Published in final edited form as:

Neuroinformatics. 2018 October ; 16(3-4): 351–361. doi:10.1007/s12021-018-9382-0.

A Robust Reduced Rank Graph Regression Method for Neuroimaging Genetic Analysis

Xiaofeng Zhu¹, Weihong Zhang², Yong Fan¹, and Alzheimer's Disease Neuroimaging Initiative

¹Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

²Peking Union Medical College Hospital, Beijing 100730, China

Abstract

To characterize associations between genetic and neuroimaging data, a variety of analytic methods have been proposed in neuroimaging genetic studies. These methods have achieved promising performance by taking into account inherent correlation in either the neuroimaging data or the genetic data alone. In this study, we propose a novel robust reduced rank graph regression based method in a linear regression framework by considering correlations inherent in neuroimaging data and genetic data jointly. Particularly, we model the association analysis problem in a reduced rank regression framework with the genetic data as a feature matrix and the neuroimaging data as a response matrix by jointly considering correlations among the neuroimaging data as well as correlations between the genetic data and the neuroimaging data. A new graph representation of genetic data is adopted to exploit their inherent correlations, in addition to robust loss functions for both the regression and the data representation tasks, and a square-root-operator applied to the robust loss functions for achieving adaptive sample weighting. The resulting optimization problem is solved using an iterative optimization method whose convergence has been theoretically proved. Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset have demonstrated that our method could achieve competitive performance in terms of regression performance between brain structural measures and the Single Nucleotide Polymorphisms (SNPs), compared with state-of-the-art alternative methods.

Keywords

Image-genetic analysis; Variable selection; Sparse learning; Graph representation

Correspondence to: Yong Fan.

Alzheimer's Disease Neuroimaging Initiative (ADNI) is a Group/Institutional Author.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Information Sharing Statement

The used dataset in this paper came from the Alzheimer's Disease Neuroimaging Initiative (ADNI, RRID:SCR_003007) database which is freely available at the website www.loni.usc.edu. The source code of the proposed method can be downloaded at <https://sites.google.com/site/seanzhuxf/>.

Introduction

Recent advances in non-invasive neuroimaging and genotyping technologies have promoted the study of neuroimaging genetics, providing opportunities to investigate interplay of the brain structure, function, connectivity, and genetic variations, as well as their relationships with human behaviors or neuropsychiatric disorders (Thompson et al. 2013; Medland et al. 2014). Quantitative measures of the brain's structure, functional activity, as well as structural and functional connectivity can be derived from imaging data as intermediate or endophenotypes (Thompson et al. 2013; Medland et al. 2014; Kong et al. 2015; Zhu et al. 2017b). Particularly, multimodal imaging techniques, such as structural magnetic resonance imaging (sMRI), functional MRI (fMRI), diffusion tensor imaging (DTI), positron emission tomography (PET), magnetoencephalography (MEG), and electroencephalography (EEG) have been widely adopted to quantitatively measure brain morphology, function, and functional and structural connectivity in studies of brain development, aging, and brain disorders (Fan et al. 2011; Fu et al. 2014, 2018; Liu et al. 2014, 2015). A variety of studies have shown that sMRI measures (volume of brain cortical and subcortical regions, cortical thickness, and cortical area), resting state fMRI (rsfMRI) measures (functional connectivity), and DTI measures (white matter integrity and structural connectivity) are genetically influenced (Wang et al. 2012a, b; Medland et al. 2014; Hibar et al. 2015; Hao et al. 2016; Greenlaw et al. 2017; Huang et al. 2017).

In neuroimaging genetic studies, exploring association between a few candidate imaging measures and genetic loci or single nucleotide polymorphisms (SNPs) with univariate statistical tools has been widely adopted, and evolves to candidate imaging measure and genome-wide association analysis, candidate gene and whole brain association analysis, and whole brain and genome-wide (WBGW) association analysis (Ge et al. 2013; Thompson et al. 2013), causing multiple testing problems. Adopting multivariate statistical tools has been an effective means to alleviate the multiple testing problems, especially those joint modeling approaches for exploring multivariate association (Wang et al. 2012a, b; Zhu et al. 2017b).

The rich information of neuroimaging and genetic data facilitates fine-grained analyses of the interplay of brain structure, function, connectivity, and genetic factors (Vounou et al. 2010, 2012; Lu et al. 2017). However, high dimensionality of the neuroimaging and genetic data has been a major challenge, causing computational and statistical problems (Thompson et al. 2013; Medland et al. 2014). Firstly, due to resource limitations, neuroimaging genetic studies often have limited samples and the number of samples is much less than the dimensions of both neuroimaging and genetic data (Zhang et al. 2011, 2012), leading to the so called "curse of dimensionality" problem (a, Peng and Fan 2017b; Zheng et al. 2017; Zhu et al. 2017a, 2018). Secondly, both noise and redundancy of the neuroimaging and genetic data could reduce the rank of the data, which may make it difficult to obtain robust machine learning models from such data (Vounou et al. 2012; Greenlaw et al. 2017; Lu et al. 2017). To alleviate the "curse of dimensionality" problem, data reduction techniques have been widely adopted in imaging-genetics studies, including linear subspace analysis methods and supervoxel methods (Fan et al. 2005, 2007; Zhang et al. 2015). In particular, linear subspace learning methods, such as principle component analysis (PCA), have been used to project voxelwise measures into a small number of components, and the supervoxel methods

parcellate the brain into regions of interest (ROIs) by adopting anatomical atlases or clustering image voxels. Anatomical atlases that define ROIs, such as Automated Anatomical Labeling (AAL) template (Tzourio-Mazoyer et al. 2002), are widely adopted for computing regional imaging measures and for defining network nodes to compute brain connectivity measures in brain network analysis (Fan et al. 2011). More recently, sparse models have been proposed to simultaneously deal with the issues of data noise and “curse of dimensionality” (Vounou et al. 2012; Lu et al. 2017; Zhu et al. 2018). For example, Lu et al. developed a Bayesian longitudinal low-rank regression model to jointly analyze correlations between high-dimensional longitudinal response variables and neuroimaging features (Lu et al. 2017). Du et al. employed sparse canonical correlation analysis to discover bi-multivariate correlations between different genetic markers such as single-nucleotide polymorphisms (SNPs) and neuroimaging quantitative traits (Du et al. 2017).

In this study, we propose a robust reduced rank graph regression (RRRGR) method to characterize correlations among genetic and neuroimaging data jointly. Specifically, we use the genetic data as a feature matrix to regress a response matrix comprising the neuroimaging data by constructing a reduced rank regression model, which jointly considers correlations among the neuroimaging data as well as correlations between the genetic data and the neuroimaging data. Furthermore, a new graph representation of genetic data is adopted to exploit their inherent correlations. Our method is further enhanced by robust loss functions for both the regression and the data representation tasks and a square-root-operator is applied to the robust loss functions for achieving adaptive sample weighting, which alleviates impact of data outliers, facilitates automatic model weighting, and eliminates the need for tuning a trade-off parameter for the robust loss functions. The resulting optimization problem is solved using an iterative optimization method whose convergence has been theoretically proved. The proposed method has been evaluated based on the Alzheimer’s Disease Neuroimaging Initiative (ADNI) structural MRI data and SNP data for exploring their associations. Experimental results showed that the proposed method could identify important SNPs that predict the brain imaging features with higher accuracy than the SNPs selected by alternative state-of-the-art methods under comparison, including Regularized Multivariate Ridge Regression (Hoerl and Kennard 1970), Multi-Task Feature Learning (Argyriou et al. 2007), Group Multi-Task Feature Learning (Wang et al. 2012a, b), and Sparse Reduced-Rank Regression (Vounou et al. 2012).

Methods

Notations

Throughout this paper, we denote matrices as boldface uppercase letters, vectors as boldface lowercase letters, and scalars as normal italic letters, respectively. For a matrix $\mathbf{X} = [x_{ij}]$, its i -th row and j -th column are denoted as $\mathbf{x}^i$ and $\mathbf{x}_j$, respectively. We denote the Frobenius norm, the $\ell_{2,1}$ -norm, and ℓ_1 - the norm of a matrix $\mathbf{X}$ as $\|\mathbf{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$, $\|\mathbf{X}\|_{2,1} = \sum_i \|\mathbf{x}^i\|_2 = \sum_i \sqrt{\sum_j x_{ij}^2}$, and $\|\mathbf{X}\|_1 = \sum_{i,j} |x_{ij}|$, respectively. We further denote the transpose operator, the trace operator, the rank, and the inverse of a matrix $\mathbf{X}$ as $\mathbf{X}^T$, $tr(\mathbf{X})$, $rank(\mathbf{X})$, and $\mathbf{X}^{-1}$, respectively.

A Robust Reduced Rank Graph Regression Method

Since not all SNPs are related to neuroimaging features, sparsity is often adopted via either an ℓ_1 -norm regularization or an $\ell_{2,1}$ -norm regularization to remove redundant SNPs for regressing the neuroimaging features (Zhu et al. 2017b). Neuroimaging features usually correlate with each other to some extent, and the same thing happens to SNPs. Hence, the correlations among the neuroimaging features and the SNPs should be taken into consideration in the neuroimaging genetic analysis. In neuroimaging studies, the anatomic ROI-based neuroimaging measures are prone to different sources of noise, such as scanning device variations, inconsistent image quality, image preprocessing errors (e.g., segmentation and registration inaccuracies), etc. The SNPs do not necessarily correlate equally with the intermediate or endophenotypes. Thus, both the neuroimaging measures and genetic data have different levels of importance/reliability in the neuroimaging genetic analysis.

By denoting $\mathbf{X} \in \mathbb{R}^{n \times p}$ and $\mathbf{Y} \in \mathbb{R}^{n \times q}$, respectively, as the feature matrix of p SNPs and the response matrix of q neuroimaging features, obtained from n unrelated subjects, we propose to characterize correlations among genetic and neuroimaging data by optimizing

$$\min_{\mathbf{S}, \mathbf{U}, \mathbf{V}} \sqrt{\|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{XS}\|_{2,1}} + \lambda_1 \|\mathbf{U}, \mathbf{S}\|_{2,1} + \lambda_2 \|\mathbf{S}\|_1, \text{ subject to } \mathbf{V}^T \mathbf{V} = \mathbf{I} \text{ and } \text{diag}(\mathbf{S}) = 0.$$

(1)

where $\mathbf{U} \in \mathbb{R}^{p \times r}$, $\mathbf{V} \in \mathbb{R}^{q \times r}$, $r = \min(n, p, q)$, $\mathbf{S} \in \mathbb{R}^{p \times p}$, and $\mathbf{I} \in \mathbb{R}^{r \times r}$ is an identity matrix. Different from conventional ridge regression using the least square loss function $\|\mathbf{Y} - \mathbf{XW}\|_F^2$ (where $\mathbf{W} \in \mathbb{R}^{p \times q}$) plus an $\ell_{2,1}$ -norm regularization on $\mathbf{W}$ to optimize the weight coefficient matrix $\mathbf{W}$, the term $\|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1}$ conducts a robust reduced rank regression in consideration of potential correlations among $\mathbf{Y}$ (Zhu et al. 2017b). Specifically, the orthogonality constraint on $\mathbf{V}$ encourages the column vectors of $\mathbf{V}$ uncorrelated, and leads to a formulation $\|\mathbf{Y} - \mathbf{XUV}^T\|_F^2 \Leftrightarrow \|\mathbf{YV} - \mathbf{XU}\|_F^2$, indicating that $\mathbf{YV}$ considers the correlations among the columns of $\mathbf{Y}$ (i.e., the correlations among neuroimaging features) to regress $\mathbf{X}$. The term $\|\mathbf{X} - \mathbf{XS}\|_{2,1}$ is a robust self-representation loss function (Hu et al. 2017), where the value of s_{ij} (s_{ij} is the element of the i -th row and the j -th column of $\mathbf{S}$) measures correlation between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$. The $\ell_{2,1}$ -norm regularization on $\mathbf{S}$ helps alleviate the impact of redundant SNPs, and the ℓ_1 -norm regularization on $\mathbf{S}$ and the constraint $\text{diag}(\mathbf{S}) = 0$ are adopted to sparsely represent each sample by other samples, excluding itself (Hu et al. 2017; Zhu et al. 2017a). In particular, $\|\mathbf{U}, \mathbf{S}\|_{2,1}$ is a group sparsity regularization on $\mathbf{U}$ and $\mathbf{S}$ in their row direction to select the same rows of these two variables for conducting variable selection on the SNPs. Therefore, Eq. (1) conducts variable selection on $\mathbf{X}$, i.e., SNPs, in consideration of potential correlations among the neuroimaging features via the first loss function and the constraints on $\mathbf{U}$ and $\mathbf{V}$, as well as makes every SNP to be sparsely represented by informative SNPs in

consideration of potential correlations among the features via the second loss function and the constraints on $\mathbf{S}$.

We adopt the square root of the $\ell_{2,1}$ -norm loss functions in the first two terms in Eq. (1), instead of the conventional Frobenius loss functions (i.e., least square loss function), because the squared root of a loss function is more robust to outliers (Peng and Fan 2016, 2017a, 2017b). To optimize Eq. (1), we firstly calculate the derivative on the first term of Eq. (1) with respect to its variables (i.e., $\mathbf{U}$ and $\mathbf{V}$) and the second term of Eq. (1) with respect to the variable $\mathbf{S}$ to have:

$$\min_{\mathbf{S}, \mathbf{U}, \mathbf{V}} \omega_1 \|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1} + \omega_2 \|\mathbf{X} - \mathbf{XS}\|_{2,1} + \lambda_1 \|\mathbf{U}, \mathbf{S}\|_{2,1} + \lambda_2 \|\mathbf{S}\|_1 \text{ subject to } \mathbf{V}^T \mathbf{V} = \mathbf{I} \text{ and } \text{diag}(\mathbf{S}) = 0, \quad (2)$$

where ω_1 and ω_2 are defined as:

$$\omega_1 = \frac{1}{2\sqrt{\|\mathbf{Y} - \mathbf{XUV}^T\|_F^2}}, \quad (3)$$

$$\omega_2 = \frac{1}{2\sqrt{\|\mathbf{X} - \mathbf{XS}\|_F^2}}$$

The values of ω_1 and ω_2 can be regarded as the model weights for the reduced rank constraint and the graph representation constraint, respectively, resulting in parameter-free model weightings. More specifically, if $\sqrt{\|\mathbf{Y} - \mathbf{XUV}^T\|_F^2}$ is small (i.e., when $\mathbf{XUV}^T$ is a good estimation of $\mathbf{Y}$), then ω_1 is assigned a large weight, and vice versa. On the other hand, it can be shown that the following equations hold (Zhu et al. 2017a, 2018),

$$\|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1} = \text{tr}\left((\mathbf{Y} - \mathbf{XUV}^T)^T \mathbf{F} (\mathbf{Y} - \mathbf{XUV}^T)\right), \quad (4)$$

$$\|\mathbf{X} - \mathbf{XS}\|_{2,1} = \text{tr}\left((\mathbf{X} - \mathbf{XS})^T \mathbf{H} (\mathbf{X} - \mathbf{XS})\right)$$

where the matrices $\mathbf{F}$ and $\mathbf{H}$ are diagonal, and their diagonal elements are defined as:

$$f_{i,i} = \frac{1}{2\|(\mathbf{Y} - \mathbf{XUV}^T)^i\|_2^2}, i = 1, \dots, q, \quad (5)$$

$$h_{j,j} = \frac{1}{2\|(\mathbf{X} - \mathbf{XS})^j\|_2^2}, j = 1, \dots, p$$

Equation (5) implies that a large weight will be assigned to a subject/sample that can be characterized with a small prediction loss (i.e., $(\mathbf{Y} - \mathbf{XUV}^T)^i$), while a small weight will be

assigned to noisy samples with high prediction loss. The optimization and a convergence analysis of Eq. (1) are presented in Appendixes 1 and 2. Some observations of Eq. (1) are following:

Remark 1. In Eq. (1), the first two terms take into consideration correlations between the SNP data $\mathbf{X}$ and the neuroimaging data $\mathbf{Y}$, the last two terms conduct variable selection to select important SNPs for neuroimaging genetic analysis, and the last three terms take the correlations among the features into account. Moreover, the reduced rank constraints on $\mathbf{U}$ and $\mathbf{V}$ take the correlations among variables of the response matrix into account.

Remark 2. Equation (1) jointly models all possible correlations among the data for neuroimaging genetic analysis. The robust loss functions (i.e., the first two loss functions) could effectively alleviate the influence of noisy samples, while the square-root-operator on the robust loss functions facilitates automatic learning of weights of different models, i.e., the regression model between the response matrix and the feature matrix in the first term as well as the reconstruction model in the second term. Therefore, less tuning parameters are needed in Eq. (1) compared with alternative modeling strategies.

Compared with the Sparse Reduced-Rank Regression (SRRR) (Vounou et al. 2012), our method has two contributions, i.e., explicitly modeling the correlations among the genetic data as a graph, and conducting a parameter-free imaging-genetic analysis. To investigate advantages of individual components of our objective function of Eq. (1), we obtain two variants of our method by making following changes. Firstly, we remove the graph based data representation of the genetic data and obtain a variant, referred to as SRRR without a regularization parameter (SRRR-P). The SRRR-P method is modeled by Eq. (6):

$$\min_{\mathbf{U}, \mathbf{V}} \sqrt{\|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1}} + \|\mathbf{U}\|_{2,1}^{2,1} \text{ subject to } \mathbf{V}^T \mathbf{V} = \mathbf{I}. \quad (6)$$

Secondly, we replace our parameter-free model with an extra regularization parameter, referred to as parametric robust reduced rank graph regression (P-RRRGR). The P-RRRGR method is modeled by Eq. (7):

$$\min_{\mathbf{S}, \mathbf{U}, \mathbf{V}} \|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1} + \alpha \|\mathbf{X} - \mathbf{XS}\|_{2,1} + \lambda_1 \|\mathbf{U}\|_{2,1} + \lambda_2^{2,1} \|\mathbf{S}\|_1 \text{ subject to } \mathbf{V}^T \mathbf{V} = \mathbf{I} \quad (7)$$

and $\text{diag}(\mathbf{S}) = 0$,

where α is a tunable parameter to regularize weights of the first two terms.

Experiments

We compared our method with four state-of-the-art methods for conducting imaging-genetic analysis on a subset of the ADNI database ('www.adni-info.org'), consisting of baseline MRI data and SNP data of 737 subjects. Particularly, the ADNI was launched in 2003 as a

public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been designed to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The ADNI MRI scans were acquired using a sagittal 3D MP-RAGE T1-w sequence (TR = 2400 ms, minimum full TE, TI = 1000 ms, FOV = 240 mm, voxel size of $1.25 \times 1.25 \times 1.2 \text{ mm}^3$) (Jack et al. 2008). For up-to-date information, see www.adni-info.org.

Data Preprocessing

The MRI scans were processed using a standard protocol, including spatial distortion correction and bias field correction, followed by skull-stripping, cerebellum removal, intensity inhomogeneity correction, segmentation, and registration. Based on the AAL template, we finally obtained gray matter volume measures of 90 cortical and subcortical regions for each MRI scan to characterize its anatomy.

Each of the MRI scans had corresponding genetic data obtained from the ADNI 1 cohort, consisting of 620,901 SNPs. The SNPs were processed by taking the following two steps, i.e., quality control and imputation (Bertram et al. 2007). The quality control step included 1) call rate check per subject and per SNP marker; 2) gender check; 3) sibling pair identification; 4) the Hardy-Weinberg equilibrium test; 5) marker removal by the minor allele frequency; and 6) population stratification. The imputation step imputed the incomplete SNPs with the modal value. Finally, we obtained 3996 SNPs, within the boundary of 20 K base pairs of the 153 Alzheimer's disease (AD) candidate genes listed on the AlzGene database (<http://www.alzgene.org/>) as of 4/18/2011. Because the dimension of SNPs was much higher than the dimension of MRI features (i.e., 90), an unsupervised variable selection method (He et al. 2006) was adopted to reduce the dimensions of SNPs to the similar level of neuroimaging features, i.e., 100 SNPs.

Experimental Settings

We compared our method with Regularized Multivariate Ridge Regression (RMRR) (Hoerl and Kennard 1970), Multi-Task Feature Learning (MTFL) (Argyriou et al. 2007), Group Multi-Task Feature Learning (GMTFL) (Wang et al. 2012a, b), and SRRR (Vounou et al. 2012). We also empirically evaluated variants of our method, namely SRRR-P and P-RRRGR.

The RMRR method adopts an ℓ_2 -norm regularization in conjunction with the least square regression for achieving robust regression results. This method could be used to explore associations between SNPs and neuroimaging measures by separately estimating individual neuroimaging measures based on a set of SNPs under study. In our experiments, we firstly conducted RMRR between p SNPs and q ROIs to select top SNPs using all the neuroimaging features (Chen et al. 2014). Particularly, the top SNPs were selected based on the absolute values of the regression coefficient matrix row-by-row. Once the top SNPs were identified, they were used as features to regress all q ROIs.

The MTFL method adopts an $\ell_{2,1}$ -norm based group sparsity regularization in conjunction with the least square regression for multi-task regression. This method could be used to characterize associations between SNPs and imaging measures in a multiple task regression setting, and at the same time to identify the most informative SNPs due to its group sparsity regularization.

The GMTFL method takes into consideration interlinked relationship among the genetic data to conduct variable selection on the feature matrix. Hence, this method considers correlations among features and correlations between measures of a response matrix and the feature matrix to select a subset of SNPs for conducting imaging-genetic analysis.

Sparse Reduced-Rank Regression (SRRR) conducts variable selection on the feature matrix by taking the correlations among the response variables and the correlations between the response matrix and the feature matrix into account, in order to select a subset of SNPs for conducting imaging-genetic analysis.

Except the RMRR method that does not consider any potential correlations among data of the feature matrix or correlations among data of the response matrix, all other methods under comparison consider potential correlation among data of the response matrix or potential correlation among data of the feature matrix. In contrast to these alternative methods under comparison, our method considers both correlations among data of the response matrix and correlation among data of the feature matrix.

In our experiments, we used 3-fold cross-validation to compare all methods. Specifically, we repeated the whole process of 3-fold cross-validation 20 times to avoid any possible bias caused by dataset partitioning for the cross-validation. The final performance was measured by averaging results of all the 20 runs of the cross-validation. We further employed a nested 3-fold cross-validation for model selection by setting the parameters in the range of $\{10^{-3}, \dots, 10^3\}$ for all methods and varying the values of r in $\{1, 2, \dots, 10\}$ for our method.

We used the Root Mean Squared Error (RMSE) as a metric to evaluate the performance of all the methods. Specifically, we firstly analyzed RMSE results of all the methods as functions of different numbers of SNPs, and then reported top 10 selected SNPs and top 10 selected ROIs of all the methods. Finally, we presented the RMSE results of our method with different numbers of ranks and different numbers of SNPs.

Experimental Results

The RMSE results (including mean and standard deviation) shown in the left side of Fig. 1 demonstrate that the proposed method achieved the overall best performance, with an average improving of 12.56% over the alternative methods under comparison. The paired t-tests between our method and each of the alternative methods under comparison showed that all p -values were smaller than 0.001.

We also evaluated our method in terms of its sensitivity to ranks of the data matrix (i.e., r) and reported the RMSE results associated with different ranks and different numbers of the SNPs to predict the neuroimaging measures in Eq. (1). The right side of Fig. 1 shows change of RMSE results with different ranks of the data matrix of SNPs, i.e., $r \in \{1, 2, \dots, 10\}$,

where the mean and the standard deviation of the RMSE were obtained from all experiments and each curve was the change of RMSE with a fixed number of SNPs to predict the neuroimaging measures, e.g., ‘top 20’ denotes the change of RMSE using top 20 SNPs to predict the imaging measures. These change plots show that the best ranks of the matrix of SNPs for our method to predict the neuroimaging measures were 6, 7, and 8. These results clearly demonstrate that a reduced rank constraint might help find the reduced rank structure of high-dimensional neuroimaging data by taking into consideration the correlations among the response variables. In summary, the results shown in the right side of Fig. 1 indicate that the promising performance achieved by our method was largely due to the fact that our method adopts the constraints of (1) reduced rankness, (2) the graph representation with sparse penalties, and (3) the parameter-free sample and model weightings, in a unified framework.

In our experiments, we averaged the absolute values of UV^T in Eq. (1) from all 60 experiments (i.e., 20 runs of the 3-fold cross-validation), then sorted the rows of the average absolute values of UV^T in a descending order to select the top 10 SNPs and sorted the columns of the average absolute values of UV^T in a descending order to select the top ROIs. We also averaged the absolute values of the weight coefficient matrix measuring the linear relationship between the SNPs and the neuroimaging features to obtain the order of both the SNPs and the neuroimaging features of all the other methods. The resulting heatmaps of the regression coefficients of the selected top 10 SNPs and the ordered ROIs of all methods are shown separately in Fig. 2.

As shown in Fig. 2, the top 10 selected SNPs came from three genes, i.e., APOlipoprotein E (APOE), Phosphatidylinositol Binding Clathrin Assembly Protein (PICALM), and Sortilin Related VPS10 Domain Containing Receptor 1 (SORCS1). These genes have been reported among the top 40 genes related to AD at the AlzGene database (Reitz et al. 2011a, b; Reitz 2012; Bettens et al. 2013). Specifically, the APOE gene has been reported to be related to either the age variants or the hippocampal volume loss of AD patients (Corder et al. 1993; Fallin et al. 2001; Schuff et al. 2009), the PICALM gene located in chromosome 11 could influence AD risk through amyloid precursor protein (APP) process via endocytic pathways, resulting in changes in A β levels (Harold et al. 2009; Chen et al. 2012), and the SORCS1 gene may increase AD risk (Rogaeva et al. 2007; Reitz et al. 2011a, b).

The selected top 10 brain regions, as shown in the right side of Fig. 2, were largely affected by AD as reported in neuroimaging studies of AD (Zhang et al. 2011, 2012; Zhu et al. 2017b). However, these brain regions might have varied degrees of correlation with AD. Particularly, the right hippocampal formation and the left uncus might be more closely related to AD than other regions.

Figure 3 shows a heatmap of the regression coefficients of the selected top 10 SNPs and the top 10 ROIs obtained by our method. This result demonstrates that there existed strong correlation between the top ranked SNPs (such as APOE gene) and the top ranked brain regions.

The experimental results for evaluating variants of our method, namely SRRR-P and P-RRRGR, demonstrate that on average the SRRR improved the SRRR-P by 0.72% and the P-RRRGR improved our method by 0.16% in terms of RMSE with varied numbers of the selected SNPs. However, the performance differences between these methods were marginal, without statistical significance (i.e., SRRR vs. SRRR-R and P-RRRGR vs. ours, paired t-tests at 95% significance level). The SRRR-P was faster ~10 times than the SRRR, while our method was faster ~100 times than the P-RRRGR. The difference in computational efficiency was caused by the extra regularization parameter which had to be tuned using cross-validation. Comparison results between the SRRR and the P-RRRGR methods reveals that the latter had a significantly better performance than the former in terms of RMSE results with different numbers of the selected SNPs (on average by 6.56% better).

Discussion and Conclusions

In this paper, we proposed a novel robust reduced rank graph regression model to explore associations between SNPs and brain imaging measures. The proposed reduced rankness constraint and sparse graph representation regularization in the SNPs along with two sparsity constraints in a linear regression framework could help characterize inherent correlation information in genetic data and brain imaging data effectively. The experimental results have demonstrated that our method could achieve promising performance in terms of characterizing associations between genetic and neuroimaging data.

Different from existing state-of-the-art methods, our method jointly considers all inherent correlation information of both genetic and neuroimaging data to characterize their associations. Our method is robust to data outliers because of its robust loss functions. Furthermore, the square-root-operator on the robust loss functions facilitates adaptive balancing between two loss functions without tuning a parameter. By contrast, the existing alternative methods need to tune a regularization parameter. Last but not the least, the optimization problem of our method could be solved using an iterative optimization method and its convergence has been theoretically proved.

We also empirically evaluated variants of the SRRR and our method, namely SRRR-P and P-RRRGR, based on the same data with the same experimental setting as presented in Section “Experimental Settings”. The experimental results have demonstrated that the regression models with an extra regularization parameter could achieve better data regression performance than those without the regularization parameter, albeit the differences were marginal. Comparison results between the SRRR and the P-RRRGR methods have also indicated that the graph representation adopted by the P-RRRGR for characterizing correlations among the SNPs could improve the data regression model compared with the SRRR method, and subsequently improve the discovery of association between the genetic and neuroimaging data.

The time complexity of our optimization method to solve Eq. (1) is cubic to the number of the SNPs. Therefore, we conducted a pre-processing process in our experiment. In our future work, we will employ sampling techniques to conduct the SNP selection. Our method is developed based on an assumption of a linear relationship between the neuroimaging data

and the SNPs, and therefore our method is not equipped to capture nonlinear associations. We will further extend our method to exploit nonlinear relationships between the genetic data and neuroimaging data by adopting kernel methods.

Acknowledgments

This work was supported in part by National Institutes of Health grants [EB022573, CA223358, DK114786, DA039215, and DA039002].

References

- Argyriou A, Evgeniou T, Pontil M. Multi-task feature learning. In: Schölkopf B, Platt J, Hoffman T, editors. *Advances in neural information processing systems*. Vol. 19. Cambridge: MIT Press; 2007. 41–48.
- Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nature Genetics*. 2007; 39(1):17–23. [PubMed: 17192785]
- Bettens K, Sleegers K, Van Broeckhoven C. Genetic insights in Alzheimer's disease. *Lancet Neurology*. 2013; 12(1):92–104. [PubMed: 23237904]
- Chen LH, Kao PYP, Fan YH, Ho DTY, Chan CSY, Yik PY, Ha JCT, Chu LW, Song YQ. Polymorphisms of CR1, CLU and PICALM confer susceptibility of Alzheimer's disease in a southern Chinese population. *Neurobiol Aging*. 2012; 33(1):210.e211–210.e217. 210.e1–210.e7.
- Chen L, Pourahmadi M, Maadooliat M. Regularized multivariate regression models with skew-t error distributions. *Journal of Statistical Planning and Inference*. 2014; 149:125–139.
- Corder E, Saunders A, Strittmatter W, Schmechel D, Gaskell P, Small Ga, Roses A, Haines J, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993; 261(5123):921–923. [PubMed: 8346443]
- Du L, Liu K, Zhang T, Yao X, Yan J, Risacher SL, Han J, Guo L, Saykin AJ, Shen L. A novel SCCA approach via truncated ℓ_1 -norm and truncated group lasso for brain imaging genetics. *Bioinformatics*. 2017.
- Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D, Schork NJ. Genetic analysis of case/control data using estimated haplotype frequencies: Application to APOE locus variation and Alzheimer's disease. *Genome Res*. 2001; 11(1):143–151. [PubMed: 11156623]
- Fan Y, Shen D, Davatzikos C. Classification of structural images via high-dimensional image warping, robust feature extraction and SVM. *Med Image Comput Comput Assist Interv*. 2005; 8(Pt 1):1–8.
- Fan Y, Shen D, Gur RC, Gur RE, Davatzikos C. COMPARE: Classification of morphological patterns using adaptive regional elements. *IEEE Trans Med Imaging*. 2007; 26(1):93–105. [PubMed: 17243588]
- Fan Y, Shi F, Smith JK, Lin W, Gilmore JH, Shen D. Brain anatomical networks in early human brain development. *Neuroimage*. 2011; 54(3):1862–1871. [PubMed: 20650319]
- Fu L, Liu L, Zhang J, Xu B, Fan Y, Tian J. Comparison of dual-biomarker PIB-PET and dual-tracer PET in AD diagnosis. *Eur Radiol*. 2014; 24(11):2800–2809. [PubMed: 25097125]
- Fu L, Liu L, Zhang J, Xu B, Fan Y, Tian J. Brain network alterations in Alzheimer's disease identified by early-phase PIB-PET. *Contrast Media & Molecular Imaging*. 2018; 2018:10.
- Ge T, Schumann G, Feng J. Imaging genetics — Towards discovery neuroscience. *Quantitative Biology*. 2013; 1(4):227–245.
- Greenlaw K, Szefer E, Graham J, Lesperance M, Nathoo FS. Initi AsDN. A Bayesian group sparse multi-task regression model for imaging genetics. *Bioinformatics*. 2017; 33(16):2513–2522. [PubMed: 28419235]
- Hao XK, Yao XH, Yan JW, Risacher SL, Saykin AJ, Zhang DQ, Shen L. Neuroimaging AsD. Identifying multimodal intermediate phenotypes between genetic risk factors and disease status in Alzheimer's disease. *Neuroinformatics*. 2016; 14(4):439–452. [PubMed: 27277494]

- Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics*. 2009; 41(10):1088–1093. [PubMed: 19734902]
- He X, Cai D, Niyogi P. Laplacian score for feature selection. *Advances in Neural Information Processing Systems*. 2006; 18:507–514.
- Hibar DP, Stein JL, Renteria ME, et al. Common genetic variants influence human subcortical brain structures. *Nature*. 2015; 520(7546):224–U216. [PubMed: 25607358]
- Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 1970; 12(1):55–67.
- Holland PW, Welsch RE. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*. 1977; 6(9):813–827.
- Hu R, Zhu X, Cheng D, He W, Yan Y, Song J, Zhang S. Graph self-representation method for unsupervised feature selection. *Neurocomputing*. 2017; 220:130–137.
- Huang C, Thompson P, Wang Y, Yu Y, Zhang J, Kong D, Colen RR, Knickmeyer RC, Zhu H. FGWAS: Functional genome wide association analysis. *NeuroImage*. 2017; 159(Supplement C):107–121. [PubMed: 28735012]
- Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. 2008; 27(4):685–691. [PubMed: 18302232]
- Kong D, Giovanello KS, Wang Y, Lin W, Lee E, Fan Y, Doraiswamy PM, Zhu H. Predicting Alzheimer's disease using combined imaging-whole genome SNP data. *Journal of Alzheimer's Disease*. 2015; 46:695–702.
- Liu L, Wang JJ, Fan Y. Morphological and functional changes in the developing brain during childhood and Adolescence OHBM Annual Meeting; Hamburg, Germany. 2014.
- Liu L, Fu L, Zhang X, Zhang J, Zhang X, Xu B, Tian J, Fan Y. Combination of dynamic (11)C-PIB PET and structural MRI improves diagnosis of Alzheimer's disease. *Psychiatry Research*. 2015; 233(2):131–140. [PubMed: 26095348]
- Lu ZH, Khondker Z, Ibrahim JG, Wang Y, Zhu HT. Initi AsDN. Bayesian longitudinal low-rank regression models for imaging genetic data from longitudinal studies. *Neuroimage*. 2017; 149:305–322. [PubMed: 28143775]
- Medland SE, Jahanshad N, Neale BM, Thompson PM. Whole-genome analyses of whole-brain data: Working within an expanded search space. *Nature Neuroscience*. 2014; 17(6):791–800. [PubMed: 24866045]
- Peng H, Fan Y. Direct sparsity optimization based feature selection for multi-class classification. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*; New York: AAAI Press; 2016. 1918–1924.
- Peng H, Fan Y. Feature selection by optimizing a lower bound of conditional mutual information. *Information Sciences*. 2017a; 418:652–667.
- Peng H, Fan Y. A general framework for sparsity regularized feature selection via iteratively reweighted Least Square minimization. *AAAI*; 2017b.
- Reitz C. Alzheimer's disease and the amyloid cascade hypothesis: a critical review. *International Journal of Alzheimer's Disease*. 2012.
- Reitz C, Brayne C, Mayeux R. Epidemiology of Alzheimer disease. *Nature Reviews Neurology*. 2011a; 7(3):137–152. [PubMed: 21304480]
- Reitz C, Tokuhiro S, Clark LN, Conrad C, Vonsattel JP, Hazrati LN, Palotás A, Lantigua R, Medrano M, Jiménez-Velázquez IZ. SORCS1 alters amyloid precursor protein processing and variants may increase Alzheimer's disease risk. *Annals of Neurology*. 2011b; 69(1):47–64. [PubMed: 21280075]
- Rogaeva E, Meng Y, Lee JH, Gu Y, Kawarai T, Zou F, Katayama T, Baldwin CT, Cheng R, Hasegawa H. The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer's disease. *Nature Genetics*. 2007; 39(2):168–177. [PubMed: 17220890]
- Schuff N, Woerner N, Boreta L, Kornfield T, Shaw L, Trojanowski J, Thompson P, Jack C Jr, Weiner M. Alzheimer's and D.N. Initiative. MRI of hippocampal volume loss in early Alzheimer's disease

- in relation to ApoE genotype and biomarkers. *Brain*. 2009; 132(4):1067–1077. [PubMed: 19251758]
- Thompson PM, Ge T, Glahn DC, Jahanshad N, Nichols TE. Genetics of the connectome. *Neuroimage*. 2013; 80:475–488. [PubMed: 23707675]
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 2002; 15(1):273–289. [PubMed: 11771995]
- Vounou M, Nichols TE, Montana G, Initia ADN. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage*. 2010; 53(3):1147–1159. [PubMed: 20624472]
- Vounou M, Janousova E, Wolz R, Stein JL, Thompson PM, Rueckert D, Montana G, Initia ADN. Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease. *Neuroimage*. 2012; 60(1):700–716. [PubMed: 22209813]
- Wang H, Nie F, Huang H, Risacher SL, Saykin AJ, Shen L. Alzheimer's Dis Neuroimaging I. Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics*. 2012a; 28(12):1127–1136. [PubMed: 22689752]
- Wang H, Nie FP, Huang H, Yan JW, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L. Initia AsDN. From phenotype to genotype: An association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs. *Bioinformatics*. 2012b; 28(18):1619–1625. [PubMed: 22962490]
- Wen Z, Yin W. A feasible method for optimization with orthogonality constraints. *Math Program*. 2013; 142(1–2):397–434.
- Zhang DQ, Wang YP, Zhou LP, Yuan H, Shen DG, Initia ADN. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011; 55(3):856–867. [PubMed: 21236349]
- Zhang DQ, Shen DG. Neuroimaging AsD. Multimodal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage*. 2012; 59(2):895–907. [PubMed: 21992749]
- Zhang Y, Caspers S, Fan LZ, Fan Y, Song M, Liu CR, Mo Y, Roski C, Eickhoff S, Amunts K, Jiang TZ. Robust brain parcellation using sparse representation on resting-state fMRI. *Brain Structure and Function*. 2015; 220(6):3565–3579. [PubMed: 25156576]
- Zheng W, Zhu X, Zhu Y, Hu R, Lei C. Dynamic graph learning for spectral feature selection. *Multimedia Tools and Applications*. 2017:1–17.
- Zhu X, Li X, Zhang S, Ju C, Wu X. Robust joint graph sparse coding for unsupervised spectral feature selection. *IEEE transactions on neural networks and learning systems*. 2017a; 28(6):1263–1275. [PubMed: 26955053]
- Zhu X, Suk HI, Huang H, Shen D. Low-rank graph-regularized structured sparse regression for identifying genetic biomarkers. *IEEE Transactions on Big Data*. 2017b; (99):1–1.
- Zhu X, Zhang S, Hu R, Zhu Y. Local and global structure preservation for robust unsupervised spectral feature selection. *IEEE Transactions on Knowledge and Data Engineering*. 2018; 30(3):517–529.

Appendix 1

Equation (1) is not jointly convex in terms of the variables $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$, but is convex for each individual variable while fixing the others. In this paper, we employ the alternative optimization strategy to optimize Eq. (1), i.e., iteratively optimizing each variable while fixing the others until the algorithm converges. Moreover, both the ℓ_1 -norm regularization on $\mathbf{U}$ as well as $\mathbf{S}$ and the ℓ_1 -norm regularization on $\mathbf{S}$ in Eq. (1) are convex but non-smooth, we adopt the framework of Iteratively Reweighted Least Square (IRLS) (Holland and Welsch 1977; Hu et al. 2017) to optimize $\mathbf{S}$ and $\mathbf{U}$. More specifically, we first reformulate

Eq. (1) as Eq. (2) by fixing the values of ω_1 and ω_2 , and then iteratively optimize 1) the variables $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$ in Eq. (1) and 2) the values of ω^1 and ω^2 . Pseudo codes for solving Eq. (1) are summarized in Algorithm 1.

(i) Update $\mathbf{U}$ by fixing the others

While fixing the other variables, Eq. (1) with respect to $\mathbf{U}$ and $\mathbf{V}$ is:

$$\min_{\mathbf{U}, \mathbf{V}} \omega_1 \|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1} \text{ subject to } \mathbf{V}^T \mathbf{V} = \mathbf{I}. \quad (8)$$

As $\mathbf{V}$ has orthogonal columns, there is a matrix $\mathbf{V}^\perp$ with orthogonal columns such that $(\mathbf{V}, \mathbf{V}^\perp)$ is an orthogonal matrix. Thus, we have

$$\begin{aligned} \|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1} &= \|(\mathbf{Y} - \mathbf{XUV}^T)(\mathbf{V}, \mathbf{V}^\perp)\|_{2,1} \quad (9) \\ &= \|\mathbf{YV} - \mathbf{XU}\|_{2,1} + \|\mathbf{YV}^\perp\|_{2,1}. \\ &= \|\mathbf{YV} - \mathbf{XU}\|_{2,1} \end{aligned}$$

The second term on the right side in Eq. (9) does not involve $\mathbf{U}$. For fixed $\mathbf{V}$ and $\mathbf{S}$, we substitute Eq. (9) into Eq. (8) to obtain:

$$\min_{\mathbf{U}} \|\mathbf{YV} - \mathbf{XU}\|_{2,1} + \lambda_1 / \omega_1 \|\mathbf{U}, \mathbf{S}\|_{2,1}. \quad (10)$$

By employing the framework of IRLS to optimize $\mathbf{U}$ and setting the derivative of Eq. (10) with respect to $\mathbf{U}$ to zero, we have:

$$\mathbf{U} = (\mathbf{X}^T \mathbf{F} \mathbf{X} + \lambda_1 / \omega_1 \mathbf{D})^{-1} \mathbf{X}^T \mathbf{F} \mathbf{Y} \mathbf{V}. \quad (11)$$

where $\mathbf{D} \in \mathbb{R}^{p \times p}$ is a diagonal matrix with $d_{jj} = \frac{1}{2\|\bar{\mathbf{u}}^j\|_2^2}$, $j = 1, \dots, p$, where $\mathbf{U}^\sim = [\mathbf{U}, \mathbf{S}]$

$\in \mathbb{R}^{p \times (p+q)}$ and the definition of $\mathbf{F}$ can be found in Eq. (10).

(ii) Update $\mathbf{V}$ and $\mathbf{S}$ by fixing the others

The objective function of Eq. (1) with respect to $\mathbf{V}$ is

$$\min_{\mathbf{V}} \omega_1 \|\mathbf{Y} - \mathbf{XUV}^T\|_{2,1}, \text{ subject to } \mathbf{V}^T \mathbf{V} = \mathbf{I}. \quad (12)$$

Equation (12) is an orthogonal optimization problem. We solve it using the method in (Wen and Yin 2013) and list the pseudo of the optimization of $\mathbf{V}$ in Algorithm 2, where $\nabla F = \mathbf{V}\mathbf{U}^T\mathbf{X}^T\mathbf{F}\mathbf{X}\mathbf{U} + \mathbf{Y}^T\mathbf{F}^T$ is the derivative of Eq. (12) with respect to $\mathbf{V}$.

Algorithm 1

Pseudo codes for solving Eq. (1).

Input: $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{Y} \in \mathbb{R}^{n \times q}$, λ_1 , λ_2 , τ ;
Output: $\mathbf{S}$, $\mathbf{U}$, $\mathbf{V}$;

- 1 Initialize $t = 1$;
- 2 Initialize $\mathbf{U}^{(1)}$ and $\mathbf{S}^{(1)}$ as two random matrices;
- 3 $\mathbf{U}^{(1)} = [\mathbf{U}^{(1)}, \mathbf{S}^{(1)}]$;
- 4 **repeat**
- 5 Calculate the diagonal matrix $\mathbf{D}^{(t+1)}$ with $d_{jj}^{(t+1)} = \frac{1}{2\|(\mathbf{U}^{(t+1)})^j\|_2^2}$, $j = 1, \dots, p$;
- 6 **for each** $i (1 \leq i \leq p)$ **do**
- 7 Calculate the diagonal matrix $\mathbf{C}_i^{(t+1)}$ with the j -th diagonal element as $\frac{1}{2\|s_{ij}^{(t+1)}\|}$;
- 8 Update $s_i^{(t+1)}$ via Eq. (A. 8);
- 9 Update $\mathbf{U}^{(t+1)}$ via Eq. (A. 4);
- 10 Update $\mathbf{U}^{(t+1)}$ via $\mathbf{U}^* = [\mathbf{U}, \mathbf{S}]$;
- 11 Update $\mathbf{V}^{(t+1)}$ via Algorithm 2;
- 12 Update ω_1 and ω_2 via Eq. (3)
- 13 Update $\mathbf{F}$ and $\mathbf{H}$ via Eq. (5);
- 14 $t = t + 1$;
- 15 **until** Eq. (1) converges;

Once having $\mathbf{U}$ and $\mathbf{V}$, we can obtain $\mathbf{F}$ via Eq. (10).

(iii) Update $\mathbf{S}$ by fixing the others

By fixing the other variables, the objective function of Eq. (1) with respect to $\mathbf{S}$ can be rewritten as:

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{X}\mathbf{S}\|_{2,1} + \lambda_1/\omega_2 \|\mathbf{U}, \mathbf{S}\|_{2,1} + \lambda_2/\omega_2 \|\mathbf{S}\|_1, \text{ subject to } \text{diag}(\mathbf{S}) = 0. \quad (13)$$

By employing the framework of IRLS to optimize $\mathbf{S}$, we take the derivative with respect to $s_i (1 \leq i \leq p)$ in Eq. (13) and then set it to zero to have

$$\min_{\mathbf{S}_i} \mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{s}_i - \mathbf{X}^T \mathbf{H} \mathbf{x}_i + \frac{\lambda_1}{\omega_2} \mathbf{D} \mathbf{s}_i + \frac{\lambda_2}{\omega_2} \mathbf{C}_i \mathbf{s}_i, \text{ subject to } \mathbf{s}_{ii} = 0. \quad (14)$$

where $\mathbf{C}_i \in \mathbb{R}^{p \times p}$ is a diagonal matrix with the j -th diagonal element as $\frac{1}{2|s_{i,j}|}$ and the definition of $\mathbf{H}$ can be found in Eq. (10). Thus, we have

$$\hat{\mathbf{s}}_i = \left(\mathbf{X}^T \mathbf{H} \mathbf{X} + \frac{\lambda_1}{\omega_2} \mathbf{D} + \frac{\lambda_2}{\omega_2} \mathbf{C}_i \right)^{-1} \mathbf{X}^T \mathbf{H} \mathbf{x}_i, \text{ and } s_{ii} = 0. \quad (15)$$

Once $\mathbf{U}$ and $\mathbf{S}$ are available, we can obtain $\mathbf{H}$ via Eq. (10).

(iv) Update ω_1 and ω_2 by fixing the others

After obtaining $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$, we can obtain the values of ω_1 and ω_2 via Eq. (8).

Algorithm 2

Pseudo code of solving the variable $\mathbf{V}$.

Input: $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{U} \in \mathbb{R}^{p \times r}$, and $\mathbf{F} \in \mathbb{R}^{q \times q}$;
Output: $\mathbf{V} \in \mathbb{R}^{q \times r}$

- 1 Initialize $t = 1$;
- 2 **repeat**
- 3 $\mathbf{E} \leftarrow \mathbf{V}^{(t)} \mathbf{V}^T \mathbf{F} \mathbf{T}^{(t)} \mathbf{T} - \mathbf{V}^T \mathbf{F} \mathbf{T}^{(t)} \mathbf{V}^{(t)}$;
- 4 $\tau \leftarrow$ non-monotonic grid search;
- 5 $\mathbf{V}^{(t+1)} \leftarrow (\mathbf{I} - \frac{\tau}{2} \mathbf{E})^{-1} (\mathbf{I} + \frac{\tau}{2} \mathbf{E}) \mathbf{X}^{(t)}$;
- 6 $t \leftarrow (t+1)$;
- 7 **until** Convergence;

Appendix 2

We prove the convergence of our proposed Algorithm 1 to solve our objective function of Eq. (1) as following.

After the t -th iteration step, we have obtained the optimal $\mathbf{U}^{(t)}$, $\mathbf{V}^{(t)}$ and $\mathbf{S}^{(t)}$. In the $(t+1)$ -th iteration step, we need to optimize $\mathbf{U}^{(t+1)}$ by fixing the others so that $\mathbf{U}^{(t+1)} = \mathbf{U}^{(t)}$. According to Eq. (11), $\mathbf{U}^{(t+1)}$ has a closedform optimal solution. Thus we have the following inequality:

$$\begin{aligned}
& \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t+1)}\mathbf{V}^{\mathbf{T}(t)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t)}\|_1 + \lambda_1\|\mathbf{U}^{(t+1)}, \mathbf{S}^{(t)}\|_{2,1} \quad (16) \\
& \leq \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t)}\mathbf{V}^{\mathbf{T}(t)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t)}\|_1 + \lambda_1\|\mathbf{U}^{(t)}, \mathbf{S}^{(t)}\|_{2,1}.
\end{aligned}$$

According to the literature (Wen and Yin 2013), the optimization of Eq. (15) converges. This implies that the following holds:

$$\begin{aligned}
& \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t+1)}\mathbf{V}^{\mathbf{T}(t+1)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t)}\|_1 + \lambda_1\|\mathbf{U}^{(t+1)}, \mathbf{S}^{(t)}\|_{2,1} \quad (17) \\
& \leq \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t+1)}\mathbf{V}^{\mathbf{T}(t)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t)}\|_1 + \lambda_1\|\mathbf{U}^{(t+1)}, \mathbf{S}^{(t)}\|_{2,1}.
\end{aligned}$$

According to conclusions in the literature (Hu et al. 2017, Zhu et al. 2017a, 2018), the framework of IRLS converges, so we have the following:

$$\begin{aligned}
& \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t+1)}\mathbf{V}^{\mathbf{T}(t+1)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t+1)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t+1)}\|_1 + \lambda_1\|\mathbf{U}^{(t+1)}, \mathbf{S}^{(t+1)}\|_{2,1} \\
& \leq \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t+1)}\mathbf{V}^{\mathbf{T}(t+1)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t)}\|_1 + \lambda_1\|\mathbf{U}^{(t+1)}, \mathbf{S}^{(t)}\|_{2,1} \\
& .
\end{aligned}$$

(18)

By integrating Eq. (16) with Eq. (17) and Eq. (18), we have:

$$\begin{aligned}
& \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t+1)}\mathbf{V}^{\mathbf{T}(t+1)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t+1)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t+1)}\|_1 + \lambda_1\|\mathbf{U}^{(t+1)}, \mathbf{S}^{(t+1)}\|_{2,1} \\
& \leq \sqrt{\|\mathbf{Y} - \mathbf{X}\mathbf{U}^{(t)}\mathbf{V}^{\mathbf{T}(t)}\|_{2,1}} + \sqrt{\|\mathbf{X} - \mathbf{X}\mathbf{S}^{(t)}\|_{2,1}} + \lambda_2\|\mathbf{S}^{(t)}\|_1 + \lambda_1\|\mathbf{U}^{(t)}, \mathbf{S}^{(t)}\|_{2,1} \\
& .
\end{aligned}$$

(19)

Equation (19) indicates that the objective function value of Eq. (1) decreases after each iteration step of Algorithm 1. Thus, our proposed Algorithm 1 to solve the proposed objective function Eq. (1) converges.

Fig. 1. RMSE of all the methods with different numbers of the selected SNPs (left) and our method with different numbers of ranks using different numbers of SNPs to predict test data (right)

Fig. 2.

Heatmaps of regression coefficients of the top 10 SNPs (left) and the top 10 ROIs (right) selected by the proposed method. The top 10 SNPs came from three genes, namely APOlipoprotein E (APOE), Phosphatidylinositol Binding Clathrin Assembly Protein (PICALM), and Sortilin Related VPS10 Domain Containing Receptor 1 (SORCS1). The top 10 ROIs included hippocampal formation right (hip.for.R), uncus left (unc.L), hippocampal formation left (hip.for.L), middle temporal gyrus left (mid.temp.gy.L), perirhinal cortex left (per.cor.L), temporal pole left (temp.pol.L), amygdala left (amy.L), middle temporal gyrus right (mid.temp.gy.R), amygdala right (amy.R), and lateral occipitotemporal gyrus left (lat.occ.gy.L). The color-bar indicates regression coefficients

Fig. 3.

Heatmaps of regression coefficients of the relationship of regression coefficients between the top 10 ROIs and the top 10 SNPs selected by the proposed method. The top 10 SNPs came from three genes, namely APOLipoprotein E (APOE), Phosphatidylinositol Binding Clathrin Assembly Protein (PICALM), and Sortilin Related VPS10 Domain Containing Receptor 1 (SORCS1). The top 10 ROIs included hippocampal formation right (hip.for.R), uncus left (unc.L), hippocampal formation left (hip.for.L), middle temporal gyrus left (mid.temp.gy.L), perirhinal cortex left (per.cor.L), temporal pole left (temp.pol.L), amygdala left (amy.L), middle temporal gyrus right (mid.temp.gy.R), amygdala right (amy.R), and lateral occipitotemporal gyrus left (lat.occ.gy.L), respectively. The colorbar indicates regression coefficients