multilevel analysis of temperature dependent near-infrared spectra. *Talanta*, 131(170):174, 2015.

[42] N. Sidiropoulos and R. Bro. On communication diversity for blind identifiability and uniqunique of low-rank decompositions of n-way arrays. *Proc. Int. Conf. Acoust. Speech and Signal Process.*, 5:2449–2452, 2000.

[43] U. Simsekli, B. Ermis, A. T. Cemgil, and E. Acar. Optimal weight learning for coupled tensor factorization with mixed divergences. In *Proc. 21st. Eur. Signal Process. Conf.*, 2013.

[44] A. Singh and G.J. Gordon. Relational learning via collective matrix factorization. In *Knowledge Discovery and Data Mining (KDD)*, Las Vegas, 2008.

[45] A.K. Smilde, R. Bro, and P. Geladi. *Multi-way analysis: Applications in the Chemical Sciences*. John Wiley & Sons Inc., 2004.

[46] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.J.A.N. Lamers, J. van der Greef, and M.E. Timmerman. Anova-simultaneous component analysis (asca): a new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13):3043–3048, 2005.

[47] A.K. Smilde, J.A. Westerhuis, and S. de Jong. A framework for sequential multiblock component methods. *Journal of Chemometrics*, 17(6):323–337, 2003.

[48] V. Srivastava, O. Obudulu, J. Bygdell, T. Lofstedt, P. Ryden, R. Nilsson, M. Ahnlund, A. Johansson, P. Jonsson, E. Freyhult, J. Qvarnstrom, J. Karlsson, M. Melzer, T. Moritz, J. Trygg, T. R. Hvidsten, and G. Wingsle. Onpls integration of transcriptomic, proteomic and metabolomic data shows multilevel oxidative stress responses in the cambium of transgenic hipi- superoxide dismutase populus plants. *Bmc Genomics*, 14:893, 2013.

[49] J. Szymanski, Y. Brotman, L. Willmitzer, and A. Cuadros-Inostroza. Linking gene expression and membrane lipid composition of arabidopsis. *Plant Cell*, 26(3):915–928, 2014.

[50] Y. Tao, D. Wu, D.W. Sun, A. Gorecki, W. Blaszczak, J. Fornal, and T. Jelinski. Quantitative and predictive study of the evolution of wine quality parameters during high hydrostatic pressure processing. *Innovative Food Science and Emerging Technologies*, 20:81, 90 2013.

[51] R. Tauler, A.K Smilde, and B.R. Kowalski. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *Journal of Chemometrics*, 9(1):31–58, 1995.

[52] J.M.F. Ten Berge. Orthogonal procrustes rotation for 2 or more matrices. *Psychometrika*, 42(2):267–276, 1977.

[53] A. Tenenhaus and M. Tenenhaus. Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of Operational Research*, 238(2):391–403, 2014.

[54] R. Tibshirani. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society Series B-statistical Methodology*, 73:273–282, 2011.

[55] M.E. Timmerman, H.C.J. Hoefsloot, A.K. Smilde, and E. Ceulemans. Scaling in asca. *Metabolomics*, accepted, 2015.

[56] M.E. Timmerman and H.A.L. Kiers. Four simultaneous component models of multivariate times series from more than one subjcet to model intraindividual and interindividual differences. *Psychometrika*, 86:105–122, 2003.

[57] A. Tomassini, A. Vitalone, F. Marini, G. Pratico, F. Sciubba, M. Bevilacqua, M. Defini, A. Di Sotto, S. Di Giacomo, P. Mariani, C.L. Mammola, E. Gaudio, A. Miccheli, and G. Mazzanti. H-1 nmr-based urinary metabolic profiling reveals changes in nicotinamide pathway intermediates due to postnatal stress model in rat. *Journal of Proteome Research*, 13(12):5848, 5859 2014.

[58] J. Trygg. O2-pls for qualitative and quantitative analysis in multivariate calibration. *Journal of Chemometrics*, 16(6):283–293, 2002.

[59] J. Trygg and S. Wold. O2-pls, a two-block (x-y) latent variable regression (lvr) method with an integral osc filter. *Journal of Chemometrics*, 17(1):53–64, 2003.

[60] J.P. Van de Geer. Linear relationns among k sets of variables. *Psychometrika*, 49(1):79–94, 1984.

[61] R.A. Van den Berg, H.C.J.. Hoefsloot, J.A. Westerhuis, A.K. Smilde, and M.J. van der Werf. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7(142), 2006.

[62] R.A. Van den Berg, C.M. Rubingh, J.A. Westerhuis, M.J. van der Werf, and A.K. Smilde. Metabolomics data exploration guided by prior knowledge. *Analytica Chimica Acta*, 651(2):173–181, 2009.

[63] E. van der Burg and G. Dijksterhuis. Generalized canonical analysis of individual sensory profiles and instrumental data. In T. Naes and E. Risvik, editors, *Multivariate analysis of data in sensory science*. Elsevier, 1996.

[64] F.M. Van der Kloet, P. Sebastian-Leon, A. Conesa, A.K. Smilde, and J.A. Westerhuis. Separating common from distinct variation. *BMC Bioinformatics*, Accepted, 2015.

[65] K. Van Deun, A.K. Smilde, L. Thorrez, H.A.L. Kiers, and I. Van Mechelen. Identifying common and distinctive processes underlying multiset data. *Chemometrics and Intelligent Laboratory Systems*, 129:40–51, 2013.

[66] K. Van Deun, A.K.. Smilde, M.J. van der Werf, H.A.L. Kiers, and I. Van Mechelen. A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, 10:246, 2009.

[67] K. Van Deun, L. Van Mechelen, I.and Thorrez, M. Schouteden, B. De Moor, M.J. van der Werf, L. De Lathauwer, A.K. Smilde, and H.A L. Kiers. Disco-sca and properly applied gsvd as swinging methods to find common and distinctive processes. *PLoS One*, 7:5, 2012.

[68] C.F. Van Loan. Generalizing the singular value decomposition. *SIAM Journal on Numerical Analysis*, 13:76–83, 1976.

[69] I. Van Mechelen and A.K. Smilde. A generic linked-mode decomposition model for data fusion. *Chemometrics and Intelligent Laboratory Systems*, 104:83–94, 2010.

[70] J. A. Westerhuis, T. Kourti, and J. F. MacGregor. Analysis of multiblock and hierarchical pca and pls models. *Journal of Chemometrics*, 12(5):301–321, 1998.

[71] T.F. Wilderjans, E. Ceulemans, I. van Mechelen, and R.A. van den Berg. Simultaneous analysis of coupled data matrices subject to different amounts of noise. *British J. Math. Stat. Psychol.*, 64:277290, 2011.

[72] H Yanai, K. Takeuchi, and Y. Takane. *Statistics for Social and Behavioral Sciences*. Springer, 2011.