A Comprehensive Guide to Multiset Canonical Correlation Analysis and its Application to Joint Blind Source Separation

Isabell Lehmann*, *Student member, IEEE*, Ben Gabrielson*, Tanuj Hasija, *Member, IEEE*, and Tülay Adali, *Fellow, IEEE*

***Abstract*—Multiset Canonical Correlation Analysis (mCCA), also called Generalized Canonical Correlation Analysis (GCCA), is a technique to identify correlated variables across multiple datasets, which can be used for feature extraction in fields like neuroscience, cross-language information retrieval, and recommendation systems, among others. Besides its wide use, there is still a lack of comprehensive understanding of its theory and implementation with different objective functions all under one umbrella. In this paper, we review the five commonly used mCCA methods `sumcor`, `maxvar`, `minvar`, `genvar`, and `ssqcor`. We provide a concise overview of their optimization problems along with their solutions and pseudocodes. After this, we discuss the application of mCCA for estimating underlying latent components in the Joint Blind Source Separation (JBSS) problem and propose the *source identification conditions* of the different mCCA methods, i.e., the conditions under which they are able to achieve JBSS. We substantiate the proposed theoretical conditions with numerical results and test the statistical efficiency of the methods for finite samples. We observe in our experiments that `genvar` appears to have the least restrictive source identification conditions and to be more statistically efficient that the other methods. This suggests that `genvar` is generally the best-performing mCCA method for JBSS except for special cases, which is an important finding, as the most commonly used mCCA methods are `maxvar` and `sumcor`.**

***Index Terms*—multiset canonical correlation analysis, generalized canonical correlation analysis, joint blind source separation, source identification conditions.**

I. Introduction

In real-world applications, data is often collected from multiple *views*, also called *modalities*, such as video and audio, or text and speech, where each view provides complementary information on the underlying phenomenon. Likewise, multiple datasets from a single view can also provide complementary information, e.g., functional Magnetic Resonance Imaging (fMRI) or electroencephalography (EEG) data collected from multiple subjects. With data fusion approaches like joint matrix decompositions, this complementary information can be combined to extract multivariate features that are directly interpretable [1], [2]. Furthermore, these features can be used for multiple purposes, such as anomaly detection, identification of biomarkers of diseases, and classification. One of the earliest joint decomposition approaches in this sense is Canonical Correlation Analysis (CCA) [3], which transforms two datasets into a new space where the correlation of the transformed variables, called the *canonical variables*, is maximized. The extension of CCA to more than two datasets, called multiset Canonical Correlation Analysis (mCCA) [4], multiview Canonical Correlation Analysis (multiview CCA) or Generalized Canonical Correlation Analysis (GCCA) [5], has been frequently used for jointly analyzing multiple datasets.

Early applications in fields like remote sensing [6] and neuroimaging data analysis [7] laid the groundwork for joint data analysis. For instance, mCCA has been used in neuroimaging to identify abnormal brain regions in patients with schizophrenia [8] and to improve understanding of how the dysregulation of specific microRNAs contributes to the loss of brain areas in major depressive disorders [9]. Recent work has also extended mCCA to address new challenges. Scalable mCCA for large-scale word embedding is proposed in [10], and cross-language information retrieval is achieved by mCCA utilizing structural regularization [11] and subspace intersection approaches [5]. SRGCCA (Sparse Regularized GCCA) has been proposed for gene expression analysis [12], and a supervised extension of Sparse Regularized GCCA for identifying multi-omics biomarkers for predicting the survival time in cancer patients [13]. In multi-feature fusion, where the different views are features extracted from the same raw data, supervised mCCA is developed for audio and video-based emotion recognition [14]. Further advancements include deep-learning based extensions of mCCA, e.g., Deep GCCA (DGCCA) for phonetic transcription of acoustic and articulatory measurements and for friend recommendation [15], and DGCCA with an attention mechanism for emotion recognition [16]. Friend recommendation has been enhanced also with graph mCCA [17] and with a graph autoencoder for mCCA [18]. Also in the emerging field of federated learning, where sharing of raw data is not allowed and only derived information (e.g., model parameters) is exchanged among computing agents and a central controller, a new formulation of DGCCA has substantially reduced communication overheads [19]. Next to the so far mentioned applications, where the canonical variables are mainly used as extracted features, another application area of mCCA has been Joint Blind Source Separation (JBSS) [20]–[22], where the canonical variables are interpreted as estimates of unobserved latent variables, called

* The first two authors contributed equally to this paper.

Isabell Lehmann and Tanuj Hasija are with the Signal and System Theory Group, Paderborn University, 33098 Paderborn, Germany.

Ben Gabrielson and Tülay Adali are with the Dept. of CSEE, University of Maryland, Baltimore County (UMBC), Baltimore, MD 21250, USA.

This work was supported by grant NSF 2316420.

Manuscript received February 24, 2025; revised September 03, 2025.

source components, that help to explain the data based on the assumption of a linear mixing model.

While for CCA, maximization of pairwise correlations can be achieved using the Singular Value Decomposition (SVD) of the data's whitened cross-covariance matrix [23], maximizing correlations in mCCA is not as straightforward because mCCA deals with more than two datasets at a time. As there are multiple pairwise correlations to consider, there are several different ways of quantifying correlation within a covariance matrix (such as the sum of the matrix's entries and the sum of the squared entries, among others), each leading to a different mCCA objective function. Kettenring [4] summarized five mCCA objective functions in 1971 along with their solutions under specific constraints and their underlying factor models. Almost 25 years later, Nielsen [24] incorporated additional constraints to these five objective functions, resulting in new methods for their solutions. Asendorf completes the summary [25] in 2015, by filling the gaps in the theoretical solutions presented by [24] and also describing how to practically find the solutions of the different mCCA methods for observed data with specific software. Tenenhaus provides alternative solutions to some of the mCCA objective functions by considering them as special cases of SRGCCA (Regularized GCCA) [26], [27], which combines multi-block analysis methods like mCCA with Partial Least Squares (PLS) path modeling. However, these references and other follow-on work fall short of providing a complete picture of this widely used set of mCCA methods. Key gaps remain in establishing their connections, providing clear implementation guidance through pseudocodes, and, more importantly, specifying the conditions for identifying underlying sources using the different mCCA methods under the assumption of the JBSS generative model. Knowing these conditions is a key prerequisite for interpretability and explainability because without having guarantees to identify the true underlying sources, it would not make sense to interpret them, i.e., to attach physical meaning to them [2].

The conditions under which methods like mCCA are able to achieve JBSS are more generally known as *source identification conditions*. They are derived from the objective function of a given method and can be used to determine what type of sources a method can separate. There are references attempting to derive these conditions, but they are limited in scope. The strongest possible source identification conditions for correlation-based JBSS methods (including mCCA) are outlined in [28], but they are not derived for specific methods such as mCCA. Reference [20] provides source identification conditions for two mCCA methods, but one of these proofs considers only a modified mCCA objective function (sum of the absolute value of entries rather than directly the sum of entries in a covariance matrix). Source identification conditions of mCCA under a different model than the more general JBSS model are derived in [29], and conditions to identify the common subspace (instead of the individual source components) are derived from an algebraic perspective in [5]. We aim to fill these gaps in the literature with the following contributions:

- We provide a concise overview of the five mCCA optimization problems, `sumcor`, `maxvar`, `minvar`, `genvar`, and `ssqcor`, along with their solutions, underlying factor models, and pseudocodes.
- We derive new insights that connect the two popular mCCA methods `sumcor` and `maxvar`.
- We propose new theorems, conjectures, and considerations for the *source identification conditions* of the mCCA methods.
- We substantiate the proposed theoretical conditions with simulations and test the *statistical efficiency* of the five mCCA methods for finite samples.

With these contributions, we help to clarify the differences between the mCCA methods, link mCCA to a generative model, which is the JBSS model, and, based on our findings, conclude that `genvar` is our preferred method for JBSS. This is a significant finding, as `maxvar` and `sumcor` are the most commonly used mCCA methods [5].

After introducing our notation in Section II, we review the optimization problems of mCCA and JBSS in Section III. We present the mCCA methods that solve these optimization problems in Section IV and derive the connection between `sumcor` and two versions of `maxvar` in Section V. We then propose the source identification conditions of the mCCA methods in Section VI, and we demonstrate these theorems and conjectures in simulations in Section VII. Finally, we summarize our findings in Section VIII.

II. Notation

We denote scalars by italicized small letters, a, vectors by bold small letters, $\mathbf{a}$, and matrices by capital bold letters, $\mathbf{A}$. The superscript $(\cdot)^{\mathsf{T}}$ denotes the transpose of a matrix, $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with the corresponding vector on the main diagonal, and $\mathrm{blkdiag}(\cdot)$ denotes a block-diagonal matrix with the corresponding matrices on the main diagonal blocks. The matrix $\mathbf{I}_{N\times N}$ denotes the $(N\times N)$-identity matrix. The covariance matrix of a random vector $\mathbf{a}_n$ is denoted as $\mathbf{C}_{\mathbf{a}_n} = \mathbb{E}\{\mathbf{a}_n\mathbf{a}_n^{\mathsf{T}}\}$, and the covariance matrix of a random vector $\mathbf{a}^{[k]}$ is denoted as $\mathbf{C}_{\mathbf{a}}^{[k]}$ for better readability (instead of $\mathbf{C}_{\mathbf{a}^{[k]}}$). We denote the big-$\mathcal{O}$ complexity of an algorithm by $\mathcal{O}(\cdot)$. Table I summarizes the most commonly used notation.

III. Background

In this section, we introduce mCCA and JBSS. We first present the optimization problems of `sumcor`, `maxvar`, `minvar`, `genvar`, and `ssqcor` and introduce the JBSS generative model. Then, we link mCCA to the JBSS model by formally introducing the terminology of source identification conditions of mCCA.

A. Multiset Canonical Correlation Analysis

Let $\mathbf{x}^{[k]} \in \mathbb{R}^N$, $k = 1, \ldots, K$, denote K datasets, represented as random vectors. Without loss of generality (w.l.o.g.), we assume that they are zero mean, i.e., $\mathbb{E}\left\{\mathbf{x}^{[k]}\right\} = \mathbf{0}_N$. MCCA aims to estimate canonical variables [4]

$$u_n^{[k]} = \left(\mathbf{t}_n^{[k]}\right)^{\mathsf{T}}\mathbf{x}^{[k]}, \quad n = 1, \ldots, N, \quad k = 1, \ldots, K,$$

such that the n^{th} canonical variables $u_n^{[1]}, \ldots, u_n^{[K]}$ are maximally correlated for each $n = 1, \ldots, N$. Here, $u_n^{[k]}$ denotes the

Table I
LIST OF SYMBOLS

Symbol	Meaning	Dimension
$\mathbf{x}^{[k]}$ / $\mathbf{X}^{[k]}$	k^{th} dataset	N / $N \times V$
$\mathbf{C}_{\mathbf{x}}^{[k]}$	covariance matrix of the k^{th} dataset	$N \times N$
$\mathbf{C}_{\mathbf{x}}$	covariance matrix of concatenated $\mathbf{x}^{[1]}, \ldots, \mathbf{x}^{[K]}$	$NK \times NK$
$\mathbf{D}_{\mathbf{x}}$	block-diagonal matrix of $\mathbf{C}_{\mathbf{x}}^{[1]}, \ldots, \mathbf{C}_{\mathbf{x}}^{[K]}$	$NK \times NK$
$\mathbf{y}^{[k]}$ / $\mathbf{Y}^{[k]}$	k^{th} whitened dataset	N / $N \times V$
$\mathbf{C}_{\mathbf{y}}$	covariance matrix of concatenated $\mathbf{y}^{[1]}, \ldots, \mathbf{y}^{[K]}$	$NK \times NK$
$u_n^{[k]}$	n^{th} canonical variable of the k^{th} dataset	1
$\mathbf{u}^{[k]}$ / $\mathbf{U}^{[k]}$	k^{th} canonical vector	N / $N \times V$
$\mathbf{u}_n$ / $\mathbf{U}_n$	n^{th} canonical component vector (CCV)	K / $K \times V$
$\mathbf{C}_{\mathbf{u}_n}$	covariance matrix of the n^{th} CCV $\mathbf{u}_n$	$K \times K$
$\boldsymbol{\ell}_n$	eigenvalues of $\mathbf{C}_{\mathbf{u}_n}$	K
$\left(\mathbf{T}^{[k]}\right)^{\mathsf{T}}$	transformation matrix for the k^{th} dataset	$N \times N$
$\mathbf{A}^{[k]}$	mixing matrix of the k^{th} dataset	$N \times N$
$\mathbf{s}^{[k]}$ / $\mathbf{S}^{[k]}$	k^{th} source vector	N / $N \times V$
$\mathbf{C}_{ss}$	covariance matrix of concatenated $\mathbf{s}^{[1]}, \ldots, \mathbf{s}^{[K]}$	$NK \times NK$
$\widehat{\mathbf{s}}^{[k]}$ / $\widehat{\mathbf{S}}^{[k]}$	estimate of $\mathbf{s}^{[k]}$ / $\mathbf{S}^{[k]}$	N / $N \times V$
$\mathbf{s}_n$ / $\mathbf{S}_n$	n^{th} source component vector (SCV)	K / $K \times V$
$\mathbf{C}_{\mathbf{s}_n}$	covariance matrix of the n^{th} SCV $\mathbf{s}_n$	$K \times K$
$\boldsymbol{\lambda}_n$	eigenvalues of $\mathbf{C}_{\mathbf{s}_n}$	K
$\mathbf{W}^{[k]}$	demixing matrix for the k^{th} dataset	$N \times N$
$\mathbf{P}$	common permutation matrix	$N \times N$
$\boldsymbol{\Gamma}^{[k]}$	diagonal scaling matrix for the k^{th} dataset	$N \times N$
$\widehat{\boldsymbol{\Gamma}}^{[k]}$	diagonal sign matrix for the k^{th} dataset, with diagonal elements equal to ± 1	$N \times N$

n^{th} canonical variable in the k^{th} dataset, and $\mathbf{t}_n^{[k]} \in \mathbb{R}^N$ is the corresponding transformation vector. Maximizing correlation across multiple random variables requires a measure of this correlation. Such a measure is not uniquely defined; however, there are different ways of empirically measuring the correlation of more than two random variables. One intuitive measure is the sum of the pairwise covariances $\mathbb{E}\left\{u_n^{[k]} u_n^{[l]}\right\}$ among the n^{th} canonical variables from all K datasets, as covariances and correlations coincide due the canonical variables being zero-mean, i.e., $\mathbb{E}\left\{u_n^{[k]}\right\} = 0$. This leads to the following optimization problem to find the first canonical variable of all datasets, $u_1^{[k]}$, $k = 1, \ldots, K$:

$$\max_{\mathbf{t}_1^{[1]}, \ldots, \mathbf{t}_1^{[K]}} \sum_{k=1}^{K} \sum_{l=1}^{K} \mathbb{E}\left\{u_1^{[k]} u_1^{[l]}\right\}. \tag{1}$$

Later we discuss an additional constraint needed on the $u_n^{[k]}$ that is necessary to avoid maximizing the objective function simply by scaling the $\mathbf{t}_n^{[k]}$, which we call the "trivial solution". To simplify our notation in the following, let all K canonical variables with index n be stacked in the n^{th} Canonical Component Vector (CCV)

$$\mathbf{u}_n = \left[u_n^{[1]}, \ldots, u_n^{[K]}\right]^{\mathsf{T}} \in \mathbb{R}^K, \quad n = 1, \ldots, N,$$

as visualized in Figure 1. Now, for the next CCV, the sum of correlations between the elements in $\mathbf{u}_2$ is maximized. Also here, an additional constraint is needed in the optimization problem, as otherwise the same CCV would be estimated as before. We will later introduce possible options for such a constraint. Finally, after estimating the N CCVs, the N canonical variables of the k^{th} dataset form the canonical vector

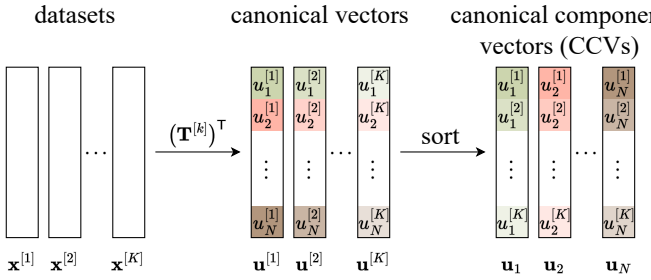


Figure 1. Schematic idea of mCCA. The datasets $\mathbf{x}^{[k]}$ are multiplied by $\left(\mathbf{T}^{[k]}\right)^{\mathsf{T}}$ to get the canonical vectors $\mathbf{u}^{[k]} = \left(\mathbf{T}^{[k]}\right)^{\mathsf{T}} \mathbf{x}^{[k]}$, $k = 1, \ldots, K$. The n^{th} elements of all canonical vectors $\mathbf{u}^{[k]}$ form the n^{th} CCV $\mathbf{u}_n$, $n = 1, \ldots, N$. mCCA finds $\left(\mathbf{T}^{[k]}\right)^{\mathsf{T}}$ such that the elements within each $\mathbf{u}_n$ are maximally correlated, which is visualized by the same color in different transparencies.

$\mathbf{u}^{[k]} = \left[u_1^{[k]}, \ldots, u_N^{[k]}\right]^{\mathsf{T}} \in \mathbb{R}^N$, which is found as

$$\mathbf{u}^{[k]} = \left(\mathbf{T}^{[k]}\right)^{\mathsf{T}} \mathbf{x}^{[k]}, \quad k = 1, \ldots, K,$$

with transformation matrix $\mathbf{T}^{[k]} = \left[\mathbf{t}_1^{[k]}, \ldots, \mathbf{t}_N^{[k]}\right] \in \mathbb{R}^{N \times N}$.

As noted before, there is more than one way to measure correlation across multiple random variables, leading to several mCCA optimization problems. Let $\mathbf{C}_{\mathbf{u}_n} = \mathbb{E}\left\{\mathbf{u}_n \mathbf{u}_n^{\mathsf{T}}\right\} \in \mathbb{R}^{K \times K}$, $n = 1, \ldots, N$, be the covariance matrix of the n^{th} CCV, which consists of the pairwise covariances $\mathbb{E}\left\{u_n^{[k]} u_n^{[l]}\right\}$, $k, l = 1, \ldots, K$, and let $\boldsymbol{\ell}_n = \left[\ell_n^{[1]} \ldots \ell_n^{[K]}\right]^{\mathsf{T}}$ be the eigenvalues of $\mathbf{C}_{\mathbf{u}_n}$. W.l.o.g., we assume for the remainder of this paper that eigenvalues are sorted in descending order, i.e., $\ell_n^{[1]} \geq \cdots \geq \ell_n^{[K]}$. Kettering [4] summarizes five mCCA objective functions, which all reduce to the CCA objective function for $K = 2$ datasets, as a function of $\mathbf{C}_{\mathbf{u}_n}$ for the n^{th} CCV. If all canonical variables have unit variance and are uncorrelated, then $\mathbf{C}_{\mathbf{u}_n} = \mathbf{I}_{K \times K}$. Ultimately, the goal of all of the following mCCA objective functions boils down to making $\mathbf{C}_{\mathbf{u}_n}$ as far away as possible from $\mathbf{I}_{K \times K}$, i.e., increasing correlation within $\mathbf{C}_{\mathbf{u}_n}$.

- mCCA-`sumcor` [30]:
  Maximize the sum of covariances,
  $$\mathcal{J}_{\texttt{sumcor}}(\mathbf{C}_{\mathbf{u}_n}) = \sum_{k=1}^{K} \sum_{l=1}^{K} c_{u_n}^{[k,l]}, \tag{F1}$$
  where $c_{u_n}^{[k,l]} = \mathbb{E}\left\{u_n^{[k]} u_n^{[l]}\right\}$ is the $(k,l)^{\text{th}}$ element of $\mathbf{C}_{\mathbf{u}_n}$. This is the expression in (1) written in terms of $\mathbf{C}_{\mathbf{u}_n}$ and is the natural extension of CCA to more than two datasets [25].
- mCCA-`ssqcor` [4]:
  Maximize the sum of squared covariances,
  $$\mathcal{J}_{\texttt{ssqcor}}(\mathbf{C}_{\mathbf{u}_n}) = \sum_{k=1}^{K} \sum_{l=1}^{K} \left(c_{u_n}^{[k,l]}\right)^2 = \sum_{k=1}^{K} \left(\ell_n^{[k]}\right)^2. \tag{F2}$$
  This removes the effect of the sign of the covariances compared with `sumcor`.
- mCCA-`maxvar` [30]:
  Maximize the largest eigenvalue of $\mathbf{C}_{\mathbf{u}_n}$,
  $$\mathcal{J}_{\texttt{maxvar}}(\mathbf{C}_{\mathbf{u}_n}) = \ell_n^{[1]}, \tag{F3}$$
  where $\ell_n^{[1]}$ is the largest eigenvalue of $\mathbf{C}_{\mathbf{u}_n}$. This $\mathbf{C}_{\mathbf{u}_n}$ is the best approximation to a rank-1 matrix, i.e., a matrix in which the covariances between all variables in the CCV $\mathbf{u}_n$ are as close as possible to 1.

- mCCA-`minvar` [4]:
  Minimize the smallest eigenvalue of $\mathbf{C}_{\mathbf{u}_n}$,
  $$\mathcal{J}_{\texttt{minvar}}(\mathbf{C}_{\mathbf{u}_n}) = \ell_n^{[K]}, \tag{F4}$$
  where $\ell_n^{[K]}$ is the smallest eigenvalue of $\mathbf{C}_{\mathbf{u}_n}$. This idea can be viewed as the reverse of `maxvar`.
- mCCA-`genvar` [31]:
  Minimize the determinant of $\mathbf{C}_{\mathbf{u}_n}$,
  $$\mathcal{J}_{\texttt{genvar}}(\mathbf{C}_{\mathbf{u}_n}) = \det(\mathbf{C}_{\mathbf{u}_n}) = \prod_{k=1}^{K} \ell_n^{[k]}. \tag{F5}$$
  The determinant of a covariance matrix is referred to as its generalized variance [32], thus the name.

As mentioned before, constraints must be added to all the presented optimization problems to avoid the trivial solution. In this paper, we consider the following two constraints on the canonical variables within the n^{th} CCV $\mathbf{u}_n$:

- $\mathbb{E}\left\{\left(u_n^{[k]}\right)^2\right\} = \left(\mathbf{t}_n^{[k]}\right)^{\mathsf{T}} \mathbf{C}_{\mathbf{x}}^{[k]} \mathbf{t}_n^{[k]} = 1,\ k = 1, \ldots, K,$ (C1)
- $\sum_{k=1}^{K} \mathbb{E}\left\{\left(u_n^{[k]}\right)^2\right\} = \sum_{k=1}^{K} \left(\mathbf{t}_n^{[k]}\right)^{\mathsf{T}} \mathbf{C}_{\mathbf{x}}^{[k]} \mathbf{t}_n^{[k]} = 1,$ (C2)

where $\mathbf{C}_{\mathbf{x}}^{[k]} = \mathbb{E}\left\{\mathbf{x}^{[k]}\left(\mathbf{x}^{[k]}\right)^{\mathsf{T}}\right\} \in \mathbb{R}^{N \times N}$ is the covariance matrix of the k^{th} dataset. The first constraint (C1) means that each canonical variable has unit variance. Constraint (C1) is used by Kettenring [4] and corresponds to the third constraint in Nielsen [24]. The second constraint (C2) means that the sum over the variances of all K canonical variables within a CCV is 1 and corresponds to the fourth constraint in [24].

Furthermore, we consider the following two constraints on the canonical variables across CCVs $\mathbf{u}_n$ and $\mathbf{u}_m$, $n \neq m$:

- $\mathbb{E}\left\{u_n^{[k]} u_m^{[k]}\right\} = 0,\ n, m = 1, \ldots, N,\ k = 1, \ldots, K,$ (H1)
- $\sum_{k=1}^{K} \mathbb{E}\left\{u_n^{[k]} u_m^{[k]}\right\} = 0,\ n, m = 1, \ldots, N.$ (H2)

Constraint (H1) from [4, eq. (9.5)] denotes that canonical variables within the same dataset are uncorrelated. Constraint (H2) from [24, eq. (3.63)] is a weaker constraint, which does not constrain the canonical variables to be uncorrelated but only constrains the sum of the covariances to be zero.

To study the source identification conditions of mCCA, we introduce the JBSS generative model and connect the canonical variables of mCCA to the source components of JBSS in the following subsections.

B. Joint Blind Source Separation

Let us assume that the datasets $\mathbf{x}^{[k]}$ are generated by the model [20]

$$\mathbf{x}^{[k]} = \mathbf{A}^{[k]} \mathbf{s}^{[k]}, \quad k = 1, \ldots, K, \tag{2}$$

where $\mathbf{A}^{[k]} \in \mathbb{R}^{N \times N}$ is an unknown invertible mixing matrix, and $\mathbf{s}^{[k]} = \left[s_1^{[k]}, \ldots, s_N^{[k]}\right]^{\mathsf{T}} \in \mathbb{R}^N$ is an unknown zero-mean source vector containing the source components $s_n^{[k]}$. The goal of JBSS is to estimate the source vectors [20]

$$\widehat{\mathbf{s}}^{[k]} = \mathbf{W}^{[k]} \mathbf{x}^{[k]} \quad k = 1, \ldots, K.$$

If JBSS is successfully achieved, i.e., if the estimated source vectors $\widehat{\mathbf{s}}^{[k]}$ recover the true source vectors $\mathbf{s}^{[k]}$ up to unavoidable permutation and scale ambiguities [33]: $\widehat{\mathbf{s}}^{[k]} = \mathbf{P}\mathbf{\Gamma}^{[k]}\mathbf{s}^{[k]}$, where $\mathbf{P} \in \mathbb{R}^{N \times N}$ is an arbitrary permutation matrix common for all datasets and $\mathbf{\Gamma}^{[k]} \in \mathbb{R}^{N \times N}$ is a diagonal scaling matrix, then, the desired $\mathbf{W}^{[k]}$ is a scaled and permuted version of the inverse of the mixing matrix $\mathbf{A}^{[k]}$ in (2), i.e., $\mathbf{W}^{[k]} = \mathbf{P}\mathbf{\Gamma}^{[k]}\left(\mathbf{A}^{[k]}\right)^{-1}$. Since matrix factorizations are generally non-unique, constraints have to be added to be able to obtain a unique solution for $\mathbf{W}^{[k]}$ and $\widehat{\mathbf{s}}^{[k]}$. These constraints typically enforce specific assumptions. The usual assumptions in JBSS are that the source components $s_n^{[k]}$ are independent within a dataset and can be dependent across datasets [34], but as we are considering only second-order statistics, dependence is simply linear dependence, i.e. correlation. By defining the n^{th} Source Component Vector (SCV) analogously to the CCV notation as

$$\mathbf{s}_n = \left[s_n^{[1]}, \ldots, s_n^{[K]}\right]^{\mathsf{T}} \in \mathbb{R}^K, \quad n = 1, \ldots, N,$$

we can formulate the following assumptions on the source components [20].

- W.l.o.g., source components have zero mean, i.e.,
  $$\mathbb{E}\left\{s_n^{[k]}\right\} = 0, \quad n = 1, \ldots, N, \quad k = 1, \ldots, K. \tag{A1}$$
- W.l.o.g., source components have unit variance, i.e.,
  $$\mathbb{E}\left\{\left(s_n^{[k]}\right)^2\right\} = 1, \quad n = 1, \ldots, N, \quad k = 1, \ldots, K. \tag{A2}$$
  Then, if unit-variance source components are estimated, the scale ambiguity reduces to a sign ambiguity, i.e., $\mathbf{W}^{[k]} = \mathbf{P}\widetilde{\mathbf{\Gamma}}^{[k]}\left(\mathbf{A}^{[k]}\right)^{-1}$, where $\widetilde{\mathbf{\Gamma}}^{[k]} \in \mathbb{R}^{N \times N}$ is a diagonal matrix with diagonal elements equal to either -1 or 1.
- Source components are uncorrelated across SCVs (but can be correlated within an SCV), i.e.,
  $$\mathbb{E}\left\{\mathbf{s}_n \mathbf{s}_m^{\mathsf{T}}\right\} = \mathbf{0}_{K \times K}, \quad n \neq m, \quad n, m = 1, \ldots, N. \tag{A3}$$
  This assumption is not w.l.o.g. and may not necessarily hold for any given dataset. However, it is a common and practical assumption for deriving solutions and studying their source identification conditions [20], [35].
- To simplify our formulations in Section VI, we further assume that all source components within an SCV are sufficiently correlated such that its covariance matrix
  $$\mathbf{C}_{\mathbf{s}_n} = \mathbb{E}\left\{\mathbf{s}_n \mathbf{s}_n^{\mathsf{T}}\right\} \in \mathbb{R}^{K \times K}, \quad n = 1, \ldots, N,$$
  cannot be partitioned into a (permuted) block-diagonal matrix. (A4)
  If Assumption (A4) does not hold, the conditions in Section VI can be extended to account for block-diagonal covariance matrices using so-called α-Gaussian SCVs, similar to the development in [34]. However, for block-diagonal covariance matrices, it may not be possible to correctly align sources across datasets.

C. Source identification conditions of mCCA for JBSS

We can see that these two problems are connected—the goal of JBSS is to estimate source vectors $\widehat{\mathbf{s}}^{[k]}$ that are maximally uncorrelated within a dataset and maximally correlated across datasets, whereas mCCA estimates canonical variables $\mathbf{u}^{[k]}$ that are maximally correlated across datasets (and typically uncorrelated within datasets). However, by its formulation, mCCA does not assume a generative model. Thus, to investigate its ability to achieve JBSS, we assume that K datasets $\mathbf{x}^{[k]} \in \mathbb{R}^N$ are generated according to the JBSS model in (2), where the source components satisfy Assumptions (A1)–(A4). Then, we define the *source identification conditions* of mCCA

as the conditions on the true source components in $\mathbf{s}^{[k]} \in \mathbb{R}^N$ under which an mCCA method achieves JBSS, i.e., under which the canonical variables $\mathbf{u}_n^{[k]}$ recover the true source vectors $\mathbf{s}^{[k]}$ up to a scaling and a common permutation:

$$\mathbf{u}^{[k]} = \mathbf{P}\mathbf{\Gamma}^{[k]}\mathbf{s}^{[k]}, \quad k = 1, \ldots, K.$$

IV. mCCA methods

In this section, we present a clear and concise overview of the solutions and the underlying factor models of the five mCCA methods `sumcor`, `maxvar`, `minvar`, `genvar`, and `ssqcor`, along with pseudocodes to make the concepts clear and easy to follow. Our notation up to this point represents the datasets $\mathbf{x}^{[k]} \in \mathbb{R}^N$, SCVs $\mathbf{s}_n \in \mathbb{R}^K$, and so on, as random vectors, and we will use this notation also when we derive these solutions. However, in practice, we handle observed datasets $\mathbf{X}^{[k]} \in \mathbb{R}^{N \times V}$, where V is the number of samples. Then, SCVs are denoted as $\mathbf{S}_n \in \mathbb{R}^{K \times V}$, and covariance matrices are replaced by their estimates, e.g., $\widehat{\mathbf{C}}_{\mathbf{x}}^{[k]} = \frac{1}{V}\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}$, $\widehat{\mathbf{C}}_{\mathbf{s}_n} = \frac{1}{V}\mathbf{S}_n\mathbf{S}_n^{\mathsf{T}}$, and so on. When providing pseudocodes for the methods, we will use the sample notation as they handle observed data.

A. mCCA-`sumcor`

For the n^{th} CCV, `sumcor` aims to find the transformation vectors $\mathbf{t}_n^{[k]}$ that maximize the objective function given in (F1), i.e., the sum of the elements of the n^{th} CCV's covariance matrix $\mathbf{C}_{\mathbf{u}_n}$ [4]. Kettenring proposed a *deflationary* numerical solution for `sumcor` under (C1) and (H1) [4], i.e., a solution where the $\mathbf{t}_n^{[1]}, \ldots, \mathbf{t}_n^{[K]}$ are estimated sequentially for $n = 1, \ldots, N$, using a numerical method for each CCV $\mathbf{u}_n$. Under (C2) and (H2), however, `sumcor` also enjoys an *all-at-once* analytical solution, i.e., a solution where all CCVs are found simultaneously using an analytical method. This method is considerably more straightforward and more efficient regarding runtime and JBSS performance than the deflationary solution. For the remainder of the paper, when discussing `sumcor`, we focus on this all-at-once analytical solution, which we present in the following.

All-at-once analytical solution. Nielsen [24] derived the following all-at-once analytical solution for (F1) under (C2) and (H2). First, the datasets are concatenated as $\mathbf{x} = \big[\big(\mathbf{x}^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{x}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$, and the covariance matrix of the concatenated datasets is $\mathbf{C}_{\mathbf{x}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathsf{T}}\} \in \mathbb{R}^{NK \times NK}$. Defining $\mathbf{D}_{\mathbf{x}} \in \mathbb{R}^{NK \times NK}$ as the block-diagonal matrix consisting of the covariance matrices of the datasets, $\mathbf{D}_{\mathbf{x}} = \operatorname{blkdiag}\big(\mathbf{C}_{\mathbf{x}}^{[1]}, \ldots, \mathbf{C}_{\mathbf{x}}^{[K]}\big)$, the transformation vectors $\mathbf{t}_n^{[k]}$ can be found by solving the Generalized EVD (GEVD) problem [24]

$$\mathbf{C}_{\mathbf{x}}\mathbf{T} = \mathbf{D}_{\mathbf{x}}\mathbf{T}\mathbf{\Phi}, \tag{3}$$

where $\mathbf{\Phi} = \operatorname{diag}\big(\phi_1, \ldots, \phi_{NK}\big) \in \mathbb{R}^{NK \times NK}$ is a diagonal matrix consisting of the eigenvalues of $\mathbf{C}_{\mathbf{x}}$ with respect to (w.r.t.) $\mathbf{D}_{\mathbf{x}}$, and $\mathbf{T} = \big[\mathbf{t}_1, \ldots, \mathbf{t}_{NK}\big] \in \mathbb{R}^{NK \times NK}$ is an orthogonal matrix with the corresponding eigenvectors as columns. Let $\widetilde{\mathbf{T}} = \big[\mathbf{t}_1, \ldots, \mathbf{t}_N\big] \in \mathbb{R}^{NK \times N}$ be the matrix consisting of the N principal eigenvectors (corresponding to the N largest eigenvalues). The transformation vectors $\mathbf{t}_n^{[k]} \in \mathbb{R}^N$, $k = 1, \ldots, K$, for the n^{th} CCV are found by partitioning $\mathbf{t}_n$ as $\mathbf{t}_n = \big[\big(\mathbf{t}_n^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{t}_n^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$, and

$$\widetilde{\mathbf{T}} = \big[\big(\mathbf{T}^{[1]}\big)^{\mathsf{T}}, \ldots \big(\mathbf{T}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK \times N} \tag{4}$$

contains the transformation matrices $\mathbf{T}^{[k]}$ for the k^{th} dataset. Then, the canonical vectors are found as $\mathbf{u}^{[k]} = \big(\mathbf{T}^{[k]}\big)^{\mathsf{T}}\mathbf{x}^{[k]} \in \mathbb{R}^N$, $k = 1, \ldots, K$, and the n^{th} CCV $\mathbf{u}_n \in \mathbb{R}^K$ consists of the n^{th} canonical variables for all K datasets, as visualized in Figure 1. In the end, we normalize the canonical variables $u_n^{[k]}$ to unit-variance and multiply $\mathbf{t}_n^{[k]}$ by the variance of $u_n^{[k]}$, $n = 1, \ldots, N$, $k = 1, \ldots, K$, for comparison with the other methods. The pseudocode for the all-at-once analytical `sumcor` method is given in Algorithm 1 in the supplementary material. The big-$\mathcal{O}$ complexity of this method is $\mathcal{O}(N^2K^2V)$, where we have assumed that $V > NK$. It is dominated by line 2.

Underlying factor model. The CCVs estimated by `sumcor` underlie the model [4]

$$\mathbf{u}_n = \mathbf{1}_K f_n + \boldsymbol{\nu}_n, \quad n = 1, \ldots, N, \tag{5}$$

where f_n is the best fitting zero-mean unit-variance common factor present in all $u_n^{[k]}$, $k = 1, \ldots, K$, $\mathbf{1}_K \in \mathbb{R}^K$ is a vector full of ones, and $\boldsymbol{\nu}_n \in \mathbb{R}^K$ is a noise vector. In the JBSS setting, this means that `sumcor` assumes an *effective rank* of one for the SCVs, i.e., all source components $s_n^{[k]}$ within the n^{th} SCV have the same underlying factor f_n [36].

B. mCCA-`maxvar`

For the n^{th} CCV, `maxvar` aims to find the transformation vectors $\mathbf{t}_n^{[k]}$ that maximize the largest eigenvalue of its covariance matrix $\mathbf{C}_{\mathbf{u}_n}$ as denoted in (F3) [4].

Deflationary analytical solution. Kettenring proposes in [4, section 10] the following deflationary solution for `maxvar` using constraints (C1) and (H1), where each CCV $\mathbf{u}_n$ is found analytically as follows. First, each dataset $\mathbf{x}^{[k]} \in \mathbb{R}^N$ is whitened using Mahalanobis whitening [37] to obtain

$$\mathbf{y}^{[k]} = \big(\mathbf{C}_{\mathbf{x}}^{[k]}\big)^{-\frac{1}{2}}\mathbf{x}^{[k]} \in \mathbb{R}^N, \quad k = 1, \ldots, K. \tag{6}$$

The whitened datasets are concatenated into the vector $\mathbf{y} = \big[\big(\mathbf{y}^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{y}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$, and the covariance matrix of the whitened datasets is $\mathbf{C}_{\mathbf{y}} = \mathbb{E}\{\mathbf{y}\mathbf{y}^{\mathsf{T}}\} \in \mathbb{R}^{NK \times NK}$. Now, the following Eigenvalue Decomposition (EVD) is performed sequentially for $n = 1, \ldots, N$:

$$\mathbf{H}_n\mathbf{C}_{\mathbf{y}}\mathbf{H}_n\mathbf{v}_n = \phi_n\mathbf{v}_n, \tag{7}$$

where ϕ_n denotes the largest eigenvalue of $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}\mathbf{H}_n \in \mathbb{R}^{NK \times NK}$, $\mathbf{v}_n \in \mathbb{R}^{NK}$ the corresponding eigenvector, and the matrix $\mathbf{H}_n$ is described in the following. For estimating the first CCV, $\mathbf{H}_1 = \mathbf{I}_{NK \times NK}$. For the next CCVs, $\mathbf{H}_n \in \mathbb{R}^{NK \times NK}$ is calculated using the eigenvectors from the previous $n-1$ EVDs as follows. First, the eigenvectors are partitioned as $\mathbf{v}_i = \big[\big(\mathbf{v}_i^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{v}_i^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}}$, and then the normalized eigenvectors of the k^{th} dataset, $\widetilde{\mathbf{v}}_i^{[k]} = \frac{\mathbf{v}_i^{[k]}}{\big\|\mathbf{v}_i^{[k]}\big\|}$, $i = 1, \ldots, n-1$, are concatenated in $\widetilde{\mathbf{V}}_{(n-1)}^{[k]} = \big[\widetilde{\mathbf{v}}_1^{[k]}, \ldots, \widetilde{\mathbf{v}}_{n-1}^{[k]}\big] \in \mathbb{R}^{N \times (n-1)}$ to form the block-diagonal matrix $\widetilde{\mathbf{V}}_{(n-1)} = \operatorname{blkdiag}\Big(\widetilde{\mathbf{V}}_{(n-1)}^{[1]}, \ldots, \widetilde{\mathbf{V}}_{(n-1)}^{[K]}\Big) \in \mathbb{R}^{NK \times (n-1)K}$. Using this matrix, $\mathbf{H}_n$ is found as the projection matrix onto the subspace that is orthogonal to $\widetilde{\mathbf{V}}_{(n-1)}$:

$\mathbf{H}_n = \mathbf{I} - \widetilde{\mathbf{V}}_{(n-1)}\big(\widetilde{\mathbf{V}}_{(n-1)}^{\mathsf{T}}\widetilde{\mathbf{V}}_{(n-1)}\big)^{-1}\widetilde{\mathbf{V}}_{(n-1)}^{\mathsf{T}}$. This is repeated until the eigenvectors of all N EVDs are found. Finally, they are concatenated in $\widetilde{\mathbf{V}}^{[k]} = \big[\widetilde{\mathbf{v}}_1^{[k]}, \ldots, \widetilde{\mathbf{v}}_N^{[k]}\big] \in \mathbb{R}^{N\times N} \in \mathbb{R}^{N\times N}$ to find the transformation matrix for the k^{th} dataset

$$\mathbf{T}^{[k]} = \big(\mathbf{C}_{\mathbf{x}}^{[k]}\big)^{-\frac{1}{2}}\widetilde{\mathbf{V}}^{[k]}, \quad k = 1, \ldots, K. \tag{8}$$

The multiplication by $\big(\mathbf{C}_{\mathbf{x}}^{[k]}\big)^{-\frac{1}{2}}$ is necessary because $\widetilde{\mathbf{V}}^{[k]}$ is for the whitened datasets.

The pseudocode for the deflationary analytical `maxvar` method is given in Algorithm 2 in the supplementary material. The big-$\mathcal{O}$ complexity of this method is $\mathcal{O}(N^2K^2V + N^4K^3)$. It is dominated by lines 4 and 11.

All-at-once analytical solution. Contrasting this deflationary analytical approach, Carroll proposed an all-at-once analytical solution for `maxvar` [38]. Therefore, he introduces an auxiliary variable z_n and seeks $u_n^{[k]}$ of all datasets that are maximally correlated to z_n, $n = 1, \ldots, N$. Instead of imposing constraints (C1) or (C2), the constraint $\mathbb{E}\left\{z_n^2\right\} = 1$ is imposed in [38], and also the $z_1, \ldots, z_N$ are made uncorrelated instead of the canonical variables across CCVs, i.e., the constraint $\mathbb{E}\left\{z_n z_m\right\} = 0, n \neq m$, $n, m = 1, \ldots, N$, is imposed instead of (H1) or (H2). Carroll's solution directly considers observed sample datasets $\mathbf{X}^{[k]} \in \mathbb{R}^{N\times V}$, where N is the dimension of the data and V is the number of samples, instead of random variables. Then, he defines

$$\mathbf{Q} = \sum_{k=1}^{K} \alpha^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big(\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big)^{-1}\mathbf{X}^{[k]} \in \mathbb{R}^{V\times V}, \tag{9}$$

where $\alpha^{[k]}$ is a scalar weighting factor that can be different for each dataset. The matrix $\mathbf{Q}$ can be interpreted as the weighted sum of the K datasets' projection matrices, representing the row space of the datasets. Then, by performing an EVD on $\mathbf{Q}$, we have

$$\mathbf{Q}\mathbf{Z} = \mathbf{Z}\mathbf{\Phi}, \tag{10}$$

with $\mathbf{Z} \in \mathbb{R}^{V\times V}$ and $\mathbf{\Phi} \in \mathbb{R}^{V\times V}$. We denote $\widetilde{\mathbf{Z}} = \big[\mathbf{z}_1, \ldots, \mathbf{z}_N\big] \in \mathbb{R}^{V\times N}$, as the matrix consisting of the N principal eigenvectors of $\mathbf{Q}$. Then, the transformation matrix for the k^{th} dataset $\mathbf{T}^{[k]}$ can be found as [38]

$$\mathbf{T}^{[k]} = \Big(\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big)^{-1}\mathbf{X}^{[k]}\widetilde{\mathbf{Z}}. \tag{11}$$

In the end, we normalize the canonical variables, i.e., the rows of $\mathbf{U}^{[k]} = \big(\mathbf{T}^{[k]}\big)^{\mathsf{T}}\mathbf{X}^{[k]} \in \mathbb{R}^{N\times V}$, to unit-variance and multiply $\mathbf{t}_n^{[k]}$ by the variance of the n^{th} row of $\mathbf{U}^{[k]}$, $n = 1, \ldots, N$, $k = 1, \ldots, K$, for comparison with the other methods. The pseudocode for the all-at-once analytical `maxvar` method is given in Algorithm 3 in the supplementary material. The big-$\mathcal{O}$ complexity of this method is $\mathcal{O}(V^3)$. It is dominated by line 2 under the assumption that $V > NK$.

Underlying factor model. The CCVs estimated by `maxvar` underlie the model [4]

$$\mathbf{u}_n = \mathbf{m}_n f_n + \boldsymbol{\nu}_n, \quad n = 1, \ldots, N, \tag{12}$$

where f_n is the best fitting zero-mean unit-variance common factor, the elements of $\mathbf{m}_n \in \mathbb{R}^K$ weight the contribution of this factor to each $u_n^{[k]}$, and $\boldsymbol{\nu}_n \in \mathbb{R}^K$ is again a noise vector. Thus, also `maxvar` assumes the SCVs to have an effective rank of one in the JBSS setting [36], but compared with `sumcor`, here the common factor has a different contribution to each source component $s_n^{[k]}$.

C. mCCA-minvar

The goal of `minvar` is very similar to `maxvar`, with the difference that now for the n^{th} CCV, `minvar` aims to find transformation vectors $\mathbf{t}_n^{[k]}$ that minimize the smallest eigenvalue of each $\mathbf{C}_{\mathbf{u}_n}$ as denoted in (F4) [4].

Deflationary analytical solution. Kettenring [4] proposed, using constraints (C1) and (H1), the same deflationary analytical solution as for `maxvar`, i.e., performing the EVD $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}\mathbf{H}_n\mathbf{v}_n = \phi_n\mathbf{v}_n$ sequentially for $n = 1, \ldots, N$, but while in `maxvar` ϕ_n denoted the largest eigenvalue, here in `minvar` ϕ_n denotes the smallest non-zero eigenvalue, and $\mathbf{v}_n \in \mathbb{R}^{NK}$ is the corresponding eigenvector. The pseudocode for the deflationary analytical `minvar` method is given in Algorithm 4 in the supplementary material. The big-$\mathcal{O}$ complexity of this method is the same as that of the deflationary analytical `maxvar` method, i.e., $\mathcal{O}(N^2K^2V + N^4K^3)$.

Underlying factor model. The CCVs estimated by `minvar` underlie the model [4]

$$\mathbf{u}_n = \sum_{r=1}^{K-1} \mathbf{m}_n^{(r)} f_n^{(r)} + \boldsymbol{\nu}_n, \tag{13}$$

where $f_n^{(r)}$, $r = 1, \ldots, K-1$, are the $K-1$ zero-mean unit-variance uncorrelated common factors, the elements of $\mathbf{m}_n^{(r)} \in \mathbb{R}^K$ weight the contribution of each $f_n^{(r)}$ to each $u_n^{[k]}$, and $\boldsymbol{\nu}_n \in \mathbb{R}^K$ is again a noise vector. Whereas `maxvar` assumes a CCV to have an effective rank-1 model (one common factor), `minvar` assumes a CCV to have an effective rank-$(K-1)$ model.

D. mCCA-genvar

For the n^{th} CCV, `genvar` aims to find the transformation vectors $\mathbf{t}_n^{[k]}$ that minimize the determinant of its covariance matrix $\mathbf{C}_{\mathbf{u}_n}$ as denoted in (F5) [4].

Deflationary numerical solution. Kettenring proposed in [4] the following deflationary solution for `genvar` using constraints (C1) and (H1), where each CCV is found using the following numerical method. First, the datasets are whitened using Mahalanobis whitening as in (6), and the joint covariance matrix $\mathbf{C}_{\mathbf{y}}$ of the whitened datasets $\mathbf{y}^{[k]}$ is calculated as in Section IV-B. Then, the transformation vectors $\mathbf{v}_n^{[k]}$ for the whitened datasets are initialized as unit-norm vectors with all elements equal to $\frac{1}{\sqrt{N}}$. As for the deflationary analytical `maxvar` method in (7), N EVDs are performed sequentially for $n = 1, \ldots, N$, but now separately for each dataset, i.e.,

$$\mathbf{H}_n^{[k]}\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]}\mathbf{H}_n^{[k]}\mathbf{v}_n^{[k]} = \phi_n^{[k]}\mathbf{v}_n^{[k]}, \quad k = 1, \ldots, K, \tag{14}$$

where $\phi_n^{[k]}$ denotes the largest eigenvalue, $\mathbf{v}_n^{[k]} \in \mathbb{R}^N$ the corresponding eigenvector, and we explain in the following how to form the k^{th} datasets' projection matrices $\mathbf{H}_n^{[k]} \in \mathbb{R}^{N\times N}$ and the matrices $\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]} \in \mathbb{R}^{N\times N}$. For estimating the first CCV, $\mathbf{H}_1^{[k]} = \mathbf{I}_{N\times N}$. For the other CCVs,

$$\mathbf{H}_n^{[k]} = \mathbf{I}_{N\times N} - \mathbf{V}_{(n-1)}^{[k]}\big(\big(\mathbf{V}_{(n-1)}^{[k]}\big)^{\mathsf{T}}\mathbf{V}_{(n-1)}^{[k]}\big)^{-1}\big(\mathbf{V}_{(n-1)}^{[k]}\big)^{\mathsf{T}},$$

where $\mathbf{V}_{(n-1)}^{[k]} = \big[\mathbf{v}_1^{[k]}, \ldots, \mathbf{v}_{n-1}^{[k]}\big] \in \mathbb{R}^{N\times(n-1)}$ contains the concatenated transformation vectors for the k^{th} whitened dataset from the $n-1$ previous EVDs. While $\mathbf{H}_n^{[k]}$ is only dependent on the $\mathbf{v}_n^{[k]}$ of the previous stages, the matrix $\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]}$

is updated in every iteration. The matrix $\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]}$ is defined as

$$\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]} = \mathbf{N}_n^{[k]} \big(\mathbf{C}_{\mathbf{u}_n}^{[-k]}\big)^{-1} \big(\mathbf{N}_n^{[k]}\big)^{\mathsf{T}} \in \mathbb{R}^{N \times N}, \tag{15}$$

where $\mathbf{N}_n^{[k]} = \big[\mathbf{C}_{\mathbf{y}}^{[k,1]}\,\mathbf{v}_n^{[1]}, \ldots, \mathbf{C}_{\mathbf{y}}^{[k,k-1]}\,\mathbf{v}_n^{[k-1]}, \mathbf{C}_{\mathbf{y}}^{[k,k+1]}\,\mathbf{v}_n^{[k+1]}, \ldots, \mathbf{C}_{\mathbf{y}}^{[k,K]}\,\mathbf{v}_n^{[K]}\big] \in \mathbb{R}^{N \times (K-1)}$ consists of the covariances of the whitened dataset $\mathbf{y}^{[k]}$ with the canonical variables $\epsilon_n^{[l]}$, $l \neq k$, $l = 1, \ldots, K$, i.e., $\mathbf{C}_{\mathbf{y}}^{[k,l]}\mathbf{v}_n^{[l]} = \mathbb{E}\big\{\mathbf{y}^{[k]}\big(\mathbf{y}^{[l]}\big)^{\mathsf{T}}\mathbf{v}_n^{[l]}\big\} = \mathbb{E}\big\{\mathbf{y}^{[k]}\epsilon_n^{[l]}\big\} \in \mathbb{R}^N$, $l \neq k$, and $\mathbf{C}_{\mathbf{u}_n}^{[-k]} \in \mathbb{R}^{(K-1)\times(K-1)}$ is the matrix obtained by removing the k^{th} row and column of the n^{th} CCV's covariance matrix $\mathbf{C}_{\mathbf{u}_n} \in \mathbb{R}^{K \times K}$, which is calculated as

$$\mathbf{C}_{\mathbf{u}_n} = \text{blkdiag}\big(\mathbf{v}_n^{[1]}, \ldots, \mathbf{v}_n^{[K]}\big)^{\mathsf{T}} \mathbf{C}_{\mathbf{y}} \text{blkdiag}\big(\mathbf{v}_n^{[1]}, \ldots, \mathbf{v}_n^{[K]}\big).$$

Using the updated $\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]}$, the EVD in (14) is performed to update $\mathbf{v}_n^{[k]}$, and $\mathbf{C}_{\mathbf{u}_n}$, and then $\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]}$, $\mathbf{v}_n^{[k]}$ and $\mathbf{C}_{\mathbf{u}_n}$ are updated again in the next iteration until either convergence or a maximum number of iterations has been reached. Convergence is achieved when the sum of the differences between the largest eigenvalues of the current iteration i and the previous iteration $i-1$ is smaller than a user-defined threshold $\epsilon > 0$, i.e., when $\sum_{k=1}^{K} \big|\psi_n^{[k]}(i) - \psi_n^{[k]}(i-1)\big| < \epsilon$, where $\psi_n^{[k]}(i) = \phi_n^{[k]}(i)$ is the largest eigenvalue of the i^{th} iteration [4]. After the iterative process is completed, the transformation matrices $\mathbf{T}^{[k]}$ are found as

$$\mathbf{T}^{[k]} = \big(\mathbf{C}_{\mathbf{x}}^{[k]}\big)^{-\frac{1}{2}} \mathbf{V}^{[k]}.$$

The multiplication by $\big(\mathbf{C}_{\mathbf{x}}^{[k]}\big)^{-1/2}$ is again necessary because $\mathbf{V}^{[k]}$ is found for the whitened datasets. The pseudocode for the deflationary numerical `genvar` method is given in Algorithms 5 and 6 in the supplementary material. The big-$\mathcal{O}$ complexity of this method is $\mathcal{O}(N^2K^2V + IN^3K^4)$, where I is the number of iterations. It is dominated by line 5 in Algorithm 5 and line 3 in Algorithm 6.

Underlying factor model. The CCVs estimated by `genvar` underlie the model [4]

$$\mathbf{u}_n = \sum_{r=1}^{K} \mathbf{m}_n^{(r)} f_n^{(r)} + \boldsymbol{\nu}_n = \mathbf{M}_n \mathbf{f}_n + \boldsymbol{\nu}_n, \tag{16}$$

with $\mathbf{M}_n = \big[\mathbf{m}_n^{(1)}, \ldots, \mathbf{m}_n^{(K)}\big] \in \mathbb{R}^{K \times K}$, $\mathbf{f}_n = \big[f_n^{(1)}, \ldots, f_n^{(K)}\big]^{\mathsf{T}} \in \mathbb{R}^K$, and $\boldsymbol{\nu}_n \in \mathbb{R}^K$. The model (16) is similar to the `maxvar` / `minvar` models in (12) and (13) but now assumes an effective rank-K model with K different $\mathbf{m}_n^{(r)}$ and $\mathbf{f}_n^{(r)}$. It was noted in [4] that `genvar` can be seen as more "unpredictable" as the determinant cost is more sensitive to the smaller eigenvalues. While this may sound like a disadvantage, we later demonstrate that `genvar` is consistently among the best-performing mCCA methods for JBSS.

E. mCCA-`ssqcor`

For the n^{th} CCV, `ssqcor` aims to find the transformation vectors $\mathbf{t}_n^{[k]}$ that maximize the sum of squared correlations within each $\mathbf{C}_{\mathbf{u}_n}$ as denoted in (F2) [4].

Deflationary numerical solution. Kettenring proposed in [4], using constraints (C1) and (H1), a solution very similar to that of `genvar`, with the difference that $\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]}$ in (15) is now defined as

$$\widetilde{\mathbf{C}}_{\mathbf{u}_n}^{[k]} = \mathbf{N}_n^{[k]} \big(\mathbf{N}_n^{[k]}\big)^{\mathsf{T}}, \tag{17}$$

The pseudocode for the deflationary numerical `ssqcor` method is given in Algorithm 7 in the supplementary material. The big-$\mathcal{O}$ complexity of this method is $\mathcal{O}(N^2K^2V + IN^3K(K+N))$, where I is the number of iterations. It is dominated by lines 5 and 20 in Algorithm 5, line 1 in Algorithm 6, and modified line 5 in Algorithm 7.

Underlying factor model. The CCVs estimated by `ssqcor` also underlie the model in (16) [4]. By comparing (16) to (12), we observe that if only one single factor heavily influences the CCVs, i.e., all except for one $\mathbf{m}_n^{(l)}$ are nearly zero, `ssqcor` will estimate CCVs that are very similar to those of `maxvar`. However, in the JBSS setting, `ssqcor` is more flexible than `maxvar` in that `ssqcor` can estimate SCVs with *any* effective rank (from 1 to K).

F. Comment on different solutions

We have seen in the previous subsections that the mCCA optimization problems can be solved in different ways. The mCCA methods presented in [4] all require estimating the transformation vectors in a deflationary way, i.e., they sequentially estimate each CCV, where the solution for each CCV can be found either analytically (typically manifesting as EVD), as for `maxvar` and `minvar`, or numerically, e.g., using gradient descent, as for `sumcor`, `genvar`, and `ssqcor`. On the other hand, `maxvar` from [38] and `sumcor` from [24] estimate the transformation vectors all-at-once for all CCVs with an analytical solution.

V. Connections between different mCCA methods

This section is dedicated to comparing the two most closely related, and perhaps the most commonly used [5], mCCA methods: the `sumcor` method and the `maxvar` method. We prove that the canonical vectors estimated by the all-at-once analytical `maxvar` method [38] are scaled versions of those from the all-at-once analytical `sumcor` method [24], and thus, the canonical variables of both methods are the same if they are normalized to unit variance. This is important to understand because the latter is computationally much faster. Furthermore, we show that the canonical vectors from the deflationary analytical `maxvar` method [4] and those of the all-at-once analytical `sumcor` method [24] are closely related but different, with the difference that the canonical variables estimated by the latter are not constrained to be uncorrelated within a dataset.

A. Connection between all-at-once analytical `sumcor` [24] and all-at-once analytical `maxvar` [38]

Comparing the `sumcor` GEVD in (3) and the `maxvar` EVD in (10), we see that we can write both EVDs in terms of the whitened datasets. By doing so, we find the connection between the EVDs and, thereby, also the connection between the methods. As Carroll formulated the `maxvar` method [38] using samples, we also use sample notation for the `sumcor` method [24] to state their connection in Theorem 1.

Theorem 1. *Let the concatenated whitened datasets be* $\mathbf{Y} = \big[\big(\mathbf{Y}^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{Y}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK \times V}$, *where* $\mathbf{Y}^{[k]} = \big(\frac{1}{V}\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\big)^{-1/2}\mathbf{X}^{[k]} \in \mathbb{R}^{N \times V}$, $k = 1, \ldots, K$, *and let*

the weighting factor $\alpha^{[k]}$ in `maxvar` *[38] be equal to one. The canonical vectors estimated by the all-at-once analytical* `maxvar` *method [38] are scaled versions of those estimated by the all-at-once analytical* `sumcor` *method [24] (before normalization), i.e.,*

$$\mathbf{U}^{[k]}_{\text{maxvar}} = \frac{1}{V}\,\widetilde{\mathbf{\Omega}}\,\mathbf{U}^{[k]}_{\text{sumcor}}, \quad k = 1, \ldots, K,$$

or in CCV notation,

$$\mathbf{U}_{\text{maxvar},n} = \frac{\omega_n}{V}\,\mathbf{U}_{\text{sumcor},n}, \quad n = 1, \ldots, N,$$

where V is the number of samples, and $\widetilde{\mathbf{\Omega}} = \operatorname{diag}\big(\omega_1, \ldots, \omega_N\big) \in \mathbb{R}^{N\times N}$ is a diagonal matrix consisting of the N largest singular values of $\mathbf{Y}$.

The proof is given in Appendix A. Note that the canonical variables estimated by these two methods are the same when they are normalized to unit variance. However, if the weighting factor $\alpha^{[k]}$ is different for some datasets, the canonical variables estimated by both methods will no longer be the same up to a scaling.

As the all-at-once analytical `maxvar` method is computationally more expensive than the all-at-once analytical `sumcor` method (because the EVD is performed on a $V \times V$ matrix instead of an $N \times N$ matrix), we do not report the all-at-once analytical `maxvar` results but only the `sumcor` results in the remainder of this paper.

B. Connection between all-at-once analytical `sumcor` *[24] and deflationary analytical* `maxvar` *[4]*

We can write the `sumcor` GEVD in (3) in terms of the whitened datasets in a deflationary way and compare it with the deflationary `maxvar` EVD in (7). By doing so, we find the connection between the EVDs and, thereby, also the connection between the methods. As the `maxvar` method [4] and the `sumcor` method [24] are formulated using random vector notation, we also use this notation to state their connection in Theorem 2.

Theorem 2. *The transformation matrices, and therefore the canonical variables, of the all-at-once analytical* `sumcor` *method [24] and the deflationary analytical* `maxvar` *method [4] are different but closely related. More precisely, let $\mathbf{V}^{[k]} = \big[\mathbf{v}_1^{[k]}, \ldots, \mathbf{v}_N^{[k]}\big] \in \mathbb{R}^{N\times N}$ be the transformation matrix for the whitened datasets $\mathbf{y}^{[k]} = \big(\mathbf{C}_{\mathbf{x}}^{[k]}\big)^{-1/2}\mathbf{x}^{[k]} \in \mathbb{R}^N$ such that $\mathbf{u}^{[k]} = \big(\mathbf{V}^{[k]}\big)^{\mathsf{T}}\mathbf{y}^{[k]}$, $k = 1, \ldots, K$. Then, for* `maxvar` *[4], the transformation vectors $\mathbf{v}_n^{[k]}$ for all datasets are orthogonal, i.e., $\big(\mathbf{v}_n^{[k]}\big)^{\mathsf{T}}\mathbf{v}_m^{[k]} = 0$, $m \neq n$, $m = 1, \ldots, N$, $k = 1, \ldots, K$, while for* `sumcor` *[24], the concatenated transformation vectors $\mathbf{v}_n = \big[\big(\mathbf{v}_n^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{v}_n^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$ are orthogonal, i.e., $\mathbf{v}_n^{\mathsf{T}}\mathbf{v}_m = \sum_{k=1}^{K}\big(\mathbf{v}_n^{[k]}\big)^{\mathsf{T}}\mathbf{v}_m^{[k]} = 0$, $m \neq n$, $n, m = 1, \ldots, N$.*

The proof is given in Appendix B.

VI. Source identification conditions of mCCA

This section is dedicated to formulating the source identification conditions for JBSS of the different mCCA methods, i.e., the conditions on the true source components under which a method is able to achieve JBSS.

A. Source identification conditions of mCCA-`sumcor`

The `sumcor` transformation vectors are the partitions of the principal eigenvectors in the GEVD in (3). We can write this GEVD in terms of the concatenated SCVs' covariance matrix to find the relationship between the `sumcor` transformation vectors and the eigenvectors of the concatenated SCVs' covariance matrix. By doing so, we find the conditions that the eigenvalues of the SCV covariance matrices must satisfy such that the canonical vectors estimated by `sumcor` recover the true source vectors (up to permutation and scaling). These conditions are the source identification conditions of `sumcor`, which we propose in Theorem 3.

Theorem 3 (Source identification conditions of `sumcor`)**.** *We assume Assumptions* (A1)–(A4)*. Let $\boldsymbol{\lambda}_n = \big[\lambda_n^{[1]}, \ldots, \lambda_n^{[K]}\big]^{\mathsf{T}}$ be the eigenvalues of $\mathbf{C}_{\mathbf{s}_n} \in \mathbb{R}^{K\times K}$, the covariance matrix of the n^{th} SCV $\mathbf{s}_n \in \mathbb{R}^K$, sorted such that $\lambda_n^{[1]} \geq \cdots \geq \lambda_n^{[K]}$. If the following two conditions hold for this SCV $\mathbf{s}_n$:*

1) *the largest eigenvalue of $\mathbf{C}_{\mathbf{s}_n}$ is distinct from the largest eigenvalue of all other $\mathbf{C}_{\mathbf{s}_m}$, i.e.,*
$$\lambda_n^{[1]} \neq \lambda_m^{[1]}, \quad m \neq n, \quad m = 1, \ldots, N,$$
2) *and the largest eigenvalue of this $\mathbf{C}_{\mathbf{s}_n}$ is greater than the second-largest (and all following) eigenvalues of all $\mathbf{C}_{\mathbf{s}_m}$, i.e.,*
$$\lambda_n^{[1]} > \lambda_m^{[k]}, \quad k = 2, \ldots, K, \quad m = 1, \ldots, N,$$

then this SCV $\mathbf{s}_n$ can be identified by `sumcor` *[24]. If these conditions hold for all $n = 1, \ldots, N$, then* `sumcor` *achieves JBSS, i.e., the canonical vectors $\mathbf{u}^{[k]}$ recover the source vectors $\mathbf{s}^{[k]}$ up to a scaling and a common permutation:*

$$\mathbf{u}^{[k]} = \mathbf{P}\mathbf{\Gamma}^{[k]}\mathbf{s}^{[k]}, \quad k = 1, \ldots, K.$$

The proof is given in Appendix C. Note that the ambiguity here is up to a scaling and not only up to a sign, as the canonical variables estimated by `sumcor` [24] are not constrained to have unit variance. However, if they are post-normalized to have unit variance, which is what we do in this paper, they can recover the true sources up to a sign and a common permutation. The conditions in Theorem 3 also hold for the all-at-once analytical `maxvar` [38] since the estimated canonical variables are the same up to a scaling.

B. Source identification conditions of mCCA-`maxvar`

In Theorem 4, we present the source identification conditions of the deflationary analytical `maxvar` [4], which are derived and proven in [20].

Theorem 4 (Source identification conditions of `maxvar` from [20])**.** *We assume Assumptions* (A1)–(A4)*. Let $\boldsymbol{\lambda}_n = \big[\lambda_n^{[1]}, \ldots, \lambda_n^{[K]}\big]^{\mathsf{T}}$ be the eigenvalues of $\mathbf{C}_{\mathbf{s}_n} \in \mathbb{R}^{K\times K}$, the covariance matrix of the n^{th} SCV $\mathbf{s}_n \in \mathbb{R}^K$, sorted such that $\lambda_n^{[1]} \geq \cdots \geq \lambda_n^{[K]}$. If the following condition holds for this SCV $\mathbf{s}_n$:*
the largest eigenvalue of $\mathbf{C}_{\mathbf{s}_n}$ is distinct from the largest eigenvalue of all other $\mathbf{C}_{\mathbf{s}_m}$, i.e.,

$$\lambda_n^{[1]} \neq \lambda_m^{[1]}, \quad m \neq n, \quad m = 1, \ldots, N,$$

then this SCV $\mathbf{s}_n$ can be identified by `maxvar` *[4]. If this condition holds for all $n = 1, \ldots, N$, then* `maxvar` *achieves JBSS, i.e., the canonical vectors $\mathbf{u}^{[k]}$ recover the source vectors*

$\mathbf{s}^{[k]}$ *up to a sign and a common permutation:*

$$\mathbf{u}^{[k]} = \mathbf{P}\widetilde{\mathbf{\Gamma}}^{[k]}\mathbf{s}^{[k]}, \quad k = 1, \ldots, K.$$

These conditions are less restrictive than the source identification conditions of `sumcor`, which also depend on the second-largest eigenvalues.

C. Source identification conditions of mCCA-`minvar`

As `minvar` minimizes the smallest eigenvalue, while `maxvar` maximizes the largest eigenvalue, we conjecture that the source identification conditions of `minvar` are identical to those of `maxvar` but depend on the *smallest* eigenvalue instead. We present the conjectured source identification conditions of `minvar` in 1.

Conjecture 1 (Source identification conditions of `minvar`)**.** *We assume Assumptions* (A1)–(A4). *Let* $\boldsymbol{\lambda}_n = \left[\lambda_n^{[1]}, \ldots, \lambda_n^{[K]}\right]^\mathsf{T}$ *be the eigenvalues of* $\mathbf{C}_{\mathbf{s}_n} \in \mathbb{R}^{K\times K}$*, the covariance matrix of the* n^{th} *SCV* $\mathbf{s}_n \in \mathbb{R}^K$*, sorted such that* $\lambda_n^{[1]} \geq \cdots \geq \lambda_n^{[K]}$*. If the following condition holds for this SCV* $\mathbf{s}_n$*:*
the smallest eigenvalue of $\mathbf{C}_{\mathbf{s}_n}$ *is distinct from the smallest eigenvalue of all other* $\mathbf{C}_{\mathbf{s}_m}$*, i.e.,*

$$\lambda_n^{[K]} \neq \lambda_m^{[K]}, \quad m \neq n, \quad m = 1, \ldots, N,$$

then this SCV $\mathbf{s}_n$ *can be identified by* `minvar` *[4]. If this condition holds for all* $n = 1, \ldots, N$*, then* `minvar` *achieves JBSS, i.e., the canonical vectors* $\mathbf{u}^{[k]}$ *recover the source vectors* $\mathbf{s}^{[k]}$ *up to a sign and a common permutation:*

$$\mathbf{u}^{[k]} = \mathbf{P}\widetilde{\mathbf{\Gamma}}^{[k]}\mathbf{s}^{[k]}, \quad k = 1, \ldots, K.$$

Note that these conditions are unlikely to be satisfied in real-world datasets, as the smallest eigenvalues in practice are typically very similar to each other.

D. Source identification conditions of mCCA-`genvar`

We discuss our considerations for the source identification conditions of `genvar` using its connection to Independent Vector Analysis using a multivariate Gaussian distribution (IVA-G) [39], [40]. More precisely, `genvar` can be considered to be a deflationary version of IVA-G with orthogonality constraints on the demixing matrices [40]. Thus, it seems plausible to assume that the source identification conditions of `genvar` are similar to those of deflationary orthogonal IVA-G. The source identification conditions of deflationary orthogonal IVA-G are currently unknown and an interesting question for future work. As the source identification conditions of unconstrained IVA-G, given in [33], are defined w.r.t. the SCV covariance matrices and therefore w.r.t. both their eigenvalues and eigenvectors, we expect that the source identification conditions of `genvar` are also defined w.r.t. both the eigenvalues and eigenvectors of the SCV covariance matrices. If true, this would mean that `genvar` can achieve JBSS even for equal eigenvalues across SCVs as long as the eigenvectors are different, and thereby would have the least restrictive source identification conditions of all mCCA methods. We will show this in experiment B in Section VII.

E. Source identification conditions of mCCA-`ssqcor`

As the `genvar` and `ssqcor` methods are the only mCCA methods with objective functions defined over all eigenvalues of the covariance matrices $\mathbf{C}_{\mathbf{u}_n}$, we expect the source identification conditions of `ssqcor` to be similar to those of `genvar`, and therefore to also depend on the eigenvalues and eigenvectors of the SCV covariance matrices.

VII. Simulation study of source identification conditions

In this section, we investigate the methods' ability to achieve JBSS in different experiments. We start with an experiment that violates the conditions of all methods and continue with experiments that substantiate the conditions of specific mCCA methods. As the methods have different underlying models and therefore assume different *effective ranks* for the SCVs, we also perform an experiment where we change the effective rank of the SCVs. In all experiments, we model the n^{th} SCV's covariance matrix $\mathbf{C}_{\mathbf{s}_n} \in \mathbb{R}^{K\times K}$ as in [36]:

$$\mathbf{C}_{\mathbf{s}_n} = \alpha_n \mathbf{L}_n \mathbf{L}_n^\mathsf{T} + (1 - \alpha_n)\mathbf{I}_{K\times K}, \tag{18}$$

where the elements of $\mathbf{L}_n \in \mathbb{R}^{K\times R}$ are drawn from $\mathcal{N}(0, 1)$, each row of $\mathbf{L}_n$ is normalized to unit-norm so that the diagonal elements of $\mathbf{C}_{\mathbf{s}_n}$ are equal to one, and $\frac{-1}{K-1} < \alpha_n \leq 1$. As the matrix $\mathbf{L}_n\mathbf{L}_n^\mathsf{T}$ has rank R, the addition of the identity matrix is necessary to make $\mathbf{C}_{\mathbf{s}_n}$ have full rank, and α_n determines the proportion of the full-rank part of $\mathbf{C}_{\mathbf{s}_n}$. The model (18) corresponds to the SCV model [36]

$$\mathbf{s}_n = \mathbf{M}_n \mathbf{f}_n + \boldsymbol{\nu}_n$$

where the elements in $\mathbf{f}_n \in \mathbb{R}^R$ are the R common factors in the n^{th} SCV, $\mathbf{M}_n \in \mathbb{R}^{K\times R}$ weights the contribution of each factor $f_n^{[r]}$ to each source component $s_n^{[k]}$, and $\boldsymbol{\nu}_n \in \mathbb{R}^K$ is a noise vector with diagonal covariance matrix. This SCV model generalizes all the mCCA factor models in (5), (12), (13), and (16). For $R = 1$, the covariance matrix $\mathbf{C}_{\mathbf{s}_n}$ in (18) has $K - 1$ identical eigenvalues with the value $\lambda_n^{[k]} = 1 - \alpha_n$, $k = 2, \ldots, K$, and one different eigenvalue with the value $\lambda_n^{[1]} = K - (K - 1)(1 - \alpha_n)$. Thus, to make the different eigenvalue $\lambda_n^{[1]} > 0$ have a specific value, we can choose

$$\alpha_n = 1 - \frac{K - \lambda_n^{[1]}}{K - 1}. \tag{19}$$

We describe later how the values of R and α_n are chosen in the different experiments.[1]

In all these experiments, it is important to note that the ability of the mCCA methods to achieve JBSS is determined not only by whether the sources satisfy the *source identification conditions* of a method but also by the method's *statistical efficiency*, i.e., how efficient the method makes use of observed data (with finite samples) when estimating the sources. A more statistically efficient estimator needs fewer samples to adequately estimate the sources. Since much of the scope of this paper focuses on theoretical conditions, we need to decouple testing *source identification conditions* from testing *statistical efficiency* when possible.

To test only the source identification conditions, the methods use the true covariance matrices. Therefore, starting from the specified $\mathbf{C}_{\mathbf{s}_n}$, true joint covariance matrix of the SCVs is

[1]The Python implementation of the mCCA methods, as well as the code for the experiments, is available at: https://github.com/SSTGroup/multiset_canonical_correlation_analysis.

Table II
OVERVIEW OF THE FOUR EXPERIMENTS TO TEST THE SOURCE IDENTIFICATION CONDITIONS OF THE METHODS. THE SECOND ROW SHOWS WHAT HOLDS FOR THE EIGENVALUES AND EIGENVECTORS IN EACH EXPERIMENT. THEN, WE SHOW WHICH METHODS ARE EXPECTED TO ACHIEVE JBSS IN WHICH EXPERIMENTS. THE SYMBOL ✓ MEANS THAT A METHOD IS EXPECTED TO ACHIEVE JBSS, ✗ MEANS THAT THE METHOD IS EXPECTED TO FAIL, AND 🅞 MEANS THAT THE EXPERIMENT IS CLOSE TO (BUT NOT EXACTLY) VIOLATING THE CONDITIONS OF A METHOD AND ITS BEHAVIOR FOR FINITE SAMPLES IS UNCLEAR.

Experiment	A	B	C	D
	$\boldsymbol{\lambda}_n = \boldsymbol{\lambda}_m$ $\boldsymbol{\Theta}_n = \mathbf{D}_m\boldsymbol{\Theta}_m$	$\boldsymbol{\lambda}_n = \boldsymbol{\lambda}_m$ $\boldsymbol{\Theta}_n \neq \mathbf{D}_m\boldsymbol{\Theta}_m$	$\lambda_n^{(\min)} \neq \lambda_m^{(\min)}$	$\lambda_n^{(\max)} \neq \lambda_m^{(\max)}$
`sumcor`	✗	✗	✗	✓
`maxvar`	✗	✗	🅞	✓
`minvar`	✗	✗	✓	🅞
`genvar`	✗	✓	✓	✓
`ssqcor`	✗	✓	✓	✓

$\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})} = \mathrm{blkdiag}\big(\mathbf{C}_{\mathbf{s}_1}, \ldots, \mathbf{C}_{\mathbf{s}_N}\big) \in \mathbb{R}^{NK \times NK}$, and the covariance matrix of the concatenated source vectors $\mathbf{s} \in \mathbb{R}^{NK}$ is found as $\mathbf{C}_{\mathbf{s}} = \boldsymbol{\Pi}\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}\boldsymbol{\Pi}^{\mathsf{T}}$, where $\boldsymbol{\Pi} \in \mathbb{R}^{NK \times NK}$ is a specific permutation matrix to sort the elements of $\mathbf{C}_{\mathbf{s}}$ in source vector ordering instead of SCV ordering. Then, the joint mixing matrix $\mathbf{A} = \mathrm{blkdiag}\big(\mathbf{A}^{[1]}, \ldots, \mathbf{A}^{[K]}\big)$ is generated, where $\mathbf{A}^{[k]}$ is an invertible mixing matrix with elements drawn from $\mathcal{N}(0,1)$. Finally, the true covariance matrix of the concatenated datasets $\mathbf{x} \in \mathbb{R}^{NK}$ is $\mathbf{C}_{\mathbf{x}} = \mathbf{A}\mathbf{C}_{\mathbf{s}}\mathbf{A}^{\mathsf{T}}$, and the covariance matrix of the concatenated whitened datasets is $\mathbf{C}_{\mathbf{y}} = \mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}}\mathbf{C}_{\mathbf{x}}\mathbf{D}_{\mathbf{x}}^{-\frac{1}{2}}$, where $\mathbf{D}_{\mathbf{x}}$ is a block-diagonal matrix consisting of the diagonal blocks of $\mathbf{C}_{\mathbf{x}}$. The methods then use the true $\mathbf{C}_{\mathbf{x}}$ and $\mathbf{C}_{\mathbf{y}}$.

To test both source identification conditions and statistical efficiency, the methods estimate the covariance matrices from finite samples. Therefore, V samples are drawn for the SCVs $\mathbf{S}_n \in \mathbb{R}^{K \times V}$ from a multivariate Gaussian distribution with zero mean and the specified $\mathbf{C}_{\mathbf{s}_n}$ as covariance matrices, and the source components are normalized to have zero mean and unit variance. Then, the source components of the SCVs $\mathbf{S}_n$ are distributed to each $\mathbf{S}^{[k]}$, and the observed datasets $\mathbf{X}^{[k]} \in \mathbb{R}^{N \times V}$ are generated by mixing the source matrices as $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\mathbf{S}^{[k]}$, where $\mathbf{A}^{[k]} \in \mathbb{R}^{N \times N}$ is generated as described previously. Finally, the methods estimate the covariance matrices $\widehat{\mathbf{C}}_{\mathbf{x}}$ and $\widehat{\mathbf{C}}_{\mathbf{y}}$ using the observed datasets $\mathbf{X}^{[k]}$.

Table II summarizes what holds for the eigenvalues and eigenvectors in each experiment and which methods are expected to achieve JBSS. We simulate $N = 5$ SCVs and $K = 100$ datasets. In the following subsections, we describe in detail how the SCV covariance matrices $\mathbf{C}_{\mathbf{s}_n}$ are generated in experiments A–E.

A. Violating the source identification conditions of all methods

We start with an experiment where we violate the source identification conditions of all methods. Consider the EVD of the SCV covariance matrices $\mathbf{C}_{\mathbf{s}_n} = \boldsymbol{\Theta}_n\boldsymbol{\Lambda}_n\boldsymbol{\Theta}_n^{\mathsf{T}}$, where $\boldsymbol{\Lambda}_n = \mathrm{diag}\big(\boldsymbol{\lambda}_n\big) \in \mathbb{R}^{K \times K}$ contains the eigenvalues of $\mathbf{C}_{\mathbf{s}_n}$ on its main diagonal, and $\boldsymbol{\Theta}_n \in \mathbb{R}^{K \times K}$ contains the corresponding eigenvectors. This experiment is designed so that all $\mathbf{C}_{\mathbf{s}_n}$ have the same eigenvalues $\boldsymbol{\lambda}_n$ and the same eigenvectors $\boldsymbol{\Theta}_n$ up to a sign (i.e., for some n^{th} and m^{th} SCVs, $\boldsymbol{\Theta}_n = \mathbf{D}_n\boldsymbol{\Theta}_m$ for any diagonal matrix $\mathbf{D}_n$ with diagonal elements equal to either -1 or 1). To achieve this, we first generate $\mathbf{C}_{\mathbf{s}_1} \in \mathbb{R}^{K \times K}$ with $R = K = 100$ and $\alpha_1 = 1$ and perform the EVD $\mathbf{C}_{\mathbf{s}_1} = \boldsymbol{\Theta}_1\boldsymbol{\Lambda}_1\boldsymbol{\Theta}_1^{\mathsf{T}}$. We then generate $\boldsymbol{\Theta}_n = \mathbf{D}_n\boldsymbol{\Theta}_1$, $n = 2, \ldots, N$, where $\mathbf{D}_n$ is a diagonal matrix with randomly generated ± 1 elements on its diagonal. Finally, we generate the remaining SCV covariance matrices as $\mathbf{C}_{\mathbf{s}_n} = \boldsymbol{\Theta}_n\boldsymbol{\Lambda}_1\boldsymbol{\Theta}_n^{\mathsf{T}}$, $n = 2, \ldots, N$. The eigenvalues for one run of the experiment are shown in Figure 2 (top). As the eigenvalues are the same for each SCV covariance matrix and the eigenvectors are also the same (up to a sign), the source identification conditions of all methods are violated.

B. Substantiating the source identification conditions of `genvar` and `ssqcor`

In this experiment, we substantiate our theoretical considerations for the source identification conditions of `genvar` and `ssqcor`. Therefore, we generate the SCV covariance matrices $\mathbf{C}_{\mathbf{s}_n}$ such that they still have the same eigenvalues but different eigenvectors. First, we generate $\mathbf{C}_{\mathbf{s}_1}$ as described in Section VII-A, perform the EVD, and use $\boldsymbol{\lambda}_1 \in \mathbb{R}^K$ as the same eigenvalues for all N SCV covariance matrices, also shown in Figure 2 (top) for one run. However, now we use the scipy function `random_correlation` [41] to generate $\mathbf{C}_{\mathbf{s}_2}, \ldots, \mathbf{C}_{\mathbf{s}_N}$ having these same eigenvalues $\boldsymbol{\lambda}_1$ but different eigenvectors. Thus, we expect `genvar` and `ssqcor` to achieve JBSS but `sumcor`/`maxvar` and `minvar` not, as their proposed source identification conditions are still violated since all SCV covariance matrices have the same largest and smallest eigenvalue.

C. Substantiating the source identification conditions of `minvar`

In this experiment, we generate $\mathbf{C}_{\mathbf{s}_n}$ such that the proposed source identification conditions of `minvar` are satisfied. Therefore, we choose $R = 1$ and specify the values of α_n according to (19) such that the smallest eigenvalues $\lambda_n^{(\min)}$ are $\lambda_n^{[1]} = \{0.1, 0.15, 0.2, 0.25, 0.3\}$ for $n = 1, \ldots, 5$. Note that here we call the smallest eigenvalue $\lambda_n^{[1]}$ instead of $\lambda_n^{[K]}$ to keep the notation as in (19). The eigenvalues for one run are again shown in Figure 2 (top). This is a very artificial scenario, as the resulting covariance matrices are very close to identity matrices, and thus the correlations between the source components are almost zero. The source identification conditions of `maxvar` are almost violated, as the largest eigenvalues are very close to each other, and those of `sumcor` are violated, as the largest eigenvalue of each SCV covariance matrix is

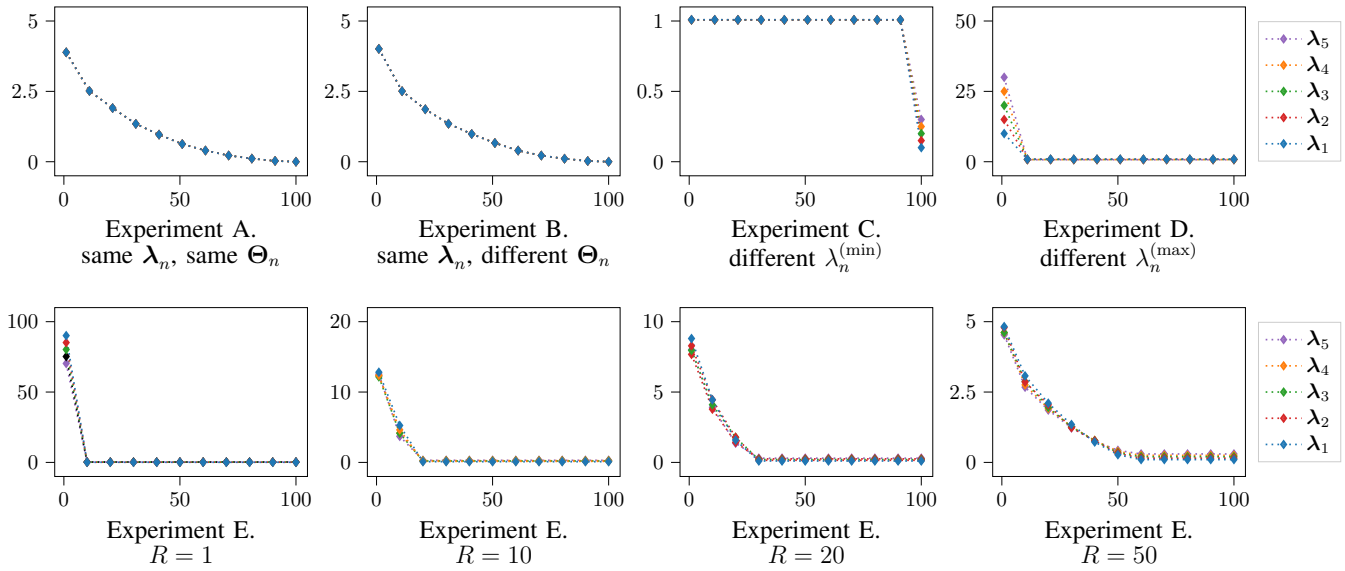


Figure 2. Example eigenvalues for experiments A–E; every tenth value is plotted. In experiment A, the source identification conditions of all methods are violated. Experiment B substantiates the conditions of `genvar` and `ssqcor`, experiment C the conditions of `minvar`, and experiment D the conditions of `sumcor` and `maxvar`. Experiment E investigates the influence of the effective rank R on the methods' ability to achieve JBSS.

not greater than all second-largest eigenvalues. However, we expect `genvar` and `ssqcor` to achieve JBSS here based on our theoretical considerations for their source identification conditions.

D. Substantiating the source identification conditions of `sumcor` and `maxvar`

In this experiment, we generate $\mathbf{C}_{\mathbf{s}_n}$ such that the source identification conditions of `sumcor` and `maxvar` are satisfied. Therefore, we choose $R = 1$ and specify the values of α_n according to (19) such that the largest eigenvalues $\lambda_n^{(\max)}$ are $\lambda_n^{[1]} = \{10, 15, 20, 25, 30\}$ for $n = 1, \ldots, 5$. The eigenvalues for one run are shown in Figure 2 (top). The source identification conditions of `minvar` are almost violated, as the smallest eigenvalues are very close to each other. Also here, we expect `genvar` and `ssqcor` to achieve JBSS.

E. Changing the effective rank R of the SCVs

In this experiment, we generate $\mathbf{C}_{\mathbf{s}_n}$ such that we can investigate the influence of the effective rank R on the methods' ability to achieve JBSS. Therefore, we vary the value of $R = \{1, 2, 5, 10, 20, 50\}$, and for each value of R we choose $\alpha_n = \{0.9, 0.85, 0.8, 0.75, 0.7\}$ for $n = 1, \ldots, 5$. To satisfy the source identification conditions of `sumcor`, we make sure that $\max_n \left(\lambda_n^{[2]}\right) < \min_n \left(\lambda_n^{[1]}\right) - \epsilon$, where $\epsilon = 0.01$, i.e., that the second-largest eigenvalue of all SCVs is smaller than the largest eigenvalue of all SCVs. The eigenvalues for one run for $R = \{1, 10, 20, 50\}$ are shown in Figure 2 (bottom). As we see, the eigenvalues of all SCV covariance matrices are distinct, but as the smallest eigenvalues are very close to each other, the conditions of `minvar` are almost violated, similar to the case in experiment D.

F. Performance metric

To measure if JBSS has been successful, we use the joint Inter Symbol Interference (jISI) from [33], which provides a measure of how well a demixing process (given by $\mathbf{W}^{[k]}$) is able to invert a mixing process (given by $\mathbf{A}^{[k]}$) w.r.t. sign and permutation ambiguities of the estimated sources and is defined as

$$\text{jISI} = \frac{1}{2N(N-1)} \left[\sum_{n=1}^{N} \left(\frac{\sum_{m=1}^{N} g_{nm}}{\max_m g_{nm}} - 1 \right) + \sum_{m=1}^{N} \left(\frac{\sum_{n=1}^{N} g_{nm}}{\max_n g_{nm}} - 1 \right) \right],$$

where g_{nm} is the $(n,m)^{\text{th}}$ element of $\mathbf{G} = \sum_{k=1}^{K} \left| \mathbf{W}^{[k]} \mathbf{A}^{[k]} \right| \in \mathbb{R}^{N \times N}$. Note that the jISI does not require the true and estimated sources to be aligned correctly—the estimated sources only need to be aligned correctly across datasets. The jISI is normalized between zero and one, where values close to zero indicate successful JBSS, i.e., source components are separated correctly within a dataset and also aligned correctly across datasets. A jISI smaller than 0.05 represents excellent JBSS performance. We report the results observed in 50 runs of the experiments.

G. Results for testing the source identification conditions

In Figure 3, we see how successful the five mCCA methods (`sumcor` from [24], and `maxvar`, `minvar`, `genvar`, and `ssqcor` from [4]) are in achieving JBSS in experiments A–E when the methods use the true covariance matrices. The diamonds and lines indicate the mean and standard deviation of the jISI across 50 runs for each experiment. The dotted gray line shows where the jISI is 0.05. In experiment A, where all SCV covariance matrices have the same eigenvalues and the same eigenvectors (up to a sign), the source identification conditions of all methods are violated, and we see that, indeed, no method is able to achieve JBSS. In experiment B, where the eigenvalues are still the same but the eigenvectors are different across SCVs, `ssqcor` and `genvar` achieve JBSS, while `sumcor`, `maxvar`, and `minvar` fail. This supports our theoretical considerations that `genvar` could achieve JBSS

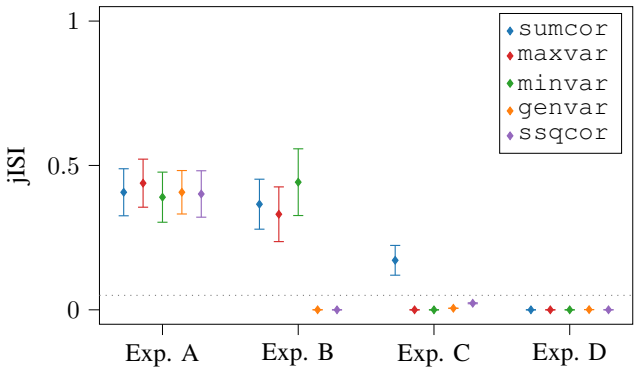


Figure 3. Joint ISI value (lower is better) for experiments A–D when the methods use the true covariance matrices. The diamonds show the mean value and the end lines plus or minus one standard deviation across 50 runs. The dotted gray line shows of jISI of 0.05, and values smaller than that indicate excellent JBSS performance. The `sumcor` method is according to Nielsen [24], and the other methods are according to Kettenring [4]. The methods perform as expected from their proposed and discussed source identification conditions.

even if the eigenvalues are the same across SCVs as long as the eigenvectors are different, and that the source identification conditions of `ssqcor` could be similar to those of `genvar`. In experiment C, where all SCVs have a different smallest eigenvalue, `minvar` achieves JBSS, supporting Conjecture 1. `Maxvar` also achieves JBSS, since the largest eigenvalue is also different for each SCV (even though the difference is very small), but `sumcor` fails as the largest eigenvalues are equal to the second-largest, substantiating Theorems 3 and 4. Finally, in experiment D, where the largest eigenvalue is different for each SCV (and the smallest eigenvalue is also different with a very small difference), all methods are successful. To conclude these experiments, our simulations support the proposed theoretical source identification conditions.

H. Results for testing source identification conditions and statistical efficiency

In Figures 4–5, we see how using finite samples affects the methods' ability to achieve JBSS. We repeat experiments C and D each with $V = \{100, 200, 500, 1000, 2000, 5000, \ldots, 1000000\}$ samples and experiment E with $V = \{10000, 100000\}$ samples.

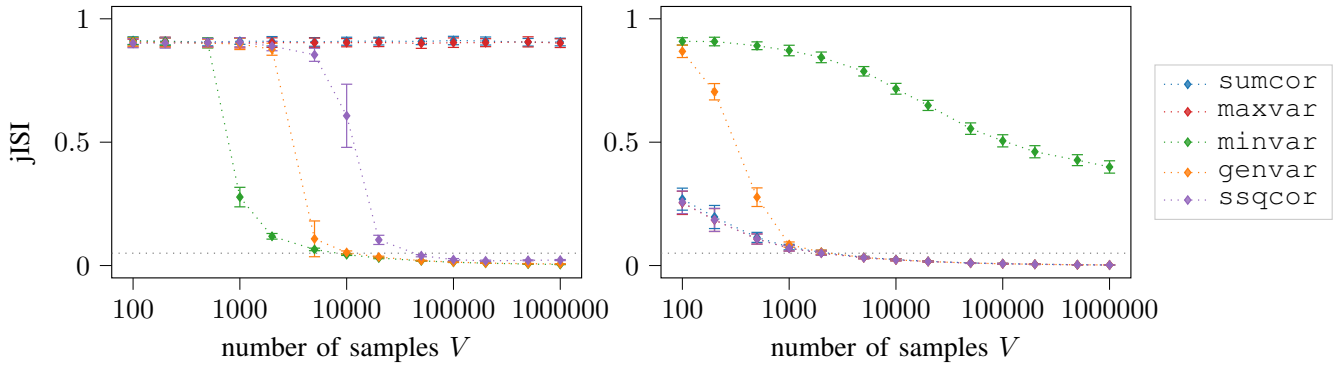
In Figure 4, we see that for $V < 500$, i.e., $V < NK$, all mCCA methods fail to achieve JBSS. This can be explained as follows. For $K = 2$ datasets, it has been shown that when $V < 2N$, some of the estimated canonical correlations are equal to one (independent of the true underlying correlation) [42], and thus the canonical variables do not recover the true sources (which have correlation smaller than one). Even when $V \approx 2N$, CCA highly overestimates the canonical correlations, and the estimated canonical variables are noisy [43]. As mCCA is an extension of CCA to more than two data sets ($K > 2$), these effects are also expected for mCCA. Note that if a small sample size occurs in practice, dimension reduction (e.g., with Principal Component Analysis (PCA)) can be carefully applied such that the sum of reduced dimensions is smaller than the number of samples while keeping all correlated sources in the dimension-reduced datasets [44]. Then, the mCCA methods can be applied to the dimension-reduced datasets to ensure accurate estimation of the underlying sources for JBSS. In our experiments, the JBSS performance of the methods generally improves with an increasing number of samples $V \gg NK$. In Figure 4(a), we see that `minvar`, `genvar`, and `ssqcor` are able to achieve JBSS in experiment C starting from $V = 10000$, 20000, 50000 samples, respectively, while `sumcor` and `maxvar` are constantly failing for all numbers of samples because the largest eigenvalues of the estimated covariance matrices are too similar, which violates their source identification conditions. In Figure 4(b), we see that `maxvar`, `genvar`, and `ssqcor` achieve JBSS in experiment D starting from $V = 5000$ samples, while `minvar` fails because the smallest eigenvalues of the estimated covariance matrices are too similar. However, it is noticeable that the JBSS performance of `minvar` improves in experiment D with a higher number of samples, while `sumcor` and `maxvar` are not improving at all in experiment C.

In Figure 5(a), where the covariance matrices are estimated from $V = 10000$ samples, we see that `minvar` fails to achieve JBSS, which is also expected as the smallest eigenvalues are very similar to each other. For $R = 1$, all other methods achieve JBSS, but for an increasing value of R, the JBSS performance of `sumcor` and `maxvar` decreases and both fail to achieve JBSS for $R \geq 2$. This can be explained by the fact that both methods assume that the SCVs have an effective rank of $R = 1$ because of their underlying factor models in (5) and (12). In contrast, `genvar` and `ssqcor` achieve JBSS for all values of R in this experiment, as they can estimate SCVs with any effective rank from 1 to K according to their underlying factor model in (16). In Figure 5(b), for $V = 100000$ samples, the trends are the same, but all methods have improved their performance. However, still only `genvar` and `ssqcor` can achieve JBSS for $R \geq 2$. Note that, when the methods use the true covariance matrices, all of them are able to achieve JBSS for all values of R (not shown here), which is expected as the eigenvalues of all SCV covariance matrices are distinct (although with only very small differences).

To conclude these experiments, `genvar` seems to be the most statistically efficient mCCA method, as it can achieve JBSS already for $V = 10000$ samples in experiments C and D and for all values of R in experiment E, and its performance is not so strongly affected by using finite samples compared with the other methods.

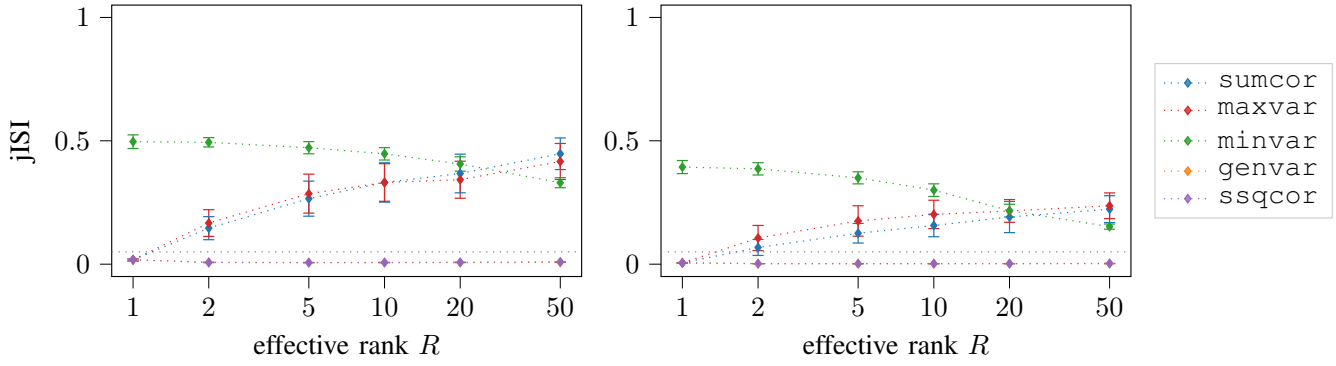
VIII. Summary and Discussion

In this paper, we have summarized the optimization problems and solutions of the five mCCA methods `sumcor`, `maxvar`, `minvar`, `genvar`, and `ssqcor` and have described their solutions including pseudocodes. We have derived the connections between the two most commonly used mCCA methods `sumcor` and `maxvar`. Thereafter, we have discussed the source identification conditions of the mCCA methods under the assumption of the JBSS generative model. We have substantiated the proposed source identification conditions of the mCCA methods with simulations, where we have



(a) Experiment C (b) Experiment D

Figure 4. Joint ISI value (mean and standard deviation across 50 runs, lower is better) for experiments C and D for a different number of samples V. In both experiments, all methods fail to achieve JBSS for $V < NK$, as expected. In experiment C, starting from 10000, 20000, and 50000 samples, `minvar`, `genvar`, and `ssqcor` achieve JBSS, respectively, while `sumcor` and `maxvar` constantly fail. In experiment D, starting from 5000 samples, all methods except for `minvar` achieve JBSS, respectively.



(a) $V = 10000$ samples (b) $V = 100000$ samples

Figure 5. Joint ISI value (mean and standard deviation across 50 runs, lower is better) for experiment E for different values of R. (a) For $V = 10000$ samples, all methods except `minvar` achieve JBSS for $R = 1$. For an increasing value of R, the performance of `minvar` improved while that of `sumcor` and `maxvar` decreases and both fail to achieve JBSS from $R \geq 2$. Lastly, `genvar` and `ssqcor` achieve JBSS for all values of R. (b) The trends for $V = 100000$ samples are the same, but all methods have improved their performance.

observed that `genvar` appears to have the least restrictive source identification among all methods conditions and thus can identify sources that the other methods may not recover. This is especially notable for a finite number of samples, where we have observed in our experiments that `genvar` seems to be the most statistically efficient method. We have furthermore observed that the ability of the methods to achieve JBSS with finite samples is related to their underlying factor models. Some of them have similar underlying factor models, and as such, we discuss them together. First, `maxvar` [4] and `sumcor` [24] both are effective in estimating SCVs with an effective rank of one, meaning that the sources of all datasets within each SCV have the same underlying factor. Their main difference is that the sources within a dataset are constrained to be uncorrelated in `maxvar`, but not in `sumcor`, which leads to better interpretability of the `maxvar` sources and makes `maxvar` the preferred method of these two. Then, `genvar` [4] and `ssqcor` [4] both can effectively estimate SCVs with an arbitrary effective rank, while it is worth mentioning that `ssqcor` performs similar to `maxvar` if the SCVs have an effective rank of one. Lastly, our experience is that `minvar` [4] is only effective for estimating SCVs in a very artificial scenario. Understanding which methods can achieve JBSS under which conditions is important for interpreting estimated sources in real-world applications. However, for real-world datasets without knowledge on the ground truth, it is often difficult to verify whether the source identification conditions of a specific method are satisfied. One attempt to do so is to calculate the eigenvalues and eigenvectors of the estimated SCV covariance matrices and inspect whether they are close to violating the conditions. If that is the case, it is likely that the given method is not a good choice for the data and perhaps another method should be preferred. To conclude our findings, we summarize our recommended uses of the mCCA methods as follows:

- In general, we recommend choosing `genvar` [4] for achieving JBSS based on our experiments, as it appears to have the least restrictive source identification conditions and to be more statistically efficient than the other methods. This is a significant finding, as `sumcor` and `maxvar`, despite being the perhaps most commonly used mCCA methods [5], are outperformed by `genvar` in our simulation study.
- However, the computational complexity of `genvar` is

the highest among all mCCA methods, with differences becoming stronger with an increasing number of datasets or dimension of data. Therefore, in specific scenarios, it makes sense to use a faster mCCA method. As the ability of the methods to achieve JBSS depends on their underlying factor models, `maxvar` [4] may also be able to achieve JBSS if it is reasonable to assume that the true SCVs have approximately an effective rank of one. This might, for example, be the case for a large number of SCVs in the analysis of functional Magnetic Resonance Imaging (fMRI) datasets [45].

The insights we provided in this paper form the basis for further research. The next logical step is to formally prove the conjecture for `minvar` and to derive and prove the source identification conditions of `genvar` and `ssqcor`. Additionally, the development of a deep learning-based extension of `genvar` represents a promising direction in JBSS, as a deep `genvar` may outperform DGCCA [15], a deep extension of `maxvar` [38], in JBSS for real-world data where source components in an SCV are assumed to be nonlinearly dependent and generated by more than one underlying factor.

APPENDIX

A. Proof of Theorem 1

The proof of Theorem 1 consists of three steps.

1) Write the `sumcor` GEVD in (3) and the `maxvar` EVD in (10) in terms of the concatenated whitened datasets $\mathbf{Y}$.
2) Derive the relation of the eigenvectors in these two EVDs.
3) Find the connection of the canonical vectors.

Proof. 1) The covariance matrix of the concatenated whitened datasets is $\widehat{\mathbf{C}}_{\mathbf{y}} = \frac{1}{V}\mathbf{Y}\mathbf{Y}^{\mathsf{T}} \in \mathbb{R}^{NK\times NK}$, and

$$\widehat{\mathbf{D}}_{\mathbf{x}} = \text{blkdiag}\Big(\frac{1}{V}\mathbf{X}^{[1]}\big(\mathbf{X}^{[1]}\big)^{\mathsf{T}}, \ldots, \frac{1}{V}\mathbf{X}^{[K]}\big(\mathbf{X}^{[K]}\big)^{\mathsf{T}}\Big) \quad (20)$$

The `sumcor` GEVD in (3), given by $\widehat{\mathbf{C}}_{\mathbf{x}}\mathbf{T} = \widehat{\mathbf{D}}_{\mathbf{x}}\mathbf{T}\mathbf{\Phi}$, can be written in terms of $\widehat{\mathbf{C}}_{\mathbf{y}}$ by multiplying it from the left with $\widehat{\mathbf{D}}_{\mathbf{x}}^{-1/2}$ [24]:

$$\widehat{\mathbf{C}}_{\mathbf{y}}\mathbf{B} = \mathbf{B}\mathbf{\Phi}, \quad (21)$$

where $\widehat{\mathbf{C}}_{\mathbf{y}} = \widehat{\mathbf{D}}_{\mathbf{x}}^{-1/2}\widehat{\mathbf{C}}_{\mathbf{x}}\widehat{\mathbf{D}}_{\mathbf{x}}^{-1/2}$, the matrix $\mathbf{B} = \widehat{\mathbf{D}}_{\mathbf{x}}^{1/2}\mathbf{T} \in \mathbb{R}^{NK\times NK}$ contains the eigenvectors of $\widehat{\mathbf{C}}_{\mathbf{y}}$, and $\mathbf{\Phi} \in \mathbb{R}^{NK\times NK}$ contains its eigenvalues on the main diagonal. Lastly, let $\widetilde{\mathbf{T}} = \big[\mathbf{t}_1, \ldots, \mathbf{t}_N\big] \in \mathbb{R}^{NK\times N}$ consist of the N principal generalized eigenvectors of $\widehat{\mathbf{C}}_{\mathbf{x}}$ w.r.t. $\widehat{\mathbf{D}}_{\mathbf{x}}$, and let $\widetilde{\mathbf{B}} = \big[\mathbf{b}_1, \ldots, \mathbf{b}_N \in \mathbb{R}^{NK\times N}$ consist of the N principal eigenvectors of $\widehat{\mathbf{C}}_{\mathbf{y}}$, with $\widetilde{\mathbf{B}}^{\mathsf{T}}\widetilde{\mathbf{B}} = \mathbf{I}_{N\times N}$.
In the `maxvar` EVD in (10), given by $\mathbf{Q}\mathbf{Z} = \mathbf{Z}\mathbf{\Phi}$, the matrix $\mathbf{Q}$ can be written in terms of $\mathbf{Y}$ as

$$\begin{aligned}\mathbf{Q} &= \sum_{k=1}^{K}\frac{1}{V}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big(\frac{1}{V}\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big)^{-1}\mathbf{X}^{[k]} \\ &= \frac{1}{V}\sum_{k=1}^{K}\big(\mathbf{Y}^{[k]}\big)^{\mathsf{T}}\mathbf{Y}^{[k]} = \frac{1}{V}\mathbf{Y}^{\mathsf{T}}\mathbf{Y} \in \mathbb{R}^{V\times V},\end{aligned}$$

where we have chosen the weighting factor to be equal to one. Let $\widetilde{\mathbf{Z}} = \big[\mathbf{z}_1, \ldots, \mathbf{z}_N\big] \in \mathbb{R}^{V\times N}$ consist of the N principal eigenvectors of $\mathbf{Q}$, with $\widetilde{\mathbf{Z}}^{\mathsf{T}}\widetilde{\mathbf{Z}} = \mathbf{I}_{N\times N}$.

2) Now, let the SVD of $\mathbf{Y}$ be $\mathbf{Y} = \mathbf{B}\mathbf{\Omega}\mathbf{Z}^{\mathsf{T}}$, where $\mathbf{B} \in \mathbb{R}^{NK\times NK}$, $\mathbf{\Omega} \in \mathbb{R}^{NK\times NK}$, and $\mathbf{Z} \in \mathbb{R}^{V\times NK}$, and let the rank-$N$ approximation of $\mathbf{Y}$ be $\mathbf{Y} = \widetilde{\mathbf{B}}\widetilde{\mathbf{\Omega}}\widetilde{\mathbf{Z}}^{\mathsf{T}}$, with $\widetilde{\mathbf{B}}$ and $\widetilde{\mathbf{Z}}$ as defined above, and $\widetilde{\mathbf{\Omega}} = \text{diag}\big(\omega_1, \ldots, \omega_N\big) \in \mathbb{R}^{N\times N}$. With the partitions of $\mathbf{Y}$ and $\widetilde{\mathbf{B}}$, we have $\begin{bmatrix}\mathbf{Y}^{[1]}\\ \vdots \\ \mathbf{Y}^{[K]}\end{bmatrix} = \begin{bmatrix}\mathbf{B}^{[1]}\\ \vdots \\ \mathbf{B}^{[K]}\end{bmatrix}\widetilde{\mathbf{\Omega}}\widetilde{\mathbf{Z}}^{\mathsf{T}}$, and thus, for each dataset $\mathbf{Y}^{[k]}\widetilde{\mathbf{Z}} = \mathbf{B}^{[k]}\widetilde{\mathbf{\Omega}}$.
3) We know from (4) that the `sumcor` transformation matrices $\mathbf{T}^{[k]}$ are the blocks of $\widetilde{\mathbf{T}} = \big[\big(\mathbf{T}^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{T}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK\times N}$. Thus, by inserting (20) in $\widetilde{\mathbf{T}} = \widehat{\mathbf{D}}_{\mathbf{x}}^{-1/2}\widetilde{\mathbf{B}}$, we find

$$\mathbf{T}^{[k]}_{\text{sumcor}} = \Big(\frac{1}{V}\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big)^{-\frac{1}{2}}\mathbf{B}^{[k]}. \quad (22)$$

Furthermore, we know from (11) that the `maxvar` transformation matrices are $\mathbf{T}^{[k]} = \Big(\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big)^{-1}\mathbf{X}^{[k]}\widetilde{\mathbf{Z}}$, and rewriting gives

$$\mathbf{T}^{[k]}_{\text{maxvar}} = \frac{1}{V}\Big(\frac{1}{V}\mathbf{X}^{[k]}\big(\mathbf{X}^{[k]}\big)^{\mathsf{T}}\Big)^{-1/2}\mathbf{Y}^{[k]}\widetilde{\mathbf{Z}} = \frac{1}{V}\mathbf{T}_{\text{sumcor}}\,\widetilde{\mathbf{\Omega}}$$

For the canonical vectors, this means

$$\mathbf{U}^{[k]}_{\text{maxvar}} = \big(\mathbf{T}^{[k]}_{\text{maxvar}}\big)^{\mathsf{T}}\mathbf{X}^{[k]} = \frac{1}{V}\widetilde{\mathbf{\Omega}}\,\mathbf{U}^{[k]}_{\text{sumcor}},$$

i.e., the canonical variables of the all-at-once analytical `maxvar` [38] and the all-at-once analytical `sumcor` [24] are scaled versions of each other.

□

B. Proof of Theorem 2

The proof of Theorem 2 consists of two steps:

1) Write the `sumcor` EVD in (3) depending on the whitened datasets $\mathbf{y}^{[k]}$ in a deflationary way.
2) Compare the deflationary `sumcor` EVD with the deflationary `maxvar` EVD in (7).

Proof. 1) Let $\mathbf{y}^{[k]} = \big(\mathbf{C}^{[k]}_{\mathbf{x}}\big)^{-\frac{1}{2}} \in \mathbb{R}^N$ be the k^{th} whitened dataset, and let $\mathbf{C}_{\mathbf{y}} = \mathbb{E}\big\{\mathbf{y}\mathbf{y}^{\mathsf{T}}\big\} \in \mathbb{R}^{NK\times NK}$ be the covariance matrix of the concatenated whitened datasets $\mathbf{y} = \big[\big(\mathbf{y}^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{y}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathrm{R}^{NK}$. As in (21), the `sumcor` EVD can be written as $\mathbf{C}_{\mathbf{y}}\mathbf{B} = \mathbf{\Phi}\mathbf{B}$, where $\mathbf{B} = \big[\mathbf{b}_1, \ldots, \mathbf{b}_{NK}\big] \in \mathbb{R}^{NK\times NK}$ contains the eigenvectors of $\mathbf{C}_{\mathbf{y}}$, and $\mathbf{b}_n = \big[\big(\mathbf{b}^{[1]}_n\big)^{\mathsf{T}}, \ldots, \big(\mathbf{b}^{[K]}_n\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$, $n = 1, \ldots, N$, contains the transformation vectors $\mathbf{b}^{[k]}_n$ for the k^{th} whitened dataset. Now we define $\mathbf{B}_{(n-1)} = \big[\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}\big] \in \mathbb{R}^{NK\times(n-1)}$ and $\mathbf{H}_n$ as [46]

$$\mathbf{H}_n = \mathbf{I} - \mathbf{B}_{(n-1)}\big(\mathbf{B}^{\mathsf{T}}_{(n-1)}\mathbf{B}_{(n-1)}\big)^{-1}\mathbf{B}^{\mathsf{T}}_{(n-1)} \in \mathbb{R}^{NK\times NK}.$$

Then, $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}$ is the projection of $\mathbf{C}_{\mathbf{y}}$ on the subspace that is orthogonal to $\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}$ [46]. Let us denote the principal eigenvector of $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}$ by $\mathbf{v}_n$. Now, it is obvious to see that $\mathbf{v}_n = \mathbf{b}_n$, i.e., that the principal eigenvector of $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}$ equals the n^{th} eigenvector of $\mathbf{C}_{\mathbf{y}}$. As $\mathbf{H}_n$ is symmetric and idempotent, the eigenvectors of $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}$ are equal to those of $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}\mathbf{H}_n$ [4]. Thus, the all-at-once `sumcor` EVD in (21) can also be performed in a deflationary way by performing the following EVD sequentially for $n = 1, \ldots, N$:

$$\mathbf{H}_n\mathbf{C}_{\mathbf{y}}\mathbf{H}_n\mathbf{v}_n = \phi_n\mathbf{v}_n,$$

where ϕ_n denotes the largest eigenvalue of $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}\mathbf{H}_n$, $\mathbf{v}_n = \big[\big(\mathbf{v}^{[1]}_n\big)^{\mathsf{T}}, \ldots, \big(\mathbf{v}^{[K]}_n\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$ denotes the corresponding

eigenvector, and $\mathbf{H}_n$ is calculated as follows. In the first EVD, $\mathbf{H}_1 = \mathbf{I}_{NK\times NK}$, and then for $n = 2, \ldots, N$,

$$\mathbf{H}_n = \mathbf{I} - \mathbf{V}_{(n-1)}\big(\mathbf{V}_{(n-1)}^{\mathsf{T}}\mathbf{V}_{(n-1)}\big)^{-1}\mathbf{V}_{(n-1)}^{\mathsf{T}},$$

where

$$\mathbf{V}_{(n-1)} = \big[\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\big] = \begin{bmatrix} \mathbf{v}_1^{[1]} & \cdots & \mathbf{v}_{n-1}^{[1]} \\ \vdots & & \vdots \\ \mathbf{v}_1^{[K]} & \cdots & \mathbf{v}_{n-1}^{[K]} \end{bmatrix}$$

$\in \mathbb{R}^{NK\times(n-1)}$ consists of the eigenvectors of the $n-1$ previous EVDs.

2) In the deflationary `maxvar` EVD in (7), also given by $\mathbf{H}_n\mathbf{C}_{\mathbf{y}}\mathbf{H}_n\mathbf{v}_n = \phi_n\mathbf{v}_n$, the definition of $\mathbf{H}_n$ is different:

$$\mathbf{H}_n = \mathbf{I} - \widetilde{\mathbf{V}}_{(n-1)}\big(\widetilde{\mathbf{V}}_{(n-1)}^{\mathsf{T}}\widetilde{\mathbf{V}}_{(n-1)}\big)^{-1}\widetilde{\mathbf{V}}_{(n-1)}^{\mathsf{T}},$$

where

$$\widetilde{\mathbf{V}}_{(n-1)} = \begin{bmatrix} \widetilde{\mathbf{v}}_1^{[1]} & \cdots & \widetilde{\mathbf{v}}_{n-1}^{[1]} & & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \ddots & \vdots & & \vdots \\ \vdots & & \vdots & \ddots & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & & \widetilde{\mathbf{v}}_1^{[K]} & \cdots & \widetilde{\mathbf{v}}_{n-1}^{[K]} \end{bmatrix}$$

$\in \mathbb{R}^{NK\times(n-1)K}$ consists of the partitioned and normalized eigenvectors $\widetilde{\mathbf{v}}_n^{[k]} = \frac{\mathbf{v}_n^{[k]}}{\|\mathbf{v}_n^{[k]}\|} \in \mathbb{R}^N$ of the $n-1$ previous EVDs. While the first eigenvectors $\mathbf{v}_1^{[k]}$ from Kettenring are the same as $\mathbf{v}_1^{[k]}$ from Nielsen, $k = 1, \ldots, K$, all following eigenvectors are different because the matrix $\widetilde{\mathbf{V}}_{(n-1)}$ is built differently than $\mathbf{V}_{(n-1)}$: Instead of concatenating the $n-1$ previous eigenvectors horizontally, Kettenring partitions the eigenvectors according to the datasets they belong to, normalizes these partitions, and then stacks all normalized partitions into a block-diagonal matrix. As a result, in `sumcor` [24], the $\mathbf{v}_n$, $\mathbf{v}_m$ are orthogonal, i.e., $\mathbf{v}_n^{\mathsf{T}}\mathbf{v}_m = \sum_{k=1}^{K}\big(\mathbf{v}_n^{[k]}\big)^{\mathsf{T}}\mathbf{v}_m^{[k]} = 0$, while in `maxvar` [4], the $\widetilde{\mathbf{v}}_n^{[k]}$, $\widetilde{\mathbf{v}}_m^{[k]}$ are orthogonal, i.e., $\big(\widetilde{\mathbf{v}}_n^{[k]}\big)^{\mathsf{T}}\widetilde{\mathbf{v}}_m^{[k]} = 0$, $k = 1, \ldots, K$. □

C. Proof of Theorem 3

In order to prove Theorem 3, we prove that if the source identification conditions of `sumcor` are satisfied, then the `sumcor` transformation matrices $\mathbf{T}^{[k]}$ and the mixing matrices $\mathbf{A}^{[k]}$ are related as follows:

$$\big(\mathbf{T}^{[k]}\big)^{\mathsf{T}} = \mathbf{P}\mathbf{\Gamma}^{[k]}\big(\mathbf{A}^{[k]}\big)^{-1},$$

which is equivalent to $\mathbf{u}^{[k]} = \mathbf{P}\mathbf{\Gamma}^{[k]}\mathbf{s}^{[k]}$ but slightly shorter to prove. The proof of Theorem 3 consists of three steps.

1) Derive the relationship of $\widetilde{\mathbf{T}}$, the concatenated `sumcor` transformation matrices in (8), and $\widetilde{\mathbf{\Delta}}$, the eigenvectors of the concatenated SCVs' covariance matrix $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$: $\widetilde{\mathbf{T}} = \mathbf{A}^{-\mathsf{T}}\mathbf{\Pi}\widetilde{\mathbf{\Delta}}$, where $\mathbf{\Pi}$ is a specific permutation matrix.
2) Show that the structure of $\widetilde{\mathbf{\Delta}}$ is $\widetilde{\mathbf{\Delta}} = \mathrm{diag}\big(\boldsymbol{\theta}_1^{[1]}, \ldots, \boldsymbol{\theta}_N^{[1]}\big)\mathbf{P}^{\mathsf{T}}$ under the conditions in Theorem 3.
3) Use 1 and 2 to show that $\big(\mathbf{T}^{[k]}\big)^{\mathsf{T}} = \mathbf{P}\mathbf{\Gamma}^{[k]}\big(\mathbf{A}^{[k]}\big)^{-1}$.

Proof. 1) First, we express $\mathbf{C}_{\mathbf{x}}$ and $\mathbf{D}_{\mathbf{x}}$ in the GEVD in (3), given by $\mathbf{C}_{\mathbf{x}}\mathbf{T} = \mathbf{D}_{\mathbf{x}}\mathbf{T}\mathbf{\Phi}$, in terms of $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$, the covariance matrix of the concatenated SCVs. Therefore, we define the concatenated source vector $\mathbf{s} = \big[\big(\mathbf{s}^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{s}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$ with covariance matrix

$$\mathbf{C}_{\mathbf{s}} = \mathbb{E}\big\{\mathbf{s}\mathbf{s}^{\mathsf{T}}\big\} = \begin{bmatrix} \mathbf{I}_{N\times N} & \mathbf{C}_{ss}^{[1,2]} & \cdots & \mathbf{C}_{ss}^{[1,K]} \\ \mathbf{C}_{ss}^{[2,1]} & \mathbf{I}_{N\times N} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{C}_{ss}^{[K-1,K]} \\ \mathbf{C}_{ss}^{[K,1]} & \cdots & \mathbf{C}_{ss}^{[K,K-1]} & \mathbf{I}_{N\times N} \end{bmatrix},$$

where we used that $\mathbf{C}_{\mathbf{s}}^{[k]} = \mathbf{I}_{N\times N}$, $k = 1, \ldots, K$, according to Assumptions (A2) and (A3), we define the joint mixing matrix $\mathbf{A} = \mathrm{blkdiag}\big(\mathbf{A}^{[1]}, \ldots, \mathbf{A}^{[K]}\big) \in \mathbb{R}^{NK\times NK}$, and we define the concatenated datasets $\mathbf{x} = \big[\big(\mathbf{x}^{[1]}\big)^{\mathsf{T}}, \ldots, \big(\mathbf{x}^{[K]}\big)^{\mathsf{T}}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$. Now, we can write the JBSS model in (2), given by $\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]}$, $k = 1, \ldots, K$, for all datasets in one equation: $\mathbf{x} = \mathbf{A}\mathbf{s}$. Then, the covariance matrix of the concatenated datasets is $\mathbf{C}_{\mathbf{x}} = \mathbb{E}\big\{\mathbf{x}\mathbf{x}^{\mathsf{T}}\big\} = \mathbf{A}\mathbf{C}_{\mathbf{s}}\mathbf{A}^{\mathsf{T}} \in \mathbb{R}^{NK\times NK}$, and $\mathbf{D}_{\mathbf{x}} = \mathbf{A}\,\mathrm{blkdiag}\big(\mathbf{I}_{N\times N}, \ldots, \mathbf{I}_{N\times N}\big)\mathbf{A}^{\mathsf{T}} = \mathbf{A}\mathbf{A}^{\mathsf{T}} \in \mathbb{R}^{NK\times NK}$. Now, we define the covariance matrix of the concatenated SCVs as the block-diagonal matrix $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})} = \mathrm{blkdiag}\big(\mathbf{C}_{\mathbf{s}_1}, \ldots, \mathbf{C}_{\mathbf{s}_N}\big) \in \mathbb{R}^{NK\times NK}$, where the off-diagonal blocks are equal to zero because the SCVs are assumed to be uncorrelated by Assumption (A3). Then, we can write $\mathbf{C}_{\mathbf{s}}$, the covariance matrix of the concatenated source vectors, as a permuted version of $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$: $\mathbf{C}_{\mathbf{s}} = \mathbf{\Pi}\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}\mathbf{\Pi}^{\mathsf{T}} \in \mathbb{R}^{N\times N}$, where $\mathbf{\Pi} \in \mathbb{R}^{NK\times NK}$ is a specific permutation matrix. This gives $\mathbf{C}_{\mathbf{x}} = \mathbf{A}\mathbf{\Pi}\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}\mathbf{\Pi}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}$, and the GEVD in (3) becomes

$$\mathbf{A}\mathbf{\Pi}\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}\mathbf{\Pi}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{T} = \mathbf{A}\mathbf{A}^{\mathsf{T}}\mathbf{T}\mathbf{\Phi}. \tag{23}$$

Multiplying (23) from the left with $\mathbf{\Pi}^{\mathsf{T}}\mathbf{A}^{-1}$ results in $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}\mathbf{\Pi}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{T} = \mathbf{\Pi}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{T}\mathbf{\Phi}$, and defining $\mathbf{\Delta} = \mathbf{\Pi}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{T}$ yields the EVD

$$\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}\mathbf{\Delta} = \mathbf{\Delta}\mathbf{\Phi}.$$

Let $\widetilde{\mathbf{T}} = \big[\mathbf{t}_1, \ldots, \mathbf{t}_N\big] \in \mathbb{R}^{NK\times N}$ consist of the N principal generalized eigenvectors of $\mathbf{C}_{\mathbf{x}}$ w.r.t. $\mathbf{D}_{\mathbf{x}}$, and let $\widetilde{\mathbf{\Delta}} = \big[\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_N\big] \in \mathbb{R}^{NK\times N}$ consist of the N principal eigenvectors of $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$. We know that $\widetilde{\mathbf{T}}$ consists of the concatenated `sumcor` transformation matrices $\mathbf{T}^{[k]}$, i.e., $\widetilde{\mathbf{T}} = \big[\mathbf{T}^{[1]}, \ldots, \mathbf{T}^{[K]}\big]^{\mathsf{T}}$, and their relationship with the eigenvectors of $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$ is

$$\widetilde{\mathbf{T}} = \mathbf{A}^{-\mathsf{T}}\mathbf{\Pi}\widetilde{\mathbf{\Delta}}. \tag{24}$$

2) We now determine the structure of $\widetilde{\mathbf{\Delta}}$. Let $\boldsymbol{\lambda}_n = \big[\lambda_n^{[1]}, \ldots, \lambda_n^{[K]}\big] \in \mathbb{R}^K$ be the eigenvalues of $\mathbf{C}_{\mathbf{s}_n}$ such that $\lambda_n^{[1]} \geq \cdots \geq \lambda_n^{[K]}$, and let $\mathbf{\Theta}_n = \big[\boldsymbol{\theta}_n^{[1]}, \ldots, \boldsymbol{\theta}_n^{[K]}\big] \in \mathbb{R}^{K\times K}$ be the corresponding eigenvectors. Let $\boldsymbol{\phi} = \big[\phi_1, \ldots, \phi_{NK}\big]^{\mathsf{T}} \in \mathbb{R}^{NK}$ be the eigenvalues of $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$ such that $\phi_1 \geq \cdots \geq \phi_{NK}$. Under the conditions given in Theorem 3, the N largest eigenvalues of $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$ are equal to the set consisting of the largest eigenvalue of each $\mathbf{C}_{\mathbf{s}_n}$, $n = 1, \ldots, N$, i.e.,

$$\big\{\phi_1, \ldots, \phi_N\big\} = \big\{\lambda_1^{[1]}, \ldots, \lambda_N^{[1]}\big\},$$

and the non-zero chunks of the N principal eigenvectors of $\mathbf{C}_{\mathbf{s}}^{(\mathrm{SCV})}$ thus consist of the principal eigenvector $\boldsymbol{\theta}_n^{[1]}$ of each $\mathbf{C}_{\mathbf{s}_n}$, $n = 1, \ldots, N$, i.e.,

$$\widetilde{\mathbf{\Delta}} = \begin{bmatrix} \boldsymbol{\theta}_1^{[1]} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\theta}_N^{[1]} \end{bmatrix}\mathbf{P}^{\mathsf{T}}, \tag{25}$$

with $\boldsymbol{\theta}_n^{[1]} = \big[\theta_n^{[1]}(1), \ldots, \theta_n^{[1]}(K)\big]^{\mathsf{T}}$. Note that we have introduced a permutation matrix $\mathbf{P} \in \mathbb{R}^{N\times N}$ as the columns of $\widetilde{\mathbf{\Delta}}$

can be in arbitrary ordering depending on the ordering of the largest eigenvalue of each $\mathbf{C}_{\mathbf{s}_n}$.

3) Now, inserting (25) in (24), the permutation matrix $\mathbf{\Pi}$ resorts the rows of $\widetilde{\mathbf{\Delta}}$, and we have

$$\widetilde{\mathbf{T}} = \mathbf{A}^{-\mathsf{T}}\left[\mathbf{\Gamma}^{[1]}, \ldots, \mathbf{\Gamma}^{[K]}\right]^{\mathsf{T}}\mathbf{P}^{\mathsf{T}},$$

where $\mathbf{\Gamma}^{[k]} = \operatorname{diag}\big(\theta_1^{[1]}(k), \ldots, \theta_N^{[1]}(k)\big) \in \mathbb{R}^{N\times N}$ and $\theta_n^{[1]}(k)$ is the k^{th} element of the principal eigenvector of $\mathbf{C}_{\mathbf{s}_n}$. Using the partitioning of $\widetilde{\mathbf{T}}$ as defined in (4), we see that

$$\begin{bmatrix}\mathbf{T}^{[1]}\\ \vdots \\ \mathbf{T}^{[K]}\end{bmatrix} = \begin{bmatrix}\left(\mathbf{A}^{[1]}\right)^{-\mathsf{T}} & & \mathbf{0}\\ & \ddots & \\ \mathbf{0} & & \left(\mathbf{A}^{[K]}\right)^{-\mathsf{T}}\end{bmatrix}\begin{bmatrix}\mathbf{\Gamma}^{[1]}\\ \vdots \\ \mathbf{\Gamma}^{[K]}\end{bmatrix}\mathbf{P}^{\mathsf{T}},$$

and thus, $\left(\mathbf{T}^{[k]}\right)^{\mathsf{T}} = \mathbf{P}\mathbf{\Gamma}^{[k]}\left(\mathbf{A}^{[k]}\right)^{-1}$, $k = 1, \ldots, K$, i.e., $\left(\mathbf{T}^{[k]}\right)^{\mathsf{T}}$ correctly estimates $\left(\mathbf{A}^{[k]}\right)^{-1}$ up to a scaling and a common permutation. □

REFERENCES

[1] T. Adali, et al., "Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1478–1493, 2015.

[2] T. Adali, et al., "Reproducibility in matrix and tensor decompositions: Focus on model match, interpretability, and uniqueness," *IEEE Signal Processing Magazine*, vol. 39, no. 4, pp. 8–24, 2022.

[3] H. Hotelling, "Relations Between Two Sets of Variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[4] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[5] M. Sorensen, et al., "Generalized Canonical Correlation Analysis: A Subspace Intersection Approach," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2452–2467, 2021.

[6] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 293–305, 2002.

[7] N. M. Correa, et al., "Canonical correlation analysis for data fusion and group inferences," *IEEE signal processing magazine*, vol. 27, no. 4, pp. 39–50, 2010.

[8] J. Sui, et al., "Three-way (N-way) fusion of brain imaging data based on mCCA+ jICA and its application to discriminating schizophrenia," *NeuroImage*, vol. 66, pp. 119–132, 2013.

[9] S. Qi, et al., "MicroRNA132 associated multimodal neuroimaging patterns in unmedicated major depressive disorder," *Brain*, vol. 141, no. 3, pp. 916–926, 2018.

[10] X. Fu, et al., "Scalable and flexible multiview MAX-VAR canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4150–4165, 2017.

[11] C. I. Kanatsoulis, et al., "Structured SUMCOR multiview canonical correlation analysis for large-scale data," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 306–319, 2018.

[12] A. Tenenhaus, et al., "Variable selection for generalized canonical correlation analysis," *Biostatistics*, vol. 15, no. 3, pp. 569–583, July 2014.

[13] A. Singh, et al., "DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays," *Bioinformatics*, vol. 35, no. 17, pp. 3055–3062, Sept. 2019.

[14] L. Gao, et al., "Discriminative multiple canonical correlation analysis for information fusion," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1951–1965, 2017.

[15] A. Benton, et al., "Deep generalized canonical correlation analysis," *arXiv preprint arXiv:1702.02519*, 2017.

[16] Y.-T. Lan, et al., "Multimodal emotion recognition using deep generalized canonical correlation analysis with an attention mechanism," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–6.

[17] J. Chen, et al., "Graph multiview canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 2826–2838, 2019.

[18] Y. Kaloga, et al., "Variational graph autoencoders for multiview canonical correlation analysis," *Signal Processing*, vol. 188, pp. 108182, 2021.

[19] S. Shrestha and X. Fu, "Communication-Efficient Federated Linear and Deep Generalized Canonical Correlation Analysis," *IEEE Transactions on Signal Processing*, vol. 71, pp. 1379–1394, 2023.

[20] Y. O. Li, et al., "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.

[21] A. de Cheveigné, et al., "Multiway canonical correlation analysis of brain data," *NeuroImage*, vol. 186, no. November 2018, pp. 728–740, 2019.

[22] R. Song, et al., "Remote Photoplethysmography With an EEMD-MCCA Method Robust Against Spatially Uneven Illuminations," *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13484–13494, 2021.

[23] P. J. Schreier, "A unifying discussion of correlation analysis for complex random vectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1327–1336, 2008.

[24] A. A. Nielsen, "Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data," *Environmental Research*, vol. 21, no. 6, pp. 555–567, 1995.

[25] N. A. Asendorf, *Informative Data Fusion: Beyond Canonical Correlation Analysis*, Ph.D. thesis, University of Michigan, 2015.

[26] A. Tenenhaus and M. Tenenhaus, "Regularized Generalized Canonical Correlation Analysis," *Psychometrika*, vol. 76, no. 2, pp. 257–284, Apr. 2011.

[27] M. Tenenhaus, et al., "Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods," *Psychometrika*, vol. 82, no. 3, pp. 737–777, Sept. 2017.

[28] J. Vía, et al., "Joint blind source separation from second-order statistics: Necessary and sufficient identifiability conditions," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 2520–2523, 2011.

[29] H. Richard, et al., "Shared Independent Component Analysis for Multi-Subject Neuroimaging," *Advances in Neural Information Processing Systems*, vol. 36, no. NeurIPS, pp. 29962–29971, 2021.

[30] P. Horst, *Generalized canonical correlations and their application to experimental data*, Number 14. Journal of clinical psychology, 1961.

[31] R. G. Steel, "Minimum generalized variance for a set of linear functions," *The Annals of Mathematical Statistics*, pp. 456–460, 1951.

[32] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, pp. 471–494, 1932.

[33] M. Anderson, et al., "Joint Blind Source Separation With Multivariate Gaussian Model: Algorithms and Performance Analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.

[34] M. Anderson, et al., "Independent Vector Analysis: Identification Conditions and Performance Bounds," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4399–4410, 2014.

[35] T. Hasija, et al., "Determining the dimension and structure of the subspace correlated across multiple data sets," *Signal Processing*, vol. 176, pp. 107613, 2020.

[36] B. Gabrielson, et al., "An Efficient Analytic Solution for Joint Blind Source Separation," *IEEE Transactions on Signal Processing*, vol. 72, pp. 2436–2449, 2024.

[37] A. Kessy, et al., "Optimal whitening and decorrelation," *The American Statistician*, vol. 72, no. 4, pp. 309–314, 2018.

[38] J. D. Carroll, "Generalization of canonical correlation analysis to three or more sets of variables," in *Proceedings of the 76th annual convention of the American Psychological Association*, 1968, pp. 227–228.

[39] T. Kim, et al., "Independent vector analysis: Definition and algorithms," in *Proc. 40th Asilomar Conference on Signals, Systems and Computers*, 2006, pp. 1393–1396.

[40] M. Anderson, et al., "Nonorthogonal independent vector analysis using multivariate Gaussian model," in *International Conference on Latent Variable Analysis and Signal Separation*. 2010, pp. 354–361, Springer.

[41] P. Virtanen, et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[42] A. Pezeshki, et al., "Empirical canonical correlation analysis in subspaces," in *2004 38th Asilomar Conference on Signals, Systems and Computers*, Nov. 2004, pp. 994–997.

[43] Y. Song, et al., "Canonical correlation analysis of high-dimensional data with very small sample support," *Signal Processing*, vol. 128, pp. 449–458, Nov. 2016.

[44] T. Hasija and T. Marrinan, "A GLRT for estimating the number of correlated components in sample-poor mCCA," in *2022 30th European Signal Processing Conference (EUSIPCO)*, Aug. 2022, pp. 2091–2095.

[45] Q. Long, et al., "Independent vector analysis for common subspace analysis: Application to multi-subject fMRI data yields meaningful subgroups of schizophrenia," *NeuroImage*, vol. 216, pp. 116872, 2020.

[46] J. H. Stapleton, *Linear statistical models*, John Wiley & Sons, 2009.