Multivariate General Linear Models (MGLM) on Riemannian Manifolds

Hyunwoo J. Kim, Nagesh Adluru, Maxwell D. Collins, Moo K. Chung, Barbara B. Bendlin, Sterling C. Johnson, Richard J. Davidson, Vikas Singh

THE UNIVERSITY of WISCONSIN MADISON

http://pages.cs.wisc.edu/∼hwkim/projects/riem-mglm/

The Motivating Problem

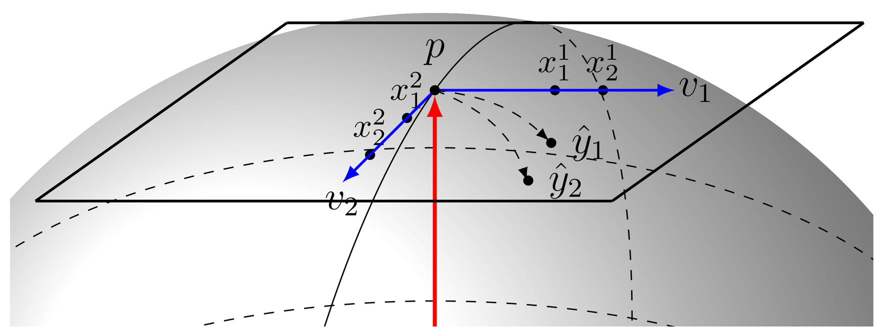
- General linear model (GLM) based statistical analysis of images where voxel-wise measurement is manifold valued.
- Identify how the relationships of voxels **differ** across clinically disparate groups, *while controlling for one or more nuisance variables*.

What is new?

- Fletcher (IJCV, 2013) provides methods for geodesic regression for $f : \mathbf{R} \to \mathcal{M}$.
- Extended for regressing $f : \mathbf{R} \to S^\infty$ by Du et al. (NeuroImage, 2014).
- This paper: multivariate multiple linear regression on manifold data, $f : \mathbf{R}^n \to \mathcal{M}$.
- The same general recipe works for both cases: $y \in \mathrm{SPD}(3)$ for diffusion tensor images (DTI) and $y \in S^\infty$ for orientation distribution functions (ODFs).
- **Main technical highlight:** Exact variational framework for $\mathbf{R}^n$ where $n > 1$ as well as a faster (but approximate) Log-Euclidean framework for optimization.

Regression on Manifolds: Basic Operations

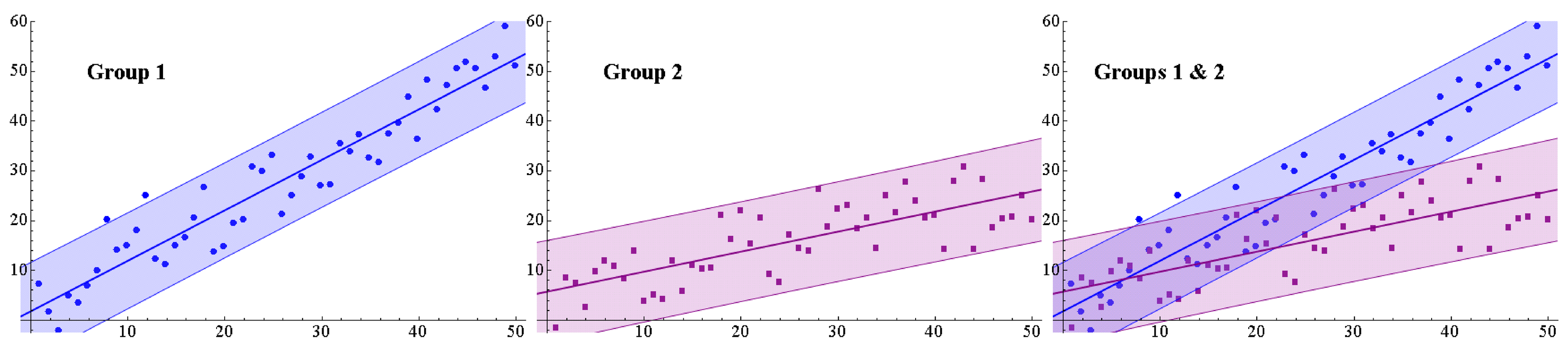
Operation	Subtraction	Addition	Distance	Mean	Covariance
Euclidean	$\overrightarrow{x_i x_j} = x_j - x_i$	$x_i + \overrightarrow{x_j x_k}$	$\lVert\overrightarrow{x_i x_j}\rVert$	$\sum_{i=1}^n \overrightarrow{\bar{x} x_i} = 0$	$\mathbb{E}\left[(x_i - \bar{x})(x_i - \bar{x})^T\right]$
Riemannian	$\overrightarrow{x_i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



v^1, v^2 are tangent vectors. Each entry of independent variables $(x^1, x^2) \in \mathbf{R}^2$, is multiplied by v_1 and v_2 respectively in $T_p\mathcal{M}$. Here, x_i^j denotes j-th entry of the i-th instance.

Multivariate General Linear Models (MGLM)

High level goal: Identify relationship of response variables with covariates; assess statistical significance of differences of regression coefficients across disparate groups.



Euclidean and Riemannian MGLMs

Euclidean

$$y = \alpha + \beta^1 x^1 + \beta^2 x^2 + \ldots + \beta^n x^n + \epsilon,$$

where $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$ and $\epsilon \sim \frac{1}{\sqrt{(2\pi)^m|\Sigma|}} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)\right)$.

Riemannian

$$y = \mathrm{Exp}\left(\mathrm{Exp}\left(p, v^1 x^1 + v^2 x^2 + \ldots + v^n x^n\right), \epsilon\right),$$

where $\epsilon \sim \frac{1}{Z(\mu,\sigma)} \exp\left(-\frac{d(y,\mu)^2}{2\sigma^2}\right)$ and $Z(\mu,\sigma) = \int_{\mathcal{M}} \exp\left(-\frac{d(y,\mu)^2}{2\sigma^2}\right)$ is the normalization factor.

Objective Function for MGLMs on Manifolds

Input: $x_1, \ldots, x_N \in \mathbf{R}^n$, $y_1, \ldots, y_N \in \mathcal{M}$
Output: $p \in \mathcal{M}$ (the anchor point), $v^1, \ldots, v^n \in T_p\mathcal{M}$

$$\min_{p \in \mathcal{M}, \forall j, v^j \in T_p\mathcal{M}} \frac{1}{2} \sum_{i=1}^N d(\mathrm{Exp}(p, Vx_i), y_i)^2, \text{ where } Vx_i := \sum_{j=1}^n v^j x_i^j$$

High Level Summary of Optimization Scheme

Main decision variables to estimate: p and $v^1, \ldots, v^n$

$$E(p, v) = \frac{1}{2}\sum_i \langle \mathrm{Log}(\hat{y}_i, y_i), \mathrm{Log}(\hat{y}_i, y_i)\rangle_{\hat{y}_i} = \frac{1}{2}\sum_i \langle \Gamma_{\hat{y}_i \to p}\mathrm{Log}(\hat{y}_i, y_i), \Gamma_{\hat{y}_i \to p}\mathrm{Log}(\hat{y}_i, y_i)\rangle_p, \quad (1)$$

where $\hat{y}_i = \mathrm{Exp}(p, \sum_j x_i^j v^j)$. The model in tangent space $T_p\mathcal{M}$ is given as

$$\min_{p, v} E^\prime(p^\prime, v) := \min_{p^\prime, v} \frac{1}{2}\sum_i \lVert (\sum_j v^j x_i^j + p^\prime) - y_i^\prime \rVert^2 \quad (2)$$

Procedure: Given a current estimate of p and V, first update p by calculating error in $\hat{y}$ followed by a parallel transport to p. Same procedure for V but slightly more involved (involves a gradient projection type step). All details in the paper.

Optimization Algorithms

Exact Variational Method

Initialize p, v, α, α_{max} and center x
while termination condition **do**
  $p_{new} = \mathrm{Exp}(p, -\alpha \nabla_p E)$
  $V_{new} = \Gamma_{p \to p_{new}}(V - \alpha \nabla_V E)$
  if $E(p_{new}, V_{new}) < E(p, V)$ **then**
    $V \leftarrow V_{new}$ and $P \leftarrow P_{new}$
    $\alpha = \min(2\alpha, \alpha_{max})$
  else
    $\alpha = \alpha/2$
  end if
end while

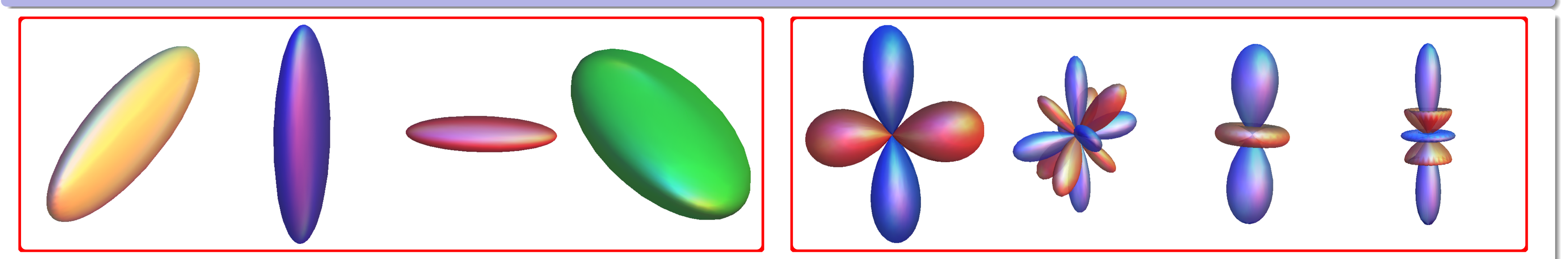
Faster Log-Euclidean Approximation

- Compute $x_i^\prime = x_i - \bar{x}$, ▷ center x
- Compute Karcher mean $\bar{y}$ of $\{y_i\}_{i=1}^n$, ▷ $p^* \approx \bar{y}$
- Compute $y_i^\prime = \mathrm{Log}(\bar{y}, y_i)$
- Parallel transport $y_i^\prime$ from $T_{\bar{y}}\mathcal{M}$ to $T_I\mathcal{M}$
- $\boldsymbol{V} = Y^\prime X^{\prime T}(X^\prime X^{\prime T})^{-1}$, ▷ $\boldsymbol{V}^* \approx \boldsymbol{V}$
- Parallel transport $\boldsymbol{V}$ from $T_I\mathcal{M}$ to $T_{\bar{y}}\mathcal{M}$

Proposition

Let $Y = \{y_1, \ldots, y_N\}$ be a subset of a manifold $\mathcal{M}$. Suppose that Y is in a sufficiently small open cover $\mathcal{B}$ such that the exponential and logarithm maps are bijections. Suppose that all $y \in Y$ are on a curve Ω that is the unique geodesic curve between some y_i and y_j in Y. Then there exists $\bar{y}$ in Ω such that $\sum_{y \in Y} Log_{\bar{y}} y = 0$ (the first order condition for Karcher mean).
Further, if $\bar{y}$ is the unique Karcher mean of $Y \subset \Omega$, and it is obtained in $\mathcal{B}$, then $\bar{y} \in \Omega$. And for some $v \in T_{\bar{y}}\mathcal{M}$ and each y, there exists $x \in \mathbf{R}$ such that $y = Exp(\bar{y}, vx)$.

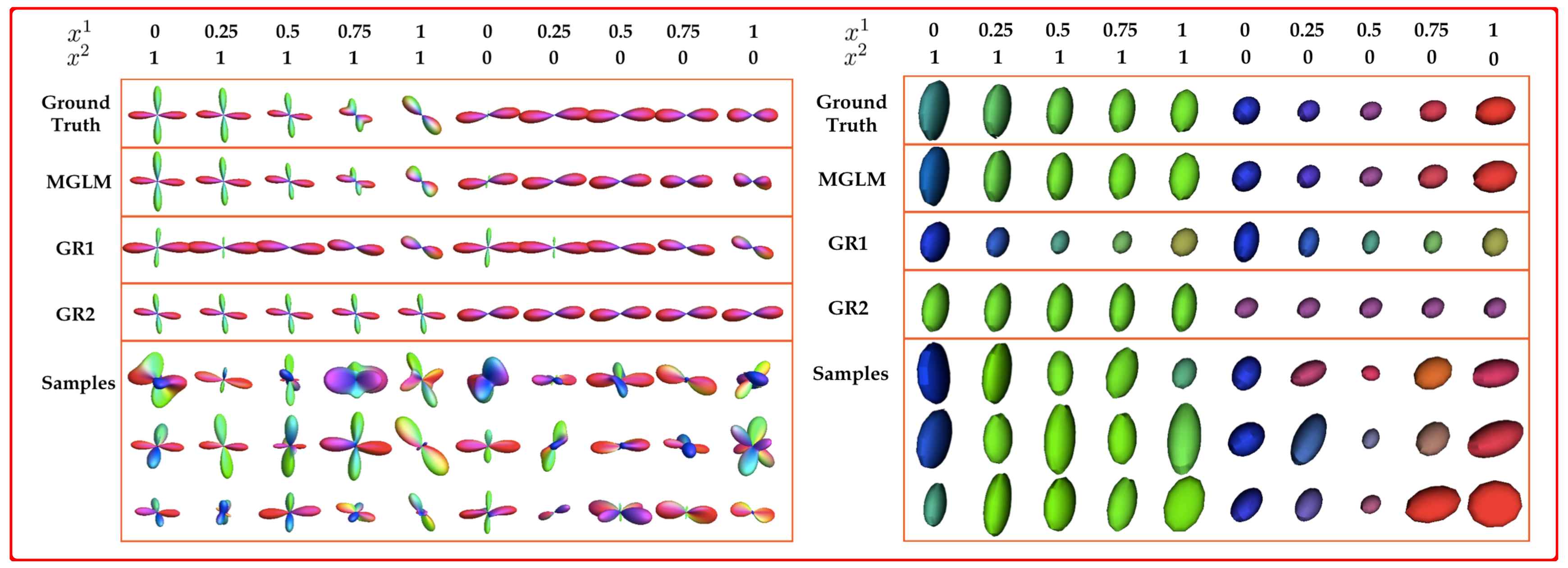
Manifolds in Diffusion Weighted Image Analysis



- Diffusion tensors: Symmetric positive definite (SPD) matrix estimated at each voxel, $\mathrm{SPD}(n)$. Forms a quotient space $\mathrm{GL}(n)/\mathrm{O}(n)$.
- Orientation distributions: Using the square root parameterization, ODFs form a unit Hilbert sphere (S^∞), $\Psi = \{\psi : \mathbb{S}^2 \to \mathbb{R}^+ | \forall s \in \mathbb{S}^2, \psi(s) \geq 0; \int_{s \in \mathbb{S}^2} \psi^2(s) ds = 1\}$.

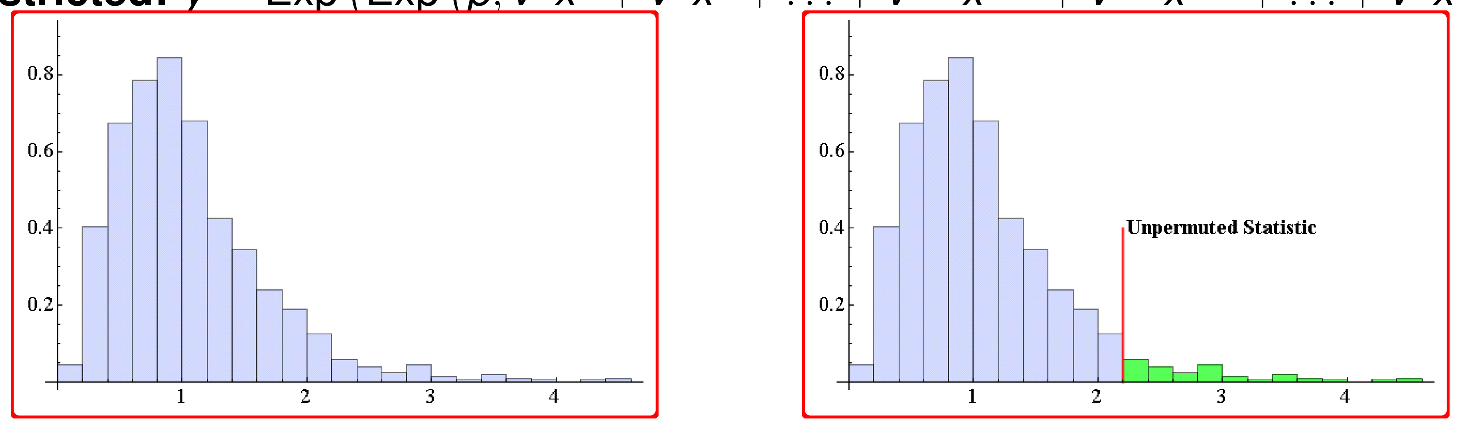
Synthetic Experiments for $f : \mathbf{R}^2 \to \mathcal{M}$

If the response variable y has a dependence on two covariates, x^1 and x^2, a multivariate GLM model is essential; geodesic regression solving for $f : \mathbf{R} \to \mathcal{M}$ will not suffice.



Permutation Testing for Statistical Significance, P-value maps

- **Full:** $y = \mathrm{Exp}\left(\mathrm{Exp}\left(p, v^1x^1 + v^2x^2 + \ldots + v^{i-1}x^{i-1} + v^i x^i + v^{i+1}x^{i+1} + \ldots + v^n x^n\right), \epsilon\right)$
- **Restricted:** $y = \mathrm{Exp}\left(\mathrm{Exp}\left(p, v^1x^1 + v^2x^2 + \ldots + v^{i-1}x^{i-1} + v^{i+1}x^{i+1} + \ldots + v^n x^n\right), \epsilon\right)$



Neuroimaging Experiments: Data and Design

- **Data**: (a) **AD Risk:** Group $\in$ {APOE+, APOE-}, $y \in \mathrm{SPD}(3)$, $N = 343$; (b) **Contemplative Neuroscience (CN):** Group $\in$ {LTM, WLC}, $y \in S^{14}$, $N = 49$.
- **MGLMs**: **Full:** $y = \mathrm{Exp}(p, v^1\mathrm{Group} + v^2\mathrm{Gender} + v^3\mathrm{Age})$; **For group effect:** $y = \mathrm{Exp}(p, v^2\mathrm{Gender} + v^3\mathrm{Age})$; **For age effect:** $y = \mathrm{Exp}(p, v^1\mathrm{Group} + v^2\mathrm{Gender})$

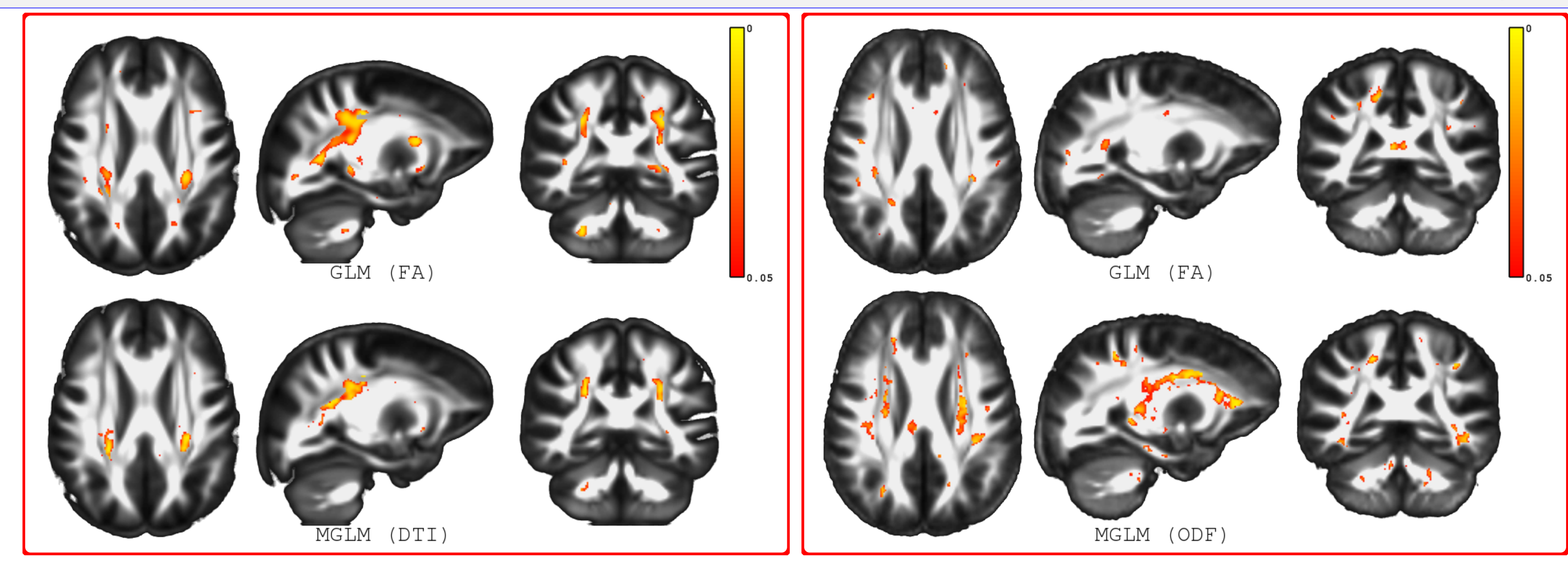


Figure : Effect of Group *controlling for age and gender.* AD study (left); CN study (right).

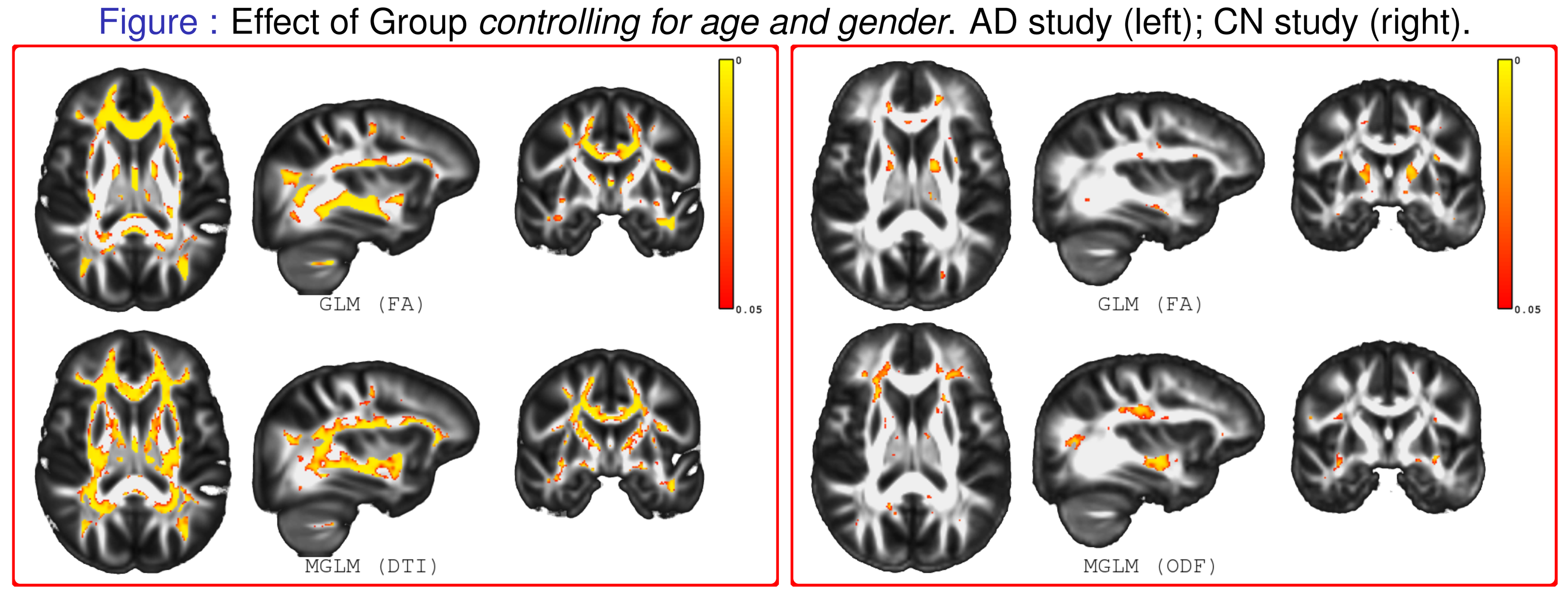


Figure : Effect of Age *controlling for group and gender.* AD study (left). CN study (right).

Computer Vision and Pattern Recognition (CVPR) 2014

Research supported in part by NIH R01 AG040396, NIH R01 AG037639, NSF CAREER award 1252725, NSF RI 1116584, Waisman Center CIHM, Wisconsin ADRC.