An efficient analytic solution for joint blind source separation

Ben Gabrielson, M. A. B. S. Akhonda, Isabell Lehmann, and Tülay Adali

***Abstract*—Joint blind source separation (JBSS) is a powerful methodology for analyzing multiple related datasets, able to jointly extract sources that describe statistical dependencies across the datasets. However, JBSS can be computationally prohibitive with high-dimensional data, thus there exists a key need for more efficient JBSS algorithms. JBSS algorithms typically rely on numerical solutions, which may be expensive due to their iterative nature. In contrast, analytic solutions follow consistent procedures that are often less expensive. In this paper, we introduce an efficient analytic solution for JBSS. Denoting a set of sources dependent across the datasets as a "source component vector" (SCV), our solution minimizes correlation among separate SCVs by minimizing distance of the SCV cross-covariance's eigenvector matrix from a block diagonal matrix. Under the orthogonality constraint, this leads to a system of linear equations wherein each subproblem has an analytic solution. We derive identifiability conditions of our solution's estimator, and demonstrate estimation performance and time efficiency in comparison with other JBSS algorithms that exploit source correlation across datasets. Results demonstrate that our solution achieves the lowest asymptotic computational complexity among JBSS algorithms, and is capable of superior estimation performance compared with algorithms of similar complexity.**

***Index Terms*—Joint Blind Source Separation, Independent Vector Analysis, Multiset Canonical Correlation Analysis.**

I. INTRODUCTION

Blind source separation (BSS) techniques have use over a wide range of applications [1]–[3], providing useful data-driven representations of latent structure within a single dataset. While single dataset applications are common, many other applications require the analysis of multiple datasets, in which these multiple datasets may be inherently related due to their similar or shared information. A classic example of this is with multi-subject functional magnetic resonance imaging (fMRI) data, where signals within each subject's dataset exhibit dependence across the datasets, typically representing similar functional network activity [4]–[8]. To represent this information within the analysis, joint blind source separation (JBSS) techniques generalize the capabilities of BSS by *jointly* analyzing the datasets, allowing JBSS to model and exploit any statistical dependencies existing across the datasets [4]–[6], [9]–[16]. This leads JBSS to achieve powerful estimators, potentially outperforming BSS not only by superior estimation performance, but also through the key benefit of uncovering the dependence relations across datasets [4]–[6], [9]. Applications of JBSS include analysis of fMRI [4]–[8], electroencephalography (EEG) [17], electrocardiography (ECG) [18], and remote sensing data [14], speech enhancement [19], molecular property prediction [20], and various others.

One general formulation of JBSS arrives from independent vector analysis (IVA), a multi-dataset extension of independent component analysis (ICA) for BSS [9]–[11]. IVA models each dataset as mixtures of latent sources, wherein each source is modeled as independent to all other sources within its own dataset *and* dependent to a single source within each of the other datasets. By formulating each set of dependent sources as a multivariate "source component vector" (SCV), IVA performs JBSS by maximizing independence among the separate SCVs, similar to how ICA maximizes independence among the univariate sources within a single dataset.

Among IVA algorithms, perhaps the most computationally practical algorithm models each SCV as generated from a multivariate Gaussian distribution (IVA-G) [10]. By modeling sources as Gaussian, statistical dependence between sources is completely described by second-order statistics (SOS), i.e. source correlation, and thus IVA-G operates by minimizing correlation among the separate SCVs. This results in several practical advantages that IVA-G enjoys compared with other IVA algorithms, such as lower computational complexity and lower data storage requirements, as IVA-G only needs to execute linear operations over covariance matrices to perform JBSS. In contrast, IVA algorithms that model SCV distributions as non-Gaussian make use of both second and higher-order statistics (HOS), resulting in IVA algorithms applying non-linear functions to the data with significantly higher computational expenses [11]. Furthermore, it is notable that regardless of the distributions of the underlying SCVs, algorithms exploiting only source correlation across datasets are fully capable of estimating the underlying SCVs, so long as the SCVs are uncorrelated (which also holds if the SCVs are statistically independent) [10], [21], [22]. These reasons make IVA-G a practical algorithm for JBSS in most scenarios.

Canonical correlation analysis (CCA) is the oldest method capable of JBSS, and is typically used to estimate correlated sources across two datasets [12]. CCA extended to multiple datasets, called multiset canonical correlation analysis (MCCA), has proven useful for obtaining correlated sources across multiple datasets [6], [13]–[15]. MCCA algorithms operate like IVA-G in that they exploit source correlation across datasets to achieve decompositions. MCCA is also fully capable of performing JBSS, likewise able to estimate underlying SCVs so long as the SCVs are uncorrelated [6].

This work is supported in part by the grants NSF 2316420, NIH R01MH118695, NIH R01MH123610, NIH R01AG073949, and by the German Research Foundation under Grant SCHR 1384/3-2. The computational hardware used is part of the UMBC High Performance Computing Facility (HPCF), supported by the U.S. NSF through the MRI and SCREMS programs (grants CNS-0821258, CNS-1228778, OAC-1726023, DMS-0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC).

Ben Gabrielson, M.A.B.S. Akhonda and Tülay Adali are with the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore MD. (bengabr1@umbc.edu).

Isabell Lehmann is with the Signal and System Theory Group at the Department of Electrical Engineering and IT at Paderborn University, Germany.

Despite the many strengths of JBSS, little work has been done to address its computational challenges. Particularly when data is high-dimensional, with large numbers of sources or datasets, computational complexity makes most JBSS algorithms infeasible. These challenges even extend to algorithms exploiting only source correlation across datasets, such as IVA-G and MCCA. One notable reason is that virtually all JBSS algorithms rely on numerical solutions, where an initial guess is iteratively refined until converging to a solution.

Thus in this paper, we propose an efficient *analytic* solution to JBSS derived from an eigendecomposition-based approach. Under assumption that the latent SCVs are uncorrelated, the SCV cross-covariance is a block diagonal matrix, and thus the SCV cross-covariance's eigenvectors also form a block diagonal matrix. We exploit this property via minimizing the distance of the SCV cross-covariance's eigenvectors from a block diagonal matrix, and show that each source's subproblem is equivalent to solving a system of linear equations from eigenvectors of the observed data's covariance matrix. This algorithm, which we call "analytic cross-correlation minimization" (ACCM), achieves the lowest asymptotic computational complexity of all JBSS algorithms. We demonstrate performance of ACCM on simulated data, showing ACCM is not only among the most efficient JBSS algorithms, but also provides generally superior separation performance compared to algorithms of similar complexities. We then demonstrate the performance of ACCM on fMRI hybrid data, and discuss the contribution of ACCM in the general JBSS setting.

The paper is organized as follows. Section II formulates the JBSS problem. Section III introduces a JBSS cost function measuring SCV cross-correlation via the "block diagonality" of the SCV cross-covariance matrix. Section IV proposes an alternative cost measuring "block diagonality" of the SCV cross-covariance's eigenvector matrix (assuming uncorrelated SCVs), from which we derive the minima of this cost given the eigenvectors of the observed data's covariance. Here we also introduce the analytic solution and derive the corresponding estimator's identifiability conditions. Section V discusses the theory of other JBSS methods and compares these to the proposed algorithm. Section VI demonstrates performance of the algorithm with respect to simulated data, and performance with respect fMRI sources within a hybrid experiment. Section VII discusses the contribution of the algorithm in the general JBSS setting. Section VIII concludes with takeaways on the proposed algorithm.

II. JBSS Problem Formulation

We first formulate the general JBSS problem where the data is modeled as a random vector. We have K datasets, each observed over T samples, where each dataset is modeled as a linear mixture of N independent sources. With $\mathbf{x}^{[k]}(t) = [x_1^{[k]}(t), \ldots, x_N^{[k]}(t)]^\top \in \mathbb{R}^N$ denoting the N observed signals within the kth dataset at sample index t, these signals are modeled as mixtures of N latent source signals $\mathbf{s}^{[k]}(t) = [s_1^{[k]}(t), \ldots, s_N^{[k]}(t)]^\top \in \mathbb{R}^N$, that are mixed by an unknown invertible matrix $\mathbf{A}^{[k]} \in \mathbb{R}^{N\times N}$ to produce the datasets $\mathbf{x}^{[k]}(t)$. Here, $(.)^\top$ denotes the transpose. The generative model is:

$$\mathbf{x}^{[k]}(t) = \mathbf{A}^{[k]}\ \mathbf{s}^{[k]}(t)\ , \quad t = 1, \ldots, T, \quad k = 1, \ldots, K \quad (1)$$

To estimate the underlying sources within the K datasets, JBSS algorithms estimate K demixing matrices $\mathbf{W}^{[k]} \in \mathbb{R}^{N\times N}$ that demix each dataset into their corresponding estimated sources $\mathbf{y}^{[k]}(t) = [y_1^{[k]}(t), \ldots, y_N^{[k]}(t)]^\top \in \mathbb{R}^N$, given by $\mathbf{y}^{[k]}(t) = \mathbf{W}^{[k]}\ \mathbf{x}^{[k]}(t)$. The nth row of demixing matrix $\mathbf{W}^{[k]}$ is given by $(\mathbf{w}_n^{[k]})^\top$, and is used to estimate the nth source within the kth dataset, given by $y_n^{[k]}(t) = (\mathbf{w}_n^{[k]})^\top\ \mathbf{x}^{[k]}(t)$.

Since the K datasets are observed over T samples, the observed datasets are represented by matrices $\mathbf{X}^{[k]} = [\mathbf{x}_1^{[k]}, \ldots, \mathbf{x}_N^{[k]}]^\top \in \mathbb{R}^{N\times T}$, and the generative model in (1) is given

K	number of datasets	(dataset index $k = 1, \ldots, K$)
N	number of SCVs	(SCV index $n = 1, \ldots, N$)
T	number of samples	(sample index $t = 1, \ldots, T$)
$\mathbf{x}^{[k]}$ / $\mathbf{X}^{[k]}$	kth dataset	
$\mathbf{s}^{[k]}$ / $\mathbf{S}^{[k]}$	kth dataset's sources	
$\mathbf{y}^{[k]}$ / $\mathbf{Y}^{[k]}$	kth dataset's estimated sources	
$x_n^{[k]}$ / $\mathbf{x}_n^{[k]}$	nth observed signal in kth dataset	
$s_n^{[k]}$ / $\mathbf{s}_n^{[k]}$	nth source in kth dataset	
$y_n^{[k]}$ / $\mathbf{y}_n^{[k]}$	nth estimated source in kth dataset	
$\mathbf{A}^{[k]}$	kth dataset's mixing matrix	
$\mathbf{W}^{[k]}$	kth dataset's estimated demixing matrix	
$(\mathbf{w}_n^{[k]})^\top$	nth demixing vector (row) in $\mathbf{W}^{[k]}$	
$\mathbf{s}_n$ / $\mathbf{S}_n$	nth SCV	
$\mathbf{y}_n$ / $\mathbf{Y}_n$	nth estimated SCV	
$\mathbf{s}$ / $\mathbf{S}$	vertical concatenation of the N SCVs	
$\mathbf{y}$ / $\mathbf{Y}$	vertical concatenation of the N estimated SCVs	
$\mathbf{x}$ / $\mathbf{X}$	vertical concatenation of the K datasets	
$\mathbf{P}_n$	nth SCV's demixing vectors $(\mathbf{w}_n^{[k]})^\top$ in a block matrix	
$\mathbf{P}$	vertical concatenation of all N SCV's $\mathbf{P}_n$	
$\mathbf{P}_{\text{opt}}$	optimal $\mathbf{P}$ matrix	
$(\mathbf{W}^{[k]})_{\text{opt}}$	optimal $\mathbf{W}^{[k]}$ demixing matrix	
$(\mathbf{w}_n^{[k]})_{\text{opt}}^\top$	optimal $(\mathbf{w}_n^{[k]})^\top$ demixing vector	
$\mathbf{C}_\mathbf{s}$	SCV cross-covariance matrix	
$\mathbf{C}_{\mathbf{s}_n}$	nth SCV covariance matrix	
$\hat{\mathbf{C}}_\mathbf{y}$	estimated SCV cross-covariance matrix	
$\hat{\mathbf{C}}_{\mathbf{y}_n}$	nth estimated SCV covariance matrix	
$\mathbf{C}_\mathbf{x}$ / $\hat{\mathbf{C}}_\mathbf{x}$	covariance / sample covariance matrix of $\mathbf{x}$	
$\mathbf{V}_\mathbf{s}$ / $\mathbf{D}_\mathbf{s}$	eigenvectors / eigenvalues of $\mathbf{C}_\mathbf{s}$	
$\mathbf{V}_{\mathbf{s}_n}$ / $\mathbf{D}_{\mathbf{s}_n}$	eigenvectors / eigenvalues of $\mathbf{C}_{\mathbf{s}_n}$	
$\hat{\mathbf{V}}_\mathbf{y}$	eigenvectors of $\hat{\mathbf{C}}_\mathbf{y}$	
$\mathbf{V}_\mathbf{x}$ / $\hat{\mathbf{V}}_\mathbf{x}$	eigenvectors of $\mathbf{C}_\mathbf{x}$ / $\hat{\mathbf{C}}_\mathbf{x}$	
$(\hat{\mathbf{v}}_\mathbf{y})_n^{[k]\top}$	row in $\hat{\mathbf{V}}_\mathbf{y}$ corresponding to $(\mathbf{w}_n^{[k]})^\top$	
$(\hat{\mathbf{V}}_\mathbf{x}^{[k]})_n$	submatrix in $\hat{\mathbf{V}}_\mathbf{x}$ corresponding to $(\mathbf{w}_n^{[k]})^\top$	
$\mathbf{V}_\mathbf{x}^{[k]}$	N rows of $\mathbf{V}_\mathbf{x}$ corresponding to kth dataset	
$\mathbf{V}_\mathbf{s}^{[k]}$	N rows of $\mathbf{V}_\mathbf{s}$ corresponding to the kth dataset	
$(\mathbf{v}_{\mathbf{s}_n}^{[k]})^\top$	kth row of $\mathbf{V}_{\mathbf{s}_n}$	
$\mathbf{G}^{[k]}$	pairwise similarity of columns in $\mathbf{V}_\mathbf{x}^{[k]}$	
$\mathbf{G}$	eigenvector similarity matrix	

Table 1. Notations used in this paper (Sections II — IV). Vectors are given as column vectors, e.g. $\mathbf{w}_n^{[k]}$ is a column vector, and $(\mathbf{w}_n^{[k]})^\top$ is a row vector, with $^\top$ denoting the transpose. Datasets and sources are represented as a random vector (e.g. $\mathbf{x}^{[k]} \in \mathbb{R}^N$), or by T observed samples of a random vector (e.g. $\mathbf{X}^{[k]} \in \mathbb{R}^{N\times T}$).

as $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\ \mathbf{S}^{[k]}$, with latent sources given by $\mathbf{S}^{[k]} = [\mathbf{s}_1^{[k]}, \ldots, \mathbf{s}_N^{[k]}]^\top \in \mathbb{R}^{N\times T}$, and estimated sources given by $\mathbf{Y}^{[k]} = \mathbf{W}^{[k]}\ \mathbf{X}^{[k]} = [\mathbf{y}_1^{[k]}, \ldots, \mathbf{y}_N^{[k]}]^\top \in \mathbb{R}^{N\times T}$.

JBSS is distinguished from BSS in that JBSS exploits dependence across the K datasets to leverage cross-information and improve overall estimation performance. JBSS formulations assume that sources of the same index n are dependent across the K datasets, thus forming N sets of K sources. In IVA terminology, each of these sets is referred to as a "source component vector" (SCV), and for simplicity we refer to these sets as SCVs for the other JBSS algorithms as well. The nth SCV is given as $\mathbf{s}_n = [s_n^{[1]}, \ldots, s_n^{[K]}]^\top \in \mathbb{R}^K$, which is estimated by $\mathbf{y}_n = [y_n^{[1]}, \ldots, y_n^{[K]}]^\top \in \mathbb{R}^K$ (for simplicity of notation, we drop sample index t when describing random variables or random vectors). Over T samples of the data, the nth SCV is represented by the matrix $\mathbf{S}_n = [\mathbf{s}_n^{[1]}, \ldots, \mathbf{s}_n^{[K]}]^\top \in \mathbb{R}^{K\times T}$, which is estimated by $\mathbf{Y}_n = [\mathbf{y}_n^{[1]}, \ldots, \mathbf{y}_n^{[K]}]^\top \in \mathbb{R}^{K\times T}$. Typically each SCV is modeled as independent (and uncorrelated) from all other SCVs, and thus any two sources across the datasets are modeled as dependent only if they correspond to the same index n (nth SCV).

JBSS algorithms can only identify demixing matrix vectors $(\mathbf{w}_n^{[k]})^\top$ and estimated sources $y_n^{[k]}$ up to scaling and permutation ambiguity within each dataset. JBSS additionally orders sources within each dataset to align with the order of SCVs, such that the nth source within a dataset corresponds to the nth SCV. In practice, the scaling ambiguity can be removed by requiring each estimated source to have unit variance.

Furthermore, most implementations of JBSS involve standardizing and pre-whitening each dataset prior to estimation. The advantage is that if sources are standardized and pre-whitened within each dataset, then the residual mixing matrices $\mathbf{A}^{[k]}$ become orthogonal, or asymptotically orthogonal for the observed datasets $\mathbf{X}^{[k]}$ as $T \to \infty$ (since $\mathrm{E}\left\{\mathbf{s}^{[k]}\ \mathbf{s}^{[k]\top}\right\} = \lim_{T\to\infty} \frac{1}{T-1}\ \mathbf{S}^{[k]}\ \mathbf{S}^{[k]\top} = \mathbf{I}$, thus $\mathrm{E}\left\{\mathbf{x}^{[k]}\ \mathbf{x}^{[k]\top}\right\} = \lim_{T\to\infty} \frac{1}{T-1}\ \mathbf{X}^{[k]}\ \mathbf{X}^{[k]\top} = \mathbf{A}^{[k]}\ \mathbf{A}^{[k]\top} = \mathbf{I}$), in which case orthogonality constraints can be imposed on $\mathbf{W}^{[k]}$ to considerably simplify the problem [23]. For the remainder of the paper, we assume that sources and datasets are standardized, and that datasets have been pre-whitened prior to JBSS.

III. Joint Covariance Block Diagonalization Cost

Before introducing the proposed cost, we first define quantities with respect to the N underlying SCVs $\mathbf{s}_n \in \mathbb{R}^K$. As each SCV is a set of standardized sources, each SCV thus has mean $\mathrm{E}\{\mathbf{s}_n\} = \mathbf{0} \in \mathbb{R}^K$ and some covariance matrix (which is also a correlation matrix) $\mathbf{C}_{\mathbf{s}_n} \triangleq \mathrm{E}\left\{\mathbf{s}_n\ \mathbf{s}_n^\top\right\} \in \mathbb{R}^{K\times K}$.

Correlation among all N SCVs is described by the "SCV cross-covariance" matrix $\mathbf{C}_\mathbf{s} = \triangleq \mathrm{E}\left\{\mathbf{s}\ \mathbf{s}^\top\right\} \in \mathbb{R}^{NK\times NK}$, where $\mathbf{s} \triangleq [(\mathbf{s}_1)^\top, \ldots, (\mathbf{s}_N)^\top]^\top \in \mathbb{R}^{NK}$ is the vertical concatenation of all N SCVs. These N SCVs are uncorrelated if and only if the pairwise cross-covariance (cross-correlation matrix) between any two SCVs is a matrix of zeros. If this holds, $\mathbf{C}_\mathbf{s}$ is equal to the direct sum of all N SCV covariances: $\mathbf{C}_\mathbf{s} = \bigoplus_{n=1}^N \mathbf{C}_{\mathbf{s}_n} \in \mathbb{R}^{NK\times NK}$, a block diagonal matrix with N blocks given by the N SCV covariances $\mathbf{C}_{\mathbf{s}_n}$. Fig. 1 (a) gives an example of $\mathbf{C}_\mathbf{s}$ for $N = 4$ uncorrelated SCVs.

Concerning the estimated SCVs, $\mathbf{y}_n$ (mean $\mathbf{0} \in \mathbb{R}^K$, covariance $\mathbf{C}_{\mathbf{y}_n} \in \mathbb{R}^{K\times K}$), we can similarly define $\mathbf{C}_\mathbf{y} \triangleq \mathrm{E}\left\{\mathbf{y}\ \mathbf{y}^\top\right\} \in \mathbb{R}^{NK\times NK}$, with $\mathbf{y} \triangleq [(\mathbf{y}_1)^\top, \ldots, (\mathbf{y}_N)^\top]^\top \in \mathbb{R}^{NK}$. Over T samples of data such that $\mathbf{y}_n$ is represented by matrix $\mathbf{Y}_n \in \mathbb{R}^{K\times T}$, each $\mathbf{C}_{\mathbf{y}_n}$ is itself estimated in practice by "estimated SCV covariance" $\hat{\mathbf{C}}_{\mathbf{y}_n} = \frac{1}{T-1}\ \mathbf{Y}_n\ \mathbf{Y}_n^\top \in \mathbb{R}^{K\times K}$, and cross-covariance $\mathbf{C}_\mathbf{y}$ is itself estimated by the "estimated SCV cross-covariance" $\hat{\mathbf{C}}_\mathbf{y} = \frac{1}{T-1}\ \mathbf{Y}\ \mathbf{Y}^\top \in \mathbb{R}^{NK\times NK}$, with $\mathbf{Y} = [(\mathbf{Y}_1)^\top, \ldots, (\mathbf{Y}_N)^\top]^\top \in \mathbb{R}^{NK\times T}$.

From estimated SCV cross-covariance $\hat{\mathbf{C}}_\mathbf{y}$, we may measure the degree of SCV cross-correlation by the distance of $\hat{\mathbf{C}}_\mathbf{y}$ from a block diagonal matrix. This is measured by the squared Frobenius norm of "off-blocks" (pairwise SCV cross-covariances) within $\hat{\mathbf{C}}_\mathbf{y}$, which we denote by the "Joint Covariance Block Diagonalization" (JCBD) cost:

$$\mathcal{J}_{\mathrm{JCBD}}(\mathbf{Y}) = \left\| \hat{\mathbf{C}}_\mathbf{y} - \bigoplus_{n=1}^N \hat{\mathbf{C}}_{\mathbf{y}_n} \right\|_\mathrm{F}^2 \tag{2}$$

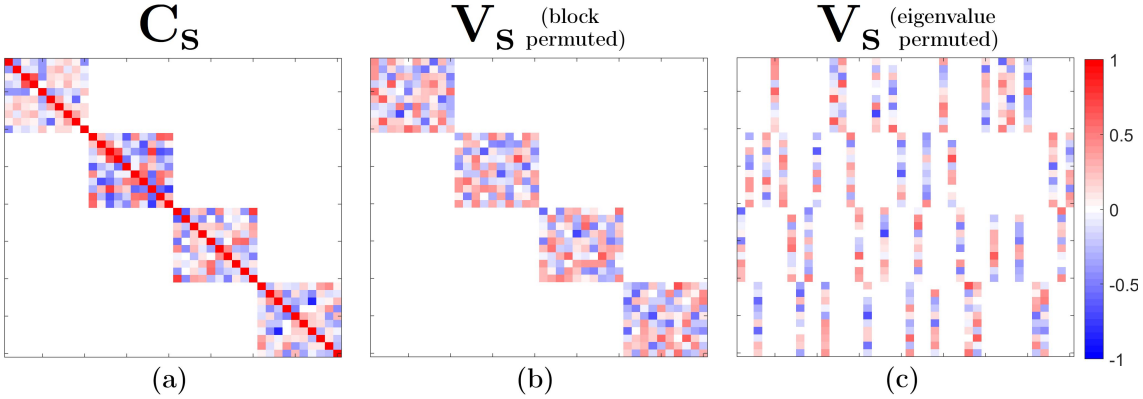

Fig. 1. SCV cross-covariance $\mathbf{C}_\mathbf{s} \in \mathbb{R}^{NK\times NK}$ of N=4 uncorrelated SCVs, and corresponding eigenvectors $\mathbf{V}_\mathbf{s}$.
(a) SCV cross-covariance $\mathbf{C}_\mathbf{s}$. Blocks on the diagonal are SCV covariances $\mathbf{C}_{\mathbf{s}_n} = \mathrm{E}\left\{\mathbf{s}_n\ \mathbf{s}_n^\top\right\} = \in \mathbb{R}^{K\times K}$. Off-blocks are pairwise cross-covariances ($\mathrm{E}\left\{\mathbf{s}_m\ \mathbf{s}_n^\top\right\} \in \mathbb{R}^{K\times K}$ for $m \neq n$), which equal $\mathbf{0} \in \mathbb{R}^{K\times K}$ if SCVs are uncorrelated.
(b) SCV cross-covariance eigenvectors $\mathbf{V}_\mathbf{s}$, with eigenvectors (columns of $\mathbf{V}_\mathbf{s}$) permuted according to SCV groupings ("block permuted"). The nth diagonal block of $\mathbf{V}_\mathbf{s}$ contains all K eigenvectors of the nth SCV covariance $\mathbf{C}_{\mathbf{s}_n}$.
(c) $\mathbf{V}_\mathbf{s}$, with eigenvectors permuted according to corresponding eigenvalue size (smallest to largest eigenvalues, left to right).

We now consider how to formulate (2) with respect to demixing vectors $\mathbf{w}_n^{[k]}$. This is done by first noting that the nth estimated SCV $\mathbf{y}_n$ is obtainable by $\mathbf{y}_n = \mathbf{P}_n\ \mathbf{x}$, where we define $\mathbf{x} \triangleq [(\mathbf{x}^{[1]})^\top, \ldots, (\mathbf{x}^{[K]})^\top]^\top \in \mathbb{R}^{NK}$ as the vertical concatenation of the K datasets, represented over T samples by $\mathbf{X} = [(\mathbf{X}^{[1]})^\top, \ldots, (\mathbf{X}^{[K]})^\top]^\top \in \mathbb{R}^{NK\times T}$, and $\mathbf{P}_n \triangleq [\mathbf{p}_n^{[1]}, \mathbf{p}_n^{[2]}, \ldots, \mathbf{p}_n^{[K]}]^\top \in \mathbb{R}^{K\times NK}$ is defined such that $\mathbf{p}_n^{[k]} \triangleq \mathbf{e}_k \bigotimes \mathbf{w}_n^{[k]} \in \mathbb{R}^{NK}$ is the Kronecker product of $\mathbf{w}_n^{[k]}$ with the kth standard basis vector $\mathbf{e}_k$. By further defining $\mathbf{P} \triangleq [\mathbf{P}_1^\top, \mathbf{P}_2^\top, \ldots, \mathbf{P}_N^\top]^\top \in \mathbb{R}^{NK\times NK}$, we can obtain all N estimated SCVs $\mathbf{y}$, and estimated SCV cross-covariance $\hat{\mathbf{C}}_\mathbf{y}$:

$$\mathbf{y}_n = \mathbf{P}_n\ \mathbf{x}\ , \quad \mathbf{y} = \mathbf{P}\ \mathbf{x}\ , \quad \hat{\mathbf{C}}_\mathbf{y} = \mathbf{P}\ \hat{\mathbf{C}}_\mathbf{x}\ \mathbf{P}^\top \tag{3}$$

where $\hat{\mathbf{C}}_\mathbf{x} = \frac{1}{T-1}\ \mathbf{X}\ \mathbf{X}^\top \in \mathbb{R}^{NK\times NK}$ is the sample estimate of $\mathbf{C}_\mathbf{x} = \triangleq \mathrm{E}\left\{\mathbf{x}\ \mathbf{x}^\top\right\}$, and $\hat{\mathbf{C}}_\mathbf{x}^{[k,j]} = \frac{1}{T-1}\ \mathbf{X}^{[k]}\ \mathbf{X}^{[j]\top} \in \mathbb{R}^{N\times N}$.

Similarly, we obtain the nth estimated SCV's covariance by $\hat{\mathbf{C}}_{\mathbf{y}_n} = \mathbf{P}_n\ \hat{\mathbf{C}}_\mathbf{x}\ \mathbf{P}_n^\top$. Thus subject to $\left\|(\mathbf{w}_n^{[k]})^\top\right\|_2 = 1$, we may write (2) with respect to all K demixing matrices $\mathbf{W}^{[k]}$ via matrix $\mathbf{P}$, or with respect to each demixing vector $(\mathbf{w}_n^{[k]})^\top$:

$$\mathcal{J}_{\text{JCBD}}\,(\mathbf{P}) = \left\|\mathbf{P}\ \hat{\mathbf{C}}_\mathbf{x}\ \mathbf{P}^\top - \bigoplus_{n=1}^{N} \mathbf{P}_n\ \hat{\mathbf{C}}_\mathbf{x}\ \mathbf{P}_n^\top\right\|_\mathrm{F}^2 \tag{4}$$

$$\mathcal{J}_{\text{JCBD}}\,(\ (\mathbf{w}_n^{[k]})^\top\ ) = \sum_{j=1}^{K}\sum_{\substack{i=1\\ i\neq n}}^{N} \left\|\ (\mathbf{w}_n^{[k]})^\top \hat{\mathbf{C}}_\mathbf{x}^{[k,j]} \mathbf{w}_i^{[j]}\ \right\|_\mathrm{F}^2 \tag{5}$$

The cost in (4) is associated with algorithms that perform "generalized joint diagonalization" [24]–[28], whose solutions require iterative procedures and cannot be solved analytically (since the cost for $(\mathbf{w}_n^{[k]})^\top$ depends on realizations of other demixing vectors $\mathbf{w}_i^{[j]}$). However in the next section, we propose a surrogate cost to (4) that can be solved analytically, by exploiting the assumption that $\mathbf{C}_\mathbf{s}$ is block diagonal.

IV. An Eigenvector-based JBSS cost

A. JBSS by eigendecomposition of the data

A key property of $\mathbf{C}_\mathbf{s}$ being a block diagonal matrix is that its eigendecomposition is also block diagonal. With $\mathbf{C}_\mathbf{s} = \bigoplus_{n=1}^{N} \mathbf{C}_{\mathbf{s}_n}$, the eigendecomposition of $\mathbf{C}_{\mathbf{s}_n}$ is given by $\mathbf{C}_{\mathbf{s}_n} = \mathbf{V}_{\mathbf{s}_n}\ \mathbf{D}_{\mathbf{s}_n}\ \mathbf{V}_{\mathbf{s}_n}^\top$, where $\mathbf{V}_{\mathbf{s}_n} \in \mathbb{R}^{K\times K}$ is the matrix with K columns containing the K eigenvectors of $\mathbf{C}_{\mathbf{s}_n}$, and $\mathbf{D}_{\mathbf{s}_n} \in \mathbb{R}^{K\times K}$ is a diagonal matrix containing the K eigenvalues of $\mathbf{C}_{\mathbf{s}_n}$. Thus, the eigendecomposition of $\mathbf{C}_\mathbf{s}$ is given by $\mathbf{C}_\mathbf{s} = \mathbf{V}_\mathbf{s}\ \mathbf{D}_\mathbf{s}\ \mathbf{V}_\mathbf{s}^\top$, where $\mathbf{V}_\mathbf{s} = \bigoplus_{n=1}^{N} \mathbf{V}_{\mathbf{s}_n} \in \mathbb{R}^{NK\times NK}$ are the NK block diagonal eigenvectors of $\mathbf{C}_\mathbf{s}$ (representing the K eigenvectors for each of the N SCV's $\mathbf{C}_{\mathbf{s}_n}$), and $\mathbf{D}_\mathbf{s} = \bigoplus_{n=1}^{N} \mathbf{D}_{\mathbf{s}_n} \in \mathbb{R}^{NK\times NK}$ is a diagonal matrix containing the NK eigenvalues. Fig. 1 (b) shows eigenvectors $\mathbf{V}_\mathbf{s}$ of an example SCV cross-covariance $\mathbf{C}_\mathbf{s}$, with eigenvectors (columns) permuted to reveal this block diagonal structure.

Another important consequence is the direct relationship between the eigendecompositions of SCV cross-covariance $\mathbf{C}_\mathbf{s}$ and data covariance $\mathbf{C}_\mathbf{x}$. As stated earlier, if the SCVs are uncorrelated, it follows from pre-whitening the datasets that the mixing matrices $\mathbf{A}^{[k]}$ are orthogonal, and thus the optimal $\mathbf{W}^{[k]}$ are also orthogonal. With all optimal $\mathbf{W}^{[k]}$ orthogonal, the optimal $\mathbf{P}$ matrix is also orthogonal, which we denote by $\mathbf{P}_\text{opt}$. This is significant concerning the noted eigendecomposition relations, because with $\mathbf{C}_\mathbf{s} = \mathbf{P}_\text{opt}\ \ \mathbf{C}_\mathbf{x}\ \ \mathbf{P}_\text{opt}^\top$, we have that $\mathbf{C}_\mathbf{s}$ is equal to $\mathbf{C}_\mathbf{x}$ under conjugation by orthogonal matrix $\mathbf{P}_\text{opt}$. Therefore, the eigenvalues of $\mathbf{C}_\mathbf{s}$ and $\mathbf{C}_\mathbf{x}$ are the same, and the eigenvectors of $\mathbf{C}_\mathbf{s}$ are obtained by rotating each eigenvector of $\mathbf{C}_\mathbf{x}$ by $\mathbf{P}_\text{opt}$. With $\mathbf{V}_\mathbf{x} \in \mathbb{R}^{NK\times NK}$ denoting the eigenvectors of $\mathbf{C}_\mathbf{x}$ and $\hat{\mathbf{V}}_\mathbf{x}$ the respective eigenvectors of $\hat{\mathbf{C}}_\mathbf{x}$, this relationship is given by:

$$\mathbf{P}_\text{opt}\ \mathbf{V}_\mathbf{x} = \mathbf{V}_\mathbf{s} \tag{6}$$

In regards to estimating parameter $\mathbf{P}$, the implications are that eigenvectors of the estimated SCV cross-covariance $\hat{\mathbf{V}}_\mathbf{y} \in \mathbb{R}^{NK\times NK}$ can be directly obtained from $\mathbf{P}\ \hat{\mathbf{V}}_\mathbf{x} = \hat{\mathbf{V}}_\mathbf{y}$ provided that demixing matrices $\mathbf{W}^{[k]}$ are constrained to be orthogonal.

This motivates a surrogate cost to (4) where instead of measuring distance of $\mathbf{C}_\mathbf{y}$ from a block diagonal matrix, we introduce a measure of "block diagonality" over the corresponding eigenvectors $\hat{\mathbf{V}}_\mathbf{y}$. In terms of parameter $\mathbf{P}$, this is given by the distance between $\mathbf{P}\ \hat{\mathbf{V}}_\mathbf{x}$ and its nearest block diagonal matrix $\mathcal{B}(\ \mathbf{P}\ \hat{\mathbf{V}}_\mathbf{x}\ )_{(N,K)}$, which we denote by the "Joint Eigenvector Block Diagonalization" (JEBD) cost:

$$\mathcal{J}_{\text{JEBD}}\,(\mathbf{P}) = \left\|\ \mathbf{P}\ \hat{\mathbf{V}}_\mathbf{x}\ -\ \mathcal{B}(\ \mathbf{P}\ \hat{\mathbf{V}}_\mathbf{x}\ )_{(N,K)}\ \right\|_\mathrm{F}^2 \tag{7}$$

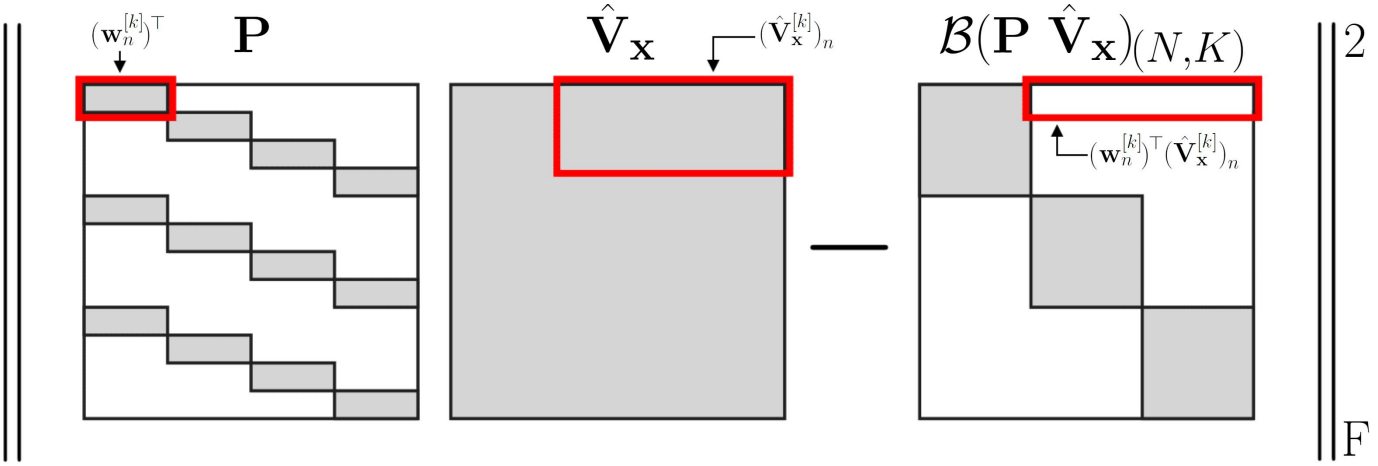

Fig. 2. Overview of the linear system of equations given in (7). Highlighted in red are the relevant quantities used to solve the subproblem for $\mathbf{w}_n^{[k]}$ given in (8), where in this example subproblem $n = 1$ and $k = 1$.

where we define $\mathcal{B}(\mathbf{V})_{(N,K)}$ as the operation that maps eigenvector matrix $\mathbf{V} \in \mathbb{R}^{NK \times NK}$ to a block diagonal matrix of N blocks each of size $K \times K$, by setting values in the "off-blocks" of $\mathbf{V}$ to zero. This is analogous to (4), where we have that $\mathcal{B}(\ \mathbf{P}\ \hat{\mathbf{C}}_{\mathbf{x}}\ \mathbf{P}^\top\ )_{(N,K)} = \bigoplus_{n=1}^{N} \mathbf{P}_n\ \hat{\mathbf{C}}_{\mathbf{x}}\ \mathbf{P}_n^\top$.

To meaningfully minimize (7), we first assume that eigenvectors (columns) in $\hat{\mathbf{V}}_{\mathbf{x}}$ are permuted so that corresponding eigenvectors (columns) in $\mathbf{V}_{\mathbf{s}}$ form a block diagonal matrix. In other words, eigenvectors are permuted such that the nth set of K eigenvectors correspond to the same SCV in $\mathbf{V}_{\mathbf{s}}$ (eigenvector columns $(n-1)K$ to nK correspond to the nth SCV). Eigendecompositon routines do not naturally do this sorting, instead they typically sort eigenvectors by size of their eigenvalues, for example as in Fig. 1 (c). We later explain in Section IV.C how one may sort $\hat{\mathbf{V}}_{\mathbf{x}}$ eigenvectors into SCV groupings, but for now we assume they are already permuted for the purpose of explaining the following solution.

With columns correctly permuted, minimizing (7) is simplified by solving for each demixing vector $(\mathbf{w}_n^{[k]})^\top$ individually. Defining $(\hat{\mathbf{v}}_{\mathbf{y}})_n^{[k]\top}$ as the row of $\hat{\mathbf{V}}_{\mathbf{y}}$ corresponding to the same index row $\mathbf{p}_n^{[k]\top}$ in $\mathbf{P}$, we have from $\mathbf{p}_n^{[k]\top}\ \hat{\mathbf{V}}_{\mathbf{x}} = (\hat{\mathbf{v}}_{\mathbf{y}})_n^{[k]\top}$ that demixing vector $(\mathbf{w}_n^{[k]})^\top$ associated with $\mathbf{p}_n^{[k]\top}$ is estimated to maximize block diagonality of $\hat{\mathbf{V}}_{\mathbf{y}}$ with respect to $(\hat{\mathbf{v}}_{\mathbf{y}})_n^{[k]\top}$. This entails minimizing the norm of $NK-K$ values in $(\hat{\mathbf{v}}_{\mathbf{y}})_n^{[k]\top}$ corresponding to the $NK-K$ eigenvectors of the $N-1$ other SCVs. These values are given by vector $(\mathbf{w}_n^{[k]})^\top(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n \in \mathbb{R}^{NK-K}$, where $(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n \in \mathbb{R}^{N \times NK-K}$ is a submatrix of $\hat{\mathbf{V}}_{\mathbf{x}}$ formed from the N rows corresponding to dataset k, and the $NK-K$ columns corresponding to the $NK-K$ eigenvectors of the $N-1$ other SCVs. Fig. 2 illustrates in red how $(\mathbf{w}_n^{[k]})^\top(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n$ is manifested in (7). Thus, the cost (7) can be defined for each individual $(\mathbf{w}_n^{[k]})^\top$:

$$\mathcal{J}_{\text{JEBD}}\left(\ (\mathbf{w}_n^{[k]})^\top\ \right) = \left\| (\mathbf{w}_n^{[k]})^\top(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n \right\|_{\text{F}}^2 \tag{8}$$

Unlike the cost in (5), the cost in (8) does not depend on realizations of the other demixing vectors, which is why minimizing this cost admits an analytic solution.

The solution to (8) is given as the minimizer of the quadratic form $(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n\ (\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n^\top \in \mathbb{R}^{N \times N}$, under the constraint that $\left\|(\mathbf{w}_n^{[k]})^\top\right\|_2 = 1$. Therefore, $(\mathbf{w}_n^{[k]})^\top$ is directly estimated by the eigenvector of $(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n\ (\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n^\top$ corresponding to its smallest eigenvalue. This procedure is repeated for each of the NK different $(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n$ to estimate all NK demixing vectors $(\mathbf{w}_n^{[k]})^\top$ across all K demixing matrices $\mathbf{W}^{[k]}$. Since this is an analytic solution to JBSS by minimizing SCV cross-correlation, we refer to the procedure as "analytic cross-correlation minimization" for JBSS (ACCM).

In the next section, we discuss the necessary and sufficient conditions on the data's generative model for which ACCM is able to uniquely identify the true sources via $(\mathbf{w}_n^{[k]})^\top$ subject to scale and permutation ambiguity, which we refer to as the identifiability conditions of the ACCM estimator.

B. Identifiability conditions of the ACCM estimator

The following theorem and proof makes use of notations and relations introduced at the beginning of section IV.A.

Theorem 1 (ACCM identifiability conditions): We assume that all datasets have been pre-whitened, that the true statistics of the data are known ($\hat{\mathbf{C}}_{\mathbf{x}} = \mathbf{C}_{\mathbf{x}}$ and $\hat{\mathbf{V}}_{\mathbf{x}} = \mathbf{V}_{\mathbf{x}}$), and that the latent SCVs are uncorrelated ($\text{E}\left\{\mathbf{s}_m\ \mathbf{s}_n^\top\right\} = \mathbf{0}$ for $m \neq n$). Then $\mathbf{W}^{[k]} = (\mathbf{A}^{[k]})^\top$ for all $k = 1, \ldots, K$, subject to scale and permutation ambiguity of the $(\mathbf{w}_n^{[k]})^\top$, if and only if for all $n = 1, \ldots, N$, there exists at least one eigenvalue within $\mathbf{D}_{\mathbf{s}_n}$ that is unique in $\mathbf{D}_{\mathbf{s}}$.

Proof: we denote $(\mathbf{W}^{[k]})_{\text{opt}}$ as the optimal orthogonal matrices $\mathbf{W}^{[k]}$ first mentioned in Section IV.A. Without loss of generality, $(\mathbf{W}^{[k]})_{\text{opt}} = (\mathbf{A}^{[k]})^\top$ subject to scale and permutation ambiguity. Our goal is thus to prove that $\mathbf{W}^{[k]} = (\mathbf{W}^{[k]})_{\text{opt}}$ subject to scale and permutation ambiguity.

We first consider the case where all eigenvalues in $\mathbf{D}_{\mathbf{s}}$ are unique, and thus all corresponding eigenvectors (columns within $\mathbf{V}_{\mathbf{x}}$ and $\mathbf{V}_{\mathbf{s}}$) are uniquely determined. Later in the proof we expand on the case when eigenvalues are not unique.

We refer to (6): $\mathbf{P}_{\text{opt}}\ \mathbf{V}_{\mathbf{x}} = \mathbf{V}_{\mathbf{s}}$, and assume that eigenvectors in $\mathbf{V}_{\mathbf{s}}$ (and corresponding eigenvectors in $\mathbf{V}_{\mathbf{x}}$) are "block permuted" as explained in section IV.A. To observe how $(\mathbf{W}^{[k]})_{\text{opt}}$ manifests in (6), we can isolate the equations in (6) that pertain only to the kth dataset:

$$(\mathbf{W}^{[k]})_{\text{opt}}\ \mathbf{V}_{\mathbf{x}}^{[k]} = \mathbf{V}_{\mathbf{s}}^{[k]} \tag{9}$$

where we define $\mathbf{V}_{\mathbf{x}}^{[k]} \in \mathbb{R}^{N \times NK}$ as the N rows of $\mathbf{V}_{\mathbf{x}}$ corresponding to the N mixtures in the kth dataset $\mathbf{x}^{[k]}$, and $\mathbf{V}_{\mathbf{s}}^{[k]} \in \mathbb{R}^{N \times NK}$ likewise as the N rows of $\mathbf{V}_{\mathbf{s}}$ corresponding to the N sources in the kth dataset $\mathbf{s}^{[k]}$. Fig. 3 illustrates this "k-centric" formulation given in (9).

The first item to note within this k-centric formulation is how the block diagonality of $\mathbf{V}_{\mathbf{s}}$ extends to $\mathbf{V}_{\mathbf{s}}^{[k]}$. Given that SCVs are uncorrelated, each source's row in $\mathbf{V}_{\mathbf{s}}^{[k]}$ has $NK-K$ values equal to 0, corresponding to the $NK-K$ eigenvectors of the $N-1$ other SCVs. $\mathbf{V}_{\mathbf{s}}^{[k]}$ can formally be written as $\mathbf{V}_{\mathbf{s}}^{[k]} = \bigoplus_{n=1}^{N} (\mathbf{v}_{\mathbf{s}_n}^{[k]})^\top$, where $(\mathbf{v}_{\mathbf{s}_n}^{[k]})^\top \in \mathbb{R}^K$ is the kth row of the eigenvector matrix $\mathbf{V}_{\mathbf{s}_n}$ corresponding to SCV covariance $\mathbf{C}_{\mathbf{s}_n}$. This reveals that all NK columns of $\mathbf{V}_{\mathbf{s}}^{[k]}$ can be represented by some scaled standard basis vector $\mathbf{e}_n$, with the nth set of K columns (of the nth SCV) corresponding to different scalings of the nth standard basis vector.

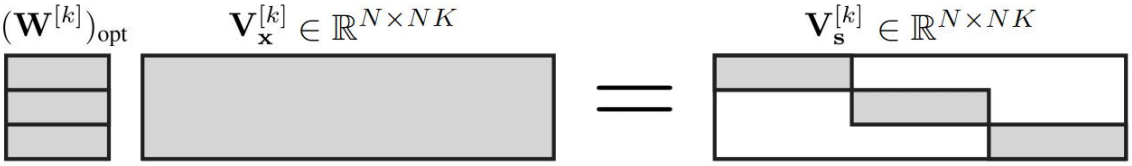

Fig. 3. k-centric formulation (9). Columns of $\mathbf{V}_{\mathbf{s}}^{[k]}$ are scaled standard basis vectors, rotated by $(\mathbf{W}^{[k]})_{\text{opt}}^\top$ to produce $\mathbf{V}_{\mathbf{x}}^{[k]}$.

Now given pre-whitened data resulting in $(\mathbf{W}^{[k]})_{\text{opt}}$ as orthogonal, each column of $\mathbf{V}_{\mathbf{x}}^{[k]}$ is equal to its respective column in $\mathbf{V}_{\mathbf{s}}^{[k]}$ rotated by orthogonal matrix $(\mathbf{W}^{[k]})_{\text{opt}}^{\top}$. With the nth set of K columns in $\mathbf{V}_{\mathbf{s}}^{[k]}$ corresponding to the nth standard basis vector $\mathbf{e}_n$, this means that the nth set of K columns in $\mathbf{V}_{\mathbf{x}}^{[k]}$ are equal to scaled versions of the nth row in $(\mathbf{W}^{[k]})_{\text{opt}}$, denoted by $(\mathbf{w}_n^{[k]})_{\text{opt}}^{\top}$, since $(\mathbf{W}^{[k]})_{\text{opt}}^{\top}\ \mathbf{e}_n = (\mathbf{w}_n^{[k]})_{\text{opt}}$. Therefore, columns of $(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n$ are composed of scaled versions of all $(\mathbf{w}_i^{[k]})_{\text{opt}}$ for $i \neq n$, thus $(\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n\ (\hat{\mathbf{V}}_{\mathbf{x}}^{[k]})_n^{\top}$ is a rank $N-1$ matrix with $(\mathbf{w}_n^{[k]})_{\text{opt}}^{\top}$ as the eigenvector corresponding to the smallest eigenvalue (eigenvalue 0).

This proves that given the previous assumptions, and given that all eigenvalues in $\mathbf{D}_{\mathbf{s}}$ are unique, then $\mathbf{W}^{[k]} = (\mathbf{W}^{[k]})_{\text{opt}}$ subject to scale and permutation ambiguity.

We now consider the case where some of the eigenvalues in $\mathbf{D}_{\mathbf{s}}$ are not unique. Uniqueness of estimating $\mathbf{V}_{\mathbf{s}}$ (and by extension $\mathbf{V}_{\mathbf{x}}$) is ultimately dependent on whether the eigenvalues of $\mathbf{C}_{\mathbf{s}}$ are unique, specifically through the eigenvalues $\mathbf{D}_{\mathbf{s}_n}$ of the SCV covariances $\mathbf{C}_{\mathbf{s}_n}$. Any two or more eigenvectors of $\mathbf{V}_{\mathbf{s}}$ are not uniquely determined if their corresponding eigenvalues are equal. If said eigenvectors correspond to two different SCVs, then these eigenvectors are not uniquely determined *and* not uniquely block diagonal within $\mathbf{V}_{\mathbf{s}}$ (i.e., corresponding columns in $\mathbf{V}_{\mathbf{s}}^{[k]}$ are not uniquely equal to scaled $\mathbf{e}_n$). Thus the estimator may fail if it includes these eigenvectors in the procedure. By ignoring all eigenvectors of $\mathbf{V}_{\mathbf{x}}$ (corresponding to eigenvectors in $\mathbf{V}_{\mathbf{s}}$) with non-unique eigenvalues, it immediately follows that any single SCV $\mathbf{s}_n$ is identifiable if and only if at least one of its eigenvalues in $\mathbf{D}_{\mathbf{s}_n}$ is unique in $\mathbf{D}_{\mathbf{s}}$. This completes the proof.

An additional outcome of this proof is that given ACCM identifiability and assumptions in Theorem 1 hold, it can be shown that the ACCM solution exactly minimizes both the JCBD and JEBD costs in (4) and (7). With the true SCVs having a block diagonal $\mathbf{C}_{\mathbf{s}}$, then there exists a block diagonal representation of $\mathbf{V}_{\mathbf{s}}$ even if some eigenvectors are not unique (non-unique eigenvectors can be rotated by an orthogonal matrix into becoming block diagonal, and still provide valid eigenvectors of $\mathbf{C}_{\mathbf{s}}$). Thus in this representation, (9) exhibits a block diagonal $\mathbf{V}_{\mathbf{s}}^{[k]}$ for all $k = 1, \ldots, K$, like in Fig. 3. As the proof describes $\mathbf{W}^{[k]} = (\mathbf{W}^{[k]})_{\text{opt}}$ subject to the aforementioned ambiguities so long as identifiability holds, then $\hat{\mathbf{V}}_{\mathbf{y}}^{[k]} = \mathbf{V}_{\mathbf{s}}^{[k]}$ for all $k = 1, \ldots, K$ subject to sign and permutation ambiguity of the rows of $\hat{\mathbf{V}}_{\mathbf{y}}^{[k]}$. Therefore for the ACCM estimated $\mathbf{W}^{[k]}$, there similarly exists a block diagonal representation of $\hat{\mathbf{V}}_{\mathbf{y}}$ even if some eigenvectors are not unique, which means that $\hat{\mathbf{C}}_{\mathbf{y}}$ must also be block diagonal. Thus, the JCBD cost (4) and the JEBD cost (7) are minimized exactly (with (7) minimized provided it is written with respect to only the unique eigenvectors in $\hat{\mathbf{V}}_{\mathbf{x}}$), both with cost value of 0.

One important characteristic to note is that the separately estimated $(\mathbf{w}_n^{[k]})^{\top}$ are only guaranteed orthogonal within each $\mathbf{W}^{[k]}$ when the underlying SCVs are uncorrelated. However in practice, real-world data encounters SCVs that are not exactly uncorrelated. Thus, separately estimated $(\mathbf{w}_n^{[k]})^{\top}$ may not form an orthogonal matrix, which becomes problematic as $\mathbf{P}\ \hat{\mathbf{V}}_{\mathbf{x}} = \hat{\mathbf{V}}_{\mathbf{y}}$ is no longer true when $\mathbf{W}^{[k]}$ are not orthogonal. To preserve orthogonality, a final step after estimating each $(\mathbf{w}_n^{[k]})^{\top}$ is to map each $\mathbf{W}^{[k]}$ to its nearest orthogonal matrix [29] by $\mathbf{W}^{[k]} \rightarrow \mathbf{W}^{[k]}(\mathbf{W}^{[k]\top}\mathbf{W}^{[k]})^{-\frac{1}{2}}$. We note that still the mapped $\mathbf{W}^{[k]}$ may not necessarily achieve the optimum of (7) when SCVs are correlated. An alternative for finding this optimum would be to split (7) into costs for each $\mathbf{W}^{[k]}$, instead of for each $(\mathbf{w}_n^{[k]})^{\top}$ as in (8). However to our knowledge, explicitly minimizing that cost does not admit an analytic solution. We emphasize that our solution may not necessarily achieve the minimum cost when SCVs are correlated, but achieves a solution typically within range of the minimum while retaining the highly efficient analytic procedure. We support this statement with results in Section VI, showing that ACCM significantly outperforms the other orthogonally constrained methods (MCCA) when SCVs are correlated.

As stated in Section IV.A, the proposed estimation procedure is only achieved meaningfully if eigenvectors in $\hat{\mathbf{V}}_{\mathbf{x}}$ can be identified according to their N SCV groupings in $\mathbf{V}_{\mathbf{s}}$. In the next section, we introduce a scheme to group eigenvectors in $\hat{\mathbf{V}}_{\mathbf{x}}$ corresponding to the different SCVs, by defining a similarity measure between the eigenvectors from the inner-products between columns within the $\hat{\mathbf{V}}_{\mathbf{x}}^{[k]}$.

C. Sorting eigenvectors for the ACCM algorithm

From the formal definition of $\mathbf{V}_{\mathbf{s}}^{[k]} = \bigoplus_{n=1}^{N} (\mathbf{v}_{\mathbf{s}_n}^{[k]})^{\top}$ in IV.B, we note that all NK columns of $\mathbf{V}_{\mathbf{s}}^{[k]}$ can be represented by a scaled standard basis vector $\mathbf{e}_n$, with the nth set of K columns (of the nth SCV) corresponding to different scalings of the nth standard basis vector. Therefore, any two columns in $\mathbf{V}_{\mathbf{s}}^{[k]}$ have an inner product of 0 if they correspond to two eigenvectors of two different SCVs. Furthermore, as each column of $\mathbf{V}_{\mathbf{x}}^{[k]}$ is its corresponding column in $\mathbf{V}_{\mathbf{s}}^{[k]}$ rotated by orthogonal matrix $(\mathbf{W}^{[k]})_{\text{opt}}^{\top}$, we likewise have that any two columns in $\mathbf{V}_{\mathbf{x}}^{[k]}$ have an inner product of 0 if they correspond to two eigenvectors of two different SCVs.

The magnitudes of pairwise inner products between any two columns in $\mathbf{V}_{\mathbf{x}}^{[k]}$ is represented by the pairwise similarity matrix $\mathbf{G}^{[k]} \triangleq \text{abs}(\ \mathbf{V}_{\mathbf{x}}^{[k]\top}\ \mathbf{V}_{\mathbf{x}}^{[k]}\ ) \in \mathbb{R}^{NK \times NK}$, where we define abs($\mathbf{V}$) as the matrix obtained from the absolute values of $\mathbf{V}$. This similarity matrix allows a sorting of the NK eigenvectors of $\mathbf{V}_{\mathbf{x}}$ into their N SCV groups, since eigenvectors of different SCVs will have similarity of 0 and thus eigenvectors of the same SCV will naturally group together. This extends from the fact that $\mathbf{G}^{[k]}$ is *invariant* with respect to orthogonal mixing matrices $\mathbf{A}^{[k]}$, since abs($\mathbf{V}_{\mathbf{s}}^{[k]\top}\ \mathbf{V}_{\mathbf{s}}^{[k]}$) = abs($\mathbf{V}_{\mathbf{x}}^{[k]\top}\ \mathbf{A}^{[k]\top}\ \mathbf{A}^{[k]}\ \mathbf{V}_{\mathbf{x}}^{[k]}$) = abs($\mathbf{V}_{\mathbf{x}}^{[k]\top}\ \mathbf{V}_{\mathbf{x}}^{[k]}$).

With $\mathbf{G}^{[k]}$ specific to the kth dataset, this naturally leads to a similarity measure aggregated across the K datasets' $\mathbf{G}^{[k]}$, denoted by eigenvector similarity matrix $\mathbf{G} \in \mathbb{R}^{NK \times NK}$:

$$\mathbf{G} \triangleq \sum_{k=1}^{K} \mathbf{G}^{[k]} = \sum_{k=1}^{K} \text{abs}\left(\mathbf{V}_{\mathbf{x}}^{[k]\top}\ \mathbf{V}_{\mathbf{x}}^{[k]} \right) \tag{10}$$

Here we likewise have that $\mathbf{G}$ is invariant with respect to $\mathbf{A}^{[k]}$. Furthermore, similarity values in $\mathbf{G}$ are bounded in [0 1], given from the fact that eigenvectors in $\mathbf{V}_{\mathbf{x}}$ are unit norm.

Table 2. Computational complexity in the limit of ACCM, IVA-G, and the 5 variants of MCCA. Here q is the number of iterations, relevant only for the iterative algorithms (SSQCOR, GENVAR, IVA-G).

ACCM	SUMCORR	MAXVAR	MINVAR	SSQCOR	GENVAR	IVA-G
O(N^3K^3)	O(N^3K^3)	O($3N^4K^3$)	O($3N^4K^3$)	O($q(3N^4K + 2N^3K^2)$)	O($q(3N^4K + N^3K^2 + NK(K-1)^2)$)	O($q(KN^4 + K^2N^3 + NK^4)$)

Estimating $\mathbf{G}$ thus makes it possible to learn the eigenvector groupings of $\mathbf{V_x}$, and thus makes it possible to meaningfully estimate the $\mathbf{W}^{[k]}$ matrices according to the procedure proposed in IV.A . However, the proposed procedure for estimating $\mathbf{W}^{[k]}$ can be simplified by the fact that each subproblem's $(\mathbf{V}_\mathbf{x}^{[k]})_n$ is highly redundant, due to any SCV's K columns in $(\mathbf{V}_\mathbf{x}^{[k]})_n$ effectively being scalar multiples of one another. Omitting all but 1 column for each SCV still leads to each subproblem's quadratic form $(\hat{\mathbf{V}}_\mathbf{x}^{[k]})_n\ (\hat{\mathbf{V}}_\mathbf{x}^{[k]})_n^\top$ as a rank $N-1$ matrix with $(\mathbf{w}_n^{[k]})_{\text{opt}}$ as the eigenvector corresponding to the smallest eigenvalue (eigenvalue 0). Therefore, $(\mathbf{V}_\mathbf{x}^{[k]})_n$ can be constructed from only $N-1$ eigenvectors in $\mathbf{V_x}$ corresponding to $N-1$ different SCVs, which means that the entire estimation procedure requires finding only N eigenvectors in $\mathbf{V_x}$ corresponding to the N different SCVs.

One possible way to find N eigenvectors of $\mathbf{V_x}$ corresponding to the N different SCVs is to find any N eigenvectors whose $N \times N$ similarity matrix in $\mathbf{G}$ is equal or as close as possible to an identity matrix. For the purposes of our procedure, we implement a greedy search to select N eigenvectors of $\hat{\mathbf{V}}_\mathbf{x}$ whose $N \times N$ similarity matrix in $\hat{\mathbf{G}}$ has the minimum Frobenius norm distance to an identity matrix.

In the next section, we overview a practical design of the proposed algorithm given the results of the previous sections.

D. Analytic algorithm overview (ACCM)

We now overview the procedure of the proposed algorithm:

Algorithm Analytic Cross-Correlation Minimization

Input: a set of K datasets $\mathbf{X}^{[k]} \in \mathbb{R}^{N\times T}$ $(k = 1, \ldots, K)$
Output: a set of K demixing matrices $\mathbf{W}^{[k]} \in \mathbb{R}^{N\times N}$

1: concatenate the datasets $\mathbf{X} = [(\mathbf{X}^{[1]})^\top, \ldots, (\mathbf{X}^{[K]})^\top]^\top$, estimate the data covariance $\hat{\mathbf{C}}_\mathbf{x} = \frac{1}{T-1}\ \mathbf{X}\ \mathbf{X}^\top$, and eigendecompose the covariance $\hat{\mathbf{C}}_\mathbf{x} = \hat{\mathbf{V}}_\mathbf{x}\ \hat{\mathbf{D}}_\mathbf{x}\ \hat{\mathbf{V}}_\mathbf{x}^\top$
2: estimate the eigenvector similarity matrix: $\hat{\mathbf{G}} = \sum_{k=1}^K \text{abs}(\ \hat{\mathbf{V}}_\mathbf{x}^{[k]\top}\ \hat{\mathbf{V}}_\mathbf{x}^{[k]}\ )$
3: find N eigenvectors $(\hat{\mathbf{V}}_\mathbf{x})_N \in \mathbb{R}^{NK\times N}$ whose $N \times N$ similarity matrix in $\hat{\mathbf{G}}$ is as close as possible to an identity matrix ("maximally independent" according to $\hat{\mathbf{G}}$).
4: estimate each $\mathbf{w}_n^{[k]}$ vector according to (8), where $(\hat{\mathbf{V}}_\mathbf{x}^{[k]})_n \in \mathbb{R}^{N\times N-1}$ is formed from the N rows of $(\hat{\mathbf{V}}_\mathbf{x})_N$ corresponding to dataset k, and from all columns of $(\hat{\mathbf{V}}_\mathbf{x})_N$ except for the nth column.
5: map all $\mathbf{W}^{[k]}$ to the nearest orthogonal matrix: $\mathbf{W}^{[k]} \to \mathbf{W}^{[k]}(\mathbf{W}^{[k]\top}\mathbf{W}^{[k]})^{-\frac{1}{2}}$

The algorithm's computational complexity is dominated by the eigendecomposition of $\hat{\mathbf{C}}_\mathbf{x}$, of complexity O($(NK)^3$). Aside from this, the second most complex step is step 3. For each $(\hat{\mathbf{V}}_\mathbf{x})_N$ candidate, computing the norm of its similarity matrix in $\hat{\mathbf{G}}$ has complexity O(N^2). If we were to evaluate all combinations of N eigenvectors within the NK total, this amounts to O($\binom{NK}{N}\ N^2$). In practice, we simplify this step only selecting a few $(\hat{\mathbf{V}}_\mathbf{x})_N$ candidates from eigenvectors corresponding to both the largest and smallest eigenvalues in $\hat{\mathbf{C}}_\mathbf{x}$, exemplifying eigenvectors describing most of the effects of correlation or partial correlation among SCVs in $\mathbf{C_y}$. For smaller complexity of this step, we perform a greedy search (starting with smallest values in $\hat{\mathbf{G}}$), and limit the number of $(\hat{\mathbf{V}}_\mathbf{x})_N$ candidates to $\frac{1}{2}\ (KN^2)^{\frac{1}{3}}$, thus the complexity of this step is always less than O($(NK)^3$). Finally, steps 4 and 5 both have complexity O(KN^3), leading to the algorithm having a total complexity of O($(NK)^3$) .

In the next section, we discuss other JBSS algorithms exploiting source correlation across datasets. We compare implementation of these algorithms with ACCM, in preparation for performance evaluations in Sections VI and VII.

V. Relationship to other JBSS algorithms exploiting source correlation across datasets

A. IVA — Multivariate Gaussian Distribution (IVA-G)

IVA-G assumes that each SCV has a probability distribution function (PDF) characterized by the i.i.d. multivariate Gaussian distribution [10]. Using the previously defined quantities, this leads to the general IVA-G cost function :

$$\mathcal{J}_{\text{IVA-G}}\ (\boldsymbol{\mathcal{W}}) = \frac{NK\ \log(2\pi e)}{2} + \frac{1}{2}\sum_{n=1}^{N} \log \left|\ \det\left(\hat{\mathbf{C}}_{\mathbf{y}_n}\right)\ \right| - \sum_{k=1}^{K} \log \left|\ \det\left(\mathbf{W}^{[k]}\right)\ \right|$$

where $\boldsymbol{\mathcal{W}} = \{\mathbf{W}^{[1]}, \ldots, \mathbf{W}^{[K]}\}$ is the set of $\mathbf{W}^{[k]}$. The term $\log \mid \det\left(\hat{\mathbf{C}}_{\mathbf{y}_n}\right) \mid$ measures correlation within the nth SCV, while the term $\log \mid \det\left(\mathbf{W}^{[k]}\right) \mid$ acts as a penalty keeping $\mathbf{W}^{[k]}$ close to orthogonal.

Unlike most other JBSS methods exploiting source correlation across datasets, IVA-G does not constrain $\mathbf{W}^{[k]}$ to be orthogonal. This gives IVA-G the ability to estimate correlated SCVs, leading to a significantly more robust estimator. IVA-G also has less stringent identifiability conditions than other methods (including ACCM): provided that no two sources within an SCV are independent, IVA-G can identify all SCVs so long as any two SCVs $\mathbf{s}_i$ and $\mathbf{s}_j$ do not have identical covariances $\mathbf{C}_{\mathbf{s}_i}$ and $\mathbf{C}_{\mathbf{s}_j}$ subject to the aforementioned ambiguities. Thus, IVA-G may be considered a "gold standard" of JBSS algorithms exploiting source correlation across datasets.

However, compared with other JBSS algorithms exploiting source correlation, IVA-G suffers considerably in terms of

computational complexity. Asides from initial costs of estimating $\hat{\mathbf{C}}_{\mathbf{x}}$, each IVA-G iteration requires updating all NK demixing vectors $\mathbf{w}_n^{[k]}$, where each of these $\mathbf{w}_n^{[k]}$ operations involves an update of $\mathbf{W}^{[k]}$ of O(N^3) complexity, an update of $\hat{\mathbf{C}}_{\mathbf{y}_n}$ of O(N^2K) complexity, and an update of $\hat{\mathbf{C}}_{\mathbf{y}_n}^{-1}$ of O(K^3) complexity. If the algorithm requires q iterations to converge, this leads to IVA-G having a total complexity of O($q(N^4K + N^3K^2 + NK^4)$) .

B. MCCA

MCCA generalizes CCA to more than two datasets, and variations of MCCA are capable of JBSS [6]. Using previously defined concepts, the goal at each stage of MCCA is to estimate demixing vectors $\mathbf{w}_n^{[k]}$ corresponding to SCV $\mathbf{y}_n$ such that correlation in the SCV's estimated $\hat{\mathbf{C}}_{\mathbf{y}_n}$ is maximized.

Five variants of MCCA exist, each employing a different empirical measure of correlation in $\hat{\mathbf{C}}_{\mathbf{y}_n}$. These costs are evaluated with respect to covariance entries $(\hat{\mathbf{C}}_{\mathbf{y}_n})_{ij}$, or eigenvalues $\hat{\lambda}_k$, where $\hat{\lambda}_k$ is the kth largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}_n}$. The five variants are given by their cost functions :

1) **SUMCORR** : maximize the sum of elements in $\hat{\mathbf{C}}_{\mathbf{y}_n}$
   $\mathcal{J}_{\text{SUMCORR}}\ (\hat{\mathbf{C}}_{\mathbf{y}_n}) = \mathbf{1}^\top\ \hat{\mathbf{C}}_{\mathbf{y}_n}\ \mathbf{1}$.
2) **MAXVAR** : maximize the largest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}_n}$
   $\mathcal{J}_{\text{MAXVAR}}\ (\hat{\mathbf{C}}_{\mathbf{y}_n}) = \hat{\lambda}_1$.
3) **MINVAR** : minimize the smallest eigenvalue of $\hat{\mathbf{C}}_{\mathbf{y}_n}$
   $\mathcal{J}_{\text{MINVAR}}\ (\hat{\mathbf{C}}_{\mathbf{y}_n}) = \hat{\lambda}_K$.
4) **SSQCOR** : maximize the sum of squared elements of $\hat{\mathbf{C}}_{\mathbf{y}_n}$, equal to sum of squared eigenvalues of $\hat{\mathbf{C}}_{\mathbf{y}_n}$
   $\mathcal{J}_{\text{SSQCOR}}\ (\hat{\mathbf{C}}_{\mathbf{y}_n}) = \sum_{i=1}^K \sum_{j=1}^K (\hat{\mathbf{C}}_{\mathbf{y}_n})_{ij}^2 = \sum_{k=1}^K \hat{\lambda}_k^2$.
5) **GENVAR** : minimize the product of eigenvalues (determinant) of $\hat{\mathbf{C}}_{\mathbf{y}_n}$
   $\mathcal{J}_{\text{GENVAR}}\ (\hat{\mathbf{C}}_{\mathbf{y}_n}) = \prod_{k=1}^K \hat{\lambda}_k$.

With pre-whitened data, all MCCA algorithms constrain $\mathbf{W}^{[k]}$ to be orthogonal. All MCCA algorithms (with exception of SUMCORR) are solved by deflationary procedures estimating each SCV one at a time, where $\mathbf{w}_n^{[k]}$ are constrained orthogonal to previously estimated $\mathbf{w}_i^{[k]}$ for $i < n$. GENVAR and SSQCOR are necessarily solved by iterative procedures where an initial guess is refined until convergence, whereas SUMCORR, MAXVAR, and MINVAR are achievable with analytic solutions. Table 2 lists the computational complexities of ACCM, IVA-G and the 5 variants of MCCA.

Several of the MCCA algorithms have noteworthy connections to the previously mentioned JBSS algorithms. GENVAR can be seen as a deflationary, orthogonally constrained version of IVA-G. Similarly, SSQCOR can be seen as a deflationary approach to minimizing the JCBD cost in (5). This connection of SSQCOR to (5) is provable from the fact that given $\mathbf{W}^{[k]}$ are constrained orthogonal, maximizing the squared norm of the SCV covariances is equivalent to minimizing the squared norm of the SCV cross-covariances in (5).

MAXVAR and SUMCORR both model each SCV $\mathbf{S}_n \in \mathbb{R}^{K\times T}$ as a common source shared across datasets, specifically a rank 1 "signal" matrix plus a full rank "noise" matrix :

$$\mathbf{S}_n = \mathbf{u}_n\mathbf{v}_n^\top + \mathbf{Z}_n, \quad 1 \le n \le N \tag{11}$$

where we define $\mathbf{v}_n \in \mathbb{R}^T$ as the common source, $\mathbf{u}_n \in \mathbb{R}^K$ as the weights of $\mathbf{v}_n$ within each $\mathbf{s}_n^{[k]}$, and $\mathbf{Z}_n \in \mathbb{R}^{K\times T}$ as the additive noise matrix (assumed to have a diagonal covariance matrix). The differences in the algorithms is on how $\mathbf{u}_n$ is modeled: SUMCORR assumes that weights in $\mathbf{u}_n$ are equivalent to one another, while MAXVAR estimates $\mathbf{u}_n$ that minimizes tr[var$\{\mathbf{S}_n - \mathbf{u}_n\mathbf{v}_n^\top\}$], where tr[.] denotes the trace and var$\{.\}$ denotes the variance [13].

SUMCORR is particularly notable for an analytic solution highly similar to ACCM. Like ACCM, the SUMCORR solution is estimated directly from eigenvectors of $\hat{\mathbf{C}}_{\mathbf{x}}$ [14], [15]. With the previously defined concepts in Section IV, provided that datasets are pre-whitened, the SUMCORR $\mathbf{w}_n^{[k]}$ are directly obtained from the N principal eigenvectors of $\hat{\mathbf{C}}_{\mathbf{x}}$ (corresponding to the N largest eigenvalues). Specifically, the nth principal eigenvector of $\hat{\mathbf{C}}_{\mathbf{x}}$, denoted by $\hat{\mathbf{v}}_{\mathbf{x}_n} \in \mathbb{R}^{NK}$, is the concatenation of scaled SUMCORR demixing vectors of the nth SCV: $\hat{\mathbf{v}}_{\mathbf{x}_n} = [(\mathbf{w}_n^{[1]})^\top, \ldots, (\mathbf{w}_n^{[K]})^\top]^\top$. This solution can be seen as performing a rank-N PCA on the vertical concatenation of datasets, and is elsewhere called group-PCA [30]. This simple analytic solution leads to SUMCORR being among the most efficient of all JBSS algorithms, as shown in Table 2. In comparison, the ACCM solution is estimated from N eigenvectors that are maximally independent according to eigenvector similarity matrix $\hat{\mathbf{G}}$ (eigenvectors that are most "block diagonal" at the optimum), which better ensures that the N eigenvectors correspond to the N different SCVs (and thus all N SCVs are identified).

While complexities of SUMCORR and ACCM are identical in the limit (both O(N^3K^3)), there are several advantages that ACCM provides over SUMCORR. To estimate the true sources, SUMCORR requires that the N principal eigenvectors of $\hat{\mathbf{C}}_{\mathbf{x}}$ correspond to the N different SCVs. Thus, SUMCORR has identifiability conditions that the largest eigenvalue of each $\mathbf{C}_{\mathbf{s}_n}$ must be both unique in $\mathbf{D}_{\mathbf{s}}$, and larger than the second largest eigenvalue of all other $\mathbf{C}_{\mathbf{s}_n}$. Thus, SUMCORR has more stringent identifiability conditions compared to ACCM. Fig. 1 (c) displays one example when ACCM is able to identify the sources but SUMCORR fails. Furthermore, when SCVs are correlated, "off-block" entries in the N principal eigenvectors are typically much farther from 0 than "off-block" entries in other eigenvectors, thus ACCM may choose better eigenvectors for minimizing SCV cross-correlation.

C. Joint Diagonalization Approaches for JBSS

Another approach to solving the BSS problem has been through joint diagonalization (JD) of multiple matrices, e.g., JADE [31] and SOBI [32] are two such early algorithms, and now have many variants. Several generalizations of these have also been proposed for JBSS [24]–[28], which are called "generalized joint diagonalization" (GJD). GJD algorithms typically estimate demixing matrices $\mathbf{W}^{[k]}$ that jointly diagonalize a set of "cumulant matrices" that describe source dependence within and across the datasets. One noteworthy

use of cumulant matrices are the covariances/pairwise cross-covariances between the K datasets ($\mathrm{E}\left\{\mathbf{x}^{[i]}\,\mathbf{x}^{[j]\top}\right\}$ for $1 \leq i, j \leq K$), for which it can be shown that the general GJD cost used in [24] is equivalent to the JCBD cost in (4), provided that $\mathbf{W}^{[k]}$ are constrained to be orthogonal. In BSS terminology, this form of GJD provides a "symmetric" iterative procedure for minimizing (4) (i.e., all demixing vectors $\mathbf{w}_n^{[k]}$ are learned in parallel), as opposed to the deflationary iterative procedure provided by SSQCOR for minimizing (4).

Other types of matrices used in GJD-type algorithms include fourth-order cumulant matrices as in JADE and auto-covariance matrices describing sample to sample dependence within and across the sources, as in SOBI. This flexibility in cumulant matrices make GJD algorithms attractive for JBSS.

In the next section, we demonstrate performance of these JBSS algorithms in Section V with simulated data, demonstrating how each algorithm's separation performance depends on the statistics that define the generative model of the data and underlying sources. After that, we demonstrate performance on real fMRI sources in the context of a hybrid experiment.

VI. Results

We use joint inter-symbol-interference (joint-ISI) to compare separation performance of JBSS algorithms when $\mathbf{A}^{[k]}$ are known, such as for simulations. Joint-ISI is given by:

$$\mathrm{ISI}_{\mathrm{JNT}}\left(\boldsymbol{\mathcal{W}}, \boldsymbol{\mathcal{A}}\right) \triangleq \frac{1}{2N(N-1)} \left[\sum_{n=1}^{N} \left(\sum_{m=1}^{N} \frac{\bar{q}_{[n,m]}}{\max_p \bar{q}_{[n,p]}} - 1 \right) + \sum_{m=1}^{N} \left(\sum_{n=1}^{N} \frac{\bar{q}_{[n,m]}}{\max_p \bar{q}_{[p,m]}} - 1 \right) \right]$$

With $\boldsymbol{\mathcal{W}}$ as the set of $\mathbf{W}^{[k]}$, $\boldsymbol{\mathcal{A}}$ as the set of $\mathbf{A}^{[k]}$, $\mathbf{Q}^{[k]} = \mathbf{W}^{[k]}\,\mathbf{A}^{[k]}$ as the "mixing-demixing matrix" of the kth dataset, $q_{[m,n]}^{[k]}$ as the $[m, n]$ entry in $\mathbf{Q}^{[k]}$, and $\bar{q}_{[m,n]} = \sum_{k=1}^{K} |q_{[m,n]}^{[k]}|$. Joint-ISI is given in [10] as an extension of the inter-symbol-interference measure (ISI) introduced in [33]. Joint-ISI is normalized in [0 1], and collectively measures how close each $\mathbf{Q}^{[k]} = \mathbf{W}^{[k]}\,\mathbf{A}^{[k]}$ matrix is to a permuted diagonal matrix, with 0 joint-ISI indicative of perfect separation.

As our paper primarily focuses on efficient JBSS, we limit our results to the source correlation-based algorithms explained in Section V. These include IVA-G, the 5 variants of MCCA, and the GJD algorithm in [24] called "JBSS-SOS", which minimizes (4) subject to the orthogonality constraint.

For all performance evaluations done in Sections VI and VII, we use the computational recources provided by the UMBC High Performance Computing Facility (HPCF), thus CPU time is reflective of HPCF's capabilities. All iterative algorithms used a maximum 1000 iterations and same stop criteria $\zeta \leq 0.0001$, defined by $\zeta = \max(1 - \mathrm{diag}(\mathbf{W}_{\mathrm{old}}^{[k]}\,\mathbf{W}^{[k]\top}))$ where $\mathbf{W}_{\mathrm{old}}^{[k]}$ is $\mathbf{W}^{[k]}$ of the previous iteration.

A. Performance with simulated data

Our SCV generative model is as follows. We model an SCV $\mathbf{S}_n \in \mathbb{R}^{K \times T}$ (mean $\mathbf{0} \in \mathbb{R}^K$, covariance $\mathbf{C}_{\mathbf{s}_n} \in \mathbb{R}^{K \times K}$) as the sum of a low rank source matrix and a full rank noise matrix:

$$\mathbf{S}_n = \mathbf{U}_n \mathbf{V}_n^{\top} + \mathbf{Z}_n, \quad 1 \leq n \leq N, \tag{12}$$

where we define $\mathbf{V}_n \in \mathbb{R}^{R_n \times T}$ as the low rank source components spanning the SCV, $\mathbf{U}_n \in \mathbb{R}^{K \times R_n}$ as the weights of the source components within each $\mathbf{s}_n^{[k]}$, and $\mathbf{Z}_n \in \mathbb{R}^{K \times T}$ as the additive noise matrix of $\mathbf{S}_n$. The number of components R_n is called the *effective rank* of SCV $\mathbf{S}_n$, which represents a level of "complexity" within the SCV. SCVs with a larger effective rank tend to have covariances $\mathbf{C}_{\mathbf{s}_n}$ closer to an identity matrix (less correlation among sources in $\mathbf{S}_n$).

This model in (12) generalizes the MAXVAR and SUMCORR model of $\mathbf{S}_n = \mathbf{u}_n\,\mathbf{v}_n^{\top} + \mathbf{Z}_n$ in (11), where our model allows the effective rank of $\mathbf{S}_n$ to be greater than 1.

In order to have full control over simulation of SCVs, we define a model analogous to (12) on the covariances $\mathbf{C}_{\mathbf{s}_n}$:

$$\mathbf{C}_{\mathbf{s}_n} = \alpha_n \mathbf{Q}_n \mathbf{Q}_n^{\top} + (1 - \alpha_n)\mathbf{I}_K, \quad 1 \leq n \leq N, \tag{13}$$

where $\mathbf{I}_K \in \mathbb{R}^{K \times K}$ is the identity matrix (covariance of $\mathbf{Z}_n$), $\mathbf{Q}_n \in \mathbb{R}^{K \times R_n}$ is a matrix where rows of $\mathbf{Q}_n$ are random unit norm Gaussian distributed vectors (thus $\mathbf{Q}_n\,\mathbf{Q}_n^{\top}$ is a random rank R_n correlation matrix), and α_n is a value in [0 1] determining the signal to noise ratio (SNR) of the SCV. By this model, we can generate a SCV following the low rank model in (12) by directly specifying its covariance in (13).

We also test algorithm performance in the event that SCVs are not exactly uncorrelated. We introduce SCV cross-correlation by simulating all $\mathbf{S}_n$ together concurrently : $\mathbf{S} \triangleq [(\mathbf{S}_1)^{\top}, \ldots, (\mathbf{S}_N)^{\top}]^{\top} \in \mathbb{R}^{NK \times T}$ (mean $\mathbf{0} \in \mathbb{R}^{NK}$, covariance $\mathbf{C}_{\mathbf{s}} \in \mathbb{R}^{NK \times NK}$). This allows us to directly specify SCV cross-correlation via the SCV cross-covariance $\mathbf{C}_{\mathbf{s}}$. Here, values in "off-blocks" (pairwise cross-covariances between SCVs) can now be varied in average value of their magnitudes, which we denote as "SCV cross-corr" γ. SCVs become more correlated as γ increases, with $\gamma = 0$ indicating uncorrelated SCVs. This model better represents real-world SCVs which are typically correlated, such as with fMRI data.

There are a number of different variables in generating the underlying SCVs that affects performance of the algorithms. These variables include the number of SCVs N, the number of samples T, the number of the datasets K, the effective rank of the SCVs R_n, the SNR within SCVs (represented by α_n), and the degree of SCV cross-correlation (represented by γ). For the following experiments, we study the effect of varying all of these variables except for N and T, since all algorithms consistently perform worse with increasing N and decreasing T. For further simplicity, we specify each SCV's α_n to be distinct and equidistant within [0.8 0.5], to represent SCVs with both high and low source correlation. Therefore, we study the performance of the algorithms with respect to the number of the datasets K, the effective rank of the SCVs R_n, and the degree of SCV cross-correlation (via γ). Each simulation concerns changing one variable while keeping the others fixed to either small or large values.

Unless otherwise specified, in all simulations we fix $N = 4$, $K = 10$, and $T = 10000$. All SCVs have $\mathbf{C}_{\mathbf{s}_n}$ specified according to (13), with α_n = [0.8 0.7 0.6 0.5] respectively for the 4 SCVs, and R_n the same for all SCVs. $\mathbf{C}_{\mathbf{s}}$ is specified such

that SCV covariances equal $\mathbf{C}_{\mathbf{s}_n}$, and the average magnitude of pairwise SCV cross-covariances equals γ. All SCVs are then generated concurrently from a multivariate Gaussian random vector $\mathbf{S}$ (mean $\mathbf{0} \in \mathbb{R}^{NK}$, covariance $\mathbf{C}_\mathbf{s} \in \mathbb{R}^{NK \times NK}$). Sources are then distributed to their corresponding datasets $\mathbf{S}^{[k]}$, then mixed according to $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\ \mathbf{S}^{[k]}$ with values in $\mathbf{A}^{[k]}$ drawn from the standard Gaussian distribution.

Later, we also test performance with noisy observations: $\mathbf{X}^{[k]} = \eta\ \mathbf{A}^{[k]}\ \mathbf{S}^{[k]} + (1-\eta)\ \mathbf{N}^{[k]}$, where $\mathbf{N}^{[k]}$ is standardized Gaussian noise, and $\eta \in [0\ 1]$ is a normalized measure of dataset SNR. We implement $\eta = 1$ (noiseless $\mathbf{X}^{[k]}$) for all earlier experiments except when later testing the effect of η.

Fig. 4 plots the algorithms' CPU time performances with varying the number of SCVs N and the number of datasets K. We first note that SUMCORR and ACCM are among the least expensive JBSS algorithms, with comparable time costs. Other MCCA algorithms have considerably higher costs (in order of MAXVAR, MINVAR, GENVAR, and SSQCOR), and IVA-G has the highest cost. JBSS-SOS has the second highest cost for small N, but is comparable to MCCA algorithms for very large N. Both plots were observed using $R_n = 2$, $\gamma = 0.0$, and $\eta = 1$, however we note that across all experiments, each algorithm's time was observed to primarily only depend on the dimensions N and K, and less on the statistics of the data.

Fig. 5 plots the algorithms' joint-ISI performances with respect to K. IVA-G is consistently the most accurate of these algorithms. ACCM is generally second to fourth most accurate (most to third most accurate orthogonally constrained algorithm), often competing closely with SSQCOR and JBSS-SOS. Given uncorrelated SCVs ($\gamma = 0.0$) in (a) and (b), orthogonally constrained algorithms such as ACCM. SSQCOR, and JBSS-SOS are nearly able to match IVA-G's performance. SUMCORR and MAXVAR perform nearly identically in all cases here, except for (c) where SUMCORR performs worse. Performance of GENVAR in (a) and (b) is difficult to ascertain: intuitively when SCVs have $R_n < K$, then the smallest K-R_n eigenvalues of SCV covariances are expected to describe SCV noise. Thus, GENVAR may perform well when $R_n = K$ and SCVs are uncorrelated as in (b) where $K = 10$, but otherwise GENVAR may not be a dependable MCCA algorithm. MINVAR is consistently among the worst performing algorithms. Finally, with $\gamma = 0.4$ in (c) and (d), ACCM significantly outperforms all MCCA algorithms, and outperforms JBSS-SOS in (d) for $R_n = 10$, indicating a relatively high-quality orthogonally constrained solution.

Fig. 6 plots the algorithms' joint-ISI performances with respect to varying the effective rank of all SCVs R_n. As in Fig. 5, we see that IVA-G is the most accurate algorithm, that ACCM is second to fourth most accurate (closely competing with SSQCOR and JBSS-SOS), and that SUMCORR and MAXVAR perform nearly identically to each other. SUMCORR and MAXVAR also perform worse with larger R_n, likely attributed to their generative model in (11) assuming $R_n = 1$. GENVAR performs comparatively well in (a) given the SCVs are close to full effective rank ($R_n = K$), but otherwise performs poorly. Finally, we note that ACCM significantly outperforms all other orthogonally constrained algorithms in (d) when γ, K, and R_n are large, representing data with a higher level of complexity in inter-SCV and intra-SCV correlation (often observed with real-world data). We do not include CPU time plots with respect to increasing R_n, since changes in time were not significant across algorithms, except for IVA-G where time was significantly higher for $R_n = 1$ and nearly constant for all other R_n.

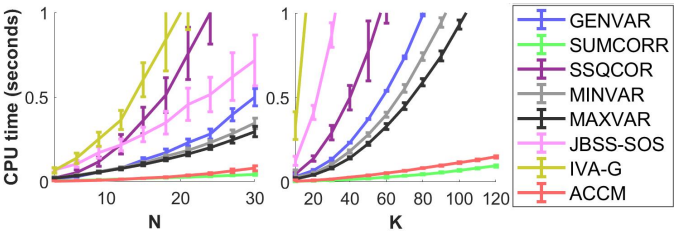

Fig. 4. CPU time (seconds) w.r.t. varying number of sources N (fixing K=10) and number of datasets K (fixing N=4).

Fig. 7 plots the algorithms' joint-ISI performances with respect to varying the average of absolute values in pairwise SCV cross-covariances ("SCV cross-corr") γ. SCVs become more correlated as γ increases, corresponding to a significantly worse performance for orthogonally constrained methods. As IVA-G does not constrain the $\mathbf{W}^{[k]}$ to be orthogonal, IVA-G significantly outperforms when γ is higher. As in Fig. 5 and 6, we note that ACCM generally the most to third most accurate orthogonally constrained algorithm. Additionally, Fig. 5, 6, and 7 all demonstrate that performance difference between ACCM and the other orthogonally constrained algorithms increases significantly when γ, R_n and K are large. Other similarities to previous figures are that GENVAR performs comparatively well here in (b) given that $R_n = K$ (see K = 10), and that MAXVAR and SUMCORR perform nearly identically to each other. We also note that MAXVAR and SUMCORR are among the worst performing MCCA algorithms observed in these simulated conditions. Regarding performance of SUMCORR, this may be due to the fact that the SUMCORR solution is estimated from the N principal eigenvectors of $\hat{\mathbf{C}}_\mathbf{x}$ as mentioned in Section V.B. . With correlated SCVs, principal eigenvectors of $\mathbf{C}_\mathbf{s}$ are farther from block diagonal, leading to worse estimation. In contrast, ACCM selects N eigenvectors that are close as possible to "block diagonal", which better ensures that the N different SCVs are identified, leading to better performance given correlated SCVs. We observed that changes in time with varying γ were not significant across algorithms, except for IVA-G where time slightly increased as γ increased.

Fig. 8 plots the algorithms' joint-ISI performances with respect to varying the SNR η, with $\eta \in [0\ 1]$, and Gaussian noise added by $\mathbf{X}^{[k]} = \eta\ \mathbf{A}^{[k]}\ \mathbf{S}^{[k]} + (1-\eta)\ \mathbf{N}^{[k]}$. If we were to observe as $T \to \infty$, adding noise this way can be seen as "shrinking" values in the data covariance $\hat{\mathbf{C}}_\mathbf{x}$, which tends $\hat{\mathbf{C}}_\mathbf{x}$ closer to an identity matrix as $\eta \to 0$, resulting in unidentifiable sources for all algorithms. An interesting note is that IVA-G is the best performing algorithm with high SNR η, but is among the worst performing for low η. The orthogonally constrained algorithms appear to perform relatively better for very low SNR; this may be due to the

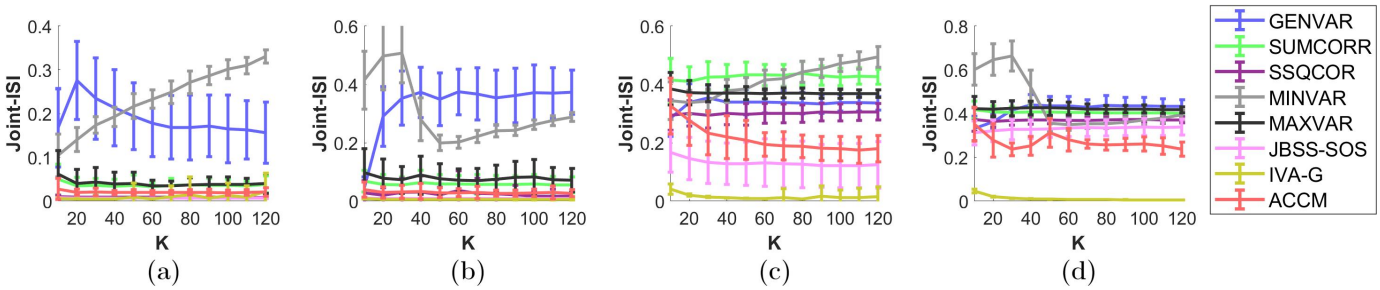

Fig. 5. Joint-ISI performance w.r.t. varying number of datasets K, while fixing SCV effective rank R_n and "SCV cross-corr" γ. (a) $R_n = 2$, $\gamma = 0.0$. (b) $R_n = 10$, $\gamma = 0.0$. (c) $R_n = 2$, $\gamma = 0.4$. (d) $R_n = 10$, $\gamma = 0.4$.

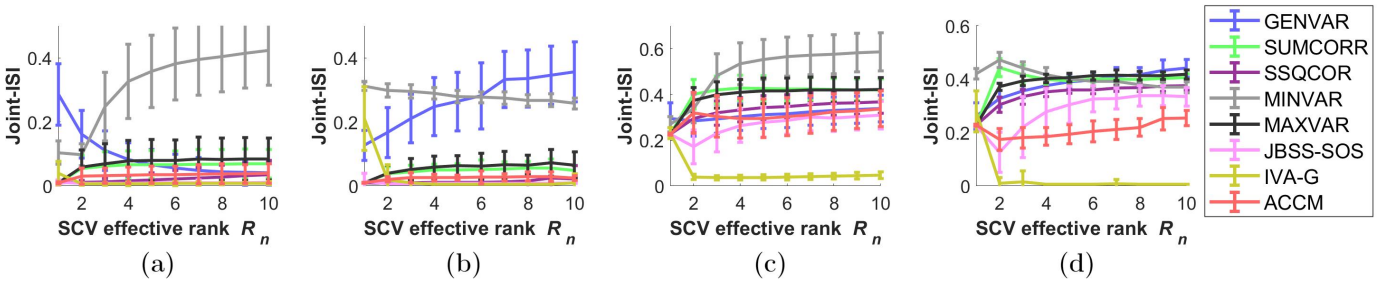

Fig. 6. Joint-ISI performance w.r.t. varying effective rank of the SCVs R_n, while fixing "SCV cross-corr" γ and number of datasets K. (a) $K = 10$, $\gamma = 0.0$. (b) $K = 100$, $\gamma = 0.0$. (c) $K = 10$, $\gamma = 0.4$. (d) $K = 100$, $\gamma = 0.4$.

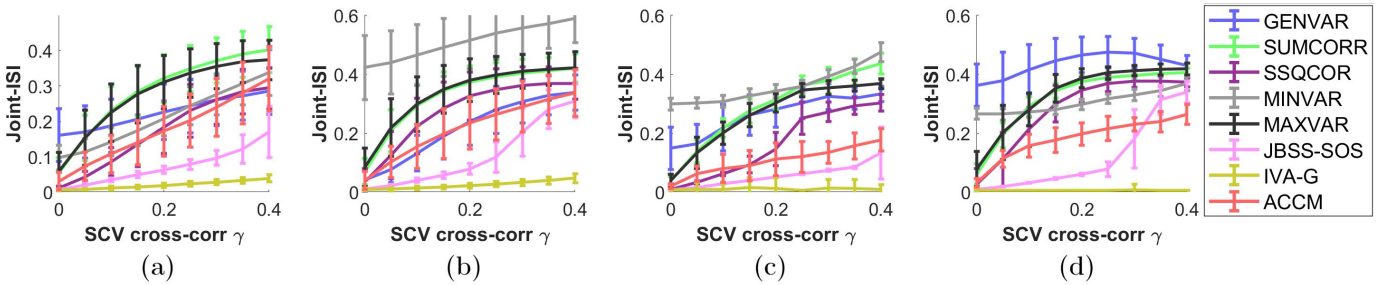

Fig. 7. Joint-ISI performance w.r.t. varying the average of absolute values in pairwise SCV cross-covariances ("SCV cross-corr") γ, while fixing SCV effective rank R_n and number of datasets K.
(a) $R_n = 2$, $K = 10$. (b) $R_n = 10$, $K = 10$. (c) $R_n = 2$, $K = 100$. (d) $R_n = 10$, $K = 100$.

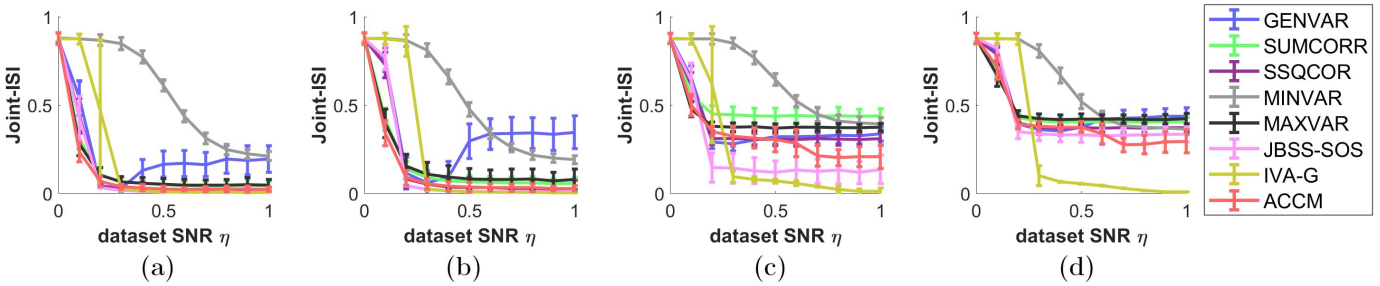

Fig. 8. Joint-ISI performance w.r.t. varying normalized dataset SNR η, while fixing SCV effective rank R_n and "SCV cross-corr" γ. (a) $R_n = 2$, $\gamma = 0.0$. (b) $R_n = 10$, $\gamma = 0.0$. (c) $R_n = 2$, $\gamma = 0.4$. (d) $R_n = 10$, $\gamma = 0.4$.

constraint's reduction of the solution space (effectively half as many parameters to estimate) being useful when the data is especially noisy. Otherwise, all other conclusions noted from the previous figures were also observed here. ACCM is among the best performing orthogonally constrained algorithms, and ACCM performs relatively well for very low SNR.

Next, we study the performance of the JBSS algorithms for fMRI analysis in the context of a hybrid data experiment.

B. fMRI hybrid experiment

One common application of JBSS is for analyzing medical imaging datasets, particularly with functional magnetic reso-

nance imaging data (fMRI) [4], [6]–[8]. There are several ways of applying JBSS to fMRI data, but for our experiments we apply JBSS to the data introduced in [34], openly available at https://coins.trendscenter.org. Having fMRI source components previously extracted from this dataset, we use these sources in a hybrid experiment to generate hybrid data. This allows us to scale the data to a very large number of datasets, while also allowing us to know the ground truth and thus directly infer the estimation performance of the algorithms.

A total of $K = 200$ datasets are generated over $N = 20$ SCVs, wherein each SCV is generated from linear mixtures of fMRI source components combined with additive Gaussian noise. The SCV generative model is given the same as in (12), with $\mathbf{V}_n \in \mathbb{R}^{R_n \times T}$ as the low rank fMRI components spanning $\mathbf{S}_n$, $\mathbf{U}_n \in \mathbb{R}^{K \times R_n}$ as the weights of the components within each $\mathbf{s}_n^{[k]}$, and $\mathbf{Z}_n \in \mathbb{R}^{K \times T}$ as the noise matrix.

Each SCV is given an effective rank R_n between 2 and 10, for which R_n fMRI components $\mathbf{V}_n \in \mathbb{R}^{R_n \times T}$ are assigned to that SCV. We introduce various levels of correlation across the SCVs by specifically introducing correlation across the different $\mathbf{V}_n$. Analogous to what is done in Section VI.A where we define the average magnitude of correlation between SCVs γ, here we define the magnitude of correlation between the SCVs' respective fMRI components $\mathbf{V}_n$.

To introduce this correlation across the $\mathbf{V}_n$, we first specify a matrix $\mathbf{H} \triangleq [(\mathbf{V}_1)^\top, \ldots, (\mathbf{V}_N)^\top]^\top \in \mathbb{R}^{R_H \times T}$ as the vertical concatenation of the $\mathbf{V}_n$, with $R_H \triangleq \sum_{n=1}^{N} R_n$, and $\hat{\mathbf{C}}_\mathbf{H} = \frac{1}{T-1}\,\mathbf{H}\,\mathbf{H}^\top \in \mathbb{R}^{R_H \times R_H}$ as the desired cross covariance matrix among all $\mathbf{V}_n$. Correlation between any two $\mathbf{V}_n$ is introduced by specifying the pairwise cross covariances $\hat{\mathbf{\Theta}}_{(m,n)} \triangleq \frac{1}{T-1}\,\mathbf{V}_m\,\mathbf{V}_n^\top \in \mathbb{R}^{R_m \times R_n}$. For each pair of SCVs, all values of the corresponding $\hat{\mathbf{\Theta}}_{m,n}$ are made to equal a value $\beta_{(m,n)}$ drawn randomly from the uniform distribution in [0 0.4], and all $\hat{\mathbf{\Theta}}_{n,n}$ are specified as matrices where diagonal values equal 1 and off-diagonal values equal 0.5 . This ensures that the SCVs are still separable from each other (maximally uncorrelated) despite the SCVs being correlated.

With desired covariance $\hat{\mathbf{C}}_\mathbf{H}$ defined by the pairwise cross covariances $\hat{\mathbf{\Theta}}_{(m,n)}$, we then obtain R_H fMRI components $\mathbf{H}$ with covariance $\hat{\mathbf{C}}_\mathbf{H}$ by applying a coloring transform on R_H uncorrelated fMRI components $\tilde{\mathbf{H}} \in \mathbb{R}^{R_H \times T}$, given by the coloring transform $\mathbf{H} = \hat{\mathbf{V}}_\mathbf{H}\,\hat{\mathbf{D}}_\mathbf{H}^{\frac{1}{2}}\,\tilde{\mathbf{H}}$. Here $\hat{\mathbf{V}}_\mathbf{H}$ are the eigenvectors of desired covariance matrix $\hat{\mathbf{C}}_\mathbf{H}$, and $\hat{\mathbf{D}}_\mathbf{H}$ the respective eigenvalues. With correlated fMRI components thus obtained by $\mathbf{H}$, we thus distribute the different $\mathbf{V}_n$ to each SCV, and then generate SCVs according to $\mathbf{S}_n = \mathbf{U}_n\mathbf{V}_n^\top + \mathbf{Z}_n$, with values in $\mathbf{U}_n$ and $\mathbf{Z}_n$ both drawn from the standard Gaussian distribution. This produces data with different levels of dependence amongst and within the SCVs, ultimately to be reflective of the complex dependence relations exhibited for fMRI and other medical imaging modalities.

Once SCVs are generated, sources within each SCV are distributed to their corresponding datasets $\mathbf{S}^{[k]}$, then mixed according to $\mathbf{X}^{[k]} = \mathbf{A}^{[k]}\,\mathbf{S}^{[k]}$ with values in $\mathbf{A}^{[k]}$ drawn from the standard Gaussian distribution. As done in the previous simulations, we run the algorithms on the mixed sources to estimate corresponding $\mathbf{W}^{[k]}$. We use joint-ISI to compare the methods' estimated $\mathbf{W}^{[k]}$ to the true $\mathbf{A}^{[k]}$. Table 3 measures joint-ISI (jISI) and CPU time of the JBSS methods averaged over 50 random variations of $\mathbf{A}^{[k]}$.

We first note here that IVA-G comes the closest to estimating the hybrid SCVs, consistent with IVA-G being considered a "gold standard" of JBSS algorithms exploiting source correlation across datasets. However, IVA-G is disproportionately burdened by high computational costs, well above that of any other algorithm. This is due to the fact that IVA-G has quartic complexity with respect to K (effectively $O(qNK^4)$ in the limit as $K \to \infty$). It is notable that the the standard deviation of CPU time for IVA-G is also extremely large, almost 2 hours for this dataset, which is disproportionately large relative to the mean time. In contrast, all other algorithms tested have very small CPU time standard deviations relative to their mean time. This suggests that the simpler cost functions of the other algorithms provide a significant advantage with respect to the relative consistency in their CPU times.

After IVA-G, ACCM is the second most accurate algorithm, with joint-ISI comparable to IVA-G. It is notable that ACCM can achieve a low joint-ISI despite being an orthogonally constrained algorithm operating on correlated SCVs, where the true $\mathbf{A}^{[k]}$ are relatively far from orthogonal. However, unlike IVA-G, ACCM is among the fastest algorithms, only beaten by SUMCORR in that regard. Whereas IVA-G takes an average of 15.8 hours to estimate the sources, ACCM takes an average of only 1.7 minutes. The other algorithms either trade off worse estimation for better time performance (SUMCORR), or have both worse estimation and time performance (JBSS-SOS, SSQCOR, and MAXVAR). This shows that ACCM is among the fastest JBSS algorithms, yet ACCM can also provide a reliable estimation performance compared with the other efficient JBSS algorithms.

VII. Discussion

These experiments demonstrate that while IVA-G is perhaps the most statistically efficient algorithm exploiting source correlation across datasets, it is also the most costly. Indeed for ICA and IVA, the most statistically efficient algorithms are costlier than simpler orthogonally constrained algorithms.

In the context of optimizing for both statistical efficiency and time of the decomposition, it is beneficial to use a faster orthogonally constrained algorithm as an *initialization* to a more statistically efficient algorithm, such as IVA-G or other variants of IVA. Orthogonally constrained algorithms like ACCM provide at least an inexpensive initial guess that is sufficiently close to the optimal solution, for which more

Table 3. mean ± std. of joint-ISI (jISI) and CPU time (minutes) of JBSS algorithms in estimating the fMRI sources, averaged over 50 random mixtures. SUMCORR, MINVAR, and GENVAR are excluded due to poorer jISI performances.

	IVA-G	ACCM	JBSS-SOS	SSQCOR	MAXVAR
jISI	0.01 ± 8e−5	0.04 ± 9e−12	0.07 ± 1e−3	0.08 ± 7e−3	0.10 ± 2e−15
time	947.3 ± 118	1.7 ± 2e−2	26.3 ± 5.2	5.6 ± 2e−1	8.6 ± 5e−1

statistically efficient algorithms can refine upon. We note that ACCM alone is a quality option if the SCVs are close to uncorrelated or if speed is an important concern.

VIII. Conclusion

This paper introduces an efficient analytic solution to JBSS, derived from the eigendecomposition of the observed data's covariance matrix. Based on assumption that the SCVs are uncorrelated, the proposed ACCM algorithm exploits the block diagonal nature of the SCV cross-covariance matrix to reformulate the problem as solving a system of linear equations from eigenvectors of the observed data's covariance matrix. Thereafter, identifiability conditions of ACCM are derived, and an efficient scheme is proposed to estimate the solution.

The ACCM algorithm is compared with other JBSS algorithms exploiting source correlation across datasets, including the MCCA algorithms, a comparable variant of GJD (JBSS-SOS), and IVA-G. Simulations varying the statistics of the data demonstrate that ACCM is among the most statistically and computationally efficient of the tested algorithms. Performance is then demonstrated on fMRI data in a hybrid experiment, which reiterates these strengths of ACCM.

Finally, we note that performing the ACCM algorithm alone is perhaps most useful when speed is a concern, and a reasonably accurate solution is desired in a very small amount of time. Otherwise, ACCM can be used to initialize a more statistically efficient algorithm such as IVA-G. This initialization is expected to lead to an overall JBSS decomposition that is both statistically and computationally efficient.

References

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[2] X. Yu, D. Hu, and J. Xu, *Blind source separation: theory and applications*. John Wiley & Sons, 2013.

[3] K. Kokkinakis and P. C. Loizou, "Advances in modern blind signal separation algorithms: theory and applications," *Synthesis lectures on algorithms and software in engineering*, vol. 2, no. 1, pp. 1–100, 2010.

[4] T. Adalı, M. Anderson, and G.-S. Fu, "Diversity in Independent Component and Vector Analyses: Identifiability, Algorithms, and Applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, 2014.

[5] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[6] Y.-O. Li, T. Adalı, W. Wang, and V. D. Calhoun, "Joint Blind Source Separation by Multiset Canonical Correlation Analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.

[7] T. Adalı, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal data fusion using source separation: Application to medical imaging," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1494–1506, 2015.

[8] Q. Long, S. Bhinge, V. D. Calhoun, and T. Adalı, "Independent Vector Analysis for common subspace analysis: Application to multi-subject fMRI data yields meaningful subgroups of schizophrenia," *NeuroImage*, vol. 216, p. 116872, 2020.

[9] T. Kim, T. Eltoft, and T.-W. Lee, "Independent Vector Analysis: An extension of ICA to multivariate components," in *International conference on independent component analysis and signal separation*. Springer, 2006, pp. 165–172.

[10] M. Anderson, T. Adalı, and X.-L. Li, "Joint Blind Source Separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2011.

[11] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent Vector Analysis: Identification conditions and performance bounds," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4399–4410, 2014.

[12] H. Hotelling, "Relations between two sets of variates," in *Breakthroughs in statistics*. Springer, 1992, pp. 162–190.

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

[14] 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.

[15] L. C. Parra, "Multi-set Canonical Correlation Analysis simply explained," *arXiv preprint arXiv:1802.03759*, 2018.

[16] T. Nakashima, R. Ikeshita, N. Ono, S. Araki, and T. Nakatani, "Fast online source steering algorithm for tracking single moving source using online independent vector analysis," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[17] D. De Ridder, S. Vanneste, and M. Congedo, "The distressed brain: a group blind source separation analysis on tinnitus," *PloS one*, vol. 6, no. 10, p. e24273, 2011.

[18] D. Sugumar, P. Vanathi, and S. Mohan, "Joint Blind Source Separation algorithms in the separation of non-invasive maternal and fetal ECG," in *2014 International Conference on Electronics and Communication Systems (ICECS)*. IEEE, 2014, pp. 1–6.

[19] L. Li, K. Koishida, and S. Makino, "Online directional speech enhancement using geometrically constrained Independent Vector Analysis." in *INTERSPEECH*, 2020, pp. 61–65.

[20] Z. Boukouvalas, D. C. Elton, P. W. Chung, and M. D. Fuge, "Independent Vector Analysis for data fusion prior to molecular property prediction with machine learning," *arXiv preprint arXiv:1811.00628*, 2018.

[21] J. Vía, M. Anderson, X.-L. Li, and T. Adalı, "Joint blind source separation from second-order statistics: Necessary and sufficient identifiability conditions," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 2520–2523.

[22] D. Lahat and C. Jutten, "An alternative proof for the identifiability of independent vector analysis using second order statistics," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 4363–4367.

[23] B. Gabrielson, M. A. Akhonda, Z. Boukouvalas, S.-J. Kim, and T. Adalı, "ICA with orthogonality constraint: Identifiability and a new efficient algorithm," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3720–3724.

[24] X.-L. Li, T. Adalı, and M. Anderson, "Joint Blind Source Separation by generalized joint diagonalization of cumulant matrices," *Signal Processing*, vol. 91, no. 10, pp. 2314–2322, 2011.

[25] M. Congedo, R. Phlypo, and J. Chatel-Goldman, "Orthogonal and non-orthogonal Joint Blind Source Separation in the least-squares sense," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 1885–1889.

[26] X.-F. Gong, Q.-H. Lin, and K. Wang, "Joint non-orthogonal joint diagonalization based on LU decomposition and Jacobi scheme," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*. IEEE, 2013, pp. 25–29.

[27] X.-F. Gong, X.-L. Wang, and Q.-H. Lin, "Generalized non-orthogonal joint diagonalization with lu decomposition and successive rotations," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1322–1334, 2015.

[28] X.-F. Gong, L. Mao, Y.-L. Liu, and Q.-H. Lin, "A Jacobi generalized orthogonal joint diagonalization algorithm for Joint Blind Source Separation," *IEEE Access*, vol. 6, pp. 38 464–38 474, 2018.

[29] J. B. Keller, "Closest Unitary, Orthogonal and Hermitian operators to a given operator," *Mathematics Magazine*, vol. 48, pp. 192–197, 1975.

[30] S. M. Smith, A. Hyvärinen, G. Varoquaux, K. L. Miller, and C. F. Beckmann, "Group-PCA for very large fMRI datasets," *Neuroimage*, vol. 101, pp. 738–749, 2014.

[31] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *IEE proceedings F (radar and signal processing)*, vol. 140, no. 6. IET, 1993, pp. 362–370.

[32] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on signal processing*, vol. 45, no. 2, pp. 434–444, 1997.

[33] S. A. Amari *et al.*, "Advances in neural information processing systems," in *Advances in neural information processing systems*, vol. 8. Cambridge, MA: MIT Press, 1996, p. 757–763.

[34] R. L. Gollub *et al.*, "The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia," *Neuroinformatics*, vol. 11, no. 3, pp. 367–388, 2013.