

SCALABLE FUSED LASSO SVM FOR CONNECTOME-BASED DISEASE PREDICTION

Takanori Watanabe¹, Clayton D. Scott¹, Daniel Kessler², Michael Angstadt², Chandra S. Sripada²

¹Dept. of EECS, ²Dept. of Psychiatry — University of Michigan, Ann Arbor, MI, 48109

ABSTRACT

There is substantial interest in developing machine-based methods that reliably distinguish patients from healthy controls using high dimensional correlation maps known as *functional connectomes* (FC's) generated from resting state fMRI. To address the dimensionality of FC's, the current body of work relies on feature selection techniques that are blind to the spatial structure of the data. In this paper, we propose to use the fused Lasso regularized support vector machine to explicitly account for the 6-D structure of the FC (defined by pairs of points in 3-D brain space). In order to solve the resulting nonsmooth and large-scale optimization problem, we introduce a novel and scalable algorithm based on the alternating direction method. Experiments on real resting state scans show that our approach can recover results that are more neuroscientifically informative than previous methods.

1. INTRODUCTION

There is substantial interest in establishing neuroimaging-based biomarkers that reliably distinguish individuals with psychiatric disorders from healthy individuals [1]. *Functional connectomes* (FC's) generated from resting state fMRI has emerged as a mainstream approach, offering robust ability to characterize the network architecture of the brain [2, 3]. FC's are typically generated by parcellating the brain into hundreds of distinct regions, and computing cross-correlation matrices [4]. However, even with a relatively coarse parcellation with several hundred regions of interest (ROI), the resulting FC contains nearly a hundred thousand connections or more, presenting critical statistical and computational challenges.

In the high dimensional setup, sparsity is a natural assumption that arises in many applications [5, 6]. Indeed, most existing methods address the dimensionality of FC's by applying some form of *a priori* feature selection (*e.g.*, *t*-test) before invoking some “off-the-shelf” classifier (*e.g.*, nearest-neighbor, support vector machine (SVM), LDA) [3]. Sparsity promoting regularizers such as Lasso [7] and Elastic-net [8] may also be considered. However, all these methods above have a major shortcoming: outside of sparsity, the structure of the data is not taken into account.

Recently, there has been strong interest in the machine learning community in designing a convex regularizer that

promotes *structured sparsity* [9, 10]. Indeed, spatially informed regularizers have been applied successfully for *decoding* in task-based fMRI, where the goal is to localize in 3-D space the brain regions that become active under an external stimulus [11, 12]. FC's exhibit rich spatial structure, as each connection comes from a pair of localized regions in 3-D space, giving each connection a localization in 6-D space (referred to as “connectome space” hereafter). However, no framework currently deployed exploits this spatial structure.

Based on these considerations, the main contribution of this paper is two-fold: (1) to account for the 6-D spatial structure of FC's, we propose to use the fused Lasso regularized SVM (FL-SVM) [13], and (2) we introduce a novel scalable algorithm based on the alternating direction method [14] for solving the nonsmooth, large-scale optimization problem that results from FL-SVM. To the best of our knowledge, this is the first application of structured sparse methods in the context of disease prediction using FC's. Experiments on real resting state scans demonstrate that our method can identify predictive features that are spatially contiguous in the connectome space, offering an additional layer of interpretability that could provide new insights about various disease processes.

2. METHODS

fMRI data consist of a time series of three dimensional volumes imaging the brain, where each 3-D volume encompasses around 10,000~100,000 voxels. The univariate time series at each voxel represents a blood oxygen level dependent (BOLD) signal, an indirect measure of neuronal activities in the brain. Traditional experiments in the early years of fMRI research involved *task-based studies*, but after it was discovered that the brain is functionally connected at rest, *resting state fMRI* became a dominant tool for studying the network architecture of the brain. As such, we used the time series from resting state fMRI to generate FC's, which are correlation maps that describe brain connectivity.

More precisely, resting state FC's were produced as follows. First, 347 spherical nodes are placed throughout the entire brain over a regularly-spaced grid with a spacing of $18 \times 18 \times 18$ mm; each of these nodes represent an ROI with a radius of 7.5 mm, which encompasses 30 voxels (the voxel size is $3 \times 3 \times 3$ mm). Next, for each of these nodes, a single representative time series is assigned by spatially averaging the BOLD signals falling within the ROI. Then, a

T.W. and C.D.S.'s work was supported by NIH grant P01CA087634, and C.S.S.'s work was supported by NIH grant K23-AA-020297.

cross-correlation matrix is generated by computing Pearson's correlation coefficient between these representative time series. Finally, a vector \mathbf{x} of length $\binom{347}{2} = 60,031$ is obtained by extracting the lower-triangular part of the cross-correlation matrix; this vector is the FC that serves as the feature vector for disease prediction.

2.1. Fused Lasso Support Vector Machine (FL-SVM)

Our goal is to learn a linear decision function $\text{sign}(\langle \mathbf{x}, \mathbf{w} \rangle)$ given a set of training input/output pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^p$ represents the FC's, and $y_i \in \{\pm 1\}$ indicates the diagnostic status of subject i . The SVM aims to learn the weight vector $\mathbf{w} \in \mathbb{R}^p$ by minimizing the following problem:

$$\arg \min_{\mathbf{w} \in \mathbb{R}^p} \sum_{i=1}^n \ell(y_i \langle \mathbf{w}, \mathbf{x}_i \rangle) + \mathcal{R}(\mathbf{w}), \quad (1)$$

where $\ell(y_i \langle \mathbf{w}, \mathbf{x}_i \rangle) = (1 - y_i \langle \mathbf{w}, \mathbf{x}_i \rangle)_+$ is the hinge loss and $\mathcal{R} : \mathbb{R}^p \rightarrow \mathbb{R}_+$ is the regularizer. For compactness, we introduce the notation $\mathbf{Y} := \text{diag}\{y_1, \dots, y_n\}$ and design matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ created from stacking the feature vectors $\{\mathbf{x}_i\}_{i=1}^n$ as rows. This allows us to express the loss term in (1) succinctly by defining a functional $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}_+$ which aggregates the total loss $\mathcal{L}(\mathbf{Y}\mathbf{X}\mathbf{w}) := \sum_{i=1}^n \ell(y_i \langle \mathbf{w}, \mathbf{x}_i \rangle)$.

In our application, we seek a model that is not only accurate but also interpretable. The standard SVM [15] uses the ℓ_2 -norm regularizer $\mathcal{R}(\mathbf{w}) = \lambda \|\mathbf{w}\|_2^2$, which is problematic for interpretation since it yields a non-sparse and dense solution. Sparsity promoting regularizers such as the Lasso $\mathcal{R}(\mathbf{w}) = \lambda \|\mathbf{w}\|_1$ and Elastic-net $\mathcal{R}(\mathbf{w}) = \lambda \|\mathbf{w}\|_1 + \frac{\gamma}{2} \|\mathbf{w}\|_2^2$ can be used for automatic feature selection [7, 8], but these approaches do not account for the spatial structure of the FC's. To address this issue, we employ the fused Lasso [13].

Fused Lasso was originally designed to encode correlations among successive variables in 1-D, but can be extended to other situations where there is a natural ordering among the feature coordinates. This is the case with our 6-D FC's due to the grid pattern in the nodes, and the FL-SVM problem reads:

$$\arg \min_{\mathbf{w} \in \mathbb{R}^p} \mathcal{L}(\mathbf{Y}\mathbf{X}\mathbf{w}) + \lambda \|\mathbf{w}\|_1 + \gamma \|\mathbf{C}\mathbf{w}\|_1, \quad (2)$$

where $\mathbf{C} \in \mathbb{R}^{e \times p}$ denotes the 6-D *differencing matrix*, and $\|\mathbf{C}\mathbf{w}\|_1 = \sum_{j=1}^p \sum_{k \in \mathcal{N}_j} |w_j - w_k|$ is a spatial penalty that accounts for the 6-D structure in the connectome by penalizing deviations among the *nearest-neighbor* edge set \mathcal{N}_j . Note that e indicates the total number of unordered pairs of adjacent coordinates. To gain a better understanding of \mathcal{N}_j , let us denote (x, y, z) and (x', y', z') the pair of 3-D points in the brain that define the 6-D connectome coordinate j . Then, the first-order neighborhood set of j can be written precisely as

$$\mathcal{N}_j := \left\{ \begin{array}{l} (x \pm 1, y, z, x', y', z'), (x, y, z, x' \pm 1, y', z'), \\ (x, y \pm 1, z, x', y', z'), (x, y, z, x', y' \pm 1, z'), \\ (x, y, z \pm 1, x', y', z'), (x, y, z, x', y', z' \pm 1) \end{array} \right\}.$$

2.2. Optimization: ADMM algorithm

Solving the optimization problem (2) is challenging since the problem size p is large and the three terms in the cost function

are each non-differentiable. To address these challenges, we propose a novel and scalable algorithm based on the alternating direction method of multipliers (ADMM) [14].

ADMM solves problems having the separable structure

$$\min_{\bar{\mathbf{x}}, \bar{\mathbf{y}}} \bar{\mathbf{f}}(\bar{\mathbf{x}}) + \bar{\mathbf{g}}(\bar{\mathbf{y}}) \quad \text{subject to } \bar{\mathbf{A}}\bar{\mathbf{x}} + \bar{\mathbf{B}}\bar{\mathbf{y}} = \mathbf{0}, \quad (3)$$

where $\bar{\mathbf{x}} \in \mathbb{R}^{\bar{p}}$ and $\bar{\mathbf{y}} \in \mathbb{R}^{\bar{q}}$ are unknown primal variables, $\bar{\mathbf{f}} : \mathbb{R}^{\bar{p}} \rightarrow \mathbb{R} \cup \{+\infty\}$ and $\bar{\mathbf{g}} : \mathbb{R}^{\bar{q}} \rightarrow \mathbb{R} \cup \{+\infty\}$ are closed convex functions, and $\bar{\mathbf{A}} \in \mathbb{R}^{c \times \bar{p}}$ and $\bar{\mathbf{B}} \in \mathbb{R}^{c \times \bar{q}}$ are matrices representing c linear constraints. ADMM exploits the separable structure in (3) by applying the following updates:

$$\bar{\mathbf{x}}^{(t+1)} \leftarrow \arg \min_{\bar{\mathbf{x}}} \bar{\mathbf{f}}(\bar{\mathbf{x}}) + \frac{\rho}{2} \|\bar{\mathbf{A}}\bar{\mathbf{x}} + \bar{\mathbf{B}}\bar{\mathbf{y}}^{(t)} + \mathbf{u}^{(t)}\|_2^2 \quad (4)$$

$$\bar{\mathbf{y}}^{(t+1)} \leftarrow \arg \min_{\bar{\mathbf{y}}} \bar{\mathbf{g}}(\bar{\mathbf{y}}) + \frac{\rho}{2} \|\bar{\mathbf{A}}\bar{\mathbf{x}}^{(t+1)} + \bar{\mathbf{B}}\bar{\mathbf{y}} + \mathbf{u}^{(t)}\|_2^2 \quad (5)$$

$$\mathbf{u}^{(t+1)} \leftarrow \mathbf{u}^{(t)} + (\bar{\mathbf{A}}\bar{\mathbf{x}}^{(t+1)} + \bar{\mathbf{B}}\bar{\mathbf{y}}^{(t+1)}), \quad (6)$$

where t denotes the iteration count, $\mathbf{u} \in \mathbb{R}^c$ is the (scaled) dual variable, and $\rho > 0$ is a user defined parameter. The convergence of the ADMM algorithm has been established in Theorem 1 of [16], which states that if matrices $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are full column-rank and the problem (3) is solvable (*i.e.*, it has an optimal objective value), the iterations (4)-(6) converges to the optimal solution. While the parameter $\rho > 0$ does not affect the convergence property of ADMM, it can impact its convergence speed. We set $\rho = 1$ in our implementations.

In order to convert the FL-SVM problem (2) into an equivalent constrained problem with the ADMM structure (3), we apply *variable splitting* [17]. Before we introduce our variable splitting scheme, we note that as it stands, our algorithm will require us to invert a matrix involving the Laplacian matrix $\mathbf{C}^T \mathbf{C} \in \mathbb{R}^{p \times p}$, which is prohibitively large. Although this matrix is sparse, it has a distorted structure due to the irregularities in the coordinates of \mathbf{x} (see Fig. 1). These irregularities arise from two reasons: (a) the nodes defining \mathbf{x} are only on the brain, not the entire rectangular field of view, and (b) \mathbf{x} lacks a complete 6-D representation since it only contains the lower-triangular part of the cross-correlation matrix.

To address this issue, we introduce an *augmentation matrix* $\mathbf{A} \in \mathbb{R}^{\bar{p} \times p}$, whose rows are either the zero vector or an element from the trivial basis $\{e_j\}_{j=1}^p$, and has the property $\mathbf{A}^T \mathbf{A} = \mathbf{I}_p$. Furthermore, we define the *augmented weight vector* $\tilde{\mathbf{w}} := \mathbf{A}\mathbf{w}$, where \mathbf{A} rectifies the irregularities in the coordinates of \mathbf{w} (and \mathbf{x}) by padding extra zero entries. This results in a new differencing matrix $\tilde{\mathbf{C}} \in \mathbb{R}^{\bar{e} \times \bar{p}}$ for $\tilde{\mathbf{w}} \in \mathbb{R}^{\bar{p}}$, whose Laplacian matrix $\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} \in \mathbb{R}^{\bar{p} \times \bar{p}}$ has a special structure known as *block-circulant with circulant-blocks* (BCCB) (see Fig. 1), which has important computational advantages.

Finally, by introducing a diagonal masking matrix $\mathbf{B} \in \{0, 1\}^{\bar{p} \times \bar{p}}$, we have $\|\mathbf{B}\tilde{\mathbf{C}}\tilde{\mathbf{w}}\|_1 = \|\mathbf{C}\mathbf{w}\|_1$. Note that this masking strategy was adopted from the recent work of Allison *et al.* [18], and has the effect of removing artifacts introduced from data augmentation when computing the spatial penalty norm $\|\cdot\|_1$. This allows us to write out the FL-SVM

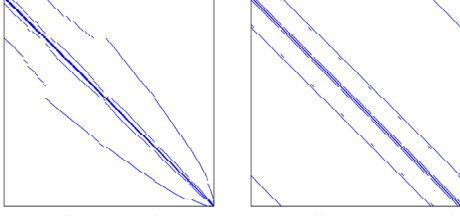


Fig. 1: Laplacian matrices corresponding to the original data $C^T C$ (left) and the augmented data $\tilde{C}^T \tilde{C}$ (right).

problem (2) in the following equivalent form:

$$\arg \min_{\mathbf{w} \in \mathbb{R}^p} \mathcal{L}(\mathbf{YX}\mathbf{w}) + \lambda \|\mathbf{w}\|_1 + \gamma \|\mathbf{B}\tilde{\mathbf{C}}\mathbf{A}\mathbf{w}\|_1. \quad (7)$$

We propose the following variable splitting scheme to convert problem (7) into an equivalent constrained problem:

$$\min_{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4} \mathcal{L}(\mathbf{v}_1) + \lambda \|\mathbf{v}_2\|_1 + \gamma \|\mathbf{B}\mathbf{v}_3\|_1 \quad (8)$$

$$\text{s.t. } \mathbf{YX}\mathbf{w} = \mathbf{v}_1, \quad \mathbf{w} = \mathbf{v}_2, \quad \tilde{\mathbf{C}}\mathbf{v}_4 = \mathbf{v}_3, \quad \mathbf{A}\mathbf{w} = \mathbf{v}_4.$$

Here, the correspondence with the ADMM formulation (3) is:

$$\begin{aligned} \bar{\mathbf{f}}(\bar{\mathbf{x}}) &= \gamma \|\mathbf{B}\mathbf{v}_3\|_1, & \bar{\mathbf{x}} &= [\mathbf{w}^T, \mathbf{v}_3^T]^T, \\ \bar{\mathbf{g}}(\bar{\mathbf{y}}) &= \mathcal{L}(\mathbf{v}_1) + \lambda \|\mathbf{v}_2\|_1, & \bar{\mathbf{y}} &= [\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_4^T]^T, \end{aligned} \quad (9)$$

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{YX} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \\ \mathbf{A} & \mathbf{0} \end{bmatrix}, \quad \bar{\mathbf{B}} = \begin{bmatrix} -\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\tilde{\mathbf{C}} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I} \end{bmatrix}.$$

The dual variables corresponding to $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, and \mathbf{v}_4 are written in block form $\mathbf{u} = [\mathbf{u}_1^T, \mathbf{u}_2^T, \mathbf{u}_3^T, \mathbf{u}_4^T]^T$. Note that this formulation satisfies the sufficient conditions for the ADMM iterations (4)-(6) to converge, as functions $\bar{\mathbf{f}}$ and $\bar{\mathbf{g}}$ are convex, and matrices $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are full column-rank.

With the variable splitting scheme (8), the ADMM update for the primal variable $\bar{\mathbf{x}}$ (4) decomposes into subproblems

$$\mathbf{w}^{(t+1)} \leftarrow \arg \min_{\mathbf{w}} \left\{ \left\| \mathbf{YX}\mathbf{w} - (\mathbf{v}_1^{(t)} - \mathbf{u}_1^{(t)}) \right\|_2^2 + \left\| \mathbf{w} - (\mathbf{v}_2^{(t)} - \mathbf{u}_2^{(t)}) \right\|_2^2 + \left\| \mathbf{A}\mathbf{w} - (\mathbf{v}_4^{(t)} - \mathbf{u}_4^{(t)}) \right\|_2^2 \right\}$$

$$\mathbf{v}_3^{(t+1)} \leftarrow \arg \min_{\mathbf{v}_3} \gamma \|\mathbf{B}\mathbf{v}_3\|_1 + \frac{\rho}{2} \left\| \mathbf{v}_3 - (\tilde{\mathbf{C}}\mathbf{v}_4^{(t)} - \mathbf{u}_3^{(t)}) \right\|_2^2$$

whereas the updates for primal variable $\bar{\mathbf{y}}$ (5) are

$$\mathbf{v}_1^{(t+1)} \leftarrow \arg \min_{\mathbf{v}_1} \mathcal{L}(\mathbf{v}_1) + \frac{\rho}{2} \left\| \mathbf{v}_1 - (\mathbf{YX}\mathbf{w}^{(t+1)} + \mathbf{u}_1^{(t)}) \right\|_2^2$$

$$\mathbf{v}_2^{(t+1)} \leftarrow \arg \min_{\mathbf{v}_2} \lambda \|\mathbf{v}_2\|_1 + \frac{\rho}{2} \left\| \mathbf{v}_2 - (\mathbf{w}^{(t+1)} + \mathbf{u}_2^{(t)}) \right\|_2^2$$

$$\mathbf{v}_4^{(t+1)} \leftarrow \arg \min_{\mathbf{v}_4} \left\{ \left\| \tilde{\mathbf{C}}\mathbf{v}_4 - (\mathbf{v}_3^{(t+1)} + \mathbf{u}_3^{(t)}) \right\|_2^2 + \left\| \mathbf{v}_4 - (\mathbf{A}\mathbf{w}^{(t+1)} + \mathbf{u}_4^{(t)}) \right\|_2^2 \right\}.$$

The closed form solutions for these are provided in Algorithm 1, which outlines the complete ADMM algorithm. We now demonstrate these updates can be carried out efficiently.

The inverse \mathbf{H}^{-1} for the \mathbf{w} update (line 4 Alg. 1) can be converted into an $(n \times n)$ inversion problem using the *matrix inversion Lemma*, where n is on the order of a hundred in our

Algorithm 1 ADMM algorithm for FL-SVM

- 1: Initialize variables $\mathbf{w}, \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}, \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$
- 2: Set $t = 0$, and precompute $\mathbf{H}^{-1} = (\mathbf{X}^T \mathbf{X} + 2\mathbf{I}_p)^{-1}$
- 3: **repeat**
- 4: $\mathbf{w}^{(t+1)} \leftarrow \mathbf{H}^{-1} \left\{ \mathbf{X}^T \mathbf{Y}^T [\mathbf{v}_1^{(t)} - \mathbf{u}_1^{(t)}] + [\mathbf{v}_2^{(t)} - \mathbf{u}_2^{(t)}] + \mathbf{A}^T [\mathbf{v}_4^{(t)} - \mathbf{u}_4^{(t)}] \right\}$
- 5: $[\mathbf{v}_3^{(t+1)}]_i \leftarrow \begin{cases} \text{Soft}_{\gamma/\rho} \left([\tilde{\mathbf{C}}(\mathbf{v}_4^{(t)} - \mathbf{u}_3^{(t)})]_i \right) & \text{if } B_{i,i} = 1 \\ [\tilde{\mathbf{C}}(\mathbf{v}_4^{(t)} - \mathbf{u}_3^{(t)})]_i & \text{if } B_{i,i} = 0 \end{cases}$
- 6: $[\mathbf{v}_1^{(t+1)}]_j \leftarrow \text{Prox}_{\ell/\rho} \left([\mathbf{YX}\mathbf{w}^{(t+1)} + \mathbf{u}_1^{(t)}]_j \right)$
- 7: $[\mathbf{v}_2^{(t+1)}]_k \leftarrow \text{Soft}_{\lambda/\rho} \left([\mathbf{w}^{(t+1)} + \mathbf{u}_2^{(t)}]_k \right)$
- 8: $\mathbf{v}_4^{(t+1)} \leftarrow \left(\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} + \mathbf{I}_{\bar{p}} \right)^{-1} \left\{ \tilde{\mathbf{C}}^T [\mathbf{v}_3^{(t+1)} + \mathbf{u}_3^{(t)}] + \mathbf{A}\mathbf{w}^{(t+1)} + \mathbf{u}_4^{(t)} \right\} \triangleright \text{solve via FFT (10)}$
- 9: $\mathbf{u}_1^{(t+1)} \leftarrow \mathbf{u}_1^{(t)} + \mathbf{YX}\mathbf{w}^{(t+1)} - \mathbf{v}_1^{(t+1)}$
- 10: $\mathbf{u}_2^{(t+1)} \leftarrow \mathbf{u}_2^{(t)} + \mathbf{w}^{(t+1)} - \mathbf{v}_2^{(t+1)}$
- 11: $\mathbf{u}_3^{(t+1)} \leftarrow \mathbf{u}_3^{(t)} + \mathbf{v}_3^{(t+1)} - \tilde{\mathbf{C}}\mathbf{v}_4^{(t+1)}$
- 12: $\mathbf{u}_4^{(t+1)} \leftarrow \mathbf{u}_4^{(t)} + \mathbf{A}\mathbf{w}^{(t+1)} - \mathbf{v}_4^{(t+1)}$
- 13: $t \leftarrow t + 1$
- 14: **until** stopping criterion is met

Note: $[\cdot]_i, [\cdot]_j, [\cdot]_k$ in line 5-7 indicate vector elements

application. The updates for $\mathbf{v}_1, \mathbf{v}_2$, and \mathbf{v}_3 (line 5-7 Alg. 1) are all separable across their coordinates and admit element-wise closed form solutions. Note $\text{Soft}_{\tau}(t) := t(1 - \tau/|t|)_+$ is the soft-threshold operator, and $\text{Prox}_{\tau\ell}(t)$ is the proximal operator [19] corresponding to the hinge-loss:

$$\text{Prox}_{\tau\ell}(t) := \begin{cases} t & \text{if } t > 1 \\ 1 & \text{if } 1 - \tau \leq t \leq 1 \\ t + \tau & \text{if } t < 1 - \tau \end{cases}$$

Finally, the update for \mathbf{v}_4 (line 8 Alg. 1) can be computed efficiently using the fast Fourier Transform (FFT). To suppress notations, let us define $\mathbf{Q} := \tilde{\mathbf{C}}^T \tilde{\mathbf{C}} + \mathbf{I}_{\bar{p}}$ and

$$\mathbf{b} := \tilde{\mathbf{C}}^T [\mathbf{v}_3^{(t)} + \mathbf{u}_3^{(t)}] + \mathbf{A}\mathbf{w}^{(t+1)} + \mathbf{u}_4^{(t)}.$$

As stated earlier, the Laplacian matrix $\tilde{\mathbf{C}}^T \tilde{\mathbf{C}}$ is a BCCB matrix, and consequently, the matrix \mathbf{Q} is BCCB as well. It is well known that a BCCB matrix can be diagonalized as $\mathbf{Q} = \mathbf{U}^H \mathbf{\Lambda} \mathbf{U}$, where \mathbf{U} is the (6-D) DFT matrix and $\mathbf{\Lambda}$ is a diagonal matrix containing the (6-D) DFT coefficients of the first column of \mathbf{Q} . As a result, the update for \mathbf{v}_4 can be carried out efficiently using the (6-D) FFT

$$\mathbf{Q}^{-1} \mathbf{b} = (\mathbf{U}^H \mathbf{\Lambda}^{-1} \mathbf{U}) \mathbf{b} = \text{ifft} \left(\text{fft}(\mathbf{b}) \odot \phi \right), \quad (10)$$

where fft and ifft denote the (6-D) FFT and inverse-FFT operation, ϕ is a vector containing the diagonal entries of $\mathbf{\Lambda}$, and \odot indicates elementwise division.

We note that the ADMM algorithm was also used to solve FL-SVM in [20] under a different variable splitting setup. However, their application focuses on 1-D data, where the Laplacian matrix corresponding to their feature vector is tridiagonal with no irregularities present. Furthermore, the vari-

able splitting scheme they propose requires an iterative algorithm to be used for one of the ADMM subproblems. In contrast, the variable splitting scheme and the data augmentation strategy we propose allow the ADMM subproblems to be decoupled in a way that all the updates can be carried out efficiently and non-iteratively in closed form.

3. EXPERIMENTS

In our experiments, we used the Center for Biomedical Research Excellence dataset made available by the Mind Research Network [21]. We analyzed resting state scans from 121 individuals consisting of 67 healthy controls and 54 schizophrenic subjects. For preprocessing details, we refer the readers to the extended version of our paper [22].

We compared the performance between the Elastic-net regularized SVM (EN-SVM) and FL-SVM; EN-SVM was also solved by ADMM, although the variable splitting and the optimization steps vary slightly from FL-SVM. The algorithms were terminated when the relative infeasibility $\frac{\|\bar{A}\bar{x}^{(t)} + \bar{B}\bar{y}^{(t)}\|}{\max\{\|\bar{A}\bar{x}^{(t)}\|, \|\bar{B}\bar{y}^{(t)}\|\}}$ of the ADMM constraint (3) fell below 5×10^{-5} or the algorithm reached 400 iterations, and 10-fold cross-validation (CV) was used to evaluate the generalizability of the classifiers. All variables were initialized at zero.

A common practice for choosing the regularization parameters is to select the choice that gives the highest prediction accuracy. However, since our goal is the discovery and validation of imaging-based biomarkers, we need a model that offers not only good classification accuracy but also interpretability (*i.e.*, sparsity). We found that the classifiers achieved a good balance between accuracy and sparsity when approximately 4% of the features (2,400) were selected out of $p = 60,031$. More specifically, EN-SVM and FL-SVM achieved classification accuracies of 71.1% and 74.4% when the regularization parameters $\{\lambda, \gamma\}$ were set at $\{2^{-9}, 2^{-5}\}$ and $\{2^{-9}, 2^{-12}\}$, and averages of 2180 and 2387 features were used across the CV folds. Hence, we further analyzed the classifiers obtained from these regularization parameters.

During CV, we learn a different weight vector for each partitioning of the dataset. In order to obtain a single representative weight vector, we re-trained the classifier using the entire dataset (121 subjects). For visualization and interpretation, we grouped the indices of these weight vectors according to the network parcellation scheme proposed by Yeo *et al.* in [23] (see Table 1), and reshaped them into 347×347 symmetric matrices with zeroes on the diagonal. Furthermore, we generated trinary representations of these matrices in order to highlight their support structures, where red, blue, and white denotes positive, negative, and zero entries respectively. The resulting matrices are displayed in Fig. 2.

From these figures, one can observe that EN-SVM yields features that are scattered throughout the connectome space, which can be problematic for interpretation. On the other hand, FL-SVM recovered much more systematic sparsity pat-

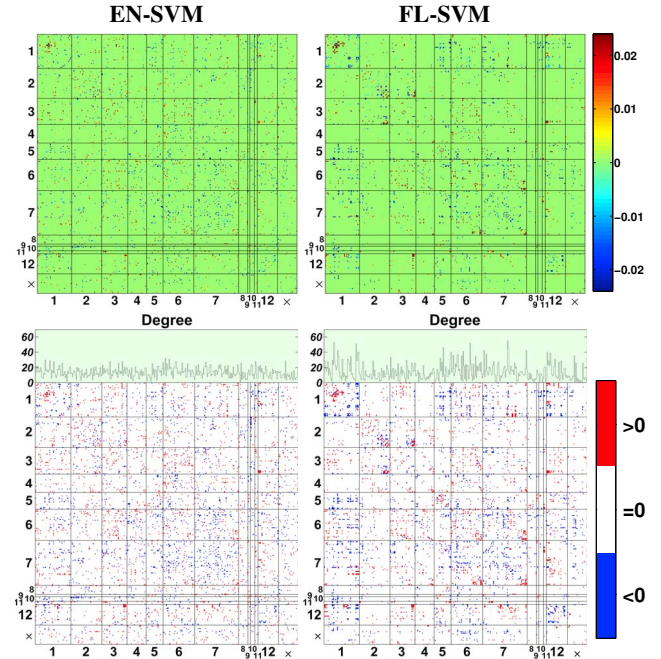


Fig. 2: Weight vectors learned using the full dataset. **Top:** heatmap of the weight vectors. **Bottom:** support structures of the weight vector, along with the degree plot of the nodes.

Network Membership Table (× is “unlabeled”)		
1. Visual	2. Somatomotor	3. Dorsal Attention
4. Ventral Attention	5. Limbic	6. Frontoparietal
7. Default	8. Striatum	9. Amygdala
10. Hippocampus	11. Thalamus	12. Cerebellum

Table 1: Network parcellation of the brain proposed in [23].

terns with multiple contiguous clusters, indicating that the predictive regions are compactly localized in the connectome space (*e.g.*, see the rich connectivity patterns in the intra-visual and intra-cerebellum network). Moreover, FL-SVM recovers multiple highly connected hubs, which is an example of a spatially contiguous cluster in the connectome space. In order to emphasize this point, the bottom row in Fig. 2 also plots the degree of the nodes, *i.e.*, the number of connections a node make with the rest of the network. This degree plot indicates that the frontoparietal network and cerebellum (among other regions) exhibited increased node degree, indicating diffuse connectivity alterations with other networks. Interestingly, these networks are among the most commonly implicated in schizophrenia [24].

4. CONCLUSION

We introduced a classification framework that explicitly accounts for the 6-D spatial structure in the FC via the fused Lasso SVM, which is solved using a novel alternating direction algorithm. We demonstrated that our method recovers sparse and highly interpretable feature patterns while maintaining predictive power, and thus could generate new insights into how psychiatric disorders impact brain networks.

5. REFERENCES

- [1] G. Atluri, K. Padmanabhan, G. Fang, M. Steinbach, J. R. Petrella, K. Lim, A. MacDonald III, N. F. Samatova, P. M. Doraiswamy, and V. Kumar, "Complex biomarker discovery in neuroimaging data: Finding a needle in a haystack," *Neuroimage: Clin.*, vol. 3, no. 0, pp. 123 – 131, 2013.
- [2] M. D. Fox and M. Greicius, "Clinical applications of resting state functional connectivity," *Front. Syst. Neurosci.*, vol. 4, no. 19, 2010.
- [3] F. X. Castellanos, A. D. Martino, R. C. Craddock, A. D. Mehta, and M. P. Milham, "Clinical applications of the functional connectome," *NeuroImage*, vol. 80, no. 0, pp. 527 – 540, 2013.
- [4] G. Varoquaux and R. C. Craddock, "Learning and comparing functional connectomes across subjects," *NeuroImage*, vol. 80, no. 0, pp. 405 – 415, Oct 2013.
- [5] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Trans. Signal Proc. Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [6] J. Fan and J. Lv, "A selective overview of variable selection in high dimensional feature space," *Stat. Sinica*, vol. 20, no. 1, pp. 101–148, 2010.
- [7] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [8] H. Zou and T. Hastie, "Regularization and variable selection via the Elastic Net," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, pp. 301–320, 2005.
- [9] J. Mairal, R. Jenatton, G. Obozinski, and F. Bach, "Convex and network flow optimization for structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 2681–2720, Nov. 2011.
- [10] C. A. Micchelli, J. M. Morales, and M. Pontil, "Regularizers for structured sparsity," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 455–489, 2013.
- [11] V. Michel, A. Gramfort, G. Varoquaux, E. Eger, and B. Thirion, "Total variation regularization for fMRI-based prediction of behavior," *IEEE Trans. Med. Imaging*, vol. 30, no. 7, pp. 1328–1340, 2011.
- [12] L. Grosenick, B. Klingenberg, K. Katovich, B. Knutson, and J. E. Taylor, "Interpretable whole-brain prediction analysis with GraphNet," *NeuroImage*, vol. 72, no. 0, pp. 304 – 321, 2013.
- [13] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused Lasso," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, no. 1, pp. 91–108, 2005.
- [14] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, 2011.
- [15] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [16] J.F.C. Mota, J.M.F. Xavier, P.M.Q. Aguiar, and M. Püschel, "A proof of convergence for the alternating direction method of multipliers applied to polyhedral-constrained functions," *arXiv:1112.2295*, 2011.
- [17] M.V. Afonso, J.M. Bioucas-Dias, and M. A T Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Proc.*, vol. 19, no. 9, pp. 2345–2356, 2010.
- [18] M.J. Allison, S. Ramani, and J.A. Fessler, "Accelerated regularized estimation of MR coil sensitivities using Augmented Lagrangian methods," *IEEE Trans. Med. Imaging*, vol. 32, no. 3, pp. 556–564, 2013.
- [19] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [20] G.B. Ye and X. Xie, "Split bregman method for large scale fused lasso," *Comput. Stat. Data Anal.*, vol. 55, no. 4, pp. 1552–1569, 2011.
- [21] The Mind Research Network (MRN), "The Center for Biomedical Research Excellence (COBRE)," http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html.
- [22] T. Watanabe, D. Kessler, C. Scott, M. Angstadt, and C. Sripada, "Disease Prediction based on Functional Connectomes using a Spatially-Informed Fused Lasso Support Vector Classifier," *arXiv:1310.5415*, 2013.
- [23] B.T.T. Yeo, F.M. Krienen, J. Sepulcre, M.R. Sabuncu, D. Lashkari, M. Hollinshead, J.L. Roffman, J.W. Smoller, L. Zöllei, J.R. Polimeni, B. Fischl, H. Liu, and R.L. Buckner, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *J. Neurophysiol.*, vol. 106, no. 3, pp. 1125–65, 2011.
- [24] Alex Fornito, Andrew Zalesky, Christos Pantelis, and Edward T. Bullmore, "Schizophrenia, neuroimaging and connectomics," *NeuroImage*, vol. 62, no. 4, pp. 2296 – 2314, 2012.