CONSTRAINED INDEPENDENT VECTOR ANALYSIS WITH REFERENCES: ALGORITHMS AND PERFORMANCE EVALUATION

Trung Vu[1]*, Francisco Laport*[1,2]*, Hanlu Yang*[1]*, and Tülay Adali*[1]

[1] Department of CSEE, University of Maryland, Baltimore County, MD 21250, USA

[2] CITIC Research Center, University of A Coruña, Campus de Elviña, 15071 A Coruña, Spain

{trungvv, flopez2, hyang3, adali}@umbc.edu

ABSTRACT

Independent vector analysis (IVA) is an effective approach to joint blind source separation that fully leverages the statistical dependence across multiple datasets. Nonetheless, its performance might suffer when the number of datasets increases. Constrained IVA is an effective way to bypass computational issues and improve the separation quality by incorporating prior information. This paper introduces a class of approaches to constrained IVA. First, besides the existing augmented Lagrange method, we introduce two novel approaches: the alternating direction method of multipliers and multi-objective optimization. Second, by exploiting the non-orthogonal decoupling of the IVA cost, we derive gradient descent and Newton's method to minimize the objective function. Finally, we demonstrate the effectiveness of algorithms for constrained IVA over unconstrained IVA with simulations.

***Index Terms*—** joint blind source separation, independent vector analysis, constrained optimization

1. INTRODUCTION

Joint blind source separation (JBSS) [1] is the separation of source signals from multiple sets of mixed signals, with fruitful applications in fields such as audio/speech processing [2–4] and medical imaging [5–7]. By leveraging the statistical dependencies across the datasets, JBSS outperforms traditional blind source separation that is applied to individual datasets. Among approaches to JBSS, group ICA [5] is one of the most popular methods in which multiple sets of data are concatenated and reduced (typically through two levels of dimension reduction using principal component analysis) to a group representation in their common subspace. Once the group data is created, independent component analysis (ICA) [8] is applied to extract group independent components (ICs), which can be then used to compute the ICs for each dataset by back-reconstruction or dual regression. The disadvantage of group ICA, however, is that it relies on the assumption of a common subspace among all datasets, and hence, its ability to capture the variability might be limited. Another powerful approach to JBSS is independent vector analysis (IVA) [9], which generalizes ICA for BSS by modeling sources within each dataset to be independent of others and each source to be dependent on a single source within each of the other datasets. Compared with group ICA and other approaches that apply ICA separately to each dataset, IVA has been shown to be a superior solution for capturing variability that exists among the datasets. Notwithstanding, one major drawback of IVA is that its performance degrades when the sample size is not large enough or when the variability among the datasets is very low [10].

To improve the performance of IVA, constrained IVA has been developed as an effective way to incorporate prior knowledge (often about the sources or the mixing matrices [11]) while also addressing the aforementioned limitations. A reliable set of constraints guides IVA optimization algorithms to avoid sub-optimal solutions, increase the quality of source separation, and improve the model match with the observation. In [12], Bhinge *et al.* introduced fixed-threshold inequality constraints to the IVA cost function and utilized a gradient-based Lagrangian framework to solve the constrained optimization problem. Later on, the authors extended this work to an adaptive scheme to select the threshold values and demonstrated the effectiveness of constrained IVA for extracting time-varying spatiotemporal networks in fMRI data analysis [13]. Beyond the gradient-based Lagrangian approach, there has been little research on optimization algorithms for constrained IVA as well as analyzing their performance. Thus, there exists a need for a comprehensive comparison of different optimization methods for constrained IVA to better understand their properties.

In this paper, we introduce two novel approaches to constrained IVA: the alternating direction method of multipliers (ADMM) and multi-objective optimization (MOO). The two approaches are promising alternatives to the augmented Lagrange (AL) method in constrained optimization: ADMM takes advantage of a dual decomposition as well as the augmented Lagrangian methods, while MOO offers a threshold-free formulation for constrained IVA. In addition, by exploit-

This work is supported in part by the grants NIH R01MH118695, NIH R01MH123610, NIH R01AG073949, NSF 2316420, and Xunta de Galicia ED481B 2022/012.

ing the non-orthogonal decoupling of the IVA cost, we derive both gradient descent and Newton's method to minimize the objective function in each formulation of constrained IVA. We highlight that the inverse of the Hessian matrix in Newton's method can be computed in a closed-form expression, making this an attractive solution with the same computational complexity per iteration as gradient descent but a faster convergence speed. Finally, we demonstrate the effectiveness of the proposed algorithms with simulations. Our numerical results provide insights into the pros and cons of each algorithm and form the basis for developing their performance analysis.

2. BACKGROUND

2.1. Independent Vector Analysis (IVA)

Consider K datasets, each formed by T independently and identically distributed (iid) samples of linear mixtures of N independent sources

$$\boldsymbol{x}^{[k]}(t) = \boldsymbol{A}^{[k]}\boldsymbol{s}^{[k]}(t),$$

for $k = 1, \ldots, K$ and $t = 1, \ldots, T$. Here, $\boldsymbol{A}^{[k]} \in \mathbb{R}^{N\times N}$ is an invertible mixing matrix for the kth dataset and $\boldsymbol{s}^{[k]}(t) = [s_1^{[k]}(t), \ldots, s_N^{[k]}(t)]^\top$ is the corresponding source vector. By stacking the nth source component across K datasets, we define the nth source component vector (SCV) as $\boldsymbol{s}_n = [s_n^{[1]}, \ldots, s_n^{[K]}]^\top$. Using this notation, each SCV is a random vector such that $\boldsymbol{s}_m$ and $\boldsymbol{s}_n$ are independent for any $m \neq n$. The goal of IVA is to identify the independent SCVs via the estimation of K demixing matrices of the form $\boldsymbol{W}^{[k]} = [\boldsymbol{w}_1^{[k]}, \ldots, \boldsymbol{w}_N^{[k]}]^\top \in \mathbb{R}^{N\times N}$. Denote $\boldsymbol{y}^{[k]}(t) = \boldsymbol{W}^{[k]}\boldsymbol{x}^{[k]}(t)$, the nth estimated SCV corresponding to the sample index t is given by $\boldsymbol{y}_n(t) = [y_n^{[1]}(t), \ldots, y_n^{[K]}(t)]^\top \in \mathbb{R}^K$.

A common distribution used to model the SCVs is the multivariate Gaussian distribution (MGD) with zero mean and covariance matrix $\boldsymbol{\Sigma}_n \in \mathbb{R}^{K\times K}$. Using the maximum likelihood principle [14], the cost function $\mathcal{J}_{\text{IVA}}$ is given by (1), where $\boldsymbol{W} = \{\boldsymbol{W}^{[k]}\}_{k=1}^K$ and $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_n\}_{n=1}^N$ are the optimization variables, and $\boldsymbol{X}^{[k]} = [\boldsymbol{x}^{[k]}(1), \ldots, \boldsymbol{x}^{[k]}(T)]$. The IVA algorithms for minimizing (1) often involve alternating between updates of $\boldsymbol{\Sigma}$ and $\boldsymbol{W}$. For each $\boldsymbol{\Sigma}_n$, solving the first-order optimality condition yields a closed-form update $\hat{\boldsymbol{\Sigma}}_n = \operatorname{argmin}_{\boldsymbol{\Sigma}_n} \mathcal{J}_{\text{IVA}} = \sum_{k,l=1}^K (\boldsymbol{w}_n^{[k]})^\top \big(\frac{1}{T}\boldsymbol{X}^{[k]}(\boldsymbol{X}^{[l]})^\top\big)\boldsymbol{w}_n^{[l]}\boldsymbol{e}_k\boldsymbol{e}_l^\top$. For each $\boldsymbol{w}_n^{[k]}$, since there is no closed-form expression for the optimal solution, we resort to iterative methods such as gradient descent or Newton's method. In particular, let $\tilde{\boldsymbol{W}}_n^{[k]}$ be the $(N-1)\times N$ matrix obtained by removing the nth row from $\boldsymbol{W}^{[k]}$ and $\boldsymbol{d}_n^{[k]} \in \mathbb{R}^N$ be the unit length vector satisfying $\tilde{\boldsymbol{W}}_n^{[k]}\boldsymbol{d}_n^{[k]} = \boldsymbol{0}_{N-1}$. The gradient $\partial\mathcal{J}_{\text{IVA}}/\partial\boldsymbol{w}_n^{[k]} = \sum_{l=1}^K (\boldsymbol{e}_l^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{e}_k)(\frac{1}{T}\boldsymbol{X}^{[k]}(\boldsymbol{X}^{[l]})^\top)\boldsymbol{w}_n^{[l]} - \boldsymbol{d}_n^{[k]}/(\boldsymbol{d}_n^{[k]})^\top\boldsymbol{w}_n^{[k]}$ or the Hessian $\partial^2\mathcal{J}_{\text{IVA}}/\partial\boldsymbol{w}_n^{[k]}\partial(\boldsymbol{w}_n^{[k]})^\top = (\boldsymbol{e}_k^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{e}_k)\boldsymbol{I}_N + \boldsymbol{d}_n^{[k]}(\boldsymbol{d}_n^{[k]})^\top/\big((\boldsymbol{d}_n^{[k]})^\top\boldsymbol{w}_n^{[k]}\big)^2$ can be used to compute the descent direction that minimizes the IVA cost function.

2.2. Constrained IVA with References

In constrained IVA, we consider a set of reference signals $\{\boldsymbol{r}_n\}_{n=1}^M \subset \mathbb{R}^T (M \leq N)$ that can be used as prior constraints to guide the separation of independent sources. For the kth dataset, denote $\boldsymbol{y}_n^{[k]} = [y_n^{[k]}(1), \ldots, y_n^{[k]}(T)]^\top$. The idea here is to ensure that $\boldsymbol{r}_n$ correlates higher with its corresponding SCV than any other SCVs in the same dataset, i.e.,

$$\epsilon(\boldsymbol{r}_n, \boldsymbol{y}_n^{[k]}) > \epsilon(\boldsymbol{r}_n, \boldsymbol{y}_m^{[k]}) \quad \forall m \neq n, \tag{2}$$

where $n = 1, \ldots, M$, $m = 1, \ldots, N$, and $\epsilon : \mathbb{R}^T \times \mathbb{R}^T \to [0, 1]$ is a similarity measure. A common approach to implementing such constraints is via a pre-defined threshold parameter [11, 12, 15]. By selecting an appropriate value ρ such that $\epsilon(\boldsymbol{r}_n, \boldsymbol{y}_n^{[k]}) \geq \rho > \epsilon(\boldsymbol{r}_n, \boldsymbol{y}_m^{[k]}), \forall m \neq n$, only one independent component is extracted as the closest to the reference signal. Thus, the thresholding-constrained formulation is proposed in [12] as

$$\min_{\boldsymbol{W},\boldsymbol{\Sigma}} \mathcal{J}_{\text{IVA}}(\boldsymbol{W}, \boldsymbol{\Sigma}) \text{ s.t. } \epsilon(\boldsymbol{r}_n, \boldsymbol{y}_n^{[k]}) \geq \rho_n^{[k]} \quad \forall n, k. \tag{3}$$

In the rest of this paper, we use the absolute value of Pearson correlation as the similarity measure, i.e., for any vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^T$, $\epsilon(\boldsymbol{a}, \boldsymbol{b}) = \big|\boldsymbol{a}^\top\boldsymbol{b}\big|/\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|$. Assume the input data is whitened, i.e., $\frac{1}{T}\boldsymbol{X}^{[k]}(\boldsymbol{X}^{[k]})^\top = \boldsymbol{I}_K$, the reference signal is normalized to zero mean and unit variance, and the demixing vector $\boldsymbol{w}_n^{[k]}$ is normalized to have unit norm after each iteration, we have $\epsilon(\boldsymbol{r}_n, \boldsymbol{y}_n^{[k]}) = |\boldsymbol{r}_n^\top(\boldsymbol{X}^{[k]})^\top\boldsymbol{w}_n^{[k]}|/T$.

3. CONSTRAINED OPTIMIZATION METHODS FOR CIVA WITH REFERENCES

3.1. Augmented Lagrangian

The augmented Lagrangian method is the first and most common approach to constrained IVA [11, 16]. Improving upon penalty methods, the AL cost adds another term mimicking a Lagrange multiplier of (3)

$$\mathcal{L}_{\gamma,\boldsymbol{\rho}}(\boldsymbol{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \mathcal{J}_{\text{IVA}}(\boldsymbol{W}, \boldsymbol{\Sigma}) + \mathcal{J}_{\text{ref}}(\boldsymbol{W}, \gamma, \boldsymbol{\rho}), \tag{4}$$

where $\mathcal{J}_{\text{ref}}(\boldsymbol{W}, \gamma, \boldsymbol{\rho}) = \frac{1}{2\gamma}\sum_{n=1}^M\sum_{k=1}^K((\max(0, \mu_n^{[k]} + \gamma(\rho_n^{[k]} - \frac{1}{T}|\boldsymbol{r}_n^\top(\boldsymbol{X}^{[k]})^\top\boldsymbol{w}_n^{[k]}|)))^2 - (\mu_n^{[k]})^2)$. Here $\mu_n^{[k]}$, for

$$\mathcal{J}_{\text{IVA}}(\boldsymbol{W}, \boldsymbol{\Sigma}) = \frac{NK}{2}\log(2\pi) + \frac{1}{2}\sum_{n=1}^N \log|\det(\boldsymbol{\Sigma}_n)| + \frac{1}{2}\sum_{n=1}^N\sum_{k,l=1}^K (\boldsymbol{e}_k^\top\boldsymbol{\Sigma}_n^{-1}\boldsymbol{e}_l)(\boldsymbol{w}_n^{[k]})^\top(\frac{1}{T}\boldsymbol{X}^{[k]}(\boldsymbol{X}^{[l]})^\top)\boldsymbol{w}_n^{[l]} - \sum_{k=1}^K \log|\det(\boldsymbol{W}^{[k]})|, \tag{1}$$

$n = 1, \ldots, M$ and $k = 1, \ldots, K$, are the Lagrange multipliers and $\gamma > 0$ is the scalar penalty parameter. At the ith iteration, we update the parameters to minimize $\mathcal{L}_{\gamma,\boldsymbol{\rho}}$ based on their current values as follows

$$
\begin{aligned}
&(\boldsymbol{W}^{i+1}, \boldsymbol{\Sigma}^{i+1}) = \operatorname{argmin}_{\boldsymbol{W},\boldsymbol{\Sigma}} \mathcal{L}_{\gamma,\boldsymbol{\rho}}\big(\boldsymbol{W}, \boldsymbol{\Sigma}, \boldsymbol{\mu}^i\big), \qquad (5)\\
&(\alpha_n^{[k]})^{i+1} = (\mu_n^{[k]})^i + \gamma\big((\rho_n^{[k]})^i - \frac{1}{T}|\boldsymbol{r}_n^\top (\boldsymbol{X}^{[k]})^\top (\boldsymbol{w}_n^{[k]})^i|\big),\\
&(\mu_n^{[k]})^{i+1} = \max\big(0, (\alpha_n^{[k]})^{i+1}\big).
\end{aligned}
$$

It can be shown [17] that for a bounded γ, the AL method recovers with the solution of (3) while avoiding the ill-conditioning problem of penalty methods.

Similar to minimizing the cost $\mathcal{J}_{\text{IVA}}$ in (1), iterative methods such as gradient descent or Newton's method can be used to solve (5). The gradient descent update of $\boldsymbol{w}_n^{[k]}$, referred as cIVA-G-AL-GD, is given by

$$
(\boldsymbol{w}_n^{[k]})^{i+1} = (\boldsymbol{w}_n^{[k]})^i - \eta\Big(\frac{\partial \mathcal{J}_{\text{IVA}}}{\partial \boldsymbol{w}_n^{[k]}} + \frac{\partial \mathcal{J}_{\text{ref}}}{\partial \boldsymbol{w}_n^{[k]}}\Big)\Big|_{(\boldsymbol{w}_n^{[k]})^i}, \qquad (6)
$$

where $\partial \mathcal{J}_{\text{ref}}/\partial \boldsymbol{w}_n^{[k]} = -\mathbb{I}_{n\le M} \operatorname{sign}(\boldsymbol{r}_n^\top \boldsymbol{y}_n^{[k]})(\frac{1}{T}\boldsymbol{X}^{[k]}\boldsymbol{r}_n) \cdot \max(0, \mu_n^{[k]} + \gamma(\rho_n^{[k]} - \frac{1}{T}|\boldsymbol{r}_n^\top (\boldsymbol{X}^{[k]})^\top \boldsymbol{w}_n^{[k]}|))$. Here, $\mathbb{I}_{(\cdot)}$ is the indicator function and $\operatorname{sign}(\cdot)$ is the sign function. Alternatively, the Newton update for $\boldsymbol{w}_n^{[k]}$, referred as cIVA-G-AL-Newton, is given by

$$
(\boldsymbol{w}_n^{[k]})^{i+1} = (\boldsymbol{w}_n^{[k]})^i - \eta (\boldsymbol{H}_n^{kk})^{-1} \frac{\partial \mathcal{L}_{\gamma,\boldsymbol{\rho}}}{\partial \boldsymbol{w}_n^{[k]}}, \qquad (7)
$$

where $\boldsymbol{H}_n^{kk} = (\boldsymbol{e}_k^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{e}_k)\boldsymbol{I}_N + \boldsymbol{u}_{nk}\boldsymbol{u}_{nk}^\top + \boldsymbol{v}_{nk}\boldsymbol{v}_{nk}^\top$. Here $\boldsymbol{u}_{nk} = \boldsymbol{d}_n^{[k]}/(\boldsymbol{d}_n^{[k]})^\top \boldsymbol{w}_n^{[k]}$ and $\boldsymbol{v}_{nk} = \mathbb{I}_{n\le M}\mathbb{I}_{(\alpha_n^{[k]})^{i+1}>0} \cdot \big(\frac{1}{T}\boldsymbol{X}^{[k]}\boldsymbol{r}_n\big)$. Using the matrix inversion lemma with a rank-2 correction, we obtain $(\boldsymbol{H}_n^{kk})^{-1}$ in (A).

3.2. Alternating Direction Method of Multipliers

Alternating direction method of multipliers can be viewed as an attempt to blend the decomposability of dual decomposition with the superior convergence properties of the AL method for constrained optimization [18]. Applying ADMM to (3), we first introduce a slack variable $\boldsymbol{Z} \in \mathbb{R}^{N\times K}$ with the (n,k)-entry $z_n^{[k]} = \frac{1}{T}\boldsymbol{r}_n^\top (\boldsymbol{X}^{[k]})^\top \boldsymbol{w}_n^{[k]}$. Then, the constraint $\epsilon(\boldsymbol{r}_n, \boldsymbol{y}_n^{[k]}) \ge \rho_n^{[k]}$ hold for all $n = 1, \ldots, M$ and $k = 1, \ldots, K$ if and only if $\boldsymbol{Z} \in \mathcal{C} = \{\mathtt{Z} \in \mathbb{R}^{N\times K} \mid |\mathtt{Z}_{nk}| \ge \rho_n^{[k]} \forall n, k\}$. Thus, we consider a reformulation of (3)

$$
\min_{\boldsymbol{W},\boldsymbol{\Sigma}} \mathcal{J}_{\text{IVA}}(\boldsymbol{W}, \boldsymbol{\Sigma}) + \mathbb{I}_{\mathcal{C}}(\boldsymbol{Z}) \text{ s.t. } \mathcal{A}(\boldsymbol{W}) - \boldsymbol{Z} = \boldsymbol{0}, \qquad (8)
$$

where $\mathcal{A} : \mathbb{R}^{N\times N\times K} \to \mathbb{R}^{N\times K}$ is a linear operator such that $[\mathcal{A}(\boldsymbol{W})]_{nk} = (\boldsymbol{a}_n^{[k]})^\top \boldsymbol{w}_n^{[k]}$ and $\boldsymbol{a}_n^{[k]} = \frac{1}{T}\boldsymbol{r}_n^\top (\boldsymbol{X}^{[k]})^\top$. Similar to AL, the objective function in ADMM is given by $\mathcal{L}_\gamma(\boldsymbol{W}, \boldsymbol{\Sigma}, \boldsymbol{Z}, \boldsymbol{\mu}) = \mathcal{J}_{\text{IVA}}(\boldsymbol{W}, \boldsymbol{\Sigma}) + \mathbb{I}_{\mathcal{C}}(\boldsymbol{Z}) + \boldsymbol{\mu}^\top(\mathcal{A}(\boldsymbol{W}) - \boldsymbol{Z}) + \frac{\gamma}{2}\|\mathcal{A}(\boldsymbol{W}) - \boldsymbol{Z}\|_F^2$, where $\boldsymbol{\mu} \in \mathbb{R}^{N\times K}$ is the Lagrange multiplier and $\gamma > 0$ is the penalty parameter. The ADMM updates are given by

$$
\begin{aligned}
&(\boldsymbol{W}^{i+1}, \boldsymbol{\Sigma}^{i+1}) = \operatorname*{argmin}_{\boldsymbol{W},\boldsymbol{\Sigma}} \mathcal{J}_{\text{IVA}}(\boldsymbol{W}, \boldsymbol{\Sigma}) + \frac{\gamma^i}{2}\left\|\mathcal{A}(\boldsymbol{W}) - \boldsymbol{Z}^i + \boldsymbol{\mu}^i\right\|_F^2,\\
&\boldsymbol{Z}^{i+1} = \operatorname*{argmin}_{\boldsymbol{Z}} \mathbb{I}_{\mathcal{C}}(\boldsymbol{Z}) + \frac{\gamma^i}{2}\left\|\mathcal{A}(\boldsymbol{W}^{i+1}) - \boldsymbol{Z} + \boldsymbol{\mu}^i\right\|_F^2,\\
&\boldsymbol{\mu}^{i+1} = \boldsymbol{\mu}^i + \big(\mathcal{A}(\boldsymbol{W}^{i+1}) - \boldsymbol{Z}^{i+1}\big). \qquad (9)
\end{aligned}
$$

To update $\boldsymbol{w}_n^{[k]}$ (as part of $\boldsymbol{W}$), we use the aforementioned gradient descent or Newton method in combination with the decoupling trick. It is noted that the inverse of the Hessian for Newton method, in this case, is similar to the rank-2 correction in the previous section. Finally, the update of $\boldsymbol{Z}$ admits a closed-form expression $\boldsymbol{Z}^{i+1} = \Pi_{\mathcal{C}}\big(\mathcal{A}(\boldsymbol{W}^{i+1}) + \boldsymbol{\mu}^i\big)$, where $\Pi_{\mathcal{C}}$ is the orthogonal projection onto $\mathcal{C}$ such that $[\Pi_{\mathcal{C}}(\mathtt{Z})]_{nk}$ equals $\mathtt{Z}_{nk}$ if $|\mathtt{Z}_{nk}| \ge \rho_n^{[k]}$ and equals $\rho_n^{[k]}$ otherwise.

3.3. Multi-Objective Optimization

Another approach to constrained IVA is viewing it as a multi-objective optimization problem. In addition to minimizing the IVA cost, one tries to maximize the similarity between the source components and their corresponding references

$$
\min_{\boldsymbol{W},\boldsymbol{\Sigma}} \begin{cases} \mathcal{J}_{\text{IVA}}\big(\boldsymbol{W}, \boldsymbol{\Sigma}\big) & \text{defined in (1)} \\ \mathcal{J}_{\text{ref}}\big(\boldsymbol{W}, \boldsymbol{\Sigma}\big) & \triangleq -\sum_{n=1}^M \epsilon^2(\boldsymbol{r}_n, \boldsymbol{y}_n^{[k]}). \end{cases}
$$

If an optimal solution for one cost function is not optimal for the other, there usually exists a Pareto optimal set [19] that cannot decrease one cost function without increasing the others. The MOO approach has been considered by Du and Fan [20] in the context of constrained ICA. Extending their approach to our cIVA context, we minimize a linear weighted sum of the two cost functions $\mathcal{J}_{\text{MOO}}\big(\boldsymbol{W}, \boldsymbol{\Sigma}\big) = \mathcal{J}_{\text{IVA}}\big(\boldsymbol{W}, \boldsymbol{\Sigma}\big) - \frac{\lambda}{2}\mathcal{J}_{\text{ref}}\big(\boldsymbol{W}, \boldsymbol{\Sigma}\big)$. Compared with (3), MOO eliminates the need for threshold parameters, which is often difficult to pre-determine an appropriate value for each component and each subject. Similar to previous approaches, gradient descent or Newton method in combination with the decoupling trick can be used to minimize $\mathcal{J}_{\text{MOO}}$. The gradient and Hessian of $\mathcal{J}_{\text{ref}}$ are given by $\partial \mathcal{J}_{\text{ref}}/\partial \boldsymbol{w}_n^{[k]} = \mathbb{I}_{n\le M}\lambda(\frac{1}{T}\boldsymbol{X}^{[k]}\boldsymbol{r}_n)(\frac{1}{T}\boldsymbol{X}^{[k]}\boldsymbol{r}_n)^\top \boldsymbol{w}_n^{[k]}$ and $\partial^2 \mathcal{J}_{\text{ref}}/\partial \boldsymbol{w}_n^{[k]} \partial (\boldsymbol{w}_n^{[k]})^\top = \mathbb{I}_{n\le M}\lambda(\frac{1}{T}\boldsymbol{X}^{[k]}\boldsymbol{r}_n)(\frac{1}{T}\boldsymbol{X}^{[k]}\boldsymbol{r}_n)^\top$.

$$
(\boldsymbol{H}_n^{kk})^{-1} = \frac{1}{\boldsymbol{e}_k^\top \boldsymbol{\Sigma}_n^{-1} \boldsymbol{e}_k}\Big(\boldsymbol{I}_N - \frac{(\boldsymbol{v}_{nk}^\top \boldsymbol{v}_{nk} + \boldsymbol{e}_k^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{e}_k)\boldsymbol{u}_{nk}\boldsymbol{u}_{nk}^\top + (\boldsymbol{u}_{nk}^\top \boldsymbol{u}_{nk} + \boldsymbol{e}_k^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{e}_k)\boldsymbol{v}_{nk}\boldsymbol{v}_{nk}^\top - (\boldsymbol{u}_{nk}^\top \boldsymbol{v}_{nk})(\boldsymbol{u}_{nk}\boldsymbol{v}_{nk}^\top + \boldsymbol{v}_{nk}\boldsymbol{u}_{nk}^\top)}{(\boldsymbol{u}_{nk}^\top \boldsymbol{u}_{nk} + \boldsymbol{e}_k^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{e}_k)(\boldsymbol{v}_{nk}^\top \boldsymbol{v}_{nk} + \boldsymbol{e}_k^\top \boldsymbol{\Sigma}_n^{-1}\boldsymbol{e}_k) - (\boldsymbol{u}_{nk}^\top \boldsymbol{v}_{nk})^2}\Big). \qquad (A)
$$

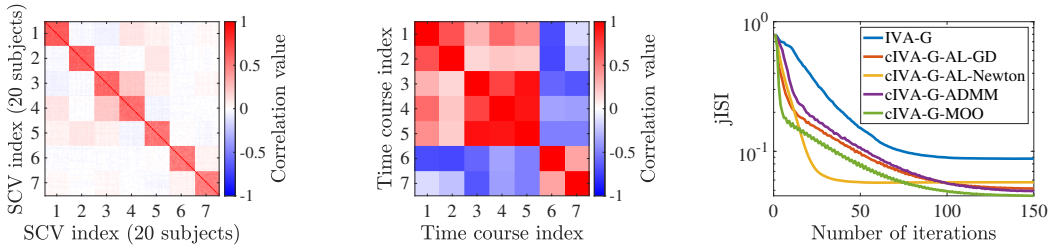

Fig. 1: Visualization of (left) the full SCV covariance matrices (with $N = 7$, $K = 20$, and $V = 5787$), (middle) an example of the correlation among the time courses used to mix the sources, and (right) a demonstration of the performance measure versus the number of iterations using different IVA algorithms (averaged over 50 runs).

4. NUMERICAL RESULTS

4.1. Setting

Extraction of template sources. We use reference signals extracted by NeuroMark [21], which includes 53 fMRI networks and is divided into seven functional domains based on their anatomical and functional properties. In the hybrid simulation experiment with varying numbers of datasets, to reduce the runtime, we only use a subset of $N = 7$ templates (with 5 from subcortical and 2 from auditory) and downsampling the original data from 57878 samples to $V = 5787$ samples. Moreover, we normalize them to zero mean and unit variance and denote the set of templates by $\{\boldsymbol{r}_n\}_{n=1}^N$.

Hybrid data generation. Given N template sources, we generate V observations of SCVs $\{\boldsymbol{S}_n\}_{n=1}^N \subset \mathbb{R}^{K\times V}$ for K datasets as $\boldsymbol{S}_n = a_n \mathbf{1}_K \boldsymbol{r}_n^\top + b_n \boldsymbol{\Phi}_n^\top \boldsymbol{Z} + c_n \boldsymbol{Z}_n + d_n \boldsymbol{E}_n$, where $a_n, b_n, c_n \in [0, 1]$ and $d_n = \sqrt{1 - a_n^2 - b_n^2 - c_n^2}$. Intuitively, a_n controls how close the nth source is to the reference $\boldsymbol{r}_n$, b_n controls the correlation across SCVs, c_n controls the variability within each SCV, and d_n ensures each source has unit variance. Here, the columns of $\boldsymbol{Z} \in \mathbb{R}^{NK\times V}$ and $\boldsymbol{Z}_n \in \mathbb{R}^{N\times V}$ are generated from multivariate Gaussian distributions (MGD) of dimensions NK and N, respectively. Each MGD has zero mean and covariance matrix with diagonal entries equal 1. Also, $\boldsymbol{\Phi}_n \in \mathbb{R}^{NK\times K}$ is a binary matrix that selects the nth row block in $\boldsymbol{Z}$. Finally, $\boldsymbol{E}_N$ is a white noise. Figure 1-(left) depicts the SCV covariance matrices of the simulated fMRI-like data. Next, for each of the K datasets, we generate an $N \times N$ matrix $\boldsymbol{A}^{[k]}$ that represents the time courses for different brain components. Figure 1-(middle) shows the block structure of the time-course covariance matrix, which reflects the two functional domains of the templates. Finally, the mixture data for each dataset is created by adding small white noise to the product $\boldsymbol{A}^{[k]}\boldsymbol{S}^{[k]}$.

4.2. Results

Five algorithms are used in our simulation, namely, unconstrained IVA (IVA-G), constrained IVA with AL using gradient descent (cIVA-G-AL-GD), AL using Newton method (cIVA-G-AL-Newton), ADMM using gradient descent (cIVA-G-ADMM), and MOO using gradient descent (cIVA-G-MOO). Figure 1-(right) illustrates the convergence behavior of the jISI of the five algorithms through iterations (averaged over 50 runs). It is noted that cIVA-G-AL-Newton exhibits the fastest convergence while IVA-G converges in the upmost iterations. There is a negligible difference among other cIVA-G algorithms. In the following, to evaluate the performance of these algorithms, we use the joint inter-symbol-interference (jISI) introduced in [9], with the value closer to 0 indicating better source separation performance.

Varying number of subjects. Figure 2-(a) and (b) demonstrates the performance of the aforementioned algorithms with $K \in \{10, 20, 30, 40, 50\}$, for fixed $N = M = 7$. Overall, cIVA-G-MOO achieves the best performance in terms of jISI while IVA-G performs the worst. When the number of subjects K is small, cIVA-G-AL-GD and cIVA-G-AL-Newton outperform cIVA-G-ADMM. However, when K is sufficiently large, cIVA-G-ADMM becomes more accurate and approaches the performance of cIVA-G-MOO. Noticeably, the performance of IVA-G and cIVA-G-ADMM improve as the number of subjects increases, whereas other cIVA algorithms are less sensitive to K. It is also worthwhile mentioning that cIVA-G-AL-Newton yields the fastest runtime[1], which can be explained by the convergence shown in Fig. 1-(right) and the fact that the iteration complexity of cIVA-G-AL-Newton is the same as other algorithms.

Varying number of references. Figure 2-(c) and (d) demonstrates the performance of the aforementioned algorithms with $M \in \{1, 2, 3, 4, 5, 6, 7\}$, with fixed $N = 7$ and $K = 20$. It can be seen that as more references are used, cIVA algorithms effectively exploit the prior information about the sources and achieve better performance. On the other hand, the performance of IVA-G is independent of M. Similar to the previous simulation, cIVA-G-MOO obtains the best per-

[1]The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). See hpcf.umbc.edu for more information on HPCF and the projects using its resources.

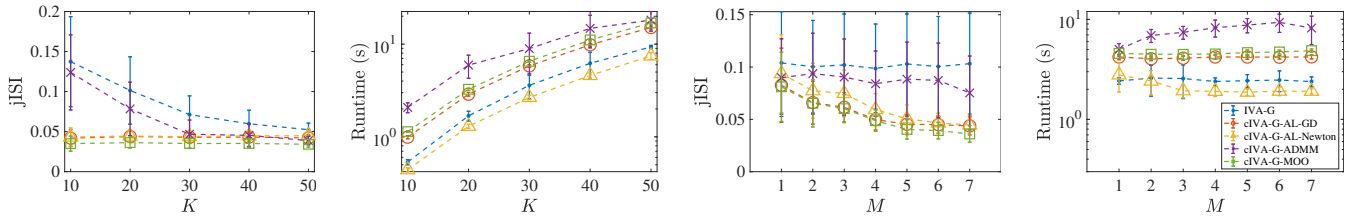

Fig. 2: Comparison of various algorithms using hybrid fMRI-like data with $N = 7$, and $V = 5787$. Plots of (a) jISI and (b) runtime (in seconds) as K varying while fixing $M = 7$; and (c) jISI and (d) runtime as M varying while fixing $K = 20$. The error bars represent one standard deviation calculated over 50 runs.

formance in terms of jISI, while IVA-G performs the worst. When fewer references are used, we observe that cIVA-G-AL-Newton performs slightly worse than cIVA-G-AL-GD, yet still better than cIVA-G-MOO and IVA-G. We also highlight the computational efficiency of cIVA-G-AL-Newton due to its fast convergence.

5. CONCLUSION

In this paper, we studied different optimization approaches for constrained independent vector analysis, namely, augmented Lagrangian, alternating direction method of multipliers, and multi-objective optimization. The multivariate Gaussian implementation along with the use of an effective constraint framework yields a desirable balance between performance and computational complexity. Simulation results show that multi-objective optimization outperforms other algorithms in terms of separation performance while augmented Lagrangian method with the Newton optimization obtains the most efficient runtime. Our study can be used as practical guidelines for the selection of constrained IVA algorithms in real fMRI data analysis.

6. REFERENCES

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

[2] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 1, pp. 70–79, 2006.

[3] Y. Liang, J. Harris, S. M. Naqvi, G. Chen, and J. A. Chambers, "Independent vector analysis with a generalized multivariate Gaussian source prior for frequency domain blind source separation," *Signal Process.*, vol. 105, pp. 175–184, 2014.

[4] L. Li, K. Koishida, and S. Makino, "Online directional speech enhancement using geometrically constrained independent vector analysis," in *Proc. Annu. Conf. Int. Speech Commun. Assoc.*, 2020, pp. 61–65.

[5] V. D. Calhoun, T. Adali, 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.

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

[7] X. Chen, H. Peng, F. Yu, and K. Wang, "Independent vector analysis applied to remove muscle artifacts in EEG data," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1770–1779, 2017.

[8] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 2010.

[9] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672–1683, 2011.

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

[11] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ICA," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 203–212, 2005.

[12] S. Bhinge, Q. Long, Y. Levin-Schwartz, Z. Boukouvalas, V. D. Calhoun, and T. Adali, "Non-orthogonal constrained independent vector analysis: Application to data fusion," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* IEEE, 2017, pp. 2666–2670.

[13] S. Bhinge, R. Mowakeaa, V. D. Calhoun, and T. Adali, "Extraction of time-varying spatiotemporal networks using parameter-tuned constrained IVA," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1715–1725, 2019.

[14] T. Adali, M. Anderson, and G.-S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 18–33, 2014.

[15] P. A. Rodriguez, M. Anderson, X.-L. Li, and T. Adali, "General non-orthogonal constrained ICA," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2778–2786, 2014.

[16] W. Lu and J. Rajapakse, "Constrained independent component analysis," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 13, 2000.

[17] E. G. Birgin and J. M. Martínez, *Practical augmented Lagrangian methods for constrained optimization*. SIAM, 2014.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[19] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscipl. Optim.*, vol. 26, pp. 369–395, 2004.

[20] Y. Du and Y. Fan, "Group information guided ICA for fMRI data analysis," *Neuroimage*, vol. 69, pp. 157–197, 2013.

[21] Y. Du, Z. Fu, J. Sui, S. Gao, Y. Xing, D. Lin, M. Salman, A. Abrol, M. A. Rahaman, J. Chen *et al.*, "NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders," *NeuroImage Clin.*, vol. 28, p. 102375, 2020.