HHS Public Access

Author manuscript

IEEE J Sel Top Signal Process. Author manuscript; available in PMC 2017 October 01.

Published in final edited form as:

IEEE J Sel Top Signal Process. 2016 October ; 10(7): 1134–1149. doi:10.1109/JSTSP.2016.2594945.

Blind Source Separation for Unimodal and Multimodal Brain Networks: A Unifying Framework for Subspace Modeling

Rogers F. Silva[*] **[Student Member, IEEE]**,
Dept. of ECE at The University of New Mexico, NM USA

The Mind Research Network, LBERI, Albuquerque, New Mexico USA

Sergey M. Plis,
The Mind Research Network, LBERI, Albuquerque, New Mexico USA

Jing Sui [Senior Member, IEEE],
Brainnetome Center & NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing China

The Mind Research Network, LBERI, Albuquerque, New Mexico USA

Marios S. Pattichis [Senior Member, IEEE],
Calhoun are with the Dept. of ECE at The University of New Mexico, NM USA

Tülay Adalı [Fellow, IEEE], and
Dept. of CSEE, University of Maryland Baltimore County, Baltimore, Maryland USA

Vince D. Calhoun [Fellow, IEEE]
Dept. of ECE at The University of New Mexico, NM USAThe Mind Research Network, LBERI, Albuquerque, New Mexico USA

Abstract

In the past decade, numerous advances in the study of the human brain were fostered by successful applications of blind source separation (BSS) methods to a wide range of imaging modalities. The main focus has been on extracting “networks” represented as the underlying latent sources. While the broad success in learning latent representations from multiple datasets has promoted the wide presence of BSS in modern neuroscience, it also introduced a wide variety of objective functions, underlying graphical structures, and parameter constraints for each method. Such diversity, combined with a host of datatype-specific know-how, can cause a sense of disorder and confusion, hampering a practitioner’s judgment and impeding further development.

We organize the diverse landscape of BSS models by exposing its key features and combining them to establish a novel unifying view of the area. In the process, we unveil important connections among models according to their properties and subspace structures. Consequently, a high-level descriptive structure is exposed, ultimately helping practitioners select the right model

[*]Corresponding author. R.F. Silva (rsilva@mrn.org).

for their applications. Equipped with that knowledge, we review the current state of BSS applications to neuroimaging.

The gained insight into model connections elicits a broader sense of generalization, highlighting several directions for model development. In light of that, we discuss emerging multi-dataset multidimensional (MDM) models and summarize their benefits for the study of the healthy brain and disease-related changes.

Index Terms

BSS; neuroimaging; unimodal; multimodality; multiset data analysis; overview; subspace; unify; modeling

I. Introduction

Blind source separation (BSS) methods [1] have been widely used in the study of the brain. They can be adapted and made compatible with a number of brain imaging, genetics, and non-imaging modalities, fueling applications to neurophysiological measurements such as magnetic resonance imaging (MRI)—both structural (sMRI) [2] and functional (fMRI) [3], [4]—magnetoencephalography (MEG) [5], electroencephalography (EEG) [6], diffusion weighted MRI (DWI) [7], copy-number variation (CNV) [8], single nucleotide polymorphism (SNP) [9], methylation [10], [11], metabolomics [12], and questionaires [13], [14] among others. Continued technological advancements in brain structure and function assessment [15]–[18] have fostered the development of a growing collection of BSS methods and tools. Their applications range from the study of human brain deficits associated with certain mental disorders to the neurophysiological associations with cognitive and behavioral measures.

Despite such wide application range, selecting the right BSS tool for a given problem—or developing a new one—can quickly turn into a daunting task if an understanding of the underlying structure of the area is missing. Our work addresses this issue. It highlights fundamental connections among BSS models and offers a novel, intuitive taxonomy, organizing BSS problems into specialized subproblems. Also, our investigation of various assumptions embedded within BSS models refines this structure, revealing hidden connections and differences.

To achieve that, we let BSS refer to any method for *simultaneous* model (or system) inversion of one or multiple datasets that uses only the observed measurements (or outputs). Then, we put a number of methods originating from different areas in perspective, focusing on three key properties: the number of datasets allowed, the grouping of sources within a dataset, and the use of second-, higher-, or all-order statistics. As a result, unanticipated connections and differences among seemingly unrelated methods are revealed, culminating in a new unified framework for BSS model selection and development. A hierarchy of increasing model complexity ensues, providing a sensible guide to practitioners in their domain-specific contexts. Model weaknesses and strengths, as well as key differences and shared features, stand out, exposing a high-level descriptive structure useful for researchers

either interested in pursuing new, unexplored directions, or simply trying to identify candidate models for an application. These properties are finally summarized in a set of systematic, yet simple, modeling choices, exposing new insight into future directions for the area and advancing a novel unifying view.

Our unifying framework describes a new, largely under-investigated class of problems, paving a way for development of new models. These models are anticipated to be highly flexible, combining and expanding key features and subspace structures from existing models. We expect that these models will play a key role in neuroimaging research and improve how we understand the intricacies of the human brain, laying out a promising new path for future research developments that could easily extend beyond neuroimaging applications.

We introduce the unifying framework in section II. We then describe how the unifying framework specializes to different models and methods in section III. We discuss future directions in section IV and provide concluding remarks in section V.

II. A Unified Framework for Subspace Modeling and Development

The wide range of applications utilizing BSS methods to capture and analyze brain networks requires a fairly broad understanding of the area by the average researcher. The terminology and notation from different fields where methods have originally been developed makes it an even harder task. Not surprisingly, the use of a BSS tool for a certain application is often unintentionally limited to what has already been applied to a certain datatype or disease. Our belief is that users and developers of BSS tools can largely benefit from a clear and intuitive description of the underlying structure among BSS methods. This is our guiding motivation throughout.

A. The Structure of BSS Problems

The BSS problem is broadly defined as “recovering unobservable source signals **s** from measurements **x** (i.e., data), with no knowledge of the parameters $\boldsymbol{\theta}$ of the generative system $\mathbf{x} = \mathbf{f}(\mathbf{s}, \boldsymbol{\theta})$.” It can be organized into subproblems depending on whether **x** contains either single or multiple datasets, and whether or not subsets of **s** (within the same dataset) group together to form one or more multidimensional sources. We propose a novel taxonomy to define four general BSS subproblems, as follows:

1. Single dataset unidimensional (SDU): **x** consists of a single dataset whose sources are not grouped, e.g., independent component analysis (ICA) [19]–[21] and second order blind identification (SOBI) [22], [23], as in section III-A1;
2. Multiple dataset unidimensional (MDU): **x** consists of one or more datasets but no multidimensional sources occur within any of the datasets, although multidimensional sources containing a single source from each dataset may occur, e.g., canonical correlation analysis (CCA) [24], partial least squares (PLS) [25], and independent vector analysis (IVA) [26], [27], discussed in III-A2;

3. Single dataset multidimensional (SDM): **x** consists of a single dataset with one or more multidimensional sources, e.g., multidimensional independent component analysis (MICA) [28], [29] and independent subspace analysis (ISA) [30], [31], discussed in III-A3;

4. Multiple dataset multidimensional (MDM): **x** contains one or more datasets, each containing one or more multidimensional sources that may group further with single or multidimensional sources from the remaining datasets, e.g., multidataset independent subspace analysis (MISA) [32], [33] and joint independent subspace analysis (JISA) [34], discussed in III-A4.

Under these definitions, subproblems are contained within each other, as described in Fig. 1, revealing a natural hierarchical structure among them. Accordingly, the *generative* models describing data generation from sources should follow the same hierarchy. Problem specification, therefore, contributes to the description of **f** (·) itself and some properties of its parameters (***θ***) and inputs (**s**) in generative models. This provides a new perspective about basic connections among models following from problem specification.

In the presence of noise, **x** = **f** (**s**, ***θ***) + **e**, where the sensor noise **e** is typically Gaussian. Noisy system inversion has been well studied, especially in linear systems, with results such as the Wiener filter, which defines a system inversion $\hat{\mathbf{s}} = \mathbf{g}(\mathbf{s}, \boldsymbol{\theta}, \mathbf{x})$ that minimizes $\|\mathrm{s} - \hat{\mathrm{s}}\|_2$ for known **s** and ***θ*** Source estimation strategies for noisy cases are different (see section II-D) and largely tailored to SDU problems. Still, noise-free BSS models are often fairly robust to noise.

B. Assumptions that Drive Model Hierarchy

A generative BSS model is completely defined only when the parameters (***θ***) and source signals (**s**) are fully described according to three sets of assumptions. The leading assumption is the presence of latent source signals in the data. Additional assumptions act to counter ill-conditions, allowing sources to be identified from data. These assumptions, combined with the problem hierarchy, induce a set of basic modeling choices, as described below and summarized in Table I.

Presence of sources—Given N observations of $M \geq 1$ datasets, we wish to identify an unobservable latent source random vector (r.v.) $\mathrm{s}=\left[\mathrm{s}_1^T \cdots \mathrm{s}_M^T\right]^T, \mathrm{s}_m=\left[s_1 \cdots s_{C_m}\right]^T$ that relates to the observed r.v. $\mathrm{x}=\left[\mathrm{x}_1^T \cdots \mathrm{x}_M^T\right]^T, \mathrm{x}_m=\left[x_1 \cdots x_{V_m}\right]^T$ via a vector function **f** (**s**, ***θ***). Learning both **s** and **f** (·, ***θ***) blindly—without prior knowledge of either of them—requires choosing the number of sources C_m in each dataset, with compound $\overline{C}=\sum_{m=1}^{M} C_m$ and $\overline{V}=\sum_{m=1}^{M} V_m$.

Mixture function—The assumptions imposed on the vector function **f** (·, ***θ***) are characterized by the following properties:

1. **Mapping.** Mixture functions are mappings of either a *linear* or *non-linear* type. A linear function is any linear transformation by square or rectangular matrices (**A**) with $\mathbf{x}=\mathbf{f}(\mathbf{s},\boldsymbol{\theta})=\mathbf{A}\mathbf{s}=\sum_{j=1}^{\overline{C}}\mathbf{a}_j s_j$, and $\boldsymbol{\theta}=\mathbf{A}=\left[\mathbf{a}_1\cdots\mathbf{a}_{\overline{C}}\right]$. For quadratic functions, $\mathbf{x}=\mathbf{f}(\mathbf{s},\mathbf{A},\mathbf{Q}_{v,\ v=1..\overline{V}})=\mathbf{A}\mathbf{s}+\mathbf{z}$, $\mathbf{z}=\left[z_1\cdots z_{\overline{V}}\right]^T$, $z_v=\mathbf{s}^T\mathbf{Q}_v\mathbf{s}$.

2. **Layout.** The layout is a cross-dataset indication of which elements of **s** generate which elements of **x**. It is *fully-connected* if **s** in all datasets contribute to all **x**. Otherwise, it is *structured*, such as in separable models where **f**() is either shared (e.g., same **A**) or dataset-specific (e.g., block-diagonal **A**). Tensor models, such as parallel factors (PARAFAC), offer further structured layout subtypes [35].

Statistical relationship among sources—This assumption regards *which* sources are (un)related, and *how*. Only stochastic relationships are considered since all deterministic relationships can be absorbed in **f**(·). Sources can also be related to themselves through sample dependence (e.g., autocorrelation).

1. **Subsets.** This choice determines *which* sources are related to each other and which are not. It leads to the notion of source *groups* ($\mathbf{s}_k$, $k = 1,\ldots, K$), i.e., related sources that go together to form one of K groups, which are broadly referred to as *subspaces*[1]. The number of subspaces (K) and their compositions are key choices in S/MDM problems.

2. **Interactions**. This specifies directed (possibly causal) and/or undirected relationships among sources. Many *graphical* structures [36] are possible, including "self-loops" to represent *sample dependence*. For brevity, this work emphasizes either undirected non-causal instantaneous relations, or those based on sample dependence.

3. **Type of statistics.** When random variables are statistically related they are said to be *dependent*. There are two types of statistical dependence to choose from: *linear* dependence (captured by second-order statistics), and *higher-order* dependence (captured by higher-order statistics). Unrelated variables, however, are statistically *independent*, meaning their joint distribution is a product of the marginal distributions, e.g., $p(\mathbf{s})=\prod_{i=1}^{C}p(s_i)$. Independence implies no grouping and, thus, sources do not interact. Consequently, both second- and higher-order dependencies are absent in that case.

   Second-order statistics (SOS): This kind of statistics refers to so-called second moments. The second moment of a zero-mean random variable is $E[X^2]$, where $E[\ ]$ is the expected value operator. It captures the "scale" (or variability) of X in the form of variance or standard deviation. For pairs of variables, however, second moments become cross-moments $E[X_1, X_2]$, which capture the level of *linear* relation between random variables in the form of a *correlation coefficient*

[1]The subspace terminology stems from the linear mixture function case [28] in which the columns of **A** corresponding to $\mathbf{s}_k$ form a linear (sub)space.

(ρ). In simple terms, SOS measures how well a straight line explains the joint statistical relationship between two random variables. Independence implies all SOS cross-moments are zero for zero-mean sources and, therefore, uncorrelation.

Higher-order statistics (HOS): This kind of statistics refers to higher-order moments. A higher-order moment of a zero-mean random variable is the expected-value of its p-th power, $E[X^p]$, for $p > 2$. It captures other properties of the distribution of X, like skewness ($p = 3$) and kurtosis ($p = 4$). For tuples of C variables, however, higher-order moments become higher-order cross-moments $E_p\left[X_1^{q_1}, \ldots, X_C^{q_C}\right], \sum_{i=1}^{C} q_i = p$, $C \leq p$; $E_p[\cdot]$ is a p-order expectation. Put simply, HOS measures joint relationships among multiple random variables beyond those explained by a single straight line. Independence implies all HOS cross-moments are zero for zero-mean sources.

SOS and HOS are at the core of statistical source relationship modeling, broadly compartmentalizing the landscape of models into those capturing only one or both types of statistics. The latter case is considered more general, as depicted in Fig. 2. Often, all-order statistical information (i.e., both SOS and HOS) can be achieved by simply choosing an effective joint probability density model for the sources/subspaces.

C. Terminology and Indeterminacies

We intentionally generalize the definition of BSS to include cases of *simultaneous* system inversion, which attempt to leverage information from multiple systems (via their outputs) to *jointly* infer their mixture functions and sources. Utilizing multiple datasets to infer each system may lead one to interpret the methods not as “blind” as a single-dataset BSS. However, our view is that, together, all these systems/datasets form a single larger system that needs to be identified only by the outputs provided by each of its parts. This view aligns with prior work coined as joint BSS [37], and even more so with multimodal brain data analysis.

Like seminal works from the SDU literature [38], certain restrictions apply with respect to the identifiability of SDM, SDU, and MDM models. In general, this is a result of how well the modeling choices (see Table I) match the true generative process. For example, the typical scale, sign, and permutation ambiguities from the linear independence-based SDU models [1] generalize to arbitrary invertible linear transformations of the subspaces in SDM models [28]. In linear second-order independence-based MDU models, identifiability is not attainable in cases were two or more subspaces share an identical block correlation structure [39], [40]. Recent initial work [41] developed for linear second-order independence-based MDM models suggests a combination of conditions from SDM and MDU. In the non-linear case, however, independence-based SDU models are unidentifiable if independence is the only assumption [1, Ch.14]. Additional constraints are required and Bayesian approaches offer a nice framework for that. A complete study of the identifiability conditions for each combination of modeling choices, especially in the case of MDM models, is valuable but exceeds the scope of this work. Note, however, that for noisy linear models optimizing for minimal error in $\mathbf{s}$ implies $\mathbf{WA} \neq \mathbf{I}$, i.e., $\mathbf{W}$ is not identifiable. Conversely, identifiability of

W does not guarantee identifiability of **s** in noisy models [1, Ch.4]. Finally, the number of components C can be estimated by model order selection approaches based on information-theoretic criteria (IC) such as Akaike's IC (AIC) [42], Bayesian IC (BIC) [43], Kullback-Leibler IC (KIC) [44], Draper's IC [45], or minimum description length (MDL) [46]. However, we are unaware of methods for direct estimation of the number of subspaces K and their compositions, except for post-hoc approaches using clustering [47] and non-linear correlations [31]. Also, these approaches do not generalize trivially to multiple datasets.

D. Turning Models into Algorithms

After characterizing the model based on the choices outlined above, three additional steps are typically required to translate it into an algorithm (Fig. 3): (i) define an inverse model for $\mathbf{s} \approx \mathbf{y} = \mathbf{g}(\mathbf{x}; \boldsymbol{\phi})$ based on the modeling choices made for $\mathbf{f}(\cdot)$, **s** and $\boldsymbol{\theta}$, where $\boldsymbol{\phi}$ denotes a system inversion parameter, (ii) select a cost function $\mathcal{J}(\boldsymbol{\phi})$ that is sensitive to the properties of **s**, and (iii) choose an optimization procedure to estimate $\boldsymbol{\phi}$ by minimizing/maximizing $\mathcal{J}(\boldsymbol{\phi})$.

The inverse model results directly from the choices in the generative model. In noise-free linear BSS models, the inverse model must also be linear, in which case $\boldsymbol{\phi}$ is also a matrix, denoted **W**, and $\mathbf{y} = \mathbf{W}\mathbf{x}$. The cost function is typically one that reflects the desired properties of **s**, i.e., it attains a minimum or a maximum when such properties are met (e.g. the likelihood function). It also changes with the structure of **A**. Thus, a numerical optimization strategy [48] is selected to identify the optimal **W** that approximately attains such minimum/maximum from data. Typical unconstrained optimization algorithms include line search methods, using regular, stochastic, relative [49], or natural [50] gradient descent, Newton or quasi-Newton descent [48], as well as trust-region methods [48]. Some variations allow for constraints on $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and/or **s** (Table I), such as regularized optimization, null-space methods [51], interior point [48], and multiobjective [52] optimization. For example, sparsity constraints on $\boldsymbol{\theta} = \mathbf{A}$ reduce the number of sources contributing to a sensor, while low total variation of **s** has a smoothing and denoising effect.

Numerical optimization is key for Bayesian estimation [1, Ch.12] as well. Using priors, Bayesian methods infer the *posterior* distribution from the likelihood function. This offers a nice, principled strategy to incorporate constraints as prior knowledge. Then, maximization of the posterior (MAP) jointly estimates **s** and $\boldsymbol{\theta}$ from **x**. Alternatively, maximization after marginalizing **s** out of the posterior is thought to prevent overfitting by avoiding arbitrary peaks of the joint posterior, leading to a better estimate of the posterior mass concentration. Variational Bayes (VB) approaches, such as ensemble learning [53], [54], carry out marginalization while approximating the joint posterior by a product of factors, each corresponding to a subset of the model (and noise) parameters $\boldsymbol{\theta}$. Marginalization can also simplify to the likelihood function when a "constant" prior is used for $\boldsymbol{\theta}$. When no analytical approximation is imposed on this likelihood, a general expectation-maximization (EM) approach emerges [55]: conditional expectations of **s** are computed in the E-step and used to update $\boldsymbol{\theta}$ in the M-step (maximizing the likelihood). This was shown [56] to be particularly useful in the case of noisy models, using the Wiener filter in the E-step. Marginalization of $\boldsymbol{\theta}$ following marginalization of **s** requires MCMC approaches for approximate integration.

E. Unified Framework

Building on the broad, general view of modeling and development in BSS problems presented above, we propose a unified framework for description of BSS models based on charting the modeling choices made for each of the three sets of assumptions outlined in II-B and the type of problem being addressed. In the following section, we illustrate the scope of this unifying framework by characterizing and reviewing a number of BSS models popular for brain data analysis.

III. Review of BSS Models for Brain Data

Following the proposed framework, in this section we review BSS models and algorithms frequently employed in brain data analysis. Throughout the description, models become progressively stronger by adopting more lenient choices, enabling simultaneous analysis of multiple datasets, allowing multidimensional sources, and exploiting both second- and higher-order statistics. High-level differences and relationships among these models become evident, as shown below. Applications to various brain data modalities are presented last.

A. BSS Models

Here, both classical algorithms and recent models are reviewed through the framework we propose. Following Table I, we first focus our attention on linear mixture models ($\mathbf{x} = \mathbf{A}\mathbf{s}$) with undirected source relationships within subspaces and/or sample dependence, and no optional constraints. In all cases, we assume that sources are zero-mean. The sections are organized by problem, following the structure in Fig. 1.

1) Single Dataset Unidimensional (SDU) Problem—The linear models for the SDU problem assume a single dataset ($M = 1$) generated by an invertible linear, fully-connected mixture (see Fig. 4). The classical model in this case is principal component analysis (PCA) [57]–[60], which assumes the sources $\mathbf{s}$ are uncorrelated, implying a diagonal source covariance matrix $\mathbf{\Sigma}^{\mathbf{s}}$. Thus, SOS dependence among estimated sources $\mathbf{y} = \mathbf{W}\mathbf{x}$ should be zero. Eigenvalue decomposition (EVD) of $\mathbf{\Sigma}^{\mathbf{x}}$ or singular value decomposition (SVD) of the observed $V \times N$ dataset $\mathbf{X}$ identify such diagonalizing $\mathbf{W}$, assigning sources to each of the principal axes of the data, i.e., the (orthogonal) directions of maximal variability. The PCA solution is unique only up to sign ambiguities. Also, sources with equal variance are unrecoverable if $\mathbf{A}$ is a rotation matrix, since $\mathbf{x}$ would then be already uncorrelated. Its focus on the top C sources with largest variability makes it prone to error when the sources of interest have low variance.

A very successful model for brain data in this category is independent component analysis (ICA) [3], [7]–[9], [61]–[70]. The ICA model assumes statistical independence among the sources in $\mathbf{s}$ and, thus, all HOS and SOS dependence among estimated sources should be zero. Cost functions for dependence assessment are abundant [1], [71]–[73], including negentropy [19] (as in FastICA [21], [74]) and information divergence [75], [76], of which mutual information [26], [77], [78] is an important special case and a natural, general choice:

$$I(\mathbf{s}) = -h(\mathbf{s}) + \sum_{i=1}^{C} h(s_i), \quad (1)$$

where $h(\mathbf{s}) = -E[\log p(\mathbf{s})]$ is the joint differential entropy, and $p(\mathbf{s})$ and $p(s_i)$ are the joint probability density function (pdf) of all sources and the marginal pdf, respectively. When $\mathbf{u} \approx \gamma(\mathbf{y}) = \gamma(\mathbf{W}\mathbf{x})$, where $\gamma(\cdot)$ is an element-wise non-linear transformation representing a fixed cumulative distribution function (cdf), and $p(u_i)$ is the Uniform distribution in the [0, 1] domain, $h(u_i)$ is zero. In this case, minimizing $I(\mathbf{u})$ is equivalent to maximizing $h(\mathbf{u})$, which is the popular Infomax approach [20] when the sigmoid function is chosen as the cdf.

Another successful method, particularly in the field of EEG-based brain-computer interface (BCI), is called second-order blind identification (SOBI) [22], [79]. SOBI considers the case of independent non-white stochastic source processes $s_i(t)$, in which sample dependence occurs within each source s_i through autocorrelation $\mathbf{R}^{\mathbf{s}}_{ii}$ (e.g., lag-correlation $\mathbf{R}^{\mathbf{s}}_{ii}(\tau)$, $\tau \in [0, T]$) but not highly among different sources (e.g., $\mathbf{R}^{\mathbf{s}}_{ij}(\tau) \approx 0$, $i \neq j$, $\forall \tau \in [0, T]$). This translates to high SOS of a source with itself and low SOS with other sources, i.e, $\mathbf{R}^{\mathbf{s}}(\tau) \approx \mathbf{I}$ for all τ. Finding $\mathbf{W}$ that simultaneously diagonalizes $\mathbf{R}^{\mathbf{x}}(\tau)$ for all τ is known as joint diagonalization (JD). This principle is shared by many other techniques, such as the algorithm for multiple unknown signals extraction (AMUSE) [80], which pursues *exact*[2] JD between $\mathbf{R}^{\mathbf{x}}(0)$ and $\mathbf{R}^{\mathbf{x}}(\tau)$, for fixed $\tau \neq 0$, and the time-delays based separation (TDSEP) [81], which uses exact JD on two weighted averages of multiple $\mathbf{R}^{\mathbf{x}}(\tau)$. Weights-adjusted SOBI (WASOBI) [23] is a variant implementation of the SOBI model that attains better performance by optimally weighting each $\mathbf{R}^{\mathbf{x}}(\tau)$ and performing *approximate* JD via weighted least-squares.

Seeking to combine the ICA and SOBI models, COMBI [82] proposes a “wrapper” algorithm that runs both efficient FastICA (EFICA) [83] and WASOBI, selecting the “best” components from each [84]. Recent developments utilizing mutual information *rate* to combine ICA and SOBI principles have inspired unique algorithms like entropy rate bound minimization (ERBM) [85] and entropy rate minimization using an autoregressive (AR) source model driven by generalized Gaussian distribution (ERM-ARG) [86]. The use of both sample correlation and HOS of sources provides significantly better performance [26], although these have not yet been applied in BCI.

The last model we consider in this category is a simple but effective one, called group ICA (GICA) [4], [87]. GICA is a popular approach for group analysis of resting-state fMRI data [88] that implements “spatial ICA.” An actively growing research area called dynamic resting-state functional network connectivity (rs-FNC) [89]–[92] analysis has been built on top of results produced with this approach. GICA models several datasets as if they were one (by temporal concatenation), assuming the same single set of sources $\mathbf{s}$ (the “aggregate” spatial maps) for all M datasets (the subjects): $\mathbf{x}_m = \mathbf{A}_m\mathbf{s}_m$, $\mathbf{s}_1 = \cdots = \mathbf{s}_M = \mathbf{s}$. Dataset- or

[2]Exact JD is only attained in the case of two symmetric matrices.

subject-specific variations in **s** are then captured by a process called back-reconstruction [93]. While effective, this assumption is also somewhat restrictive and recent work suggests that MDU-type models may capture subject-specific source variability better than GICA when $\mathbf{s}_m$ are highly distinct [26], [94] across datasets.

The class of SDU models includes those that extend beyond noise-free linear unconstraint approaches. We go over some of them. Factor analysis (FA) assumes Gaussian sources with uncorrelated sensor noise [95]. Probabilistic ICA (PICA) [96] is a FastICA approach that utilizes a preprocessing step (namely, standardize each observation $\mathbf{x}_n$ to zero mean and unit variance) to condition the data prior to PCA reduction. This is based on an extended FA [38] to select the number of components, assuming sensor noise is very low. Expanding Bayesian methods for linear ICA [97]–[99], non-linear FA (NFA) and non-linear independent FA (NIFA) approaches [100], [101] have been proposed under a Bayesian framework too.

In tensor ICA (TICA) [102], the PARAFAC tensor model $\mathbf{x}_m = \mathbf{A}\mathbf{D}_m\mathbf{s}$, $m = 1M$, where $\mathbf{D}_m$ is a diagonal matrix, is used to achieve a variation of GICA. Identically to GICA on M datasets, shared sources **s** are recovered using PICA. However, $\mathbf{A}_m = \mathbf{A}\mathbf{D}_m$ is obtained in a post-processing step of the concatenated mixing matrices. Thus, the columns of $\mathbf{A}_m$ are *perfectly* correlated across datasets (subjects). While this can be useful for task-elicited activation in which the task stimulus timing is identical, in the case of spontaneous activity, such as resting fMRI, a model that does not assume perfectly correlated timecourses across datasets (subjects) is preferred.

The use of tensor representations is also common in cumulant-based SDU methods [103], but does not relate to modeling datasets as tensors. Instead, they seek to diagonalize second- and fourth-order cumulant tensors by applying the same transformation matrix **W** to all dimensions of the tensor, effectively achieving second- and fourth-order independence by driving cross-moments toward zero. As for constraints, ICA-R [104]–[106] was proposed to allow references to be incorporated and help extract certain specific patterns from data while pursuing independence. The reference “guides” the decomposition, potentially overcoming high noise issues.

2) Multiple Dataset Unidimensional Problem—MDU-type linear models assume $M \geq 1$ datasets generated by a linear, structured mixture. As such, the combined mixing matrix **A** has a block-diagonal structure that confines sources to their respective datasets: $\mathbf{x}_m = \mathbf{A}_m\mathbf{s}_m$, $m = 1 \ldots M$. The number of sources is typically the same for all datasets ($C_m = C$). Although no multidimensional (grouped) sources occur within any single dataset, sources are allowed to group across datasets (see Fig. 4). The properties of these $K = C$ cross-dataset M-dimensional source groupings (or subspaces, $\mathbf{s}_k$) mark the major differences among models in this category, particularly their choice of SOS, HOS, or both to describe (un)relatedness.

SOS-only models underly classical algorithms such as canonical correlation analysis (CCA) [24] and partial least squares (PLS) [25], [107], as well as more recent models such as multiset CCA (mCCA) [108] and second-order independent vector analysis (IVA) [26], [27]. Thus, in these models, unrelated sources are linearly independent (uncorrelated) and related sources are linearly dependent (correlated). In the case of CCA and PLS, exactly $M = 2$

datasets are considered, meaning only corresponding sources from each dataset form related pairs. CCA seeks a particular solution that maximizes the *correlation* between related source pairs $\mathbf{s}_{k=i} = [s_{1i}, s_{2i}]^T$, $s_{1i} = \mathbf{W}_{1i}\mathbf{x}_1$ and $s_{2i} = \mathbf{W}_{2i}\mathbf{x}_2$, $i = 1 \ldots C$, where $\mathbf{W}_{mi}$ is the i-th row of $\mathbf{W}_m$, while PLS maximizes their *covariance* instead. For $i = 1$, CCA solves the following constrained optimization [109],

$$\begin{array}{ll} \underset{\mathbf{W}_{1i},\mathbf{W}_{2i}}{max} & \mathbf{W}_{1i}\sum_{12}^{\mathbf{x}}\mathbf{W}_{2i}^T \\ s.t. & \mathbf{W}_{mi}\sum_{mm}^{\mathbf{x}}\mathbf{W}_{mi}^T=1, \forall m \in \{1,2\}\,, \end{array} \quad (2)$$

where $\sum_{ml}^{\mathbf{x}}$ is the (cross-) covariance matrix between datasets $\mathbf{x}_m$ and $\mathbf{x}_l$, while PLS solves a different constraint [109],

$$\begin{array}{ll} \underset{\mathbf{W}_{1i},\mathbf{W}_{2i}}{max} & \mathbf{W}_{1i}\sum_{12}^{\mathbf{x}}\mathbf{W}_{2i}^T \\ s.t. & \mathbf{W}_{mi}\mathbf{W}_{mi}^T=1, \forall m \in \{1,2\}\,. \end{array} \quad (3)$$

In either case, by Lagrange multipliers, the constrained optimizations reduce to generalized eigenvalue (GEV) problems $\mathbf{E}\mathbf{v}_i = \lambda\mathbf{F}\mathbf{v}_i$, where

$$\mathbf{E}=\begin{bmatrix} \mathbf{0} & \sum_{12}^{\mathbf{x}} \\ \sum_{21}^{\mathbf{x}} & \mathbf{0} \end{bmatrix},\ \mathbf{F}_{CCA}=\begin{bmatrix} \sum_{11}^{\mathbf{x}} & \mathbf{0} \\ \mathbf{0} & \sum_{22}^{\mathbf{x}} \end{bmatrix},\ \mathbf{v}_i=\begin{bmatrix} \mathbf{W}_{1i}^T \\ \mathbf{W}_{2i}^T \end{bmatrix},$$

for CCA[3], and $\mathbf{F}_{PLS} = \mathbf{I}$ for PLS, both solvable by the power method (PM) of numerical linear algebra. The structure of $\mathbf{F}_{PLS}$ implies $\mathbf{W}_{11}$ and $\mathbf{W}_{21}$ are the left and right singular vectors of the largest singular value of $\sum_{12}^{\mathbf{x}}$, respectively [109]. It also enables a popular variant of PM called non-linear iteration partial least squares (NIPALS) [111], which converges in a single step when either $\mathbf{x}_1$ or $\mathbf{x}_2$ is univariate ($V_m = 1$).

For $i > 1$, CCA requires additional constraints to enforce diagonal structure on source covariances $\sum_{11}^{\mathbf{s}}$, $\sum_{22}^{\mathbf{s}}$, and $\sum_{12}^{\mathbf{s}}$

$$\sum_{ml,j\bar{j}}^{\mathbf{s}}=\mathbf{W}_{mj}\sum_{ml}^{\mathbf{x}}\mathbf{W}_{l\bar{j}}^T=0, 1 \le \bar{j}<j \le i, \forall m,l \in \{1,2\}\,,$$

meaning all non-corresponding source pairs are uncorrelated. While sequential estimation in CCA simply yields the remaining eigenvectors of the GEV problem [112], the same is not always the case in PLS [109]. This is because different deflation strategies can be employed between iterations i and i+1, giving rise to a wide range of PLS variants [109], [110], [113], [114]:

Name	Deflation

[3]CCA is sometimes referred to as PLS Mode B [110].

PLS-SVD

$$\mathbf{X}_m^{i+1}=\mathbf{X}_m^i-\mathbf{W}_{mi}^T\mathbf{W}_{mi}\mathbf{X}_m^i$$

PLS Mode A

$$\mathbf{X}_m^{i+1}=\mathbf{X}_m^i-\mathbf{X}_m^i\mathbf{B}_m^i$$

PLS1/PLS2

$$\mathbf{X}_1^{i+1}=\mathbf{X}_1^i-\mathbf{X}_1^i\mathbf{B}_1^i,\ \mathbf{X}_2^{i+1}=\mathbf{X}_2^i-\mathbf{X}_2^i\mathbf{B}_1^i$$

where $\mathbf{B}_m^i=\mathbf{X}_m^{iT}\mathbf{W}_{mi}^T\left(\mathbf{W}_{mi}\mathbf{X}_m^i\mathbf{X}_m^{iT}\mathbf{W}_{mi}^T\right)^{-1}\mathbf{W}_{mi}\mathbf{X}_m^i$, $\mathbf{X}_m^i$ is the m-th $V_m \times N$ data matrix, containing N observations of the measurement vector $\mathbf{x}_m$ at the i-th iteration, and $\mathbf{X}_m^1=\mathbf{X}_m$. Alternating the PLS-SVD deflation with the basic PLS optimization in (3), for $i > 1$, is equivalent to sequential rank-one deflations of $\sum_{12}^{\mathbf{x}}$ [109], or to additional orthogonal constraints $\mathbf{W}_{mj}\mathbf{W}_{m\bar{j}}^T=0, 1\le\bar{j}<j\le i$, $\forall m\in\{1, 2\}$, all simply yielding the remaining eigenvectors of the initial GEV problem. As a result, non-corresponding source pairs are uncorrelated in PLS-SVD (as in CCA). This is generally not the case in PLS Mode A, although sources are uncorrelated to each other within the same dataset [110] (since the residual $\mathbf{X}_m^{i+1}$ is orthogonal to $\mathbf{X}_m^i$). The deflation in PLS Mode A is inspired by least squares [115], seeking to remove the variability in $\mathbf{X}_m^i$ explained by $\mathbf{S}_{mi}^i=\mathbf{W}_{mi}\mathbf{X}_m^i$ (the m-th $1 \times N$ vector containing N observations of the source s_{mi}) at the i-th iteration. PLS-SVD and PLS Mode A are exploratory approaches recommended for modeling unobservable relationships between datasets. PLS1 and PLS2, however, are intended for regression, univariate ($V_2 = 1$) and multivariate ($V_2 > 1$), respectively [113], with $\mathbf{X}_2$ acting as the response variable and using the information contained in $\mathbf{X}_1$ to make predictions about $\mathbf{X}_2$. To that end, s_{1i} is assumed to be a good surrogate for s_{2i} [113], thus replacing it in the deflation of $\mathbf{X}_2^i$. After C deflations, a "global" multiple regression equation is derived which can be used for prediction of $\mathbf{X}_2$ from new observations of $\mathbf{X}_1$ [109], [112]:

$$\mathbf{x}_{2,\text{pred}}=\mathbf{P}_2(\mathbf{W}_1\mathbf{P}_1)^{-1}\mathbf{W}_1\mathbf{x}_{1,\text{pred}},$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are matrices whose columns are the projections $\mathbf{p}_{1i}=\mathbf{X}_1^i\mathbf{B}_1^i$ and $\mathbf{p}_{2i}=\mathbf{X}_2^i\mathbf{B}_1^i$, respectively, $1\le i\le C$, and $\mathbf{W}_1$ contains $\mathbf{W}_{1i}$ at the i-th row.

Overall, for sources $\mathbf{s}$ estimated with a known inverse model $\mathbf{g}(\cdot)$, [116] suggests that the generative model $\mathbf{f}()$ conveying optimal interpretation is the one which minimizes the data reconstruction error $\|(\mathbf{s})-\mathbf{x}\|$. When the L_2-norm is selected, the least squares solution ensues. For separable linear models, $\mathbf{f}_m(\mathbf{s}_m)=\mathbf{X}_m\mathbf{X}_m^T\mathbf{W}_m^T\left(\mathbf{W}_m\mathbf{X}_m\mathbf{X}_m^T\mathbf{W}_m^T\right)^{-1}\mathbf{W}_m\mathbf{X}_m$, and, thus, $\mathbf{A}_m\approx\sum_{mm}^{\mathbf{x}}\mathbf{W}_m\sum_{mm}^{\mathbf{s}\ -1}$. This is quite similar in nature to the PLS Mode A deflation, except that *all* sources $\mathbf{s}_m$ contribute simultaneously to the estimation of $\mathbf{A}_m$. This idea was explored in source power comodulation (SPoC) [117], which pursued a CCA-type analysis between windowed variance profiles of $\mathbf{s}_1$ (in dataset $m = 1$) and a single known fixed reference source s_{21} (in dataset $m = 2$), canonical SPoC (cSPoC) [118], which pursued CCA between "envelope" transformations of $\mathbf{s}_m$, and multimodal SPoC (mSPoC) [119],

which pursued CCA between $\mathbf{s}_1$ and temporally filtered windowed variance profiles of $\mathbf{s}_2$. The key differences between CCA and SPoC-type approaches are that $\mathbf{s}_1$ and $\mathbf{s}_2$ can have different number of observations and at least one set of sources undergoes a non-linear transformation.

Since covariance is $E[s_{1i}, s_{2i}] = \rho^s_{12i}\sigma^s_{1i}\sigma^s_{2i}$, where ρ^s_{12i} is correlation between s_{1i} and s_{2i}, and $\sigma^s_{1i}\,\sigma^s_{2i}$ are their standard deviations, respectively, maximizing covariance will not necessarily maximize correlation[4]. Thus, covariance might not detect sources with high linear dependence (high correlation) if either of their standard deviations is low (possibly the case in genetic data [120]). Because PLS does not prioritize correlation over scale, we consider it "deficient" from an optimization perspective as it will always be biased by the scale of the data. Nevertheless, it may still be advantageous to rely on this property, depending on the importance of scale for a given application, such as regression, where larger scale translates into larger explained variability.

Multiset CCA (mCCA) [108] extends CCA to $M \geq 2$ datasets. In this case, sources are organized into $K = C$ subspaces, each one spanning over M datasets, forming K tuples made of *one* corresponding source from each dataset. The M sources contained in each subspace ($\mathbf{s}_k$, $k = i$) share an $M \times M$ correlation matrix $\mathbf{R}^{\mathbf{s}_k}$. The solution sought in mCCA maximizes the entire correlation structure of each $\mathbf{R}^{\mathbf{s}_k}$. This is typically achieved by either minimizing its determinant $\det(\mathbf{R}^{\mathbf{s}_k})$ or maximizing some norm $\|\mathbf{R}^{\mathbf{s}_k}\|$. Consequently, sources in the same subspace $\mathbf{s}_k$ are (potentially) highly linearly related to each other and, therefore, statistically *dependent*. Typical cost functions for mCCA include sum of squares of all entries of the correlation matrix (SSQCOR) and the generalized variance (GENVAR). For $i = 1$, SSQCOR mCCA solves the following constrained optimization [37], [108], [121], with $\mathbf{R}^{\mathbf{s}_k} = \sum^{\mathbf{s}_k}$, when the constraints are enforced, and $k = i$:

$$\begin{aligned} \underset{\mathbf{W}_{\cdot i}}{max} \quad & \mathrm{tr}\left(\sum^{\mathbf{s}_i}\sum^{\mathbf{s}_i T}\right),\ \sum^{\mathbf{s}_i} = \mathbf{W}_{\cdot i}\sum^{\mathbf{x}}\mathbf{W}^T_{\cdot i} \\ s.t. \quad & \mathbf{W}_{mi}\sum^{\mathbf{x}}_{mm}\mathbf{W}^T_{mi} = 1, m = 1 \ldots M, \end{aligned} \tag{4}$$

where, tr(·) is the trace operator, $\sum^{\mathbf{s}_i}$ is the cross-covariance among the i-th source from all datasets, $\Sigma^{\mathbf{x}}$ is the crosscovariance matrix among all datasets, and $\mathbf{W}_{\cdot i}$ is a $M \times \overline{V}$, $\overline{V} = \sum_{m=1}^{M} V_m$ block-diagonal matrix whose m-th block contains the row vector $\mathbf{W}_{mi}$. GENVAR mCCA, on the other hand, solves [108]:

$$\begin{aligned} \underset{\mathbf{W}_{\cdot i}}{min} \quad & \det\left(\sum^{\mathbf{s}_i}\right),\ \sum^{\mathbf{s}_i} = \mathbf{W}_{\cdot i}\sum^{\mathbf{x}}\mathbf{W}^T_{\cdot i} \\ s.t. \quad & \mathbf{W}_{mi}\sum^{\mathbf{x}}_{mm}\mathbf{W}^T_{mi} = 1, m = 1 \ldots M, \end{aligned} \tag{5}$$

[4]The constraint in CCA enforces $\sigma^s_{1i} = \sigma^s_{2i} = 1$. Likewise for mCCA.

which is a function of the product of eigenvalues of $\mathbf{R}_i$. For $i > 1$, mCCA also requires additional constraints:

$$\sum\nolimits^{\mathbf{s}}_{ml,j\bar{j}} = \mathbf{W}_{mj} \sum\nolimits^{\mathbf{x}}_{ml} \mathbf{W}^{T}_{l\bar{j}} = 0, 1 \leq \bar{j} < j \leq i,\ 1 \leq m, 1 \leq M,$$

meaning all non-corresponding source pairs are uncorrelated. Multi-block PLS approaches exist [122] but are not common in brain data analysis.

IVA is a model that seeks to minimize the mutual information between subspaces $\mathbf{s}_k$ [26], [27], similarly to (1):

$$I(\mathbf{s}) = -h(\mathbf{s}) + \sum_{k=1}^{K} h(\mathbf{s}_k), \tag{6}$$

where $K = C$, $h(\mathbf{s}) = -E[\log p(\mathbf{s})]$, and $p(\mathbf{s})$ and $p(\mathbf{s}_k)$ are the joint probability density function (pdf) of all sources and the marginal subspace pdf, respectively. The choice of distribution for $p(\mathbf{s}_k)$ distinguishes between different IVA algorithms. In second-order IVA (IVA-G) [123], $p(\mathbf{s}_k)$ is modeled as multivariate Gaussian, simplifying (6) to:

$$I_{\text{IVA- G}}(\mathbf{s}) = \frac{CM\ log(2\pi e)}{2} + \frac{1}{2} log \left(\prod_{i=1}^{C} \prod_{m=1}^{M} \lambda_{mi} \right) - \sum_{m=1}^{M} log\, |\det(\mathbf{W}_m)| - const, \tag{7}$$

where λ_{mi} are the M eigenvalues of the source covariance matrices $\sum^{\mathbf{s}_k}$, $k=i$, and e is Napier's constant. When $\mathbf{W}_m$ is constrained to be orthogonal, IVA-G is equivalent to GENVAR mCCA [123].

IVA assumes all sources are independent within each dataset, such that $p(\mathbf{s}_m) = \prod_{i=1}^{C} p(s_{mi})$, $m = 1 \ldots M$. This implies that second- and higher-order cross-moments between non-corresponding sources are driven to zero and, likewise, become potentially non-zero for sources in the *same* subspace. For IVA-G, only SOS is considered so that independence translates to uncorrelation. Thus, IVA-G shares the same base model as CCA, PLS-SVD, and mCCA, except the latter three seek particular solutions that explicitly maximize some measure of correlation inside each subspace $\mathbf{s}_k$. The relationship with IVA becomes clearer when we consider that the cost functions in these algorithms assume zero-correlation among sources in different subspaces. From an optimization theory perspective, this can be achieved, for example, by null-space methods [51], which first project into a space that satisfies the constraints (i.e., uncorrelates non-corresponding sources) and then find a solution within that space. It should be clear that the constraints in CCA, PLS-SVD, and mCCA correspond exactly to IVA-G and that maximization of correlation/covariance gives a particular solution within that space.

IVA-L [27], on the other hand, uses the multivariate Laplace distribution to model HOS dependence within subspaces. Due to its HOS nature, it leads to independence, in addition to uncorrelation, among subspaces. Uncorrelation is imposed within subspaces, however, due to the assumed identity *dispersion* matrix [107]:

$$I_{\text{IVA- L}}(\mathbf{s}) = \frac{CM\ log(4\pi)}{2} - Clog\left(\sqrt{\pi}\Gamma\left(\frac{M+1}{2}\right)\right) + \sum_{i=1}^{C}\|\mathbf{s}_i\|_2 - \sum_{m=1}^{M} log\,|\det(\mathbf{W}_m)|, \quad (8)$$

with $k = i$, $\|\mathbf{s}_i\|_2 = \mathbf{s}_i^T\mathbf{s}_i = \sqrt{\sum_{m=1}^{M}\mathbf{s}_{mi}^2}$, $s_{mi} = \mathbf{W}_{mi}\mathbf{x}_m$, and $\Gamma(\cdot)$ is the Gamma function. IVA can explicitly pursue independence in each dataset while retaining dependence across corresponding sources. Estimating dataset-specific $\mathbf{W}_m$ and dependent, rather than identical, sources $\mathbf{s}_m$ enables greater flexibility to capture dataset-specific variability. Under similar motivation, GIG-ICA [124] uses the source estimates $\mathbf{s}$ from GICA as *reference* for dataset-specific estimation of $\mathbf{W}_m$ and $\mathbf{s}_m$ via ICA-R, effectively using two SDU approaches to attain an MDU result.

Next, we consider the joint ICA (jICA) model [125], commonly utilized in multimodal data fusion [121], [126] and "temporal ICA" of temporally concatenated fMRI [127], [128]. Originally proposed for (but not limited to) exactly two datasets, jICA's hallmark assumption is that the same mixing matrix $\mathbf{A}$ generates both datasets. It also assumes *none* of the sources are statistically related, i.e., $p(\mathbf{s}) = \prod_{i=1}^{C}\prod_{m=1}^{M} p(s_{mi})$, and that $p(\cdot)$ is the same for all sources. Comparing to the IVA model, this is equivalent to constraining the block-diagonal structure to $\mathbf{A}_m = \mathbf{A}$, $m = 1 \ldots M$, and modeling subspaces with $p(\mathbf{s}_k) = \prod_{m=1}^{M} p(s_{mk})$. The key difference between such a variant of IVA and JICA is that the first uses an M-dimensional independent joint pdf for $\mathbf{s}_k$ while the second combines corresponding sources into a single one-dimensional pdf. Thus, JICA conveniently follows optimization of (1).

In a similar fashion, the approach of common spatial patterns (CSP) [129], [130] assumes the same generative system for two different datasets (typically from the same subject in different conditions). Like exact JD approaches for SDU models, it uses SOS from two covariance matrices, $\sum^{\mathbf{x}_1}$ and $\sum^{\mathbf{x}_+} \triangleq \sum^{\mathbf{x}_1} + \sum^{\mathbf{x}_2}$, to form a generalized eigenvalue problem $\sum^{\mathbf{x}_1}\mathbf{v}_i = \lambda\sum^{\mathbf{x}_+}\mathbf{v}_i$, $\mathbf{v}_i = \mathbf{W}_{1i}^T = \mathbf{W}_{2i}^T$, and identify uncorrelated sources $\mathbf{s}_1$ and $\mathbf{s}_2$ with maximum *variance contrast* between corresponding sources. That is, for C sources, s_{11} has maximal variance (eigenvalue) and the corresponding s_{21} has *minimal* variance; likewise, s_{1C} has minimal variance and s_{2C} has maximal variance. The motivation for $\sum^{\mathbf{x}_+}$ stems from classification: maximizing variance contrast between corresponding sources helps enhance source-based classification of different conditions. Equivalently, maximization of the Jensen-Shannon divergence (or symmetric Kullback-Leibler (KL) divergence) leads to the same CSP solution when the sources are assumed to be uncorrelated

Gaussian [131]. Extensions of CSP focus on regularizing the problem with additional constraints [132], [133], leading to GEV problems of the form $\sum^{\mathbf{x}_m}\mathbf{v}_{mi}=\lambda\left(\sum^{\mathbf{x}_+}+\sum_p\alpha_p\mathbf{K}_p\right)\mathbf{v}_{mi}$, $\mathbf{v}_{mi}=\mathbf{W}_{mi}^T$, where $\mathbf{K}_p$ is a positive definite penalty term representing any undesired properties detected in the data (typically, non-stationarity due to multiple subjects, sessions, or artifacts). A general regularization framework based on divergence measures is also possible, with Beta divergence offering additional robustness against outliers [131]. The benefits of regularization come at a cost: $\mathbf{W}_{mi}$ is different for each dataset and, thus, requires dataset-specific optimization. Consequently, this implies a different generative system per dataset, which places CSP closer to CCA (and mCCA) but not quite since each dataset is optimized separately rather than jointly.

Expanding on the jICA model, linked ICA (LICA) [134] applies Bayesian ICA to a "joint" TICA model, establishing a more flexible structure on the shared $\mathbf{A}$. Specifically, TICA is applied to D groups of M_d datasets of equal spatial resolution. For each $d = 1..D$, $\mathbf{x}_{dm} = \mathbf{A}\mathbf{D}_{dm}\mathbf{s}_d$, $m = 1..M_d$. Like jICA, sources $\mathbf{s}_d$ are assumed independent across the D groups of datasets but, like GICA, they are identical for the M_d datasets in the same group, i.e., $\mathbf{s}_{d1}=\cdots=\mathbf{s}_{dM_d}=\mathbf{s}_d$. Thus, sources are modeled as $p(\mathbf{s})=\prod_{i=1}^{C}\prod_{d=1}^{D}p(s_{di})$, finally combining the corresponding sources $\mathbf{s}_k$, $k = i$, into a single one-dimensional pdf that is a weighted average of each source's pdf: $p(s_k)=\sum_{d=1}^{D}\alpha_d p(s_{dk})$, $k = i$, where α_d is determined by the number of observations in each $\mathbf{s}_d$. As a result of the PARAFAC model, the mixing matrices from each dataset share a common structure but are not identical, unlike jICA. However, their columns are *perfectly* correlated, which may or may not be a benefit depending on the problem. Finally, the Bayesian strategy provides a nice framework to handle noise and employ re-estimation of the source distribution parameters, which is selected as a mixture of Gaussian (MOG) for each s_k.

To conclude, a nice example of generalization are the wide class of models contained in the IVA-Kotz family [107] (and a version that includes HOS and sample dependence [135]), which cleverly use the Kotz distribution to capture both SOS and HOS simultaneously.

3) Single Dataset Multidimensional (SDM) Problem—The linear models for the SDM problem assume a single dataset generated by an invertible linear, fully-connected mixture with multidimensional (grouped) sources occurring within the dataset (see Fig. 4), forming $K \leq C$ subspaces $\mathbf{s}_k$. The most common model in this category was originally introduced as multidimensional ICA (MICA) [28] and later investigated under the independent subspace analysis (ISA) [30], [31], [136] nomenclature. In this model, sources in the same subspace $\mathbf{s}_k$ are related and, thus, statistically *dependent*. Sources in different subspaces are assumed independent. The simplest ISA approach post-processes ICA results to form subspaces post-hoc as in [28], [31], [47].

A more principled approach is to minimize (6), redefining $\mathbf{s}_k$ according to user-defined source groupings. Applying this approach, Hyvärinen et al. [136] assume $\mathbf{W}$ is orthogonal, using a Laplacian-like multivariate distribution for independence among subspaces and HOS-only dependence within subspaces, enforcing uncorrelation both within and among

subspaces. Similarly, Silva et al. [33] use a scale-adjusted multivariate Laplace distribution and introduces reconstruction error constraints to bypass the typical PCA reduction step. Focusing only on SOS among subspaces, Lahat et al. [29] use a multivariate Gaussian distribution to model subspace dependences. Likewise, Silva et al. [137] improve on [33] by not enforcing uncorrelation within subspaces.

Stationary subspace analysis (SSA) is a model designed to find one group of sources whose *mean and SOS* remain unchanged over time, hence the name. KL-SSA [138] uses an AJD of $\mathbf{\Sigma}^{\mathbf{s}}(\tau)$, accounting for different $\boldsymbol{\mu}^{\mathbf{s}}(\tau)$, to identify stationary sources based on KL divergence, very similar to Pham's AJD criterion [139]. It models each window as a Gaussian source $\mathcal{N}(\mu^{\mathrm{s}}(\tau), \sum^{\mathrm{s}}(\tau))$, and the collection of all windows as $\mathcal{N}(\mu^{\mathrm{s}}, \sum^{\mathrm{s}})$. Then the sum of divergences between each window and the "aggregate" source is minimized. By definition, $K = 2$, though only the group of stationary sources is retrieved. Analytic SSA (ASSA) [140] approximates the KL divergence with a "variance of covariance" function, simplifying the optimization to a generalized eigenvalue problem. Like other SDM models, sources within the retrieved subspace are still unmixed. Consequently, SSA is only intended to act as a filtering tool, removing non-stationarities from the data.

4) Multiple Dataset Multidimensional Problem—MDM models were developed only recently. In linear MDM, $M \geq 1$, datasets are generated by a linear, structured mixture where the combined mixing matrix $\mathbf{A}$ has a block-diagonal structure: $\mathbf{x}_m = \mathbf{A}_m\mathbf{s}_m$, $m = 1 \ldots M$. The number of sources C_m is typically *not* the same for all datasets. Multidimensional (grouped) sources occur within any single dataset and are allowed to group (or not) across datasets (see Fig. 5), forming K d_k-dimensional subspaces $\mathbf{s}_k$.

Sources in the same subspace $\mathbf{s}_k$ are considered statistically *dependent*, and independent otherwise. Thus, $p(\mathrm{s}) = \prod_{k=1}^{K} p(\mathrm{s}_k)$, and minimization of (6) with user-defined source groupings for $\mathbf{s}_k$ follows. Variations include models that either exploit a) only SOS, like joint independent subspace analysis (JISA) [34], b) only HOS, like multidataset independent subspace analysis (MISA) [32], [33] (emphasizing uncorrelation within subspaces), or c) both, like MISA with SOS [137].

B. Applications and Data Modalities

The "blind" property of BSS models makes them a powerful tool in applications lacking a precise model of the measured system and with data confounded by noise of unknown characteristics. Brain imaging is an area where these properties are especially emphasized. We now briefly review some applications of BSS models to brain imaging data. The first case focuses on data from a single imaging modality (unimodal) while the other considers current attempts to identify common motifs from two or more imaging modalities (multimodal).

1) Unimodal—First we consider the analysis of a single dataset. If it contains data from a single subject, the model captures subject-specific information. If instead it contains subject-specific summaries from a certain population, then the model characterizes underlying group patterns. Both cases resort to SDU models. The second example regards the processing of

two or more datasets simultaneously. When each dataset contains information from a single subject, shared or similar patterns across subjects are identified and group trends derived. Although group trends offer useful conclusions about population differences, preserving subject-specific information is crucial for clinical diagnosis and personalized treatment. Both SDU and MDU models have been used to capture inter-subject, and sometimes inter-trial, variability.

a) SDU Problem: ICA has been successfully utilized in a number of exciting applications, especially those that have proven challenging with the standard regression-type approaches [141], [142]. For fMRI [4], [143] and EEG [6], [144]–[146], ICA reveals dynamics for which a temporal model is not available [147], finding largely non-overlapping, temporally coherent brain regions without constraining the shape of the temporal response. The Infomax algorithm with a sparse prior is particularly well suited for spatial analysis [148]. Besides fMRI, the brain grey [2], [149] and white matter [149], as well as functional near-infrared spectroscopy (fNIRS) [69] have also been analyzed by ICA to study the diseased brain. In addition to its wide use for recovering spatial networks as independent components —in diffusion tensor imaging (DTI) [7], positron emission tomography (PET) [67], and even MEG [150]—ICA is also promising for temporal [128], [151] and spectral (e.g., MR spectroscopic imaging (MRSI) [70]) domains of brain imaging.

ICA was shown to be useful in modalities such as EEG and MEG. It has also been successfully applied to multi-neuronal recordings [68]. Mostly, it is used for artifact reduction [61], [152] or for real-time control in BCI applications [65]. However, for artifact reduction in MEG, SOBI was shown to be superior to other BSS methods available at the time [5] (also see [153], [154] for another comparative study).

GICA [4] of fMRI [3] is one of the most successful BSS tools for neuroimaging analysis, providing a means to handle multi-subject datasets. It enables network identification under task and rest regimes [89]. The connection patterns between these networks is very interesting and useful for differentiation [90], [155]–[157] and neurodiagnostic discovery [87].

b) MDU Problem: Both CCA and mCCA have been successfully utilized in applications ranging from single- [158] and multi-subject analysis of fMRI [159], [160], as well as BCI [161], [162]. PLS, on the other hand, has also been successfully applied to neuropsychological and MRI relationships [163], while temporal ICA was investigated in [127], [128], [151]. Finally, IVA has found great value in leveraging the success of independence-based methods while better characterizing inter-subject variability [94], [164], [165].

2) Multimodal—Multimodal analyses are intended to leverage information contained in multiple data streams by modeling the relationship among modalities. Convergent evidence suggests that combining functional and structural information is useful in clinical research [166]–[168]. Moreover, multi-modal studies often demonstrate some congruent effects across modalities and different brain pathologies [169]–[173]. Any remaining complementary information typically contributes to increased differentiation power among

diseases. Unlike multimodal approaches that use one modality to constrain or filter the other, *multimodal data fusion* seeks hidden shared information underlying the signals from both modalities simultaneously. Multimodal data fusion applications are by and large viewed as MDU-type problems [174], [175]. However, for the most part, current neuroimaging multimodal fusion schemes have focused on pairs of modalities.

IV. Future Directions

A. Emerging Modalities

Improvements in imaging instrumentation and signal acquisition continue to provide researchers with novel and higher quality information about the physiology of the brain. Naturally, the use and combination of such emerging modalities has great potential for producing new findings and applications.

We begin with a summary of promising, emerging modalities. First, high-resolution quantitative MR imaging of tissue-specific parameters such as longitudinal relaxation (T1) provides good indication of cortical myelination [176], with the benefit of allowing direct comparison of images across scanners and sites, as well as longitudinally, for most cortical brain areas [177]. Also, a measure of local variation in grey matter called voxel-based cortical thickness (VBCT) provides higher grey matter sensitivity than typical voxel-based morphometry (VBM) [178]. Following recent breakthroughs in DWI [179], [180], crossing white matter fibers can now be resolved and tracked, providing thrilling details about the orientation and structural connectivity of fiber bundles. Similar improvements have led to ultra-fast fMRI sequences [16] at higher spatial resolution, simplifying the filtering of certain physiological noise sources, such as breathing and cardiac pulsation.

We also highlight the emergence of new devices that combine different imaging modalities. Simultaneous PET/MR devices offer multiple opportunities for multimodal research. Particularly, with the advent of functional PET (fPET) by constant infusion of 2-[(18)F]-fluorodeoxyglucose (FDG), functional changes in glucose utilization by the brain can be observed with better time resolution than traditional PET [181]. Consequently, simultaneous functional imaging using fPET and fMRI (by arterial spin labeling (ASL) and blood oxygen level dependent (BOLD) contrasts) enables the study of neurovascular coupling [182], especially in the study of drug challenges. Also, sMRI scans interleaved with PET acquisition can enable improved PET resolution and SNR following MRI-guided attenuation and partial-volume corrections, motion compensation, and reconstruction [183].

Similarly, recent advances in cap and probe design allow for high-density simultaneous fNIRS and EEG experiments with reduced motion artifacts and for extended periods of time [184], [185]. These could be combined with PET/MR systems, offering a huge opportunity for neurodiscovery, and indicating an impending need for multiple-dataset methods oriented to multimodal fusion.

B. Emerging Applications

The past decade has witnessed a growing interest in multimodal analyses, especially N-way multimodal fusion [33], [186], for their potential to leverage hidden multimodal interactions

(i.e., cross-dataset dependence) and construct a more complete view of brain function and structure. More modalities typically means increased confidence in statements about the neural determinants of healthy [187] and disease states. Moreover, multiple-dataset BSS has demonstrated potential to identify endophenotypes from brain imaging data for genetic association studies, eliciting candidate biomarkers for several mental illnesses [188], [189].

While multiple-dataset BSS provides a pristine opportunity for neurodiscovery in multimodal studies of mental disorders [190], the benefits for unimodal studies involving longitudinal or multi-site data can be more immediate, as they may identify recurrent features across time and study location.

Structure-function connectivity analyses [171], [191]–[193] are also expected to largely benefit from continued research in multiple-dataset BSS. For example, graph-oriented BSS may elicit the underlying breakdown of networks into modules and meta-states [194], offering a fresh new look at connectivity.

C. Emerging Techniques

Typically, linear BSS on a $V \times N$ dataset $\mathbf{X} = \mathbf{AS}$ captures information along the dimension of N. In order to leverage information contained along the other dimension, some approaches have been exploiting the data transpose $\mathbf{X}^T$ as well. Utilizing a two-step approach, mCCA+jICA [190] computes mCCA among $\mathbf{X}_m^T$ to find $\mathbf{A}_m^T$ with high correlation among corresponding rows, followed by jICA on $\mathbf{S}_m$; conversely, [151] uses temporal ICA on the temporally concatenated $\mathbf{A}^T$ that results from GICA. On the other hand, approaches like non-negative matrix factorization (NMF) [195] alternate between $\mathbf{X}$ and $\mathbf{X}^T$ at every step of the alternating least-squares (ALS) [196], [197] optimization. This approach generalizes well to the tensor case [198], [199]. Parallel ICA [9], however, alternates between ICA (on $\mathbf{X}_1$ and $\mathbf{X}_2$, separately) and CCA (between $\mathbf{X}_1^T$ and $\mathbf{X}_2^T$) at every step. Such techniques are likely to play an important role in the development of new methods for linear MDM problems.

V. Conclusion

We have presented a unifying view of BSS, providing new insight into the connections among many traditional and modern BSS models. The new perspective leads to a broader sense of generalization, highlighting several directions for further development. Particularly, MDM models emerge as an organic confluence of three major trends: simultaneous multi-dataset analysis, grouping of sources within a dataset, and use of all-order statistics. Therefore, they capture all key features and subspace structures common to their predecessors. Their benefit is in the flexibility to simultaneously model dataset-specific as well as cross-dataset subspace associations.

The demands from advances in multimodal brain imaging technology are likely to propel the development of BSS methods strongly into the direction of N-way multimodal fusion. However, a better and more general solution becomes possible. The availability of such advanced BSS approaches will provide novel ways to investigate inter-subject covariation across multiple populations, sites, and longitudinal inquiries.

Acknowledgments

This work was primarily funded by NIH grant R01EB005846 and NIH NIGMS Center of Biomedical Research Excellent (COBRE) grant 5P20RR021938/P20GM 103472 to Vince Calhoun (PI), and NSF 1539067.

Biographies

Rogers F. Silva (Student Member, IEEE) received the B.Sc. degree in Electrical Engineering in 2003 from the Catholic University (PUCRS), Porto Alegre, Brazil, the M.S. degree in Computer Engineering (with minors in Statistics and in Mathematics) in 2011, and the Ph.D. degree in computer engineering in 2016 (expected), both from The University of New Mexico, Albuquerque, NM, USA. He worked as an engineer, lecturer, and consultant, and is currently a senior graduate research assistant at The Mind Research Network, Albuquerque, NM, USA. As a multidisciplinary research scientist, he develops algorithms for statistical and machine learning, image analysis, numerical optimization, memory efficient large scale data reduction, and distributed analyses, focusing on multimodal, multi-subject neuroimaging data from thousands of subjects. His interests are: multimodal data fusion, statistical and machine learning, image, video and data analysis, multiobjective, combinatorial and constrained optimization, signal processing, and neuroimaging.

Sergey M. Plis received the Ph.D. degree in Computer Science in 2007 from The University of New Mexico, Albuquerque, NM, USA. He is the Director of Machine Learning at the Mind Research Network, Albuquerque, NM, USA. His research interests lie in developing novel and applying existing techniques and approaches to analyzing large scale datasets in multimodal brain imaging and other domains. He develops tools that fall within the fields of machine learning and data science.

One of his key goals is to take advantage of the strengths of imaging modalities and infer structure and patterns that are hard to obtain non-invasively and/or that are unavailable for direct observation. In the long term this amounts to developing methods capable of revealing mechanisms used by the brain to form task-specific transient interaction networks and their cognition-inducing interactions via multimodal fusion at features and interaction levels.

His ongoing work is focused on inferring multimodal probabilistic and causal descriptions of these function-induced networks based on fusion of fast and slow imaging modalities. This includes feature estimation via deep learning-based pattern recognition and learning causal graphical models.

Jing Sui (Senior Member, IEEE) received her B.S. and PhD. Degree in optical engineering with honors from Beijing Institute of Technology in 2002 and 2007 respectively. Then she worked at the Mind Research Network, NM, USA as a postdoctor fellow and got promoted to Research Scientist in 2010 and Assistant Professor of Translational Neuroscience in 2012. She is currently a full professor at the National Laboratory of Pattern Recognition & Brainnetome Center, Institute of Automation, Chinese Academy of Science (CAS). Dr Suis research interests include machine learning, multi-modal data fusion (fMRI, sMRI, DTI, EEG, genetics), pattern recognition and biomarker identification of mental illnesses. Her ongoing work in predictive data mining attempts to meet clinical challenges of making early intervention possible based on fundamental neuroimaging data. She is a recipient of One Hundred Talents plan of CAS and has been program committee member of IEEE BIBM since 2011. Now she serves as grant reviewer for China Natural Science Foundation, China Scholarship Council etc and regular reviewers for more than 30 peer-reviewed journals.

Tülay Adalı (Fellow, IEEE) received the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, NC, USA, in 1992.

She joined the faculty at the University of Maryland Baltimore County (UMBC), Baltimore, MD, USA, in 1992. She is currently a Distinguished University Professor in the Department of Computer Science and Electrical Engineering at UMBC. Her research interests are in the areas of statistical signal processing, machine learning for signal processing, and biomedical data analysis.

Prof. Adalı assisted in the organization of a number of international conferences and workshops including the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), the IEEE International Workshop on Neural Networks for Signal

Processing (NNSP), and the IEEE International Workshop on Machine Learning for Signal Processing (MLSP). She was the General Co-Chair, NNSP (2001–2003); Technical Chair, MLSP (2004–2008); Program Co-Chair, MLSP (2008, 2009, and 2014), 2009 International Conference on Independent Component Analysis and Source Separation; Publicity Chair, ICASSP (2000 and 2005); and Publications Co-Chair, ICASSP 2008. She is Technical Program Co-Chair for ICASSP 2017 and Special Sessions Co-Chair for ICASSP 2018. She chaired the IEEE Signal Processing Society (SPS) MLSP Technical Committee (2003–2005, 2011–2013), served on the SPS Conference Board (1998–2006), and the Bio Imaging and Signal Processing Technical Committee (2004–2007). She was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2003–2006), the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING (2007–2013), the IEEE JOURNAL OF SELECTED AREAS IN SIGNAL PROCESSING (2010–2013), and *Elsevier Signal Processing Journal* (2007–2010). She is currently serving on the Editorial Boards of the PROCEEDINGS OF THE IEEE and *Journal of Signal Processing Systems for Signal, Image, and Video Technology*, and is a member of the IEEE Signal Processing Theory and Methods Technical Committee. She is a Fellow of the American Institute for Medical and Biological Engineering (AIMBE), a Fulbright Scholar, recipient of a 2010 IEEE Signal Processing Society Best Paper Award, 2013 University System of Maryland Regents' Award for Research, and an NSF CAREER Award. She was an IEEE Signal Processing Society Distinguished Lecturer for 2012 and 2013.

Marios S. Pattichis (Senior Member, IEEE) received the B.Sc. degree (high and special honors) in Computer Sciences and the B.A. degree (high honors) in Mathematics, both in 1991, M.S. degree in Electrical Engineering in 1993, and the Ph.D. degree in Computer Engineering in 1998, all from the University of Texas at Austin, Austin, TX, USA. He is currently a Professor with the Department of Electrical and Computer Engineering, University of New Mexico (UNM), Albuquerque, NM, USA. He was a founding Co-PI of COSMIAC at UNM, where he is currently Director of the image & video Processing and Communications Lab (ivPCL, ivpcl.org). His current research interests include digital image, video processing, communications, dynamically reconfigurable computer architectures, and biomedical and space image processing applications.

Dr. Pattichis is currently a Senior Associate Editor of the IEEE SIGNAL PROCESSING LETTERS. He has been an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and has also served as a Guest Associate Editor for the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE. He was the General Chair of the 2008 IEEE Southwest Symposium on Image Analysis and Interpretation. He was a recipient

of the 2004 Electrical and Computer Engineering Distinguished Teaching Award at UNM. For his development of the digital logic design labs at UNM he was recognized by the Xilinx Corporation in 2003 and by the UNM School of Engineering's Harrison faculty excellent award in 2006.

Vince D. Calhoun (Fellow, IEEE & AAAS) received the B.S. degree in electrical engineering from the University of Kansas, Lawrence, KS, USA, in 1991, the M.S. degrees in biomedical engineering and information systems from The Johns Hopkins University, Baltimore, MD, USA, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, Baltimore, in 2002.

He worked as a Research Engineer in the Psychiatric Neuroimaging Laboratory, Johns Hopkins, from 1993 until 2002. He then served as the Director of Medical Image Analysis at the Olin Neuropsychiatry Research Center and as an Associate Professor at Yale University. He is currently Executive Science Officer and Director of Image Analysis and MR Research at the Mind Research Network, Albuquerque, NM, USA, and is a Distinguished Professor in the Departments of Electrical and Computer Engineering (primary), Biology, Computer Science, Neurosciences, and Psychiatry at the University of New Mexico, Albuquerque. He is the author of more than 400 full journal articles and over 500 technical reports, abstracts, and conference proceedings. Much of his career has been spent on the development of data-driven approaches for the analysis of brain imaging data. He has won over $85 million in NSF and NIH grants on the incorporation of prior information into ICA for functional magnetic resonance imaging, data fusion of multimodal imaging and genetics data, and the identification of biomarkers for disease, and leads a P20 COBRE center grant on multimodal imaging of schizophrenia, bipolar disorder, and major depression.

Dr. Calhoun is a Fellow of the American Association for the Advancement of Science (AAAS), the American Institute for Medical and Biological Engineering (AIMBE), and the International Society for Magnetic Resonance in Medicine (ISMRM). He is also a member and regularly attends the Organization for Human Brain Mapping, the International Society for Magnetic Resonance in Medicine, the International Congress on Schizophrenia Research, and the American College of Neuropsychopharmacology. He is also a regular grant reviewer for NIH and NSF. He has organized workshops and special sessions at multiple conferences. He is currently chair of the IEEE Machine Learning for Signal Processing (MLSP) technical committee. He is a reviewer for many journals, is on the editorial board of the Brain Connectivity and Neuroimage journals, and serves as an Associate Editor for Journal of Neuroscience Methods and several other journals.

References

1. Comon, P., Jutten, C. Handbook of Blind Source Separation. 1st. Oxford, UK: Academic Press; 2010.
2. Xu L, Groth K, Pearlson G, Schretlen D, Calhoun V. Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia. Hum Brain Mapp. 2009; 30(3):711–724. [PubMed: 18266214]
3. Calhoun V, Liu J, Adalı T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. NeuroImage. 2009; 45(1 Supplement 1):S163–S172. [PubMed: 19059344]
4. Calhoun V, Adalı T, Pearlson G, Pekar J. A method for making group inferences from functional MRI data using independent component analysis. Hum Brain Mapp. 2001; 14(3):140–151. [PubMed: 11559959]
5. Tang A, Pearlmutter B, Zibulevsky M, Carter S. Blind source separation of multichannel neuromagnetic responses. Neurocomputing. 2000; 3233:1115–1120.
6. Eichele T, Rachakonda S, Brakedal B, Eikeland R, Calhoun V. EEGIFT: Group independent component analysis for event-related EEG data. Comput Intell Neurosci. 2011; 2011:129365. [PubMed: 21747835]
7. Arfanakis K, Cordes D, Haughton V, Carew J, Meyerand M. Independent component analysis applied to diffusion tensor MRI. Magn Reson Med. 2002; 47(2):354–363. [PubMed: 11810680]
8. Boutte D, Calhoun V, Chen J, Sabbineni A, Hutchison K, Liu J. Association of genetic copy number variations at 11 q14.2 with brain regional volume differences in an alcohol use disorder population. Alcohol. 2012; 46(6):519–527. [PubMed: 22732324]
9. Liu J, Demirci O, Calhoun V. A parallel independent component analysis approach to investigate genomic influence on brain function. IEEE Signal Process Lett. 2008; 15:413–416. [PubMed: 19834575]
10. Liu J, Morgan M, Hutchison K, Calhoun VD. A study of the influence of sex on genome wide methylation. PLoS ONE. Apr; 2010 5(4):1–8.
11. Renard E, Teschendorff AE, Absil P-A. Capturing confounding sources of variation in DNA methylation data by spatiotemporal independent component analysis. Proc ESANN 2014. :195–200.
12. Bartel J, Krumsiek J, Theis FJ. Statistical methods for the analysis of high-throughput metabolomics data. Comput Struct Biotechnol J. 2013; 4(5):1–9.
13. Spearman, C., Jones, L. Human Ability: A Continuation of “The Abilities of Man”. Macmillan; 1950.
14. Eysenck SBG, Eysenck HJ. The validity of questionnaire and rating assessments of extraversion and neuroticism, and their factorial stability. Brit J Psychol. 1963; 54(1):51–62.
15. Feinberg D, Moeller S, Smith S, Auerbach E, Ramanna S, Glasser M, Miller K, Ugurbil K, Yacoub E. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS ONE. Dec; 2010 5(12):1–11.
16. Feinberg D, Setsompop K. Ultra-fast MRI of the human brain with simultaneous multi-slice imaging. J Magn Reson. 2013; 229:90–100. [PubMed: 23473893]
17. Bianciardi M, Toschi N, Edlow B, Eichner C, Setsompop K, Polimeni J, Brown E, Kinney H, Rosen B, Wald L. Toward an in vivo neuroimaging template of human brainstem nuclei of the ascending arousal, autonomic, and motor systems. Brain Connect. Dec; 2015 5(10):597–607. [PubMed: 26066023]
18. van der Zwaag W, Schfer A, Marques J, Turner R, Trampel R. Recent applications of UHF-MRI in the study of human brain function and structure: a review. NMR Biomed. 2015
19. Comon P. Independent component analysis, a new concept? Signal Process. 1994; 36(3):287–314.
20. Bell A, Sejnowski T. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. Nov; 1995 7(6):1129–1159. [PubMed: 7584893]
21. Hyvärinen A, Erkki O. A fast fixed-point algorithm for independent component analysis. Neural Comput. 1997; 9(7):1483–1492.

22. Belouchrani, A., Abed-Meraim, K., Cardoso, J-F., Moulines, E. Proc ICDSP 1993. Nicosia, Cyprus: 1993. Second-order blind separation of temporally correlated sources; p. 346-351.
23. Yeredor A. Blind separation of gaussian sources via second-order statistics with asymptotically optimal weighting. IEEE Signal Process Lett. Jul; 2000 7(7):197–200.
24. Hotelling H. Relations between two sets of variates. Biometrika. Dec; 1936 28(3/4):321–377.
25. Wold, H. Nonlinear estimation by iterative least squares procedures. In: David, F., editor. Research Papers in Statistics Festschrift for J Neyman. New York, NY: Wiley; 1966. p. 411-444.
26. Adalı T, Anderson M, Fu G-S. Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging. IEEE Signal Process Mag. May; 2014 31(3):18–33.
27. Kim, T., Eltoft, T., Lee, T-W. Proc ICA 2006. Charleston, SC: Independent vector analysis: An extension of ICA to multivariate components; p. 165-172.
28. Cardoso, J-F. Proc IEEE ICASSP 1998. Vol. 4. Seattle, WA: 1998. Multidimensional independent component analysis; p. 1941-1944.
29. Lahat D, Cardoso J, Messer H. Second-order multidimensional ICA: Performance analysis. IEEE Trans Signal Process. Sep; 2012 60(9):4598–4610.
30. Hyvärinen, A., Köster, U. Proc ESANN 2006. Bruges, Belgium: 2006. FastISA: A fast fixed-point algorithm for independent subspace analysis; p. 371-376.
31. Szabó Z, Póczos B, Lőrincz A. Separation theorem for independent subspace analysis and its consequences. Pattern Recognit. 2012; 45(4):1782–1791.
32. Silva, R., Plis, S., Adalı, T., Calhoun, V. Proc OHBM 2014. Hamburg, Germany: Multidataset independent subspace analysis. Poster 3506
33. Silva, R., Plis, S., Adalı, T., Calhoun, V. Proc IEEE ICIP 2014. Paris, France: 2014. Multidataset independent subspace analysis extends independent vector analysis; p. 2864-2868.
34. Lahat, D., Jutten, C. Proc LVA/ICA 2015. Liberec, Czech Republic: 2015. Joint independent subspace analysis: A quasi-newton algorithm; p. 111-118.
35. Zhou G, Zhao Q, Zhang Y, Adalı T, Xie S, Cichocki A. Linked component analysis from matrices to high-order tensors: Applications to biomedical data. Proc IEEE. Feb; 2016 104(2):310–331.
36. Koller, D., Friedman, N. Probabilistic Graphical Models: Principles and Techniques. 1st. Cambridge, MA: MIT Press; 2009. ser Adaptive computation and machine learning
37. Li Y-O, Adalı T, Wang W, Calhoun V. Joint blind source separation by multiset canonical correlation analysis. IEEE Trans Signal Process. Oct; 2009 57(10):3918–3929. [PubMed: 20221319]
38. Rao C. A decomposition theorem for vector variables with a linear structure. Ann Math Stat. 1969; 40(5):1845–1849.
39. Anderson M, Fu G, Phlypo R, Adalı T. Independent vector analysis: Identification conditions and performance bounds. IEEE Trans Signal Process. Sep; 2014 62(17):4399–4410.
40. Lahat, D., Jutten, C. Proc IEEE ICASSP 2016. Shanghai, China: 2016. An alternative proof for the identifiability of independent vector analysis using second order statistics. In Press
41. Lahat D, Jutten C. Joint independent subspace analysis using second-order statistics. IEEE Trans Signal Process. 2016 In Press.
42. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. Dec; 1974 19(6):716–723.
43. Schwarz G. Estimating the dimension of a model. Ann Statist. Mar; 1978 6(2):461–464.
44. Cavanaugh JE. A large-sample model selection criterion based on Kullback's symmetric divergence. Stat Probabil Lett. 1999; 42(4):333–343.
45. Draper D. Assessment and propagation of model uncertainty. J Roy Statist Soc Ser B. 1995; 57(1): 45–97.
46. Rissanen J. Modeling by shortest data description. Automatica. 1978; 14(5):465–471.
47. Ma, S., Li, X-L., Correa, N., Adalı, T., Calhoun, V. Proc IEEE ICASSP 2010. Dallas, TX: 2010. Independent subspace analysis with prior information for fMRI data; p. 1922-1925.
48. Nocedal, J., Wright, S. Numerical Optimization. 2nd. New York, NY: Springer; 2006. ser Springer Series in Operations Research and Financial Engineering

49. Cardoso J, Laheld B. Equivariant adaptive source separation. IEEE Trans Signal Process. Dec; 1996 44(12):3017–3030.
50. Amari S-I. Natural gradient works efficiently in learning. Neural Comput. Feb; 1998 10(2):251–276.
51. Strang, G. Computational Science and Engineering. Wellesley-Cambridge Press; 2007.
52. Boyd, S., Vandenberghe, L. Convex Optimization. 1st. Cambridge, UK: Cambridge University Press; 2004.
53. Honkela A, Valpola H, Ilin A, Karhunen J. Blind separation of nonlinear mixtures by variational bayesian learning. Digit Signal Process. 2007; 17(5):914–934.
54. Choudrey R, Roberts S. Variational mixture of bayesian independent component analyzers. Neural Comput. 2003; 15(1):213–252. [PubMed: 12590826]
55. Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. Ann Stat. 1999; 27(1):94–128.
56. Delabrouille J, Cardoso J-F, Patanchon G. Multidetector multicomponent spectral matching and applications for cosmic microwave background data analysis. Mon Not R Astron Soc. 2003; 346(4):1089–1102.
57. Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag. 1901; 2(11):559–572.
58. Friston K, Frith C, Liddle P, Frackowiak R. Functional connectivity: The principal-component analysis of large (PET) data sets. J Cereb Blood Flow Metab. 1993; 13(1):5–14. [PubMed: 8417010]
59. Andersen A, Gash D, Avison M. Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework. Magn Reson Imaging. 1999; 17(6):795–815. [PubMed: 10402587]
60. Leonardi N, Richiardi J, Gschwind M, Simioni S, Annoni J-M, Schluep M, Vuilleumier P, Ville DVD. Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. NeuroImage. 2013; 83:937–950. [PubMed: 23872496]
61. Vigário R. Extraction of ocular artefacts from EEG using independent component analysis. Electroencephalogr Clin Neurophysiol. 1997; 103(3):395–404. [PubMed: 9305288]
62. Makeig S, Jung T, Bell A, Ghahremani D, Sejnowski T. Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci. 1997; 94(20):10979–10984. [PubMed: 9380745]
63. McKeown M, Jung T, Makeig S, Brown G, Kindermann S, Lee T, Sejnowski T. Spatially independent activity patterns in functional MRI data during the stroop color-naming task. Proc Natl Acad Sci. 1998; 95(3):803–810. [PubMed: 9448244]
64. McKeown M, Makeig S, Brown G, Jung T, Kindermann S, Bell A, Sejnowski T. Analysis of fMRI data by blind separation into independent spatial components. Hum Brain Map. 1998; 6(3):160–188.
65. Makeig S, Enghoff S, Jung T-P, Sejnowski T. A natural basis for efficient brain-actuated control. IEEE Trans Rehabil Eng. Jun; 2000 8(2):208–211. [PubMed: 10896189]
66. Vigário R, Sarela J, Jousmiki V, Hämäläinen M, Oja E. Independent component approach to the analysis of EEG and MEG recordings. IEEE Trans Biomed Eng. May; 2000 47(5):589–593. [PubMed: 10851802]
67. Park H-J, Kim J-J, Youn T, Lee D, Lee M, Kwon J. Independent component model for cognitive functions of multiple subjects using [15O]H2O PET images. Hum Brain Mapp. 2003; 18(4):284–295. [PubMed: 12632466]
68. Takahashi S, Anzai Y, Sakurai Y. Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes. J Neurphysiol. 2003; 89(4):2245–2258.
69. Morren G, Wolf M, Lemmerling P, Wolf U, Choi J, Gratton E, De Lathauwer L, Van Huffel S. Detection of fast neuronal signals in the motor cortex from functional near infrared spectroscopy measurements using independent component analysis. Med Biol Eng Comput. 2004; 42(1):92–99. [PubMed: 14977228]

70. Pulkkinen J, Häkkinen A-M, Lundbom N, Paetau A, Kauppinen R, Hiltunen Y. Independent component analysis to proton spectroscopic imaging data of human brain tumours. Eur J Radiol. 2005; 56(2):160–164. [PubMed: 16233889]
71. Cichocki, A., Amari, SI. Adaptive Blind Signal and Image Processing. Chichester, UK: Wiley; 2003. revised and corrected ed.
72. Hyvärinen, A., Karhunen, J., Oja, E. Independent Component Analysis. 1st. New York, NY: Wiley; 2002.
73. Kirshner S, Póczos B. ICA and ISA using Schweizer-Wolff measure of dependence. 2008:464–471.
74. Haykin, S. Neural Networks and Learning Machines. 3rd. Upper Saddle River, NJ: Prentice Hall; 2008.
75. Cichocki A, Amari S-I. Families of alpha- beta- and gamma-divergences: Flexible and robust measures of similarities. Entropy. 2010; 12(6):1532–1568.
76. Amari S-I, Cichocki A. Information geometry of divergence functions. Bull Pol Acad Sci, Tech Sci. 2010; 58(1):183–195.
77. Cover, T., Thomas, J. Elements of Information Theory. 2nd. Wiley; 2006. ser Telecommunications and Signal Processing
78. Stögbauer H, Kraskov A, Astakhov S, Grassberger P. Least-dependent-component analysis based on mutual information. Phys Rev E. Dec.2004 70:066123.
79. Belouchrani A, Abed-Meraim K, Cardoso J-F, Moulines E. A blind source separation technique using second-order statistics. IEEE Trans Signal Process. Feb; 1997 45(2):434–444.
80. Tong, L., Soon, V., Huang, Y., Liu, R. Proc IEEE ISCS 1990. Vol. 3. New Orleans, LA: 1990. AMUSE: a new blind identification algorithm; p. 1784-1787.
81. Ziehe, A., Müller, K-R. Proc ICANN 98. Skövde, Sweden: 1998. TDSEP – an efficient algorithm for blind separation using time structure; p. 675-680.
82. Tichavsky, P., Koldovsky, Z., Doron, E., Yeredor, A., Gomez-Herrero, G. Proc 14h Euro SPC. Florence, Italy: 2006. Blind signal separation by combining two ICA algorithms: HOS-based EFICA and time structure-based WASOBI; p. 1-5.
83. Koldovsky Z, Tichavsky P, Oja E. Efficient variant of algorithm FastICA for independent component analysis attaining the Cramer-Rao lower bound. IEEE Trans Neural Netw. Sep; 2006 17(5):1265–1277. [PubMed: 17001986]
84. Tichavsky P, Koldovsky Z, Yeredor A, Gomez-Herrero G, Doron E. A hybrid technique for blind separation of non-gaussian and time-correlated sources using a multicomponent approach. IEEE Trans Neural Netw. Mar; 2008 19(3):421–430. [PubMed: 18334362]
85. Li, X-L., Adalı, T. Proc IEEE ICASSP 2010. Dallas, TX: 2010. Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization; p. 1934-1937.
86. Fu, G-S., Phlypo, R., Anderson, M., Li, X-L., Adalı, T. Proc IEEE ICASSP 2013. Vancouver, BC: 2013. Algorithms for markovian source separation by entropy rate minimization; p. 3248-3252.
87. Calhoun V, Adalı T. Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE Rev Biomed Eng. 2012; 5:60–73. [PubMed: 23231989]
88. Allen E, Erhardt E, Damaraju E, Gruner W, Segall J, Silva R, Havlicek M, Rachakonda S, Fries J, Kalyanam R, et al. A baseline for the multivariate comparison of resting-state networks. Front Syst Neurosci. 2011; 5
89. Allen E, Damaraju E, Plis S, Erhardt E, Eichele T, Calhoun V. Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex. 2012
90. Yaesoubi M, Allen E, Miller R, Calhoun V. Dynamic coherence analysis of resting fMRI data to jointly capture state-based phase, frequency, and time-domain information. NeuroImage. Oct.2015 120:133–142. [PubMed: 26162552]
91. Calhoun V, Miller R, Pearlson G, Adalı T. The Chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. Neuron. 2014; 84(2):262–274. [PubMed: 25374354]
92. Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Penna SD, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold

DA, de Pasquale F, Sporns O, Walter M, Chang C. Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage. 2013; 80:360–378. [PubMed: 23707587]

93. Erhardt E, Rachakonda S, Bedrick E, Allen E, Adalı T, Calhoun V. Comparison of multi-subject ICA methods for analysis of fMRI data. Hum Brain Mapp. 2011; 32(12):2075–2095. [PubMed: 21162045]
94. Michael A, Anderson M, Miller R, Adalı T, Calhoun V. Preserving subject variability in group fMRI analysis: Performance evaluation of GICA versus IVA. Front Syst Neurosci. 2014; 8(106)
95. Lawley D, Maxwell A. Factor analysis as a statistical method. J R Stat Soc. 1962; 12(3):209–229.
96. Beckmann C, Smith S. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imag. Feb; 2004 23(2):137–152.
97. Attias H. Independent factor analysis. Neural Comput. May; 1999 11(4):803–851. [PubMed: 10226184]
98. Choudrey, R. PhD dissertation. University of Oxford; 2002. Variational methods for bayesian independent component analysis.
99. Roberts, S., Choudrey, R. Proc DSMML Workshop 2004. Sheffield, UK: 2004. Bayesian independent component analysis with prior constraints: An application in biosignal analysis; p. 159-179.
100. Lappalainen, H., Honkela, A. Bayesian non-linear independent component analysis by multi-layer perceptrons. In: Girolami, M., editor. Advances in Independent Component Analysis. London, UK: Springer; 2000. p. 93-121.ch. 6
101. Valpola H, Karhunen J. An unsupervised ensemble learning method for nonlinear dynamic state-space models. Neural Comput. Nov; 2002 14(11):2647–2692.
102. Beckmann C, Smith S. Tensorial extensions of independent component analysis for multisubject FMRI analysis. NeuroImage. 2005; 25(1):294–311. [PubMed: 15734364]
103. Cardoso J, Souloumiac A. Blind beamforming for non-Gaussian signals. IEE Proc F. Dec; 1993 140(6):362–370.
104. Lu, W., Rajapakse, J. Proc NIPS2000. Denver, CO: 2000. Constrained independent component analysis; p. 570-576.
105. Lu W, Rajapakse J. ICA with reference. Neurocomputing. 2006; 69(16–18):2244–2257.
106. Lin Q-H, Zheng Y-R, Yin F-L, Liang H, Calhoun V. A fast algorithm for one-unit ICA-R. Inform Sciences. 2007; 177(5):1265–1275.
107. Anderson, M., Fu, G-S., Phlypo, R., Adalı, T. Proc IEEE ICASSP 2013. Vancouver, BC: 2013. Independent vector analysis, the Kotz distribution, and performance bounds; p. 3243-3247.
108. Kettenring J. Canonical analysis of several sets of variables. Biometrika. Dec; 1971 58(3):433–451.
109. De Bie, T., Cristianini, N., Rosipal, R. Eigenproblems in pattern recognition. In: Corrochano, E., editor. Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics. 1st. New York, NY: Springer; 2005. p. 129-167.ch. 5
110. Wegelin, J. A survey of partial least squares (PLS) methods, with emphasis on the two-block case. Dept. of Statistics, University of Washington; Seattle, WA, 98195, USA: Mar. 2000 Tech. Rep. 371
111. Wold, H. Nonlinear iterative partial least squares (NIPALS) estimation procedures. In: Wold, H., Lyttkens, E., editors. Bulletin of the International Statistical Institute. Vol. 42. London, UK: International Statistical Institute (ISI); 1969. p. 91-424.
112. Lu, H., Plataniotis, K., Venetsanopoulos, A. Multilinear Subspace Learning: Dimensionality Reduction of Multidimensional Data. 1st. Boca Raton, FL: CRC Press; Dec. 2013 ser Machine Learning & Pattern Recognition
113. Rosipal, R., Krämer, N. Proc SLSFS 2005. Bohinj, Slovenia: 2005. Overview and recent advances in partial least squares; p. 34-51.
114. Höskuldsson A. PLS regression methods. J Chemometrics. Jun; 1988 2(3):211–228.
115. Stigler S. Gauss and the invention of least squares. Ann Statist. May; 1981 9(3):465–474.

116. Haufe S, Meinecke F, Görgen K, Dähne S, Haynes J-D, Blankertz B, Bießmann F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. NeuroImage. 2014; 87:96–110. [PubMed: 24239590]
117. Dähne S, Meinecke F, Haufe S, Höhne J, Tangermann M, Müller K-R, Nikulin V. SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. NeuroImage. 2014; 86:111–122. [PubMed: 23954727]
118. Dähne S, Nikulin V, Ramírez D, Schreier P, Müller K-R, Haufe S. Finding brain oscillations with power dependencies in neuroimaging data. NeuroImage. 2014; 96:334–348. [PubMed: 24721331]
119. Dähne S, Bießmann F, Meinecke F, Mehnert J, Fazli S, Müller K-R. Integration of multivariate data streams with bandpower signals. IEEE Trans Multimedia. Aug; 2013 15(5):1001–1013.
120. Chen, J., Calhoun, V., Liu, J. Proc IEEE EMBC 2012. San Diego, CA: 2012. ICA order selection based on consistency: Application to genotype data; p. 360-363.
121. Dähne S, Bießmann F, Samek W, Haufe S, Goltz D, Gundlach C, Villringer A, Fazli S, Müller K-R. Multivariate machine learning methods for fusing multimodal functional neuroimaging data. Proc IEEE. Sep; 2015 103(9):1507–1530.
122. Arteaga, F., Gallarza, M., Gil, I. A new multiblock PLS based method to estimate causal models: application to the post-consumption behavior in tourism. In: Esposito Vinzi, V.Chin, W.Henseler, J., Wang, H., editors. Handbook of Partial Least Squares: Concepts, Methods and Applications. 1st. New York, NY: Springer; 2010. p. 141-169.ch. 6
123. Anderson, M., Li, X-L., Adalı, T. Proc LVA/ICA 2010. St Malo, France: 2010. Nonorthogonal independent vector analysis using multivariate gaussian model; p. 354-361.
124. Du Y, Fan Y. Group information guided ICA for fMRI data analysis. NeuroImages. 2013; 69:157–197.
125. Calhoun V, Adalı T, Kiehl K, Astur R, Pekar J, Pearlson G. A method for multi-task fMRI data fusion applied to schizophrenia. Hum Brain Map. 2006; 27(7):598–610.
126. Calhoun V, Adalı T. Feature-based fusion of medical imaging data. IEEE Trans Inf Technol Biomed. Sep; 2009 13(5):711–720. [PubMed: 19273016]
127. Calhoun V, Adalı T, Pearlson G, Pekar J. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Hum Brain Mapp. 2001; 13(1): 43–53. [PubMed: 11284046]
128. Baker, B., Silva, R., Calhoun, V., Sarwate, A., Plis, S. IEEE MLSP 2015. Boston, MA: 2015. Large scale collaboration with autonomy: Decentralized data ICA; p. 1-6.
129. Koles Z. The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroen Clin Neuro. 1991; 79(6):440–447.
130. Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng. Dec; 2000 8(4):441–446. [PubMed: 11204034]
131. Samek W, Kawanabe M, Müller KR. Divergence-based framework for common spatial patterns algorithms. IEEE Rev Biomed Eng. 2014; 7:50–72. [PubMed: 24240027]
132. Lotte F, Guan C. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. IEEE Trans Biomed Eng. Feb; 2011 58(2):355–362. [PubMed: 20889426]
133. Samek W, Vidaurre C, Müller K-R, Kawanabe M. Stationary common spatial patterns for braincomputer interfacing. J Neural Eng. 2012; 9(2):026013. [PubMed: 22350439]
134. Groves AR, Beckmann CF, Smith SM, Woolrich MW. Linked independent component analysis for multimodal data fusion. NeuroImage. 2011; 54(3):2198–2217. [PubMed: 20932919]
135. Fu, G-S., Anderson, M., Adalı, T. Proc CISS. Baltimore, MD: 2015. Independent vector analysis by entropy rate bound minimization.
136. Hyvärinen, A., Hurri, J., Hoyer, P. Natural Image Statistics: A Probabilistic Approach to Early Computational Vision. 1st. Vol. 39. Springer; 2009. ser Computational Imaging and Vision
137. Silva, R., Plis, S., Pattichis, M., Adalı, T., Calhoun, V. Proc OHBM 2015. Honolulu, HI: Incorporating second-order statistics in multidataset independent subspace analysis. Poster 3743

138. von Bünau P, Meinecke FC, Király FC, Müller K-R. Finding stationary subspaces in multivariate time series. Phys Rev Lett. Nov.2009 103:214101. [PubMed: 20366040]
139. Pham DT. Joint approximate diagonalization of positive definite Hermitian matrices. SIAM J Matrix Anal Appl. 2001; 22(4):1136–1152.
140. Hara, S., Kawahara, Y., Washio, T., von Bünau, P. Proc ICONIP 2010. Sydney, Australia: 2010. Stationary subspace analysis as a generalized eigenvalue problem; p. 422-429.
141. McKeown M, Hansen L, Sejnowski T. Independent component analysis of functional MRI: what is signal and what is noise? Curr Opin Neurobiol. 2003; 13(5):620–629. [PubMed: 14630228]
142. Calhoun V, Adalı T. Unmixing fMRI with independent component analysis. IEEE Eng Med Biol Mag. Mar; 2006 25(2):79–90.
143. Biswal B, Ulmer J. Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. J Comput Assist Tomogr. 1999; 23(2):265–271. [PubMed: 10096335]
144. Eichele T, Specht K, Moosmann M, Jongsma M, Quiroga R, Nordby H, Hugdahl K. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. Proc Natl Acad Sci. 2005; 102(49):17798–17803. [PubMed: 16314575]
145. Bridwell D, Kiehl K, Pearlson G, Calhoun V. Patients with schizophrenia demonstrate reduced cortical sensitivity to auditory oddball regularities. Schizophr Res. 2014; 158(13):189–194. [PubMed: 25034764]
146. Onton J, Westerfield M, Townsend J, Makeig S. Imaging human EEG dynamics using independent component analysis. Neurosci Biobehav Rev. 2006; 30(6):808–822. [PubMed: 16904745]
147. Calhoun V, Adalı T, Pearlson G, van Zijl P, Pekar J. Independent component analysis of fMRI data in the complex domain. Magnet Reson Med. 2002; 48(1):180–192.
148. Petersen, K., Hansen, L., Kolenda, T., Rostrup, E., Strother, S. Proc ICA 2000. Helsinki, Finland: 2000. On the independent components of functional neuroimages; p. 615-620.
149. Caprihan A, Abbott C, Yamamoto J, Pearlson G, Perrone-Bizzozero N, Sui J, Calhoun V. Source-based morphometry analysis of group differences in fractional anisotropy in schizophrenia. Brain Connect. 2011; 1(2):133–145. [PubMed: 22180852]
150. Brookes M, Woolrich M, Luckhoo H, Price D, Hale J, Stephenson M, Barnes G, Smith S, Morris P. Investigating the electrophysiological basis of resting state networks using magnetoencephalography. Proc Natl Acad Sci. 2011; 108(40):16783–16788. [PubMed: 21930901]
151. Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, Van Essen DC, Feinberg DA, Yacoub ES, Ugurbil K. Temporally-independent functional modes of spontaneous brain activity. Proc Natl Acad Sci. 2012; 109(8): 3131–3136. [PubMed: 22323591]
152. Makeig, S., Bell, A., Jung, T-P., Sejnowski, T. Proc NIPS1995. Denver, CO: 1995. Independent component analysis of electroencephalographic data; p. 145-151.
153. Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage. 2007; 34(4):1443–1449. [PubMed: 17188898]
154. Tang A, Sutherland M, McKinney C. Validation of SOBI components from high-density EEG. NeuroImage. 2005; 25(2):539–553. [PubMed: 15784433]
155. Calhoun V, Sui J, Kiehl K, Turner J, Allen E, Pearlson G. Exploring the psychosis functional connectome: aberrant intrinsic networks in schizophrenia and bipolar disorder. Front Psychiatry. 2012; 2(75)
156. Jafri M, Pearlson G, Stevens M, Calhoun V. A method for functional network connectivity among spatially independent resting-state components in schizophrenia. NeuroImage. 2008; 39(4):1666–1681. [PubMed: 18082428]
157. Damaraju E, Allen E, Belger A, Ford J, McEwen S, Mathalon D, Mueller B, Pearlson G, Potkin S, Preda A, Turner J, Vaidya J, van Erp T, Calhoun V. Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage: Clin. 2014; 5:298–308. [PubMed: 25161896]

158. Friman O, Carlsson J, Lundberg P, Borga M, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. Magn Reson Med. Feb; 2001 45(2):323–330. [PubMed: 11180440]

159. Correa N, Adalı T, Li Y-O, Calhoun V. Canonical correlation analysis for data fusion and group inferences. IEEE Signal Process Mag. 2010; 27(4):39–50. [PubMed: 20706554]

160. Li Y-O, Eichele T, Calhoun V, Adalı T. Group study of simulated driving fMRI data by multiset canonical correlation analysis. J Signal Process Syst. 2012; 68(1):31–48. [PubMed: 23750290]

161. Lin Z, Zhang C, Wu W, Gao X. Frequency recognition based on canonical correlation analysis for SSVEP-Based BCIs. IEEE Trans Biomed Eng. Dec; 2006 53(12):2610–2614. [PubMed: 17152442]

162. Zhang Y, Zhou G, Jin J, Wang X, Cichocki A. Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis. Int J Neural Syst. 2014; 24(4):1450013. [PubMed: 24694168]

163. Nestor P, O'Donnell B, McCarley R, Niznikiewicz M, Barnard J, Jen Shen Z, Bookstein F, Shenton M. A new statistical method for testing hypotheses of neuropsychological MRI relationships in schizophrenia: partial least squares analysis. Schizophr Res. Jan; 2002 53(1–2): 57–66. [PubMed: 11728838]

164. Lee J-H, Lee T-W, Jolesz F, Yoo S-S. Independent vector analysis (IVA): Multivariate approach for fMRI group study. NeuroImage. 2008; 40(1):86–109. [PubMed: 18165105]

165. Gopal S, Miller R, Michael A, Adalı T, Cetin M, Rachakonda S, Bustillo J, Cahill N, Baum S, Calhoun V. Spatial variance in resting fMRI networks of schizophrenia patients: An independent vector analysis. Schizophr Bull. 2016; 42(1):152–160. [PubMed: 26106217]

166. Schlösser R, Nenadic I, Wagner G, Güllmar D, von Consbruch K, Köhler S, Schultz C, Koch K, Fitzek C, Matthews P, Reichenbach J, Sauer H. White matter abnormalities and brain activation in schizophrenia: A combined DTI and fMRI study. Schizophr Res. 2007; 89(13):1–11. [PubMed: 17085018]

167. Wang F, Kalmar J, He Y, Jackowski M, Chepenik L, Edmiston E, Tie K, Gong G, Shah M, Jones M, Uderman J, Constable R, Blumberg H. Functional and structural connectivity between the perigenual anterior cingulate and amygdala in bipolar disorder. Biol Psychiat. 2009; 66(5):516–521. [PubMed: 19427632]

168. Kolb B, Whishaw I. Brain plasticity and behavior. Annu Rev Psychol. 1998; 49(1):43–64. [PubMed: 9496621]

169. Keightley M, Chen J-K, Ptito A. Examining the neural impact of pediatric concussion: a scoping review of multimodal and integrative approaches using functional and structural MRI techniques. Curr Opin Pediatr. 2012; 24(6):709716.

170. Toosy A, Ciccarelli O, Parker G, Wheeler-Kingshott C, Miller D, Thompson A. Characterizing functionstructure relationships in the human visual system with functional MRI and diffusion tensor imaging. NeuroImage. 2004; 21(4):1452–1463. [PubMed: 15050570]

171. Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci. 2009; 10(3):186–198. [PubMed: 19190637]

172. Camchong J, MacDonald A, Bell C, Mueller B, Lim K. Altered functional and anatomical connectivity in schizophrenia. Schizophrenia Bull. 2011; 37(3):640–650.

173. Sui J, Pearlson G, Caprihan A, Adalı T, Kiehl K, Liu J, Yamamoto J, Calhoun V. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA + joint ICA model. NeuroImage. 2011; 57(3):839–855. [PubMed: 21640835]

174. Lahat D, Adali T, Jutten C. Multimodal data fusion: An overview of methods, challenges, and prospects. Proc IEEE. Sep; 2015 103(9):1449–1477.

175. Chen X, Wang ZJ, McKeown M. Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods. IEEE Signal Process Mag. May; 2016 33(3):86–107.

176. Lutti A, Dick F, Sereno MI, Weiskopf N. Using high-resolution quantitative mapping of R1 as an index of cortical myelination. NeuroImage. 2014; 93(Part 2):176–188. [PubMed: 23756203]

177. Sereno M, Lutti A, Weiskopf N, Dick F. Mapping the human cortical surface by combining quantitative T1 with retinotopy. Cereb Cortex. 2013; 23(9):2261–2268. [PubMed: 22826609]

178. Hutton C, Draganski B, Ashburner J, Weiskopf N. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. NeuroImage. 2009; 48(2):371–380. [PubMed: 19559801]

179. Johansen-Berg, H., Behrens, T. Diffusion MRI. San Diego, CA: Academic Press; 2009.

180. Li J, Shi Y, Toga AW. Mapping brain anatomical connectivity using diffusion magnetic resonance imaging: Structural connectivity of the human brain. IEEE Signal Process Mag. May; 2016 33(3):36–51. [PubMed: 27212872]

181. Villien M, Wey H-Y, Mandeville JB, Catana C, Polimeni JR, Sander CY, Zürcher NR, Chonde DB, Fowler JS, Rosen BR, Hooker JM. Dynamic functional imaging of brain glucose utilization using fPET-FDG. NeuroImage. 2014; 100:192–199. [PubMed: 24936683]

182. Villien M. Simultaneous functional imaging using fPET and fMRI. EJNMMI Physics. 2015; 2(1)

183. Zaidi H, Becker M. The promise of hybrid PET/MRI: Technical advances and clinical applications. IEEE Signal Process Mag. 2016; 33(3):67–85.

184. Yücel MA, Selb J, Boas DA, Cash SS, Cooper RJ. Reducing motion artifacts for long-term clinical NIRS monitoring using collodion-fixed prism-based optical fibers. NeuroImage. 2014; 85(Part 1):192–201. [PubMed: 23796546]

185. Giacometti P, Diamond SG. Correspondence of electroencephalography and near-infrared spectroscopy sensitivities to the cerebral cortex using a high-density layout. Neurophotonics. 2014; 1(2):025001. [PubMed: 25558462]

186. Sui J, Yu Q, He H, Pearlson G, Calhoun V. A selective review of multimodal fusion methods in schizophrenia. Front Hum Neurosci. 2012; 6(27)

187. Hao X, Xu D, Bansal R, Dong Z, Liu J, Wang Z, Kangarlu A, Liu F, Duan Y, Shova S, Gerber A, Peterson B. Multimodal magnetic resonance imaging: The coordinated use of multiple, mutually informative probes to understand brain structure and function. Hum Brain Mapp. 2013; 34(2): 253–271. [PubMed: 22076792]

188. Liu J, Ghassemi M, Michael A, Boutte D, Wells W, Perrone-Bizzozero N, Macciardi F, Mathalon D, Ford J, Potkin S, Turner J, Calhoun V. An ICA with reference approach in identification of genetic variation and associated brain networks. Front Hum Neurosci. 2012; 6(21)

189. Meda S, Narayanan B, Liu J, Perrone-Bizzozero N, Stevens M, Calhoun V, Glahn D, Shen L, Risacher S, Saykin A, Pearlson G. A large scale multivariate parallel ICA method reveals novel imaginggenetic relationships for alzheimer's disease in the ADNI cohort. NeuroImage. 2012; 60(3):1608–1621. [PubMed: 22245343]

190. Sui J, He H, Yu Q, Chen J, Rogers J, Pearlson G, Mayer A, Bustillo J, Canive J, Calhoun V. Combination of resting state fMRI, DTI and sMRI data to discriminate schizophrenia by N-way MCCA+jICA. Front Hum Neurosci. 2013; 7(235)

191. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. NeuroImage. 2010; 52(3):1059–1069. [PubMed: 19819337]

192. Yu Q, Allen E, Sui J, Arbabshirani M, Pearlson G, Calhoun V. Brain connectivity networks in schizophrenia underlying resting state functional magnetic resonance imaging. Curr Top Med Chem. 2012; 12(21):2415–2425. [PubMed: 23279180]

193. Yu Q, Sui J, Liu J, Plis S, Kiehl K, Pearlson G, Calhoun V. Disrupted correlation between low frequency power and connectivity strength of resting state brain networks in schizophrenia. Schizophr Res. 2013; 143(1):165–171. [PubMed: 23182443]

194. Miller RL, Yaesoubi M, Turner JA, Mathalon D, Preda A, Pearlson G, Adalı T, Calhoun VD. Higher dimensional meta-state analysis reveals reduced resting fMRI connectivity dynamism in schizophrenia patients. PLoS ONE. Mar; 2016 11(3):1–24.

195. Lee D, Seung H. Learning the parts of objects by non-negative matrix factorization. Nature. 1999; 401(6755):788–791. [PubMed: 10548103]

196. Yates F. The analysis of replicated experiments when the field results are incomplete. Empire J Exp Agr. 1933; 1(2):129–142.

197. Leeuw, Jde, Young, FW., Takane, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. Psychometrika. 1976; 41(4):471–503.

198. Bro R. PARAFAC. Tutorial and applications. Chemometr Intell Lab. 1997; 38(2):149–171.

199. Phan, AH., Cichocki, A. Proc ISNN 2008, Part II. Beijing, China: 2008. Fast and efficient algorithms for non-negative tucker decomposition; p. 772-782.

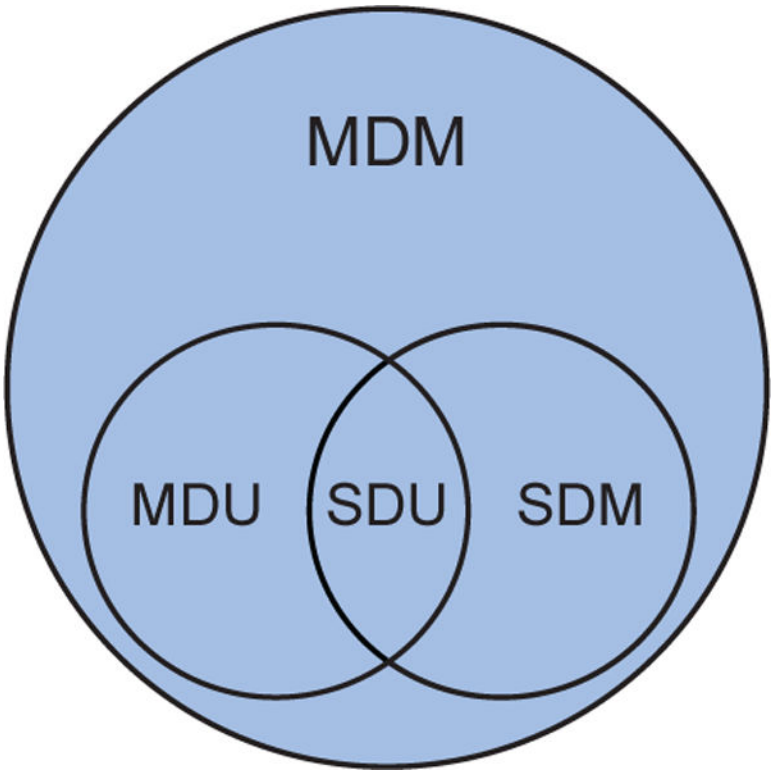

Fig. 1. Venn diagram of the structure of BSS subproblems

Single dataset unidimensional (SDU) problems are special cases of multiple dataset unidimensional (MDU), single dataset multidimensional (SDM), and multiple dataset multidimensional (MDM) problems, while MDU and SDM problems are special cases of MDM problems.

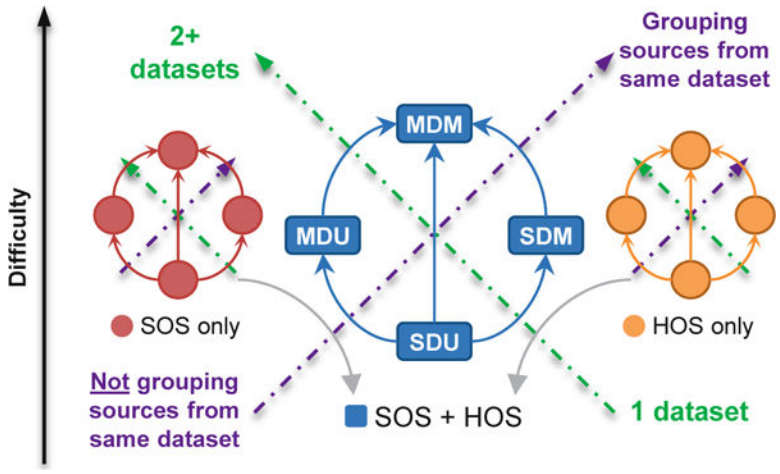

Fig. 2. Hierarchy of linear BSS models

Historically, BSS models have been made more general by: A) increasing the number of datasets which can be *jointly* analyzed [see Layout and Subsets]; B) moving from *isolated* sources to *groups* of sources in the same dataset [see Subsets]; C) exploiting SOS, HOS, or both [see Type of Statistics]. The arrows indicate the directions of increasing difficulty, model complexity, and generality. Highly general models can address MDM problems by incorporating lenient modeling choices.

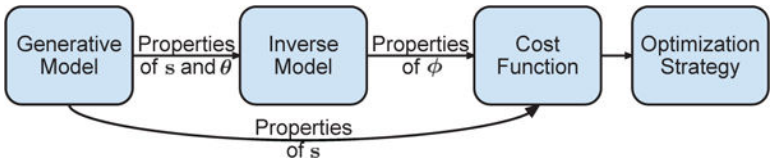

Fig. 3. Steps involved in moving from a model to an algorithm

The properties of **s** and $\boldsymbol{\theta}$, selected during the modeling step, impose requirements on the inverse model and cost function. The parameters $\boldsymbol{\phi}$ of the inverse model are the inputs to the cost function and change according to the selected optimization strategy.

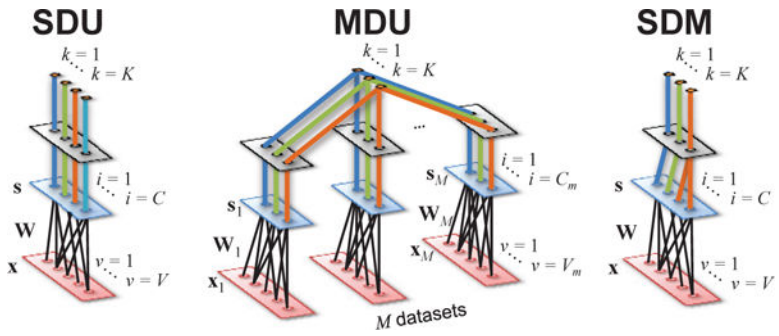

Fig. 4. General inverse models for linear SDU, MDU, and SDM problems

The models are presented as graphical structures where the lower layer corresponds to one $1 \times V$ observation of the input data **x**. The middle layer represents the C sources **s** obtained by linear transformation of **x** through the unmixing matrix **W**. The top layer establishes the type of interaction between sources as described in Section II-B, forming K *subspaces* $\mathbf{s}_k$. The problems are described in Section II-A.

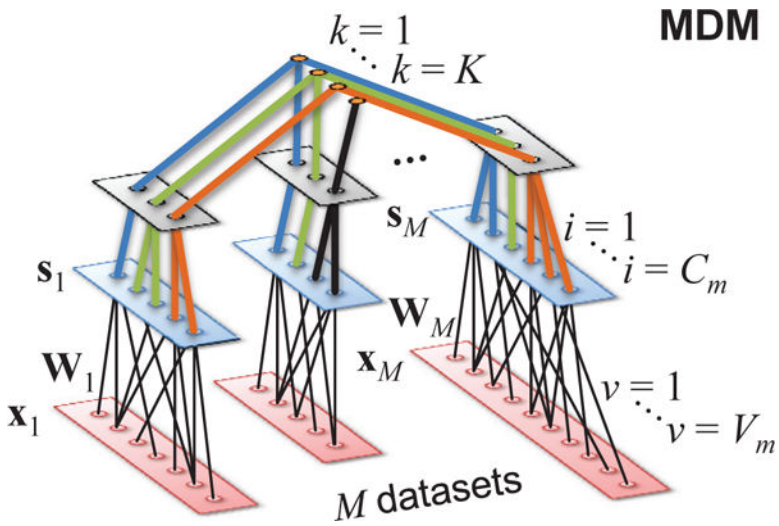

Fig. 5. General inverse model for the linear MDM problem

The lower layer corresponds to one $V_m \times 1$ observation of each input data stream $\mathbf{x}_m$, reflecting the different intrinsic dimensionality (V_m) of each dataset. The middle layer represents the C_m sources. The top layer establishes the K *subspaces* $\mathbf{s}_k$, which may be dataset-specific ($k = K$ in the figure) or span through many datasets, illustrating the different compositions permitted.

TABLE I

Typical choices in BSS modeling. Choices relevant to this work are color-coded to match Fig. 2.

Model Property	**Choice to Make**	**Options**
Sources	Model order	$C \in \mathbb{N}$ (per dataset)
Mixture	Mapping	linear or non-linear
	Layout	**fully-connected or structured**
(optional)	Parameter constraints	Orthogonality, sparsity, match a template, min. total variation
Statistical Relationship	Subsets (*which*)	grouping of sources and $K \leq C$
	Interactions (*how*)	graph (directed or not, acyclic/cyclic, tree/hierarchy), and/or sample dependence
	Type of Stats. (*how*)	SOS, HOS, or both
(optional)	Source constraints	Sparsity, match a template, min. total variation