DECENTRALIZED INDEPENDENT VECTOR ANALYSIS

Nikolas P. Wojtalewicz[⋆] *Rogers F. Silva*[†] *Vince D. Calhoun*[†] *Anand D. Sarwate*[‡] *Sergey M. Plis*[†]

[⋆] New College of Florida [†] Mind Research Network [‡] Rutgers University

ABSTRACT

Independent vector analysis (IVA) is an approach for joint blind source separation of several data sets that learns simultaneous unmixing transforms for each set. It assumes corresponding sources from different data sets to be statistically dependent. One of the main advantages is IVA's ability to retain subject-specific differences while simplifying comparison across subjects as the resulting components have the same order. The latter is an instrumental property for enabling collaboration between remote sites without sharing their data, which may be required because of ethical, privacy or efficiency concerns. This paper proposes a new decentralized algorithm for IVA that exploits the structure of the objective function. A centralized aggregator coordinates IVA algorithms at multiple sites using message passing, parallelizing the computation and limiting the amount of communication. Thus, the algorithm enables a plausibly private collaboration across multiple sites. Besides enabling analysis of decentralized data, our approach improves the running time of IVA when used locally.

***Index Terms*—** IVA, distributed signal processing, blind source separation, decentralized data

1. INTRODUCTION

Brain imaging provides a wealth of information about the function and structure of the brain, offering unique views into the complexities of this highly interconnected organ. In order to parse this information into useful bits, numerous approaches have been developed to decompose brain images. Methods such as principal component analysis (PCA) [1] and independent component analysis (ICA) [2, 3, 4] have found much success in applications ranging from functional [5, 6, 7] and structural [8] magnetic resonance imaging (fMRI and sMRI, respectively) to electro- [9, 10, 11] and magneto-encephalography [12, 13] (EEG and MEG, respectively).

PCA is a second-order statistical method, and as a result, excels when the data can be expressed with just the mean and covariance. It fails, however, when the data has higher order relations. ICA, on the other hand, is a higher-order statistical method and, as such, can solve problems which PCA cannot. Furthermore, challenges arise when we consider the utility of either of these methods for group analyses. Quite commonly, the data from each subject is organized into separate datasets but methods such as PCA and ICA are not designed to account for multiple datasets.

One approach to address this challenge is to apply ICA separately to each subject, which is known to return independent sources in arbitrary order and require post-processing matching, a daunting and intractable task for large number of subjects. Modifications of PCA and ICA have been proposed to account for multiple subjects, notably group PCA [14] and group ICA (GICA) [6]. Both methods rely on temporal concatenation of the datasets and make a fairly strong assumption that the underlying source maps are identical across subjects. Clearly, it is far more desirable (and expected) that the source maps contain subject-specific features. Such features could help determine personalized treatment options, guide therapeutic decision making, and determine patient outcomes in populations with disabling mental illnesses. With that in mind, two approaches have emerged: group information-guided ICA (GIG-ICA) [15], which employs subject-specific ICA based on the initial results from GICA and, more recently, independent vector analysis [16, 17]. IVA's strength is in its inherent structure, which allows corresponding sources from different subjects to be *similar* (i.e., dependent) rather than identical. IVA gives subject-specific information while enabling identification of dependent sources straightforwardly. This allows for subject maps to contain unique information while still being linked across different subjects.

Despite strong trends in the neuroimaging community towards data-sharing, privacy and ethical concerns preclude many datasets from being shared. Often these are the valuable data that either come from rare disorders or from historical data collected without a consent that allows wide sharing. This situation calls for algorithms which can process data stored at remote sites without requiring raw data transmission and limiting the information. IVA's property of ordering the components for all datasets could help research under aforementioned circumstances of unshareable data: metanalysis of the component properties becomes possible even without sharing the data. However, the algorithm is centralized. Therefore, it would be very useful to find a method to distribute the IVA algorithm such that it could operate on decentralized data.

In this work, we introduce the decentralized IVA (dIVA) algorithm to accomplish this task. dIVA allows numerous institutions to not only collaborate on the same IVA problem but also spread the computational load any single institution must carry out in order to preform such high dimensional analysis, improving its execution time. dIVA enables the previously not available use case of analyzing data without the need of pulling it to a central location. Furthermore, it improves upon the speed of local IVA when used locally instead. We introduce the traditional IVA algorithm in section 2. We then describe our dIVA algorithm in section 3. Finally, we present performance results in section 4, and provide concluding remarks and future directions in section 5.

Notation: Matrices and vectors will be given in boldface, with dimensions specified to clarify. Also, $[N] = \{1, 2, \ldots, N\}$, $\circ$ is the Hadamard product, $\mathbf{C}^{\circ -1}$ is element-wise inverse, and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm.

This work was sponsored in part by NSF award CCF-1453432, NIH award 1R01DA040487-01A1, and DARPA and SSC Pacific under contract No. N66001-15-C-4070.

2. INDEPENDENT VECTOR ANALYSIS

Model. We first describe Independent Vector Analysis (IVA) operating on a single collection of data sets. IVA is a well-studied method for joint (simultaneous) blind source separation (BSS) of multiple data sets [18]. We assume the site has K data sets, one per subject, each organized in a matrix $\mathbf{x}_k \in \mathbb{R}^{N\times R}$ representing R observations from N sensors. These can be compactly represented as a data tensor $\mathbf{X} \in \mathbb{R}^{N\times R\times K}$. The signal model for each data set observation is $\mathbf{x}_k = \mathbf{A}_k\mathbf{s}_k$, where $\mathbf{A}_k \in \mathbb{R}^{N\times N}$ is a mixing matrix for the k-th data set, and $\mathbf{s}_k$ are the N underlying source signals that were mixed by $\mathbf{A}_k$ to form the observations.

As in ICA, IVA models the N sources $\mathbf{s}_k = [s_{k,1}\ s_{k,2}\ \cdots\ s_{k,N}]^\top$ as independent random variables. To perform the separation of a single data set, the algorithm tries to minimize a multiway information functional (I) among the latent source groups $\{\mathbf{s}^n \in \mathbb{R}^{K\times R} : n \in [N]\}$:

$$I = \int p(\mathbf{s}) \log\left(\frac{p(\mathbf{s})}{\prod_{n=1}^{N} p(\mathbf{s}^n)}\right) d\mathbf{s}, \tag{1}$$

where $\mathbf{s}^n = [s_{1,n}\ s_{2,n}\ \cdots\ s_{K,n}]^\top$ is the n-th set (or group) of corresponding sources across the data sets, $\mathbf{s}$ is the entire collection of all N sources from all K data sets, and $p(\cdot)$ is the probability density function (pdf). As with ICA, the overall objective is to estimate *unmixing matrices* $\{\mathbf{W}_k = \mathbf{A}_k^{-1} : k \in [K]\}$ that recover the sources $\mathbf{s}_k = \mathbf{W}_k\mathbf{x}_k$: in IVA however, $p(\mathbf{s}^n)$ can account for dependencies between the n-th sources across data sets. We denote the collection of unmixing matrices by the unmixing tensor $\mathbf{W} \in \mathbb{R}^{N\times N\times K}$. Then,

$$I(\mathbf{W}) = \sum_{n=1}^{N} h(\mathbf{s}^n) - h(\mathbf{X}) - \sum_{k=1}^{K} \log|\det \mathbf{W}_k|, \tag{2}$$

where $h(\cdot)$ denotes the differential entropy of a random variable. In the minimization of I with respect to $\mathbf{W}$, $h(\mathbf{X})$ acts as a constant term, so we arrive at a modified objective,

$$\tilde{I}(\mathbf{W}) = \sum_{n=1}^{N} h(\mathbf{s}^n) - \sum_{k=1}^{K} \log|\det \mathbf{W}_k|. \tag{3}$$

This highlights that the goal in IVA is to separate each of the K data sets in a way such that the n-th sources across data sets are *dependent*. Thus, IVA forms N K-dimensional source *groups*.

Algorithm. Algorithms for solving the IVA unmixing problem make assumptions about the dependence structure of the underlying source groups $\mathbf{s}^n$. For example, IVA-Laplace (IVA-L) [16, 19] and IVA-Gaussian (IVA-G) [20] make multivariate Laplace and Gaussian assumptions on the source groups, respectively. In this paper we develop algorithms for IVA-L on decentralized data; the traditional IVA-L algorithm is given in Algorithm 1.

The IVA-L algorithm iteratively refines an estimate of the underlying sources in a state variable $\mathbf{Y} \in \mathbb{R}^{N\times R\times K}$, which collects the sources $\mathbf{Y}_k \in \mathbb{R}^{N\times R}$ for each data set k. The algorithm assumes a Laplacian prior on the source groups, such that $h(\mathbf{s}_r^n) \approx \|\mathbf{y}_r^n\|$, and independent observations across $r \in [R]$. This yields a proxy for the information functional:

$$\mathsf{cost}(\mathbf{W}) = \frac{1}{R}\sum_{r=1}^{R}\sum_{n=1}^{N}\|\mathbf{y}_r^n\| - \sum_{k=1}^{K}\log|\det \mathbf{W}_k|. \tag{4}$$

For a given $\mathbf{Y}$, it computes the norms $\|\mathbf{y}_r^n\|$ across the K data sets for each source and observation, and updates the objective estimate. For convenience, we define the matrix $\mathbf{C} \in \mathbb{R}^{N\times R}$ by $\mathbf{C} = \sqrt{\sum_{k=1}^{K}\mathbf{Y}_k \circ \mathbf{Y}_k}$, containing the norms $\{\|\mathbf{y}_r^n\|\}$, and $\mathbf{C}^{\circ -1}$, the element-wise inverse of $\mathbf{C}$, containing $\{1/\|\mathbf{y}_r^n\|\}$. Then the cost is

$$\mathsf{cost}(\mathbf{W}) = \frac{1}{R}\sum_{r=1}^{R}\sum_{n=1}^{N} C_{n,r} - \sum_{k=1}^{K}\log|\det \mathbf{W}_k|. \tag{5}$$

Next, we take a relative gradient step ($\nabla\mathbf{W}$) with a modified step size α. Finally, the unmixing matrix $\mathbf{W}$ is updated, and the process continues iteratively until either the maximum number of iterations is reached or the change in $\mathbf{W}$ is sufficiently small.

Algorithm 1 Independent Vector Analysis - Laplace IVA-L

Require: data $\mathbf{X} \in \mathbb{R}^{N\times R\times K}$, tolerance level $t = 10^{-6}$, maximum iterations J, Armijo condition constant $c = 10^{-10}$, $\alpha = 1$
$\mathbf{W} \in \mathbb{R}^{N\times N\times K}$, ▷ e.g., $\mathbf{W}_k = \mathbf{I}$
$\|\nabla\mathbf{W}_k\|_{\mathrm{F}} = 1$ for $k \in [K]$, $\|\nabla\mathbf{W}_{\mathrm{prev}}\|_{\mathrm{F}}^2 = \|\nabla\mathbf{W}\|_{\mathrm{F}}^2$
$\mathsf{cost}(0) = \infty$, $j = 1$
while $j < J$, $\max_k\|\alpha\nabla\mathbf{W}_k\|_{\mathrm{F}} > t$ **do**
  $\mathbf{Y}_k = \mathbf{W}_k\mathbf{x}_k$ for $k \in [K]$ ▷ Source estimates
  $\mathbf{C} = \sqrt{\sum_{k=1}^{K}\mathbf{Y}_k \circ \mathbf{Y}_k}$ and $\mathsf{cost}(j)$ using (5).
  while $\mathsf{cost}(j) > \mathsf{cost}(j-1) - \alpha c\|\nabla\mathbf{W}\|_{\mathrm{F}}^2$ **do**
    $\alpha = \frac{3}{4}\alpha$ ▷ Initiate backtracking
    $\mathbf{W} = \mathbf{W}_{\mathrm{prev}} + \alpha\nabla\mathbf{W}$
    $\mathbf{Y}_k = \mathbf{W}_k\mathbf{x}_k$ for $k \in [K]$
    $\mathbf{C} = \sqrt{\sum_{k=1}^{K}\mathbf{Y}_k \circ \mathbf{Y}_k}$ and $\mathsf{cost}(j)$ using (5).
  end while
  for all datasets $k = 1 : K$ **do**
    $\phi = \mathbf{C}^{\circ -1} \circ \mathbf{Y}_k$
    $\nabla\mathbf{W}_k = \mathbf{W}_k - \phi\mathbf{Y}_k^T\mathbf{W}_k$ ▷ Relative gradient
  end for
  if $j > 1$ **then**
    $\alpha = \alpha\frac{\|\nabla\mathbf{W}_{\mathrm{prev}}\|_{\mathrm{F}}^2}{\|\nabla\mathbf{W}\|_{\mathrm{F}}^2}$ ▷ Update step size
  end if
  $\mathbf{W}_{\mathrm{prev}} = \mathbf{W}$, $\|\nabla\mathbf{W}_{\mathrm{prev}}\|_{\mathrm{F}}^2 = \|\nabla\mathbf{W}\|_{\mathrm{F}}^2$
  $\mathbf{W} = \mathbf{W}_{\mathrm{prev}} + \alpha\nabla\mathbf{W}$
  $j = j + 1$
end while

3. DECENTRALIZED IVA-L

Our main contribution in this paper is designing an approach for IVA that can operate on decentralized data. In this setting we have P sites, each of which has its own data tensor $\mathbf{X}_p \in \mathbb{R}^{N\times R\times K_p}$ for $p \in [P]$. Note that each site p may have a different number of data sets K_p. As before, the goal is to perform a separation into sources at each site and to find unmixing matrices $\{\mathbf{W}_{k,p} \in \mathbb{R}^{N\times N} : k \in [K_p], p \in [P]\}$ in order to perform the separation $\mathbf{s}_{k,p} = \mathbf{W}_{k,p}\mathbf{x}_{k,p}$. The key challenge in decentralized IVA is that the dependence between the sources is now across the sites as well, i.e. the $K = \sum_{p=1}^{P} K_p$ data sets collectively.

We take a computational model in which a master node, or centralized aggregator, wishes to perform IVA-L on all K data sets. Due to privacy and ethical constraints, the sites cannot transmit their data sets directly to the aggregator, but instead can only send data derivatives. For example, element-wise squaring is a non-linear operation and, thus, the original data is unrecoverable after that. We therefore want to identify computations at the local sites and aggregator that

can effectively find the unmixing matrices minimizing (5). We can rewrite the cost in terms of sums over the sites:

$$\mathsf{cost}(\mathbf{W}) = \frac{1}{R}\sum_{r=1}^{R}\sum_{n=1}^{N}\sqrt{\sum_{p=1}^{P}\|\mathbf{y}_{r,p}^{n}\|^2} - \sum_{p=1}^{P}\sum_{k=1}^{K_p}\log|\det \mathbf{W}_{k,p}|. \tag{6}$$

The aggregator can compute the cost using the squared norms $\|\mathbf{y}_{r,p}^{n}\|^2$ and $d_p = \sum_{k=1}^{K_p}\log|\det \mathbf{W}_{k,p}|$, which are sums over the K_p data sets at each site.

The procedure at the master node is given in Algorithm 6. After initializing, the aggregator asks the local sites to compute their summaries using LocCost (Algorithm 2) and recomputes $\mathbf{C}$ and the cost using (6). To run, the master node needs to know $\mathbf{C}$, the cost, and both the maximum and sum of gradient norms, $\{\|\nabla\mathbf{W}_{k,p}\|_{\mathrm{F}}^2 : k \in [K_p]\}$, from each site, which are computed using LocGrad (Algorithm 3). As in IVA-L (Algorithm 1), the aggregator backtracks if the cost did not meet the Armijo condition for sufficient decrease. Once it finds an appropriate step size, it sends the matrix of inverse norms to the sites, which take a relative gradient step on their local data using LocUpdate (Algorithm 5) after α is updated.

Algorithm 2 LocCost

Input: site index p
Require: local data $\mathbf{X}_p \in \mathbb{R}^{N\times R\times K_p}$, $\mathbf{W}_p \in \mathbb{R}^{N\times N\times K_p}$
$\mathbf{Y}_{k,p} = \mathbf{W}_{k,p}\mathbf{X}_{k,p}$ for $k \in [K_p]$ ▷ Source estimates
$\mathbf{C}_p = \sum_{k=1}^{K_p}\mathbf{Y}_{k,p}\circ\mathbf{Y}_{k,p}$
$d_p = \sum_{k=1}^{K_p}\log|\det \mathbf{W}_{k,p}|$
Return $\mathbf{C}_p$, d_p

Algorithm 3 LocGrad

Input: site index p, $\mathbf{C}^{\circ-1} \in \mathbb{R}^{N\times R}$
Require: $\mathbf{Y}_p \in \mathbb{R}^{N\times R\times K_p}$, $\mathbf{W}_p \in \mathbb{R}^{N\times N\times K_p}$
for all $k \in [K_p]$ **do**
$\phi = \mathbf{C}^{\circ-1}\mathbf{Y}_{k,p}$
$\nabla\mathbf{W}_{k,p} = \mathbf{W}_{k,p} - \phi\mathbf{Y}_{k,p}^{\top}\mathbf{W}_{k,p}$ ▷ Relative gradient
end for
Return $\sum_{k=1}^{K_p}\|\nabla\mathbf{W}_{k,p}\|_{\mathrm{F}}, \max_k\|\nabla\mathbf{W}_{k,p}\|_{\mathrm{F}}$

Algorithm 4 LocApply

Input: site index p, step size α
Require: $\mathbf{W}_p, \nabla\mathbf{W}_p, \mathbf{W}_{\mathrm{prev},p} \in \mathbb{R}^{N\times N\times K_p}$
$\mathbf{W}_p = \mathbf{W}_{\mathrm{prev},p} + \alpha\nabla\mathbf{W}_p$

Algorithm 5 LocUpdate

Input: site index p, step size α
Require: $\mathbf{W}_p, \nabla\mathbf{W}_p, \mathbf{W}_{\mathrm{prev},p} \in \mathbb{R}^{N\times N\times K_p}$
$\mathbf{W}_{\mathrm{prev},p} = \mathbf{W}_p$
$\mathbf{W}_p = \mathbf{W}_{\mathrm{prev},p} + \alpha\nabla\mathbf{W}_p$

Dimensionality reduction. Typically, the data for a single subject is $\mathbf{x}_k \in \mathbb{R}^{T\times R}$, where $T > N$. Before IVA-L can be applied, dimensionality reduction via group PCA is often required. This can be very computationally expensive on large data sets. Thus, in this work we explore the use of random projections to attain dimensionality reduction. Specifically, let $\mathbf{U}$ be a $T \times N$ random matrix sampled from normal distribution. The objective is to obtain a matrix whose rows are orthonormal and whose entries are sampled from Normal distribution. To this end, let $\mathbf{U} = \mathbf{QR}$ be the QR-Decomposition of $\mathbf{U}$, where $\mathbf{Q}$ is a $T \times N$ matrix with orthonormal columns and R is an upper traingular $N \times N$ matrix. Then $\mathbf{Q} = \mathbf{UR}^{-1}$ is a matrix with orthonormal columns whose entries are sampled from the Normal distribution, effectively sampling uniformly on an $(N-1)$-dimensional hypersphere [21]. Thus, $\mathbf{Q}^T$ is a matrix whose rows are orthonormal, as desired. The same projection $\mathbf{Q}$ is used across all subjects.

Algorithm 6 dIVAAgg

Require: $[P]$, each site containing $\mathbf{X}_p \in \mathbb{R}^{N\times R\times K_p}$ and $\mathbf{W}_p \in \mathbb{R}^{N\times N\times K_p}$, tolerance level $t = 10^{-6}$, maximum iterations J, Armijo condition constant $c = 10^{-10}$, $\alpha = 1$
Locally initialize $\{\mathbf{W}_p : p \in [P]\}$ ▷ e.g., $\mathbf{W}_{p,k} = \mathbf{I}$
$\|\nabla\mathbf{W}_p\|_{\mathrm{F}} = 1$, $\|\nabla\mathbf{W}_p\|_{\mathrm{F,max}} = 1$ for $p \in [P]$
$\mathsf{cost}(0) = \infty$, $\|\nabla\mathbf{W}_{\mathrm{prev}}\|_{\mathrm{F}}^2 = \|\nabla\mathbf{W}\|_{\mathrm{F}}^2$, $j = 1$
while $j < J$, $\max_p \alpha\|\nabla\mathbf{W}_p\|_{\mathrm{F,max}} > t$ **do**
$(\mathbf{C}_p, d_p) = \mathsf{LocCost}(p)$ for $p \in [P]$.
$\mathbf{C} = \sqrt{\sum_{p=1}^{P}\mathbf{C}_p}$
$\mathsf{cost}(j) = \frac{1}{R}\sum_{n,r}C_{n,r} - \sum_{p=1}^{P}d_p$
while $\mathsf{cost}(j) > \mathsf{cost}(j-1) - \alpha c\|\nabla\mathbf{W}\|_{\mathrm{F}}^2$ **do**
$\alpha = \frac{3}{4}\alpha$ ▷ Initiate backtracking
$\mathsf{LocApply}(p, \alpha)$ for $p \in [P]$
$(\mathbf{C}_p, d_p) = \mathsf{LocCost}(p)$ for $p \in [P]$.
$\mathbf{C} = \sqrt{\sum_{p=1}^{P}\mathbf{C}_p}$
$\mathsf{cost}(j) = \frac{1}{R}\sum_{n,r}C_{n,r} - \sum_{p=1}^{P}d_p$
end while
$(\|\nabla\mathbf{W}_p\|_{\mathrm{F}}, \|\nabla\mathbf{W}_p\|_{\mathrm{F,max}}) = \mathsf{LocGrad}(p, \mathbf{C}^{\circ-1})$ for $p \in [P]$
$\|\nabla\mathbf{W}\|_{\mathrm{F}}^2 = \sum_{p=1}^{P}\|\nabla\mathbf{W}_p\|_{\mathrm{F}}^2$
if $j > 1$ **then**
$\alpha = \alpha\frac{\|\nabla\mathbf{W}_{\mathrm{prev}}\|_{\mathrm{F}}^2}{\|\nabla\mathbf{W}\|_{\mathrm{F}}^2}$ ▷ Update step size
end if
$\|\nabla\mathbf{W}_{\mathrm{prev}}\|_{\mathrm{F}}^2 = \|\nabla\mathbf{W}\|_{\mathrm{F}}^2$
$\mathsf{LocUpdate}(p, \alpha)$ for $p \in [P]$
$j = j + 1$
end while

4. RESULTS

To test the efficiency of the dIVA algorithm, we run three experiments on 1024 synthetic fMRI data sets (each representing a subject), and assess their performance via the joint Moreau-Amari intersymbol interference (jISI) index [20], which is a function of the square matrices $\mathbf{H}_k = \hat{\mathbf{W}}_k\mathbf{A}_k$, where $\hat{\mathbf{W}}_k = \mathbf{W}_k\mathbf{Q}^{\top}$.

Synthetic sources. 20 synthetic source maps were generated for each subject, $\mathbf{s}_k$. These maps were generated with the SimTB fMRI Simulation Toolbox [22]. Each source map was set to be a 2D image of dimensions 206×206, each with 32968 pixels representing in-brain voxels. Each source map was identical in shape across subjects. However, subject-specific variability was introduced by random scaling, translating and rotating the spatial features on each subject. Thus, corresponding sources should be dependent (rather than identical) across subjects.

Mixing process. Each of the 20 columns of the mixing matrices $\mathbf{A}_k$ were selected as the timecourses from the work in [23]. They

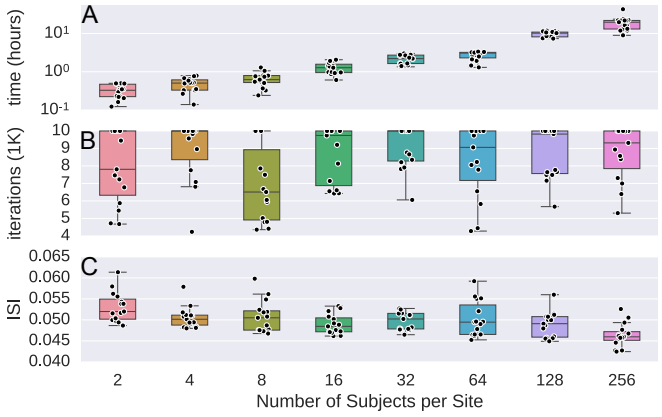

Fig. 1. For 2 sites we increase the number of subjects as each site keeping it equal for both. The running time increases with more data (A) as jISI decreases (C), with iteration number staying put.

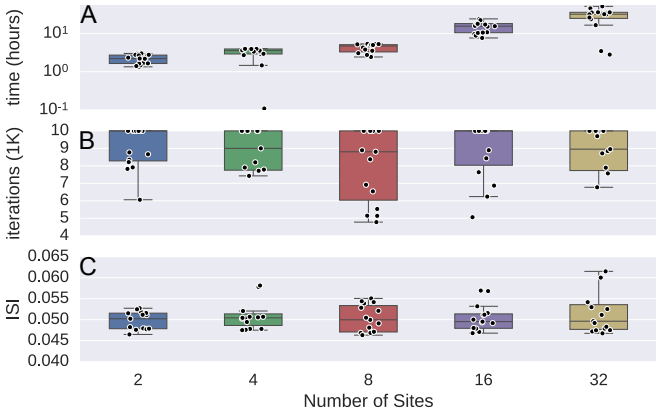

Fig. 2. Fixing the number of subjects per site to 32, we increase the number of sites. jISI and the number of iterations stay virtually unchanged.

were simulated using a generalized autoregressive (AR) conditional heteroscedastic (GARCH) model [24, 25], which has shown use in causal source separation [26] and time-series analyses of neuroscience data [26, 27], especially resting-state fMRI [28, 29]. This was achieved by generating random AR processes (no moving average terms) such that their AR series converged. The AR order was randomly selected between 1 and 10, as well as random coefficients $\{\alpha[l]\}$, such that $\{\alpha[0]\} \in [0.55, 0.8]$ and $\{\alpha[l]\} \in [-0.35, 0.35]$ for $l > 0$. For the error terms $\delta_t = \sigma_t \epsilon_t$, an ARMA model driven by ϵ_t from a generalized Normal distribution with shape parameter 100 (so it was approximately uniform on $[-1, 1]$) and $\sigma_t^2 = 0.1 + 0.1y[t-1]^2 + 0.75\sigma[t-1]^2$ was used. Each of the time courses had $T = 250$ time points, obtained after a burn-in period of 20000 samples, checking that all pair-wise correlations between the 20 time courses was below 0.35. Finally, the mixing process from each subject was applied to the collection of source maps from the same subject as $\mathbf{x}_k = \mathbf{A}_k \mathbf{s}_k$.

Experiments. For the first experiment, the total number of sites remains fixed at $P = 2$, and we test how dIVA preforms as we increase the number of subjects available at each site (Fig. 1). For the second experiment, we test how dIVA performs when the number of subjects available at each site remains fixed at $K_p = 32$ for $p \in [P]$, while the number of sites involved in the analysis increases (Fig. 2). The tests thus provide a measure of how dIVA changes as the number of subjects, the number of sites, and the number of subjects per site varies.

To compare the speed of locally run dIVA to IVA, we preform a time test. Both dIVA and IVA are run for 100 iterations using the same subjects and initial unmixing matrix. After the 100 iterations, we divide the wall clock time it took IVA to run by the time dIVA took to run to get the speed up factor of dIVA over IVA. Repeating this process for 10 different initial starting points and increasing number of subjects gives a rough idea of how running times of locally run dIVA compare to IVA, as seen in Fig. 3.

The experiments show that dIVA does indeed work on decentralized datasets. Its accuracy does not deteriorate with the additional number of subjects and may even slightly improve (Fig. 1(C)). Although the running time steadily grows with more data the situation may change when the sites are really decentralized and perform the computation truly in parallel. We leave discussion of many factors (such as network latency) that can affect the speed for future. The

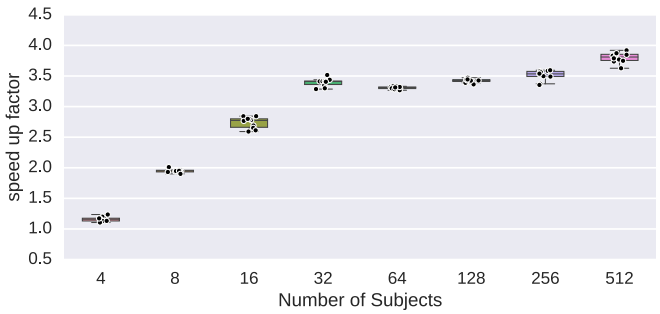

Fig. 3. Comparison of the running times of IVA and dIVA as a function of the total number of subjects. The higher the factor the faster dIVA is compared to IVA.

main goal of dIVA is not to speed up the computation by putting data on remote resources but rather bring the computation to otherwise inaccessible data. However, as Fig. 3 shows, dIVA when used locally may be able to improve efficiency of the current practice of using IVA up to a factor of 4, which for long runs with large datasets may amount to a couple of days versus a week of computation.

5. CONCLUSION

Independent vector analysis (IVA) is an appealing way to leverage multiple data sets for performing blind source separation. Unfortunately, in contexts such as collaborative research consortia, the data sets may be held by different parties who wish to collaborate in an IVA analysis but are unable to share their data, precluding centralized processing. In addition, the running time of IVA is quadratic in the number of data sets: this prevents truly large-scale analysis on high-dimensional data such as MRI. However, as we show in this paper, the objective function optimized by IVA can be split across the sites, allowing the bulk of the computation to be parallelized with the aid of an aggregator that collects summaries from individual sites. We proposed a new algorithm, dIVA, that operates on decentralized data and indeed provides high estimation accuracy. Our experimental results show that the running time of the method is significantly improved over a centralized approach. Future work includes addition of differential privacy to further protect the data at local sites.

6. REFERENCES

[1] Karl Pearson, "On lines and planes of closest fit to systems of points in space," *Philos Mag*, vol. 2, no. 11, pp. 559–572, 1901.

[2] P. Comon and C. Jutten, *Handbook of Blind Source Separation*, Academic Press, Oxford, UK, 1st edition, 2010.

[3] Andrzej Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing*, Wiley, Chichester, UK, revised and corrected edition, 2003.

[4] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis*, Wiley, New York, NY, 1st edition, 2002.

[5] V.D. Calhoun, J. Liu, and T. Adalı, "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data," *NeuroImage*, vol. 45, no. 1, Supplement 1, pp. S163–S172, 2009.

[6] V.D. Calhoun, T. Adalı, G.D. Pearlson, and J.J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Hum Brain Mapp*, vol. 14, no. 3, pp. 140–151, 2001.

[7] M.J. McKeown, S. Makeig, G.G. Brown, T.P. Jung, S.S. Kindermann, A.J. Bell, and T.J. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Hum Brain Map*, vol. 6, no. 3, pp. 160–188, 1998.

[8] Lai Xu, K.M. Groth, G.D. Pearlson, D.J. Schretlen, and V.D. Calhoun, "Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia," *Hum Brain Mapp*, vol. 30, no. 3, pp. 711–724, 2009.

[9] Tom Eichele, Srinivas Rachakonda, Brage Brakedal, Rune Eikeland, and V.D. Calhoun, "EEGIFT: Group independent component analysis for event-related EEG data," *Comput Intell Neurosci*, vol. 2011, pp. 129365, 2011.

[10] R.N. Vigário, "Extraction of ocular artefacts from EEG using independent component analysis," *Electroencephalogr Clin Neurophysiol*, vol. 103, no. 3, pp. 395–404, 1997.

[11] S. Makeig, T.P. Jung, A.J. Bell, D. Ghahremani, and T.J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," *Proc Natl Acad Sci*, vol. 94, no. 20, pp. 10979–10984, 1997.

[12] R. Vigário, J. Sarela, V. Jousmiki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE Trans Biomed Eng*, vol. 47, no. 5, pp. 589–593, May 2000.

[13] A.C. Tang, B.A. Pearlmutter, Michael Zibulevsky, and S.A. Carter, "Blind source separation of multichannel neuromagnetic responses," *Neurocomputing*, vol. 32–33, pp. 1115–1120, 2000.

[14] Srinivas Rachakonda, Rogers F. Silva, Jingyu Liu, and Vince D. Calhoun, "Memory efficient PCA methods for large group ICA," *Front Neurosci*, vol. 10, pp. 17, 2016.

[15] Yuhui Du and Yong Fan, "Group information guided ICA for fMRI data analysis," *NeuroImages*, vol. 69, pp. 157–197, 2013.

[16] J.-H. Lee, T.-W. Lee, F.A. Jolesz, and S.-S. Yoo, "Independent vector analysis (IVA): Multivariate approach for fMRI group study," *NeuroImage*, vol. 40, no. 1, pp. 86–109, 2008.

[17] S. Gopal, R.L. Miller, A. Michael, T. Adalı, M. Cetin, S. Rachakonda, J.R. Bustillo, N. Cahill, S.A. Baum, and V.D. Calhoun, "Spatial variance in resting fMRI networks of schizophrenia patients: An independent vector analysis," *Schizophr Bull*, vol. 42, no. 1, pp. 152–160, 2016.

[18] R.F. Silva, S.M. Plis, J. Sui, M.S. Pattichis, T. Adalı, and V.D. Calhoun, "Blind source separation for unimodal and multimodal brain networks: A unifying framework for subspace modeling," *IEEE J Sel Topics Signal Process*, vol. 10, no. 7, pp. 1–16, Oct 2016.

[19] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Proc ICA 2006*, J. Rosca, D. Erdogmus, J.C. Príncipe, and S. Haykin, Eds., Charleston, SC, 2006, vol. 3889 of *Lecture Notes in Computer Science*, pp. 165–172, Springer.

[20] Matthew Anderson, X.-L. Li, and Tülay Adalı, "Nonorthogonal independent vector analysis using multivariate Gaussian model," in *Proc LVA/ICA 2010*, Vincent Vigneron, Vicente Zarzoso, Eric Moreau, Rémi Gribonval, and Emmanuel Vincent, Eds., vol. 6365 of *Lecture Notes in Computer Science*, pp. 354–361. Springer, St. Malo, France, 2010.

[21] Nir Ailon and Bernard Chazelle, "The fast johnsonlindenstrauss transform and approximate nearest neighbors," *SIAM J Comput*, vol. 39, no. 1, pp. 302–322, 2009.

[22] Erik B. Erhardt, Elena A. Allen, Yonghua Wei, Tom Eichele, and Vince D. Calhoun, "Simtb, a simulation toolbox for fmri data under a model of spatiotemporal separability," *NeuroImage*, vol. 59, no. 4, pp. 4160–4167, 2012.

[23] B.T. Baker, R.F. Silva, V.D. Calhoun, A.D. Sarwate, and S.M. Plis, "Large scale collaboration with autonomy: Decentralized data ICA," in *IEEE MLSP 2015*, Boston, MA, 2015, pp. 1–6.

[24] Robert F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[25] Tim Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J Econom*, vol. 31, pp. 307–327, 1986.

[26] Kun Zhang and Aapo Hyvarinen, "Source separation and higher-order causal analysis of MEG and EEG," in *Proc UAI 2010*, Catalina Island, California, 2010, AUAI Press.

[27] Tohru Ozaki, *Time Series Modeling of Neuroscience Data*, CRC Press, 2012.

[28] Qiang Luo, Tian Ge, Fabian Grabenhorst, Jianfeng Feng, and Edmund T. Rolls, "Attention-dependent modulation of cortical taste circuits revealed by Granger causality with signal-dependent noise," *PLoS Comput Biol*, vol. 9, no. 10, pp. e1003265, 10 2013.

[29] Martin A. Lindquist, Yuting Xu, Mary Beth Nebel, and Brain S. Caffo, "Evaluating dynamic bivariate correlations in resting-state fmri: A comparison study and a new approach," *NeuroImage*, vol. 101, pp. 531–546, 2014.