Canonical Correlation Analysis on Riemannian Manifolds and its Applications

Hyunwoo J. Kim[1], Nagesh Adluru[1], Barbara B. Bendlin[1], Sterling C. Johnson[1], Baba C. Vemuri[2], Vikas Singh[1]

http://pages.cs.wisc.edu/~hwkim/projects/riem-cca

[1]University of Wisconsin-Madison [2]University of Florida

The motivating problem

- Canonical correlation analysis (CCA) based statistical analysis of images where voxel-wise measurement is manifold valued.
- CCA on the *product Riemannian manifold* representing SPD matrix-valued fields
- Correlations between diffusion tensor images (DTI) and Cauchy deformation tensor fields derived from T1-weighted magnetic resonance (MR) images

What's new?

- Extensions of CCA to nonlinear spaces e.g., KCCA, ANN, Deep CCA (ICML 2013)
- Adaptive CCA (ICML 2012) uses matrix manifold optimization but it's NOT for manifold-valued data
- **This paper : generalization of CCA for Riemannian manifolds**
- **Main technical highlight:** Exact iterative method as well as single path algorithms with approximate projections (inner product and log-Euclidean)
- Relationship between transformations on SPD(n) for more accurate approximation

CCA in Euclidean space

Pearson correlation for $x \in \mathbf{R}$ and $y \in \mathbf{R}$

$$\rho_{x,y} = \frac{\mathrm{COV}(x,y)}{\sigma_x\sigma_y} = \frac{\mathbb{E}[(x-\mu_x)(y-\mu_y)]}{\sigma_x\sigma_y} = \frac{\sum_{i=1}^N (x_i-\mu_x)(y_i-\mu_y)}{\sqrt{\sum_{i=1}^N (x_i-\mu_x)^2}\sqrt{\sum_{i=1}^N (y_i-\mu_y)^2}} \quad (1)$$

Canonical Correlation for $\mathbf{x} \in \mathbf{R}^m$ and $\mathbf{y} \in \mathbf{R}^n$

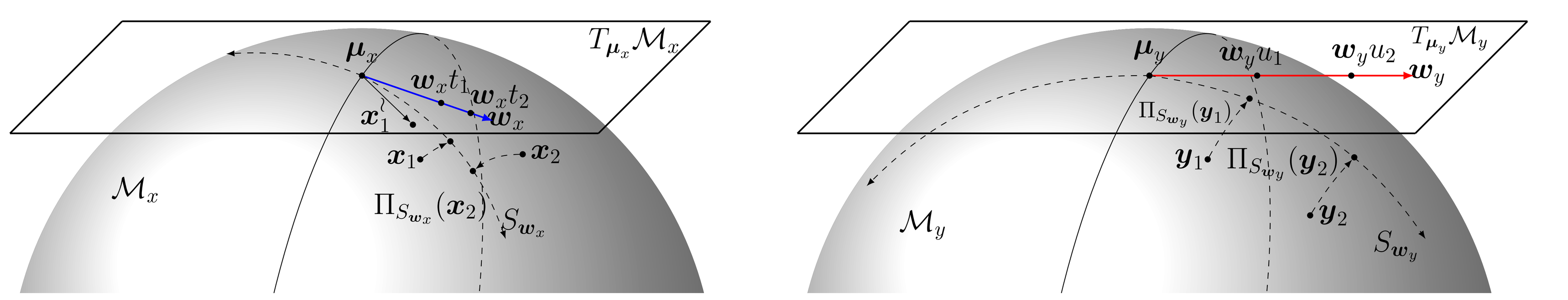
$$\max_{\mathbf{w}_x,\mathbf{w}_y} \mathrm{corr}(\pi_{\mathbf{w}_x}(\mathbf{x}), \pi_{\mathbf{w}_y}(\mathbf{y})) = \max_{\mathbf{w}_x,\mathbf{w}_y} \frac{\sum_{i=1}^N \mathbf{w}_x^T(\mathbf{x}_i-\mu_x)\mathbf{w}_y^T(\mathbf{y}_i-\mu_y)}{\sqrt{\sum_{i=1}^N (\mathbf{w}_x^T(\mathbf{x}_i-\mu_x))^2}\sqrt{\sum_{i=1}^N (\mathbf{w}_y^T(\mathbf{y}_i-\mu_y))^2}} \quad (2)$$

where *projection coefficient* $\pi_{\mathbf{w}_x}(\mathbf{x}) := \arg\min_{t\in\mathbf{R}} \mathrm{d}(t\mathbf{w}_x + \mu_x, \mathbf{x})^2$.

CCA on Manifolds: Basic Operations

Operation	Subtraction	Addition	Distance	Mean	Covariance
Euclidean	$\overrightarrow{x_ix_j} = x_j - x_i$	$x_i + \overrightarrow{x_jx_k}$	$\lVert\overrightarrow{x_ix_j}\rVert$	$\sum_{i=1}^n \overrightarrow{\bar{x}x_i} = 0$	$\mathbb{E}[(x_i-\bar{x})(x_i-\bar{x})^T]$
Riemannian	$\overrightarrow{x_ix_j} = \mathrm{Log}(x_i, x_j)$	$\mathrm{Exp}(x_i, \overrightarrow{x_jx_k})$	$\lVert\mathrm{Log}(x_i,x_j)\rVert_{x_i}$	$\sum_{i=1}^n \mathrm{Log}(\bar{x}, x_i) = 0$	$\mathbb{E}[\mathrm{Log}(\bar{x},x_i)\mathrm{Log}(\bar{x},x_i)^T]$

$$\pi_{\mathbf{w}_x}(\mathbf{x}) := \arg\min_{t\in\mathbb{R}} \mathrm{d}(\mathrm{Exp}(\mu_x, t\mathbf{w}_x), \mathbf{x})^2 \quad (3)$$



CCA on manifolds

Input: $x_1,\dots,x_N \in \mathcal{M}$, $y_1,\dots,y_N \in \mathcal{M}$
Output: $\mathbf{w}_x \in T_{\mu_x}\mathcal{M}$, $\mathbf{w}_y \in T_{\mu_y}\mathcal{M}$ (submanifolds), $t_i, u_i \in \mathbf{R}$ (projection coefficients)

$$\rho_{\mathbf{x},\mathbf{y}} = \max_{\mathbf{w}_x,\mathbf{w}_y,\mathbf{t},\mathbf{u}} \frac{\sum_{i=1}^N (t_i-\bar{t})(u_i-\bar{u})}{\sqrt{\sum_{i=1}^N (t_i-\bar{t})^2}\sqrt{\sum_{i=1}^N (u_i-\bar{u})^2}}$$

$$\text{s.t.} \quad t_i = \arg\min_{t_i\in(-\epsilon,\epsilon)} \lVert\mathrm{Log}(\mathrm{Exp}(\mu_x, t_i\mathbf{w}_x), \mathbf{x}_i)\rVert^2, \forall i \in \{1,\dots,N\}$$
$$u_i = \arg\min_{u_i\in(-\epsilon,\epsilon)} \lVert\mathrm{Log}(\mathrm{Exp}(\mu_y, u_i\mathbf{w}_y), \mathbf{y}_i)\rVert^2, \forall i \in \{1,\dots,N\} \quad (4)$$

Formulation with first order optimality conditions

$$\rho(\mathbf{w}_x,\mathbf{w}_y) = \max_{\mathbf{w}_x,\mathbf{w}_y,\mathbf{t},\mathbf{u}} f(\mathbf{t},\mathbf{u}) \text{ s.t. } \nabla_{t_i} g(t_i,\mathbf{w}_x) = 0, \nabla_{u_i} g(u_i,\mathbf{w}_y) = 0, \forall i \in \{1,\dots,N\} \quad (5)$$

$$f(\mathbf{t},\mathbf{u}) := \frac{\sum_{i=1}^N (t_i-\bar{t})(u_i-\bar{u})}{\sqrt{\sum_{i=1}^N (t_i-\bar{t})^2}\sqrt{\sum_{i=1}^N (u_i-\bar{u})^2}}$$

$$g(t_i,\mathbf{w}_x) := \lVert\mathrm{Log}(\mathrm{Exp}(\mu_x, t_i\mathbf{w}_x), \mathbf{x}_i)\rVert^2, \; g(u_i,\mathbf{w}_y) := \lVert\mathrm{Log}(\mathrm{Exp}(\mu_y, u_i\mathbf{w}_y), \mathbf{y}_i)\rVert^2$$

Objective function for CCA on Manifolds

Given a constrained optimization problem $\max f(\mathbf{x})$ s.t. $c_i(\mathbf{x}) = 0, \forall i$, the **augmented Lagrangian method (ALM)** solves a sequence of the following models increasing ν_k.

$$\max f(\mathbf{x}) + \sum_i \lambda_i c_i(\mathbf{x}) - \nu^k \sum_i c_i(\mathbf{x})^2 \quad (6)$$

The augmented Lagrangian formulation for our CCA formulation (5) is

$$\max_{\mathbf{w}_x,\mathbf{w}_y,\mathbf{t},\mathbf{u}} \mathcal{L}_A(\mathbf{w}_x,\mathbf{w}_y,\mathbf{t},\mathbf{u},\lambda^k;\nu^k) = \max_{\mathbf{w}_x,\mathbf{w}_y,\mathbf{t},\mathbf{u}} f(\mathbf{t},\mathbf{u}) + \sum_i^N \lambda_{t_i}^k \nabla_{t_i} g(t_i,\mathbf{w}_x) + \sum_i^N \lambda_{u_i}^k \nabla_{u_i} g(u_i,\mathbf{w}_y) - \frac{\nu^k}{2}\left(\sum_{i=1}^N \nabla_{t_i} g(t_i,\mathbf{w}_x)^2 + \nabla_{u_i} g(u_i,\mathbf{w}_y)^2\right) \quad (7)$$

Iterative algorithm by augmented Lagrangian method

$\mathbf{x}_1,\dots,\mathbf{x}_N \in \mathcal{M}_x$, $\mathbf{y}_1,\dots,\mathbf{y}_N \in \mathcal{M}_y$
Given $\nu^0 > 0$, $\tau^0 > 0$, starting points $(\mathbf{w}_x^0, \mathbf{w}_y^0, \mathbf{t}^0, \mathbf{u}^0)$ and λ^0
for $k = 0, 1, 2\dots$ **do**
 Start at $(\mathbf{w}_x^k, \mathbf{w}_y^k, \mathbf{t}^k, \mathbf{u}^k)$
 Find an approximate minimizer $(\mathbf{w}_x^k, \mathbf{w}_y^k, \mathbf{t}^k, \mathbf{u}^k)$ of $\mathcal{L}_A(\cdot, \lambda^k; \nu^k)$, and terminate when $\lVert\nabla\mathcal{L}_A(\mathbf{w}_x^k, \mathbf{w}_y^k, \mathbf{t}^k, \mathbf{u}^k, \lambda^k; \nu^k)\rVert \le \tau^k$
 if a convergence test is satisfied **then**
  Stop with approximate feasible solution
 end if
 $\lambda_{t_i}^{k+1} = \lambda_{t_i}^k - \nu^k \nabla_{t_i} g(t_i, \mathbf{w}_x), \forall i$ and $\lambda_{u_i}^{k+1} = \lambda_{u_i}^k - \nu^k \nabla_{u_i} g(u_i, \mathbf{w}_y), \forall i$
 Choose new penalty parameter $\nu^{k+1} \ge \nu^k$
 Set starting point for the next iteration
 Select tolerance τ^{k+1}
end for

Single path algorithm with approximate projection

$$\Pi_S(\mathbf{x}) \approx \mathrm{Exp}(\mu, \sum_{i=1}^d \mathbf{v}_i \langle \mathbf{v}_i, \mathrm{Log}(\mu, \mathbf{x})\rangle_\mu) \quad (8)$$

Input $X_1,\dots,X_N \in \mathcal{M}_y$, $Y_1,\dots,Y_N \in M_y$
Compute intrinsic mean μ_x, μ_y of $\{X_i\}, \{Y_i\}$
Compute $X_i' = \mathrm{Log}(\mu_x, X_i)$, $Y_i' = \mathrm{Log}(\mu_y, Y_i)$
Transform (using **group action**) $\{X_i'\}, \{Y_i'\}$ to the $T_I\mathcal{M}_x$, $T_I\mathcal{M}_y$
Perform CCA between $T_I\mathcal{M}_x$, $T_I\mathcal{M}_y$ and get axes $W_a \in T_I\mathcal{M}_x$, $W_b \in T_I\mathcal{M}_y$
Transform (using **group action**) W_a, W_b to $T_{\mu_x}\mathcal{M}_x$, $T_{\mu_y}\mathcal{M}_y$

Why group action is needed?

- Transformations to the Identity of SPD(n) for more accurate projection
- Group action is equivalent to the parallel transport from $T_p\mathcal{M}$ to $T_I\mathcal{M}$ on SPD(n)
- Group action is computationally more efficient

Theorem

On SPD manifold, let $\Gamma_{p\to I}(w)$ denote the parallel transport of $w \in T_p\mathcal{M}$ along the geodesic from $p \in \mathcal{M}$ to $I \in \mathcal{M}$. The parallel transport is equivalent to group action by $p^{-1/2}wp^{-T/2}$, where the inner product $\langle u, v\rangle_p = \mathrm{tr}(p^{-1/2}up^{-1}vp^{-1/2})$.

Synthetic Experiments: Riemannian vs. Euclidean CCA

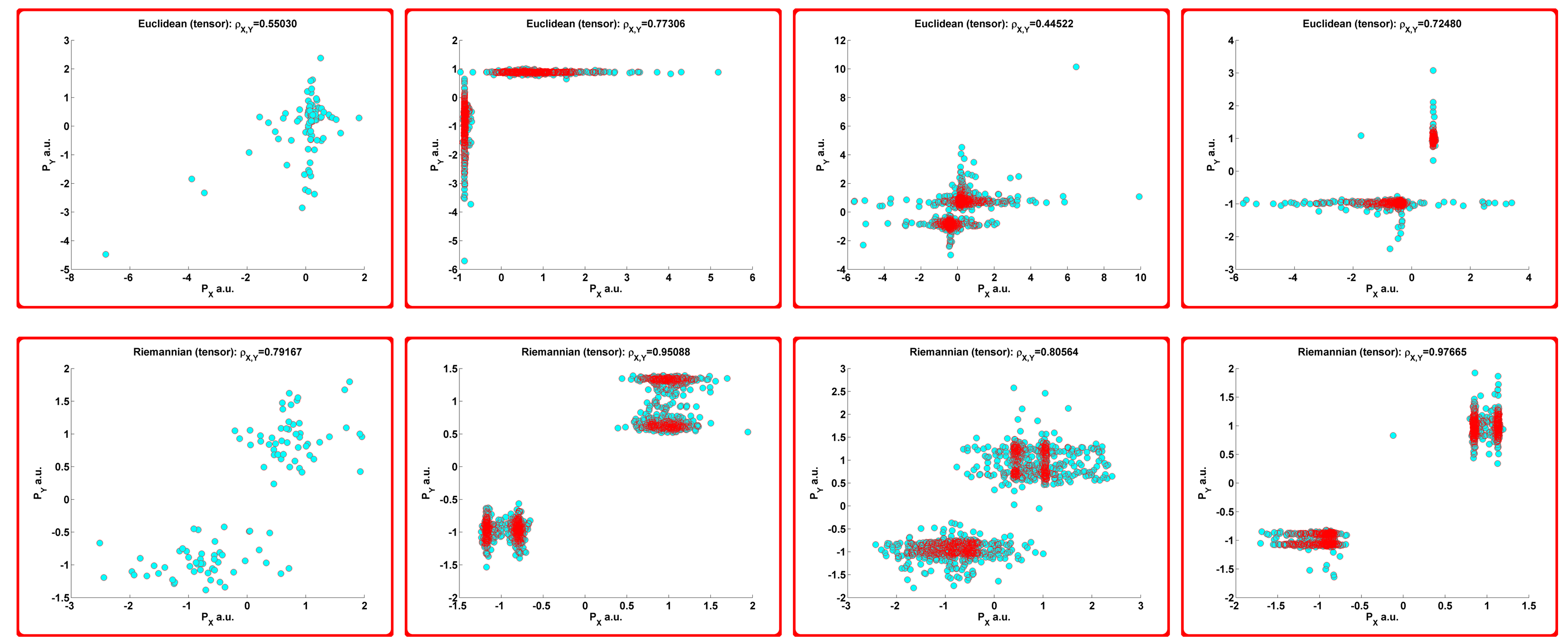


Figure : Improvements in $\rho_{x,y}$ using Riemannian CCA. Each column represents a synthetic experiment with a specific set of $\{\mu_{x_j}, \epsilon_{x_j}; \mu_{y_j}, \epsilon_{y_j}\}$. The first column has 100 samples while the other columns have 1000 samples.

Neuroimaging Experiments: CCA for Multi-modal Analysis

Task: Detect gender and age effects on white and gray matter interactions. The white matter is characterized by diffusion tensors (Y_{DTI}) while the gray matter by Cauchy deformation tensors ($X_{\mathrm{T1W}} = \sqrt{J^TJ}$).

$$Y_{\mathrm{DTI}} = \beta_0 + \beta_1 \mathrm{Gender} + \beta_2 X_{\mathrm{T1W}} + \beta_3 X_{\mathrm{T1W}} \cdot \mathrm{Gender} + \varepsilon,$$
$$Y_{\mathrm{DTI}} = \beta_0' + \beta_1' \mathrm{AgeGroup} + \beta_2' X_{\mathrm{T1W}} + \beta_3' X_{\mathrm{T1W}} \cdot \mathrm{AgeGroup} + \varepsilon,$$

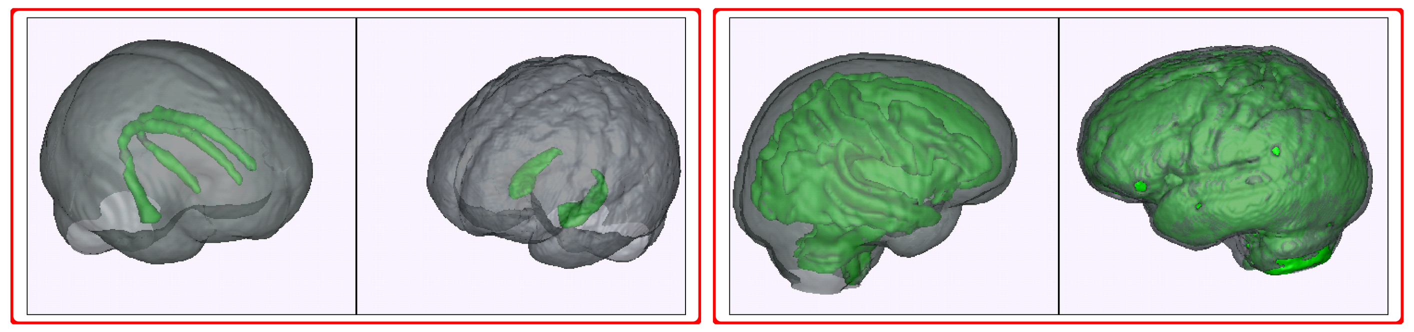


Figure : Regions of interest in which the interaction models are tested. From left to right, cingulum bundle (DTI), hippocampus (T1W), entire white matter (DTI) and entire gray matter (T1W).

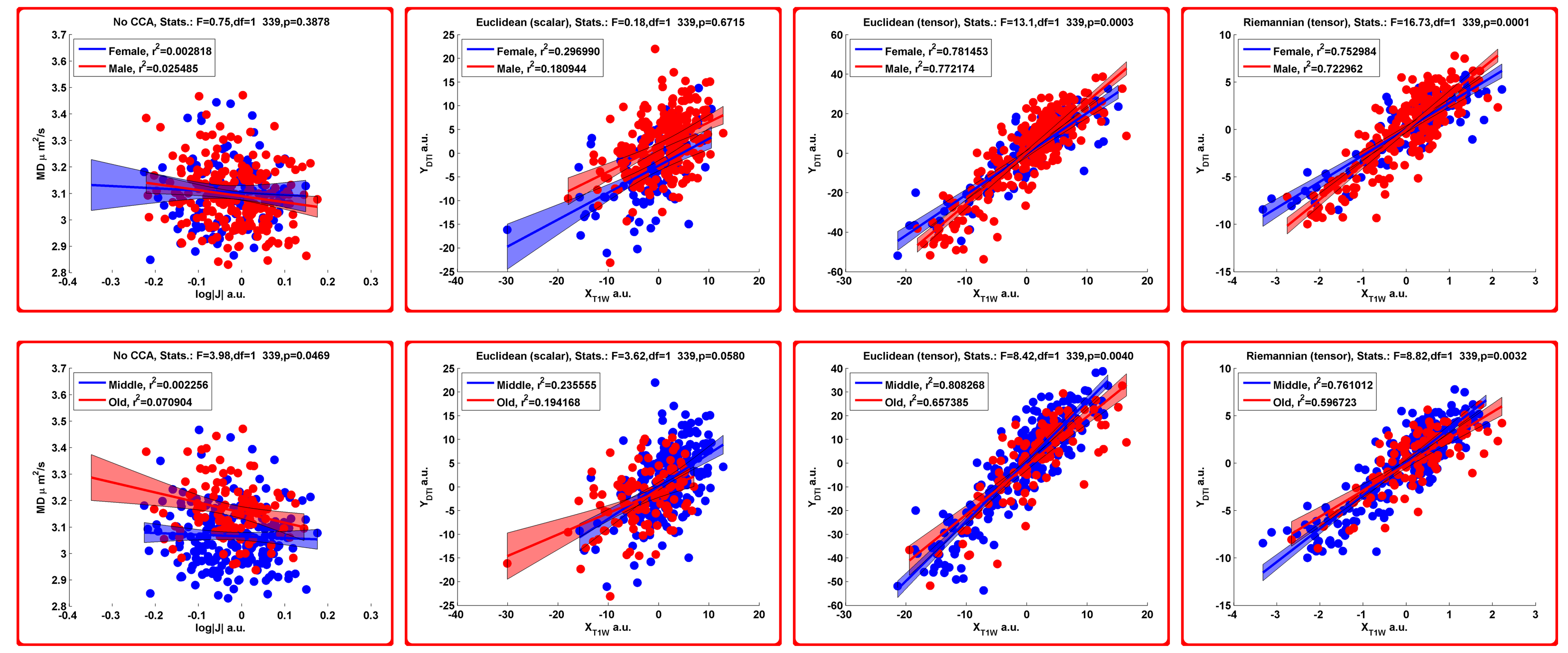


Figure : Riemannian CCA weight vectors for DTI ($\mathbf{w}_y$) and T1W ($\mathbf{w}_x$) when using the cingulum (left) and hippocampal (right) ROIs. Bottom row shows the same when using entire white and gray matter ROIs.

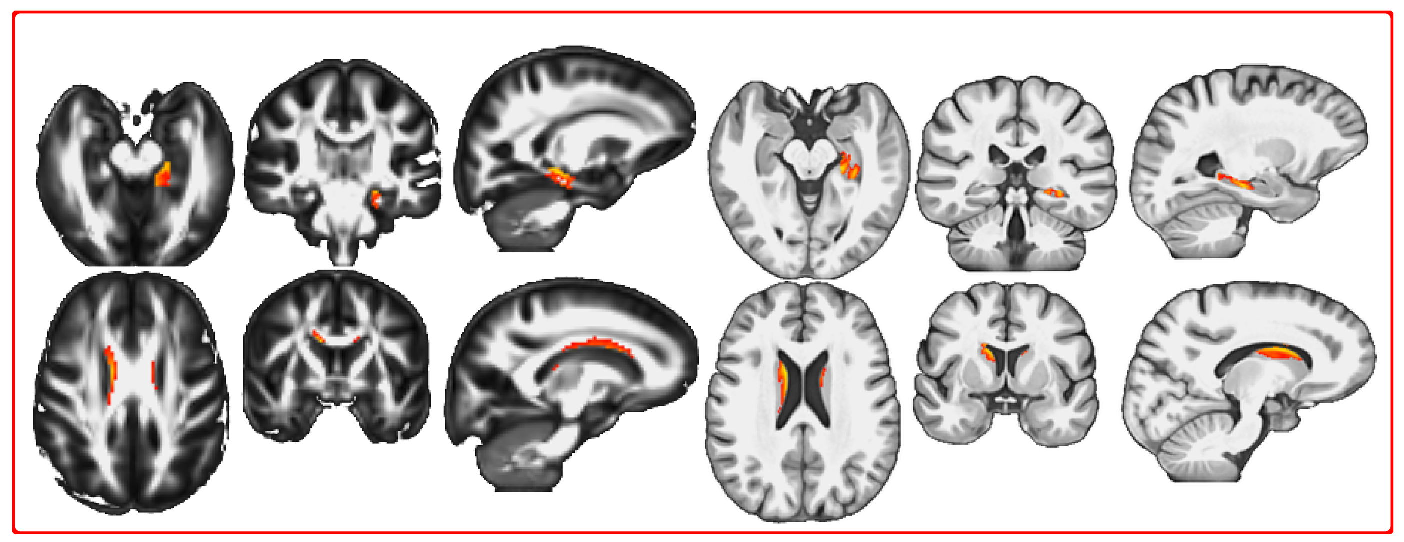


Figure : Riemannian CCA weight vectors for DTI ($\mathbf{w}_y$) and T1W ($\mathbf{w}_x$) when using the cingulum (left) and hippocampal (right) ROIs. Bottom row shows the same when using entire white and gray matter ROIs.

European Conference on Computer Vision (ECCV) 2014 Research supported in part by NIH R01 AG040396, NIH R01 AG037639, NSF CAREER award 1252725, NSF RI 1116584, Waisman Center CIHM, Wisconsin ADRC.