# Improved Dictionary Learning for FMRI Data Analysis Capturing Common and Individual Activation Maps

Shuai Xu† Rui Jin† Seung-Jun Kim†* Tülay Adali† Vince D. Calhoun‡

†Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County, USA
‡Tri-institutional Center for Translational Research in Neuroimaging and Data Science,
Georgia State University, Georgia Institute of Technology, and Emory University, USA

***Abstract*—A dictionary learning (DL)-based method for multisubject functional magnetic resonance imaging (fMRI) data analysis is proposed. The method can incorporate group attributes to find the neural activation maps that are common across groups as well as those that are explanatory of the group differences. The key contribution is to significantly improve the map estimation performance by mitigating undesirable local optima obtained from a supervised DL formulation for fMRI data analysis. We introduce two novel ideas, namely, developing a variant of Fisher's discriminant cost, which is imposed on the non-discriminative features, and optimally permuting the dictionary and the sparse factor such that the solution is not trapped to unwanted local optima. Preliminary tests on synthetic and real data sets verify the effectiveness of the proposed approach.**

## I. Introduction

Leveraging the relationship between neuronal activity and brain hemodynamics, functional magnetic resonance imaging (fMRI) enables non-invasive monitoring of brain functional activity [1]. Methods for analyzing fMRI data can be categorized into hypothesis-driven methods, such as the general linear model (GLM) [2], and data-driven methods, such as the ones based on independent component analysis (ICA) and dictionary learning (DL). The latter methods rely on minimal prior assumptions such as statistical independence and sparsity to extract component neural activations [3], [4]. Recently, deep learning-based methods are also being investigated [5], [6].

As large-scale fMRI studies involving hundreds of subjects or more become prevalent, there is a growing demand for robust data-driven methods for multisubject fMRI data analysis. The DL algorithms were tailored to capture group- and subject-specific features in multisubject fMRI data [7], [8]. However, they often did not directly incorporate informative group attributes such as the diagnosis, gender, and handedness, into the analysis. Therefore, a supervised DL approach was developed using Fisher's discriminative cost to extract the neural activations that can explain the group attributes in [9]. This was achieved by splitting the dictionary (and the corresponding sparse factor) into two subdictionaries that capture the features common across groups and those that are discriminative of the differences. However, due to the DL formulation's nonconvexity and the resulting difficulty in obtaining global optima, the algorithm often yielded features that did not have the designated properties. For example, many of the activation maps obtained in the common subdictionary turned out to be highly discriminative.

To mitigate this issue, we propose a simple yet effective modification of Fisher's cost for the features in the common subdictionary, which encourages traits that are opposite to those imposed by the conventional Fisher's cost. Furthermore, capitalizing on the decomposability of the cost function, permutations are performed on the sparse component maps and the corresponding weight vectors (dictionary atoms) across the common/discriminative subfactors. The optimal permutation is efficiently found by solving a linear assignment problem. Our preliminary experimental results show that the improved DL method can effectively avoid undesirable local optima and obtain neural activation maps with higher quality and explainability.

The rest of the paper is organized as follows. The supervised DL method for fMRI data analysis is briefed in Sec. II. The proposed improvements are delineated in Sec. III, and the algorithm is derived in Sec. IV. The proposed method is evaluated in Sec. V. The conclusion is provided in Sec. VI.

## II. FMRI Data Analysis Using Supervised DL

First, the supervised DL formulation for fMRI data analysis is summarized [9]. Let $\mathbf{X} \in \mathbb{R}^{M\times V}$ be a collection of fMRI data volumes from $M$ subjects. We aim at obtaining $K$ component spatial activation maps $\mathbf{Z} \in \mathbb{R}^{K\times V}$ that are sparse and the corresponding weight matrix (dictionary) $\mathbf{D} \in \mathbb{R}^{M\times K}$ such that $\mathbf{X} \approx \mathbf{DZ}$. To leverage the subject group labels and estimate the component maps that are characteristic of the differences between groups (as well as the maps that are common), the spatial maps $\mathbf{Z}$ is split into $\mathbf{Z} := [\overline{\mathbf{Z}}^\top, \widetilde{\mathbf{Z}}^\top]^\top$ ($^\top$ denotes transposition), where $\widetilde{\mathbf{Z}} \in \mathbb{R}^{\widetilde{K}\times V}$ contains $\widetilde{K}$ discriminative maps and $\overline{\mathbf{Z}} \in \mathbb{R}^{\overline{K}\times V}$ contains $\overline{K}$ common ones, with $\overline{K}+\widetilde{K} = K$. Likewise, $\mathbf{D}$ is partitioned as $\mathbf{D} := [\overline{\mathbf{D}}, \widetilde{\mathbf{D}}]$, where the rows of $\widetilde{\mathbf{D}} \in \mathbb{R}^{M\times\widetilde{K}}$ are the discriminative features for individual subjects and the rows of $\overline{\mathbf{D}} \in \mathbb{R}^{M\times\overline{K}}$ are the common ones.

The group labels can be incorporated through Fisher's discriminant cost [10]. Suppose that $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ collects the features used to classify $N$ data samples into $C$ classes. Denote the set of indices of samples that belong to class $c$ as $\mathcal{N}_c$, which has cardinality $N_c$. Then, the so-called within-class

This work was supported in part by NSF grants 1631838 and 2242412. *Corresponding author. E-mail: sjkim@umbc.edu.

scatter matrix $\mathbf{S}_{\rm w}$ and the between-class scatter matrix $\mathbf{S}_{\rm b}$ are defined as

$$\mathbf{S}_{\rm w}(\mathbf{Y}) := \sum_{c=1}^{C} \sum_{n\in\mathcal{N}_c} (\mathbf{y}_n - \mathbf{m}_c)(\mathbf{y}_n - \mathbf{m}_c)^\top \quad (1)$$

$$\mathbf{S}_{\rm b}(\mathbf{Y}) := \sum_{c=1}^{C} N_c(\mathbf{m}_c - \mathbf{m})(\mathbf{m}_c - \mathbf{m})^\top \quad (2)$$

respectively, where $\mathbf{m}_c := \frac{1}{N_c}\sum_{n\in\mathcal{N}_c}\mathbf{y}_n$ is the feature mean for class $c$ and $\mathbf{m} := \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n$ is the overall feature mean. Fisher's discriminant criteria aims at minimizing the scatter within each class and maximizing the distance between scatters from different classes. One choice of the cost function for this purpose is

$$f(\mathbf{Y}) := \mathrm{tr}\{\mathbf{S}_{\rm w}(\mathbf{Y})\} - \mathrm{tr}\{\mathbf{S}_{\rm b}(\mathbf{Y})\} + \|\mathbf{Y}\|_F^2 \quad (3)$$

where the last term with Frobenius norm ensures the convexity of the cost with respect to (w.r.t.) $\mathbf{Y}$. In fact, upon defining $\mathbf{H}_1 \in \mathbb{R}^{N\times N}$ with the $(i,j)$-entry equal to $1/N_c$ if $i,j\in\mathcal{N}_c$ and 0 otherwise, and $\mathbf{H}_2 \in \mathbb{R}^{N\times N}$ with all entries equal to $1/N$, it can be shown that $f(\mathbf{Y}) = \mathrm{tr}\{\mathbf{Y}\widetilde{\mathbf{H}}\mathbf{Y}^\top\}$ with $\widetilde{\mathbf{H}} := 2\mathbf{I} - 2\mathbf{H}_1 + \mathbf{H}_2$ and all the eigenvalues of $\widetilde{\mathbf{H}}$ are nonnegative [11].

Since the rows of the subdictionary $\widetilde{\mathbf{D}}$ carry the features for discriminating subject groups, the Fisher's cost is employed on $\widetilde{\mathbf{D}}$ only as $f(\widetilde{\mathbf{D}}^\top)$. Thus, the supervised DL problem for fMRI data analysis is formulated as

$$\min_{\mathbf{D}\in\mathcal{D},\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Z}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \frac{\lambda_2}{2} f(\widetilde{\mathbf{D}}^\top) \quad (4)$$

where $\mathbf{D}\in\mathcal{D} := \{[\mathbf{d}_1,\ldots,\mathbf{d}_K] : \|\mathbf{d}_k\|_2 \leqslant 1, k=1,\ldots,K\}$ is imposed due to the scaling ambiguity of bi-factorization, and $\lambda_1$ and $\lambda_2$ are nonnegative regularization parameters. An efficient block coordinate descent (BCD) algorithm can be derived based on the convexity of the problem when one of $\{\mathbf{D},\mathbf{Z}\}$ is fixed, which leads to a local optimum of (4) [9]. However, since the problem is nonconvex in $(\mathbf{D},\mathbf{Z})$ jointly, globally optimal solutions are hard to obtain. Even when solved with multiple initial points, the formulation often leads to undesirable solutions, such as the ones with discriminative features and maps ending up in $\overline{\mathbf{D}}$ and $\overline{\mathbf{Z}}$, respectively, or non-discriminative (common) features and maps in $\widetilde{\mathbf{D}}$ and $\widetilde{\mathbf{Z}}$. When this happens, the explainability and the quality of the estimated maps are diminished.

## III. Proposed Formulation

To mitigate this issue, a few improvements are made to the supervised DL algorithm. First, a variant of Fisher's cost is applied to the common features to explicitly encourage the commonness of features across classes. Secondly, the maps in $\mathbf{Z}$ and the corresponding weights in $\mathbf{D}$ are permuted during optimization to prevent being trapped in undesirable local optima.

### A. Proposed Variant of Fisher's Cost

Instead of singly relying on the Fisher's cost imposed on $\widetilde{\mathbf{D}}$ alone for ensuring both the common and the discriminative maps/features to show up in the correct subfactors, an additional Fisher-like cost is imposed on the common dictionary $\overline{\mathbf{D}}$ as well. The idea is to encourage all features in $\overline{\mathbf{D}}$ to be similar, which can be achieved by imposing a prior in the opposite direction of the conventional Fisher's cost, namely, by maximizing the scatter within each class and minimizing the same across different classes. Thus, a reasonable cost function is

$$g(\mathbf{Y}) := \mathrm{tr}\{\mathbf{S}_{\rm b}(\mathbf{Y})\} - \mathrm{tr}\{\mathbf{S}_{\rm w}(\mathbf{Y})\} + 2\|\mathbf{Y}\|_F^2 \quad (5)$$

where the last term again ensures the convexity of $g$ w.r.t. $\mathbf{Y}$, which is important to derive a convergent BCD algorithm. That is, it can be again shown that $g(\mathbf{Y}) = \mathrm{tr}\{\mathbf{Y}\bar{\mathbf{H}}\mathbf{Y}^\top\}$, with $\bar{\mathbf{H}} := 2\mathbf{H}_1 - \mathbf{H}_2 + \mathbf{I}$, and that all the eigenvalues of $\bar{\mathbf{H}}$ are nonnegative. Thus, with regularization parameters $\{\lambda_i \geqslant 0\}_{i=1}^3$, the proposed DL formulation is given as

$$\min_{\mathbf{D}\in\mathcal{D},\mathbf{Z}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{Z}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \frac{\lambda_2}{2} f(\widetilde{\mathbf{D}}^\top) + \frac{\lambda_3}{2} g(\bar{\mathbf{D}}^\top). \quad (6)$$

### B. Finding Optimal Permutation

To further aid in avoiding undesirable local optima, where the estimated features in $\overline{\mathbf{D}}$ may turn out to be discriminative and those in $\widetilde{\mathbf{D}}$ non-discriminative (and the same way with the corresponding maps in $\overline{\mathbf{Z}}$ and $\widetilde{\mathbf{Z}}$), here it is proposed to proactively permute the columns of $\mathbf{D}$ and the rows of $\mathbf{Z}$ to minimize the objective cost in (6) during the optimization process. Such permutations do not alter the values of the reconstruction term $\frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{Z}\|_F^2$ or the sparsity prior $\lambda_1\|\mathbf{Z}\|_1$ in (6), but only affect the terms involving $f$ and $g$.

It is noted first that one can decompose $f(\mathbf{Y})$ and $g(\mathbf{Y})$ as $f(\mathbf{Y}) = \sum_k f_k(\mathbf{Y})$ and $g(\mathbf{Y}) = \sum_k g_k(\mathbf{Y})$, respectively, where

$$f_k(\mathbf{Y}) := \sum_{c=1}^{C}\Big[\sum_{n\in\mathcal{N}_c}(y_{kn} - m_{c,k})^2 - N_c(m_{c,k} - m_k)^2\Big] + \sum_{n=1}^{N} y_{kn}^2 \quad (7)$$

$$g_k(\mathbf{Y}) := \sum_{c=1}^{C}\Big[N_c(m_{c,k} - m_k)^2 - \sum_{n\in\mathcal{N}_c}(y_{kn} - m_{c,k})^2\Big] + 2\sum_{n=1}^{N} y_{kn}^2 \quad (8)$$

where $y_{kn}$ is the $(k,n)$-entry of $\mathbf{Y}$, and $m_{c,k}$ and $m_k$ are the $k$-th entries of $\mathbf{m}_c$ and $\mathbf{m}$, respectively. Due to the decomposability, given $\mathbf{D}$ and $\mathbf{Z}$, the optimal permutation can be found by solving over permutations $\pi : \{1,\ldots,K\} \to \{1,\ldots,K\}$

$$\min_{\pi} \sum_{k=1}^{K} \omega(k,\pi(k)) \quad (9)$$

where

$$\omega(k,\pi(k)) := \begin{cases} \lambda_3 g_{\pi(k)}(\mathbf{D}^\top) & \text{if } 1 \leqslant \pi(k) \leqslant \overline{K} \\ \lambda_2 f_{\pi(k)}(\mathbf{D}^\top) & \text{if } \overline{K}+1 \leqslant \pi(k) \leqslant K. \end{cases} \quad (10)$$

In the cost function of (9), the costs for the individual atoms are summed, where each atom can be assessed with either the modified or the original Fisher's cost, depending on which subdictionary the permutation puts the atom into. Problem (9) is a linear assignment problem (LAP), which can be solved efficiently [12]. Once the optimal $\pi^*$ is obtained, the $k$-th column of $\mathbf{D}$ is moved to the $\pi^*(k)$-th column of $\mathbf{D}_{\pi^*}$ and the $k$-th row of $\mathbf{Z}$ to the $\pi^*(k)$-th row of $\mathbf{Z}_{\pi^*}$. The permuted

Input: $\mathbf{X}, \mathbf{D}^{(0)}, \mathbf{Z}^{(0)}, \lambda, \mu$
Output: $\mathbf{D}^{(\infty)}$, $\mathbf{Z}^{(\infty)}$

1: Initialize $\mathbf{D}^{(0)}$ and $\mathbf{Z}^{(0)}$ randomly. Set $\ell = 0$.
2: While not converged
3: Set $i = 0$, $t^{(0)} = 1$, $\mathbf{Z}^{(\ell,i)} = \mathbf{Z}^{(\ell)}$, $\mathbf{W}^{(\ell,i)} = \mathbf{Z}^{(\ell)}$, and $L^{(\ell)} = \lambda_{max}\left((\mathbf{D}^{(\ell)})^\top \mathbf{D}^{(\ell)}\right)$
/* *Update* $\mathbf{Z}$ */
4: While not converged
5: $\mathbf{G}^{(\ell,i)} \leftarrow -(\mathbf{D}^{(\ell)})^\top(\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z}^{(\ell,i)})$
6: $\mathbf{Z}^{(\ell,i+1)} \leftarrow \mathcal{S}_{\lambda/L}(\mathbf{W}^{(\ell,i)} - \mathbf{G}^{(\ell,i)}/L^{(\ell)})$
7: $t^{(i+1)} \leftarrow (1 + \sqrt{1 + 4(t^{(i)})^2})/2$
8: $\mathbf{W}^{(\ell,i+1)} \leftarrow \mathbf{Z}^{(\ell,i+1)} + \frac{t^{(i)}-1}{t^{(i+1)}}(\mathbf{Z}^{(\ell,i+1)} - \mathbf{Z}^{(\ell,i)})$
9: $i \leftarrow i + 1$
10: End While
11: Set $s = 0$, $\mathbf{Z}^{(\ell+1)} = \mathbf{Z}^{(\ell,i)}$, and $\mathbf{D}^{(\ell,s)} = \mathbf{D}^{(\ell)}$
/* *Update* $\mathbf{D}$ */
12: Set $\mathbf{A} = \mathbf{Z}^{(\ell+1)}(\mathbf{Z}^{(\ell+1)})^\top$ and $\mathbf{B} = \mathbf{X}(\mathbf{Z}^{(\ell+1)})^\top$
13: While not converged
14: For $k = 1, 2, \ldots, \bar{K}$
15: $\mathbf{u}_k \leftarrow [a_{kk}\mathbf{I} + \lambda_3 \bar{\mathbf{H}}]^{-1}\left(\mathbf{b}_k - \sum_{k'=1,k'\neq k}^{K} a_{kk'}\mathbf{d}_{k'}^{(\ell,s)}\right)$
16: $\mathbf{d}_k^{(\ell,s+1)} \leftarrow \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k$
17: Set the $k$-th column of $\mathbf{D}^{(\ell,s)}$ to $\mathbf{d}_k^{(\ell,s+1)}$
18: End For
19: For $k = \bar{K} + 1, \ldots, K$
20: $\mathbf{u}_k \leftarrow [a_{kk}\mathbf{I} + \lambda_2 \widetilde{\mathbf{H}}]^{-1}\left(\mathbf{b}_k - \sum_{k'=1,k'\neq k}^{K} a_{kk'}\mathbf{d}_{k'}^{(\ell,s)}\right)$
21: $\mathbf{d}_k^{(\ell,s+1)} \leftarrow \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k$
22: Set the $k$-th column of $\mathbf{D}^{(\ell,s)}$ to $\mathbf{d}_k^{(\ell,s+1)}$
23: End For
24: $s \leftarrow s + 1$
25: End While
26: Set $\mathbf{D}^{(\ell+1)} = \mathbf{D}^{(\ell,s)}$
/* *Find the optimal permutation* */
27: Compute $\pi$* via (9).
28: Permute $\mathbf{D}^{(l+1)}$ and $\mathbf{Z}^{(l+1)}$ according to $\pi$*
29: $\ell \leftarrow \ell + 1$
30: End While
31: $\mathbf{D}^{(\infty)} \leftarrow \mathbf{D}^{(\ell)}$ and $\mathbf{Z}^{(\infty)} \leftarrow \mathbf{Z}^{(\ell)}$

TABLE I: Algorithm for solving (6).

$\mathbf{D}_{\pi*}$ and $\mathbf{Z}_{\pi*}$ yield the cost function value that is no larger than that of $\mathbf{D}$ and $\mathbf{Z}$. The BCD-based DL algorithm is then re-run using the permuted factors as the initial point. This process can be repeated until the cost function no longer decreases.

## IV. Algorithm Derivation

Though (6) is nonconvex jointly w.r.t. the optimization variables, when any two of $\mathbf{Z}, \bar{\mathbf{D}}$, and $\widetilde{\mathbf{D}}$ are fixed, the problem is convex w.r.t. the remaining one. Thus, we can find a locally optimal solution through iterative updates. Specifically, when $\mathbf{D}$ is fixed, the formulation reduces to the least absolute shrinkage and selection operator (LASSO) problem for $\mathbf{Z}$. On the other hand, when $\mathbf{Z}$ and either of $\bar{\mathbf{D}}$, or $\widetilde{\mathbf{D}}$ are fixed, the formulation is a convex quadratic optimization problem. We employ the BCD method, which alternates among $\mathbf{Z}, \bar{\mathbf{D}}$, and $\widetilde{\mathbf{D}}$, and provably converges to a locally optimal solution.

Firstly, when $\mathbf{D}$ is fixed, the update of $\mathbf{Z}$ can be done by

$$\mathbf{Z}^{(\ell+1)} := \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 \qquad (11)$$

where $\ell$ is the iteration index. The fast iterative shrinkage-thresholding algorithm (FISTA) [13] can be applied to solve this problem. The FISTA linearly approximates the differentiable part of the objective, i.e. $g(\mathbf{Z}; \mathbf{D}^{(\ell)}) := \frac{1}{2}\|\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z}\|_F^2$, and iteratively solves a proximal problem. The detail of the solution is given in lines 4-10 in TABLE I, where $-(\mathbf{D}^{(\ell)})^\top(\mathbf{X} - \mathbf{D}^{(\ell)}\mathbf{Z})$ is the gradient of $g$ w.r.t. $\mathbf{D}^{(\ell)}$, and $L^{(\ell)} = \lambda_{\max}\left((\mathbf{D}^{(\ell)})^\top\mathbf{D}^{(\ell)}\right)$ is the largest eigenvalue of $(\mathbf{D}^{(\ell)})^\top\mathbf{D}^{(\ell)}$. The operator $\mathcal{S}_\beta(\mathbf{A})$ performs element-wise soft-thresholding on the entire $\mathbf{A}$, i.e. $[\mathcal{S}_\beta(\mathbf{A})]_{ij} = \text{sign}(a_{ij})\max\{0, |a_{ij}| - \beta\}$

When $\mathbf{Z}$ is fixed at $\mathbf{Z}^{(\ell+1)}$, the update for $\mathbf{D}$ is done via

$$\mathbf{D}^{(\ell+1)} := \arg\min_{\mathbf{D}\in\mathcal{D}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{Z}^{(\ell+1)}\|_F^2 + \frac{\lambda_2}{2}\text{tr}\left\{\widetilde{\mathbf{D}}^\top\widetilde{\mathbf{H}}\widetilde{\mathbf{D}}\right\} + \frac{\lambda_3}{2}\text{tr}\left\{\bar{\mathbf{D}}^\top\bar{\mathbf{H}}\bar{\mathbf{D}}\right\} \qquad (12)$$

The data fidelity term in (12) can be replaced by $\frac{1}{2}\text{tr}\{\mathbf{A}\mathbf{D}^\top\mathbf{D}\} - \text{tr}\{\mathbf{B}\mathbf{D}^\top\}$ with $\mathbf{A} := \mathbf{Z}^{(\ell+1)}(\mathbf{Z}^{(\ell+1)})^\top$ and $\mathbf{B} := \mathbf{X}(\mathbf{Z}^{(\ell+1)})^\top$. Thus, (12) can be rewritten as

$$\mathbf{D}^{(\ell+1)} = \arg\min_{\mathbf{D}\in\mathcal{D}} \frac{1}{2}\text{tr}\{\mathbf{A}\mathbf{D}^\top\mathbf{D}\} - \text{tr}\{\mathbf{B}\mathbf{D}^\top\} + \frac{\lambda_2}{2}\text{tr}\{\widetilde{\mathbf{H}}\widetilde{\mathbf{D}}\widetilde{\mathbf{D}}^\top\} + \frac{\lambda_3}{2}\text{tr}\left\{\bar{\mathbf{D}}^\top\bar{\mathbf{H}}\bar{\mathbf{D}}\right\} \qquad (13)$$

We can again apply the BCD method to solve (13) by regarding each atom of $\mathbf{D}$ as a block variable, which can be updated sequentially [14]. Specifically, at the iteration $s+1$, to update the $k$-th atom, fix other atoms to the values at the iteration $s$, i.e. $\mathbf{d}_{k'}^{(\ell)} = \mathbf{d}_{k'}^{(\ell,s)}$ for $k' = 1, \ldots, k-1, k+1, \ldots, K$. Since $\mathbf{D}$ consists of two sub-dictionaries $\bar{\mathbf{D}}$ and $\widetilde{\mathbf{D}}$, the update of atoms can be done differently depending on whether they belong to $\bar{\mathbf{D}}$ or $\widetilde{\mathbf{D}}$. The common dictionary atoms $\{\mathbf{d}_k\}$ for $k = 1, 2, \ldots, \bar{K}$ are updated via

$$\mathbf{u}_k = (a_{kk}\mathbf{I} + \lambda_3\bar{\mathbf{H}})^{-1}\left[\mathbf{b}_k - \sum_{k'\neq k} a_{kk'}\mathbf{d}_{k'}^{(\ell,s)}\right] \qquad (14)$$

$$\mathbf{d}_k^{(\ell,s+1)} = \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k. \qquad (15)$$

Likewise, the discriminative dictionary atoms $\{\mathbf{d}_k\}$ for $k = \bar{K}+1, \ldots, K$ are updated by

$$\mathbf{u}_k = (a_{kk}\mathbf{I} + \lambda_2\widetilde{\mathbf{H}})^{-1}\left[\mathbf{b}_k - \sum_{k'\neq k} a_{kk'}\mathbf{d}_{k'}^{(\ell,s)}\right] \qquad (16)$$

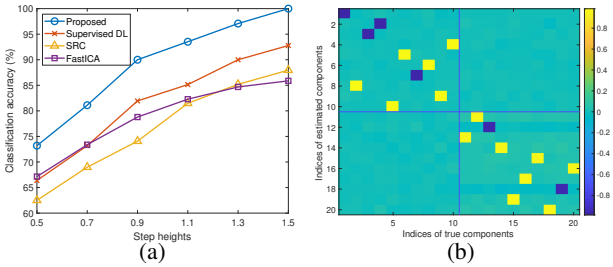
$$\mathbf{d}_k^{(\ell,s+1)} = \frac{1}{\max\{\|\mathbf{u}_k\|_2, 1\}}\mathbf{u}_k. \qquad (17)$$

The overall updates of the dictionary is given in lines 13-23 in TABLE I. After $\mathbf{Z}^{(\ell+1)}$ and $\mathbf{D}^{(\ell+1)}$ are updated, the optimal permutation $\pi$* will be computed via (9). Then, the columns of $\mathbf{D}^{(\ell+1)}$ and the rows of $\mathbf{Z}^{(\ell+1)}$ are rearranged via $\pi$*. Finally, the converged $\mathbf{D}^{(\infty)}$ and $\mathbf{Z}^{(\infty)}$ are output as the solution to (6).

## V. Numerical Tests

### A. Test with Synthetic Data

The proposed method was first tested on a synthetic data set with $\widetilde{K} = \overline{K} = 10$ and $V = 10,000$. The sparse factor was generated from a Bernoulli ($p = 0.5$)-Gaussian distribution. The dictionary for $M = 271$ subjects was generated from




(a) (b)

Fig. 1: Performance on the synthetic data set. (a) Comparison of classification accuracies. (b) Correlation coefficients between true and estimated maps.

| Step | 0.5 | 0.7 | 0.9 | 1.1 | 1.3 | 1.5 |
|---|---|---|---|---|---|---|
| Proposed | 90% | 100% | 100% | 100% | 100% | 100% |
| Supervised DL | 50% | 60% | 70% | 70% | 70% | 80% |

TABLE II: Percentage of correctly determined map types.

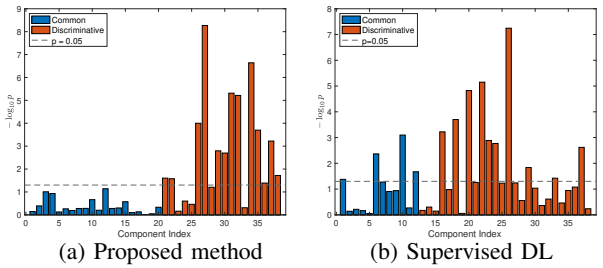
$\mathcal{N}(0,1)$. A step of height $h$ was added to the entries of 121 subjects in the columns associated with discriminative maps to simulate two groups. The synthetic data set was then obtained as the product of the dictionary and the sparse factor plus Gaussian noise. We performed a grid search for $\{\lambda_i\}_{i=1}^3$ by splitting the data into training and validation sets and choosing the parameters with the best classification accuracy on the validation set. Then, (6) was solved with 75 different initializations and the most stable run was found [9].

Fig. 1(a) shows the prediction accuracy of the group labels from the estimated $\widetilde{\mathbf{D}}$ as a function of the step heights, for the proposed and the benchmark methods. For comparison, the results from the supervised DL in (4), fast ICA [15], and sparse representation classification (SRC) [16] methods are also depicted. The kernel support vector machines (SVMs) were trained for fast ICA, and the $K$-nearest neighbor (KNN) classifiers were used for the supervised DL and our method, which yielded the best performances in the respective cases. It can be seen from the figure that our proposed method outperforms the other methods.

Fig. 1(b) depicts the correlation coefficients between the true and the estimated component maps for the data set with the step height of 1.5. It can be observed that all the component maps are faithfully estimated although there are sign and permutation ambiguities. Notably, it can be seen that all the true common maps are estimated as common ($k = 1, \ldots, \overline{K}$), and all the true discriminative maps are estimated as discriminative ($k = \overline{K}+1, \ldots, K$). Table II lists the percentage of the maps with correctly identified types. It can be seen that except for the smallest step height of 0.5, the map types are all correctly determined by the proposed method, while the supervised DL method produces type mismatches.

### *B. Test with Real Data*

We tested the proposed algorithm on the real fMRI data from the MIND Clinical Imaging Consortium (MCIC). The data set involved 150 healthy control subjects and 121 subjects with schizophrenia performing the auditory oddball (AOD) task [9]. After parameter tuning via cross-validation, our method yielded a classification accuracy of 70.1% (with $\overline{K}$ = 20, $\widetilde{K}$ = 18, $\lambda_1 = 0.0027$, $\lambda_2 = 0.175$, and $\lambda_3 = 0.34$), higher than 68% of the supervised DL algorithm (with $\overline{K}$ = 12 and $\widetilde{K}$ = 26) and 66% of the ICA-based method [17]. It should be noted that only the discriminative features are used for classification and our method beats supervised DL with fewer discriminative features.




(a) Proposed method (b) Supervised DL

Fig. 2: Comparison of $p$-values.

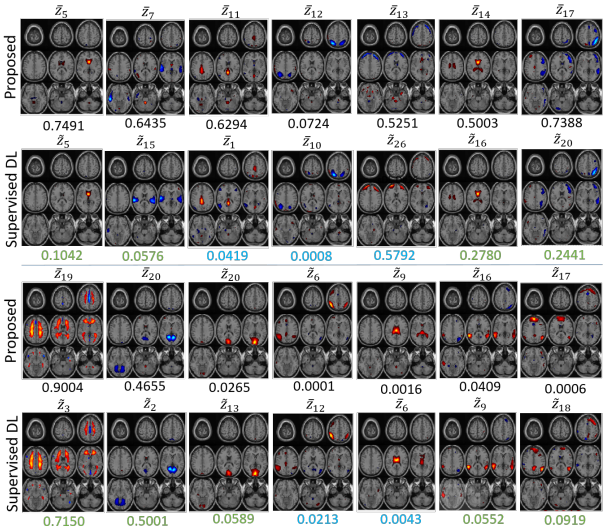



Fig. 3: Visualization of selected maps and their $p$-values from two-sample $t$-tests. The maps from the proposed method (in the 1st and the 3rd rows) are paired with the highly correlated maps (2nd and 4th) from supervised DL. The warm colors in the maps indicate higher activation in the controls and the cold colors in the patients. They are samples of maps correctly identified of their types by the proposed method but not by supervised DL. Specifically, the $p$-values in blue are significant ($p$-values less than 0.05) although the maps are estimated as non-discriminative in $\overline{\mathbf{Z}}$. Likewise, the $p$-values in green are larger than 0.05 although they are estimated as discriminative in $\widetilde{\mathbf{Z}}$.

Fig. 2 shows the $p$-values of the two-sample $t$-tests performed using columns of $\mathbf{D}$. It is observed that the $p$-values of the common maps (in blue) are all larger than 0.05 ($-\log_{10} p \leqslant 1.3$) for the proposed method, while this is not the case for the supervised DL, even though the common map order of the former ($\overline{K}$ = 20) is larger than that of the latter ($\overline{K}$ = 12). It can also be seen that the fraction of the discriminative maps with the $p$-values larger than 0.05

is much smaller for the proposed method compared to the supervised DL. The geometric mean (and the geometric median $\exp(\text{med}\{\log p_i\})$) of the $p$-values of the common and discriminative maps from the proposed method are 0.43 (0.53) and 1.5e-3 (7.5e-3), respectively, while the same for the supervised DL are 0.091 (0.11) and 0.017 (0.07), respectively. Thus the proposed method can obtain the maps of the desired types more effectively.

Next, a subset of the spatial activation maps estimated from the proposed algorithm is shown in Fig. 3. For comparison, they are paired with the maps from the supervised DL algorithm that are highly correlated. The map pairs in Fig. 3 are examples of the supervised DL algorithm producing maps whose $p$-values do not match the designated map types, while the proposed method could estimate maps with correct map types. For example, map $\widetilde{\mathbf{z}}_9$ from the proposed algorithm (or $\overline{\mathbf{z}}_6$ from supervised DL) shows a part of the thalamus region where some recent studies found connections to schizophrenia [18], [19]. Also, it can be seen that the supervised DL algorithm estimated it as a common map, but the $p$-value is smaller than 0.05. On the other hand, our method estimated it as a discriminative map, and the $p$-value is indeed much smaller than 0.05. Another example is the map $\widetilde{\mathbf{z}}_{17}$ from the proposed method (or $\overline{\mathbf{z}}_{18}$ from the supervised DL), which is the default mode network (DMN) region, known to be associated with schizophrenia [20]. Although both methods estimate the map as discriminative, only the map from our algorithm is found to be discriminative according to the $p$-value.[1]

## VI. Conculsion

A DL-based fMRI data analysis algorithm has been proposed, which could estimate the component neural activation maps that are common across subject groups as well as those that are discriminative of group differences. An existing supervised DL algorithm for fMRI data analysis often yielded maps that do not fall into the correct map category due to the local optima present in the nonconvex formulation. A modified version of Fisher's discriminant cost was employed to mitigate this issue on the common map weights. Furthermore, the component maps were permuted during the optimization exploiting the decomposability of the cost function to prevent the algorithm from being stuck in undesirable local optima. Numerical tests with synthetic and real data sets showed that the proposed method yields significantly improved performance compared to existing methods in terms of classification accuracy, correctness of map types, and interpretability of estimated maps.

## References


[1] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proc. Natl. Acad. Sci.*, vol. 87, no. 24, pp. 9868–9872, Dec. 1990.


[1] We noticed that map $\overline{\mathbf{z}}_{13}$ from our method has higher activation in the patient group than the control group, whereas the opposite happens in map $\widetilde{\mathbf{z}}_{26}$ from supervised DL. It turns out that the magnitude of the $t$-statistic associated with these maps is very small. Thus, the sign correction for this case is not reliable.


[2] S. M. Smith, "Overview of fMRI analysis," *British J. Radiology*, vol. 77, no. suppl_2, pp. S167–S175, 2004.

[3] V. D. Calhoun and T. Adali, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Rev. Biomed. Eng.*, vol. 5, pp. 60–73, Aug. 2012.

[4] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Trans. Med. Imaging*, vol. 30, no. 5, pp. 1076–1089, May 2011.

[5] X. Li, N. C. Dvornek, J. Zhuang, P. Ventola, and J. S. Duncan, "Brain biomarker interpretation in ASD using deep learning and fMRI," in *Proc. Int. Conf. Med. Imag. Comput. Computer-Assisted Intervention*, Granada, Spain, Sep. 2018, pp. 206–214.

[6] R. Jin and S.-J. Kim, "FMRI data analysis perserving map variability via unsupervised object-centric learning," in *Proc. Ann. Int. Conf. of the IEEE Eng. Med. Biol. Soc.*, Orlando, FL, Jul. 2024.

[7] A. Iqbal and A.-K. Seghouane, "A dictionary learning algorithm for multi-subject fMRI analysis based on a hybrid concatenation scheme," *Dig. Signal Process.*, vol. 83, pp. 249–260, 2018.

[8] M. U. Khalid, "Sparse group bases for multisubject fMRI data," *IEEE Access*, vol. 10, pp. 83379–83397, 2022.

[9] R. Jin, K. Dontaraju, S.-J. Kim, M. A. B. S. Akhonda, and T. Adali, "Dictionary learning-based fMRI data analysis for capturing common and individual neural activation maps," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1265–1279, May 2020.

[10] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Computer Vis.*, Barcelona, Spain, Nov. 2011, pp. 543–550.

[11] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE Trans. Neural Net. Learn. Syst.*, vol. 27, no. 11, pp. 2160–2173, 2015.

[12] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, Mar. 1995.

[13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learning Res.*, vol. 11, no. Jan, pp. 19–60, 2010.

[15] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, pp. 411–30, May - Jun. 2000.

[16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, Feb. 2009.

[17] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia," *IEEE Trans. Med. Imaging*, vol. 36, no. 7, pp. 1385–1395, 2017.

[18] V. Mancini, D. Zöller, M. Schneider, M. Schaer, and S. Eliez, "Abnormal development and dysconnectivity of distinct thalamic nuclei in patients with 22q11.2 deletion syndrome experiencing auditory hallucinations," *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 5, no. 9, pp. 875–890, 2020.

[19] J. Hua, N. Blair, A. Paez, et al., "Altered functional connectivity between sub-regions in the thalamus and cortex in schizophrenia patients measured by resting state BOLD fMRI at 7T," *Schizophrenia Res.*, vol. 206, pp. 370–377, 2019.

[20] M. L. Hu, X. F. Zong, J. Mann, et al., "A review of the functional and anatomical default mode network in schizophrenia," *Neurosci. Bulletin*, vol. 33, pp. 73–84, 2017.