Appendices

Supplementary Materials for "Structured Sparse Canonical Correlation Analysis for Brain Imaging Genetics: An Improved GraphNet Method" by Lei Du, Heng Huang, Jingwen Yan, Sungeun Kim, Shannon L. Risacher, Mark Inlow, Jason H. Moore, Andrew J. Saykin and Li Shen, for the Alzheimer's Disease Neuroimaging Initiative

A PROOF OF THEOREM 1

THEOREM 1. *Given two datasets* $\mathbf{X}$ *and* $\mathbf{Y}$*, and the pre-tuned parameters* (λ, β, γ)*. Let* $\tilde{\mathbf{u}}$ *be the solution to our SCCA problem of Eq. (10-11). Without loss of generality, we consider the* u_i*-th and* u_j*-th feature are only linked to each other on the graph, i.e.,* $e_{i,j} = 1$*. Let* ρ_{ij} *is the sample correlation between them,* $w_{i,j}$ *is their edge weight. Then the estimated canonical loading* $\mathbf{u}$ *satisfies,*

$$\begin{aligned} |\tilde{u}_i - \tilde{u}_j| &\leq \frac{1}{\gamma_1 + 2\lambda_1 w_{i,j}} \sqrt{2(1-\rho_{ij})}, \;\; \text{if } \rho_{ij} \geq 0, \\ |\tilde{u}_i + \tilde{u}_j| &\leq \frac{1}{\gamma_1 + 2\lambda_1 w_{i,j}} \sqrt{2(1+\rho_{ij})}, \;\; \text{if } \rho_{ij} < 0, \end{aligned} \tag{14}$$

and the estimated canonical loading $\mathbf{v}$ *satisfies,*

$$\begin{aligned} |\tilde{v}_i - \tilde{v}_j| &\leq \frac{1}{(\gamma_2 + 2\lambda_2 w'_{i,j})} \sqrt{2(1-\rho'_{ij})}, \;\; \text{if } \rho'_{ij} \geq 0, \\ |\tilde{v}_i + \tilde{v}_j| &\leq \frac{1}{(\gamma_2 + 2\lambda_2 w'_{i,j})} \sqrt{2(1+\rho'_{ij})}, \;\; \text{if } \rho'_{ij} < 0. \end{aligned} \tag{15}$$

where $w'_{i,j}$ *is the weight between the i-th and j-th feature of* $\mathbf{v}$*, and* ρ'_{ij} *is their sample correlation coefficient.*

PROOF. (1) We first prove the upper bound for $\rho_{ij} \geq 0$, i.e., the u_i-th and u_j-th features are positively correlated. Since $\tilde{\mathbf{u}}$ is the solution to Eq. (10), we have,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial u_i} &= (\lambda_1 \hat{D}_{1,i} + \beta_1 D_{1,i} + \gamma_1 \mathbf{x}_i^T \mathbf{x}_i) \tilde{u}_i = \mathbf{x}_i^T \mathbf{Y} \mathbf{v}, \\ \frac{\partial \mathcal{L}}{\partial u_j} &= (\lambda_1 \hat{D}_{1,j} + \beta_1 D_{1,j} + \gamma_1 \mathbf{x}_j^T \mathbf{x}_j) \tilde{u}_j = \mathbf{x}_j^T \mathbf{Y} \mathbf{v}. \end{aligned}$$

Now that u_i and u_j are the only linked features, we have $D(i,i) = D(j,j) = A(i,j) = w_{i,j}$. We also know that $\text{sgn}(u_i) = \frac{u_i}{|u_i|}$, $||\mathbf{x}_i||_2^2 = \rho_{ii} = 1$ and $||\mathbf{x}_j||_2^2 = \rho_{jj} = 1$. In addition, $\rho_{ij} \geq 0$ implies that $\text{sgn}(u_i) = \text{sgn}(u_j)$. Then according to the definition of $\mathbf{D}_1$ and $\hat{\mathbf{D}}_1$, we arrive at,

$$\begin{aligned} \lambda_1 w_{i,j}(|\tilde{u}_i| - |\tilde{u}_j|)\text{sgn}(\tilde{u}_i) + \beta_1 \text{sgn}(\tilde{u}_i) + \gamma_1 \tilde{u}_i &= \mathbf{x}_i^T \mathbf{Y}\mathbf{v}, \\ \lambda_1 w_{i,j}(|\tilde{u}_j| - |\tilde{u}_i|)\text{sgn}(\tilde{u}_j) + \beta_1 \text{sgn}(\tilde{u}_j) + \gamma_1 \tilde{u}_j &= \mathbf{x}_j^T \mathbf{Y}\mathbf{v}. \end{aligned}$$

i.e.,

$$\begin{aligned} \lambda_1 w_{i,j}(\tilde{u}_i - \tilde{u}_j) + \beta_1 \text{sgn}(\tilde{u}_i) + \gamma_1 \tilde{u}_i &= \mathbf{x}_i^T \mathbf{Y}\mathbf{v}, \\ \lambda_1 w_{i,j}(\tilde{u}_j - \tilde{u}_i) + \beta_1 \text{sgn}(\tilde{u}_i) + \gamma_1 \tilde{u}_j &= \mathbf{x}_j^T \mathbf{Y}\mathbf{v}. \end{aligned} \tag{16}$$

Then we have the following equation by subtracting these two equations,

$$(\gamma_1 + 2\lambda_1 w_{i,j})(\tilde{u}_i - \tilde{u}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{Y}\mathbf{v} \tag{17}$$

Taking ℓ_2-norm on both sides, and using $||(\mathbf{x}_i - \mathbf{x}_j)||_2^2 = 2(1-\rho_{ij})$ can yield,

$$(\gamma_1 + 2\lambda_1 w_{i,j})|\tilde{u}_i - \tilde{u}_j| = \sqrt{2(1-\rho_{ij})}||\mathbf{Y}\mathbf{v}||_2 \tag{18}$$

The constraint function of our model implies $||\mathbf{Y}\mathbf{v}||_2 \leq 1$, thus we arrive at,

$$|\tilde{u}_i - \tilde{u}_j| \leq \frac{1}{(\gamma_1 + 2\lambda_1 w_{i,j})} \sqrt{2(1-\rho_{ij})} \tag{19}$$

(2) We now investigate the upper bound for $\rho_{ij} < 0$, i.e., the u_i-th and u_j-th features are negatively correlated. This implies that $\text{sgn}(u_i) = -\text{sgn}(u_j)$. Thus the Eq. (**??**) becames,

$$\begin{aligned} \lambda_1 w_{i,j}(\tilde{u}_i + \tilde{u}_j) + \beta_1 \text{sgn}(\tilde{u}_i) + \gamma_1 \tilde{u}_i &= \mathbf{x}_i^T \mathbf{Y}\mathbf{v}, \\ \lambda_1 w_{i,j}(\tilde{u}_j + \tilde{u}_i) + \beta_1(-\text{sgn}(\tilde{u}_i)) + \gamma_1 \tilde{u}_j &= \mathbf{x}_j^T \mathbf{Y}\mathbf{v}. \end{aligned} \tag{20}$$

We add the two equations in Eq. (**??**) other than subtracting them,

$$(\gamma_1 + 2\lambda_1 w_{i,j})(\tilde{u}_i + \tilde{u}_j) = (\mathbf{x}_i + \mathbf{x}_j)^T \mathbf{Y}\mathbf{v} \tag{21}$$

Similarly, by taking ℓ_2-norm, and using $||(\mathbf{x}_i - \mathbf{x}_j)||_2^2 = 2(1-\rho_{ij})$ and $||\mathbf{Y}\mathbf{v}||_2 \leq 1$, we arrive at,

$$|\tilde{u}_i + \tilde{u}_j| \leq \frac{1}{(\gamma_1 + 2\lambda_1 w_{i,j})} \sqrt{2(1+\rho_{ij})} \tag{22}$$

which completes the proof.

Since AGN-SCCA is symmetric for $\mathbf{u}$ and $\mathbf{v}$, the proof regarding the upper bound of $\mathbf{v}$ can be obtained via the same strategy.

B PROOF OF THEOREM 2

THEOREM 2. *The problem Eq. (8) is lower bounded by -1.*

PROOF. We can define Lagrange dual function of problem Eq. (8),

$$g(\Gamma) = \min_{\mathbf{u},\mathbf{v}} \mathcal{L}(\mathbf{u}, \mathbf{v}, \Gamma) \tag{23}$$

Since the dual function is the pointwise minimum of affine functions of Γ, it is concave. Obviously, $g(\Gamma) \leq -\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v}$ according to Eq. (6) and Eq. (7).

Now that $g(\Gamma)$ is concave, its maximum exists and the following equations holds when it is maximized.

$$\frac{\partial g(\Gamma)}{\partial \gamma_1} = ||\mathbf{X}\mathbf{u}||_2^2 - 1 = 0, \;\; \frac{\partial g(\Gamma)}{\partial \gamma_2} = ||\mathbf{Y}\mathbf{v}||_2^2 - 1 = 0 \tag{24}$$

This implies that: (1) $-\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} \in [-1, 1]$ because the canonical correlation lies in [-1,1]; and (2) $\max_{\Gamma \succeq 0} g(\Gamma) = -\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v}$.

Now we define the dual problem regarding our primal problem,

$$d^* = \max_{\Gamma \succeq 0} g(\Gamma) = \max_{\Gamma \succeq 0} \min_{\mathbf{u},\mathbf{v}} \mathcal{L}(\mathbf{u}, \mathbf{v}, \Gamma) = -\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} \tag{25}$$

This means $d^* \in [-1, 1]$. According the Lagrange duality, the weak duality always holds for our problem, i.e.,

$$p^* \geq d^* \geq -1 \tag{26}$$

Therefore, the problem Eq. (8) is lower bounded by -1.

C PROOF OF THEOREM 3

To prove the convergence of AGN-SCCA algorithm, we introduce the following lemma from (?).

LEMMA 1. *Any two nonzero vectors* $\tilde{\mathbf{u}}$, $\mathbf{u}$ *with the same length satisfies,*

$$||\tilde{\mathbf{u}}||_2 - \frac{||\tilde{\mathbf{u}}||_2^2}{2||\mathbf{u}||_2} \le ||\mathbf{u}||_2 - \frac{||\mathbf{u}||_2^2}{2||\mathbf{u}||_2}. \quad (27)$$

PROOF. See (?).

Then we have the following lemma.

LEMMA 2. *Any two nonzero numbers* u *and* $\tilde{u}$ *satisfy,*

$$||\tilde{u}||_1 - \frac{||\tilde{u}||_1^2}{2||u||_1} \le ||u||_1 - \frac{||u||_1^2}{2||u||_1}. \quad (28)$$

PROOF. Obviously, for any nonzero real numbers u and $\tilde{u}$, given Lemma **??**, we have $||u||_1 = ||u||_2$, and $||\tilde{u}||_1 = ||\tilde{u}||_2$. This completes the proof.

THEOREM 3. *In each iteration, the AGN-SCCA algorithm monotonously decreases the objective value till it converges.*

PROOF. Now we prove Theorem 3 in two stages.

(1) Stage 1: From Steps 3-6 in Algorithm 1, we fix $\mathbf{v}$ to solve for $\mathbf{u}$. The objective function Eq. (7) is equivalent to,

$$\mathcal{L}(\mathbf{u}, \mathbf{v}) = -\mathbf{u}^T\mathbf{X}^T\mathbf{Y}\mathbf{v} + \frac{1}{2}||\mathbf{u}||_{AGN} + \frac{\gamma_1}{2}||\mathbf{X}\mathbf{u}||_2^2$$

We denote the updated value as $\tilde{\mathbf{u}}$. Then from the Step 5 we have

$$\begin{aligned} &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + \lambda_1\tilde{\mathbf{u}}^\mathbf{T}\hat{\mathbf{D}}_\mathbf{1}\tilde{\mathbf{u}} + \beta_1\tilde{\mathbf{u}}^\mathbf{T}\mathbf{D}_\mathbf{1}\tilde{\mathbf{u}} + \gamma_1\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{X}\tilde{\mathbf{u}} \\ \le &-\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + \lambda_1\mathbf{u}^\mathbf{T}\hat{\mathbf{D}}_\mathbf{1}\mathbf{u} + \beta_1\mathbf{u}^\mathbf{T}\mathbf{D}_\mathbf{1}\mathbf{u} + \gamma_1\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{X}\mathbf{u} \end{aligned}$$

According to the definition of $\hat{\mathbf{D}}_\mathbf{1}$ and $\mathbf{D}_\mathbf{1}$, we arrive at,

$$\begin{aligned} &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + \lambda_1\sum_{k_1} 2\mathbf{L}_1^{k_1}|\mathbf{u}|\frac{||\tilde{u}^{k_1}||_1^2}{2||u^{k_1}||_1} \\ &+\beta_1\sum_{k_1}\frac{||\tilde{u}^{k_1}||_1^2}{2||u^{k_1}||_1} + \gamma_1\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{X}\tilde{\mathbf{u}} \\ \le &-\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + \lambda_1\sum_{k_1} 2\mathbf{L}_1^{k_1}|\mathbf{u}|\frac{||u^{k_1}||_1^2}{2||u^{k_1}||_1} \\ &+\beta_1\sum_{k_1}\frac{||u^{k_1}||_1^2}{2||u^{k_1}||_1} + \gamma_1\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{X}\mathbf{u} \end{aligned} \quad (29)$$

We first multiply $2\lambda_1\mathbf{L}_1^{k_1}|\tilde{\mathbf{u}}|$ on both sides of Eq. (**??**) for each $k_1 \in [1, p]$, and multiply β_1 on both sides of Eq. (**??**). Finally we sum them with Eq. (**??**) on both sides,

$$\begin{aligned} &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + 2\lambda_1|\tilde{\mathbf{u}}|^\mathbf{T}\mathbf{L}_\mathbf{1}|\tilde{\mathbf{u}}| + \beta_1||\tilde{\mathbf{u}}||_1 + \gamma_1||\mathbf{X}\tilde{\mathbf{u}}||_2^2 \\ \le &-\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + 2\lambda_1|\mathbf{u}|^\mathbf{T}\mathbf{L}_\mathbf{1}|\tilde{\mathbf{u}}| + \beta_1||\mathbf{u}||_1 + \gamma_1||\mathbf{X}\mathbf{u}||_2^2 \end{aligned}$$

Again, we first multiply $2\lambda_1\mathbf{L}_1^{k_1}|\mathbf{u}|$ for each $k_1 \in [1, p]$ on both sides of Eq. (**??**), and multiply β_2 on both sides of Eq. (**??**). By summing them together with Eq. (**??**), we arrive at,

$$\begin{aligned} &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + 2\lambda_1|\mathbf{u}|^\mathbf{T}\mathbf{L}_\mathbf{1}|\tilde{\mathbf{u}}| + \beta_1||\tilde{\mathbf{u}}||_1 + \gamma_1||\mathbf{X}\tilde{\mathbf{u}}||_2^2 \\ \le &-\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + 2\lambda_1|\mathbf{u}|^\mathbf{T}\mathbf{L}_\mathbf{1}|\mathbf{u}| + \beta_1||\mathbf{u}||_1 + \gamma_1||\mathbf{X}\mathbf{u}||_2^2 \end{aligned}$$

Obviously, according to the transitive property of inequalities, the following inequality holds,

$$\begin{aligned} &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + 2\lambda_1|\tilde{\mathbf{u}}|^\mathbf{T}\mathbf{L}_\mathbf{1}|\tilde{\mathbf{u}}| + \beta_1||\tilde{\mathbf{u}}||_1 + \gamma_1||\mathbf{X}\tilde{\mathbf{u}}||_2^2 \\ \le &-\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + 2\lambda_1|\mathbf{u}|^\mathbf{T}\mathbf{L}_\mathbf{1}|\mathbf{u}| + \beta_1||\mathbf{u}||_1 + \gamma_1||\mathbf{X}\mathbf{u}||_2^2 \end{aligned}$$

i.e.,

$$\begin{aligned} &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + \frac{1}{2}||\tilde{\mathbf{u}}||_{AGN} + \frac{\gamma_1^*}{2}||\mathbf{X}\tilde{\mathbf{u}}||_2^2 \\ \le &-\mathbf{u}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + \frac{1}{2}||\mathbf{u}||_{AGN} + \frac{\gamma_1^*}{2}||\mathbf{X}\mathbf{u}||_2^2 \end{aligned} \quad (30)$$

where $\lambda_1^* = 4\lambda_1$, $\gamma_1^* = 2\gamma_1$,$\beta_1^* = 2\beta_1$.

Therefore, our algorithm will decrease in each iteration during the this phase, i.e., $\mathcal{L}(\tilde{\mathbf{u}}, \mathbf{v}) \le \mathcal{L}(\mathbf{u}, \mathbf{v})$.

(2) Phase 2: From the Step 7 to Step 10, we fix $\mathbf{u}$ to solve for $\mathbf{v}$. Applying the same steps above, we can arrive at,

$$\begin{aligned} &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\tilde{\mathbf{v}} + \frac{1}{2}||\tilde{\mathbf{v}}||_{AGN} + \frac{\gamma_2^*}{2}||\mathbf{Y}\tilde{\mathbf{v}}||_2^2 \\ \le &-\tilde{\mathbf{u}}^\mathbf{T}\mathbf{X}^\mathbf{T}\mathbf{Y}\mathbf{v} + \frac{1}{2}||\mathbf{v}||_{AGN} + \frac{\gamma_2^*}{2}||\mathbf{Y}\mathbf{v}||_2^2 \end{aligned} \quad (31)$$

Thus our algorithm also decreases in each iteration during the second phase, i.e., $\mathcal{L}(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \le \mathcal{L}(\tilde{\mathbf{u}}, \mathbf{v})$.

Now that $\mathcal{L}(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \le \mathcal{L}(\tilde{\mathbf{u}}, \mathbf{v})$ and $\mathcal{L}(\tilde{\mathbf{u}}, \mathbf{v}) \le \mathcal{L}(\mathbf{u}, \mathbf{v})$, it is obvious to have $\mathcal{L}(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) \le \mathcal{L}(\mathbf{u}, \mathbf{v})$. Therefore, Algorithm 1 monotonically decreases the objective function in each iteration.