Bioinformatics, 34(2), 2018, 278–285
doi: 10.1093/bioinformatics/btx594
Advance Access Publication Date: 18 September 2017
Original Paper

OXFORD

Bioimage informatics

A novel SCCA approach via truncated ℓ_1-norm and truncated group lasso for brain imaging genetics

Lei Du[1,*], Kefei Liu[2], Tuo Zhang[1], Xiaohui Yao[2], Jingwen Yan[2], Shannon L. Risacher[2], Junwei Han[1], Lei Guo[1], Andrew J. Saykin[2] and Li Shen[2,*], for the Alzheimer's Disease Neuroimaging Initiative[†]

[1]Department of Control and Information, School of Automation, Northwestern Polytechnical University, Xi'an 710072, China and [2]Department of Radiology and Imaging Sciences, Indiana University School of Medicine, IN 46202, USA

*To whom correspondence should be addressed.

[†]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.

Associate Editor: Robert Murphy

Received on February 21, 2017; revised on August 21, 2017; editorial decision on September 14, 2017; accepted on September 15, 2017

Abstract

Motivation: Brain imaging genetics, which studies the linkage between genetic variations and structural or functional measures of the human brain, has become increasingly important in recent years. Discovering the bi-multivariate relationship between genetic markers such as single-nucleotide polymorphisms (SNPs) and neuroimaging quantitative traits (QTs) is one major task in imaging genetics. Sparse Canonical Correlation Analysis (SCCA) has been a popular technique in this area for its powerful capability in identifying bi-multivariate relationships coupled with feature selection. The existing SCCA methods impose either the ℓ_1-norm or its variants to induce sparsity. The ℓ_0-norm penalty is a perfect sparsity-inducing tool which, however, is an NP-hard problem.

Results: In this paper, we propose the truncated ℓ_1-norm penalized SCCA to improve the performance and effectiveness of the ℓ_1-norm based SCCA methods. Besides, we propose an efficient optimization algorithms to solve this novel SCCA problem. The proposed method is an adaptive shrinkage method via tuning τ. It can avoid the time intensive parameter tuning if given a reasonable small τ. Furthermore, we extend it to the truncated group-lasso (TGL), and propose TGL-SCCA model to improve the group-lasso-based SCCA methods. The experimental results, compared with four benchmark methods, show that our SCCA methods identify better or similar correlation coefficients, and better canonical loading profiles than the competing methods. This demonstrates the effectiveness and efficiency of our methods in discovering interesting imaging genetic associations.

Availability and implementation: The Matlab code and sample data are freely available at http://www.iu.edu/~shenlab/tools/tlpscca/.

Contact: dulei@nwpu.edu.cn or shenli@iu.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Brain imaging genetics has gained more and more attentions in recent years. Using the quantitative endophenotype, other than the error-prone cutoffs to define case-control status, as the target measure can improve the performance in identifying disease related genetic markers (Kim *et al.*, 2013; Potkin *et al.*, 2009; Saykin *et al.*, 2015; Shen *et al.*, 2010, 2014). A major task of imaging genetics is to identify bi-multivariate associations between single nucleotide polymorphisms (SNPs) and imaging quantitative traits (QTs). As an ideal computational model, sparse canonical correlation analysis (SCCA), which is powerful in discovering bi-multivariate relationships and selecting relevant features, has become a popular technique in brain imaging genetics (Chen and Liu, 2012; Chen *et al.*, 2012, 2013; Du *et al.*, 2014, 2016a, b; Parkhomenko *et al.*, 2009; Witten *et al.*, 2009).

The canonical correlation analysis (CCA) technique has been proposed to learn the associations between two views of data (Hotelling, 1936). However, CCA suffers from severe overfitting issue because its performance decreases significantly once the number of features is larger than that of the observations. (Witten *et al.*, 2009) introduced the ℓ_1-norm into the CCA model to pursuit sparse canonical loadings. This is a desired property because it only selects a small proportion of the features which are meaningful for further investigation. After that, many sparse CCA methods using the ℓ_1-norm or its variants such as the fused lasso are proposed, and they have been successfully applied to different biology applications (Shen *et al.*, 2014). However, there are two major concerns regarding these SCCA methods. First, the ℓ_1-norm pursuits sparsity at the price of shrinking large features by a constant. This may incurs estimation bias and thus decrease the prediction power. Second, the ℓ_0-norm is the most ideal sparsity-inducing constraint, which only includes those nonzero features. Unfortunately, the ℓ_0-norm function is neither non-convex nor discontinuous. Nevertheless, it is known to be an NP-hard (nondeterministic polynomial time) problem (Fung and Mangasarian, 2011). In addition, the ℓ_1-norm constraint is not a stable feature selector (Fan and Li, 2001). Group lasso is widely used to model the group structure by considering each group of variables as a whole (Yuan and Lin, 2006). Actually, group lasso is a ℓ_1-norm constraint at the group level, and within each group a ℓ_2-norm is used. Therefore, group lasso is limited by the same problems to that of ℓ_1-norm. For example, group lasso shrinks large groups by a constant which may lead to biased results.

In order to address these issues, the likelihood-based selector, named the truncated ℓ_1-norm penalty (TLP) (Shen *et al.*, 2012; Yang *et al.*, 2012), is proposed. The TLP is defined as $J_\tau(|x|) = \min(\frac{|x|}{\tau}, 1)$ where τ is a small positive tuning parameter. TLP plays the role of the surrogate of ℓ_0-norm. It could approximate the ℓ_0-norm function and permits the desirable sparsity by providing a reasonable τ. At the same time, TLP can be equivalently transferred to a piecewise linear function, and thus is easier to handle than the ℓ_0-norm function. Besides, the TLP can be easily extended to the truncated group lasso to provide an improved group feature selector.

With the above observations, we explore the effect of using TLP and TGL in the SCCA model. We first propose the TLP based SCCA (TLP-SCCA) which embraces the TLP norm instead of the ℓ_0-norm or the ℓ_1-norm. The TLP-SCCA has the following advantages (Shen *et al.*, 2012). First, the TLP function performs as a tradeoff between the ℓ_0 and ℓ_1 function. This means that it not only has better feature selection ability, but also can be solved effectively like the ℓ_1-norm. Second, it is an adaptive shrinkage method if τ is tuned appropriately. Third, using an appropriate τ, it can discriminate the small coefficients from large coefficients, and leave those large ones less penalized. We propose an effective optimization algorithm to solve the TLP-SCCA problem. In addition, we extend TLP-SCCA to the truncated group lasso penalized SCCA, aiming to boost the existing methods using group lasso. The experimental results, compared with three state-of-the-art ℓ_1-SCCA methods (Chen and Liu, 2012; Du *et al.*, 2014; Witten *et al.*, 2009) and one group lasso penalized SCCA method (Du *et al.*, 2014), show that both TLP-SCCA and TGL-SCCA exhibit better canonical loading profiles than these benchmark methods, and also obtain higher or similar correlation coefficients.

2 Materials and methods

We use the boldface lowercase letter to denote a vector, and use the boldface uppercase letter to denote a matrix. $||\cdot||$ indicates the ℓ_2-norm of a vector. Let $\mathbf{X}\in\mathbb{R}^{n\times p}$ denote the SNP data, and $\mathbf{Y}\in\mathbb{R}^{n\times q}$ denote the QT data, where n is the number of participants, and p and q are the number of SNPs and QTs, respectively. Then the penalized CCA model using ℓ_1-norm is defined

$$\begin{aligned} &\min_{\mathbf{u},\mathbf{v}} -\mathbf{u}^\top\mathbf{X}^\top\mathbf{Y}\mathbf{v} \\ &s.t. \quad ||\mathbf{u}||^2 \le 1, ||\mathbf{v}||^2 \le 1, \Omega(\mathbf{u}) \le c_1, \Omega(\mathbf{v}) \le c_2, \end{aligned} \tag{1}$$

where $\Omega(\mathbf{u})$ and $\Omega(\mathbf{v})$ are the ℓ_1-norm functions which can control model sparsity (Parkhomenko *et al.*, 2009; Witten and Tibshirani, 2009).

2.1 The truncated ℓ_1-norm

The ℓ_1 regularization for lasso is widely used as a convex relaxation of the ℓ_0-norm function (Yang *et al.*, 2012). Though lasso regularizer yields sparse solutions, it may incur estimation bias because it overpenalizes large coefficients. The truncated ℓ_1-norm is defined as follows (Zhang, 2013):

$$\Omega_{\mathrm{TLP}}(\mathbf{u}) = \sum_{i=1}^{p} J_\tau(|u_i|), \quad \text{where} \quad J_\tau(|u_i|) = \min\left(\frac{|u_i|}{\tau}, 1\right). \tag{2}$$

The parameter τ is a predefined threshold or tuning parameter. Given an appropriate τ, TLP balances between the ℓ_1-norm and the ℓ_0-norm according to the magnitude of u_is. It is equivalent to the ℓ_1-norm if the coefficient is small enough ($|u_i| \le \tau$), while it becomes the ℓ_0-norm as the coefficient goes beyond the threshold τ ($|u_i| > \tau$).

2.2 Smoothing the penalty

Since TLP is non-convex, we first decompose it into two continuous functions, and then show how to approximate TLP by a quadratic function. Let $f_1(u_i) = |u_i| + \tau$ and $f_2(u_i) = ||u_i| - \tau|$, TLP can be represented as:

$$J_\tau(|u_i|) = \frac{1}{2\tau}[f_1(u_i) - f_2(u_i)]. \tag{3}$$

Obviously, $f_1(u_i)$ is convex. Although $f_2(u_i)$ is non-convex, it is piecewise continuously differentiable. Therefore, the TLP function is piecewise continuously differentiable (Zhang, 2013).

The first-order Taylor expansion of the function $\sqrt{x}$ at x_0 is

$$\sqrt{x} \approx \sqrt{x_0} + \frac{1}{2\sqrt{x_0}}(x - x_0) = \frac{x + x_0}{2\sqrt{x_0}}. \tag{4}$$

Substituting $x = u_i^2$ and $x_0 = (u_i^{(t)})^2$ into the right of (4), we define

$$g_1(u_i) = \frac{u_i^2 + \left(u_i^{(t)}\right)^2}{2|u_i^{(t)}|} + \tau \tag{5}$$

where $u_i^{(t)}$ is the t-th iteration in the optimizing task.

THEOREM 1. $g_1(u_i)$ is a point-by-point approximation to $f_1(u_i)$, and they have the same stationary point.

The proof is available in the Supplementary Material due to space limitation. Using the same idea, we can define

$$g_2(u_i) = \text{sign}\left(|u_i^{(t)}| - \tau\right)\left(\frac{u_i^2 + \left(u_i^{(t)}\right)^2}{2|u_i^{(t)}|} - \tau\right) \tag{6}$$

as the surrogate of $f_2(u_i)$. Similarly, $g_2(u_i)$ is a pointwise approximation to $f_2(u_i)$, and they have the same stationary point, too.

Based on Equation (3), we define the surrogate penalty of $\Omega_{\text{TLP}}(\mathbf{u})$ by combining Equations (5) and (6) across all u_i's

$$\begin{aligned}\Omega_{\text{TLP}}^{\text{App}}(\mathbf{u}) &= \frac{1}{2\tau}\sum_{i=1}^{p}[g_1(u_i) - g_2(u_i)] \\ &= \frac{1}{2\tau}\sum_{i=1}^{p}\left[\frac{u_i^2}{2|u_i^{(t)}|} - \frac{\text{sign}\left(|u_i^{(t)}| - \tau\right)}{2|u_i^{(t)}|}u_i^2\right] + C,\end{aligned} \tag{7}$$

where

$$C = \frac{1}{2\tau}\sum_{i=1}^{p}\left[\frac{|u_i^{(t)}|}{2} + \tau - \text{sign}\left(|u_i^{(t)}| - \tau\right)\left(\frac{|u_i^{(t)}|}{2} - \tau\right)\right] \tag{8}$$

is an irrelevant constant that is not a function of $\mathbf{u}$. It is easy to verify that $\Omega_{\text{TLP}}^{\text{App}}(\mathbf{u})$ is also a point-by-point approximation to $\Omega_{\text{TLP}}(\mathbf{u})$, and it is continuously differentiable and convex in terms of $\mathbf{u}$. The surrogate penalty $\Omega_{\text{TLP}}^{\text{App}}(\mathbf{v})$ can be constructed in the same way.

2.3 The TLP-SCCA model and optimization

Now using the TLP, we formally define the TLP-SCCA model as

$$\begin{aligned}&\min_{\mathbf{u},\mathbf{v}} -\mathbf{u}^\top\mathbf{X}^\top\mathbf{Y}\mathbf{v} + \lambda_1\Omega_{\text{TLP}}(\mathbf{u}) + \lambda_2\Omega_{\text{TLP}}(\mathbf{v}) \\ &s.t.\quad ||\mathbf{X}\mathbf{u}||^2 \le 1, ||\mathbf{Y}\mathbf{v}||^2 \le 1,\end{aligned} \tag{9}$$

where $\Omega_{\text{TLP}}(\mathbf{u})$, $\Omega_{\text{TLP}}(\mathbf{v})$ are the TLP penalties, and λ_1 and λ_2 are positive parameters. Though $\mathbf{X}^\top\mathbf{X}$ is generally assumed to be an identity matrix (Chen and Liu, 2012; Witten and Tibshirani, 2009), the performance can be improved when using $||\mathbf{X}\mathbf{u}||^2 \le 1$ other than $||\mathbf{u}||^2 \le 1$ (Du *et al.*, 2014). Thus we do not assume $\mathbf{X}^\top\mathbf{X}$ to be an identity.

According to the definition, TLP-SCCA transforms between ℓ_0-SCCA and ℓ_1-SCCA depending on the relationship between τ and the magnitude of the coefficients. Thus its merits are 3-fold. Firstly, it has both better sparsity-inducing ability and lower estimation bias than the ℓ_1-SCCA since it approximates the optimal ℓ_0-SCCA. Secondly, TLP-SCCA is guaranteed to perform well as long as τ is small enough. This is desirable because it avoids the time-consuming procedure to fine-tune τ, especially in high-dimensional biology study. Thirdly, it is easier to solve than the intractable ℓ_0-SCCA.

The unconstrained formulation of the TLP-SCCA problem writes

$$\begin{aligned}\mathcal{L}(\mathbf{u},\mathbf{v}) &= -\mathbf{u}^\top\mathbf{X}^\top\mathbf{Y}\mathbf{v} + \lambda_1\Omega_{\text{TLP}}(\mathbf{u}) + \lambda_2\Omega_{\text{TLP}}(\mathbf{v}) \\ &+ \frac{\gamma_1}{2}\left(||\mathbf{X}\mathbf{u}||^2 - 1\right) + \frac{\gamma_2}{2}\left(||\mathbf{Y}\mathbf{v}||^2 - 1\right),\end{aligned} \tag{10}$$

where γ_1, and γ_2 are the tuning parameters. Equation (10) is non-convex due to the non-convexness of TLP terms. In this paper, we define the following biconvex function

$$\begin{aligned}\mathcal{L}^{\text{App}}(\mathbf{u},\mathbf{v}) &= -\mathbf{u}^\top\mathbf{X}^\top\mathbf{Y}\mathbf{v} + \lambda_1\Omega_{\text{TLP}}^{\text{App}}(\mathbf{u}) + \lambda_2\Omega_{\text{TLP}}^{\text{App}}(\mathbf{v}) \\ &+ \frac{\gamma_1}{2}\left(||\mathbf{X}\mathbf{u}||^2 - 1\right) + \frac{\gamma_2}{2}\left(||\mathbf{Y}\mathbf{v}||^2 - 1\right).\end{aligned} \tag{11}$$

as the surrogate function of Equation (10).

THEOREM 2. Minimizing Equation (11) is equivalent to minimizing Equation (10).

The proof is contained in the Supplementary Material. Now that Equation (11) is quadratic and bi-convex in $\mathbf{u}$ and $\mathbf{v}$, at the stationary points it satisfies that:

$$\begin{aligned}0 &\in -\mathbf{X}^\top\mathbf{Y}\mathbf{v} + \left(\frac{\lambda_1}{\tau}\mathbf{D}_1^{(t)} - \frac{\lambda_1}{\tau}\tilde{\mathbf{D}}_1^{(t)} + \gamma_1\mathbf{X}^\top\mathbf{X}\right)\mathbf{u}, ||\mathbf{X}\mathbf{u}||^2 - 1 = 0, \\ 0 &\in -\mathbf{Y}^\top\mathbf{X}\mathbf{u} + \left(\frac{\lambda_2}{\tau}\mathbf{D}_2^{(t)} - \frac{\lambda_2}{\tau}\tilde{\mathbf{D}}_2^{(t)} + \gamma_2\mathbf{Y}^\top\mathbf{Y}\right)\mathbf{v}, ||\mathbf{Y}\mathbf{v}||^2 - 1 = 0,\end{aligned}$$

where $\mathbf{D}_1^{(t)}$ is a diagonal matrix with the i-th diagonal element being $\frac{1}{2|u_i^{(t)}|}$ ($i \in [1, p]$), and $\tilde{\mathbf{D}}_1^{(t)}$ is a diagonal matrix with the i-th diagonal element as $\frac{\text{sign}(|u_i^{(t)}| - \tau)}{2|u_i^{(t)}|}$. Likewise, $\mathbf{D}_2^{(t)}$ is a diagonal matrix with entry $\frac{1}{2|v_j^{(t)}|}$ at the diagonal, and $\tilde{\mathbf{D}}_2^{(t)}$ is a diagonal matrix with diagonal entry as $\frac{\text{sign}(|v_j^{(t)}| - \tau)}{2|v_j^{(t)}|}$ (see Table 2 footnote).

Then we obtain the alternatively updating rules for $\mathbf{u}$ and $\mathbf{v}$:

$$\begin{aligned}\mathbf{u}^{(t+1)} &= \left(\frac{\lambda_1}{\tau}\mathbf{D}_1^{(t)} - \frac{\lambda_1}{\tau}\tilde{\mathbf{D}}_1^{(t)} + \gamma_1\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}\mathbf{v}^{(t)}, \\ \mathbf{u}^{(t+1)} &= \mathbf{u}^{(t+1)}/||\mathbf{X}\mathbf{u}^{(t+1)}||_2, \\ \mathbf{v}^{(t+1)} &= \left(\frac{\lambda_2}{\tau}\mathbf{D}_2^{(t)} - \frac{\lambda_2}{\tau}\tilde{\mathbf{D}}_2^{(t)} + \gamma_2\mathbf{Y}^\top\mathbf{Y}\right)^{-1}\mathbf{Y}^\top\mathbf{X}\mathbf{u}^{(t+1)}, \\ \mathbf{v}^{(t+1)} &= \mathbf{v}^{(t+1)}/||\mathbf{Y}\mathbf{v}^{(t+1)}||_2,\end{aligned}$$

where $\mathbf{D}_1^{(t)}$ and $\tilde{\mathbf{D}}_1^{(t)}$ are repeatedly constructed from the latest $\mathbf{u}^{(t)}$, and $\mathbf{D}_2^{(t)}$ and $\tilde{\mathbf{D}}_2^{(t)}$ from $\mathbf{v}^{(t)}$. The procedure of the TLP-SCCA algorithm is shown in Algorithm 1. In every iteration, $\mathbf{u}$ and $\mathbf{v}$ are alternately updated till the algorithm converges or arrives at the specific termination condition.

2.3.1 Computational analysis

In Algorithm 1, Step 3 and Step 6 are linear in the dimension of $\mathbf{u}$ and $\mathbf{v}$, and they are easy to calculate. Steps 4 and 7 are the most time-consuming parts of the algorithm. To obtain $\mathbf{u}^{t+1}$ and $\mathbf{v}^{t+1}$, we solve a system of linear equations using the Gauss–Jordan elimination method. This avoids computing the matrix inverse which is computationally much more intensive. Step 5 and 8 are rescale steps and quite simple. Hence the whole algorithm is efficient and runs quickly.

2.4 Extension to truncated group lasso based SCCA

The group lasso pursuits group sparsity by imposing ℓ_2-norm on within-group variables and ℓ_1-norm on between-group variables

Algorithm 1 The TLP-SCCA Algorithm

Require:
$\mathbf{X} \in \mathbb{R}^{n\times p}$, $\mathbf{Y} \in \mathbb{R}^{n\times q}$, pre-tuned $\Lambda = \{\lambda_1, \gamma_1, \lambda_2, \gamma_2\}$, and τ.
Ensure:
Canonical loadings u and v.
1: Initialize $t = 0$, $\mathbf{u} = \mathbf{u}^0 \in \mathbb{R}^{p\times 1}$, $\mathbf{v} = \mathbf{v}^0 \in \mathbb{R}^{q\times 1}$;
2: while not converged do
3: Calculate $\mathbf{D}_1^{(t)}$ and $\tilde{\mathbf{D}}_1^{(t)}$ from the latest $\mathbf{u}^{(t)}$;
4: $\mathbf{u}^{(t+1)} = (\frac{\lambda_1}{\tau}\mathbf{D}_1^{(t)} - \frac{\lambda_1}{\tau}\tilde{\mathbf{D}}_1^{(t)} + \gamma_1\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}\mathbf{v}^{(t)}$;
5: $\mathbf{u}^{(t+1)} = \mathbf{u}^{(t+1)}/||\mathbf{X}\mathbf{u}^{(t+1)}||_2$;
6: Calculate $\mathbf{D}_2^{(t)}$ and $\tilde{\mathbf{D}}_2^{(t)}$ from the latest $\mathbf{v}^{(t)}$;
7: $\mathbf{v}^{(t+1)} = (\frac{\lambda_2}{\tau}\mathbf{D}_2^{(t)} - \frac{\lambda_2}{\tau}\tilde{\mathbf{D}}_2^{(t)} + \gamma_2\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top\mathbf{X}\mathbf{u}^{(t+1)}$;
8: $\mathbf{v}^{(t+1)} = \mathbf{v}^{(t+1)}/||\mathbf{Y}\mathbf{v}^{(t+1)}||_2$;
9: $t \leftarrow t+1$.
10: end while

(Yuan and Lin, 2006). Inspired by this, we extend TLP to the truncated group lasso (TGL):

$$\Omega_{\mathrm{TGL}}(\mathbf{u}) = \sum_{k=1}^{K} J_\tau(||\mathcal{G}_k||), \quad \text{with} \quad J_\tau(||\mathcal{G}_k||) = \min\left(\frac{||\mathcal{G}_k||}{\tau}, 1\right). \tag{12}$$

$\mathcal{G}_k$ is a subset of $\mathbf{u}$, which denotes k-th group, and $\mathbf{u}$ is the concatenation of all $\mathcal{G}_k$'s ($k \in [1, K]$).

Similar to $J_\tau(|u_i|)$, $J_\tau(||\mathcal{G}_k||)$ can be decomposed as

$$J_\tau(||\mathcal{G}_k||) = \frac{1}{2\tau}[f_1(||\mathcal{G}_k||) - f_2(||\mathcal{G}_k||)], \tag{13}$$

where $f_1(||\mathcal{G}_k||) = ||\mathcal{G}_k|| + \tau$, and $f_2(||\mathcal{G}_k||) = |||\mathcal{G}_k|| - \tau|$. Using the subgradient, we define $g_1(\mathcal{G}_k)$ with respect to each group $\mathcal{G}_k$ as

$$g_1(||\mathcal{G}_k||) = \sum_{i\in\mathcal{G}_k} \frac{u_i^2 + \left(u_i^{(t)}\right)^2}{2||\mathcal{G}_k^{(t)}||} + \tau, \tag{14}$$

and

$$g_2(||\mathcal{G}_k||) = \mathrm{sign}\left(||\mathcal{G}_k^{(t)}|| - \tau\right)\left(\frac{u_i^2 + \left(u_i^{(t)}\right)^2}{2||\mathcal{G}_k^{(t)}||} - \tau\right). \tag{15}$$

From Theorem 1, we immediately have the following lemma.

LEMMA 1. *$g_1(||\mathcal{G}_k||)$ is pointwise approximated to $f_1(||\mathcal{G}_k||)$, and $g_2(||\mathcal{G}_k||)$ is pointwise approximated to $f_2(||\mathcal{G}_k||)$.*

The proof is presented in the Supplementary Material. Then according to Equation (7), we define the surrogate penalty of $\Omega_{\mathrm{TGL}}^{\mathrm{App}}(\mathbf{u})$ as:

$$\Omega_{\mathrm{TGL}}^{\mathrm{App}}(\mathbf{u}) = \frac{1}{2\tau}\sum_{k=1}^{K}\sum_{i\in G_k}\left[\frac{u_i^2}{2||G_k^{(t)}||} - \frac{\mathrm{sign}\left(||G_k^{(t)}|| - \tau\right)}{2||G_k^{(t)}||}u_i^2\right] + C, \tag{16}$$

where

$$C = \frac{1}{2\tau}\sum_{k=1}^{K}\sum_{i\in\mathcal{G}_k}\left[\frac{\left(u_i^{(t)}\right)^2}{2||\mathcal{G}_k^{(t)}||} + \tau - \mathrm{sign}\left(||\mathcal{G}_k^{(t)}|| - \tau\right)\left(\frac{\left(u_i^{(t)}\right)^2}{2||\mathcal{G}_k^{(t)}||} - \tau\right)\right] \tag{17}$$

is a constant that does not make contribution towards optimization.

We now propose the TGL-SCCA model

$$\begin{aligned} &\min_{\mathbf{u},\mathbf{v}} -\mathbf{u}^\top\mathbf{X}^\top\mathbf{Y}\mathbf{v} + \lambda_1\Omega_{\mathrm{TGL}}(\mathbf{u}) + \lambda_2\Omega_{\mathrm{TGL}}(\mathbf{v}) \\ &s.t. \quad ||\mathbf{X}\mathbf{u}||^2 \le 1, ||\mathbf{Y}\mathbf{v}||^2 \le 1. \end{aligned} \tag{18}$$

The constrained problem (18) can be reformulated equivalently as the following unconstrained problem

$$\begin{aligned} \mathcal{L}(\mathbf{u},\mathbf{v}) &= -\mathbf{u}^\top\mathbf{X}^\top\mathbf{Y}\mathbf{v} + +\lambda_1\Omega_{\mathrm{TGL}}(\mathbf{u}) + \lambda_2\Omega_{\mathrm{TGL}}(\mathbf{v}) \\ &+\frac{\gamma_1}{2}\left(||\mathbf{X}\mathbf{u}||^2 - 1\right) + \frac{\gamma_2}{2}\left(||\mathbf{Y}\mathbf{v}||^2 - 1\right), \end{aligned} \tag{19}$$

which can be solved by optimizing its pointwise approximation $\mathcal{L}^{\mathrm{App}}(\mathbf{u},\mathbf{v})$,

$$\begin{aligned} \mathcal{L}^{\mathrm{App}}(\mathbf{u},\mathbf{v}) &= -\mathbf{u}^\top\mathbf{X}^\top\mathbf{Y}\mathbf{v} + \lambda_1\Omega_{\mathrm{TLP}}^{\mathrm{App}}(\mathbf{u}) + \lambda_2\Omega_{\mathrm{TLP}}^{\mathrm{App}}(\mathbf{v}) \\ &+\frac{\gamma_1}{2}\left(||\mathbf{X}\mathbf{u}||^2 - 1\right) + \frac{\gamma_2}{2}\left(||\mathbf{Y}\mathbf{v}||^2 - 1\right). \end{aligned} \tag{20}$$

This objective is quadratic, continues and biconvex with respect to $\mathbf{u}$ and $\mathbf{v}$. Thus the updating expressions for $\mathbf{u}$ and $\mathbf{v}$ are obtained:

$$\begin{aligned} \mathbf{u}^{(t+1)} &= \left(\frac{\lambda_1}{\tau}\mathbf{D}_1^{(t)} - \frac{\lambda_1}{\tau}\tilde{\mathbf{D}}_1^{(t)} + \gamma_1\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}\mathbf{v}^{(t)}, \\ \mathbf{u}^{(t+1)} &= \mathbf{u}^{(t+1)}/||\mathbf{X}\mathbf{u}^{(t+1)}||_2, \\ \mathbf{v}^{(t+1)} &= \left(\frac{\lambda_2}{\tau}\mathbf{D}_2^{(t)} - \frac{\lambda_2}{\tau}\tilde{\mathbf{D}}_2^{(t)} + \gamma_2\mathbf{Y}^\top\mathbf{Y}\right)^{-1}\mathbf{Y}^\top\mathbf{X}\mathbf{u}^{(t+1)}, \\ \mathbf{v}^{(t+1)} &= \mathbf{v}^{(t+1)}/||\mathbf{Y}\mathbf{v}^{(t+1)}||_2, \end{aligned}$$

where $\mathbf{D}_1^{(t)}$ is a diagonal matrix with i-th diagonal element being $\frac{1}{2||\mathcal{G}_k^{(t)}||}$ if $u_i \in \mathcal{G}_k (i \in [1, p])$, and $\tilde{\mathbf{D}}_1^{(t)}$ is a diagonal matrix with i-th diagonal element as $\frac{\mathrm{sign}\left(||\mathcal{G}_k^{(t)}|| - \tau\right)}{2||\mathcal{G}_k^{(t)}||}$. $\mathbf{D}_2^{(t)}$ and $\tilde{\mathbf{D}}_2^{(t)}$ are diagonal matrices similar to $\mathbf{D}_1^{(t)}$ and $\tilde{\mathbf{D}}_1^{(t)}$ but deduced from $\mathbf{v}$. Therefore, the solution is obtained by the similar iterating procedure as shown in Algorithm 1.

3 Experiments

3.1 Experimental setup

3.1.1 Benchmarks

We intend to investigate whether TLP-SCCA can indeed improve the performance of the ℓ_1-SCCA methods, and that whether TGL-SCCA outperforms GL-SCCA (group lasso based SCCA) or not. Therefore, we use both ℓ_1-SCCA and GL-SCCA as benchmarks. We are aware of that there are three kinds of ℓ_1-SCCA which are implemented by different optimization techniques. To make a thorough comparison, we employ all three methods in this paper. They are the singular value decomposition based method (Witten *et al.*, 2009), the the excessive gap method which pursuits the optimum by keeping decreasing the dual gap (Chen and Liu, 2012), and the alternative search based method (Du *et al.*, 2014). To facilitate the description, we call the three ℓ_1-SCCA methods L1-SCCA (Witten *et al.*, 2009), L1-NSCCA (Chen and Liu, 2012) and L1-S2CCA (Du *et al.*, 2014). As for group SCCA methods, except S2CCA (group lasso based SCCA) (Du *et al.*, 2014), those GL-SCCA methods assume that $\mathbf{X}^\top\mathbf{X} = \mathbf{Y}^\top\mathbf{Y} = \mathbf{I}$ which is an overly strong assumption. The reason is that in real applications neighboring SNPs are correlated with one another, which implies that both in-set covariance matrices are not identical. Thus we compare TGL-SCCA with S2CCA in this study. For both methods, the group information is provided in advance.

3.1.2 Parameter tuning

There are four parameters, i.e. λ_1, λ_2, γ_1 and γ_2, in both TLP-SCCA and TGL-SCCA algorithms except for the threshold parameter τ. We note that τ is the most important parameter in our models as it balances between the ℓ_0-norm and ℓ_1-norm based penalties. According to the definition, TLP-SCCA reduces to L1-SCCA if given a too large τ, i.e. $\tau > \max_i|u_i|$. On the contrary, a too small τ, i.e. $\tau < \min_i|u_i|$, makes TLP-SCCA reduce to the unpenalized CCA. Therefore, we search the optimal τ from a moderate range, i.e. [0.01, 0.05, $\cdots$, 1] for TLP-SCCA in this paper. In the Supplementary Material section, we will conduct a series of experiments to show the performance trajectories under different τ's. Similarly, TGL-SCCA reduces to the conventional group lasso based SCCA if $\tau > \max_k||\mathcal{G}_k||$; or becomes the group lasso based CCA if $\tau < \min_k||\mathcal{G}_k||$, which will not yield sparse results. Therefore, the τ for TGL-SCCA is tuned from [0.5, 0.55, $\cdots$, 1.5]. Further, we know that the regularization parameters for both TLP and TGL functions are $\frac{\lambda_1}{\tau}$ and $\frac{\lambda_2}{\tau}$. That is, simultaneously changing λ_1 and τ will not affect penalization strength, e.g. $\frac{\lambda_1}{\tau} = \frac{2\lambda_1}{2\tau}$. Therefore, we fix $\lambda_1 = \lambda_2 = 1$ in this paper. Finally, we optimally tune γ_1 and γ_2 by the grid search strategy spanning from 10^{-2} to 10^2. The 5-fold **nested cross-validation** strategy is employed, in which the inner loop pursuits the best parameters and training and testing results are obtained from the outer loop. All methods run on the same partition and the same platform to make a fair comparison.

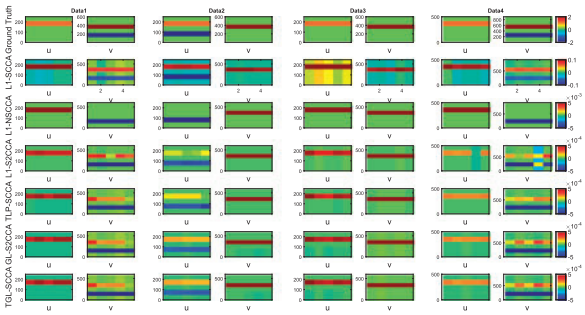

Fig. 1. Canonical loadings estimated on four synthetic datasets. The first column is for Dataset 1, and the second column is for Dataset2, and so forth. For each data, the estimated weight of **u** is shown on the left subfigure, and **v** is on the right. The first row is the ground truth, and each remaining one corresponds to an SCCA method: (i) Ground Truth; (ii) L1-SCCA; (iii) L1-NSCCA; (iv) L1-S2CCA; (v) TLP-SCCA; (vi) GL-S2CCA; (vii) TGL-SCCA

We set the stopping criterion of Algorithm 1 to $\max_i|u_i^{t+1} - u_i^t| \leq \epsilon$ and $\max_j|v_j^{t+1} - v_j^t| \leq \epsilon$, where ϵ is the desirable estimation error bound. ϵ is empirically set to 10^{-5} according to experiments.

3.2 Results on synthetic data

We generate four datasets with different properties to assess the performance of these SCCA methods. The four datasets have different signals and correlation coefficients. We also set the number of observations to be smaller than the number of variables. The datasets are generated as follows. We first set up **u** and **v**, respectively. Then we generate a latent variable **z** from Gaussian distribution $N(0, \mathbf{I}_{n\times n})$. After that, data matrix **X** is generated according to $\mathbf{x}_i \sim N(z_i\mathbf{u}, \sum_x)$ with $(\sum_x)_{jk} = \exp^{-|u_j - u_k|}$. The data matrix **Y** is created similarly by $\mathbf{y}_i \sim N(z_i\mathbf{v}, \sum_y)$, where $\left(\sum_y\right)_{jk} = \exp^{-|v_j - v_k|}$. Now we list the details of the four datasets. (i) The first dataset has 250 variables for **X** and 600 ones for **Y**. Its true correlation coefficient is 0.62. (ii) The second one also has 250 variables and 600 ones for **X** and **Y**, respectively. Its correlation coefficient is 0.84, and its true signal pattern is different. (iii) The third dataset has 250 variables for **X** and 600 variables for **Y**. But its correlation coefficient (0.75) and true signal pattern are different from the previous two datasets. (iv) The fourth dataset has 500 variables for **X** side and 900 for **Y** side. Its correlation coefficients is 0.65. The true signal of each dataset is shown in Figure 1 (top row).

We present the trained canonical loadings of all SCCA methods in Figure 1. Obviously, both TLP-SCCA and L1-SCCA successfully identify the true signals embedded in the data. However, L1-SCCA has more small irrelevant signals than our method. The L1-NSCCA only find out incomplete signals for most **u** and **v**, and L1-S2CCA performs inconsistently across 5-folds. This demonstrates that TLP-SCCA outperforms those benchmarks in identifying accurate canonical loadings. For structured methods, they both perform consistently on all four datasets, showing their ability in extraction of grouping features. In Table 1, we also show the correlation coefficients estimated from both training and testing data, where a higher value stands for a better performance. We use the boldface to highlight the best values. On both training and testing sets, we observe that TLP-SCCA obtains almost all the best correlation coefficients, which means it performs better than those ℓ_1-SCCA methods consistently. In addition, we observe that TGL-SCCA performs similarly to the GL-S2CCA on the training set. However, it outperforms GL-S2CCA on the testing set. This is more desirable since the better testing performance is truly what we want to use to evaluate a learning method. In summary, on these diverse simulation data, for the

Table 1. Performance comparison on synthetic data

	Dataset	data1	data2	data3	data4
Training	L1-SCCA	0.651 ± 0.04	0.831 ± 0.01	0.640 ± 0.05	0.672 ± 0.03
	L1-NSCCA	0.620 ± 0.04	0.805 ± 0.02	0.751 ± 0.04	0.650 ± 0.02
	L1-S2CCA	0.624 ± 0.04	0.830 ± 0.02	0.751 ± 0.04	0.543 ± 0.25
	TLP-SCCA	**0.665 ± 0.04**	0.836 ± 0.02	0.759 ± 0.04	0.683 ± 0.02
	GL-S2CCA	0.662 ± 0.06	**0.847 ± 0.01**	**0.763 ± 0.01**	**0.688 ± 0.01**
	TGL-SCCA	0.662 ± 0.06	0.846 ± 0.01	**0.763 ± 0.01**	**0.688 ± 0.01**
Testing	L1-SCCA	0.638 ± 0.11	0.835 ± 0.04	0.558 ± 0.23	0.666 ± 0.12
	L1-NSCCA	0.615 ± 0.16	0.809 ± 0.05	0.716 ± 0.16	0.664 ± 0.05
	L1-S2CCA	0.617 ± 0.16	0.834 ± 0.05	0.713 ± 0.16	0.552 ± 0.27
	TLP-SCCA	**0.644 ± 0.14**	**0.838 ± 0.04**	0.715 ± 0.16	**0.683 ± 0.07**
	GL-S2CCA	0.622 ± 0.20	0.833 ± 0.05	0.746 ± 0.05	0.678 ± 0.06
	TGL-SCCA	0.622 ± 0.20	0.834 ± 0.05	**0.748 ± 0.05**	0.680 ± 0.06

Note: Training and testing correlation coefficients (mean ± std) of 5-fold cross-validation are shown. The best values are shown in boldface.

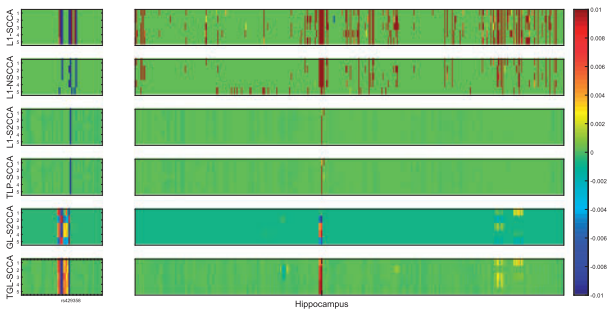

Fig. 2. Canonical loadings estimated on real imaging genetics data. Each row corresponds to an SCCA method: (i) L1-SCCA; (ii) L1-NSCCA; (iii) L1-S2CCA; (iv) TLP-SCCA; (v) GL-S2CCA; (vi) TGL-SCCA. For each method, the estimated weights of **u** are shown on the left panel, and those of **v** are on the right one

individual sparsity, TLP-SCCA performs better than or equal to the ℓ_1-norm based SCCA methods on not only the correlation coefficients but also the canonical loadings profiles. For the group sparsity, TGL-SCCA and GL-S2CCA hold similar canonical loading profiles, but TGL-SCCA obtains better testing canonical correlation coefficients.

3.3 Results on real neuroimaging genetics data

We also compared TLP-SCCA with three ℓ_1-SCCA methods using real neuroimaging and genetics (SNP) data. The real brain imaging genetics data used in the preparation of this article were obtained from the Alzheime's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. One crucial goal of the ADNI is to test if 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 AD. For up-to-date information, please refer to www.adni-info.org.

The ADNI cohort studied in this work contained 176 AD, 363 MCI and 304 healthy control (HC) non-Hispanic Caucasian participants. Structural MRI scans had been processed with voxel-based morphometry (VBM) in SPM. Generally, all scans had been aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) maps, normalized to MNI space and had been smoothed with an 8 mm FWHM kernel. In the experiment setup we subsampled the whole brain, and aimed to examine associations between the voxels (GM density measures) and SNPs, thus we did not merely use ROI summary statistics. On the account of this point, we could obtain more detailed results. Totally 465 voxels spanning all brain ROIs were extracted as imaging data. In addition, the impacts of the baseline age, gender, education and handedness had been also eliminated through regression analysis. For the genotyping data, we used 58 SNP biomarkers from the AD-related genes such *APOE*, *TOMM40* and *PVRL2*, in which *APOE* e4 SNP rs429358 is the best-known AD genetic risk factor (Ramanan *et al.*, 2014). We intend to evaluate the correlations between the voxels data and the *APOE* SNP data using the SCCA methods.

We first show the heat maps of the canonical loadings estimated from the training set in Figure 2. In this figure, each row corresponds to a participant method. The estimated canonical loading **u**, indicating weights for SNPs, is shown on the left panel and the estimated **v** containing weights for the imaging markers is shown on the right. We observe that TLP-SCCA shows a very clear picture with

Table 2. Performance comparison on real data

			Training			Mean ± Std.
L1-SCCA	0.24	0.28	0.25	0.24	0.24	0.25 ± 0.02
L1-NSCCA	0.15	0.30	0.14	0.26	0.27	0.23 ± 0.07
L1-S2CCA	0.25	0.29	0.26	0.27	0.30	0.27 ± 0.02
TLP-SCCA	0.26	0.30	0.26	0.27	0.28	**0.28 ± 0.02**
GL-S2CCA	0.18	0.25	0.25	0.26	0.24	0.23 ± 0.03
TGL-SCCA	0.22	0.26	0.24	0.24	0.27	0.24 ± 0.02
			Testing			Mean ± Std.
L1-SCCA	0.24	0.09	0.22	0.21	0.24	0.20 ± 0.06
L1-NSCCA	0.10	0.10	0.00	0.25	0.22	0.13 ± 0.10
L1-S2CCA	0.36	0.19	0.33	0.29	0.17	0.27 ± 0.09
TLP-SCCA	0.34	0.16	0.32	0.27	0.27	**0.27 ± 0.07**
GL-S2CCA	0.18	0.22	0.18	0.16	0.18	0.18 ± 0.02
TGL-SCCA	0.25	0.14	0.21	0.21	0.27	0.22 ± 0.05

Note: Training and testing correlation coefficients (each fold and Mean ± Std.) of 5-fold cross-validation are shown. The best Mean ± Std. is shown in boldface.

The i-th element of $\mathbf{D}_1^{(t)}$ does not exist if $u_i^{(t)} = 0$. We regularize it as $\frac{1}{2\sqrt{|u_i^{(t)}|^2+\zeta}}$ with ζ being a very small positive number. Then the objective function about **u** conveys to $\ell(\mathbf{u}) = -\mathbf{u}^\top \mathbf{X}^\top \mathbf{Y}\mathbf{v} + \frac{\gamma_1}{2}||\mathbf{X}\mathbf{u}||^2 + \sum^{i=1p} \frac{\lambda_1}{2\tau}\left[\sqrt{u_i^2+\zeta} + \tau - |\sqrt{u_i^2+\zeta} - \tau|\right]$. $\ell(\mathbf{u})$ will reduce to the original problem (10) if ζ approaches zero. We can use the same strategy for constructing $\mathbf{D}_2^{(t)}$ when $v_j^{(t)} = 0$.

respect to both canonical loadings. It only highlights the *APOE* e4 SNP rs429358 on the genetic side. This locus is well known to be related to AD. On the imaging side, TLP-SCCA only captures one voxel out of 465 ones. The highlight region is the hippocampus which is highly correlated with AD (Hampel *et al.*, 2008). However, the benchmark methods report many signals which may misguide further investigation. We can also observe that the performances of these competing methods are not stable as they perform differently on the 5-fold results. While TLP-SCCA has more stable and consist results than those ℓ_1-SCCA methods.

In Figure 2, the last two rows, which shows the canonical weights from two structured SCCA methods, clearly exhibit a group structure on both genetic side and imaging side. They both identify all the SNPs in the 7th group for the genotypic data, showing their relevance to the AD. This group of loci are rs283814 (*PVRL2*), rs157580 (*TOMM40*), rs2075650 (*TOMM40*), rs157582 (*TOMM40*), rs8106922 (*TOMM40*), rs1160985 (*TOMM40*), rs405697 (*TOMM40*), rs405509 (*APOE*) and rs769449 (*APOE*). They all locate in the *APOE* region (i.e. the best known AD risk region) and most of them have been independently reported to be associated with Alzheimer's disease or human longevity (Abraham *et al.*, 2008; Denny *et al.*, 2013; Harold *et al.*, 2009; Kalra *et al.*, 2008; Kamboh *et al.*, 2012; Ma *et al.*, 2013; Soerensen *et al.*, 2013). Strangely, both group methods do not report the locus rs429358 as a risk factor. One possible reason could be that all the nine loci in the 7th group are associated with AD, and their combined effects might be much higher than that of the 8th group where the rs429358 is located. Since both structured methods are sparse at the group level, they only consider the 7th group of loci with the highest influence as risk factors. Now we investigate the imaging markers. TGL-SCCA and GL-S2CCA find out the group of signals from the hippocampus, in which atrophy is an evidence of AD (Hampel *et al.*, 2008). Besides, they identify small signals from the frontal region,

the inferior temporal gyrus and the precentral gyrus, and all of these three regions have been reported to be associated with AD (Chan *et al.*, 2001; Mattsson *et al.*, 2017; Stuss *et al.*, 1992; Woodward *et al.*, 2010). To show the canonical loadings clearly, we map the averaged canonical loadings regarding the imaging measurements onto the brain. Supplementary Figure S3 clearly shows that L1-SCCA highlights most parts of the whole brain followed by L1-NSCCA, while TLP-SCCA only highlights a small region. Both GL-SCCA and TGL-SCCA highlight more voxels than TLP-SCCA due to their group constraint.

Table 2 exhibits both training and testing correlation coefficients where each individual result and its *mean* and *standard deviation* are shown. Obviously, TLP-SCCA yields the best $mean \pm std$ on both training and testing. Moreover, TLP-SCCA is more stable than L1-S2CCA and L1-NSCCA. Interestingly, both group SCCA methods obtain smaller but more stable than those ℓ_1-norm and TLP based methods. And, TGL-SCCA reveals higher correlation than the GL-S2CCA, indicating its improved performance. In summary, TLP-SCCA accurately reveals a biologically meaningful correlation between *APOE* SNP rs429358 and the GM density of hippocampus, an AD related brain region, and TGL-SCCA identify group associations between genetic loci and imaging markers. Therefore, SCCA methods using the truncated lasso or group lasso have better detection ability than those with lasso and group lasso.

4 Conclusion

In this paper, we have proposed a robust and efficient SCCA method (TLP-SCCA) which is flexible and can avoid time consuming parameter tuning. The proposed method employs the truncated ℓ_1-norm other than the ℓ_1-norm, and could reduce the estimation bias. We have proposed an effective algorithm using the alternate convex search (ACS) strategy. The TLP-SCCA algorithm is guaranteed to converge, and we have analyzed its computation complexity. We also have extended it to the truncated group lasso based SCCA (TGL-SCCA) to improve the existing group SCCA methods.

We compared TLP-SCCA, with three ℓ_1-SCCA with different optimization techniques, and compared TGL-SCCA with the GL-SCCA. The results on four diverse synthetic datasets revealed that TLP-SCCA obtained comparable correlation coefficients on the training sets and higher correlation coefficients on the testing sets. Besides, TLP-SCCA identified clearer canonical loadings than three ℓ_1-SCCA methods, and accurately recovered the true signals embedded in the data. The TGL-SCCA obtained better results than the GL-SCCA in terms of both correlation coefficients and canonical weights. The result on the real imaging genetic data showed that TLP-SCCA generated higher correlation coefficients on both training and testing sets. It also had the best canonical weights because the patterns are very clear and clean. That is, TLP-SCCA discovered a biologically meaningful bi-multivariate association between the hippocampus region and the *APOE* e4 genetic locus. As for the group methods, our TGL-SCCA outperformed GL-SCCA on both canonical coefficient and the canonical weights patterns. Besides individual and group sparsity, the network structure is also of great interest and importance in imaging genetics. Therefore, in the future work, we will impose related structure penalty into our model to accommodate this issue.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Bio-gen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Funding

This research was supported by the National Natural Science Foundation of China [61602384]; the Natural Science Basic Research Plan in Shaanxi Province of China [2017JQ6001]; the China Postdoctoral Science Foundation [2017M613202]; and the Fundamental Research Funds for the Central Universities [3102016OQD0065] at Northwestern Polytechnical University. This work was also supported by the National Institutes of Health [R01 EB022574, R01 LM011360, U01 AG024904, RC2 AG036535, R01 AG19771, P30 AG10133, UL1 TR001108, R01 AG 042437, R01 AG046171]; the Department of Defense [W81XWH-14-2-0151, W81XWH-13-1-0259, W81XWH-12-2-0012]; the National Collegiate Athletic Association [14132004]; and the CTSI SPARC Program at Indiana University.

Conflict of Interest: none declared.

References

Abraham,R. *et al.* (2008) A genome-wide association study for late-onset Alzheimer's disease using dna pooling. *BMC Med. Genomics*, **1**, 44.

Chan,D. *et al.* (2001) Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease. *Ann. Neurol.*, **49**, 433–442.

Chen,J. *et al.* (2013) Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis. *Biostatistics*, **14**, 244–258.

Chen,X., and Liu,H. (2012) An efficient optimization algorithm for structured sparse cca, with applications to EQTL mapping. *Stat. Biosci.*, **4**, 3–26.

Chen,X. *et al.* (2012). Structured sparse canonical correlation analysis. In *International Conference on Artificial Intelligence and Statistics*, pp. 199–207. JMLR Proceedings.

Denny,J.C. *et al.* (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.*, **31**, 1102–1111.

Du,L. *et al.* (2014). A novel structure-aware sparse learning algorithm for brain imaging genetics. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2014, pp. 329–336. Springer, Lecture Notes in Computer Science 8675.

Du,L. *et al.* (2016a). Sparse canonical correlation analysis via truncated ℓ1-norm--norm with application to brain imaging genetics. In *IEEE International Conference on Bioinformatics and Biomedicine*, pp. 707–711. IEEE.

Du,L. *et al.* (2016b) Structured sparse canonical correlation analysis for brain imaging genetics: An improved graphnet method. *Bioinformatics*, **32**, 1544–1551.

Fan,J., and Li,R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.*, **96**, 1348–1360.

Fung,G., and Mangasarian,O. (2011) Equivalence of minimal ℓ_0-and ℓp-norm solutions of linear equalities, inequalities and linear programs for sufficiently small p. *J. Optim. Theory Appl.*, **151**, 1–10.

Hampel,H. *et al.* (2008) Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimers Dement.*, **4**, 38–48.

Harold,D. *et al.* (2009) Genome-wide association study identifies variants at clu and picalm associated with Alzheimer's disease. *Nat. Genet.*, **41**, 1088–1093.

Hotelling,H. (1936) Relations between two sets of variates. *Biometrika*, **28**, 321–377.

Kalra,M. *et al.* (2008) Association of apoe genetic variants with obstructive sleep apnea in children. *Sleep Med.*, **9**, 260–265.

Kamboh,M. *et al.* (2012) Genome-wide association study of Alzheimer's disease. *Transl. Psychiatry*, **2**, e117.

Kim,S. *et al.* (2013) Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLoS One*, **8**, e70269.

Ma,X.-Y. *et al.* (2013) Association of tomm40 polymorphisms with late-onset alzheimers disease in a northern han chinese population. *Neuromol. Med.*, **15**, 279–287.

Mattsson,N. *et al.* (2017) 18f-av-1451 and csf t-tau and p-tau as biomarkers in Alzheimer's disease. *EMBO Mol. Med.*, **9**, 1212.

Parkhomenko,E. *et al.* (2009) Sparse canonical correlation analysis with application to genomic data integration. *Stat. Appl. Genet. Mol. Biol.*, **8**, 1–34.

Potkin,S.G. *et al.* (2009) Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cogn. Neuropsychiatry*, **14**, 391–418.

Ramanan,V.K. *et al.* (2014) APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir PET genome-wide association study. *Mol. Psychiatry*, **19**, 351–357.

Saykin,A.J. *et al.* (2015) Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. *Alzheimers Dement.*, **11**, 792–814.

Shen,L. *et al.* (2010) Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage*, **53**, 1051–1063.

Shen,L. *et al.* (2014) Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav.*, **8**, 183–207.

Shen,X. *et al.* (2012) Likelihood-based selection and sharp parameter estimation. *J. Am. Stat. Assoc.*, **107**, 223–232.

Soerensen,M. *et al.* (2013) Evidence from case–control and longitudinal studies supports associations of genetic variation in apoe, cetp, and il6 with human longevity. *Age*, **35**, 487–500.

Stuss,D.T. *et al.* (1992) " no longer gage": frontal lobe dysfunction and emotional changes. *J. Consult. Clin. Psychol.*, **60**, 349.

Witten,D.M., and Tibshirani,R.J. (2009) Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat. Appl. Genet. Mol. Biol.*, **8**, 1–27.

Witten,D.M. *et al.* (2009) A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, **10**, 515–534.

Woodward,M. *et al.* (2010) Differentiating the frontal variant of Alzheimer's disease. *Int. J. Geriatr. Psychiatry*, **25**, 732–738.

Yang,S. *et al.* (2012). Feature grouping and selection over an undirected graph. In *SIGKDD*, pp. 922–930. ACM.

Yuan,M., and Lin,Y. (2006) Model selection and estimation in regression with grouped variables. *J. Roy. Stat. Soc. B*, **68**, 49–67.

Zhang,T. (2013) Multi-stage convex relaxation for feature selection. *Bernoulli*, **19**, 2277–2293.