INVITED PAPER

Component- and Factor-Based Models for Data Fusion in the Behavioral Sciences

This paper considers multiblock and factor-based methods for data fusion on coupled data matrices that all share the same data mode and demonstrates how to capture both similarities and differences in the data matrices by varying parameters in the objective function.

By Iven Van Mechelen and Eva Ceulemans

ABSTRACT | We start from a few examples of coupled behavioral sciences data, along with associated research questions and data-analytic methods. Linking up with these, we introduce a few concepts and distinctions, by means of which we specify the focus of this paper: 1) data that take the form of a collection of coupled matrices that are linked in either the experimental unit or the variable data mode; 2) associated with questions about the mechanisms underlying these data matrices; 3) which are to be addressed by data-analytic methods that rely on a submodel per data matrix, with a common parameterization of the shared data mode. Next, we outline the principles of two closely related families within this focus: the families of multiblock component- and factor-based models for data fusion (while considering both deterministic and stochastic model variants). Then, we review developments within these families to capture both similarities and differences between the different data matrices under study. We follow with a discussion on recent attempts to address quite a few challenges in data fusion based on multiblock component and factor models, including whether and how to differentially weigh the different data matrices under study, and problems such as dealing with large numbers of variables, outliers, and missing values. While the focus of this paper is on data and modeling contributions from the behavioral sciences, we point in a concluding section at their relevance for other domains and at the importance of related methods developed in those domains.

Manuscript received November 20, 2014; revised February 20, 2015; accepted May 21, 2015. Date of publication August 13, 2015; date of current version August 20, 2015. This work was supported in part by the Research Fund of KU Leuven (GOA/15/003) and by the Interuniversity Attraction Poles program financed by the Belgian government (IAP/P7/06).
The authors are with the University of Leuven, 3000 Leuven, Belgium (e-mail: Iven.VanMechelen@ppw.kuleuven.be; Eva.Ceulemans@ppw.kuleuven.be).

Digital Object Identifier: 10.1109/JPROC.2015.2442652

KEYWORDS | Behavioral science; data integration; principal component analysis (PCA)

I. DATA, RESEARCH QUESTIONS, AND A CLASS OF MODELS

A. Data

Today, complex data are ubiquitous within the behavioral sciences in general and psychology in particular, not in the least because of significant advances in measurement technology. An especially challenging class of complex data is that of so-called multiblock, multiset, coupled, or linked(-mode) data. This type of data is illustrated by the two hypothetical examples outlined in Fig. 1. Fig. 1(a) displays schematically the data blocks obtained in a functional magnetic resonance imaging (fMRI) study in which participants were presented emotion-eliciting stimulus images. The contents of these images were experimentally varied with regard to a set of stimulus features, including arousal and valence. Upon the presentation of each image, the hemodynamic response was registered in all voxels of the brain across a series of time points. Fig. 1(b) shows a second example of multiblock data, now taken from a study in (contextualized) personality psychology. On the one hand, participants in this study were rated with regard to their level of aggressive behavior in a series of frustrating situations, which varied in terms of a set of so-called active psychological features (such as whether the frustration was deliberately caused by someone else). On the other hand, self-report data on the same participants were collected with regard to a set of dispositional variables (personality traits, abilities, etc.).

The basic constituents of multiblock data are data blocks, which can be formalized as mappings f from

0018-9219 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

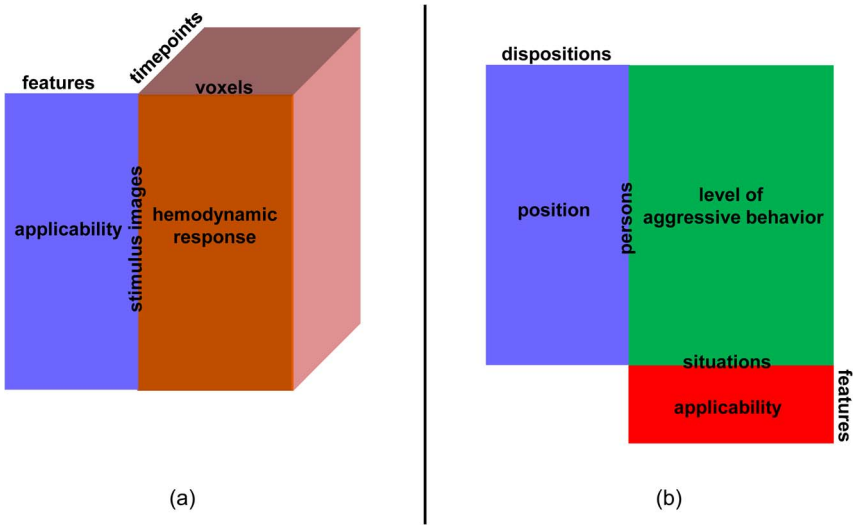


Fig. 1. *Two hypothetical examples of multiblock data from psychology. (a) Two linked data blocks from an fMRI study on emotions. (b) Three linked data blocks from a study on individual differences in aggressive behavior in contextualized personality psychology.*

Cartesian products of sets (called data modes) to some univariate range Y, $f : (S_1 \times S_2 \times \cdots \times S_N) \rightarrow Y$. Linked data then comprise a connected whole of data blocks, with two data blocks being connected if their corresponding Cartesian products have at least one data mode in common. For example, in Fig. 1(a), the two data blocks are connected because they share the stimulus image data mode; from its part, Fig. 1(b) displays three connected data blocks, the first two of which share the person data mode, and the last two the situation data mode. (For a more comprehensive introduction into multiblock data, see [1].)

The focus of this paper is on coupled data matrices that all share the same data mode, that is, either the row or the column mode. Without loss of generality we assume that the rows of each matrix correspond to a data mode of experimental units (or objects) and the columns to a data mode of variables. Fig. 2 gives a schematic representation of the two resulting data structures. Fig. 2(a) displays data from a hypothetical study of emotions experienced in daily life (obtained with a so-called experience-sampling methodology). In this case, the data comprise situation by emotion matrices, one for each participant in the study (three of which are displayed in the panel), which all share the data mode of the variables (i.e., the emotions under study). Fig. 2(b) displays data from a hypothetical study with children on the link between loneliness and depression on the one hand and hypervigilance for social threat on the other hand. The data comprise three matrices which pertain to: 1) percent fixation time on socially threatening stimuli in film clips (obtained with an advanced eye-tracking methodology); 2) measurements of loneliness; and 3) measures of depression; these matrices share the data mode of the experimental units (i.e., the children under study).

B. Research Questions

Coupled matrices, and multiblock data more in general, may go with a plethora of questions. For example, in the fMRI study in Fig. 1(a), interest may be in predicting hemodynamic responses to stimuli on the basis of the manipulated stimulus characteristics. In the study from personality psychology in Fig. 1(b), the primary interest of the researcher may be in the understanding of the structure of individual differences in aggressive behavior in context, with the two other data blocks playing an auxiliary role only. In the study on emotions experienced in daily life in Fig. 2(a), one may wish to understand the common structure and the common underlying mechanisms that underlie the covariation between the emotions under study in each of the person-specific data blocks. In the study from child psychology in Fig. 2(b), the research question may pertain to the association between individual differences on overall measures of fixation time, loneliness, and depression.

To structure the different questions that may go with multiblock data, the answers to three meta-questions may be most helpful. 1) Do all data blocks under study have similar levels of importance or priority? (For instance, in the example above on aggressive behavior in context, this obviously is not the case, unlike in the other examples.) 2) Do all data blocks under study have the same role from the viewpoint of the research question? (For instance, in the fMRI study, this is not the case, with the stimulus feature information appearing in the role of predictor block, and the hemodynamic response information in the role of criterion block.) 3) Is the primary research interest in global summary measures derived from each of the data blocks under study (such as a sum score per experimental unit or a rank order of the variables), or in the full block

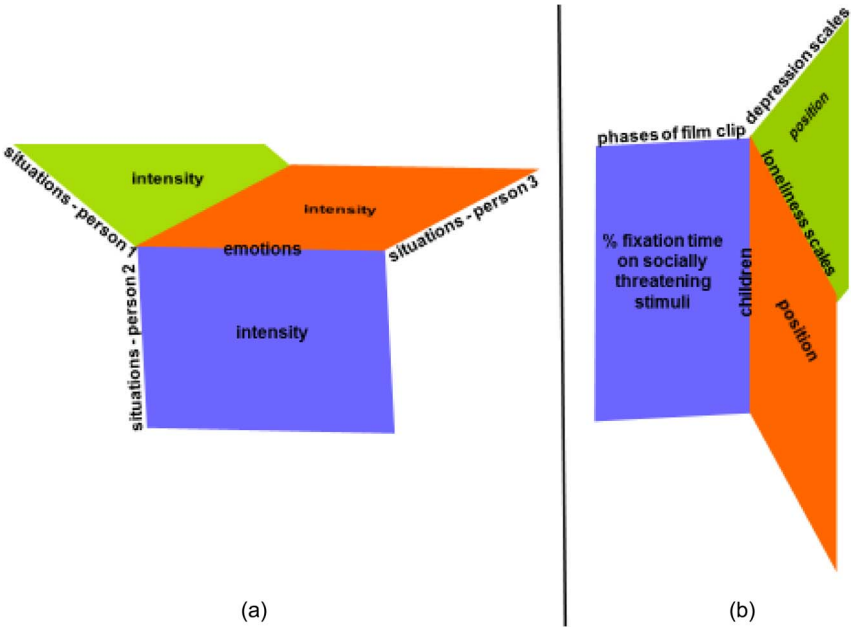


Fig. 2. *Examples of the two multiblock data structures focused on in this paper. Data matrices all of which share (a) the variable or (b) the experimental unit data mode.*

information? (For instance, in the example from child psychology, the research interest may be limited to sum scores of hypervigilance, loneliness, and depression for each child; in contrast, in the example of emotions experienced in daily life, interest may be in the full within-block covariation in all person-specific data blocks under study.)

In the remainder of this paper, we will exclusively focus on research questions which imply that: 1) all data matrices under study have equal levels of priority or importance; 2) all data matrices under study are exchangeable in terms of their formal role; and 3) interest is in the full information included in the data matrices rather than in derived summary measures.

C. A Class of Models

Both inside and outside the behavioral sciences, a wealth of methods and models has been introduced to analyze the full information included in multiblock data that consist of a collection of exchangeable coupled data matrices. Examples include various individual differences versions of spatial distance models (such as INDSCAL or individual differences multidimensional scaling) reviewed in [2], generalized singular value decomposition (SVD) [3], (regularized) generalized or multiset canonical correlation analysis [4], [5], and O2-PLS[1] [6].

These methods differ in terms of: 1) the number of coupled data matrices they can deal with (i.e., 2 versus > 2, with, e.g., generalized SVD originally being developed for two data blocks only); 2) whether they can be applied to matrices coupled in the object or in the variable mode (with, e.g., O2-PLS being developed for data matrices that are coupled in the object mode); 3) whether they include a submodel for each individual matrix that is part of the multiblock data, with this submodel implying a quantification (i.e., a unidimensional or multidimensional parameterization) of each data mode (as is the case for all methods mentioned above); 4) whether the shared data mode is given the same quantification in all submodels in which it is involved (as is the case in INDSCAL and generalized SVD, unlike in regularized generalized canonical correlation analysis and O2-PLS); and 5) whether the model estimation involves the optimization of a single overall objective function (which holds for INDSCAL and regularized generalized canonical correlation analysis, unlike for generalized SVD and O2-PLS). (Note that a description of a powerful generic model for coupled data for which properties 3)–5) hold true is given in [1].)

In this paper, we will focus on two closely related families of models for coupled data matrices, all of which share either the experimental unit or the variable data mode, and which originate to a large extent from research in the behavioral sciences: the families of multiset component- and factor-based models for data fusion. All members of these families satisfy properties 3)–5) and can generally be applied to more than two coupled data matrices; most of them have been developed for the analysis of either matrices coupled in the object mode or matrices coupled in the variable mode, although several of them can be used in both cases. In the behavioral sciences, models of these families are typically

[1]O2-PLS = orthogonal projections to latent structures with separation of variation in each of two blocks that is orthogonal to the other block.

used to identify the mechanisms underlying the data matrices under study. Examples of these include dimensions of (connotative) meaning (underlying stimulus by rating scale data to capture the emotional or attitudinal associations with the stimuli under study), and abilities/personality traits (underlying person by intelligence test or person by personality descriptor data).

The remainder of this paper is structured as follows. In Section II, we will outline the principles of the two families in question. Next, we will review a number of developments within the families to capture both similarities and differences between the different data matrices under study (Section III). We will conclude with a discussion of recent attempts to address quite a few challenges in data fusion based on multiblock component and factor models (Section IV).

The focus of our review will be on the structural characteristics of the different models under study (while considering both deterministic and stochastic model variants), and on the associated loss or objective functions. Although for a small subset of the multiset component models the associated loss or objective functions can be optimized by applying SVD to the suitably preprocessed and concatenated data, the optimization of the loss or objective functions is in general a fairly involved problem. For this problem, a very broad range of algorithms has been proposed (for factor analysis even already in the one-block case [7]). A discussion of these would imply a paper in its own and falls outside the scope of this review.

II. THE FAMILY OF MULTISET COMPONENT- AND FACTOR-BASED MODELS FOR DATA FUSION

Before presenting an outline of the families of multiset component and factor models, we briefly recapitulate their one-block counterparts.

A. The One-Block Case

Principal component analysis (PCA) is a well-established technique in multivariate data analysis. It originates from work by Pearson [8], and was introduced in psychometrics by Hotelling [9]. PCA can be considered the solution to several more or less equivalent optimization problems. Here we focus on an optimization aiming at a low-dimensional approximation of the data in a Q-dimensional space as a whole rather than in terms of principal directions within that space. PCA then can be looked at as the solution to

$$\mathbf{X} = \mathbf{1}_{I\times 1}\overline{\mathbf{x}} + \mathbf{F}\mathbf{B}' + \mathbf{E}, \quad \mathbf{F} = \mathbf{X}\mathbf{W}, \qquad \underset{\mathbf{W},\mathbf{B}}{\text{minimize}} \|\mathbf{E}\|^2 \tag{1}$$

with $\mathbf{X}$ denoting the $I \times J$ data matrix, $\mathbf{1}_{I\times 1}$ the $I \times 1$ vector of ones, $\overline{\mathbf{x}}$ the $1 \times J$ vector with the variable means, $\mathbf{F}$ an $I \times Q$ matrix containing the scores of the I experimental units on Q variables called components (which in turn are weighted combinations of the observed variables, with $\mathbf{W}$ denoting a $J \times Q$ matrix of weights), $\mathbf{B}$ a $J \times Q$ matrix with the loadings of the J variables on the Q components, and $\mathbf{E}$ an $I \times J$ matrix of residuals.

Factor analysis (FA) [10] originates in psychometrics. In part, it results from seminal work by Spearman [11] in the intelligence domain. The data are now considered I realizations of J random variables that are denoted by an $1 \times J$ vector $\mathbf{x}$. The factor-analytic model then reads as follows:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{f}\mathbf{B}' + \mathbf{e} \tag{2}$$

with $\boldsymbol{\mu}$ denoting an $1 \times J$ vector with the population means of the J variables, $\mathbf{f}$ an $1 \times Q$ vector representing Q latent variables called (common) factors (i.e., a vector of factor scores), $\mathbf{B}$ a $J \times Q$ matrix with the loadings of the J variables on the Q factors, and $\mathbf{e}$ an $1 \times J$ vector of residuals, comprising both unique factors (one for each variable) and errors. It is further assumed that $E[\mathbf{f}] = \mathbf{0}$, $E[\mathbf{e}] = \mathbf{0}$, $\boldsymbol{\Sigma}_{\mathbf{f}} = \boldsymbol{\Phi}$, $\boldsymbol{\Sigma}_{\mathbf{e}} = \boldsymbol{\Psi}$, and $\text{corr}(\mathbf{f}, \mathbf{e}) = \mathbf{0}$ (with $\boldsymbol{\Sigma}$ denoting a covariance matrix, $\boldsymbol{\Phi}$ the $Q \times Q$ covariance matrix of the common factors, and $\boldsymbol{\Psi}$ the $J \times J$ diagonal covariance matrix of the residuals with as diagonal elements the sum of the unique and the noise variance for each variable). It then follows that:

$$E[\mathbf{x}] = \boldsymbol{\mu}, \qquad \Sigma_{\mathbf{x}} = \mathbf{B}\boldsymbol{\Phi}\mathbf{B}' + \boldsymbol{\Psi}. \tag{3}$$

Often additional assumptions are made on the distribution of $\mathbf{f}$ and $\mathbf{e}$, such as $\mathbf{f} \sim N(\mathbf{0}, \boldsymbol{\Phi})$ and $\mathbf{e} \sim N(\mathbf{0}, \boldsymbol{\Psi})$. It then follows that:

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Phi}\mathbf{B}' + \boldsymbol{\Psi}) \tag{4}$$

which further allows for maximum-likelihood estimation of the model parameters. Otherwise, within the factor-analytic domain a key distinction is drawn between confirmatory and exploratory approaches: Whereas in the exploratory case all parameters are estimated freely, in confirmatory FA prior hypotheses on the loadings in $\mathbf{B}$ are tested, by imposing constraints on some or all of them and by subsequently examining whether the resulting model sufficiently fits the data.

At this point, five remarks should be made. First, comparing (1) and (2) shows that PCA and FA are fairly similar, the primary difference between the two being that PCA was originally proposed as a purely descriptive dimension reduction approach on the level of a sample, whereas FA is based on a stochastic population model. However, some authors considered sample principal components as estimates of their population counterparts (see, e.g., [12]).

Moreover, as shown in [13], the loadings resulting from a PCA are closely related to the maximum-likelihood estimates of the loadings of a factor-analytic model with $\mathbf{\Psi} = \sigma^2\mathbf{I}$; this implies that the most important difference between PCA and FA reads that the latter, unlike the former, allows for heteroscedastic residual variances. However, one should note that both methods largely yield very similar results for most data sets [14]. Second, neither for the PCA approach as introduced here nor for the factor-analytic approach the solution is unique, with any solution being subject to a so-called rotational freedom. Next to possible orthogonality constraints on $\mathbf{B}$, $\mathbf{W}$, or $\mathbf{F}$, various additional criteria have been proposed in the literature to solve this lack of identifiability, most of which formalize some type of "simple structure" for the loadings [15]. Third, whereas solving (1) implies retrieving the component scores $\mathbf{F}$ as weighted linear combinations of the observed variables $\mathbf{X}$, the estimation of the parameters of the factor-analytic model (2) can be dealt with without estimating the scores on the common and unique factors. If one nevertheless wants to estimate those, a so-called indeterminacy problem shows up, which means that the solution to this problem is typically far from unique, over and above the identifiability problem mentioned in the previous point (see, e.g., [16]). This indeterminacy problem, and its possible epistemological implications, has led to lengthy debates within the psychometric community (see, e.g., [17] and subsequent discussion). Fourth, as already mentioned above, in the behavioral sciences one often wishes components or factors to represent some kind of latent mechanisms underlying the data (dimensions of meaning, abilities, personality traits, etc.), with the loadings being of critical importance to interpret the mechanisms in question. In both cases, claims about underlying mechanisms should be handled with care, however, partly because of the identifiability and indeterminacy problems referred to above, and partly because external evidence may be needed to validate such claims. Fifth, in absence of strong prior hypotheses, the choice of the "correct" number Q of components or factors is a challenging problem, for which a host of solutions has been proposed [18].

B. The Multiblock Case

In the multiblock case, the data comprise K data blocks $\mathbf{X}_k$, $k = 1, \ldots, K$. In case the data blocks are coupled or linked through the variable data mode, $\mathbf{X}_k$ is of size $I_k \times J$; in case they are linked through the object data mode, it is of size $I \times J_k$. Below we will successively discuss component- and factor-based models for these two types of linkage. One might wonder why type of linkage is used here as a primary organizing principle as one could expect methods for the two types of linkage to be identical upon a transposition of the data (possibly upon suitable preprocessing steps). However, the latter is far from true in that, apart from a few exceptions, most multiset component and factor models for data fusion deal with objects and variables in a nonexchangeable way. This lack of exchangeability is typically implied by aspects of the model structure (including constraints put on model constituents) and statistical model characteristics (including distributional assumptions).

1) Linkage via Variable Data Mode: Simultaneous component analysis (SCA) generalizes PCA to the multiblock case. In case of blocks linked via the variable mode, the most common SCA variant (SCA-P) constrains the component loadings of the variables (i.e., the so-called pattern matrix) to be identical across blocks [19], [20]. This implies that (1) generalizes to

$$\mathbf{X}_k = \mathbf{1}_{I_k \times 1}\overline{\mathbf{x}}_k + \mathbf{F}_k\mathbf{B}' + \mathbf{E}_k, \qquad \mathbf{F}_k = \mathbf{X}_k\mathbf{W}_k, \quad \underset{\mathbf{W}_k,\mathbf{B}}{\text{minimize}} \sum_k \|\mathbf{E}_k\|^2 \tag{5}$$

with the loading matrix $\mathbf{B}$ not bearing a subscript k, that is to say, being constrained to be identical across the K data blocks. A number of variants of (5) have been proposed in which different restrictions are put on the variances of the block-specific component scores $\mathbf{F}_k$ and their intercorrelations [21]. Moreover, also other SCA variants have been introduced in which the quantification (parameterization) of the variable data mode that is kept constant across the data blocks (possibly upon a proportionality constant) pertains either to the weight matrix $\mathbf{W}$ (SCA-W) or to the so-called structure matrix of correlations between variables and components (SCA-S), rather than to the loading matrix $\mathbf{B}$ [20], [22].

In [23], exploratory and confirmatory simultaneous factor models have been proposed. These imply that each data block $\mathbf{X}_k$ corresponds to some subpopulation, with for the variables in that subpopulation $\mathbf{x}_k$

$$\mathbf{x}_k = \boldsymbol{\mu}_k + \mathbf{f}_k\mathbf{B}_k' + \mathbf{e}_k. \tag{6}$$

From the assumptions $\mathbf{f}_k \sim N(\mathbf{0}, \mathbf{\Phi}_k)$, $\mathbf{e}_k \sim N(\mathbf{0}, \mathbf{\Psi}_k)$ (with $\mathbf{\Psi}_k$ representing a diagonal variance–covariance matrix), and $\text{corr}(\mathbf{f}_k, \mathbf{e}_k) = \mathbf{0}$, it then further follows that:

$$\mathbf{x}_k \sim N\left(\boldsymbol{\mu}_k, \mathbf{B}_k\mathbf{\Phi}_k\mathbf{B}_k' + \mathbf{\Psi}_k\right). \tag{7}$$

Unlike [23] which deals with a more general case, in the present context we focus on the case in which all blocks pertain to the same J variables and in which the loading matrices $\mathbf{B}_k$ are constrained to be identical across all data blocks

$$\mathbf{x}_k \sim N(\boldsymbol{\mu}_k, \mathbf{B}\mathbf{\Phi}_k\mathbf{B}' + \mathbf{\Psi}_k). \tag{8}$$

Otherwise, it should be noted that in a number of simultaneous factor approaches the assumption $\mathbf{f}_k \sim N(\mathbf{0}, \mathbf{\Psi}_k)$ is relaxed by allowing for differences between blocks in the means of the factor scores (which may reveal between-block differences in the dimensions of meaning, abilities, personality traits, etc., underlying the data). Note that the latter implies that the $\boldsymbol{\mu}_k$ parameters get the meaning of intercepts rather than of within-block variable means). The resulting model (in case of no constraint of identical $\mathbf{B}_k$ matrices) then can be written as follows:

$$\begin{aligned}&\mathbf{x}_k = \boldsymbol{\alpha}_k + \mathbf{f}_k\mathbf{B}_k' + \mathbf{e}_k \sim N(\boldsymbol{\beta}_k, \mathbf{\Phi}_k), \quad \mathbf{e}_k \sim N(\mathbf{0}, \mathbf{\Psi}_k)\\ &\mathbf{x}_k \sim N\left(\boldsymbol{\alpha}_k + \boldsymbol{\beta}_k\mathbf{B}_k', \mathbf{B}_k\mathbf{\Phi}_k\mathbf{B}_k' + \mathbf{\Psi}_k\right).\end{aligned} \tag{9}$$

2) Linkage via Object Data Mode: In case of K data blocks linked through the object mode, SCA-P relies on the following model equation:

$$\mathbf{X}_k = \mathbf{1}_{I\times 1}\overline{\mathbf{x}}_k + \mathbf{F}\mathbf{B}_k' + \mathbf{E}_k, \quad \mathbf{F} = [\mathbf{X}_1 \ldots \mathbf{X}_k \ldots \mathbf{X}_K]\mathbf{W}, \quad \underset{\mathbf{W},\mathbf{B}}{\text{minimize}} \sum_k \|\mathbf{E}_k\|^2 \tag{10}$$

with the score matrix $\mathbf{F}$ now not bearing a subscript k, that is to say, being constrained to be identical across the K data blocks.

A factor-analytic counterpart of this approach has been proposed in [24]–[26] under the name "interbattery FA." This name stems from a context in which different tests batteries are given to the same sample of individuals (associated with the question whether factor solutions, and, hence, the ability or personality structures underlying the data, are stable across test batteries). The interbattery factor model then reads as follows:

$$\mathbf{x}_k = \boldsymbol{\mu}_k + \mathbf{f}\mathbf{B}_k' + \mathbf{e}_k \tag{11}$$

with the residuals $\mathbf{e}_k$ now spanning contributions from block-specific factors, unique factors, and errors. It is further assumed that $\mathbf{f} \sim N(\mathbf{0}, \mathbf{\Phi})$, $\mathbf{e}_k \sim N(\mathbf{0}, \mathbf{\Psi}_k)$, with $\mathbf{\Psi}_k$ now representing a square (and usually nondiagonal) variance–covariance matrix, and that $\mathbf{f}$ and $\mathbf{e}_k$ are independent. It then again follows that:

$$\mathbf{x}_k \sim N(\boldsymbol{\mu}_k, \mathbf{B}_k\mathbf{\Phi}\mathbf{B}_k' + \mathbf{\Psi}_k). \tag{12}$$

It may be noted that the remarks made before on the close relation between component and factor approaches, on identifiability problems, on the interpretation of components/factors in terms of latent mechanisms, and on the choice of the "correct" number of components/factors largely also hold in the multiblock case (the most important exception perhaps being that identifiability depends on the particular model variant under study).

III. CAPTURING SIMILARITIES AND DIFFERENCES BETWEEN THE DATA MATRICES UNDER STUDY

The overall aim of the multiset component- and factor-based models outlined in Section II is to capture the within-block (co)variation in all data matrices under study by means of a set of common components or factors. In the behavioral sciences, one usually wishes these to represent a set of common mechanisms (dimensions of meaning, abilities/personality factors, etc.) underlying all matrices in question. In quite a few cases, however, one may wish to capture both similarities and differences between the coupled data matrices under study; otherwise, sometimes the data at hand may imply a need to do so. In this section, we will review a number of approaches within the families of component and factor models for this purpose.

A. Disentangling Common and Distinctive Within-Block Mechanisms

When thinking about possible mechanisms underlying coupled data matrices, one could draw a distinction between mechanisms underlying all matrices and mechanisms underlying a single matrix or a few of them only. Making use of a terminology introduced by Tversky [27] in a theoretical analysis of similarity data, one could refer to these, in case of $K = 2$ coupled data blocks, as common and distinctive mechanisms, respectively. More specifically, starting from the two sets of underlying mechanisms of the $K = 2$ data blocks under study, one could refer to their intersection as the set of common mechanisms and to their two set-theoretical differences as two sets of distinctive mechanisms that each exclusively underlie one data block only. Note that in the $K > 2$ case $2^K - 1$ different types of mechanisms may be distinguished, one for each nonempty subset of the K data blocks.

To formalize the concepts of common and distinctive mechanisms represented by components or factors, one could appeal to the notion of variance accounted for, with common components (factors) accounting for a nonnegligible amount of variance in all matrices under study, and distinctive mechanisms only accounting for variance in the matrices to which they apply. In addition, in order not to blur the concepts in question, one may wish to assume that common components (factors) are uncorrelated with their distinctive counterparts, and that different types of distinctive components are uncorrelated as well.

At this point, however, two remarks may be useful. First, the concept of variance accounted for by a single component (factor) in a single data block goes with a few complications, in particular in case of coupled data matrices that are linked through the variable data mode; indeed, in that case, a

coherent formalization of the concept in question is not readily available. Second, in literature the concept of distinctive mechanisms is somewhat obscured by a loose use of terms as distinctive, specific, and unique, with these terms often not being clearly defined and with a lack of clarity about whether they refer to different concepts or not.

One might already have noted an implicit reference to a distinction between common and distinctive factors in Section II, notably in the description of interbattery FA. Indeed, in the formalization of the interbattery factor model a distinction was drawn between the (common) factors denoted by $\mathbf{f}$ and block-specific factors, with the latter, however, being hidden in the covariance matrices $\mathbf{\Psi}_k$ (which represent the variances/covariances stemming from a sum of block-specific and unique factors). To be sure, however, it is possible to make the factors in question explicit (see, e.g., [28]).

In general, multiset component and factor models primarily aim at capturing commonalities in the structure of the within-block (co)variation of the coupled matrices under study. In practice, however, the resulting components and factors typically represent a mix of common and distinctive mechanisms. Therefore, as a consequence, one may wish to look for component and factor solutions in which common and distinctive mechanisms are represented more purely (see, e.g., [28] and [29]). Below, we briefly review two methods that have been proposed for this purpose within the behavioral sciences.

The first method, proposed in [30] and [31], can be used in conjunction with SCA-P for data matrices linked via the variable or the object data mode and to interbattery FA. It makes use of the nonidentifiability of the component- and factor-analytic solutions in question to rotate these to solutions with components/factors with a clearer common or distinctive status. In case of data matrices coupled via the variable mode, the target for this rotation consists of a partial specification of the block-specific component score matrices $\mathbf{F}_k$ in (5), by entering for a distinctive component zeroes in the corresponding column of $\mathbf{F}_k$ for all blocks k to which the component in question does not apply. A similar partial specification of the block-specific loading matrices $\mathbf{B}_k$ in (10) and (11) is to be used in case of data matrices coupled via the object mode. In [32], DISCO–SCA (i.e., SCA in conjunction with a partially specified target rotation toward a structure with clearer DIStinctive and COmmon components) is shown to have excellent performance, both in an absolute sense and in comparison with a state-of-the-art SVD-based method (GSVD) from a neighboring research domain. It should be noted, however, that DISCO–SCA, just like related methods to tease apart common and distinctive mechanisms, goes with the nontrivial model selection problem of choosing the correct numbers of common and distinctive components.

The second method proposed in [33] (with acronym ECO-POWER) exclusively focuses on the detection of common components in the case of data matrices coupled via the object mode. The exclusive focus on common components resides in the fact that in all other available approaches, common components, unlike their distinctive counterparts, have a residual status only. Indeed, in these approaches distinctive components are forced to have (near-)zero loadings or scores in the blocks to which they do not apply, whereas the loadings or scores of the common components are not controlled. Common components are extracted in ECO-POWER by replacing the loss function in (10) by a so-called power loss function. The latter equals a sum across all components of a product per component of the variances accounted for by that component in all data blocks under study.

B. Between-Block Differences in the Set of Underlying Mechanisms

So far we assumed the same set of mechanisms to underlie all coupled data matrices under study. However, in quite a few cases this may be an unrealistic assumption. For instance, for emotion data of the type outlined in Fig. 2(a), recent work on emotional granularity and emotion differentiation suggests that the number of underlying components (as well as their nature) may differ across participants (see, e.g., [34]), with a lower number reflecting a higher vulnerability and a lower well-being. A possible strategy to trace such differences could be to fit a separate PCA model to each block and to compare the resulting loading matrices. However, with a larger number of blocks, such a strategy would quickly become quite infeasible.

To parsimoniously model differences between blocks with regard to within-block mechanisms, De Roover *et al.* [35]–[39] proposed a clusterwise SCA modeling framework for data matrices coupled through the variable mode; this framework encompasses SCA and separate PCA for each block as special cases. The key idea of clusterwise SCA is that the blocks belong to C mutually exclusive clusters (which have to be induced from the data), and that a separate SCA model holds per cluster of blocks. The clusterwise SCA model and loss function then read as follows:

$$\mathbf{X}_k = \mathbf{1}_{I_k\times 1}\overline{\mathbf{x}}_k + \sum_{c=1}^{C} p_{kc}\mathbf{F}_k^{(c)}\mathbf{B}^{(c)\prime} + \mathbf{E}_k, \quad \mathbf{F}_k^{(c)} = \mathbf{X}_k\mathbf{W}_k, \quad \underset{\mathbf{P},\mathbf{W}_k,\mathbf{B}}{\text{minimize}} \sum_k \|\mathbf{E}_k\|^2 \qquad (13)$$

where p_{kc} denotes an entry of the $K \times C$ binary partition matrix $\mathbf{P}$ that equals 1 when the kth block belongs to the cth cluster and 0 otherwise, $\mathbf{F}_k^{(c)}$ denotes the $I_k \times Q^{(c)}$ component score matrix of the kth block when assigned to the cth cluster, and $\mathbf{B}^{(c)}$ denotes the $J \times Q^{(c)}$ loading matrix of the cth cluster. Note that clusterwise SCA comes down to SCA when $C = 1$ and to a separate PCA per block when $C = K$.

Within the clusterwise SCA framework, several model variants have been proposed. These differ with respect to:

1) the SCA variant used within the clusters (in line with [21]); 2) whether the number of components per cluster $Q^{(c)}$ is kept constant across clusters; and 3) whether some of the components are restricted to be the same across clusters. If the latter is the case, the model equation part in (13) can be rewritten as follows:

$$\mathbf{X}_k = \mathbf{1}_{I_k\times 1}\overline{\mathbf{x}}_k + \mathbf{F}_{k,\text{comm}}\mathbf{B}'_{\text{comm}} + \sum_{c=1}^{C} p_{kc}\mathbf{F}^{(c)}_{k,\text{spec}}\mathbf{B}^{(c)\prime}_{\text{spec}} + \mathbf{E}_k \quad (14)$$

with the subscript comm referring to the scores and loadings of the components that are common to all clusters and spec to components that specifically pertain to a single cluster only, and with the further constraint that $\mathbf{F}'_{k,\text{comm}}\mathbf{F}^{(c)}_{k,\text{spec}} = \mathbf{0}$ to ensure separation of the common and cluster-specific components. Note that cluster-specific is to be distinguished from cluster-distinctive in that cluster-specific components may represent underlying mechanisms that partly also apply to other clusters.

C. Measurement Invariance

In the previous section, we discussed the issue of major between-block differences in the set of underlying mechanisms. In this section, we assume that the set of underlying mechanisms is largely the same for all blocks, to subsequently focus on more subtle between-block differences. For this purpose, we will limit ourselves to data matrices that are coupled via the variable mode.

The guiding example will be that of a set of items or tests (viz., the variables) that are presented to respondents (viz. the objects or experimental units) from groups that differ in terms of ethnicity, sex, or age, with each data block corresponding to a distinct group of respondents. Whereas within this context the set of underlying mechanisms is often assumed to be largely the same across groups, no seldom doubt is thrown on the role of specific items. This issue is often rather hot as it immediately relates to concerns about bias, fairness, and equity in test practices.

In the psychometric literature, these problems have been given extensive attention in the study of measurement invariance [40], [41]. Measurement invariance pertains to a set of latent variables (formally represented by, e.g., a set of common factors) one wishes to measure by means of a set of manifest variables $\mathbf{X}$ (e.g., items of questionnaires or surveys). Measurement invariance with regard to some subdivision of the population under study in a number of subpopulations then means that in each of these subpopulations the conditional distribution of the manifest variables given the latent variables equals the corresponding distribution in the overall population.

Within the context of model (9), the study of measurement invariance includes three different types of tests: 1) whether the loading matrices $\mathbf{B}_k$ are equal across groups (subpopulations); 2) whether the intercepts $\boldsymbol{\alpha}_k$ are equal across groups; and 3) whether the covariance matrices of the residuals $\boldsymbol{\Psi}_k$ are equal across groups. In this section, we focus on the first type of tests only.

If the hypothesis that the loading matrices $\mathbf{B}_k$ are equal across groups has to be rejected, one often will try to rescue the constrained version of model (9) with equal loading matrices, by allowing the loadings to differ across groups for a few items (variables) only. In this regard, one usually recurs to sequential modification procedures. For this purpose, one may rely either on parameter-specific modification indices [42], [43], or on a series of invariance tests [44], all of which are grounded in FA. As a third alternative, one may consider a component-analysis-based procedure proposed in [45]. In this procedure, first a clusterwise SCA is run along with a rotation of all cluster-specific loading matrices to some reference structure. Next, for each component the minimum congruence value is calculated across all pairs of groups. Finally, in a stepwise greedy procedure, items are selected for which the loadings on the component under study are allowed to vary across groups, until the minimum congruence exceeds some critical value.

D. Between-Block Differences in Means

In this section, we revisit the SCA model variants for data matrices coupled through the variable mode. The SCA variants discussed so far for this type of data entirely focused on the identification of mechanisms underlying the within-block covariation. Indeed, in (5), the block-specific variable means are taken care of by the $\mathbf{1}_{I_k\times 1}\overline{\mathbf{x}}_k$ term and are further not involved in the componential representation.

In some cases, however, one may wish to gain simultaneously insight in the mechanisms underlying the between-block differences in means and in the mechanisms underlying the within-block covariation. At this point, the multilevel SCA framework introduced by Timmerman [46] may be most useful.

The core idea behind multilevel SCA is that each data matrix can be uniquely decomposed into a sum of three matrices: a matrix with the grand means of the variables, a matrix with the deviations of the block-specific means from the grand means, and a matrix with the deviations of the actual data entries from the block-specific means. Next, a PCA model can be fitted to the second type of matrices and an SCA model to the matrices of the third type, yielding the following overall model

$$\mathbf{X}_k = \mathbf{1}_{I_k\times 1}\overline{\mathbf{x}} + \mathbf{1}_{I_k\times 1}\mathbf{f}_k^{\text{between}}\mathbf{B}^{\text{between}\prime} + \mathbf{F}_k^{\text{within}}\mathbf{B}^{\text{within}\prime} + \mathbf{E}_k \quad (15)$$

with $\mathbf{f}_k^{\text{between}}$ denoting the $1 \times Q^{\text{between}}$ vector with the scores of block k on the Q^{between} between-block components, $\mathbf{B}^{\text{between}}$ the $J \times Q^{\text{between}}$ matrix with the between-block component loadings, $\mathbf{F}_k^{\text{within}}$ the $I_k \times Q^{\text{within}}$ matrix with the

scores of block k on the Q^{within} within-block components, and $\mathbf{B}^{within}$ the $J \times Q^{within}$ matrix with the within-block component loadings. A heuristic for choosing the optimal numbers of between- and within-block components as well as the optimal SCA variant is discussed in [47].

If the different data blocks are reconceived as different levels of some independent variable under study, the decomposition outlined above suggests that multilevel SCA combines features of a one-way ANOVA with features of a component analysis. The between components shed light on how the independent variable affects the measured variables, whereas the within components represent dependencies among the ANOVA residuals. This type of model has been generalized to cases where the data blocks are cells of more intricate experimental designs (such as, e.g., designs in which two or more independent variables are fully crossed). In case of a very large number of dependent variables (as, e.g., in omics research), traditional data-analytic approaches to the analysis of the resulting data, such as MANOVA, break down. For this type of data, Smilde *et al.* [48] proposed an extension of SCA, called ANOVA–SCA (ASCA). In ASCA, the matrices with the deviations of the block-specific means from the grand means are further split into matrices with estimates of the different main and interaction effects under the "usual" identifiability constraints (including, e.g., that the sum of all main effects should equal zero). Each type of these matrices then is represented in terms of a separate set of components. As an aside, one may note that ASCA with SCA-P of the ANOVA residuals could be considered a combination of particular types of partial reduced rank regression and partial PCA.

IV. CHALLENGES IN DATA FUSION WITH MULTISET COMPONENT- AND FACTOR-BASED MODELS

Data fusion with multiblock component- and factor-based models goes with quite a few challenges. In this section, we review recent attempts to deal with several of these.

A. Loss Functions With Differential Weighing

The loss functions included in (5) and (10) boil down to a summation of residuals of all data entries of all data blocks, implying that in the model estimation each entry in any of the data matrices under study is given the same weight. Yet, in order to identify the substantively most relevant mechanisms (or in order to retrieve the "true" mechanisms), some alternative type of weighing might be preferable. At this point, three important remarks are to be made.

First, in this paper we focused on problems with coupled data blocks (matrices) that are considered exchangeable, not only in terms of their formal role but also in terms of their priority or importance. We stick to this. This implies that we leave aside the possibility of substantive *a priori* arguments in favor of a differential weighing of the data blocks. Rather we will consider possibilities of a differential weighing based on data-analytic (viz., formal and substantively nonspecific) principles.

Second, differential weighing cannot be detached from preprocessing the data in one way or another (e.g., autoscaling each variable under study). Indeed, a differential weighing can be achieved in three different ways, which are partly exchangeable: by including weights in the loss function, by making a weight precede each data block in the model equation, and by preprocessing the data. Moreover, the influence of some particular type of preprocessing may also depend on the type of coupling in the data. Consider, as an example, autoscaling of all variables under study per data block: When data blocks are coupled via the variable mode, this corrects for between-block differences in size (which there reflect between-block differences in the number of experimental units), whereas this is not the case for data blocks coupled via the experimental units.

Third, a fairly comprehensive review of methods for simultaneous component-based data integration [49] identified differences between the methods under study in terms of which type of coupled data they originally focused on (i.e., data matrices coupled via objects versus via variables), and in terms of the nature of the orthogonality and identifiability constraints they put on the component solution. Yet, the most consequential differences between the methods pertained to preprocessing and differential weighing.

Below we will review a number of principles that can be relied on when choosing a suitable preprocessing and weighing strategy in component- and factor-based data fusion. The primary emphasis will be on simultaneous component-based methods, many of which have been originally devised for data matrices coupled via the object mode. The principles are further not always mutually exclusive, in that some of them could be combined.

1) All Data Entries Should Be Given Equal Weights: As already mentioned above, the loss functions included in (5) and (10) imply that all data entries are given equal weights. Such a weighing scheme is used in most methods of SCA, insofar they are not used in conjunction with some rescaling type of preprocessing that would imply an implicit differential weighing. Arguments in favor of this scheme include that it links up with a likelihood-based estimation of a minimal stochastic extension of the models under study (with homoscedastic error within and between data blocks), and that this scheme appeared to outperform a scheme with equal weights for all data blocks in a simulation study on a multiblock component model for coupled multiway data [50].

2) All Variables Should Be Given Equal Weights: A second principle could be to give all variables under study equal weights in the data analysis. This could be achieved by particular types of preprocessing such as dividing each variable (optionally within each block) by its standard deviation or range. In fact, the majority of the simultaneous

component methods reviewed in [49] include some form of variable autoscaling. Arguments in favor of autoscaling each variable under study include that this may remove irrelevant differences between variables in measurement scale. Arguments against include that autoscaling erroneously blows up variables that do not display a significant amount of substantive variation, or, stated differently, that all variables under study are treated as if they display the same amount of substantively meaningful variation, which is often not the case.

3) All Data Blocks Should Be Given Equal Weights: To give all data blocks equal weights, one could consider to preprocess the data by rescaling each data block to a variance of one. Alternatively, for matrices coupled via the variable mode, equal weights could also be achieved by dividing each variable in each block by its block-specific standard deviation. Rescaling each data block to a variance of one is done in the SUM–PCA procedure from chemometrics (for matrices coupled via the object mode) [51], after an autoscaling of all variables. Arguments in favor of this weighing scheme include that between-block differences in size may be incidental and that the different types, levels, or sources of information that are represented by the different data blocks should be given equal importance when searching for the mechanisms underlying the blocks. Arguments against this scheme have been mentioned in the subsection on equal weights for all data entries.

4) Less Redundant Data Blocks Should Be Given Higher Weights: One could argue that matrices with richer (i.e., less redundant) information should be given higher weights in the data fusion. For data blocks that are coupled through the object data mode this could be realized by weighing each data block in the loss function by the inverse of the proportion of variance accounted for in the block in question by its first component. This is done in a method proposed in the French literature under the name Analyse Factorielle Multiple (multiple FA) [52]. In fact, multiple FA uses this weighing along with a correction for the number of variables in each block and an autoscaling of each variable. Arguments pro this scheme include that data blocks that include many variables that are (almost) replications of one another should be punished for this; arguments con that, if the underlying processes are more coherent in some data blocks, this is not a valid reason to dismiss the data blocks in question.

5) Data Blocks That Share More Covariation Information Overlap With Other Data Blocks Should Be Given Higher Weights: Data fusion could be looked at as an exercise to find a compromise or consensus of the different coupled data blocks under study. From this perspective, it may make sense to give more weight in the search for a consensus to data matrices that are in general more similar to the other data matrices. This principle is relied on in the STATIS method for fusion of data matrices coupled through the object data mode [53]. Specifically, STATIS starts by converting each $I \times J$ data block $\mathbf{X}_k$ to an $I \times I$ matrix $\mathbf{S}_k$ of between-object cross-products, with $\mathbf{S}_k = \mathbf{X}_k\mathbf{X}_k'$. Next an $I^2 \times K$ matrix $\mathbf{F}$ is constructed with, as kth column, the vectorized matrix $\mathbf{S}_k$. Subsequently, $\mathbf{F}$ is subjected to a PCA. The block-specific weights used by STATIS then are put equal to the loadings of the K columns of $\mathbf{F}$ on the first principal component of $\mathbf{F}$. Otherwise, one may note that the weights defined in this way do not purely represent overlap in between-object covariation with the other blocks, but are also influenced by the number of variables in the block under study, the magnitude of the values in that block, and the strength of the between-object covariations within that block. Arguments in favor of this weighing scheme may include the desire to robustify the output of the analysis against outlying data blocks (see also Section IV-C), whereas arguments against may include that by dismissing discrepant data blocks one may arrive at an illusory consensus.

6) Less Noisy Data Blocks Should Be Given Higher Weights: Finally, one could argue that less noisy data blocks should be given higher weights. Unless external information on the amount of noise in each data block is available, such a weighing strategy requires tools to estimate this information from the data at hand. For this purpose, Wilderjans *et al.* [54] propose a minimal stochastic extension of the SCA model (5) that includes a homoscedastic normally distributed error term with a block-specific error variance [in [55], a similar extension was proposed for (10)]. Maximum-likelihood estimation of the resulting models then automatically implies a down-weighing of more noisy data blocks. In [44] and [55], results of (synthetic and realistic) simulation studies are presented that show that the proposed weighing scheme implies a better recovery of the underlying true SCA structure (which is an argument in favor of the weighing scheme in question). The gain in recovery is sizeable only, however, if the amount of noise is larger in the data blocks that are larger in size.

B. Dealing With Very Large Numbers of Variables

In case of high-dimensional multiblock data, that is, coupled data matrices that pertain to high numbers of variables, the interpretation of the components or factors in terms of the underlying mechanisms they represent may become daunting. In this case, sparse models may offer a neat way out. A powerful generic framework for this purpose has been outlined in [56]. The framework pertains to SCA models for data matrices coupled via the object mode. It implies that the loss function in (10) is augmented with the sum of a subset of four possible penalty terms, with each penalty term being multiplied with a corresponding tuning parameter. The penalties can be put either on the weights in $\mathbf{W}$ or on the loadings in the matrices $\mathbf{B}_k$. In terms of the loadings in $\mathbf{B}_k$, with b_{jqk} denoting the loading

of the jth variable in the kth block on the qth component, the four possible penalty terms read as follows[2]:

$$\begin{aligned}
&\text{(a)} \quad \sum_{k,q}\sum_{j=1}^{J_k} |b_{jqk}| \\
&\text{(b)} \quad \sum_{k,q}\sum_{j=1}^{J_k} b_{jqk}^2 \\
&\text{(c)} \quad \sum_{k,q}\sqrt{J_k \sum_{j=1}^{J_k} b_{jqk}^2} \\
&\text{(d)} \quad \sum_{k,q}\left(\sum_{j=1}^{J_k} |b_{jqk}|\right)^2 .
\end{aligned} \tag{16}$$

Term (a) is a lasso penalty; it implies variable selection and loading shrinkage at the level of the full set of combinations of components and variables as a whole. Term (b) is a ridge penalty; apart from loading shrinkage, it leads to more or less equal loadings for correlated variables. Term (c) is a group lasso penalty; it targets, for each component, selection and shrinkage of entire block-specific sets of variables or loadings. Finally, term (d) is an elitist lasso penalty; it enforces, for each component, selection and shrinkage within each block-specific set of variables or loadings. Importantly, terms (c) and (d) neatly take advantage of the principle of penalizing in terms of groups of variables in the multiblock context of data fusion. Both of them yield some type of group-based variable selection and loading shrinkage, unlike term (a) which implies that one has no control over where selection and shrinkage take place.

The generic framework (which goes with a generic first-order proximal estimation method that can efficiently deal with over 1000 variables) encompasses existing sparse approaches that rely on a specific penalty term or specific combinations thereof [such as the elastic net $-$(a) + (b)$-$, and the sparse group lasso $-$(a) + (c)$-$]. In addition, it also allows to investigate novel combinations of such terms.

C. Dealing With Outliers

In the previous section, we dealt with variable selection, with various sparse penalties or penalty combinations resulting in a sidelining of a number of variables (at least for some of the components under study). In this section, we will take a closer look at various forms of "object selection." The reason for this is that the loss functions of multiblock component models such as the ones included in (5) and (10) are of a least squares type, implying that component analysis estimates are fairly sensitive to outlying objects. Similarly, multiblock FA parameter estimates may be seriously distorted by outliers as well.

[2]The formulas (c) and (d) are corrected versions of their counterparts in [56] and line up with the code associated with the paper in question.

To solve for this, one may consider to robustify the methods in question, meaning that the loss functions or model estimation procedures are adjusted in such a way that outlying observations no longer disturb the analysis results. The outcome of such methods can be used for two different kinds of purposes. First, one can interpret the resulting model estimates directly. Second, one can use robust approaches as screening methods for detecting whether the data contain outliers. Otherwise, with regard to the latter, a useful feature of robust approaches is that they usually go with visual tools to detect outliers. Such a detection can further lead to a categorical output in terms of a classification of outlying versus non-outlying objects. Subsequently, one can remove the outliers and perform a classical (i.e., nonrobust) analysis on the remaining objects. Alternatively, a screening on the basis of robust methods can lead to a continuous output in terms of characterizing the objects according to a gradual outlyingness score. Subsequently, these scores could be used to differentially weigh the objects in a weighted classical analysis.

Robust multiblock work has mainly focused on component-based models. Here we will briefly review the most important results. To begin with, for the one block case, a number of robust PCA methods have been proposed that can be readily extended to SCA-P models for multiblock data (coupled via the variable or via the object mode). In a first type of approaches, the L_2-norm-based loss function in (1) is replaced by its L_1-norm-based counterpart, which is less sensitive to the presence of outliers (see, e.g., [57]). In a second type of approaches, classical estimators, such as the empirical covariance matrix, are replaced by robust ones, such as the minimum covariance determinant (MCD) estimator (see, e.g., [58]). A limitation of these approaches is that they cannot be used for high-dimensional data, however. A third type of robust PCA/SCA approaches [59], [60] uses projection pursuit techniques and searches for the directions in which the projected observations have the largest robustified variance. They can also be used when the number of variables is larger than the number of objects. Fourth, Hubert *et al.* [61] proposed ROBPCA which combines the ideas of the previous two types of approaches. Moreover, ROBPCA yields a diagnostic outlier map that plots for each object its orthogonal distance to the PCA subspace (i.e., the sum of the squared differences between an object's score vector and the projection thereof on the subspace spanned by the Q first eigenvectors of the robust covariance matrix) versus the score distance (i.e., the Mahalanobis distance between the object's score vector and the center, computed using the MCD estimates of the center and the covariance matrix). Based on these two distances, the objects can be classified into four categories: regular observations (low score and low orthogonal distance), good leverage points (high score distance, but low orthogonal distance), orthogonal outliers (high orthogonal distance, but low score distance), and bad leverage points (high score and high

orthogonal distance). Especially the latter two types of observations may seriously distort classical PCA/SCA results as a minimization of the PCA/SCA loss functions implies a minimization of all orthogonal distances.

Furthermore, a specific robust multiblock component-based method has been proposed in [62]: robust nultilevel SCA. This method uses the spatial median rather than the arithmetic mean when splitting the data in a between-block part and a within-block part, and applies ROBPCA to both parts to obtain robust between-block and within-block components. Robust multilevel SCA further goes with two outlier plots: a between-block plot that reveals which blocks have an outlying center, and a within-block plot that reveals which objects are outlying with regard to their within-block center and/or covariance structure.

D. Dealing With Missing Values

In practice, it often happens that some of the entries of the coupled data matrices are missing. To decide how to handle these missing data, it is important to keep Rubin's three mechanisms of missingness in mind [63]: When the missingness is related neither to observed nor to unobserved data, data are "missing completely at random" (MCAR). In case missingness depends on observed variables in the data set but not on unobserved variables, data are "missing at random" (MAR). "Not missing at random" (NMAR) refers to missingness that depends on the values of unobserved variables.

When confronted with missingness in coupled data, researchers often recur to listwise deletion, that is, discarding all objects that have at least one missing value. Such a strategy may result in large losses of information [64], however, and requires the missing data to be MCAR.

As a way out, within the context of uniblock and multiblock component analysis, several authors have proposed to use iterative imputation methods, which are less restrictive in that they assume MAR only [65]–[67]. These methods yield estimates of both the missing observations and the model parameters. To this end, the least squares component-based loss functions have to be adapted by introducing entry-specific weights, which equal zero when the corresponding entries are missing and one if this is not the case. Such weighted loss functions can be optimized using the estimation methods discussed by Kiers [68].

A drawback of such iterative procedures is that they may be prone to overfitting, in that they lead to low values of the weighted loss functions, indeed, yet also to poor parameter estimates. To avoid overfitting, one can recur to regularization. Recent results on PCA and multilevel SCA suggest that shrinking singular values using a shrinkage term based on [13] may be an effective approach [66], [67].

V. CONCLUDING REMARKS

In this paper, we reviewed the broad domain of multiblock component- and factor-based methods for data fusion. We did so primarily from the angle of applications in the behavioral sciences. Yet, the relevance of the methods under review for a broad range of other fields of application (including, e.g., signal processing) should be obvious. Also, our primary focus was on methodological developments in the behavioral sciences (notwithstanding the fact that we also reported a few developments from other domains, including, in particular, chemistry and systems biology). Apart from the fact that the thus chosen focus of this paper meets the overall goal of the present special issue to bring together contributions from diverse application domains, it also reflects the fact that much work on component and factor models historically originates from the behavioral sciences in general and psychology/psychometrics in particular. It should be clear, however, that in the meantime a considerable amount of work on component- and factor-based methods for data fusion has been done in other domains as well, examples including several contributions on so-called coupled or collective matrix factorization as studied in computer science [69]–[71]. In fact, an in-depth study of the interrelations of the models and methods reviewed in this paper and those developed in domains other than the behavioral sciences would be an interesting and important challenge for future research.

The focus of this paper was on coupled data matrices that all share the same data mode, that is, either the row or the column mode. Yet, many core ideas in the paper can be extended to more complex multiblock data structures (and associated modeling approaches) that may include blocks of order > 2 and more sophisticated forms of between-block linkage (as discussed in [1]). In addition, quite a few concepts, distinctions, and approaches reviewed in this paper may have relevance for the domain of data fusion as a whole. ■

REFERENCES

[1] I. Van Mechelen and A. K. Smilde, "A generic linked-mode decomposition model for data fusion," *Chemometrics Intell. Lab. Syst.*, vol. 104, pp. 84–93, Nov. 2010.

[2] J. D. Carroll and P. Arabie, "Multidimensional scaling," *Annu. Rev. Psychol.*, vol. 31, pp. 609–641, Jan. 1980.

[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," *Proc. Nat. Acad. Sci. USA*, vol. 100, pp. 3351–3356, Mar. 2003.

[4] Y.-O. Li, T. Adalı, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3918–3929, Oct. 2009.

[5] A. Tenenhaus and M. Tenenhaus, "Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis," *Eur. J. Oper. Res.*, vol. 238, pp. 391–403, Oct. 2014.

[6] J. Trygg, "O2-PLS for qualitative and quantitative analysis in multivariate calibration," *J. Chemometrics*, vol. 16, pp. 283–293, Jun. 2002.

[7] M. W. Browne, "A comparison of factor analytic techniques," *Psychometrika*, vol. 33, pp. 267–334, Sep. 1968.

[8] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosoph. Mag.*, ser. 6, vol. 2, pp. 559–572, 1901.

[9] H. Hotelling, "Analysis of a complex of statistical variables into principal

components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, Sep. 1933.

[10] D. N. Lawley and A. E. Maxwell, "Factor analysis as a statistical method," *J. Roy. Stat. Soc. D, Stat.*, vol. 12, no. 3, pp. 209–229, 1962.

[11] C. Spearman, "'General intelligence,' objectively determined and measured," *Amer. J. Psychol.*, vol. 15, pp. 201–292, Apr. 1904.

[12] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[13] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Stat. Soc. B*, vol. 61, pp. 611–622, 1999.

[14] W. F. Velicer, A. C. Peacock, and D. N. Jackson, "A comparison of component and factor patterns: A Monte Carlo approach," *Multivariate Behav. Res.*, vol. 17, pp. 371–388, May 1982.

[15] M. W. Browne, "An overview of analytic rotation in exploratory factor analysis," *Multivariate Behav. Res.*, vol. 36, pp. 111–150, Mar. 2001.

[16] L. Guttman, "The determinacy of factor score matrices with implications for five other basic problems of common-factor theory," *Brit. J. Stat. Psychol.*, vol. 8, pp. 65–81, Nov. 1955.

[17] M. D. Maraun, "Metaphor taken as math: Indeterminacy in the factor analysis model," *Multivariate Behav. Res.*, vol. 31, pp. 517–538, 1996.

[18] U. Lorenzo-Seva, M. E. Timmerman, and H. A. L. Kiers, "The Hull method for selecting the number of common factors," *Multivariate Behav. Res.*, vol. 46, pp. 340–364, Apr. 2011.

[19] J. Levin, "Simultaneous factor analysis of several Gramian matrices," *Psychometrika*, vol. 31, pp. 413–419, Sep. 1966.

[20] H. A. L. Kiers and J. M. F. ten Berge, "Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure," *Brit. J. Math. Stat. Psychol.*, vol. 47, pp. 109–126, May 1994.

[21] M. E. Timmerman and H. A. L. Kiers, "Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences," *Psychometrika*, vol. 68, pp. 105–121, Mar. 2003.

[22] R. E. Millsap and W. Meredith, "Component analysis in cross-sectional and longitudinal data," *Psychometrika*, vol. 53, pp. 123–134, Mar. 1988.

[23] K. G. Jöreskog, "Simultaneous factor analysis in several populations," *Psychometrika*, vol. 36, pp. 409–426, Dec. 1971.

[24] L. R. Tucker, "An inter-battery method of factor analysis," *Psychometrika*, vol. 23, pp. 111–136, Jun. 1958.

[25] R. P. McDonald, "Three common factor models for groups of variables," *Psychometrika*, vol. 35, pp. 111–128, Mar. 1970.

[26] M. W. Browne, "Factor analysis of multiple batteries by maximum likelihood," *Brit. J. Math. Stat. Psychol.*, vol. 33, pp. 184–199, Nov. 1980.

[27] A. Tversky, "Features of similarity," *Psychol. Rev.*, vol. 84, pp. 327–352, Jul. 1977.

[28] A. Klami, S. Virtanen, and S. Kaski, "Bayesian canonical correlation analysis," *J. Mach. Learn. Res.*, vol. 14, pp. 965–1003, Apr. 2013.

[29] 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.*, vol. 7, pp. 523–542, Mar. 2013.

[30] M. Schouteden, K. Van Deun, S. Pattyn, and I. Van Mechelen, "SCA with rotation to distinguish common and distinctive information in linked data," *Behav. Res. Methods*, vol. 45, pp. 822–833, Sep. 2013.

[31] M. Schouteden, K. Van Deun, T. F. Wilderjans, and I. Van Mechelen, "Performing DISCO-SCA to search for distinctive and common information in linked data," *Behav. Res. Methods*, vol. 46, pp. 576–587, Jun. 2014.

[32] K. Van Deun *et al.*, "DISCO-SCA and properly applied GSVD as swinging methods to find common and distinctive processes," *PLoS ONE*, vol. 7, pp. 1–13, May 2012, Art. ID. e37840.

[33] M. Schouteden, K. Van Deun, and I. Van Mechelen, "ECO-POWER: A novel method to reveal common mechanisms underlying linked data," in *Proc. COMPSTAT*, Limassol, Cyprus, 2012, pp. 757–768.

[34] Y. Erbas, E. Ceulemans, M. Pe, P. Koval, and P. Kuppens, "Negative emotion differentiation: Its personality and well-being correlates and a comparison of different assessment methods," *Cogn. Emotion*, vol. 28, pp. 1196–1213, Jan. 2014.

[35] K. De Roover *et al.*, "Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data," *Psychol. Methods*, vol. 17, pp. 100–119, 2012.

[36] K. De Roover, E. Ceulemans, M. E. Timmerman, J. B. Nezlek, and P. Onghena, "Modeling differences in the dimensionality of multiblock data by means of clusterwise simultaneous component analysis," *Psychometrika*, vol. 78, pp. 648–668, Mar. 2013.

[37] K. De Roover, M. E. Timmerman, B. Mesquita, and E. Ceulemans, "Common and cluster-specific simultaneous component analysis," *PLoS ONE*, vol. 8, pp. 1–14, May 2013, Art. ID. e62280.

[38] K. De Roover, M. E. Timmerman, I. Van Mechelen, and E. Ceulemans, "On the added value of multiset methods for three-way data analysis," *Chemometrics Intell. Lab. Syst.*, vol. 129, pp. 98–107, Nov. 2013.

[39] K. De Roover, E. Ceulemans, M. E. Timmerman, and P. Onghena, "A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations," *Brit. J. Math. Stat. Psychol.*, vol. 86, pp. 81–102, Feb. 2013.

[40] W. Meredith, "Measurement invariance, factor analysis and factorial invariance," *Psychometrika*, vol. 58, pp. 525–543, Dec. 1993.

[41] R. E. Millsap, *Statistical Approaches to Measurement Invariance*. New York, NY, USA: Routledge, 2011.

[42] B. M. Byrne, R. J. Shavelson, and B. Muthén, "Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance," *Psychol. Bull.*, vol. 105, pp. 456–466, May 1989.

[43] R. C. MacCallum, "Specification searches in covariance structure modeling," *Psychol. Bull.*, vol. 100, pp. 107–120, Jul. 1986.

[44] G. W. Cheung and R. B. Rensvold, "Testing factorial invariance across groups: A reconceptualization and proposed new method," *J. Manage.*, vol. 25, pp. 1–27, Jan. 1999.

[45] K. De Roover, M. E. Timmerman, J. De Leersnyder, B. Mesquita, and E. Ceulemans, "What's hampering measurement invariance: Detecting non-invariant items using clusterwise simultaneous component analysis," *Front. Psychol.*, vol. 5, Jun. 2014, doi: 10.3389/fpsyg.2014.00604.

[46] M. E. Timmerman, "Multilevel component analysis," *Brit. J. Math. Stat. Psychol.*, vol. 59, pp. 301–320, Nov. 2006.

[47] E. Ceulemans, M. E. Timmerman, and H. A. L. Kiers, "The CHull procedure for selecting among multilevel component solutions," *Chemometrics Intell. Lab. Syst.*, vol. 106, pp. 12–20, Mar. 2011.

[48] A. K. Smilde *et al.*, "ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data," *Bioinformatics*, vol. 21, pp. 3043–3048, Jul. 2005.

[49] 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 Bioinf.*, vol. 10, Aug. 2009, DOI: 10.1186/1471-2105-10-246.

[50] T. F. Wilderjans, E. Ceulemans, and I. Van Mechelen, "Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes," *Comput. Stat. Data Anal.*, vol. 53, pp. 1086–1098, Feb. 2009.

[51] A. K. Smilde, J. A. Westerhuis, and S. de Jong, "A framework for sequential multiblock component methods," *J. Chemometrics*, vol. 17, pp. 323–337, Jun. 2003.

[52] B. Escoffier and J. Pagès, *Analyses Factorielles Simples et Multiples: Objectifs, Méthodes et Interprétation*, 4th ed. Paris, France: Dunod, 2008.

[53] H. l'Hermier des Plantes and B. Thiébaut, "Étude de la pluviosité au moyen de la méthode S.T.A.T.I.S," *Rev. Stat. Appl.*, vol. 25, no. 2, pp. 57–81, Mar. 1977.

[54] 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," *Brit. J. Math. Stat. Psychol.*, vol. 64, pp. 277–290, May 2011.

[55] R. A. van den Berg *et al.*, "Integrating functional genomics data using maximum likelihood based simultaneous component analysis," *BMC Bioinf.*, vol. 10, Oct. 2009, DOI: 10.1186/1471-2105-10-340.

[56] K. Van Deun *et al.*, "A flexible framework for sparse simultaneous component based data integration," *BMC Bioinf.*, vol. 12, Nov. 2011, doi: 10.1186/1471-2105-12-448.

[57] J. P. Brooks, J. H. Dulá, and E. L. Boone, "A pure L_1-norm principal component analysis," *Comput. Stat. Data Anal.*, vol. 61, pp. 83–98, May 2013.

[58] C. Croux and G. Haesbroeck, "Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies," *Biometrika*, vol. 87, pp. 603–618, Sep. 2000.

[59] C. Croux, P. Filzmoser, and M. R. Oliveira, "Algorithms for projection-pursuit robust principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 87, pp. 218–225, Jun. 2007.

[60] M. Hubert, P. Rousseeuw, and S. Verboven, "A fast method for robust principal components with applications to chemometrics," *Chemometrics Intell. Lab. Syst.*, vol. 60, pp. 101–111, Jan. 2002.

[61] M. Hubert, P. Rousseeuw, and K. Vanden Branden, "ROBPCA: A new approach to robust principal components analysis," *Technometrics*, vol. 47, pp. 64–79, Feb. 2005.

[62] E. Ceulemans, M. Hubert, and P. Rousseeuw, "Robust multilevel simultaneous component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 129, pp. 33–39, Nov. 2013.

[63] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, pp. 581–592, Dec. 1976.

[64] J. O. Kim and J. Curry, "The treatment of missing data in multivariate analysis," *Sociol. Method Res.*, vol. 6, pp. 215–241, Nov. 1977.

[65] K. De Roover, E. Ceulemans, and M. E. Timmerman, "How to perform multiblock component analysis in practice," *Behav. Res. Methods*, vol. 44, pp. 41–56, Mar. 2012.

[66] J. Josse and F. Husson, "Multiple imputation in PCA," *Adv. Data Anal. Classification*, vol. 5, pp. 231–246, Oct. 2011.

[67] J. Josse, M. E. Timmerman, and H. A. L. Kiers, "Missing values in multi-level simultaneous component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 129, pp. 21–32, Nov. 2013.

[68] H. A. L. Kiers, "Weighted least squares fitting using ordinary least squares algorithms," *Psychometrika*, vol. 62, pp. 251–266, Jun. 1997.

[69] A. P. Singh and G. J. Gordon, "Relational learning via collective matrix factorization," in *Proc. 14th Int. Conf. Knowl. Disc. Data Mining*, 2008, pp. 650–658.

[70] E. Acar *et al.*, "Coupled matrix factorization with sparse factors to identify potential biomarkers in metabolomics," *Int. J. Knowl. Disc. Bioinf.*, vol. 3, pp. 22–43, Jul. 2012.

[71] G. Bouchard, S. Guo, and D. Yin, "Convex collective matrix factorization," in *Proc. 16th Int. Conf. Artif Intell. Stat.*, 2013, pp. 144–152.

ABOUT THE AUTHORS



Iven Van Mechelen received the M.Sc. degree in mathematics from the University of Antwerp, Antwerpen, Belgium, in 1980, the M.Psych. degree from the University of Leuven, Leuven, Belgium, in 1984, and the Ph.D. degree in psychology from the University of Leuven, in 1989.

From 1985 to 1991, he was a Fellow of the National Fund for Scientific Research. Since 1993, he has been a Professor of Quantitative Psychology and Individual Differences at the University of Leuven. He is the author of more than 120 papers in international peer-reviewed journals. His research interests include clustering methods, models for multiway and multiset data, and the contextualized study of individual differences in personality, emotions, and affective dynamics.

Prof. Van Mechelen was a recipient of the Award of the Research Council of the University of Leuven in 1993. He was the advisor of the doctoral dissertation of Iwin Leenen, who was given the Dissertation Award of the Psychometric Society in 2000. He was a recipient of the Award for the best 2005 paper in the *Journal of Research in Personality* (for a paper coauthored by K. Vansteelandt and J. Nezlek) in 2006. He was elected Fellow of the Association for Psychological Science in 2007. In 2012–2013, he was President of the International Federation of Classification Societies.



Eva Ceulemans received the M.Psych. degree and the Ph.D. degree in psychology from the University of Leuven, Leuven, Belgium, in 1999 and 2003, respectively.

From 1999 to 2007, she was a Fellow of the Fund for Scientific Research, Flanders. Since 2007, she has been a Professor of Methodology of Educational Sciences at the University of Leuven. She is the author of more than 70 papers in international peer-reviewed journals. Her research interests include the modeling of multiset and multiway data, combining clustering with dimension reduction and/or regression, dynamic network modeling, the analysis of dyadic data, and the modeling of emotion dynamics.

Prof. Ceulemans was a recipient of the Chikio Hayashi Award of the International Federation of Classification Societies in 2002, and she was the advisor of the doctoral dissertation of Kim De Roover, who was given the Dissertation Award of the Classification Society in 2014.